



## **Business Intelligence em Dados de Inscrições**

**TIAGO JOSÉ MATOS PEREIRA**

novembro de 2016

# **Business Intelligence em Dados de Inscrições de Alunos em Engenharia de Informática do Instituto Superior de Engenharia do Porto**

**Tiago José Matos Pereira**

**Dissertação para obtenção do Grau de Mestre em Engenharia Informática, Área de Especialização em Sistemas Computacionais**

**Orientador: Paulo Jorge Machado Oliveira**

**Co-orientador: Ângelo Manuel Rego e Silva Martins**

**Júri:**

Presidente:

Vogais:



# Resumo

Esta dissertação aborda a necessidade de melhorar as capacidades analíticas da Direção de curso da Licenciatura de Engenharia de Informática (LEI), do Instituto Superior de Engenharia do Porto (ISEP), sendo neste documento identificado pelo Cliente. Para realizar essa melhoria foi decidido conceber uma solução de *Business Intelligence* à medida, por forma a responder a essas necessidades.

Ao longo do documento são abordados vários conceitos teóricos considerados importantes para a realização da solução, bem como várias ferramentas de ETL (extração, transformação e carregamento), ferramentas de análise de dados e alternativas à solução elaborada. É de salientar que nas ferramentas de ETL é feita uma comparação entre si, utilizando a mesma metodologia nas ferramentas de análises de dados e soluções alternativas.

Após a apresentação do estado da arte, é exposto o desenho e análise da estrutura do Armazém de Dados por forma a dar resposta às várias análises pretendidas pelo cliente e, também, outras que até aqui eram complicadas de realizar. De seguida é descrito todo o detalhe de implementação, orientado às ferramentas utilizadas.

Por fim são demonstrados os resultados obtidos, efetuando uma comparação entre as análises disponibilizadas pelo cliente e as obtidas no armazém de dados, permitindo assim demonstrar a veracidade dos dados obtidos nesta nova solução. Como se pôde constatar, as discrepâncias obtidas foram significativas, levando a uma análise exaustiva para comprovar que a nova solução contém os dados corretos.

Não obstante, esta dissertação tem como principal contributo a melhoria contínua na obtenção de análises que demonstram a performance do curso na instituição. Também, contribui para garantir a credibilidade e uniformização de dados bem como diminuir a margem de erro no tratamento dos mesmos, pois não requer intervenção humana.

**Palavras-chave:** *Business Intelligence*, Armazém de Dados, Análises, Cliente



# Abstract

This master thesis addresses the need to improve the analytical capabilities of the current Direction of Computer Engineering Degree (LEI), in the Instituto Superior de Engenharia do Porto (ISEP), that is identified by the client in this document. To achieve this improvement, it was decided to design a customized Business Intelligence solution to meet those needs.

In this document are presented several theoretical concepts that were considered important for the realization of the solution, as well as several ETL tools (extraction, treatment, and loading), data analysis tools and alternatives to this new solution. It should be noted that in ETL tools is made a comparison between them, using the same methodology in data analysis tools and alternative solutions.

After the state of the art presentation, is exposed the design and analysis of the structure of the Data Warehouse to respond to the various analyzes intended by the client and others that until now were complicated to perform. Next, we describe the entire implementation detail, oriented to the tools used.

Finally, the results obtained are shown, making a comparison between the analyzes made available by the client and those obtained in the data warehouse, thus allowing to demonstrate the veracity of the data obtained in this new solution. As it turned out, the discrepancies obtained were significant, leading to an exhaustive analysis to prove that the new solution contains the correct data.

Nevertheless, it's important to refer that main contribution of this master thesis, is the continuous improvement in the process of obtaining analysis, that demonstrate the performance of the course, in the institution. It also contributes to ensure the credibility and standardization of data in the data warehouse, removing the human intervention in the process of data handling.

**Keywords:** *Business Intelligence, Data Warehouse, Analysis, Client*



# Agradecimentos

A realização desta dissertação de mestrado contou com o apoio de várias pessoas, que me ajudaram, direta ou indiretamente, a alcançar os objetivos a que me propus.

Primeiramente, gostaria de agradecer ao meu Orientador Paulo Jorge Machado Oliveira pelo apoio prestado durante a realização desta dissertação, que contribuiu para o enriquecimento da minha formação académica.

De seguida gostaria de agradecer aos meus pais e namorada por todo o apoio e incentivo disponibilizado durante a realização desta dissertação, pois sem eles nada disto seria possível.

Também, gostaria de agradecer especialmente à minha colega Gisela Couto pelo apoio, amizade, força e dedicação demonstradas ao longo do meu percurso académico.

Por fim, um último agradecimento ao Co-orientador Ângelo Manuel Rego e Silva Martins pelos dados fornecidos e pelo esclarecimento prestado sobre as regras de negócio existentes no curso, que foram fulcrais para a realização da solução.



# Índice

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Enquadramento	1
1.2	Motivação	2
1.3	Objetivos	2
1.4	Organização do Documento	3
<b>2</b>	<b>Análise de Valor</b>	<b>5</b>
<b>3</b>	<b>Estado de Arte</b>	<b>9</b>
3.1	Conceitos Elementares	9
3.1.1	Business Intelligence	9
3.1.2	Armazém de Dados	10
3.1.3	Armazenamento de dados - Armazém de Dados versus Sistema Operacional	11
3.1.4	Data Mart	12
3.2	Modelação Dimensional	12
3.2.1	Tabela de Factos	13
3.2.2	Tabela Dimensão	13
3.2.3	Dimensão conforme	13
3.2.4	Role Playing Dimensions	14
3.2.5	Tabelas de factos agregadas	14
3.2.6	Tipos de modelos dimensionais	14
3.2.7	Slowly Changing Dimensions	17
3.3	Arquiteturas de Armazém de Dados	19
3.3.1	Kimball BI Architecture	19
3.3.2	Inmon BI Architecture	21
3.3.3	Comparação entre arquiteturas	22
3.4	Online Analytical Processing	23
3.5	Extração, Transformação e Carregamento de Dados	24
3.5.1	Processo de ETL	25
3.5.2	Ferramentas	25
3.5.3	Comparação de ferramentas ETL	31
3.6	Ferramentas de apresentação/analíticas	32
3.6.1	Microstrategy	32
3.6.2	Microsoft Power BI	33
3.6.3	Phocas	34
3.6.4	Comparação entre ferramentas analíticas de apresentação de dados	34
3.7	Análise de Mercado	35
3.7.1	Ferramentas de Gestão para Ensino	35
3.7.2	AD na área de ensino/educação	38
3.7.3	Indicadores utilizados no Ensino Superior	40
<b>4</b>	<b>Análise e Desenho</b>	<b>43</b>

4.1	Análise e descrição do problema .....	43
4.2	Definição das análises a implementar .....	44
4.3	Arquitetura da solução .....	46
4.3.1	Fontes de dados .....	47
4.3.2	Modelação do Armazém de Dados .....	48
4.3.3	Staging Area .....	50
4.3.4	Armazém de Dados.....	52
4.3.5	Ferramentas a utilizar para Dashboards e Reporting .....	54
4.3.6	Manutenção de Histórico.....	55
4.4	Avaliação de Resultados.....	56
<b>5</b>	<b>Implementação.....</b>	<b>59</b>
5.1	Ambiente de desenvolvimento e configurações iniciais .....	59
5.1.1	Acesso às Fontes de Dados.....	61
5.1.2	Mecanismo de deploy .....	62
5.1.3	Envio de Emails .....	64
5.1.4	Estrutura do projeto .....	66
5.2	Auditoria de dados .....	68
5.2.1	Mecanismo de extração de dados .....	69
5.2.2	Mecanismo de Rollback em caso de falha.....	70
5.3	Processo de ETL.....	71
5.3.1	Staging Area .....	72
5.3.2	Armazém de Dados.....	74
5.4	Orquestrador do processo de ETL .....	79
5.5	Cubos .....	80
5.5.1	Cubo de Inscrições .....	80
5.5.2	Cubo de Inscrições e Avaliações .....	82
<b>6</b>	<b>Avaliação de resultados .....</b>	<b>85</b>
6.1	Análises no Microsoft Excel .....	86
6.1.1	Número de alunos inscritos por unidade curricular e ano curricular, em cada ano formativo .....	86
6.1.2	Alunos inscritos a cada unidade curricular em cada ano formativo .....	96
6.1.3	Inscrições versus o número de cadeiras em atraso .....	99
6.1.4	Unidades curriculares em atraso, por ano curricular e por ano de entrada do aluno.....	103
6.1.5	Inscrições por ano curricular versus o número de cadeiras aprovadas.....	104
6.2	Análises complementares .....	107
6.2.1	Inscrições a unidades curriculares por tipo de inscrição .....	107
6.2.2	Alunos inscritos por regime em cada turma .....	108
6.2.3	Alunos inscritos por tipo de horário por cada ano curricular .....	109
6.2.4	Alunos inscritos por ano curricular versus alunos inscritos em atraso .....	110
6.2.5	Alunos inscritos por ano curricular versus alunos inscritos em avanço .....	111
6.2.6	Alunos inscritos em cada regime por tipo de inscrição .....	112
6.2.7	Alunos inscritos por ano de entrada em cada regime.....	113

6.3	Avaliação do Sistema .....	114
<b>7</b>	<b>Conclusão .....</b>	<b>117</b>
7.1	Objetivos Alcançados .....	117
7.2	Trabalho Futuro .....	118
<b>8</b>	<b>Anexos.....</b>	<b>119</b>
8.1	Análises do Cliente .....	119



# Lista de Figuras

Figura 1 – Modelo de negócio de Canvas.....	5
Figura 2 – Aplicação e vantagem na utilização de <i>Business Intelligence</i> (Dean, 2015).....	10
Figura 3 – Modelo Dimensional em Estrela (“Fundamentos e Modelagem de Bancos de Dados Multidimensionais,” 2015).....	15
Figura 4 – Modelo Dimensional em Floco de Neve (“Fundamentos e Modelagem de Bancos de Dados Multidimensionais,” 2015).....	16
Figura 5 - Modelo dimensional de constelação de factos (“Fundamentos e Modelagem de Bancos de Dados Multidimensionais,” 2015).....	17
Figura 6 - Exemplo de SCD do tipo dois (Kimball and Ross, 2013).....	18
Figura 7 - SCD do tipo três (Kimball and Ross, 2013).....	18
Figura 8 - Conceito <i>The Back and Front Room</i> (Kimball and Caserta, 2004).....	20
Figura 9 - Bill’s Inmon architecture (Oracle, 2002).....	22
Figura 10 - Processo de ETL (Ferreira et al., 2016).....	25
Figura 11 – Oracle ODI (ARSON Group SAC, n.d.).....	27
Figura 12 - IBM InfoSphere DataStage.....	28
Figura 13 – Exemplo utilizando <i>Microsoft SSIS IDE</i> (Anoop Kumar, 2013).....	29
Figura 14 – Pentaho IDE.....	30
Figura 15 – Exemplo de utilização <i>MicroStrategy Analytics™ desktop</i> (MicroStrategy, 2015) .	33
Figura 16 – Exemplo de utilização <i>Microsoft Power BI</i> .....	34
Figura 17 – EdVantage (SchoolCity Inc., 2015) .	36
Figura 18 - Exemplo de utilização do EdVantage na Escola Elementar de Buffalo (School Buffalo, 2015).....	36
Figura 19 – Exemplo de utilização da aplicação <i>Skedula - School / Teacher Management Portal (CaseNex, 2010)</i> .....	38
Figura 20 - Arquitetura do DW da Universidade de Nova Iorque (New York University, 2014).	39
Figura 21 - Arquitetura da solução.....	46
Figura 22 - Folha de cálculo com informação das turmas inscritas.....	47
Figura 23 - Folha de cálculo com informação de ECTS das disciplinas.....	48
Figura 24 - Folha de cálculo com a informação de inscrição dos alunos (apenas com classificação final).....	48
Figura 25 – Esquema de dados da <i>Staging Area</i> .....	52
Figura 26 – Modelo Dimensional do <i>Data Mart</i> de Inscrições.....	53
Figura 27 - Modelo de Dados Global do AD.....	54
Figura 28 – Power Bi versão <i>Desktop</i> .....	55
Figura 29 - Coleção de Projetos no Team Foundation Server.....	60
Figura 30 - Local de configuração de logins.....	61
Figura 31 - Linked Servers de acesso à fonte de dados.....	62
Figura 32 - Funcionalidade de publicação da solução.....	63
Figura 33 -Exemplo de script de <i>deploy</i> .....	64

Figura 34 - Configurações Email .....	65
Figura 35 - Configuração do perfil de email.....	66
Figura 36 - Estrutura dos projetos .....	67
Figura 37 - Estrutura principal dos projetos de Base Dados .....	67
Figura 38 - Estrutura do Projeto de ETL.....	68
Figura 39 - Projeto <i>Analysis Services</i> .....	68
Figura 40 – Esquema da tabela de auditoria .....	69
Figura 41 - Exemplo de registos inseridos na dimensão aluno .....	70
Figura 42 – Registo auditado para o conjunto.....	70
Figura 43 – Processo de remoção de registos na dimensão aluno .....	71
Figura 44 - Exemplo de <i>Staging Area</i> .....	72
Figura 45 – Estrutura da tabela <i>DQP TipoInscricao</i> .....	73
Figura 46 - Exemplo de Processo Extração e Transformação de Dados .....	73
Figura 47 - Processo carregamento das dimensões .....	75
Figura 48 - Processo carregamento da dimensão Aluno .....	76
Figura 49 - Processo principal do carregamento da tabela de factos de Inscrições .....	77
Figura 50 - Processo de carregamento da tabela de factos de inscrições .....	77
Figura 51 – Modelo ER do Orquestrador.....	79
Figura 52 - Estrutura do Cubo Inscrições.....	81
Figura 53 - Adicionar dimensão manualmente no cubo.....	82
Figura 54 - Esquema do Cubo de Inscrições e Avaliações.....	83
Figura 55 - Inscrições por disciplina em cada ano curricular no ano letivo 2012/2013 (AD) ....	87
Figura 56 - Inscrições por disciplina em cada ano curricular no ano letivo 2012/2013 (Cliente) .....	88
Figura 57 - Comparação entre a análise do AD e do Cliente para o ano letivo 2012/2013 .....	89
Figura 58 - Inscrições efetuadas em PESTI por alunos do 1º ano no ano letivo 2012/2013 .....	90
Figura 59 - Inscrições efetuadas em IARTI por alunos do 1º ano no ano letivo 2012/2013.....	91
Figura 60 - Alunos Inscritos a COMPA no ano letivo 2012/2013 .....	92
Figura 61 - Alunos inscritos a CORGA no ano letivo 2012/2013 .....	92
Figura 62 - Inscrições por disciplina em cada ano curricular no ano letivo 2013/2014 (AD) ....	93
Figura 63 - Inscrições por disciplina em cada ano curricular no ano letivo 2013/2014 (Cliente) .....	94
Figura 64 - Comparação entre a análise do AD e do Cliente para o ano letivo 2013/2014 .....	94
Figura 65 - Alunos Inscritos a LPROG no ano letivo 2013/2014.....	95
Figura 66 - Alunos Inscritos a EAPLI no ano letivo 2013/2014.....	95
Figura 67 - Comparação entre a análise do AD e do Cliente para o ano letivo 2014/2015 .....	96
Figura 68 – Análise proveniente do AD do ano letivo 2012/2013 .....	97
Figura 69 – Análise fornecida pelo cliente referente ao ano letivo 2012/2013.....	97
Figura 70 - Comparação entre análise do AD e do cliente para o ano letivo 2012/2013.....	97
Figura 71 – Análise proveniente do AD do ano letivo 2013/2014 .....	97
Figura 72 - Análise fornecida pelo cliente referente ao ano letivo 2013/2014.....	97
Figura 73 - Comparação entre análise do AD e do cliente para o ano letivo 2013/2014.....	98
Figura 74 - Análise proveniente do AD do ano letivo 2014/2015 .....	98

Figura 75 - Análise fornecida pelo cliente referente ao ano letivo 2013/2014 .....	98
Figura 76 - Comparação entre análise do AD e do cliente para o ano letivo 2014/2015 .....	98
Figura 77 - Análise proveniente do AD do ano letivo 2012/2013 .....	99
Figura 78 - Análise fornecida pelo cliente referente ao ano letivo 2012/2013 .....	99
Figura 79 - Comparação entre análise do AD e do cliente para o ano letivo 2012/2013 .....	100
Figura 80 - Aluno Inscrito a 14 UCs e Reprovado a 0 UCs .....	101
Figura 81 - Aluno Inscrito a 13 UCs e Reprovado a 1 UCs .....	101
Figura 82 – Épocas existentes na fonte de dados .....	102
Figura 83 - Análise comparativa do Ano Letivo 2013/2014 .....	103
Figura 84 - Análise comparativa do Ano Letivo 2014/2015 .....	103
Figura 85 - Análise com base no AD para o ano letivo 2012/2013 .....	104
Figura 86 – Análise do cliente para o ano letivo 2012/2013 .....	104
Figura 87 - Comparação Ano Letivo 2012/2013 .....	104
Figura 88 - Análise com base no AD para o ano letivo 2013/2014 .....	105
Figura 89 - Análise do cliente para o ano letivo 2013/2014 .....	105
Figura 90 - Comparação Ano Letivo 2013/2014 .....	105
Figura 91 - Alunos inscritos a 12 UCs e aprovados a 8, no ano letivo 2013/2014 .....	106
Figura 92 – Inscrições a unidades curriculares por tipo de inscrição para o ano letivo 2013/2014 .....	108
Figura 93 – Alunos inscritos por regime em cada turma, no ano letivo 2013/2014 .....	109
Figura 94 – Alunos inscritos por tipo de horário em cada ano curricular para o ano letivo 2013/2014 .....	110
Figura 95 – Alunos inscritos versus inscrições realizadas em atraso, no ano letivo 2013/2014 .....	111
Figura 96 - Alunos inscritos versus inscrições realizadas em avanço, no ano letivo 2014/2015 .....	112
Figura 97 – Alunos inscritos em cada regime por tipo de inscrição, no ano letivo 2014/2015 .....	113
Figura 98 - Alunos inscritos por ano de entrada em cada regime, no ano letivo 2014/2015 .....	114
Figura 99 - Logs do Processo de ETL e atualização dos Cubos .....	115
Figura 100 – Alunos Inscritos a cada disciplina em cada ano curricular (análise 1) .....	119
Figura 101 - Número de inscrições efetuadas por alunos em cada ano curricular (análise 2) .....	120
Figura 102 - Inscrições versus Unidades Curriculares Reprovadas (análise 3) .....	120
Figura 103 - Inscrições por ano letivo versus Unidades Curriculares Reprovadas (análise 4) .....	120
Figura 104 - Inscrições versus unidades curriculares aprovadas (análise 5) .....	121



# Lista de Tabelas

Tabela 1 - Comparação entre arquiteturas <i>Kimball</i> e <i>Inmon</i> (Inmon, 2002);(Kimball and Caserta, 2004);(Kimball and Ross, 2013);(Sansu George, 2012) .....	22
Tabela 2 - OLAP versus OLTP (Atanazio, 2013) .....	24
Tabela 3 - Comparação de ferramentas ETL (“ETL Tools comparison,” 2016),(McBurney, 2007) . .....	31
Tabela 4 - Comparação de ferramentas analíticas. ....	35
Tabela 5 – Listagem de alguns indicadores retirados da FenProf (FenProf, 2012). ....	41
Tabela 6 - Matriz entre atributos de dimensões e métricas a calcular .....	49
Tabela 7 - Software Instalado.....	59



# Acrónimos e Símbolos

## Lista de Acrónimos

<b>BI</b>	Business Intelligence
<b>AD</b>	Armazém de Dados
<b>OLTP</b>	Online Transaction Processing System
<b>OLAP</b>	Online Analytical Processing
<b>DBMS</b>	Database Management Systems
<b>SQL</b>	Structured Query Language
<b>CRUD</b>	Create Read Update and Delete
<b>ER</b>	Entity-Relationship modeling
<b>SCD</b>	Slowly Changing Dimensions
<b>ETL</b>	Extract, Transform, Load
<b>SGBD</b>	Sistema de Gestão de Base de Dados
<b>KPI</b>	Key Performance Indicator
<b>ECTS</b>	European Credit Transfer and Accumulation System
<b>CIF</b>	Corporate Information Factory
<b>LEI-ISEP</b>	Licenciatura de Engenharia de Informática
<b>SSIS</b>	SQL Server Integration Services
<b>SSDT</b>	SQL Server Data Tools
<b>UC</b>	Unidade Curricular



# 1 Introdução

No capítulo corrente será abordado o tema desta dissertação, no âmbito da unidade curricular Tese de Mestrado de Engenharia de Informática, lecionada no Instituto Superior de Engenharia do Porto.

Em primeiro lugar será feito, de uma forma sucinta, o enquadramento desta dissertação com o problema atual. De seguida é apresentada a motivação para a realização da mesma, bem como os objetivos gerais que se pretendem alcançar.

Por fim será apresentada a estrutura deste documento, por forma a indicar o conteúdo de cada capítulo.

## 1.1 Enquadramento

Esta dissertação nasce da necessidade de melhorar as capacidades analíticas da Direção de curso da Licenciatura de Engenharia de Informática (LEI), do Instituto Superior de Engenharia do Porto (ISEP), sendo neste documento identificado pelo Cliente.

A cada ano letivo os alunos efetuam novas inscrições em unidades curriculares (UCs) que já frequentaram anteriormente ou irão frequentar, pela primeira vez, no curso da LEI, no ISEP. Posto isto, a cada ano letivo que passa a informação histórica de inscrições dos alunos é maior, sendo que esta é gerida e armazenada pelo portal da instituição de acordo com um modelo relacional cujo esquema é desconhecido.

O acesso ao sistema operacional para obtenção dos dados necessários não foi concedido, sendo os dados necessários fornecidos pelo Diretor de Curso da LEI, em folhas de cálculo.

Por forma a obter os indicadores de desempenho do curso, o cliente construiu com base nas folhas de cálculo as várias análises pretendidas, como por exemplo, o número de alunos inscritos às várias UCs (Unidades Curriculares) de acordo com o ano curricular em que se encontra. Estas análises tem uma grande importância para a instituição pois são as que definem

o estatuto do curso na instituição e a nível nacional. Também, tem uma grande importância nas creditações que o curso obtém ao longo do tempo.

Pretende-se com esta dissertação demonstrar que é possível obter as mesmas análises de uma forma mais flexível, rápida e fiável. Com a flexibilidade e rapidez obtida será potenciada a identificação de determinados comportamentos ou padrões, se estes fogem do expectável ou não e tomar ações de acordo com os objetivos da instituição.

Para além destas vantagens, acresce a possibilidade de consultar os dados de uma forma mais integrada e criar novas análises que até aí eram complexas de realizar, permitindo extrair novas conclusões.

## 1.2 Motivação

No contexto desta dissertação pretende-se desenvolver uma solução que responda a análises sobre as inscrições dos alunos, de acordo com o modelo de negócio da instituição. As inscrições dos alunos contêm informação relevante sobre quais as disciplinas a que um aluno se inscreveu, em que turma ficou colocado, tipo de inscrição que efetuou, num determinado curso e num determinado ano letivo. Ao cruzar esta informação com a informação das avaliações permite obter os vários indicadores que o cliente pretende e assim visualizar o desempenho da instituição.

Com esta solução pretende-se obter as análises já existentes, a capacidade de criar novas de uma forma mais flexível, confiável e fácil de operar. Pretende-se ainda que o tempo entre a obtenção dos dados e a geração dos relatórios de desempenho seja diminuído, devido à utilização de processos de tratamento de dados e criação de análises automatizadas. Estes processos automáticos garantem uma maior precisão dos dados minimizando os erros de intervenção humana.

## 1.3 Objetivos

De acordo com o problema e motivação mencionados anteriormente, pretende-se então:

- Desenvolver um Armazém de Dados (AD) que modele a informação relativa a inscrições, de acordo com as especificidades da instituição e do curso
- Elaborar um processo de extração, tratamento e carregamento adequados, por forma a disponibilizar informação mais precisa, mais rápida e confiável
- Assegurar a retrocompatibilidade das análises feitas pelos utilizadores
- Assegurar a qualidade de dados presentes no AD
- Disponibilizar em modo *self-service* a criação de novas análises

## 1.4 Organização do Documento

Primeiramente, este documento começa pela introdução onde é feita uma breve descrição do contexto e do problema que deu origem a esta dissertação. De seguida no Capítulo 2, é apresentada a análise de valor onde é descrito, detalhadamente, o valor que este projeto cria ao ser realizado e o modelo de negócio devidamente explicitado.

Por conseguinte no Capítulo 3 Estado da Arte, são apresentados os vários conceitos da área, diferentes ferramentas de processamento de dados, possíveis ferramentas analíticas para construir e visualizar análises, soluções existentes no mercado e indicadores utilizados no Ensino Superior Português.

Por forma a descrever a solução numa ótica mais tecnológica, no Capítulo 4 é apresentada a componente de arquitetura da solução de acordo com a descrição do problema. É de salientar que este documento encontra-se estruturado de acordo com as regras de escrita técnico-científicas (Pereira, 2016).

De seguida, no Capítulo 5 é apresentado o detalhe técnico dos processos utilizados na implementação do Armazém de Dados (AD), bem como metodologias de consistência de dados caso ocorra algum erro durante o processo de Extração, Transformação e Carregamento (ETL).

De modo a comprovar a veracidade do Armazém de Dados, no Capítulo 6 são apresentadas as várias análises fornecidas pelo cliente, as análises reproduzidas no próprio AD bem como as diferenças entre ambas. Adicionalmente, são apresentadas análises complementares por forma a enriquecer as análises disponibilizadas pelo AD.

Por fim, no Capítulo 7 é apresentada a conclusão sobre o trabalho realizado e com base na mesma, é apresentado o possível trabalho futuro a realizar.



## 2 Análise de Valor

A Licenciatura em Engenharia de Informática (LEI), no Instituto Superior de Engenharia do Porto, possui um ciclo de programa denominado por processo de Bolonha, com três anos e cento e oitenta créditos que foi adotado por Portugal em 2006/2007 (Martins and Costa, 2010). O programa de três anos, está dividido em seis semestres, baseia-se na *ACM Computing Curricula* e por forma a visualizar os resultados desta implementação é necessário analisar vários indicadores de desempenho (Martins and Costa, 2010).

O cliente (Direção de Curso da LEI), com este projeto, pretende um ambiente de armazenamento de dados orientado ao negócio, por forma a obter resultados com uma maior fiabilidade, qualidade e no menor tempo possível. Na Figura 1, encontra-se o modelo de negócio de Canvas que descreve de uma forma sucinta o negócio.

Parceiros Chave	Atividades Chave	Proposta de Valor	Relacionamento com o Cliente	Segmentos de Cliente
Técnicos do departamento, no que toca a manutenção do hardware necessário (servidores)	Manutenção dos Servidores Assistência técnica Suporte ao que foi desenvolvido  Manutenção do data warehouse (carregamento de nova informação, atualização de dados)  Geração de reports <b>Recursos Chave</b> Alunos: responsáveis pelo desenvolvimento de novos reports/funcionalidades Know how: propriedade intelectual criada  Servidores: permitir centralizar o sistema, permitir o trabalho por mais do que um aluno em paralelo.	Armazém de dados com informação carregada centralizada  Informação segmentada Aumento de Performance Solução escalável Probabilidade de erro de cálculos associados a avaliações dos alunos reduzidos  Acessibilidade na criação de análises baseadas no armazém de dados Redução de custos na medida em que o utilizador consegue poupar tempo na criação e estudo dos resultados inerentes às análises que pretende	Assistência na realização de novos reports e novas funcionalidades   <b>Canais</b> Divulgação do armazém de dados criado aos docentes responsáveis pelo projeto  Divulgação do tipo de sistema criado a outros possíveis departamentos do ISEP interessados.	Cientes diretos/finais - Docentes
<b>Estruturas de Custo</b>		<b>Fontes de Receita</b>		
Manutenção do hardware (servidores e terminais onde o sistema vai estar disponível para utilização por parte dos docentes)		Disponibilidade para criação de novos reports e funcionalidades Disponibilidade para resolução de eventuais problemas, englobando resolução de problemas no data warehouse, no processo de carregamento de dados, nos possíveis reports gerados e a nível de manutenção de servidores.		

Figura 1 – Modelo de negócio de Canvas.

Em primeiro lugar, tem-se como parceiros principais os técnicos informáticos do departamento, que auxiliam na manutenção do *hardware* dos servidores onde a solução está alocada. Como indicado nas atividades-chave, é feita uma manutenção dos mesmos servidores, mas numa ótica de manutenção da solução implementada/alojada nos mesmos. Relativamente às atividades chave que foram definidas, a assistência técnica, o suporte ao que foi desenvolvido neste projeto, carregamento de nova informação e a geração/atualização de relatórios, de modo à solução estar atualizada e refletir os dados mais atuais.

De acordo com o contexto desta dissertação, os recursos-chave são os alunos responsáveis por desenvolver novas funcionalidades/gerar relatórios, o *know-how* criado ao desenvolver a solução e os servidores que alocam a mesma. Quanto à proposta de valor é referida a informação que se encontra segmentada no próprio armazém de dados, o aumento de performance na obtenção de resultados, a escalabilidade e a redução de custos na medida em

que não é necessário esperar muito tempo para se obter a informação. Também, no que toca à relação com os clientes é disponibilizado uma assistência na realização de novos relatórios ou novas funcionalidades no AD, por forma a garantir que este corresponde às necessidades dos utilizadores desta solução.

Por fim, como canais de divulgação é apresentado a divulgação do armazém de dados criado para o cliente, que também é passível de ser usado pelos docentes do departamento. Devido à sua flexibilidade este pode ser adaptado para ser utilizado por outras instituições de ensino superior. Estes mesmos docentes e o diretor de curso correspondem à vertente de segmentos de clientes presentes no modelo.

*“A análise de valor tem como principal objetivo aumentar o valor de um item ou serviço com o preço mais baixo possível, sem necessitar de sacrificar a qualidade” (Nicola, 2016).*

Depois de definido o negócio, torna-se importante definir uma proposta de valor com o objetivo de desenvolver um produto que se encontre dentro das expectativas do cliente, utilizando os recursos de forma adequada sem que a qualidade do mesmo seja posta em causa e que responda às necessidades anteriormente identificadas (neste contexto específico). Depois do consumo do produto, ainda é possível analisar o valor percebido pela perspetiva do cliente (tendo em conta o valor total esperado do produto e o custo total), para que se consiga assim concluir se a estratégia utilizada garantiu o sucesso.

O valor criado por soluções (neste caso específico de *software*) pode ser modelado recorrendo ao modelo conceptual de decomposição de valor para o cliente. Este modelo baseia-se na combinação do conceito de formas de valor e posições temporais, no conceito de uma Rede de Valor. Esta é uma rede de troca de resultados tangíveis e intangíveis sobre determinadas regras, criando valor ativo na empresa. Também se baseia no conceito de ativos endógenos e exógenos e no conceito de benefícios percebidos (PBI) versus os sacrifícios (PSI). Ao combinar os conceitos deste modelo obtém-se a formalização de um modelo quantitativo que utiliza técnicas que resultam da área multicritério de tomada de decisão (Nicola et al., 2014).

O modelo Conceptual de decomposição de valor para o cliente contém, como referido anteriormente, formas de valor. Essas formas de valor são o *Net Value to Customer*, o valor racional, valor de vendas, valor de marketing e valor derivado para o cliente. O valor racional para o cliente significa a diferença em relação ao preço de objetivo. De seguida o valor de vendas para o cliente, como a própria forma de valor indica e sendo preocupação principal é o preço, em que o valor de *marketing* para o cliente baseia-se nos atributos percebidos do produto. Por fim, o valor derivado consiste na experiência dos utilizadores ao longo do tempo e o *Net Value to Customer* consiste no balanço dos benefícios e sacrifícios de modo a disponibilizar o melhor ou o pior valor para o cliente (Nicola et al., 2014).

Quanto às posições temporais estas podem ser divididas em quatro partes: *Ex Ante Value to Customer* (Pré-compra), *Transaction Value to Customer* (Durante a transação), *Ex Post Value* (pós-compra) *to Customer* e *Disposition Value To Customer* (Experiência de utilização). A primeira forma de valor é baseada numa fase de tentar prever como as pessoas entendem os

seus serviços. Já a segunda forma de valor implica um sentido de valor para o cliente experiente numa ótica comercial. Por conseguinte, a terceira forma de valor fornece resultados de experiências realizadas com base em escolhas dos clientes e fornecedores. Por fim, a última forma de valor referida é uma fase que reflete o ponto de descarte ou de venda (Nicola et al., 2014)

Numa primeira fase, o modelo conceptual de decomposição de valor para o cliente tem como objetivo entender como o valor possa ser dividido em componentes mais simples, podendo assim integrar o valor percebido. A criação de uma rede de valor fornece a identificação de produtos tangíveis e intangíveis trocados com o cliente, bem como os ativos (endógenos e exógenos) construídos e/ou utilizados na prestação dessa integração. Desta maneira, fica-se a perceber a relevância dos ativos envolvidos e como se relacionam com os Benefícios Percebidos/Sacrifícios (Nicola et al., 2014).

Na segunda fase, o objetivo é obter mais informações de um período particular de tempo relacionado com a percepção de benefícios e sacrifícios de uma empresa. Nesta fase e na seguinte, tomam-se as conclusões da análise anterior bem como os ativos mais relevantes para avaliar a forma como o cliente percebe a proposição de valor da empresa. Também, é avaliado nesta fase a relevância de cada produto a cada PBI/PSI usando a escala de *Saaty* (Nicola et al., 2014).

Por fim, numa terceira fase é avaliado a proposição de valor da empresa e dos seus ativos de apoio. Esta análise combina os dois fluxos descritos, na perspetiva da empresa de um lado e a perspetiva do cliente do outro. Para estes fluxos serem analisados recorre-se a modelos quantitativos, como por exemplo o método da teoria dos jogos, o método multicritério AHP (*Analytic Hierarchy Process*) e o método *Fuzzy*. O método da teoria dos jogos é utilizado quando no processo de decisão há mais do que um interveniente, ajuda analisar situações de forma mais racional e formular uma alternativa aceitável de acordo com as consequências. Quanto ao método multicritério AHP este tem como objetivo ser uma *framework* compreensiva e racional no que toca à estruturação do problema, representação e quantificação dos seus elementos e relacioná-los com os objetivos gerais para avaliar as soluções alternativas. Por último, o método *Fuzzy* é um método utilizado quando se pretende lidar com o pensamento humano (Nicola et al., 2014).

No contexto desta dissertação este modelo não se aplica pois baseia-se em critérios de vendas e de marketing que não vão de encontro com o objetivo desta solução. Esta contém um valor racional, não monetário e derivado do consumo por parte dos utilizadores que pretendam analisar este tipo de informação (Nicola et al., 2014).

Por forma a chegarmos à solução pretendida, foram feitas negociações para o que seria implementado. Como (Filzmoser and Vetschera, 2008) explicitam:

*“[...] negotiations are dynamic processes in which the parties involved communicate to exchange offers, make concessions, raise threats, or otherwise influence each other in order to reach an agreement”.*

No contexto deste projeto, pode-se classificar que teve um cenário de negociação do tipo “*WIN-WIN*”, ou seja, ambas as partes saem beneficiadas pois o autor desta dissertação sai beneficiado pelo conhecimento adquirido sobre o modelo negócio, bem como a aquisição e melhoria das qualidades técnicas que foram absorvidas durante a elaboração da solução. Por outro lado, o cliente, que neste caso é diretor de curso da LEI-ISEP, sai beneficiado pois obtém uma solução com melhor performance, mais adequada para a análise de indicadores que pretende, e possui a possibilidade de efetuar outras análises mais sofisticadas que até agora seriam bastante complexas. Com esta solução é-lhes permitido a construção de novos indicadores/visualização de existentes, de acordo com as regras de avaliação existentes do curso, de forma fácil e rápida.

## 3 Estado de Arte

Este capítulo refere-se ao estado da arte, a base que sustenta todo o trabalho desenvolvido. No decorrer deste capítulo será apresentado todo o estudo efetuado desde o conhecimento de arquiteturas e técnicas de modelação, incluindo ferramentas para recolha e análise de dados. Os critérios utilizados para a escolha têm como base o suporte oferecido, *standards* seguidos, bem como funcionalidades oferecidas, documentação e licenças. Performance e estabilidade da ferramenta não serão contabilizados pois isso requeria a implementação da solução pretendida nas várias ferramentas que serão expostas e para além de não ser o âmbito desta dissertação, estas necessitam de ter licenças para funcionarem legalmente.

De notar que este capítulo foi desenvolvido com a colaboração da colega Gisela Couto, que se encontra a desenvolver um *data mart* relativo aos dados de avaliações de alunos.

### 3.1 Conceitos Elementares

Neste ponto são apresentados, entre outros, vários conceitos relacionados com AD, o próprio BI, modelação dimensional e as diferentes arquiteturas que este pode ter.

#### 3.1.1 Business Intelligence

“[...] o negócio é um conjunto de atividades realizadas por qualquer fim, seja ele ciência, tecnologia, comércio, indústria, lei, governo, defesa, etc. [...] A noção de inteligência é definida também aqui, num sentido mais geral, como “a capacidade de apreender a inter-relação dos factos apresentados de modo a orientar a ação para o objetivo desejado.”” (Traduzido de (Luhn, 1958)).

Podemos definir BI como um conjunto de técnicas e ferramentas usadas sobre grandes volumes de dados com o objetivo de obter conhecimento sobre o negócio em questão, através de

análises históricas e correntes sobre os dados. Atualmente, existem diversas metodologias que permitem recolher dados de sistemas internos/externos a uma organização para posteriormente armazená-los, prepará-los para análise e assim criar relatórios capazes de evidenciar ao utilizador os principais indicadores de que este pretende, sem que conheça toda a arquitetura técnica que tem por base todo este mecanismo (Rouse, 2015).

Este tipo de análises é construído com base nos dados previamente carregados no sistema (designado por AD), onde o utilizador é livre de definir o tipo de métricas que pretende, bem como o tipo de informação a ter em conta.

Uma das grandes vantagens deste processo é a rapidez com que os resultados são calculados e facilmente partilhados. Estes sistemas estão preparados para receber e processar grandes quantidades de dados, tornando por si só a tomada de decisões mais facilitada e rápida na medida em que o utilizador escolhe que tipo de análises pretende fazer aos dados. Depois de obtidos os dados necessários e de tomadas as decisões, torna-se possível ter uma visão concreta e fiável sobre os seus fundamentos, na medida em que todos os dados que entram no sistema sofreram um tratamento prévio. Para responsáveis de empresas pode ser considerada uma ferramenta muito útil, permitindo uma rápida evolução não só a nível de tomada de decisões futuras como foi descrito, mas também a nível de análise do comportamento dos seus potenciais concorrentes e clientes, identificando possíveis melhorias nos produtos e segmentos de mercado que ainda não foram explorados pela mesma. Na Figura 2 é apresentado um resumo dos benefícios descritos, que facilitam nas operações do quotidiano.



Figura 2 – Aplicação e vantagem na utilização de *Business Intelligence* (Dean, 2015).

### 3.1.2 Armazém de Dados

Quando se pretende armazenar um grande volume de dados centralizado e efetuar as respetivas análises é necessário um AD. É um sistema computacional capaz de armazenar

grandes quantidades de dados e manter a performance na consulta devido à sua estrutura desnormalizada com poucas dependências entre domínios de informação. Armazena todo o conjunto de informação em modelos multidimensionais/dimensionais, constituídos por dimensões e tabelas de factos. Neste tipo de modelo, as dimensões armazenam todos os dados que pertencem ao seu domínio. Como exemplo deste tipo de armazenamento temos o caso do aluno, onde seria especificada uma estrutura capaz de armazenar os dados básicos de cada aluno existente. No que toca às tabelas de factos, estas são responsáveis por relacionar os registos existentes em cada domínio, armazenando acontecimentos e KPIs previamente determinados na fase de análise do sistema (Ballard et al. 2006).

No que toca às características de um AD este possui quatro características bastante particulares (Sansu George, 2012):

- **Orientado ao assunto:** toda a modelação realizada é de acordo com os vários temas existentes nas organizações, compartimentando assim toda a informação neste contida.
- **Variável no tempo:** é muito importante guardar as várias alterações que ocorrem nos dados ao longo do tempo por forma a permitir histórico.
- **Não volátil:** por forma a obter-se histórico dos dados inseridos num AD, estes nunca são eliminados ou reescritos
- **Integrado:** pois contém os dados dos vários sistemas operacionais que uma organização possui

Desta forma é possível obter informação consistente, integrada e criar análises com base no histórico existente, permitindo assim obter uma visão geral do desempenho da organização.

### 3.1.3 Armazenamento de dados – Armazém de Dados versus Sistema Operacional

Num armazém de dados após a extração da informação, é necessário armazená-la em repositórios de dados especializados, permitindo que possam ser preservados e consumidos à posteriori. Dependendo do tipo de características existentes e do contexto em que se insere, a estrutura do armazenamento pode ser distinta. Generalizando, existem dois tipos de sistemas: sistemas orientados à transação (sistemas OLTP) e sistemas orientados a análise/assunto (sistemas OLAP) (datawarehouse4u 2009).

Os sistemas operacionais têm como base uma arquitetura orientada à transação, isto é, tem como objetivo registar todas as transações efetuadas num determinado momento e em cada domínio. Os AD (armazéns de dados) possuem uma estrutura relacional, onde os dados são armazenados em tabelas de acordo com o contexto e relacionados entre os restantes artefactos existentes, criando dependências em rede. Dado o número máximo de dependências que pode existir em determinado contexto e a quantidade de dados adjacente, este tipo de sistemas pode perder a performance na consulta desses mesmos dados (Editorial Team+2007).

Como já foi enunciado, cada um dos tipos de armazenamento enunciados possuem as suas especificidades. Devido ao processo prévio de limpeza e tratamento de dados (ETL), o AD permite despistar em grande alcance as inconsistências, bem como definir qual a melhor resolução do problema perante o caso em questão (Tech-FAQ 2013). No sistema operacional, o utilizador é que necessita de tomar a iniciativa de limpeza das inconsistências de forma

manual. Não é um processo estruturado que faça parte do carregamento propriamente dito, fica ao critério do criador do sistema.

O acesso à informação nas análises pode atingir tempos de resposta muito expectantes, ao contrário do que acontece com os sistemas operacionais. No AD, as consultas efetuadas pelo consumidor não são efetuadas diretamente sobre o sistema de armazenamento. Depois do sistema se encontrar limpo e carregado, é criada uma camada de abstração para que todas as consultas incidam sobre a camada. No caso dos sistemas operacionais, todas as interrogações e consultas que se efetuam ao sistema, são inteiramente efetuados diretamente ao sistema.

De notar que, apesar da execução de análises sobre o AD obter bons resultados, a desvantagem aparece no carregamento/tratamento de dados para este sistema e respetiva manutenção, sendo um processo lento e trabalhoso (Tech-FAQ 2013). O AD permite que o carregamento de dados seja efetuado a partir de fontes de dados de tipos diferentes nomeadamente ficheiros Excel, ficheiros de texto, bases de dados, etc. No entanto e tendo em conta as desvantagens enunciadas relativamente aos sistemas operacionais, é de notar que estes podem ser usados como fontes de dados para carregar o AD (Ballard et al. 2006).

Para implementar um sistema de armazenamento, é necessário ter em conta as necessidades do meio e o benefício/custo que cada tipo de solução terá e escolher qual a que melhor se ajusta.

#### **3.1.4 Data Mart**

Um Data Mart é um sistema de divisão lógica mais pequena que fornece suporte a tomada de decisões para uma determinada área de negócio em específico (por exemplo: Vendas, Marketing, etc.). O AD pode ser dividido/composto por várias áreas deste tipo, tornando-se mais fácil de gerir e manter na medida em que as operações necessárias a serem efetuadas apenas incidem naquele domínio em específico, mantendo os restantes operacionais e com impacto reduzido (Ballard et al., 2006).

Relativamente à forma como os dados são carregados nestas áreas, podem categorizar-se em dependentes ou independentes. Este fator depende do tipo de fonte de dados utilizada desde um outro AD ou bases de dados operacionais, serviços ou ficheiros de texto (Oracle, 2007).

### **3.2 Modelação Dimensional**

Atualmente os modelos multidimensionais/dimensionais constituem uma base sólida de armazenamento e de gestão de dados nas soluções de BI (Elias, 2015). Estes modelos permitem a definição do relacionamento dos dados, concebendo um suporte a consultas em todas as vertentes de negócio, bem como a extração de detalhes sobre esses mesmos dados. Desta forma torna-se mais fácil visualizar e relacionar os dados de uma organização, de uma forma mais intuitiva e eficaz.

Nos tópicos seguintes vão ser abordados os tipos de tabelas capazes de armazenar informação neste tipo de sistema, bem como, as possíveis arquiteturas e modelos existentes.

### 3.2.1 Tabela de Factos

“Uma tabela de factos contém métricas numéricas produzidas por um evento de medição operacional no mundo real. Com o menor nível de granularidade, uma linha da tabela de factos corresponde ao evento de medição e vice-versa. Assim, o *design* fundamental de uma tabela de factos é inteiramente baseado numa atividade física e não influenciada pelos eventuais relatórios” ( traduzido de Kimball and Ross, 2013).

Esta tabela, como o próprio nome indica, contém os eventos que originam um ou vários factos que ocorreram num determinado tipo de negócio, podendo daí retirar valor. Esta possui métricas, valores, por forma a indicar a evolução do negócio (Henrique, 2012). Por fim, é importante referir que esta tabela possui uma quantidade considerável de factos segundo uma granularidade bem definida, sendo que a granularidade consiste no nível de detalhe de informação definido para ser armazenado na tabela.

### 3.2.2 Tabela Dimensão

“As dimensões disponibilizam o contexto de quem, o quê, onde, quando, porquê e como o contexto em torno de um evento do processo de negócio. As tabelas dimensão contém os atributos descritivos utilizados pelas aplicações de BI por forma a filtrar e agrupar os factos. Com a granularidade dos factos bem em mente, todas as possíveis dimensões podem ser identificadas. Sempre que possível, uma dimensão deve conter apenas um registo quando este é associado a uma linha da tabela de factos” ( traduzido de Kimball and Ross, 2013).

Como referido anteriormente, uma tabela dimensão consiste em adicionar um contexto aos factos inseridos numa tabela de factos. Este tipo de tabela contém uma relação com a tabela de factos através de uma chave natural, denominada chave substituta (*surrogate key*). Esta chave natural identifica inequivocamente um registo na tabela e serve para relacionar a dimensão com a tabela de factos como anteriormente referido. Uma vantagem de usar este tipo de chave é o aumento da performance no que toca ao relacionamento entre o registo da dimensão, com os registos da (s) tabela (s) no qual se insere este tipo de contexto.

### 3.2.3 Dimensão conforme

“Tabelas Dimensão são conformes quando os atributos em tabelas de dimensões distintas têm os mesmos nomes de colunas e conteúdo de domínio. Informação contida em tabelas de factos separadas pode ser combinada num único relatório utilizando atributos conformes de uma dimensão que estão associados a cada tabela de factos.” ( traduzido de Kimball and Ross, 2013).

De acordo com a definição de dimensão conforme é exequível realizar relatórios detalhados, porque é possível obter numa só linha os resultados de várias tabelas de factos, conseguindo assim a integração de dados que é uma das características que um AD possui.

Desta forma, estas dimensões disponibilizam consistência analítica e impacto reduzido de custos de desenvolvimento porque a roda não precisa de ser novamente recriada, ou seja, os dados encontram-se integrados.

### 3.2.4 Role Playing Dimensions

“Uma dimensão física pode ser referenciada várias vezes numa tabela de factos em que cada referência ligada representa logicamente uma regra distinta para a dimensão” (traduzido de Kimball and Ross, 2013).

Um exemplo comum de *Role Playing Dimensions* é quando uma tabela de factos possui vários campos que contém datas e cada campo presente possui uma chave estrangeira para a dimensão data. Assim, cada ligação representa uma diferente visão para a dimensão data pelo qual se denomina regra e cada campo representa um contexto diferente (Kimball and Ross, 2013).

### 3.2.5 Tabelas de factos agregadas

“Tabelas de factos agregadas são simples *rollups* numéricos de dados de tabelas de factos atómicas, construídas exclusivamente para acelerar a consulta aos dados. Estas tabelas de factos devem estar disponíveis na camada de BI ao mesmo tempo que as tabelas de factos atómicas para que as ferramentas de BI escolham em tempo real o nível de agregação apropriado ao realizar consultas (*query's*).” (traduzido de Kimball and Ross, 2013).

Posto isto, é importante realçar que as tabelas de factos agregadas também contêm chaves estrangeiras para as respetivas dimensões conformes e, contêm factos agregados criados pela soma das medidas de tabelas de factos atómicas. Também é importante referir que ao criar este tipo de tabelas é importante criar índices para uma maior performance na obtenção dos dados (traduzido de Kimball and Ross, 2013).

### 3.2.6 Tipos de modelos dimensionais

#### 3.2.6.1 Modelo em Estrela

Uma das características que distingue um sistema operacional de um AD é a sua modelação. Por norma, um AD assenta num modelo dimensional que consiste num modelo tipicamente em estrela, denominado *star-schema* em que a informação é relacionada de forma que pode ser representada num cubo. Desta forma, permite visualizar a informação de uma forma simples e relacioná-la com diferentes domínios (Moreira, 2006). Também é de referir que este modelo

dimensional é composto por uma tabela ou mais tabelas de factos relacionadas com diferentes dimensões, como se pode verificar na Figura 3.

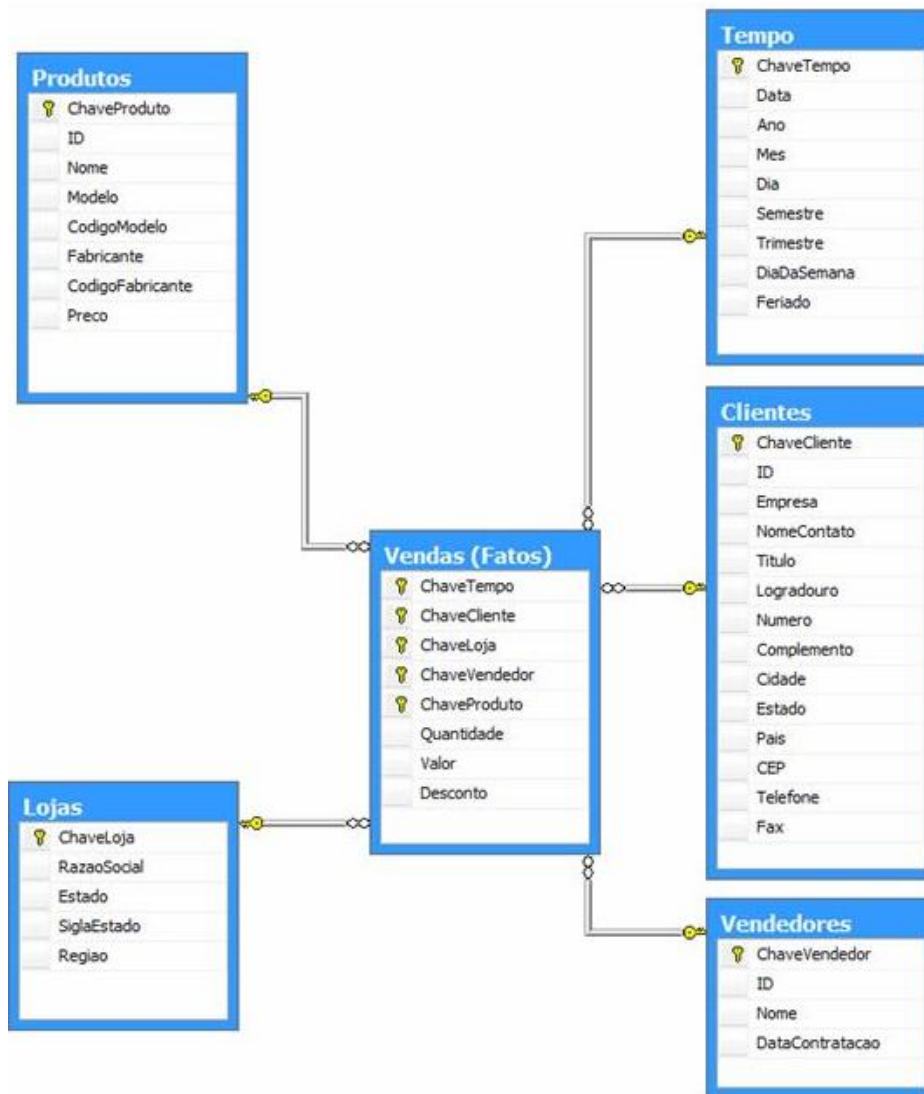


Figura 3 – Modelo Dimensional em Estrela (“Fundamentos e Modelagem de Bancos de Dados Multidimensionais,” n.d.)

### 3.2.6.2 Modelo dimensional em Floco de Neve

O modelo em Floco de neve, *snow-flake*, presente na Figura 5 baseia-se no modelo *star-schema*, referido anteriormente, e distingue-se pelo facto de as dimensões poderem ter ligações com outras dimensões, ou seja, há relação entre dimensões. Este tipo de modelo dimensional é usado para eliminar redundância de dados, no entanto, é desaconselhado usar pois adiciona complexidade acrescida para relacionar a informação, sendo que o propósito de um AD é a informação poder ser acedida de uma forma mais *user-friendly* (Kimball and Ross, 2013).

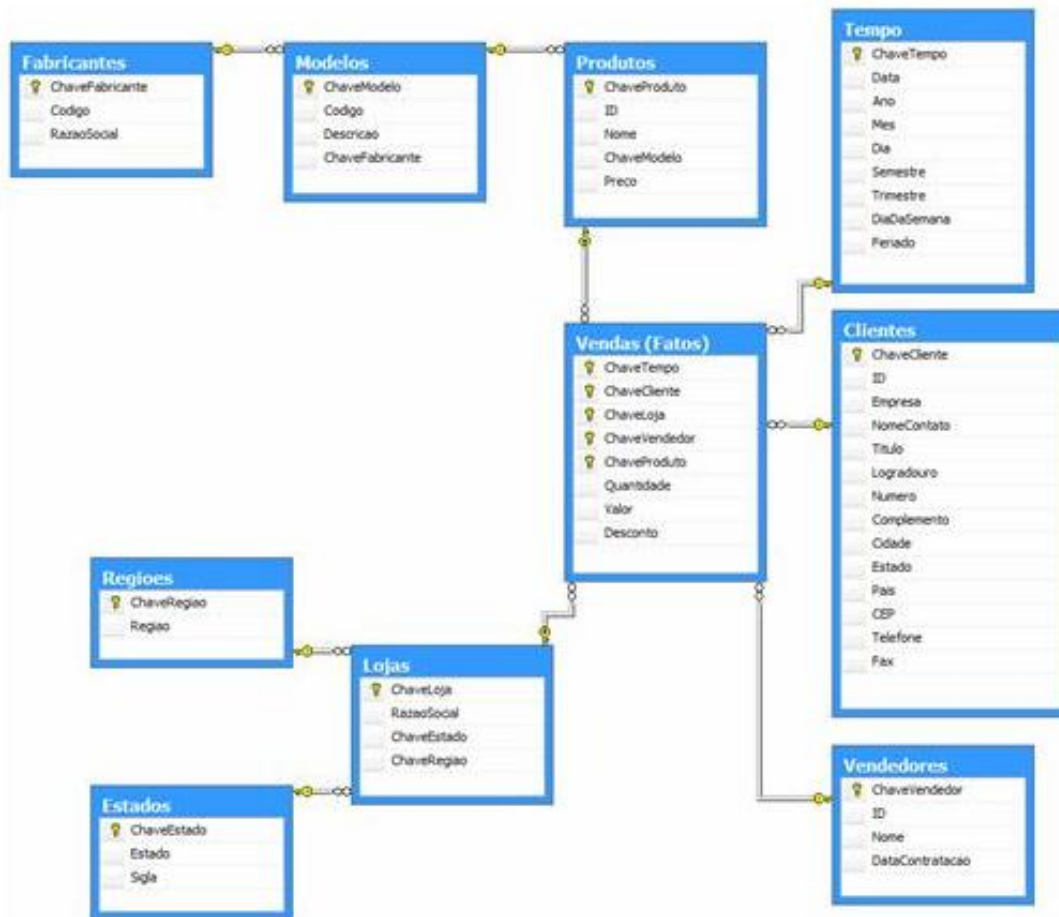


Figura 4 – Modelo Dimensional em Floco de Neve (“Fundamentos e Modelagem de Bancos de Dados Multidimensionais,” n.d.)

### 3.2.6.3 Modelo Constelação de Factos

Como se pode visualizar na Figura 5, o modelo de constelação de factos baseia-se no modelo em estrela, que foi anteriormente explicado, e diferencia-se na medida em que as dimensões são partilhadas pelas diferentes tabelas de factos (“Fundamentos e Modelagem de Bancos de Dados Multidimensionais,” n.d.).

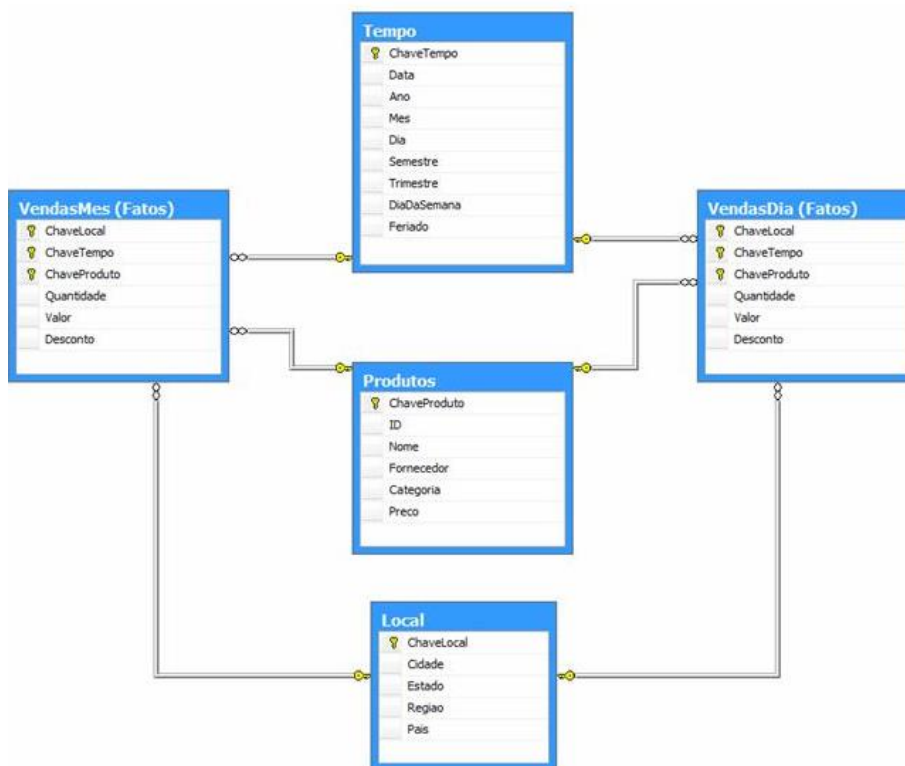


Figura 5 - Modelo dimensional de constelação de factos (“Fundamentos e Modelagem de Bancos de Dados Multidimensionais,” n.d.)

Assim, permite reduzir a manutenção no AD pois as dimensões são partilhadas entre as tabelas de factos e não é necessário criar dimensões com dados do mesmo contexto (“Fundamentos e Modelagem de Bancos de Dados Multidimensionais,” n.d.).

### 3.2.7 Slowly Changing Dimensions

Como o próprio nome indica, esta técnica de modelação aplica-se apenas em dimensões. Tem como propósito captar mudanças existentes nos valores dos atributos que geralmente são estáticos quando comparado com tabelas de factos. Para cada dimensão existente deve ser delineada uma estratégia de mudança, por forma a corresponder às alterações que acontecem nos sistemas operacionais (Kimball and Ross, 2013).

Segundo *Ralph Kimball*, existem vários tipos de SCD (*Slowly Changing Dimensions*) sendo que os mais conhecidos e utilizados são SCD do tipo um, dois e três (Kimball and Ross, 2013). Uma SCD do tipo um consiste em reescrever os atributos que foram alterados e os registos existentes na tabela de factos são associados com este novo (s) campo (s) campos. Uma SCD do tipo dois, é o tipo mais utilizado, consiste em adicionar uma nova linha que reflete as novas alterações, colocando o novo registo como corrente e inserir a data corrente como a data efetiva do registo. Quanto ao registo antigo, este é colocado como expirado e a data em que foi adicionado o novo

registro é colocada na data do antigo registro no campo da data em que expirou, como ilustrado na Figura 6.

Original row in Product dimension:

Product Key	SKU (NK)	Product Description	Department Name	...	Row Effective Date	Row Expiration Date	Current Row Indicator
12345	ABC922-Z	IntelliKidz	Education	...	2012-01-01	9999-12-31	Current

Rows in Product dimension following department reassignment:

Product Key	SKU (NK)	Product Description	Department Name	...	Row Effective Date	Row Expiration Date	Current Row Indicator
12345	ABC922-Z	IntelliKidz	Education	...	2012-01-01	2013-01-31	Expired
25984	ABC922-Z	IntelliKidz	Strategy	...	2013-02-01	9999-12-31	Current

Figura 6 - Exemplo de SCD do tipo dois (Kimball and Ross, 2013).

Por fim, uma SCD do tipo três consiste em adicionar um campo novo para capturar uma nova alteração. Um exemplo de utilização deste tipo de SCD é nos casos em que por exemplo estamos a guardar informação sobre produtos e a que departamento pertence, sendo que o produto pode, por ventura, mudar de departamento ao qual pertence. Neste caso, os utilizadores podem querer visualizar as vendas por departamento e reciprocamente capturar os dados em termos do antigo departamento, sendo que SCD do tipo dois não conseguiria responder a este tipo de requisito. Encontra-se ilustrado o exemplo anteriormente enunciado na Figura 7.

Original row in Product dimension:

Product Key	SKU (NK)	Product Description	Department Name
12345	ABC922-Z	IntelliKidz	Education

Updated row in Product dimension:

Product Key	SKU (NK)	Product Description	Department Name	Prior Department Name
12345	ABC922-Z	IntelliKidz	Strategy	Education

Figura 7 - SCD do tipo três (Kimball and Ross, 2013).

Não obstante, este tipo de SCD não é muito utilizado pois não consegue capturar alterações de atributos que podem mudar imprevisivelmente (Kimball and Ross, 2013).

### 3.3 Arquiteturas de Armazém de Dados

Cada AD desenvolvido possui uma estrutura distinta, com características específicas e relacionadas com o meio em que está inserido. Neste subcapítulo serão abordadas as principais teorias existentes atualmente, defendidas por *Bill Inmon* e *Ralph Kimball*.

#### 3.3.1 Kimball BI Architecture

“Pense como um restaurante. Imagine que os clientes do restaurante são os utilizadores finais e a comida são os dados. Quando os alimentos são oferecidos a todos os clientes na sala de jantar, estes são servidos exatamente no local onde esperam receber e a forma: limpos, organizados e apresentados de uma forma em que cada peça pode ser facilmente distinguida e consumida. [...]. Na cozinha, a comida é selecionada, limpa, cortada, cozinhada e preparada para apresentação.” (Traduzido de Kimball and Ross, 2013).

Segundo *Ralph Kimball*, um AD encontra-se dividido em duas áreas lógicas, podendo em alguns casos encontrarem-se separadas fisicamente: camada de implementação (*The Back Room*) e a camada de apresentação (*The Front Room*). De uma forma breve, na primeira área encontra-se toda a lógica criada incluindo ligações às fontes de dados (*Source Systems*), processos para extração e tratamento desses mesmos dados e processos para armazenamento num sistema de base de dados de *staging* (*The Staging Area*). Esta seção também inclui AD, que vai ser carregado a partir dos dados resultantes do processo de transformação de dados.

Até aqui, o acesso é interdito por parte da camada de apresentação, sendo que os dados passam a estar disponíveis para os utilizadores finais através da camada de apresentação que engloba todas as aplicações de suporte ao uso e análise dos dados (*The Presentation Area*) (Kimball and Ross, 2013). Retirada da mesma fonte de informação, de seguida é apresentada na Figura 8 uma ilustração da composição das divisões lógicas descritas.

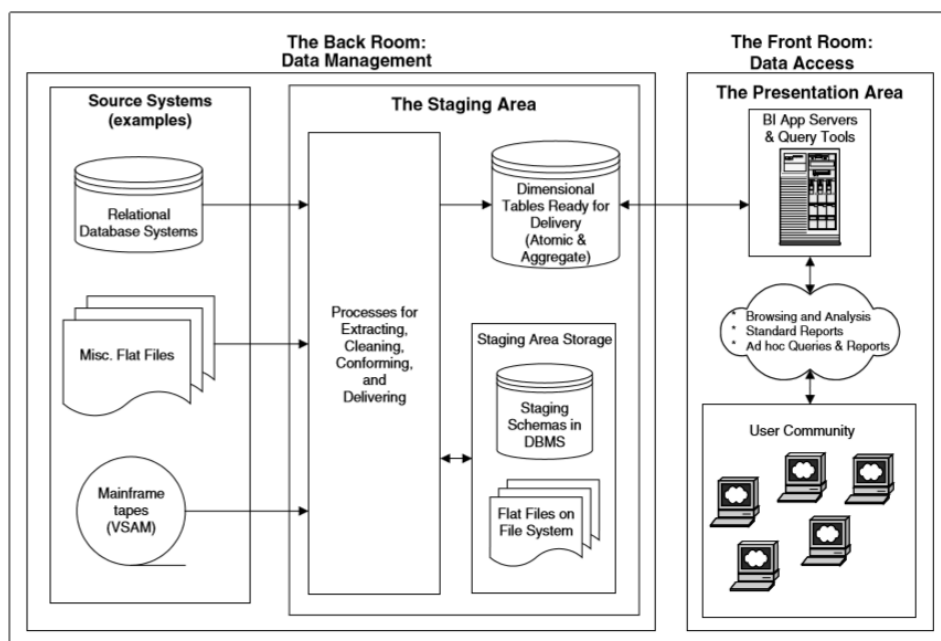


Figura 8 - Conceito *The Back and Front Room* (Kimball and Caserta, 2004).

Tomando como foco a primeira camada, todo o processo começa a partir das fontes de dados. Estas fontes são as que possuem os dados necessários que a posteriori vão alimentar o AD podendo assumir formatos variados desde uma base de dados operacional, utilizada para a gestão do negócio, até a ficheiros de texto. Primeiramente é feita uma extração dos dados em bruto, de uma ou mais fontes, podendo ser armazenada previamente facilitando o recomeço do processo de carregamento sem ter a necessidade que sobrecarregar novamente o sistema na extração seguinte. Depois de extraídos os dados, estes passam para uma área denominada por *Staging Area* (Kimball and Caserta, 2004); (IBM, 1999).

“A cozinha é uma área de trabalho, fora do alcance dos clientes do restaurante.” (traduzido de (Kimball and Caserta, 2004)).

Continuando com a analogia, este define-a como uma área de limpeza e tratamento dos dados, sendo composta por um processo de extração, tratamento, carregamento (processo ETL) e por um sistema de armazenamento interno invisível para o utilizador, com o fim de armazenar os dados tratados antes de passar para o AD final. Após estes dados serem carregados em *staging*, recebem o último tratamento antes do armazenamento no modelo multidimensional, denominado por *Enterprise Bus Architecture*. Estes dados encontram-se na forma mais atómica possível para permitir dar resposta quer a *query's* imprevisíveis, bem como a agregações (*roll-up*) ou conseguir ir a um maior nível de detalhe (*drill-down*) (Oracle, 2007). Depois de estes dados serem inseridos no AD, podem ser feitas manutenções a nível de histórico e atualizações aos registos.

Esta arquitetura defende uma metodologia do tipo *bottom-up*, considerando que a criação de valor para o AD pode ser incrementada ao longo do tempo através de novos *data marts*,

acrescentando assim novos domínios de informação. Não requer que todos estes domínios sejam criados logo na fase inicial.

### 3.3.2 Inmon BI Architecture

“O armazém de dados é o coração do ambiente arquitetado, e é o fundamento de todo o processamento do sistema de suporte à decisão, DSS. [...]. Existe uma única fonte integrada de dados [...] está orientado para as principais áreas de negócio que foram definidas no modelo de dados corporativo de alto nível. Cada área de negócio está fisicamente implementada como uma série de tabelas relacionadas no armazém de dados.” (Traduzido de (Inmon, 2002).

*Inmon* defende que todos os dados operacionais que existem numa organização são parte integrante do AD. Este tipo de arquitetura assenta numa metodologia do tipo *top-down*, onde primeiramente se procura identificar todos os domínios de dados existentes com o objetivo de criar relações lógicas entre si, resultando assim num modelo de dados essencialmente relacional. Desta forma, a estrutura total do AD é definida e só depois possivelmente se dividirá em diferentes *data marts*, incluindo toda a estrutura numa área denominada por CIF (*Corporate Information Factory*).

“Um armazém de dados é um conjunto de dados de suporte a decisões de administração orientado ao assunto, integrado, não volátil e variante no tempo” (Traduzido de (Inmon, 2002).

Este tipo de arquitetura possui algumas especificidades, distintas da arquitetura apresentada anteriormente. O AD é orientado ao assunto, ou seja, a informação encontra-se organizada de acordo com as relações que possui, resultando numa base de dados operacional normalizada na terceira forma normal. Estas relações existem dado que esta arquitetura defende que todos os dados que possuam o mesmo domínio de informação devem ser relacionados. Relativamente à atualização dos dados, não existe qualquer possibilidade de o fazer. Cada um dos registos é preservado e guardado tal e qual como foi inserido para efeitos de análise futuras, registando apenas o momento de inserção para existir a perceção da variável tempo perante os restantes dados. Por último, este tipo de AD é também integrado dado que pode conter informação de mais do que um sistema relacional existente, tornando a informação nele existente consistente por seguir uma metodologia igualmente relacional.

Como apresenta a Figura 9, a nível de extração, tratamento e carregamento de dados, é igualmente utilizada uma área de *staging* como a primeira camada a receber informação das diferentes fontes de dados. Toda a informação passa desta área diretamente para o AD, sendo a partir deste sistema que cada *data mart* de informação é logicamente criado e carregado.

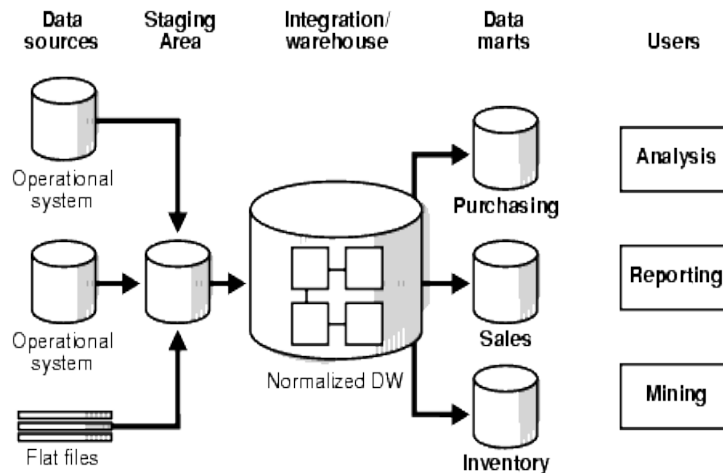


Figura 9 - Bill's Inmon architecture (Oracle, 2002).

### 3.3.3 Comparação entre arquiteturas

Na Tabela 1, é feita uma comparação entre as duas arquiteturas enunciadas, apresentando os pontos positivos e negativos em cada uma.

Tabela 1 - Comparação entre arquiteturas *Kimball* e *Inmon* (Inmon, 2002);(Kimball and Caserta, 2004);(Kimball and Ross, 2013);(Sansu George, 2012)

	<b>Arquitetura Ralph Kimball</b>	<b>Arquitetura Bill Inmon</b>
<b>Dados de negócio e o AD</b>	O AD é composto por vários <i>data marts</i> em que cada um é responsável por um segmento no global do negócio. Todos os <i>data marts</i> resultam assim no AD.	Todos os dados são parte integrante do AD. É definida uma estrutura global e só depois, se existir essa necessidade, podem ser criados segmentos à parte.
<b>Staging area</b>	Defende o conceito.	Defende o conceito.
<b>ETL</b>	Defende o conceito.	Defende o conceito.
<b>Data Marts</b>	Numa primeira fase, são definidos cada um dos segmentos, dando origem a <i>data marts</i> distintos. Só depois é que o AD é definido (abordagem <i>bottom-up</i> ).	Logo à partida é definida a estrutura do AD. Apenas se necessário, esta estrutura pode ser segmentada em <i>data marts</i> distintos (abordagem <i>top-down</i> ).
<b>Variante no tempo</b>	Defende o conceito.	Defende o conceito.
<b>Modelo de AD</b>	Segue o modelo dimensional.	Essencialmente relacional (na terceira forma normal).
<b>Orientação aos processos</b>	Não, é orientado ao assunto.	Sim.
<b>Complexidade de desenvolvimento</b>	Simple, na medida em que cada estrutura de dados existente é pensada e segmentada logo na	Complexa, dado que primeiramente se constrói uma estrutura global para todo o tipo

	fase da construção do AD. O tipo de relações que este tipo de estruturas pode ter é favorável, dado que as dependências são poucas, baseando-se no geral em relações entre dimensões e tabelas de factos.	de informação da organização. As reações entre dados podem ser de perceção complexa e podem deteorar a performance a nível de pesquisas de dados se a dependência entre dados for muito grande.
<b>Registo de alterações nos dados (SCD)</b>	Suporta. É a arquitetura defensora.	Não defende.
<b>Tempo de desenvolvimento</b>	Menor tempo de desenvolvimento.	Maior tempo de desenvolvimento.
<b>Custo</b>	Menor custo de desenvolvimento inicial. Cada segmento que seja construído numa fase posterior terá exatamente o mesmo custo.	Maior esforço inicial. Os desenvolvimentos seguintes terão menor custo de desenvolvimento.
<b>Conhecimentos requeridos</b>	Não são requeridos conhecimentos especialistas, apenas generalistas.	Equipa especializada, dada a complexidade do modelo.

Concluindo, cada uma das arquiteturas possui as suas especificidades. Aquando da escolha da melhor arquitetura deve ser tido em linha de conta o contexto e as necessidades por satisfazer, optando por uma solução progressiva de *Kimball* ou uma solução mais tradicional como a de *Inmon*.

### 3.4 Online Analytical Processing

OLAP (*Online Analytical Processing*) é o mecanismo de análise de sistemas multidimensionais que disponibiliza a capacidade de realizar operações complexas e sofisticadas sobre este tipo de modelos. Desta forma, permite aos utilizadores finais realizar *queries ad-hoc* em múltiplas dimensões, disponibilizando a informação que necessitam para tomarem as suas decisões.

A vantagem de utilizar OLAP está na velocidade de acesso à informação armazenada no modelo multidimensional criando agregações e cálculos muito rapidamente em múltiplos conjuntos de dados. A implementação deste tipo de tecnologia depende do tipo de *software* que está a utilizar mas também do tipo das fontes de dados e dos objetivos do negócio em que se insere (OLAP.com, 2016).

Existem dois tipos de OLAP: MOLAP e ROLAP, sendo que HOLAP é uma mistura dos dois. No MOLAP, *Multidimensional Online Analytical Processing*, os dados estão armazenados num cubo multidimensional, proporcionando a rapidez na obtenção dos dados e nas operações de *slicing*

e *dicing*<sup>1</sup> em cubos. Por sua vez, o ROLAP, *Relational Online Analytical Processing*, consiste em efetuar análises em dados armazenados num modelo relacional e efetuar as operações como no MOLAP. Por fim, o HOLAP, *Hybrid Online Analytical Processing*, consiste na junção do MOLAP e ROLAP, sendo que o HOLAP aproveita a tecnologia do cubo para um processamento mais rápido (1keydata, 2015).

Por fim, na Tabela 2, encontra-se ilustrado as diferenças entre um sistema OLAP, anteriormente explicado, e um sistema OLTP (*Online Transaction Processing*), que é um sistema tipicamente relacional (Atanazio, 2013).

Tabela 2 - OLAP versus OLTP (Atanazio, 2013)

	OLAP	OLTP
<b>Fontes de dados</b>	Os dados provêm dos sistemas relacionais	Bases de dados operacionais
<b>Propósito dos dados</b>	Ajudar na tomada de decisão e planeamento	Controlar e executar as operações de negócio
<b>Queries</b>	Complexas	Simples
<b>Velocidade de processamento</b>	Depende do volume de dados a analisar	Rápido Processamento
<b>Requisitos de espaço</b>	Grande capacidade	Relativamente pequeno
<b>Modelo da Base de Dados</b>	Multidimensional	Relacional
<b>Backup e Recuperação</b>	Backup não regular	Backup regular
<b>Idade dos dados</b>	Histórico	Corrente
<b>Operações efetuadas</b>	Ler	Adicionar, atualizar, ler e eliminar
<b>O que os dados revelam</b>	Vistas multidimensionais de vários tipos de atividades de negócio	Imagem do processo de negócio corrente

### 3.5 Extração, Transformação e Carregamento de Dados

Nesta subsecção será apresentado o processo o processo de extração, transformação e carregamento (ETL) de dados num AD.

De seguida, será feita uma apresentação de algumas ferramentas de ETL existentes no mercado e, por fim, uma comparação entre ambas por forma a classificar cada uma, de modo a facilitar a escolha da ferramenta a utilizar.

<sup>1</sup> Tipo de operações efetuadas sobre o cubo de dados. Quando é analisada informação apenas para um determinado valor de uma dimensão, estamos perante uma operação do tipo *slice*. A operação *dice* consiste em analisar um conjunto de valores entre múltiplas dimensões.

### 3.5.1 Processo de ETL

O processo de ETL, como anteriormente referido, consiste no processo de Extração, Carregamento dos dados oriundos da respetiva fonte e encontra-se ilustrado na Figura 10.

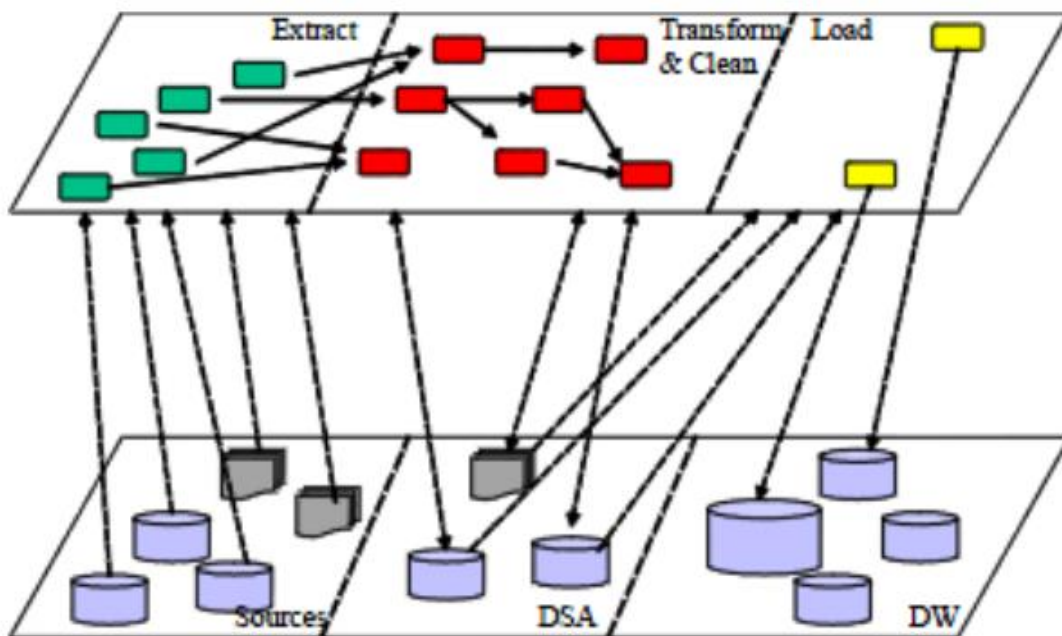


Figura 10 - Processo de ETL (Ferreira et al., 2016)

Como se pode constatar na Figura 10 no primeiro passo é efetuada a extração (*Extract*) dos dados das respetivas fontes (*sources*) de modo a alimentar a *Staging Area* (DSA) (Ferreira et al., 2016).

De seguida, após extrair os dados é feita uma transformação e limpeza dos dados (*Transform & Clean*) por forma a corresponder ao formato dos dados necessário. Esta transformação ocorre para eliminar possíveis erros e converter para um formato homogéneo os dados dos vários sistemas fonte (Ferreira et al., 2016).

Por fim, ocorre o processo de carregamento (*Load*) que consiste em carregar para o AD os dados processados na *Staging Area*, estando já com os dados com o formato desejado (Ferreira et al., 2016). É de realçar que nesta fase pode ainda haver alguma transformação necessária aos dados.

### 3.5.2 Ferramentas

Para implementar um AD, é fulcral que a fase de ETL seja bem implementada e flexível. Para isso a escolha de uma ferramenta para o processo é fundamental sendo que, a versatilidade e

configurabilidade que se pode adicionar num processo de ETL é sempre bem-vindo de modo a este adaptar-se a vários cenários e evoluir no tempo.

Nos tópicos que se seguem serão apresentadas algumas ferramentas de ETL que estão no top dez de utilização pelas empresas, sendo que as últimas três são *freeware* (ETL, 2015).

#### 3.5.2.1 Oracle Data Integrator

O ODI é uma ferramenta desenvolvida pela *Oracle* que permite a integração, transformação, replicação, gestão de meta dados, serviços e qualidade de dados num AD. Recentemente, foi lançada uma versão que permite a integração com a *cloud* e com possibilidade de efetuar análises em soluções *Big Data*<sup>2</sup>, sendo por isso uma mais-valia (Oracle, 2016); (ETL, 2015).

Permite interagir com vários sistemas heterogéneos através da utilização de outras ferramentas complementares, aumentando a performance das soluções de BI. Como consequência da interação, através do *Oracle GoldenGate* é possível obter dados em tempo real, através de processos de sincronização de dados nestes sistemas. Desta forma, o desempenho é maximizado aquando da migração de dados, sem tempo de inatividade, possibilitando a recuperação em caso de desastre e respetiva sincronização ativa e contínua nas bases de dados. Esta ferramenta também possui componentes *drag and drop*, que abstraem da maior parte de implementação apenas dando pequenas instruções e comandos SQL para efetuar o processo de ETL (Narasimharajan], 2011). Para além dos benefícios, a documentação sobre a ferramenta é escassa para ajudar a integrar e a criar processos de carregamento e extração de dados, existindo por outro lado formações da *Oracle* mas com um custo associado para as adquirir (InformationWeek, 2015).

Na Figura 11 é apresentado um exemplo do ambiente de utilização da ferramenta, na criação de um processo de extração e carregamento de dados relativos a clientes.

---

<sup>2</sup> Termo utilizado para descrever a existência de uma grande quantidade de informação.

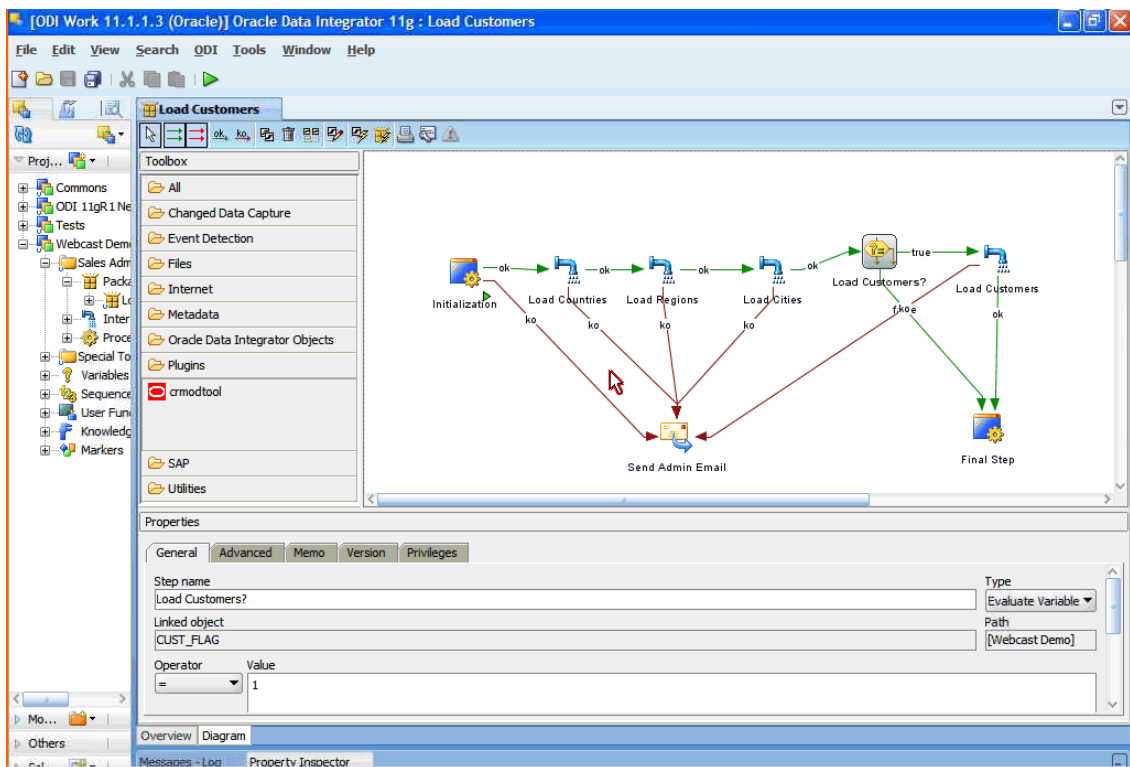


Figura 11 – Oracle ODI (ARSON Group SAC, n.d.).

### 3.5.2.2 IBM – InfoSphere Information Services

O *InfoSphere* da IBM destina-se a soluções de BI, sendo um software otimizado que permite não só a integração com várias fontes de dados, mas também permite realizar análises com um elevado nível de detalhe. Adicionalmente, permite também proteger os dados recorrendo a metodologias de segurança.

O desenvolvimento de soluções usando esta ferramenta é bastante intuitivo, na medida em que, possui uma interface gráfica com componentes *drag and drop* que facilitam o desenho de todo o processo de ETL, como se pode verificar na Figura 12. Não obstante, permite integração com a *cloud*, graças ao *IBM dashDB*, que consiste em gerir todos os dados armazenando-os na *cloud* permitindo o acesso instantâneo à informação, não sendo necessário recorrer a uma ferramenta de análise para visualizar os dados (IBM, 2015). Na Figura 12 é apresentado um exemplo de carregamento de dados para o AD, a partir de fontes de dados no formato xml.

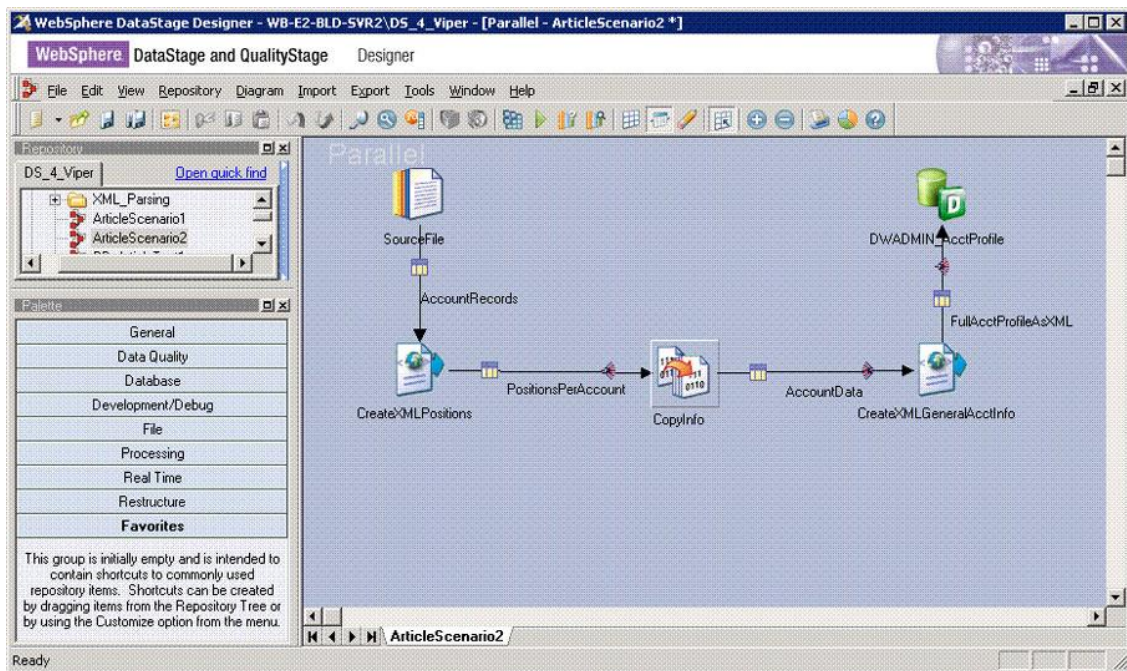


Figura 12 - IBM InfoSphere DataStage

### 3.5.2.3 Microsoft SQL Server Integration Services

O *Microsoft Integration Services*, mais conhecido por SSIS, é uma plataforma de integração de dados utilizada a nível empresarial. Este tipo de ferramenta é utilizado para trabalhar soluções com alguma complexidade de lógica de negócio, permitindo o armazenamento, limpeza e gestão de dados. Para ajudar na resolução de problemas, permite copiar ou descarregar ficheiros, responder a eventos com correio eletrónico de entrada de *log* de possíveis erros ou apenas a título informativo. É de salientar que esta trabalha com pacotes que podem correr em simultâneo e suportam dados provenientes de várias fontes de dados (Microsoft, 2015).

A Figura 13 apresenta um exemplo de um fluxo de dados criado no âmbito da *staging* para o carregamento dos dados de avaliações dos alunos.

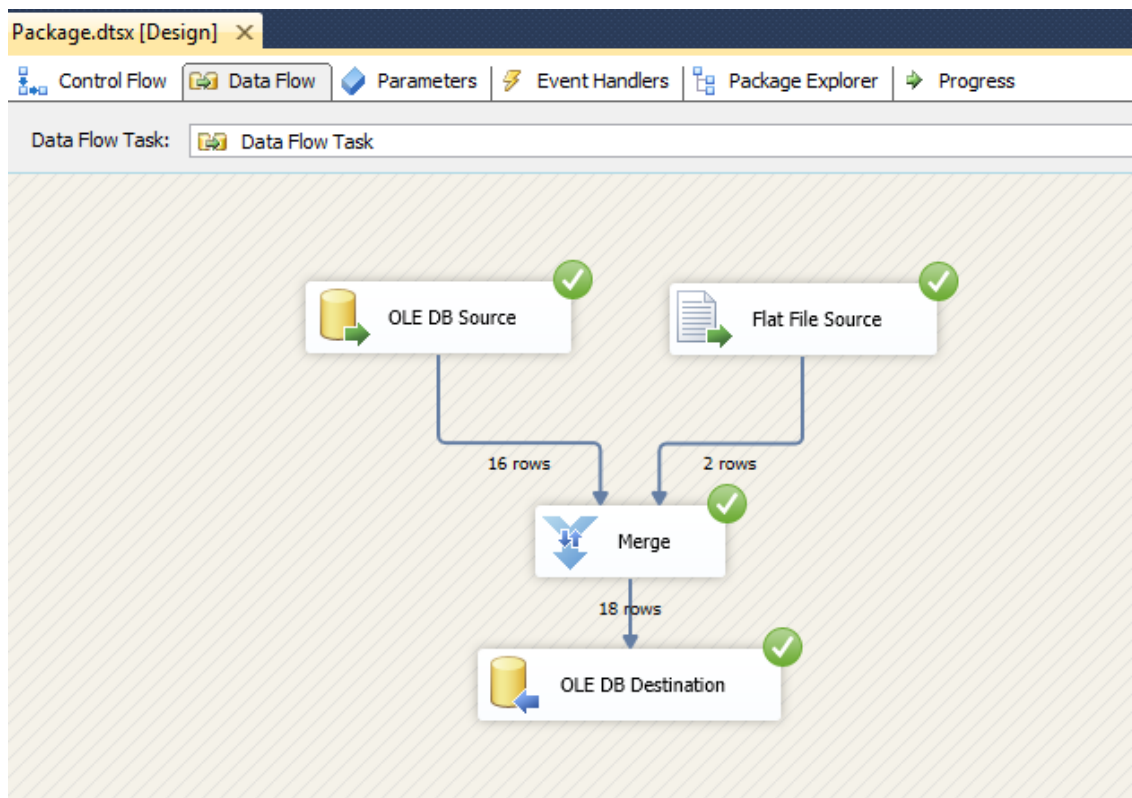


Figura 13 – Exemplo utilizando *Microsoft SSIS IDE* (Anoop Kumar, 2013).

Esta ferramenta permite criar todo o processo de carregamento, extração e transformação através de grafos, de forma simples e dinâmica, em que cada um dos seus nós é composto por um componente para efetuar uma entre as várias operações aos dados disponíveis naquele fluxo de informação. No exemplo apresentado anteriormente, foram utilizados componentes primeiramente para extrair os dados de um SGBD e de um ficheiro, para de seguida juntar as informações das duas fontes com o objetivo de inserir toda essa informação num determinado armazém. Ao executar o pacote, é possível facilmente identificar a quantidade de registos que passa por cada caminho do grafo, bem como compreender que todo o processo ocorreu como esperado.

Para desenvolver no SSIS é necessário instalar a ferramenta SSDT (*SQL Server Data Tools*) para criar as várias camadas que constituem a solução de BI. É necessário ter instalado o *SQL Server Management Studio* para gerir e/ou executar num ambiente de produção (Microsoft, 2015).

Por fim, também é possível realizar as seguintes operações via SSDT (Microsoft, 2015):

- Importar e Exportar pacotes via assistente para criar cópia de dados de uma fonte para um destino;
- Criar pacotes com um fluxo de controlo complexo, fluxo de dados, lógica orientada e registo de eventos;
- Realizar testes e *debug* sobre os diferentes pacotes existentes, bem como monitorizar os diferentes recursos que o SSIS utiliza;

- Criar configurações que permitam alterar as definições/propriedades dos pacotes em tempo de execução;
- Utilitário que permite instalar os pacotes e suas dependências noutras máquinas;
- Gravar cópias dos pacotes numa base de dados específica, do sistema do SSIS e do sistema de ficheiros.

#### 3.5.2.4 Pentaho Data Integration

O Pentaho Data Integration é uma ferramenta poderosa de ETL com uma abordagem orientada por meta dados. Como referido nas ferramentas anteriores, esta também não foge à regra e possui uma interface gráfica intuitiva, com componentes *drag and drop* como é possível verificar na figura Figura 14 (Pentaho, 2015).

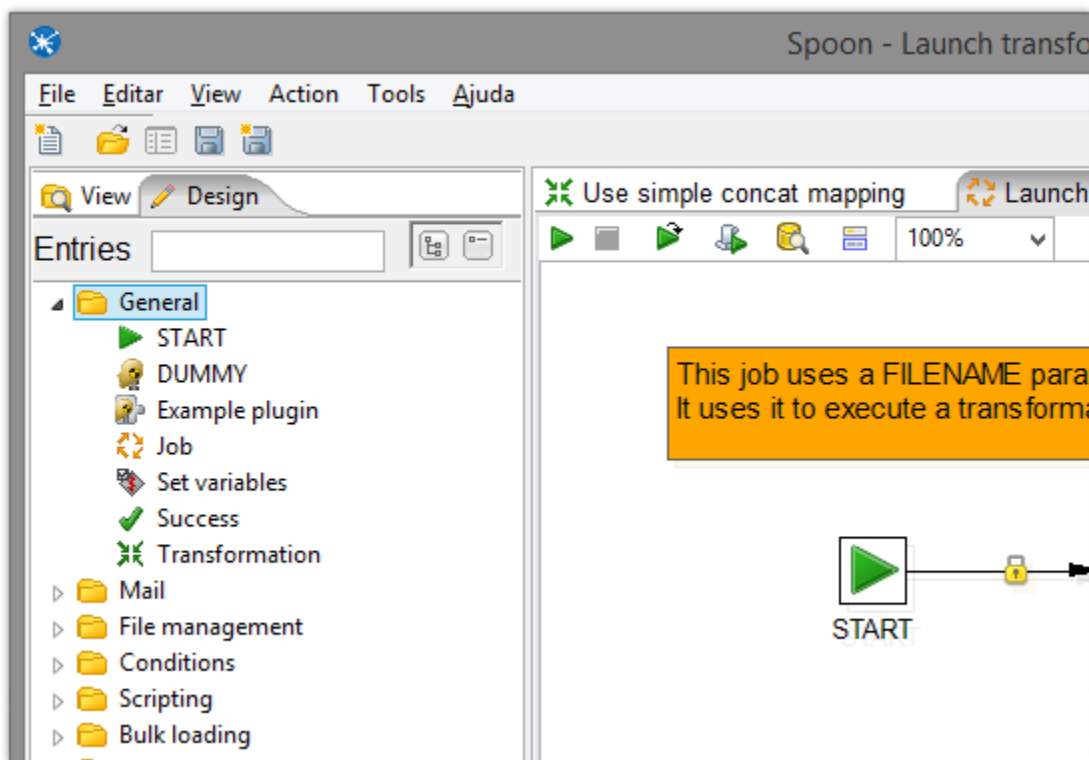


Figura 14 – Pentaho IDE.

Por fim, possui funcionalidades de monitorização, criação de perfil de dados, mecanismo de balanceamento de carga, mecanismo para reverter alterações efetuadas por *jobs* e obter dados de múltiplas fontes de dados. Esta ferramenta possui uma loja de *plugins* que se pode adicionar à solução base, permitindo assim complementar a aplicação fornecida pela Pentaho e integrar ou transformar dados de diferentes fontes. Não obstante, esta ferramenta possui também capacidade de elaborar soluções recorrendo ao *Big Data* (Pentaho, 2015).

### 3.5.3 Comparação de ferramentas ETL

Na Tabela 3, encontra-se a comparação das ferramentas de ETL apresentadas na secção anterior. Esta tabela apresenta os prós e contras da utilização de cada ferramenta (“ETL Tools comparison,” 2016),(McBurney, 2007).

Tabela 3 - Comparação de ferramentas ETL (“ETL Tools comparison,” 2016),(McBurney, 2007) .

	Vantagens	Desvantagens
<b>Microsoft Integration Services (SSIS)</b>	Vasta Documentação e Suporte; Boa performance perante grandes volumes de dados; Possui interação gráfica a nível de criação nos mapeamentos dos dados.	Não é possível executar em ambientes não Windows; Manuseamento da ferramenta pode tornar-se complexo; Visão e estratégia pouca claras.
<b>Oracle Data Integrator (ODI)</b>	Melhor conexão com aplicações de AD da Oracle; Complexidade de utilização média; Possibilidade de integração de todas as ferramentas num só ambiente aplicacional.	Complexidade de utilização; Pouca documentação; Esta ferramenta apenas se foca no processo ETL; Apenas é utilizada em <i>back-end</i> a nível de tratamento e transformação de dados, sem qualquer preocupação na disponibilização dos dados.
<b>Pentaho Data Integration</b>	Bom suporte fornecido pela Pentaho; Fácil utilização; Permite integração com outros produtos (BI, EII e EAI); Performance razoável perante grandes volumes de dados; Vários objetivos de transformação e suporte aumentando as <i>slow changing dimensions</i> ; Permite integração com serviços <i>web</i> .	Permite transformar os dados num <i>cluster</i> , mas não executa eficientemente o reparticionamento dos dados em paralelo; Não possui componente adequado para realizar testes de qualidade; Só consegue lidar com pequenos volumes de dados quando efetua <i>lookup's</i> ; Baixo desempenho quando necessita de executar vários <i>scripts</i> de SQL em <i>lookup's</i> .
<b>IBM – InfoSphere Information Services</b>	Performance razoável perante grandes volumes de dados; Permite integração com serviços <i>web</i> ; Visão mais forte e flexível do mercado;	Processamento em paralelo; Aprendizagem pode tornar-se complexa e difícil; Necessita de muito bom equipamento para que o processamento seja razoável.

Como se pode verificar, cada uma das ferramentas apresentam mais valias no contexto de desenvolvimento de um processo de tratamento e carregamento de dados. A ferramenta da Microsoft possui capacidades real-time que não são muito evidenciadas pelo ODI, no entanto a ferramenta da Oracle possui um âmbito mais focado no ETL e processamento paralelo. No caso da ferramenta Pentaho e IBM, apesar de ser uma ferramenta com um bom suporte fornecido por parte da empresa que o desenvolve, este contém problemas de performance quando lida com um maior volume de dados.

## 3.6 Ferramentas de apresentação/analíticas

### 3.6.1 Microstrategy

A *MicroStrategy Analytics™* é uma ferramenta analítica comercializada pela MicroStrategy que permite a análise e a partilha de conhecimento a partir de grandes volumes de dados. Possui uma interface familiar ao utilizador comum, permitindo de forma simples e prática a geração de relatórios e *dashboards*, apresentando e medindo a informação pretendida. Atualmente, é bastante utilizada por várias organizações a nível mundial, como o Ebay, a Zurich, o Novo Banco, a Portugal Telecom, entre outras organizações (MicroStrategy, 2015).

Numa primeira fase, são apresentados conjuntos de opções com o objetivo de definir quais as fontes de dados alvo. Tomando como exemplo uma importação de dados de um ficheiro do tipo Excel e depois de escolhidas as folhas onde é necessário retirar informação, a ferramenta retira os atributos existentes e determina possíveis métricas a serem medidas pelo tipo de estrutura apresentado. É possível alterar a estrutura pré-deduzida.

Numa fase seguinte, o utilizador é livre de definir quais os atributos e métricas a analisar. Possui uma interface apelativa e fácil de utilizar, na medida em que os atributos e métricas são selecionados para o *dashboard* através de *drag and drop*. Cada um destes pode ser personalizado com vários tipos de gráficos, por forma a embelezar e tornar perceptível o que se está a analisar, para posterior avaliação e tomada de decisão.

Também, contém métodos internos que realizam algumas transformações aos dados por forma a realizar as análises com base na estrutura de informação definida, nomeadamente tarefas de *Data Mining*<sup>3</sup> e análises preditivas.

De seguida na Figura 15 é apresentado um exemplo de utilização, disponibilizado pela própria empresa, relacionado com o ciclismo. Apresenta a média de utilizadores por hora do dia, dia da semana, por mês e ainda a comparação do número de utilizadores versus a temperatura do dia.

---

<sup>3</sup> *Data Mining*: é o processo de descoberta de conhecimento em bases de dados.

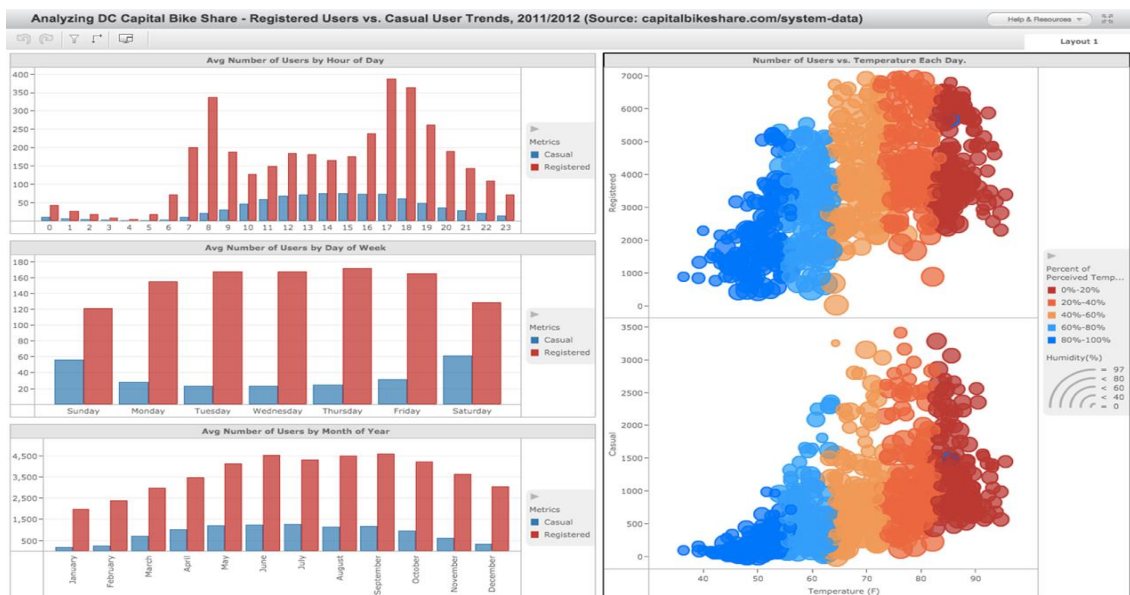


Figura 15 – Exemplo de utilização *MicroStrategy Analytics™ desktop* (MicroStrategy, 2015) .

Para além da aplicação *desktop*, também permite que toda esta informação seja acedida por dispositivos móveis, facilitando assim essencialmente a partilha de informação entre grupos de trabalho.

Por fim, é de realçar que os preços de licenças desta ferramenta são elevados, não permitindo que qualquer empresa possa usufruir destas funcionalidades. Também, como limitação, não existe aplicação móvel para *Windows Phone*, o que a impede ser uma solução *cross-platform*.

### 3.6.2 Microsoft Power BI

Esta é outra das soluções analíticas encontradas, simples de usar e gratuita. O *Microsoft Power BI* é um conjunto de funcionalidades disponibilizadas online ou localmente em aplicação *desktop*, que permite a criação de *dashboards* interativos através de conjuntos de informação, independentemente do tipo e do tamanho. As criações feitas podem ser partilhadas entre utilizadores e publicadas para que qualquer pessoa consiga usufruir (Microsoft, 2016). Numa primeira abordagem, o utilizador começa por definir qual a fonte de dados. Possui uma lista variada de tipos de fontes que são aceites, permitindo ainda combinar dados entre fontes diferentes.

Na Figura 16 é apresentado um exemplo de utilização da ferramenta, no contexto da gestão do desenvolvimento feita a partir do *Visual Studio* (*framework* de desenvolvimento).

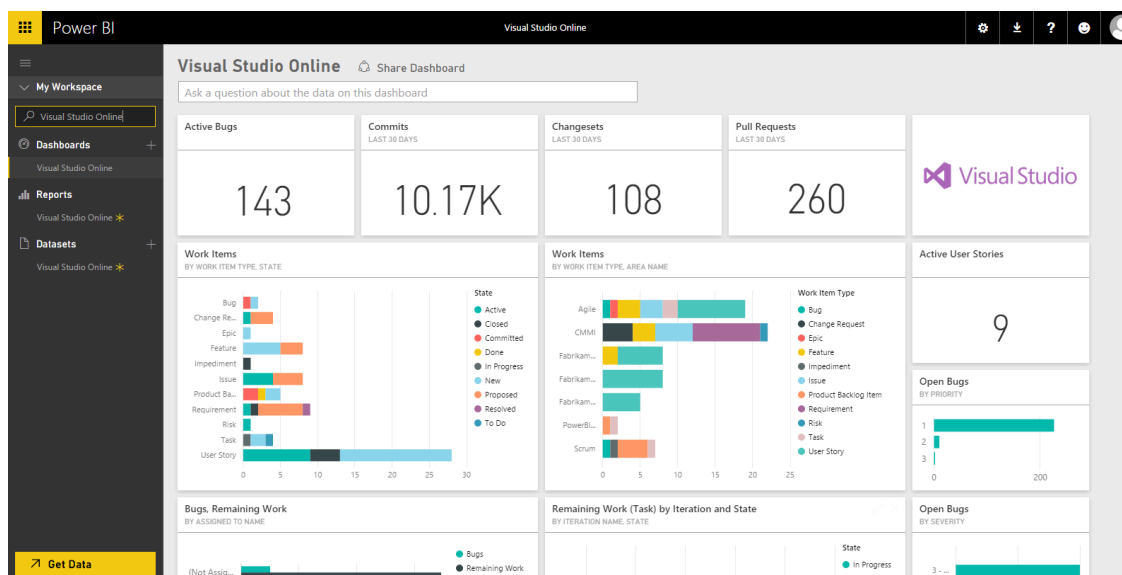


Figura 16 – Exemplo de utilização *Microsoft Power BI*.

### 3.6.3 Phocas

Como alternativa à solução apresentada anteriormente, o *software* Phocas disponibiliza funcionalidades capazes de efetuar análises e partilhar resultados. É composto por um módulo de análise de dados, onde o utilizador seleciona o tipo de informação a importar, permitindo realizar todo o processo de análise na nuvem de dados e a partir daí mostrar os dados aos utilizadores finais em diferentes plataformas personalizáveis. Apresenta uma interface simples e prática, apta a ser utilizada mesmo por utilizadores sem qualquer tipo de formação técnica específica (Phocas, 2015).

### 3.6.4 Comparação entre ferramentas analíticas de apresentação de dados

Na Tabela 4, encontra-se a comparação entre diferentes ferramentas analíticas existentes no mercado. Esta comparação segue os seguintes parâmetros, que são considerados importantes para uma ferramenta deste tipo (G2 Crowd, 2016):

- Usabilidade: facilidade com que se cria novas análises;
- Configuração: configuração da ferramenta na máquina onde se realiza as análises;
- Manutenção: facilidade com que se mantém as análises existentes;
- Suporte: suporte fornecido à ferramenta por parte do fabricante;
- Fácil para o negócio: facilidade de utilizar para o negócio;
- *Self-Service*: ser o mais fácil de ser utilizado pelo utilizador;
- Análises avançadas: possibilidade de efetuar análises complexas;
- Construção de novos relatórios: facilidade de construção de novos relatórios.

Tabela 4 - Comparação de ferramentas analíticas.

	<b>Microstrategy</b>	<b>Phocas</b>	<b>Power BI</b>
<b>Usabilidade</b>	Complexo	Intuitivo e Fácil	Intuitivo e Fácil
<b>Configuração</b>	Difícil de configurar	Simples de configurar	Simples de configurar
<b>Manutenção</b>	Difícil de manter	Fácil de Manter	Fácil de Manter
<b>Suporte</b>	Contém um bom suporte	Suporte muito bom	Suporte Reduzido
<b>Fácil para negócio</b>	Contém alguma complexidade	Muito fácil	Fácil
<b>Self-Service</b>	Não é uma ferramenta fácil de utilizar	Muito intuitiva	Intuitiva
<b>Análises Avançadas</b>	Suporta análises complexas	Bom suporte em análises complexas	Não lida muito bem com análises complexas
<b>Construção de relatórios</b>	Um pouco complexo	Bastante fácil e intuitivo	Fácil e intuitivo
<b>Mais utilizado</b>	Empresas	Pequenas- Médias Empresas	Pequenas- Médias Empresas

Desta forma, consegue-se apurar que a ferramenta Phocas é mais completa e simples, desde a interação por parte do utilizador, até à manutenção da informação utilizada nas análises de dados criadas.

### 3.7 Análise de Mercado

Esta secção tem como objetivo apresentar possíveis soluções ou estudos que foram desenvolvidos e postos em prática no âmbito da criação e manutenção de um AD no contexto escolar. Numa segunda abordagem são apresentadas aplicações de mercado, sendo vistas como uma possível solução para apresentação dos dados armazenados. Destacam-se pela sua baixa complexidade de compreensão/uso, bem como pela completa gama de funcionalidades que disponibilizam, de acordo com as necessidades do meio em que são utilizadas.

Para além destes, existem outros indicadores que são específicos das instituições, como é o caso da LEI-ISEP. Esse conjunto de indicadores serão apresentados numa próxima secção.

#### 3.7.1 Ferramentas de Gestão para Ensino

Nesta secção descrevem-se alguns dos produtos existentes atualmente no mercado.

### 3.7.1.1 EdVantage

Desenvolvido pela *VersiFit Technologies*, este é um produto que tem a gestão de dados curricular como objetivo principal. Como se pode verificar através da Figura 17, é possível avaliar todo o tipo de informação referente aos alunos, aos professores e a cada programa de disciplina que foi efetivamente lecionado. Para além da disponibilização do armazém de dados para armazenamento da informação, este produto permite ainda o uso de um conjunto de funcionalidades tais como: gerar *reports* de acordo com o tipo de informação que o utilizador pretende consumir, permite definir métricas comparativas, *drill-down* de análises, interrogar as consultas através de uma ferramenta avançada incluída no produto e é possível adaptar o produto às necessidades do consumidor (SchoolCity Inc., 2015).

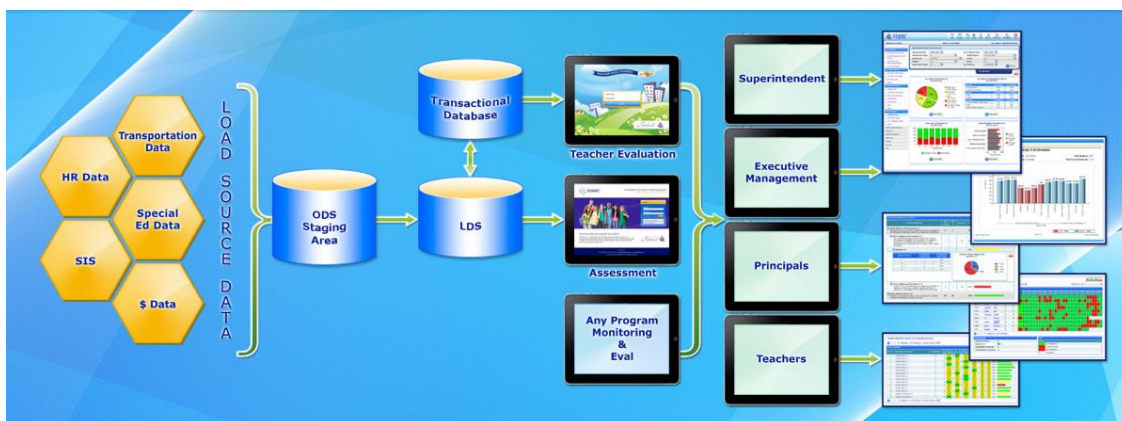


Figura 17 – EdVantage (SchoolCity Inc., 2015) .

Foi encontrada a utilização deste produto por parte da Escola Elementar de *Buffalo*, como se pode verificar na Figura 18 (School Buffalo, 2015) .

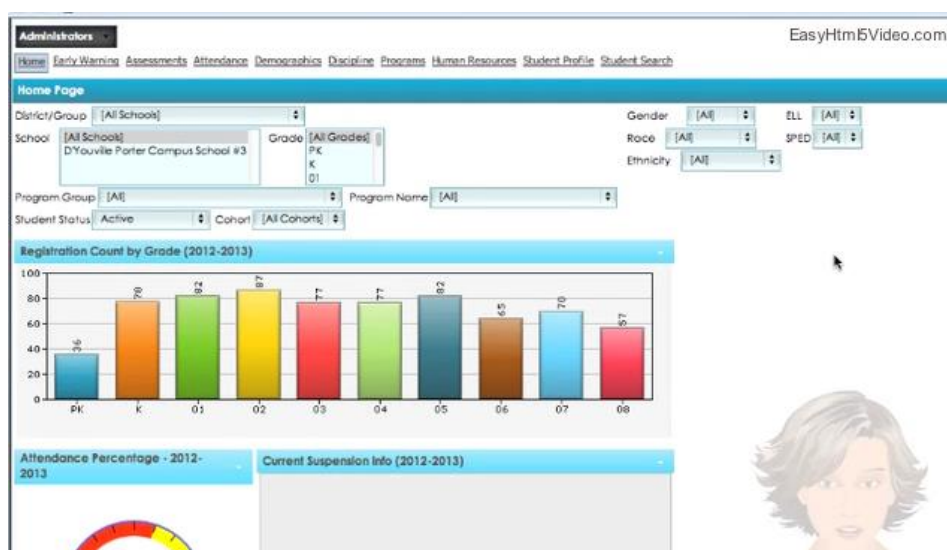


Figura 18 - Exemplo de utilização do EdVantage na Escola Elementar de Buffalo (School Buffalo, 2015).

### 3.7.1.2 DataCation

DataCation (CaseNex, 2010) é um AD desenvolvido pela CaseNex. Permite integrar dados num local centralizado, dados estes do contexto escolar e que normalmente se encontra separados fisicamente (informação dos alunos, dos Recursos Humanos, Finanças). Os dados são carregados no sistema, tendo como origem os sistemas atualmente existentes na instituição e utiliza o processo ETL para esse efeito. Desta forma consegue-se garantir a integridade e confiança dos dados(CaseNex, 2010).

Abstraindo do sistema, a empresa disponibiliza um conjunto de ferramentas que se baseia nos dados armazenados que serão apresentados de forma agregada. Dependendo do tipo de utilizador final que cada ferramenta vai ter, foram consideradas métricas *standard* essenciais que servem como base para as tomadas de decisões e alinhamento dos objetivos/alterações necessárias na instituição. No caso em que sejam necessários novos parâmetros de análise, esta também disponibiliza uma solução em específico, tornando esta solução escalável e adaptável às necessidades (CaseNex, 2010).

Em suma, na Figura 19 encontra-se um exemplo de utilização desta solução. Destacam-se as seguintes ferramentas/módulos disponibilizadas por parte da empresa:

- **Skedula - School/Teacher Management Portal:** Ferramenta utilizada por parte dos docentes que permite seguir o percurso do aluno a nível de avaliações, trabalhos de casa, evoluções e eventos que tenha associado;
- **Block sckedula - SIS:** orientada para os docentes administrativos, permitindo criar o processo de admissão do aluno, consultar e fazer a manutenção do registo (disciplinas a que está inscrito, faltas, registos de saúde, entre outros itens). Permite a criação de relatórios estatísticos sobre qualquer informação do aluno;
- **Graduation Eligibility Tracking System:** permite fazer a gestão da informação de resultados de avaliações a nível do aluno e das escolas. A partir desta informação é possível analisar a performance de uma determinada escola de um agrupamento por tipo de resultado, por etnia, por género, entre outros parâmetros;
- **Pupil Path-Parent/Student Portal:** Plataforma *online* que permite aceder à informação do aluno por parte dos pais ou mesmo pelo próprio aluno. Permite consultar as avaliações, os programas que o aluno frequenta, objetivos de aprendizagem, entre outros. Esta informação pode também ser extraída em formato de relatório, sendo este designado como um relatório de progresso do aluno;
- **Scorecard/Dashboard:** permite a criação de novas métricas associadas ao agrupamento/escolas, com o objetivo de detetar novos pontos de interesse e assim melhorar os serviços oferecidos pelas instituições de ensino. Permite analisar em tempo real as novas métricas definidas e analisar os KPI entre estudantes, agrupamentos e escolas;
- **K-12 Diagnostic Management System:** orientada para os administradores de cada agrupamento, permite auxiliar os professores nas análises que efetuam relativamente a avaliações mediocres/elevadas, identificando logo à partida os estudantes em risco (medindo o seu desempenho numa unidade curricular em particular);

- **PADS-Grade Management System:** Solução *online* que permite gerir todos os relatórios possíveis de serem gerados;
- **Assessment Management System:** Sistema responsável por recolher todos os dados possíveis sobre o desempenho dos alunos nas entregas ao longo do ano curricular;
- **Teacher Performance Record:** Sistema responsável por recolher todos os dados possíveis sobre o desempenho dos docentes em sala de aula.



Figura 19 – Exemplo de utilização da aplicação *Skedula - School / Teacher Management Portal* (CaseNex, 2010).

### 3.7.2 AD na área de ensino/educação

#### 3.7.2.1 University Data Warehouse Plus

A Universidade de Nova Iorque desenhou um sistema de apoio à decisão, presente na Figura 20, interno à instituição. Este sistema gere diferentes tipos de informação vindos de departamentos/áreas diferentes e, para isso é disponibilizado aos utilizadores finais um conjunto de ferramentas analíticas e de *reporting*, capazes de responder às necessidades operacionais (New York University, 2014). Possui quatro áreas de destaque:

- **Informação de Finanças:** este módulo foi o primeiro a avançar no que toca ao desenvolvimento e engloba a criação de relatórios e *dashboards* de informação agregada financeira. Tem como utilizadores alvo presidentes, chefes de departamento, investigadores, administradores, entre outros. Para além de usufruírem do sistema, podem ser atores importantes na criação/desenvolvimento do sistema, na medida em que podem definir as necessidades que precisam de ser satisfeitas, lista de KPI que são importantes de medir e que tipo de informação é importante reter. Numa fase posterior é criado o novo relatório ou componentes que podem ser usufruídos por qualquer um dos utilizadores que pertença a esta área;
- **Métricas por Departamento da Instituição:** apresenta informação sobre professores, investigadores e alunos, aplicando métricas definidas pelo *Provost's Council on Science and Technology*, orientado para chefes de departamento, reitores e outras instituições;

- **Informação de Recursos Humanos:** utiliza informação sobre os docentes, as posições atribuídas a cada docente na instituição, a distribuição de trabalho e os salários dos colaboradores;
- **Informação referente aos alunos:** O módulo corrente é integrado com os anteriores, focando-se apenas sobre os dados dos alunos e respetivo historial curricular.

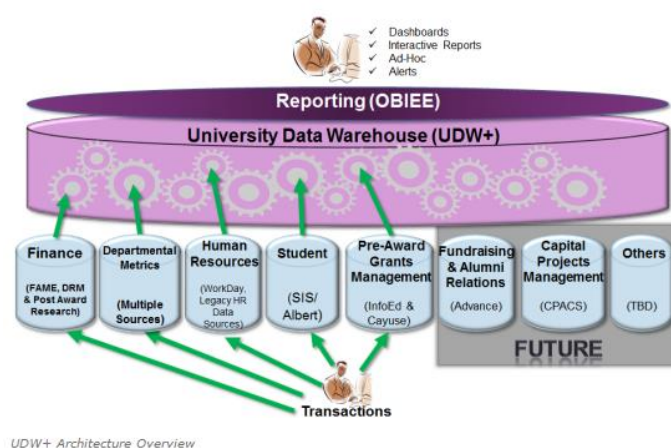


Figura 20 - Arquitetura do DW da Universidade de Nova Iorque (New York University, 2014).

### 3.7.2.2 School District of Philadelphia

O agrupamento de escolas do distrito de Filadélfia necessitava de captar e armazenar os dados curriculares uma forma simples e centralizada com o objetivo de melhorar todo o processo escolar, apoiando os docentes nas suas operações diárias. Devido ao número de escolas existentes no agrupamento (encontra-se dividido em 11 regiões /270 escolas no total) a gestão e o acesso à informação por parte de todo o ambiente docente é difícil, sendo considerada uma uma das sete maiores escolas dos Estados Unidos. Desta forma, a IBM uniu-se com a Universidade de Pensilvânia e com a própria escola com o objetivo de criar um AD que respondesse às suas necessidades, criando atualmente relatórios e *dashboards* mensais/anuais, que otimizam o processo e resumem cerca de 700 indicadores identificados como essenciais (IBM, 2006).

Este novo sistema trouxe com ele melhorias no ambiente escolar significativas: permitiu que a taxa de absentismo dos professores diminuísse em 2%; A taxa de comparência nas aulas por parte dos alunos aumentou em 57%; As suspensões diminuíram em 23%; Incidentes/conflitos graves diminuíram em 18% (IBM, 2006).

Para que todos os utilizadores tirem melhor partido das vantagens do novo sistema, foi criada também uma ferramenta de acesso, que gera os relatórios de acordo com as métricas estipuladas, dada pelo nome de SchoolStat (IBM, 2006) , orientando-se para resultados e para o desempenho da escola. Apesar da ferramenta ter a capacidade de deduzir possíveis métricas

interessantes a partir dos dados carregados no sistema, foi definida uma lista de KPI base por parte do agrupamento, destacando-se os seguintes (Patusky et al., 2007) :

- Percentagem de respostas corretas nos testes de *benchmark*. A elaboração dos testes por parte dos alunos é feita na disciplina de Matemática, Leitura e Ciências em cada 6 semanas, apenas destinada a alunos do ensino secundário que possuam uma classificação de seis a oito;
- Percentagem de alunos com um determinado nível atribuído pelo professor a nível da leitura;
- Número de alunos que se encontram na categoria dois (baixa aderência às aulas lecionadas, baixo desempenho ou determinado comportamento do aluno) do *Comprehensive Student Assistance Process (CSAP)* – este processo ajuda na identificação da necessidade de ensino especial para os alunos que se incluam nesta categoria;
- Taxa de ausência dos funcionários da instituição (docentes e não docentes) por dia;
- Taxa de aderência às aulas por parte dos alunos;
- Percentagem de suspensões.

Em suma, a informação recolhida de cada uma das escolas entra para o AD existente (introduzidos por parte do secretariado, pelos professores ou outros membros administrativos, no máximo mensalmente) e é a partir desta que os dados são agregados e apresentados em forma de relatórios, fáceis de interagir e compreender por parte do utilizador final (Patusky et al., 2007).

### **3.7.3 Indicadores utilizados no Ensino Superior**

No ecossistema do ensino superior existem várias instituições de ensino com diferentes metodologias. Segundo o documento elaborado por (Saúde et al., 2014), existem várias referências de indicadores de desempenho nas instituições de ensino superior. Estes são utilizados de acordo com o(s) objetivos(s) específicos no que toca à fórmula para efetuar o cálculo, fontes de informação e necessidade de evidências complementares.

No documento são apresentados alguns indicadores gerais utilizados no ensino superior. Foram escolhidos os indicadores que melhor se adequam ao cliente com o objetivo de preparar o sistema para analisar indicadores utilizados por outras instituições. Na Tabela 5, são apresentados esses indicadores.

Tabela 5 – Listagem de alguns indicadores retirados da FenProf (FenProf, 2012).

<b>Indicador</b>	<b>Descrição</b>
<b>Número de inscrições na instituição num determinado ano letivo.</b>	Número de inscrições de alunos por ano letivo.
<b>Número de inscritos por graus e por ano letivo.</b>	Número de inscrições de alunos por graus (Licenciatura, Mestrado, Doutoramento) e por ano letivo.
<b>Número de vagas por grau.</b>	Número de vagas de alunos por graus (Licenciatura, Mestrado, Doutoramento) e por ano letivo.
<b>Número de Inscritos por grau, por tipo de acreditação, num determinado ano letivo.</b>	Número de inscrições de alunos por Graus (Licenciatura, Mestrado, Doutoramento), por tipo de acreditação e por ano letivo.
<b>Número de vagas por grau e por tipo de acreditação.</b>	Número de inscrições de alunos por graus (Licenciatura, Mestrado, Doutoramento), por tipo de acreditação e por ano letivo.
<b>Número de inscritos por curso num determinado ano letivo.</b>	Número de alunos inscritos por curso, num determinado ano letivo.
<b>Número de vagas por curso.</b>	Número de vagas por curso, num determinado ano letivo.
<b>Número de alunos inscritos por fases num determinado letivo.</b>	Número de alunos inscritos por fases (1ª, 2ª e 3ª fase) num determinado ano letivo.
<b>Taxa de ocupação de um curso.</b>	Taxa de ocupação por curso num determinado ano letivo.
<b>Número de matriculados por fases.</b>	Número de alunos matriculados por fases num determinado ano letivo.

Para além destes, existem outros indicadores que são específicos das instituições como é o caso da LEI-ISEP. Esses indicadores serão apresentados num próximo capítulo.



## **4 Análise e Desenho**

Neste capítulo será apresentado todo o detalhe sobre a análise do problema, a solução a que chegou, bem como o desenho da sua arquitetura. Na primeira secção, é feita a análise e descrição do problema existente de acordo com o contexto desta dissertação.

Na segunda secção serão apresentados os indicadores que foram recolhidos dos relatórios do cliente. Em cada indicador é apresentado uma breve explicação do que se pretende calcular, de modo a esclarecer o leitor.

Na terceira secção é apresentada a arquitetura inicial da solução de acordo com as camadas que foram definidas para a mesma, desde as fontes de dados até ao desenho do modelo dimensional do AD.

Por fim, é apresentado a secção de avaliações de resultados em que é especificado como será avaliada a solução, de modo a comprovar que esta está de acordo com os objetivos propostos.

### **4.1 Análise e descrição do problema**

Como referido anteriormente, ao longo do tempo é necessário armazenar o histórico de inscrições dos alunos para se poder medir a performance da instituição de acordo com os indicadores utilizados pela mesma. Neste momento, esta informação está a ser obtida através de ficheiros Excel extraídos do portal da instituição e onde são efetuadas todas as análises, recorrendo às capacidades analíticas da ferramenta (Excel) de uma forma muito arcaica. Para obter os indicadores pretendidos são criadas colunas auxiliares com as transformações necessárias, aplicando por vezes fórmulas matemáticas recorrendo a várias colunas, o que facilita a ocorrência de erros e a sua propagação, tornando também complexo a resolução dos mesmos.

Também, á medida que se efetuam cada vez mais análises que recorram a vários campos auxiliares ou implique a criação de novos vai impossibilitando a manutenção da mesma, sendo que é importante referir que o Excel não foi desenhado para este tipo de soluções. A identificação de padrões e comportamentos torna-se complexo e suscetível a erro devido ao não tratamento e uniformização prévia dos dados. Aliado a estes fatores, tem-se a baixa performance disponibilizada pela ferramenta quando esta lida com uma quantidade considerável de registos. Desta maneira, não se consegue garantir a integridade e autenticidade dos dados obtidos.

Devido a todos estes fatores, surgiu a necessidade de criar uma solução capaz de colmatar todas estas falhas, permitindo o armazenamento do histórico pretendido, disponibilizar dados confiáveis e, sobretudo, que responda às necessidades do cliente. Também, pretende-se que esta seja uma solução escalável, fácil de manter e com uma maior performance, possibilitando o apoio na tomada de decisão e a apresentação de resultados de uma forma *self-service*.

## 4.2 Definição das análises a implementar

Por forma a responder às necessidades do cliente, foi necessário realizar o levantamento das análises já efetuadas por este. Assim, foi possível identificar os seguintes indicadores para disponibilizar nesta versão do AD:

1. **Alunos inscritos por unidade curricular, ano curricular, em cada ano formativo:** indica o número de alunos que estão inscritos a uma determinada unidade curricular e ano formativo, de acordo com o ano formativo em que se encontra correntemente (engloba os casos em que, por exemplo, o aluno está inscrito no segundo ano da Licenciatura e no entanto ainda se encontra inscrito em cadeiras por concluir do primeiro ano). Esta análise encontra-se ilustrada na Figura 100 , no anexo 8.1;
2. **Alunos inscritos a cada unidade curricular em cada ano formativo:** indica o número de alunos que estão inscritos em unidades curriculares, ou seja, o número de alunos que se encontram inscritos a pelo menos uma cadeira, duas ou mais. Esta análise encontra-se presente na Figura 101, no anexo 8.1.
3. **Inscrições versus o número de cadeiras em atraso:** consiste em obter informação sobre o número de alunos por número de reprovações, tendo em linha de conta o número de unidades curriculares que esteve inscrito. Esta análise encontra-se presente na Figura 102, no anexo 8.1.
4. **Unidades curriculares em atraso, por ano curricular e por ano de entrada do aluno:** Contagem do número de inscrições por número de disciplinas reprovadas, por ano

curricular e ano de entrada do aluno. Esta análise encontra-se presente na Figura 103, no anexo 8.1.

- 5. Inscrições por ano curricular versus o número de cadeiras aprovadas:** exhibe as inscrições por ano curricular versus o número de unidades curriculares que obtiveram aprovação. Esta análise encontra-se presente na Figura 104, no anexo 8.1.

Adicionalmente, foram consideradas outras possíveis análises como mais valia para a visualização do desempenho da instituição, complementado assim o leque das já existentes.

- 6. Inscrições a unidades curriculares por tipo de inscrição:** permite visualizar o número de alunos inscritos a cada disciplina no respetivo ano a que pertence, por tipo de inscrição (em atraso, em avanço ou normal). Esta análise adiciona uma mais valia na medida em que é possível visualizar e analisar o tipo de inscrições que se efetuam nas diferentes disciplinas, podendo retirar daí conclusões sobre cadeiras que não estão a obter resultados positivos.
- 7. Alunos inscritos por regime em cada turma:** permite visualizar o número de alunos em cada turma por regime em que se encontra inscrito. Esta análise adiciona valor porque, o regime a que o aluno se inscreveu pode ter influência na realização da UC. No caso de um aluno que vem de ERASMUS este provavelmente necessitará de mais apoio para se integrar e acompanhar o conteúdo da UC.
- 8. Alunos inscritos por tipo de horário por cada ano curricular:** contempla os alunos que se inscreveram por exemplo em Laboral ou Pós-Laboral de acordo com o ano curricular em que se encontra.
- 9. Alunos inscritos por ano curricular versus alunos inscritos em atraso:** possibilita ver quantos alunos inscritos a cada ano curricular, a um determinado número de unidades curriculares versus o número de unidades curriculares em atraso. Por outras palavras permite ver quantos alunos se inscreveram pelo menos a uma ou mais unidades curriculares e quantos fizeram zero ou mais inscrições em atraso. No caso de ser zero é o caso normal em que está a inscrever no período regular à unidade curricular, ou seja, não a está a repetir.
- 10. Alunos inscritos por ano curricular versus alunos inscritos em avanço:** permite visualizar quantos alunos inscritos a um determinado número de unidades curriculares, por ano curricular versus o número de unidades curriculares em avanço. É semelhante à análise anterior, diferindo no caso desta permitir visualizar unidades curriculares em avanço e não em atraso.
- 11. Alunos inscritos em cada regime por tipo de inscrição:** possibilita visualizar os alunos inscritos em cada regime por cada tipo de inscrição (avanço, atraso ou normal). Desta forma é possível verificar a influência do regime em que o aluno está inscrito no tipo de inscrições que faz.
- 12. Alunos inscritos por ano de entrada em cada regime:** permite obter uma visão geral do número de alunos inscritos em cada regime, de acordo com o seu ano de entrada.

### 4.3 Arquitetura da solução

Para esta solução foi feita uma modelação dimensional de acordo com a arquitetura definida por *Ralph Kimball* pois esta segue uma metodologia *bottom-up*, em que permite separar diferentes áreas de negócio através de *Data Marts* e adiciona a possibilidade de serem adicionados outros mais tarde sem comprometer os que já existem. Também a escolha desta arquitetura se deve ao facto de possuir uma estrutura desnormalizada que potencia a performance no carregamento e no acesso da informação no AD.

Por conseguinte, aliado a todas estas vantagens, é possível ainda acrescentar os baixos custos de manutenção da solução pois este tipo de modelação permite realizar facilmente alterações ou adicionar novas funcionalidades, tornando-a escalável. Na Figura 21 encontra-se a arquitetura concebida para esta solução.

É também importante ter em conta a dimensão de dados associado às inscrições dos alunos. Ao longo do tempo, as inscrições efetuadas, quer por novos alunos quer por alunos que já frequentam a instituição, fazem com que o histórico de inscrições cresça em grande número. Deste modo é essencial utilizar mecanismos por forma a manter a performance no acesso aos dados e não comprometer o armazenamento dos mesmos.

Não obstante, para a implementação da solução foram elegidas tecnologias Microsoft, nomeadamente *SQL Server* para o armazenamento dos dados e o *Microsoft Integration Services* para definir todo o processo de extração, transformação e carregamento de dados para o AD. Esta escolha deve-se ao facto destas ferramentas conterem capacidades e performance adequadas, suporte e documentação disponibilizado pela Microsoft, seguirem os *standards* existentes no mercado e por terem sido disponibilizadas licenças de forma gratuita ao abrigo do acordo existente entre a Microsoft e o ISEP.

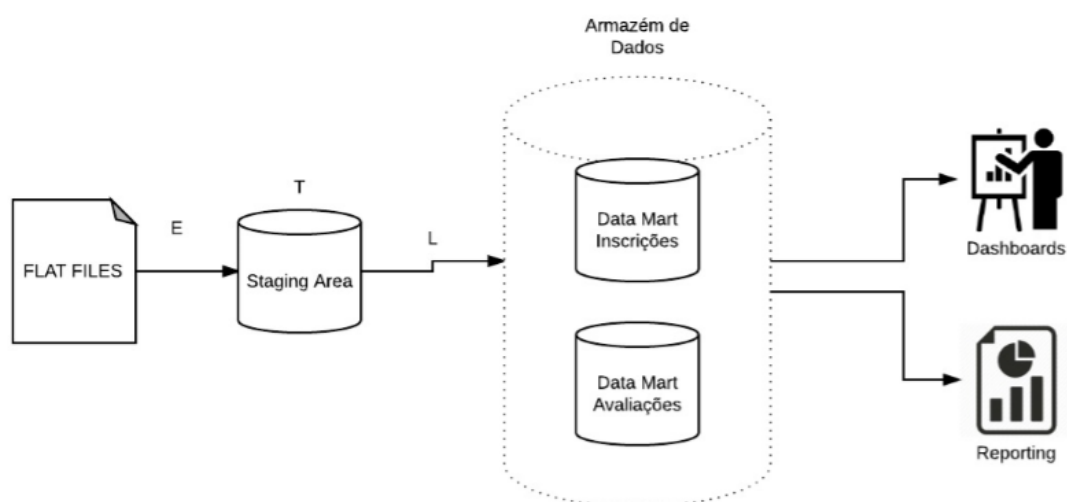


Figura 21 - Arquitetura da solução.

### 4.3.1 Fontes de dados

Relativamente às fontes de dados, apenas foram disponibilizados ficheiros Excel que contêm os dados e as análises efetuadas aos mesmos. Nestas folhas de cálculo encontram-se os dados das inscrições dos alunos às diferentes unidades curriculares ao longo dos vários anos letivos e, como foram extraídas do portal da instituição contêm conteúdo não necessário a ser armazenado no AD. De acordo com as indicações fornecidas pelo cliente foram apenas consideradas as seguintes folhas: página de relacionamento entre alunos/turmas, uma página que contém a informação de inscrição dos alunos (apenas com classificação final) e a folha com as configurações das disciplinas existentes no curso.

A folha presente na Figura 22 contém a informação sobre as turmas, as unidades curriculares, o regime e respetiva data a que o aluno se inscreveu. É importante salientar que inscrições a disciplinas sem turma podem acontecer, como por exemplo uma disciplina que um aluno não obteve aprovação num ano letivo anterior, mas mantém a nota de frequência no ano letivo presente ou no caso da UC de PESTI que não tem turma associada. Para marcar uma inscrição à disciplina sem turma é usado o caracter asterisco e para os outros casos é utilizado um traço.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	
		Regime Freq.	Data de Insc.	Ano Curr	D/N	ALGAN	APROG	AMATA	LAPR1	PRCMP	ARQCP	BDDAD	ESINF	FSIAP	LAPR3	ASIST	ALGAV	ARGSI	
NúmerC	Nome																		
1000122		Parcial	#####	3 N														3NC	
1000121		Parcial	#####	3 D														3DJ	
1000120		Parcial	#####	1 N		1NB													
1010119		Integral	#####	3 D													3DA	3DA	3DA
1010118		Integral	#####	3 D														3DH	
1010117		Parcial	#####	3 N															
1010116		Parcial	#####	3 N															
1010115		Parcial	#####	3 N				1DQ					-	2NC		3NA		3NA	

Figura 22 - Folha de cálculo com informação das turmas inscritas.

De seguida é apresentada a folha de “configurações” de disciplinas existentes no curso, e tem como objetivo estabelecer o número de ECTS por UC e em que ano curricular e semestre é lecionada.

Disciplinas	Ano	FAA	Semestre	NM	RE	EE	TE	EF	Acrescentar	ID	ECTS
ALGAN	1		1							1	5
APROG	1		1							2	5
AMATA	1		1							3	5
LAPR1	1		1							4	8
PRCMP	1		1							5	6
ARQCP	2		3							6	5
BDDAD	2		3							7	6
ESINF	2		3							8	6
FSIAP	2		3							9	5
LAPR3	2		3							10	8
ASIST	3		5							11	5
ALGAV	3		5							12	5
ARQSI	3		5							13	5
GESTA	3		5							14	4
LAPR5	3		5							15	6
SGRAI	3		5							16	5
ESOFT	1		2							17	6
LAPR2	1		2							18	8
MATCP	1		2							19	5
MDISC	1		2							20	5
PPROG	1		2							21	6

Figura 23 - Folha de cálculo com informação de ECTS das disciplinas.

Por último, na Figura 24 é presente a última fonte de dados que contempla inscrições e avaliações de alunos. Esta folha contém as várias unidades curriculares a que o aluno se inscreveu, a data de finalização das mesmas, a classificação obtida e a informação sobre creditação de competências (se existir). É de salientar que apesar desta folha conter dados de anos anteriores e não reunir toda a informação necessária, após conversações com o cliente, estes não vão ser considerados como inscrições, apenas avaliações de alunos.

Número	Disciplina	No	Data	TN	AnoEnt	AnoDisciplina	AuxNumUC	ECTS	NotaPesad	AuxTotNoti	CreditosAu
1000122	ALGAN	12	22/01/2009		00		1	5	60	60	5
1000122	ALGAV	10	02/09/2014		00		3	2	50	110	10
1000122	AMATA	11	10/02/2010		00		1	3	55	165	15
1000122	APROG	13	13/02/2007		00		1	4	78	243	21
1000122	ARQCP	11	09/02/2011		00		2	5	55	298	26
1000122	ARQSI	12	15/02/2013		00		3	6	60	358	31
1000122	ASIST	10	07/02/2013		00		3	7	50	408	36
1000122	BDDAD	12	08/02/2012		00		2	8	72	480	42
1000122	COMPA	10	12/04/2014		00		3	9	40	520	46
1000122	CORGA	12	02/07/2013		00		3	10	48	568	50
1000122	EAPLI	13	11/07/2012		00		2	11	78	646	56
1000122	ESINF	13	07/12/2012		00		2	12	78	724	62
1000122	ESOFT	12	29/06/2010		00		1	13	72	796	68
1000122	GESTA	12	28/01/2013		00		3	14	48	844	72
1000122	IARTI	10	04/07/2014		00		3	15	40	884	76

Figura 24 - Folha de cálculo com a informação de inscrição dos alunos (apenas com classificação final).

#### 4.3.2 Modelação do Armazém de Dados

Para implementar um armazém de dados é necessário efetuar previamente a sua modelação para se ter a noção das estruturas necessárias para dar resposta às análises pretendidas. Estas estruturas são baseadas nessas análises que se pretendem efetuar. É importante salientar que à medida que são necessárias novas análises a estrutura do armazém pode não conseguir responder e então será necessário adicionar novas métricas e campos que satisfaçam as novas necessidades, sem comprometer as já existentes.

De acordo com as análises anteriormente apresentadas, é exposto na Tabela 6 os campos e as métricas criadas para responder às análises. Desta forma consegue-se perceber quais os atributos utilizados e métricas a calcular com base nesses mesmos atributos. Na primeira coluna encontram-se os atributos que contextualizam as métricas, no cabeçalho da tabela encontram-se as métricas e em cada célula da tabela encontra-se o número de cada análise (número definido na secção definição das análises) que utiliza a métrica e atributo correspondente.

Tabela 6 - Matriz entre atributos de dimensões e métricas a calcular

	Número Inscrições em Atraso (AT)	Número Inscrições em Avanço (AA)	Número de Inscrições	Número Unidades Curriculares Aprovadas	Número Unidades Curriculares Reprovadas
<b>Ano Curricular</b>	9	10	1,2,4,8,9,10	5	3
<b>Ano Formativo</b>	9	10	1,2,3,4,5,6,7,8,9,10,11,12	5	3,4
<b>Aluno</b>	9	10	1,2,3,4,5,6,7,8,9,10,11,12	5	3,4
<b>Turma</b>			7		
<b>Disciplina</b>			1,2		
<b>Regime</b>			11,12,7		
<b>Tipo Inscrição</b>	9	10	2,6,9,10,11		
<b>Ano Entrada (do aluno)</b>			12		
<b>Tipo Horário</b>			8		
<b>Curso</b>	9	10	1,2,3,4,5,6,7,8,9,10,11,12	5	3,4
<b>Tipo Classificação</b>				5	3,4

Por conseguinte, para construir o modelo dimensional recorreu-se aos quatro passos chave definidos por *Ralph Kimball*:

- **Processo de negócio:** no âmbito desta dissertação o processo de negócio baseia-se em analisar informação relativa a inscrições de alunos no curso de Engenharia de Informática no ISEP.
- **Granularidade:** por forma a dar resposta às análises anteriormente explicitadas e tendo como base a matriz presente na Tabela 6 a granularidade é definida pela inscrição de um aluno a uma turma, disciplina, regime (Integral, parcial, entre outros), curso, tipo

de inscrição que efetuou (avanço, em atraso ou normal), num determinado horário (Laboral ou Pós-Laboral), no respetivo ano curricular, numa determinada data e ano letivo.

- **Identificar dimensões:**
  - **Aluno:** contém informação relativa aos dados dos alunos, como por exemplo o ano de entrada
  - **Ano Curricular:** contém os anos curriculares existentes
  - **Turma:** possui informação relativas às várias turmas existentes
  - **Disciplina:** engloba informação relativa às disciplinas
  - **Regime:** detém os regimes existentes na instituição
  - **Tipo Inscrição:** contempla os tipos de inscrição que um aluno pode efetuar
  - **Horário:** se está inscrito em Laboral ou Pós-Laboral
  - **Curso:** Cursos existentes na instituição
  - **Data:** possui detalhes sobre uma determinada data (ano, mês, dia, semestre, trimestre, ano curricular, entre outros)
  - **Tipo Classificação:** armazena detalhes sobre o tipo de classificação obtida (aprovado, reprovado, entre outros).
- **Identificar os factos:** de acordo com a granularidade definida, foi desenhada uma tabela de factos de inscrições e inscrições com métricas agregadas . Em primeiro lugar, a base que permite armazenar as diferentes inscrições é tabela de factos *Inscrição* que apresenta as diferentes chaves substitutas de acordo com a granularidade anteriormente definida. Nesta tabela não foi encontrada a necessidade de ter medidas, já que tem o único objetivo de manter o histórico de inscrições. Para responder às métricas de **Número de Inscrições Atrasadas** e **Número de Inscrições em avanço** foi criada a tabela de factos denominada *Fact\_Inscricao\_Aggreg* possui uma granularidade de acordo com as várias inscrições que um aluno efetua num curso, num determinado ano letivo, tipo de horário, ano curricular. Desta forma, possibilita obter várias métricas com base na tabela *Core*, permitindo assim responder a análises mais complexas como por exemplo **Número Inscrições de alunos com cadeiras em atraso**. Também, tem como objetivo aumentar a performance na disponibilização destas métricas, fornecendo-as já pré-calculadas.

Por fim, quanto às métricas *Número de Unidades Curriculares Aprovadas* e *Número de Unidades Curriculares Reprovadas* apenas estão presentes na Tabela 6 para evidenciar as análises que se baseiam nestas métricas e que são calculadas com base em avaliações e inscrições.

Não obstante, nesta dissertação apenas será explicado em detalhe as dimensões, tabelas de *staging* e tabelas de factos relacionadas com inscrições.

### 4.3.3 Staging Area

Depois de efetuado um estudo às diferentes fontes de dados anteriormente descritas e de acordo com as matrizes que correlacionam atributos de dimensões e métricas para dar resposta

às diferentes análises, foi criada uma estrutura de armazenamento para a *Staging Area*. Nesta fase são realizadas as transformações necessárias a estes mesmos dados por forma a serem mais tarde carregados para o AD.

De acordo com o que foi explicitado anteriormente, é apresentado na Figura 25 o modelo definido para a *Staging Area*. Por forma a responder às necessidades, foram identificadas várias entidades que contêm informação de detalhe:

- **Aluno:** contém os detalhes sobre os alunos da LEI-ISEP (Licenciatura de Engenharia de Informática no Instituto Superior de Engenharia do Porto).
- **AnoCurricular:** contém informação sobre os vários anos curriculares. Os valores possíveis são: primeiro ano, segundo ano e terceiro ano
- **Turma:** contém as várias turmas às quais os alunos estão inscritos
- **Horário:** contém os diferentes horários possíveis: diurno ou noturno
- **TipInscricao:** armazena informação relativa ao tipo de inscrição efetuada pelo aluno às diferentes unidades curriculares. Estes podem inscrever-se a uma UC em avançado (AV), em atraso (AT) ou normal (NM).
- **Disciplina:** esta entidade contém informação sobre as diferentes unidades curriculares existentes no curso. É possível obter a sigla da disciplina, um breve descritivo, número de ECTS e em que ano curricular e semestre é lecionada.
- **Regime:** possui informação sobre os diferentes regimes existentes a que os alunos se podem inscrever. Estes regimes podem ser: Erasmus, Extracurricular, Extraordinário, Gratuito, Integral, Internacional-Integral, Mobilidade, Parcial e Vasco da Gama;
- **Curso:** permite armazenar os diferentes cursos existentes na instituição. No contexto desta dissertação apenas terá o curso de Engenharia de Informática.

Por fim, para armazenar os dados relativos às inscrições foi criada a entidade *Inscricao* que reflete a granularidade definida.

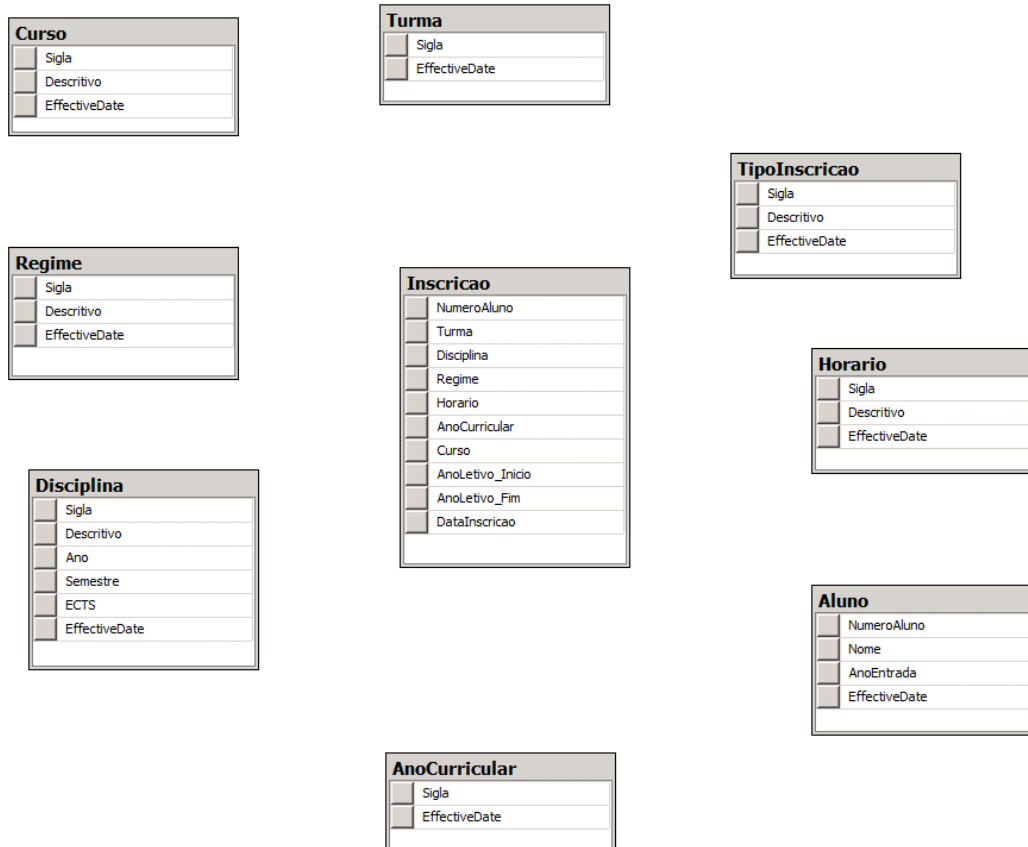


Figura 25 – Esquema de dados da *Staging Area*

Não obstante, é importante referir que não foram definidas relações entre as tabelas por forma a potenciar a velocidade com que é feita a extração e tratamento dos dados da respetiva fonte.

#### 4.3.4 Armazém de Dados

Após os dados terem sido inseridos na *Staging Area* estes ainda necessitam de ser carregados para o AD.

Como foi referido anteriormente, foi feita uma modelação de acordo com a arquitetura definida por *Ralph Kimball*. A estrutura de armazenamentos de dados é tipicamente em estrela, que contempla a tabela de factos e as várias dimensões ligadas através das chaves substitutas (*Surrogate Keys*).

Nesta fase deve haver apenas o carregamento dos dados para o AD com as respetivas chaves substitutas, no entanto, pode ainda haver alguma transformação final aos mesmos. No caso das Inscrições toda a transformação é efetuada na *staging area* havendo apenas, nesta fase do processo, o carregamento dos dados para o AD.

Na Figura 26 encontra-se o modelo em estrela definido para o AD. Este modelo tem como base o que foi delineado na secção de modelação do armazém de dados, nomeadamente na Tabela

6 e na granularidade definida para as diferentes tabelas de factos identificadas para a área de inscrições. É de notar que possui uma estrutura muito semelhante à da *Staging Area* com a adição de campos para permitir histórico, respetivas chaves substitutas em cada dimensão e um campo comum que permite indicar quando é que o registo foi carregado para o AD.

No caso específico da tabela de factos *Fact\_Inscricao\_Aggreg* esta é carregada com base na tabela de factos *Fact\_Inscricao*. Este processo é efetuado desta forma, pois os dados já se encontram devidamente estruturados na tabela de factos *Fact\_Inscricao*, sendo apenas necessário efetuar um *roll-up* à granularidade em que os dados se encontram na tabela fonte, por forma a corresponder à granularidade definida. Adicionalmente, foi definido criar um índice para maior performance na obtenção dos dados quando é feita uma consulta a esta tabela.

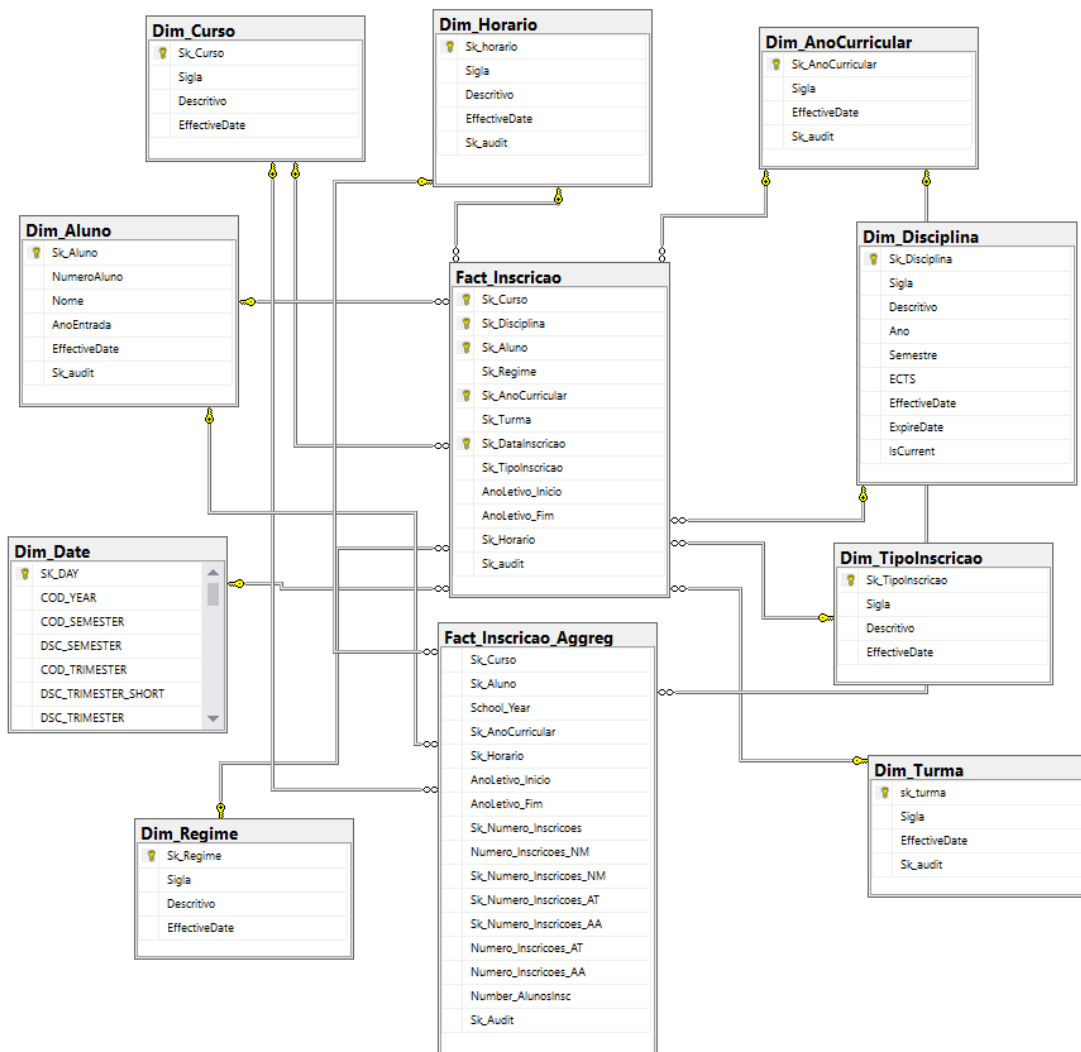


Figura 26 – Modelo Dimensional do *Data Mart* de Inscrições

Na Figura 27 está presente o modelo dimensional global do AD. É importante referir que algumas métricas no capítulo de modelação são partilhadas. Quanto às métricas, dimensões e tabelas de factos que estão presentes no diagrama e não são explicadas neste documento, estas serão contextualizadas na tese desenvolvida pela colega com o tema das avaliações de alunos. Posto isto, estas estão presentes nesta dissertação para se obter uma visão integral do AD, contextualizar a fonte das métricas que são utilizadas nas análises anteriormente apresentadas e que são calculadas com base em avaliações.

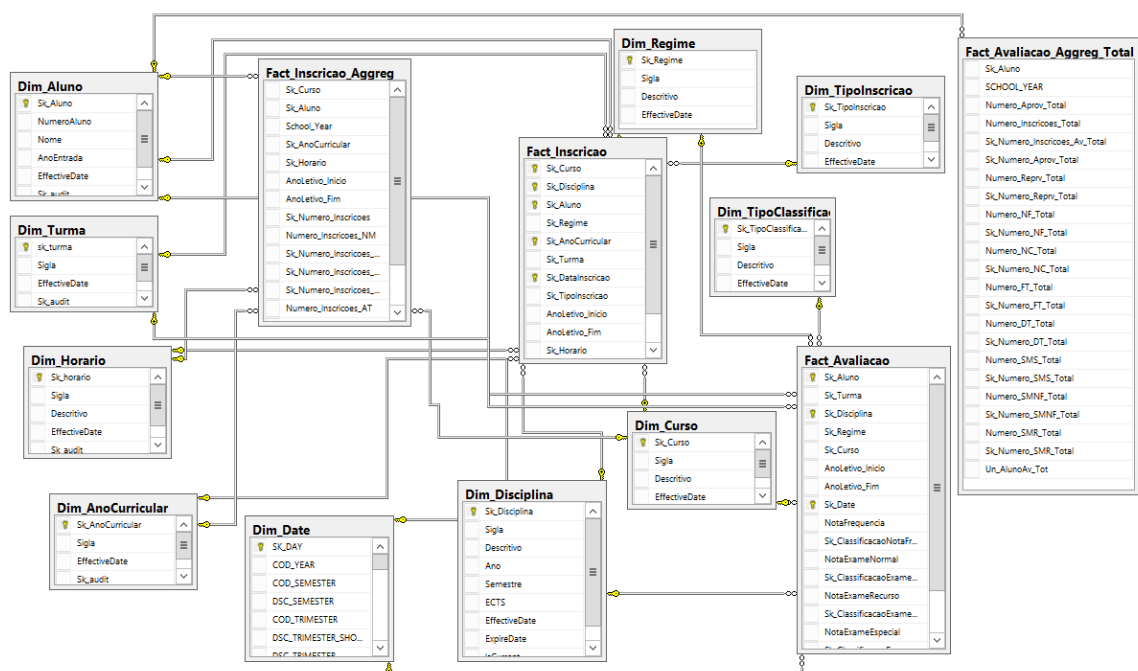


Figura 27 - Modelo de Dados Global do AD

### 4.3.5 Ferramentas a utilizar para Dashboards e Reporting

Na última camada desta solução tem-se a parte de *reporting* e *dashboards*. É uma camada bastante importante na medida em que é aqui que os utilizadores irão realizar e visualizar as análises pretendidas com base nalgumas métricas já pré-calculadas.

No âmbito desta dissertação são calculadas e apresentadas as análises já calculadas pelo cliente utilizando a ferramenta já adotada pelo mesmo, Microsoft Excel, e apresentam-se algumas no *Microsoft Power BI*.

A escolha do *Power BI*, como ilustrado na Figura 28, deve-se pelas capacidades analíticas disponibilizadas por esta nova ferramenta, bem como ser disponibilizado de forma gratuita para a sua utilização. Também o facto de permitir ser acessível por vários dispositivos via aplicação ou página web é uma mais-valia que foi tida em consideração.

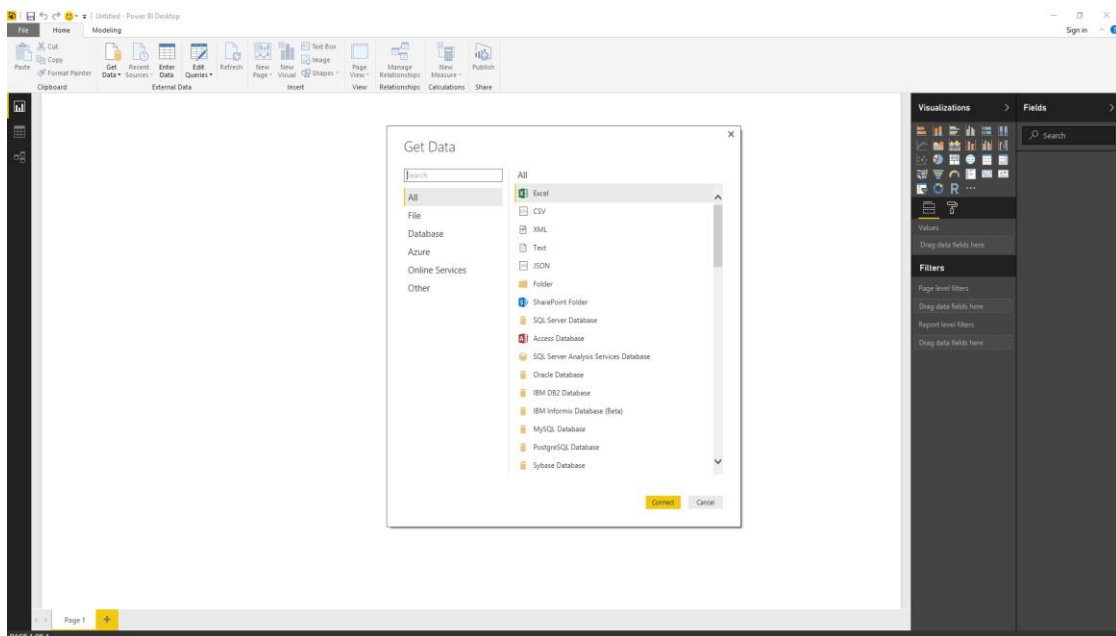


Figura 28 – Power Bi versão *Desktop*

#### 4.3.6 Manutenção de Histórico

Para a realização de análises faz-se necessário manter histórico de dados no AD. A manutenção de histórico requer uma maior capacidade de processamento e processos de ETL mais complexos que pode provocar lentidão quando são executadas as análises pretendidas.

Posto isto, por forma a adequar a manutenção de histórico para cada dimensão e tabela de factos é avaliado se é necessário ou não manter histórico de acordo com as análises que se poderão realizar.

No caso das dimensões, a nível de SCD, temos:

- **Aluno:** nesta dimensão não é necessário manter histórico da informação do aluno pois não são efetuadas análises com base na informação deste, e o ano de entrada do mesmo mantém-se sempre.
- **Turma:** no caso da dimensão turma só podem ser adicionadas novas turmas, ou no caso de nesse ano as turmas não serem preenchidas estas não são utilizadas. Se houver alguma alteração a nível dos atributos da mesma, estes não têm relevância para as análises efetuadas no momento ou no futuro. Deste modo, apenas basta utilizar SCD do tipo 1.
- **Ano Curricular:** nesta dimensão, apenas contém informação relativa aos anos curriculares que o curso possui. Estes não mudam com muita frequência, mas caso mudem apenas podem ser adicionados ou retirados anos ao curso. Para além disso, o rastreamento de mudanças nesta dimensão não é algo que seja contemplado nas

análises pretendidas ou em futuras análises sendo que, desta forma, não é necessário armazenar histórico nesta dimensão.

- **Horário:** apesar de haver análises que incidem sobre os diferentes tipos de horários, estes não alteram com muita frequência. Após algumas conversações com o cliente, este não achou relevante efetuar manutenção de histórico.
- **Data:** dimensão que armazena informação detalhada de datas ao longo dos vários anos. Nesta não é aplicável qualquer tipo de SCD.
- **Disciplina:** dimensão que armazena informação relativa às unidades curriculares existentes. Nesta, é importante manter histórico sobre as mudanças eventuais de ECTS, pois estes afetam no cálculo da média ao longo dos vários anos letivos. Desta forma, para guardar essas alterações foi escolhido SCD do tipo dois.
- **Tipo Inscrição:** dimensão que, como próprio nome indica, armazena o tipo de inscrições que um aluno pode efetuar na instituição. A manutenção de histórico nesta dimensão não é necessário pois o tipo de informação armazenado, ou é deixado de utilizar ou é adicionado um novo tipo de inscrição sendo que essa alteração não contém valor relevante para as análises efetuadas.

Por fim, na tabela *Fact\_Inscricao* é feito o armazenamento ao nível mais detalhado de inscrições, contendo todo o histórico necessário para consultar mais tarde sem quaisquer métricas calculadas contendo apenas a data de inscrição, sendo assim denominada de *FactLess Fact Table*. Já a tabela de factos *Fact\_Inscricao\_Aggreg* é um *snapshot* realizado ao longo dos vários anos letivos de acordo com a granularidade definida. Estas tabelas são denominadas de *Aggregate Fact Tables* porque é feito um *rollup* dos registos contidos na tabela *Fact\_Inscricao* para uma melhor performance na obtenção dos resultados.

## 4.4 Avaliação de Resultados

Avaliação de resultados pode ser definida como um conjunto de critérios definidos que serão avaliados mais tarde quando a solução for preenchida com os dados devidamente tratados. Estes critérios, de acordo com o contexto, devem ser definidos segundo possíveis hipóteses que se pretendem alcançar, detalhar por passos as possíveis condições a serem alcançadas, tipo de medidas utilizadas (independentes ou dependentes), equipamento necessário e o plano de análise de dados (Ferreira, 2015).

Como já foi referido, no contexto desta dissertação pretende-se validar a veracidade dos dados importados no AD através da comparação dos indicadores calculados na nova solução com os da que existe atualmente. Também, é importante verificar se todos os registos foram carregados e não há perda de dados. A nível de medidas utilizadas, as grandezas utilizadas nesta dissertação são o tempo que leva a realizar uma ronda completa de ETL e a satisfação do utilizador quanto à utilização da ferramenta analítica e de criação de novos relatórios (tendo em conta quando se compara os resultados obtidos atualmente com a antiga solução pode haver um desvio não significativo).

No que toca à grandeza tempo utilizada para medir a performance da solução, é importante salientar que se considera cada ronda de ETL desde o processo de extração da *staging area* até ao carregamento de dados no AD.



## 5 Implementação

Neste capítulo será descrito todo o detalhe de execução do projeto para possibilitar uma melhor compreensão de como este foi implementado.

Inicialmente serão apresentadas as configurações iniciais efetuadas no projeto, bem como aspectos relevantes de instalação dos ambientes de desenvolvimento e versionamento de código.

De seguida, é descrito o processo de carregamento desde da área de *staging* ao próprio AD possibilitando uma melhor compreensão do fluxo de dados e do processo implementado.

Na última fase é exposto o cubo analítico de inscrições e Inscrições/Avaliações. Através deste são realizadas análises de uma forma mais intuitiva, *self-service* e sem necessitar de ter conhecimentos técnicos de *SQL*.

### 5.1 Ambiente de desenvolvimento e configurações iniciais

Para iniciar o desenvolvimento do AD foi necessário instalar o seguinte Software, presente na Tabela 7:

Tabela 7 - Software Instalado

Software	Descrição
<b>Microsoft Visual Studio Professional 2013</b>	Ferramenta de desenvolvimento de todo o processo ETL, Base de dados e Cubos
<b>Microsoft SQL Server Data Tools (SSDT) 2013</b>	Extensão do Microsoft Visual Studio que disponibiliza ferramentas de desenvolvimento adicionais para base de dados SQL Server

<b>Microsoft SQL Server Data Tools for Business Intelligence (SSDT-BI) 2013</b>	Extensão do Microsoft Visual Studio que disponibiliza ferramentas de desenvolvimento de soluções de <i>Business Intelligence</i>
<b>Microsoft SQL Server 2014</b>	Base dados onde fica alocada a solução
<b>Microsoft Office 2013 (Excel)</b>	Análises do cliente replicadas
<b>Microsoft Power BI</b>	Análises e Dashboards
<b>Microsoft Team Foundation Server 2013</b>	Versionamento do projeto

A instalação de cada software foi efetuada de acordo com as configurações por defeito do fabricante.

No que toca a acessos ao servidor, foi configurado o acesso RDP (*Remote Desktop*) do Windows Server para efetuar o *deploy* da solução e configurações locais do ambiente.

Para versionamento da solução desenvolvida foi instalado o *Team Foundation Server 2013* da *Microsoft*, como se pode visualizar na Figura 29. Neste, foi criada uma coleção denominada *BiTese* que contém o projeto de desenvolvimento.

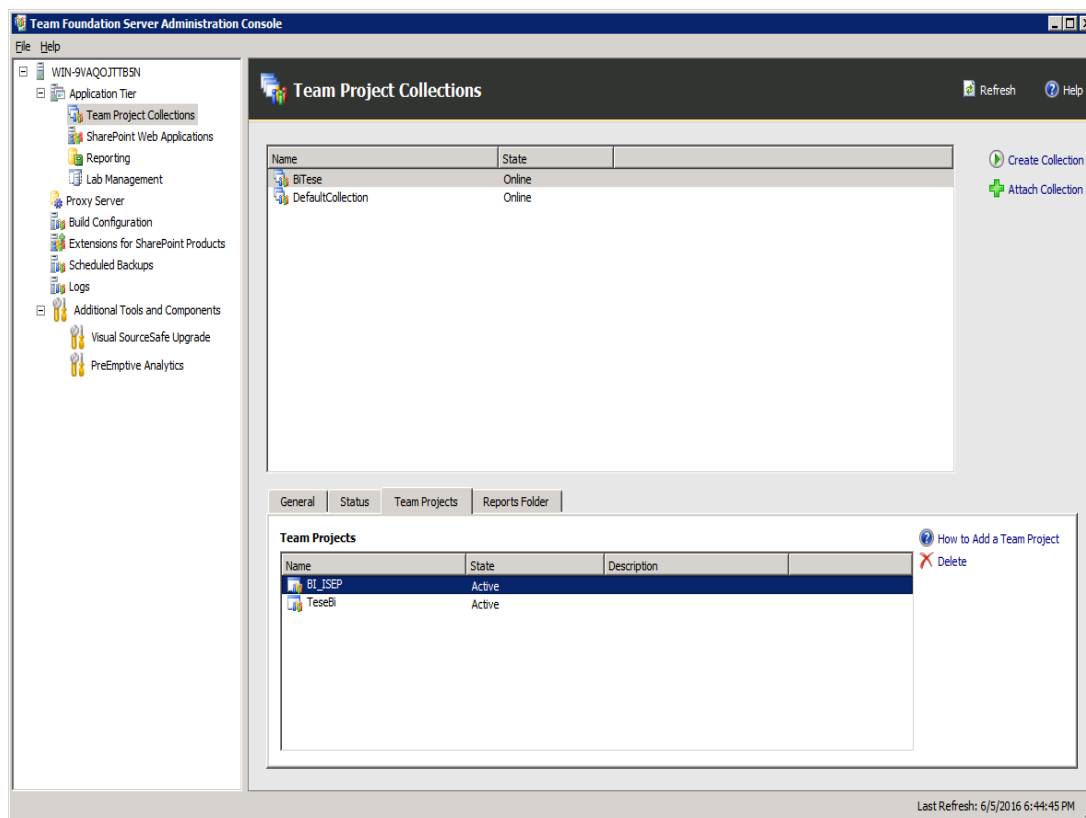


Figura 29 - Coleção de Projetos no Team Foundation Server

Por conseguinte, para permitir um desenvolvimento no exterior do servidor foi necessário configurar acessos externos para as bases de dados, através de autenticação *SQL Server Login*, como é possível verificar na Figura 30, e a abertura da porta correspondente na Firewall do sistema operativo.

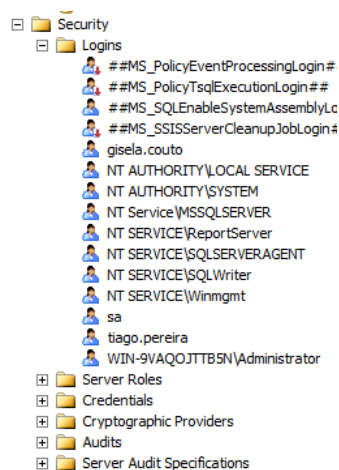


Figura 30 - Local de configuração de logins

Para desenvolver esta solução foram definidas algumas regras:

- Mapeamento da área de trabalho (*workspace*) do projeto alocado no TFS (*Team Foundation Server*) deve ser feito para a unidade “C:\” do computador
- Ficheiros que alimentam a área de *Staging* devem estar localizados no diretório “C:\data\_source”.

Estas regras servem para garantir que quem desenvolve nesta solução tem os mesmos mapeamentos.

Por último, nas próximas subsecções serão apresentadas as configurações específicas dos ambientes, nomeadamente a estrutura e bases de dados criadas para o projeto.

### 5.1.1 Acesso às Fontes de Dados

Para aceder à informação contida nas fontes de dados foram criados *Linked Servers*<sup>4</sup>. Recorre-se a esta funcionalidade para colmatar a dependência e as falhas inerentes ao componente do Microsoft Excel disponibilizado no *Microsoft SSIS*.

Por conseguinte, outra vantagem de utilizar esta funcionalidade é uma maior facilidade de troca do ficheiro de fonte de dados comparativamente à configuração que é necessária efetuar no componente Excel existente no *Microsoft SSIS*, por forma a realizar a extração dos dados da respetiva fonte.

Posto isto, como ilustrado na Figura 31 , para satisfazer as necessidades foram criados os seguintes *Linked Servers*:

<sup>4</sup> *Linked Servers*: permitem aceder e executar comandos SQL fora da instância local do SQL Server, como por exemplo aceder a uma base de dados Oracle, ficheiros Excel, entre outros.

- **FichasAluno:** conecta-se à fonte de dados que contém informação sobre avaliações de alunos já finalizadas este ano e em outros anos.
- **InfoEstatica:** contém informação relativa a configurações do curso, como por exemplo as turmas, tipos de classificação, tipos de inscrição, entre outros.
- **NotasAlunos:** possui informação relativamente às inscrições e avaliações do ano letivo em questão. Também, possui configurações sobre as disciplinas como anteriormente detalhado no subcapítulo de fontes de dados.

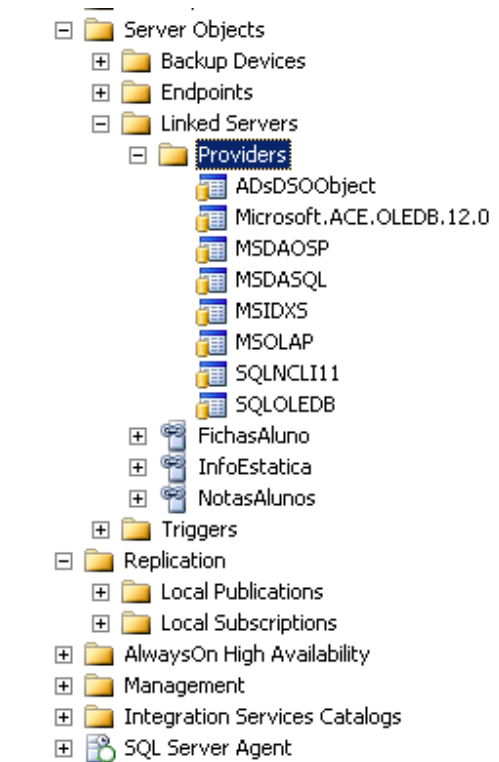


Figura 31 - Linked Servers de acesso à fonte de dados

Por fim, é de salientar que para esta funcionalidade opere é necessário instalar o *driver* “Microsoft.ACE.OLEDB.12.0”. Este é o driver que permite conectar entre a fonte de dados (Ficheiros Excel) e o *SQL Server*.

### 5.1.2 Mecanismo de deploy

Para efetuar a publicação (*deploy*) da solução, as ferramentas utilizadas oferecem uma interface bastante intuitiva/*user-friendly*. Com base no estudo efetuado no estado da arte, o *plugin SSDT* do Visual Studio para além de permitir desenhar e realizar controlo de versões de todas as estruturas de dados, procedimentos, índices, entre outros, também disponibiliza a funcionalidade de publicar os mesmos projetos nas bases de dados criadas, como ilustrado na Figura 32.

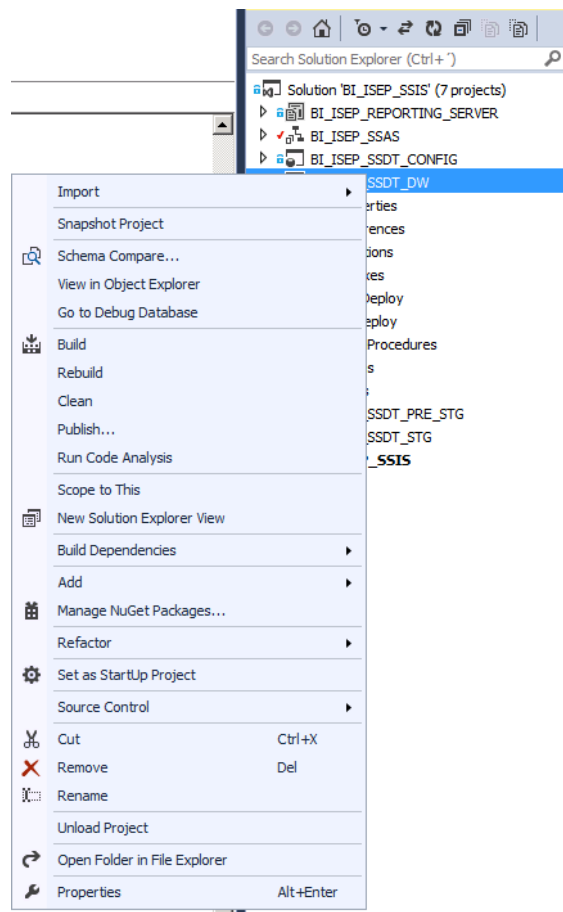


Figura 32 - Funcionalidade de publicação da solução

Conjuntamente, permite publicar configurações com valores pré-definidos ou correr scripts de criação de regras após a estrutura de dados ser publicada. Esta funcionalidade chama-se *PostDeploy* e está ilustrada na Figura 33. Também, permite efetuar o oposto, ou seja, correr scripts antes das estruturas de dados serem publicadas, recorrendo desta vez à funcionalidade de *PreDeploy*.

```

BI_ISEP_SSDT_DW_1.publish.sql * -> X
SET NUMERIC_ROUNDABORT OFF;

GO

:setvar BI_ISEP_SSDT_STG "DEV_STG"
:setvar DatabaseName "DEV_DW"
:setvar DefaultFilePrefix "DEV_DW"
:setvar DefaultDataPath "C:\Program Files\Microsoft SQL Server\MSSQL12.MSSQLSERVER\MSSQL\DATA\"
:setvar DefaultLogPath "C:\Program Files\Microsoft SQL Server\MSSQL12.MSSQLSERVER\MSSQL\DATA\"

GO
:on error exit
GO
/*
Detect SQLCMD mode and disable script execution if SQLCMD mode is not supported.
To re-enable the script after enabling SQLCMD mode, execute the following:
SET NOEXEC OFF;
*/
:setvar __IsSqlCmdEnabled "True"
GO
IF N'$(__IsSqlCmdEnabled)' NOT LIKE N'True'
    BEGIN
        PRINT N'SQLCMD mode must be enabled to successfully execute this script.';
        SET NOEXEC ON;
    END

GO
USE [$(DatabaseName)];

GO
CREATE TABLE [dbo].[PRE_DIM_DATE](
    [SK_DAY] AS CONVERT(INTEGER, CONVERT([char](4), datepart(year, [DATE]), (0))
        + SUBSTRING('0' + CONVERT([char](2), datepart(month, [DATE]), (0)), LEN(CONVERT([char](2), datepart(month, [DATE])), 2)
        + SUBSTRING('0' + CONVERT([char](2), datepart(day, [DATE]), (0)), LEN(CONVERT([char](2), datepart(day, [DATE])), 2))), 2),
    [COD_YEAR] AS (datepart(year, [DATE])),
    [COD_SEMESTER] AS (CONVERT(char(6), CONVERT([char](4), datepart(year, [DATE]), 0) +
        (CASE WHEN datename(quarter, [DATE])=1 OR datename(quarter, [DATE])=2
            THEN 'S1'
            ELSE 'S2' END))),
    [DSC_SEMESTER] AS (CONVERT(char(10), CONVERT([char](4), datepart(year, [DATE]), 0) +
        (CASE WHEN datename(quarter, [DATE])=1 OR datename(quarter, [DATE])=2
            THEN ' | 1S'
            ELSE ' | 2S' END))))).

```

Figura 33 -Exemplo de script de *deploy*

Em suma, é importante de salientar que ao efetuar uma publicação, esta ferramenta consulta o que existe atualmente na base de dados e verifica se necessita de criar ou alterar estruturas de dados, procedimentos, vistas, entre outros. No caso das tabelas conterem dados, este faz a gestão automática de criar uma tabela com a nova estrutura e migra os dados da antiga estrutura para esta nova, sendo que no fim apaga a estrutura anterior e renomeia a que foi criada para o nome pretendido.

### 5.1.3 Envio de Emails

Quando se manipula um volume de dados considerável é necessário ter uma solução de ETL complementada com um serviço de avisos via email para uma melhor monitorização e aviso de eventuais problemas que ocorram.

Para facultar esta funcionalidade é necessário efetuar a configuração do servidor de emails no próprio SQL Server. Para ativar esta funcionalidade, como é possível visualizar na Figura 34, é necessário criar uma conta de email com as configurações específicas do provedor de serviços.

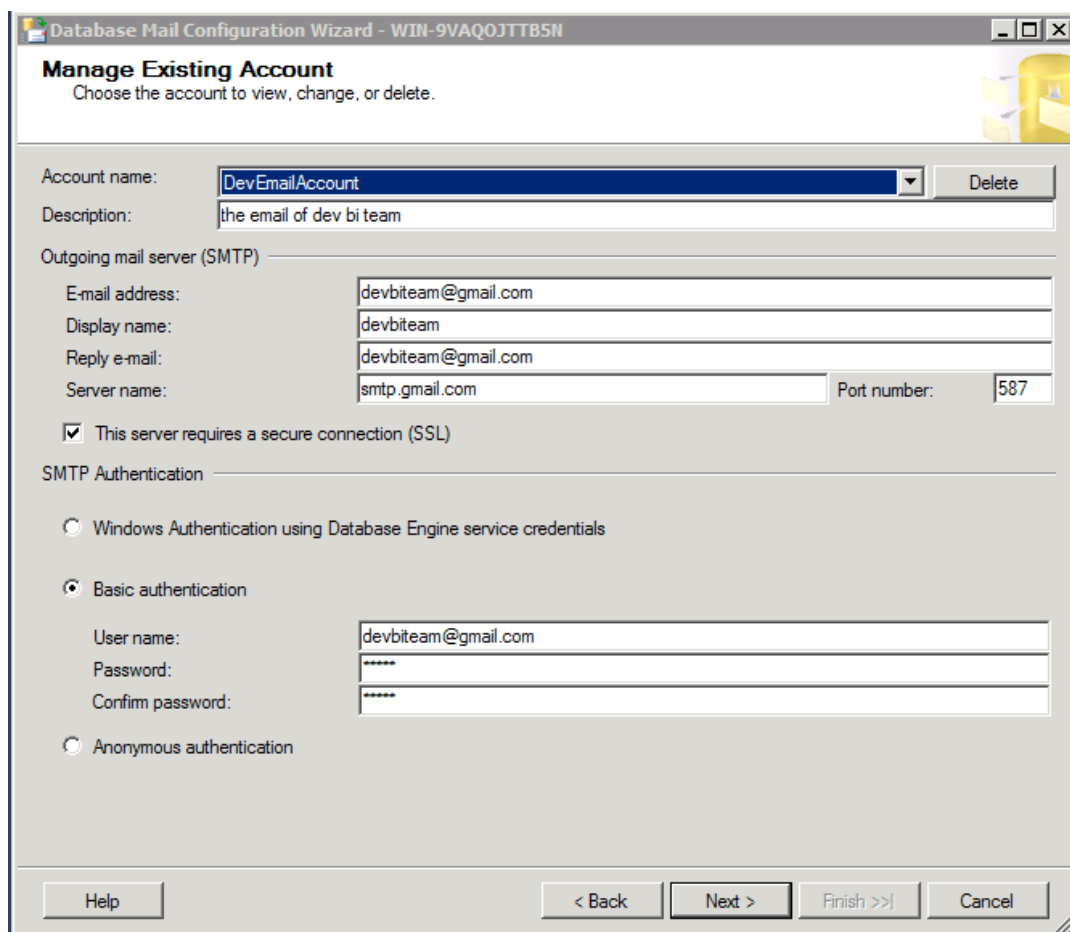


Figura 34 - Configurações Email

Por fim, é necessário criar um perfil de e-mail público para possibilitar a utilização das configurações anteriormente referidas. A criação deste, nasce da necessidade de controlar o acesso a configurações de e-mail efetuadas, como ilustrado na Figura 35.

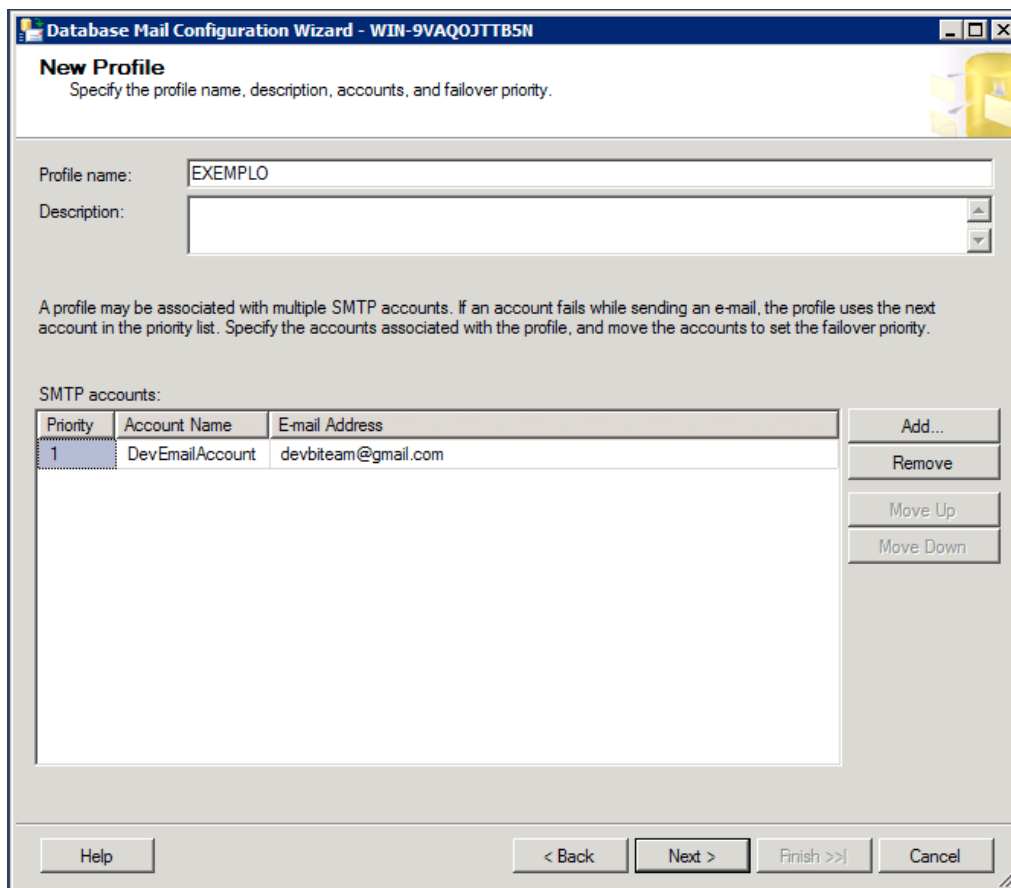


Figura 35 - Configuração do perfil de email

#### 5.1.4 Estrutura do projeto

O desenho da estrutura de um projeto de desenvolvimento é fulcral para o sucesso do próprio desenvolvimento. Este armazena todas as estruturas de dados, lógica de negócio envolvida e a sua organização metódica agiliza o desenvolvimento de novas funcionalidades e a manutenção de existentes.

Desta forma, para agilizar o versionamento e localização das funcionalidades dividiu-se o projeto em quatro subprojectos, como ilustrado na Figura 36:

- **BI\_ISEP\_SIS:** Projeto de Business *Intelligence* onde está alocado todo o processo de ETL que por sua vez contém todo o workflow e lógica de negócio associado desde a área *Staging* ao AD.
- **BI\_ISEP\_SSAS:** Projeto de Business *Intelligence* que comporta a parte analítica do projeto. Aqui é onde está desenvolvido o modelo que suporta o cubo analítico que mais tarde servirá para construir as análises pretendidas pelo cliente.
- **BI\_ISEP\_SSDT\_STG:** Projeto de Base dados SQL *Server* que contém todas as estruturas de dados e procedimentos associados à área de *Staging*.

- **BI\_ISEP\_SSDT\_DW:** Projeto de Base dados SQL Server que contém todas as estruturas de dados e procedimentos associados à área do AD.

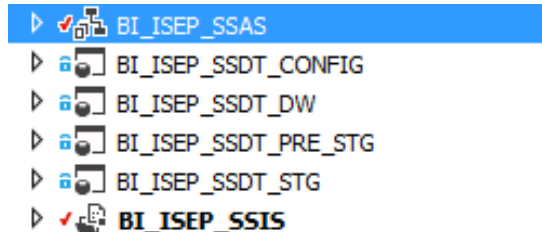


Figura 36 - Estrutura dos projetos

Como se pode verificar na Figura 37, encontram-se diferentes projetos de base dados. Estes encontram-se estruturados desta forma para incluírem apenas as estruturas de dados, procedimentos, vistas e configurações específicas de acordo com o seu destino. Desta forma, a alteração, criação e manutenção de estruturas no projeto torna-se mais intuitiva e simples de realizar.

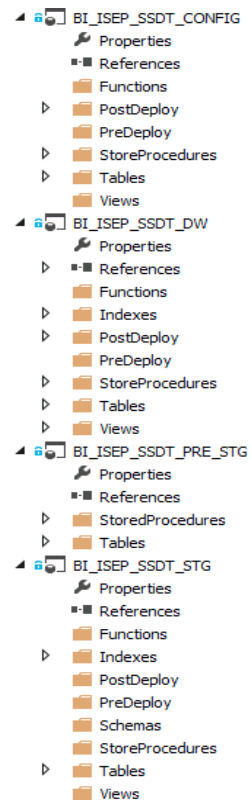


Figura 37 - Estrutura principal dos projetos de Base Dados

Por conseguinte, para alocar toda a estrutura relativa ao processo de ETL foi criado o projeto **BI\_ISEP\_SSIS**, como apresentado na Figura 38. A lógica de negócio e área a que se destina estão contemplados no nome do próprio pacote e no seu interior, ou seja, cada pacote trata do domínio específico em que se insere. Posteriormente será apresentada, em mais detalhe, a metodologia utilizada para o carregamento dos dados.

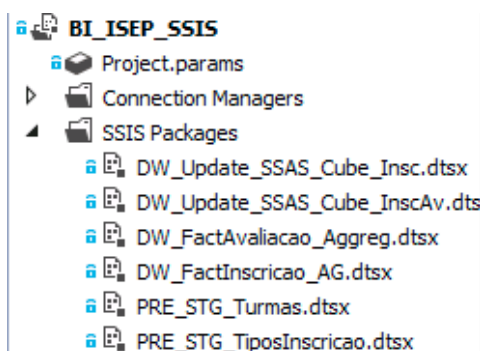


Figura 38 - Estrutura do Projeto de ETL

Por fim, para armazenar os cubos analíticos foi criado o projeto de *Analysis Services* (**BI\_ISEP\_SSAS**). Neste, como se pode constatar na Figura 39, são armazenadas as estruturas dos cubos e modelos que contém as relações entre as dimensões e tabelas de factos criadas, permitindo assim criar análises de uma forma mais *self-service*.

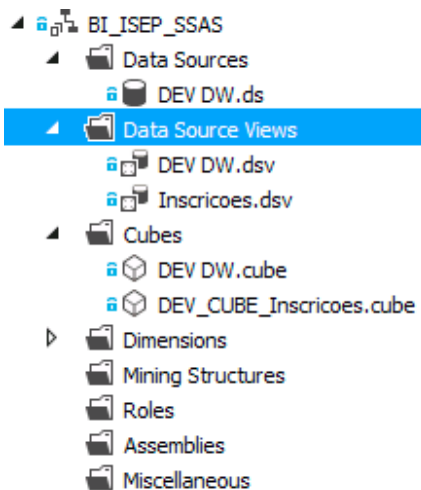


Figura 39 - Projeto *Analysis Services*

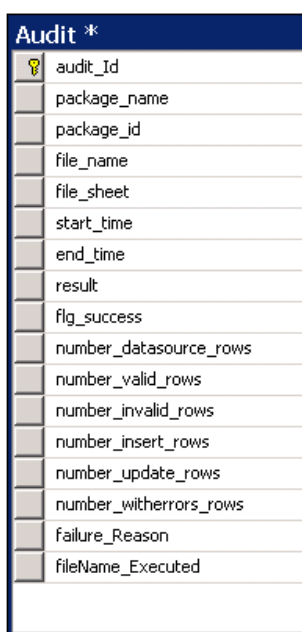
## 5.2 Auditoria de dados

Para controlar com um maior detalhe os dados importados para a área de *staging* e do AD, foi criado um processo de auditoria de dados.

Neste, como se pode visualizar na Figura 40, é armazenada informação detalhada sobre os registos que foram selecionados, inseridos, atualizados ou, em caso extremo, eliminados num

determinado pacote, durante um determinado carregamento efetuado, com uma data de início e data de fim. Também, são registados os erros que possam eventualmente ocorrer nos pacotes que fazem parte do processo de ETL, para um melhor auxílio na resolução dos mesmos.

É de salientar que, em cada pacote executado é inserido um novo registo de auditoria com um identificador único da sua execução.



Audit *	
audit_Id	
package_name	
package_id	
file_name	
file_sheet	
start_time	
end_time	
result	
flg_success	
number_datasource_rows	
number_valid_rows	
number_invalid_rows	
number_insert_rows	
number_update_rows	
number_witherrors_rows	
failure_Reason	
fileName_Executed	

Figura 40 – Esquema da tabela de auditoria

### 5.2.1 Mecanismo de extração de dados

Todos os dados inseridos no AD são auditados de acordo com a chave primária da tabela de auditoria, para identificar inequivocamente os registos importados para a tabela destino presente no próprio AD.

Desta forma, audita-se quantos registos foram inseridos, atualizados durante o processo de carregamento de uma determinada tabela para posterior análise. Assim, permite um melhor controlo de erros ou apenas controlo de volume de dados importados a cada carregamento versus o tempo que leva a carregar, possibilitando assim determinar problemas de performance.

Para exemplificar melhor este processo está presente na Figura 41 um conjunto de registos inseridos na dimensão aluno.

Sk_Aluno	NumeroAluno	Nome	AnoEntrada	EffectiveDate	Sk_audit
982	1100369	Unknown	2010	2016-08-01 22:26:18.2410000	1088
983	1110509	Unknown	2011	2016-08-01 22:26:18.2410000	1088
984	1130121	Unknown	2013	2016-08-01 22:26:18.2410000	1088
985	1970412	Unknown	1997	2016-08-01 22:26:18.2410000	1088
986	1110443	Unknown	2011	2016-08-01 22:26:18.2410000	1088

Figura 41 - Exemplo de registos inseridos na dimensão aluno

Como se pode verificar na Figura 41, os registos possuem um *Sk\_audit* que corresponde ao identificador único inserido para a execução específica do pacote SSIS da dimensão aluno. De seguida, como se pode comprovar pela imagem da Figura 42 o identificador único de auditoria 1088, que corresponde ao campo *audit\_id*, está contido na execução do pacote de SSIS da dimensão aluno em que foram inseridos 1190 registos no total.

audit_id	package_name	start_time	end_time	result	number_datasource_rows	fig_success	number_insert_rows
1088	DW_DimAluno	2016-08-01 22:26:43.0000000	2016-08-01 22:26:45.5570000	Success	1190	1	1190

Figura 42 – Registo auditado para o conjunto

Desta forma, permite rastrear os registos inseridos numa determinada ronda de ETL de acordo com a data em que ocorreu e o tempo que demorou a executar o processo.

## 5.2.2 Mecanismo de Rollback em caso de falha

Nos pacotes que contém a lógica de negócio embebida, podem ocorrer falhas e ser necessário eliminar a informação que foi inserida no carregamento de dados, para que no próximo este não duplique e/ou torne inconsistente os mesmos. Como exemplo disto, tem-se o carregamento de registos para uma tabela de factos em que, se pode determinar que um campo de data que esteja em falta é crítico para a inserção de registos.

Posto isto, no caso este estar em falta marca-se o estado do processo de carregamento deste pacote como falhado e, no passo seguinte, a remoção dos registos inseridos é feita através do identificador único previamente criado de auditoria e inserido conjuntamente com os registos no destino. Por forma a ilustrar este processo pode consultar-se a Figura 43.

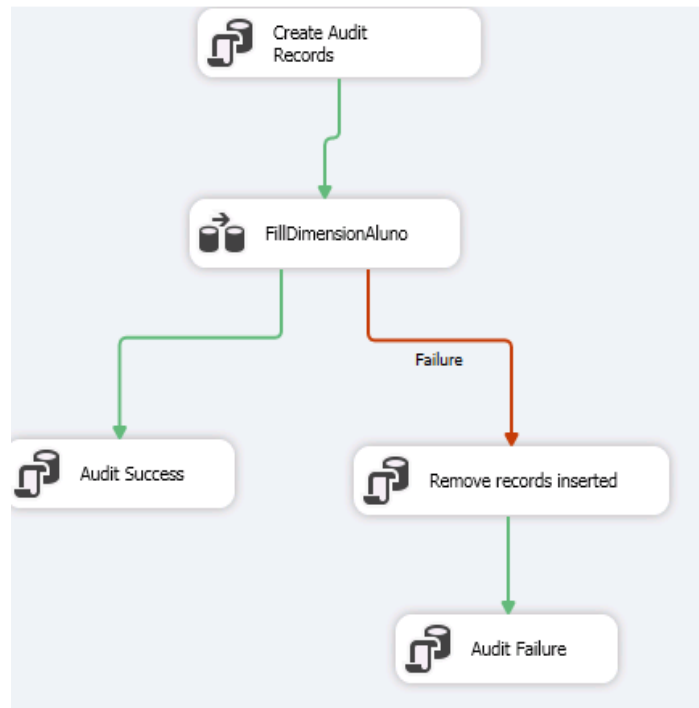


Figura 43 – Processo de remoção de registos na dimensão aluno

Por fim, é de evidenciar que este processo de remoção acontece no componente *remove records inserted*. Como anteriormente explicitado, este baseia-se no identificador único de auditoria criado no início do processo, no componente *create audit records*, selecionando os registos inseridos nesta execução. Desta forma, garante-se que a dimensão ou tabela de factos não contém inconsistências nos dados inseridos, permitindo reiniciar o processo sem quaisquer riscos.

### 5.3 Processo de ETL

Como anteriormente explicitado, o processo de ETL consiste em extrair os dados existentes das respetivas fontes, transformá-los de acordo com as regras de negócio e, por fim, serem carregados no AD para serem analisados.

De acordo com esta metodologia, será apresentado em primeiro lugar a área de *staging* que abrange o processo de extração e transformação de dados. Neste, será apresentado o processo utilizado para extrair e transformar os dados nos vários pacotes.

Por último, será apresentado o processo de carregamento de dados para o AD. Durante a apresentação deste processo serão exibidos os detalhes existentes no mesmo, tanto ao nível das dimensões como ao nível das tabelas de factos.

### 5.3.1 Staging Area

A fase inicial do processo de ETL decorre na *Staging Area* onde é efetuada a extração e tratamento dos dados oriundos da respetiva fonte.

Numa visão geral, como é possível consultar na Figura 44, é efetuada uma limpeza à tabela de *Staging* que é alvo do processo corrente, para que esta não contenha dados de uma ronda de ETL anterior. De seguida, é efetuada a extração dos dados da respetiva fonte, tratamento dos mesmos e são inseridos na tabela destino, de acordo com o domínio a que pertencem. No fim do processo, este insere um registo na tabela de auditoria com os dados retirados no processo executado e o estado com que terminou.

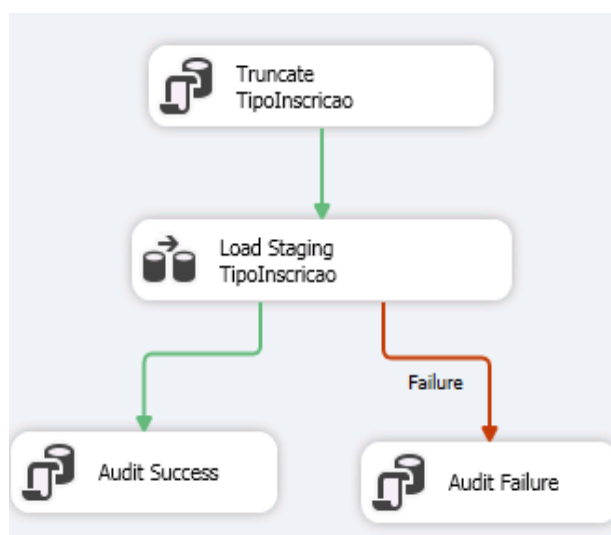


Figura 44 - Exemplo de *Staging Area*

De acordo no processo ilustrado na Figura 46, inicialmente é feita uma limpeza da tabela destino para esta não inserir dados de uma execução anterior.

No passo seguinte, é feita a extração e transformação de dados necessário para corresponder às necessidades posteriores no AD. Neste processo está também embebido o tratamento de eventuais erros que possam ocorrer em componentes que convertem dados e/ou efetuam transformações. Esta metodologia tem como objetivo proteger o processo de ETL de eventuais falhas e permitir ao máximo a sua disponibilidade, ou seja, continuar a processar outros pacotes de *Staging*. Por forma a enriquecer o processo e possibilitar um melhor rastreamento dos registos com problemas, foram criadas tabelas denominadas *DQP (Data Quality Problem)* para cada tabela de *Staging* em que cada uma contém uma estrutura semelhante à tabela de *Staging* correspondente com a adição de novas colunas, nomeadamente *DQP*, *Component*, *ErrorCode* e *ErrorDesc*. A coluna *DQP* contém uma mensagem específica do possível problema sendo que a coluna *Component* armazena o nome do componente onde ocorreu o erro. No caso das colunas *ErrorCode* e *ErrorDesc* contém o código do erro e a mensagem de erro respetivamente, que são facultados pelo componente do SSIS. Na figura Figura 45 encontra-se um exemplo da estrutura de uma tabela *DQP*.

TipoInscricaoDQP	
Sigla	
Descritivo	
EffectiveDate	
DQP	
Component	
ErrorCode	
ErrorDesc	

Figura 45 – Estrutura da tabela *DQP TipoInscricao*

No exemplo de extração e tratamento de dados presente na Figura 46, que corresponde ao componente *Load Staging TipoInscricao* presente na Figura 44, é possível verificar que o primeiro passo executado é extrair os dados da sua fonte e contar o número de registos obtidos para mais tarde ser auditado.

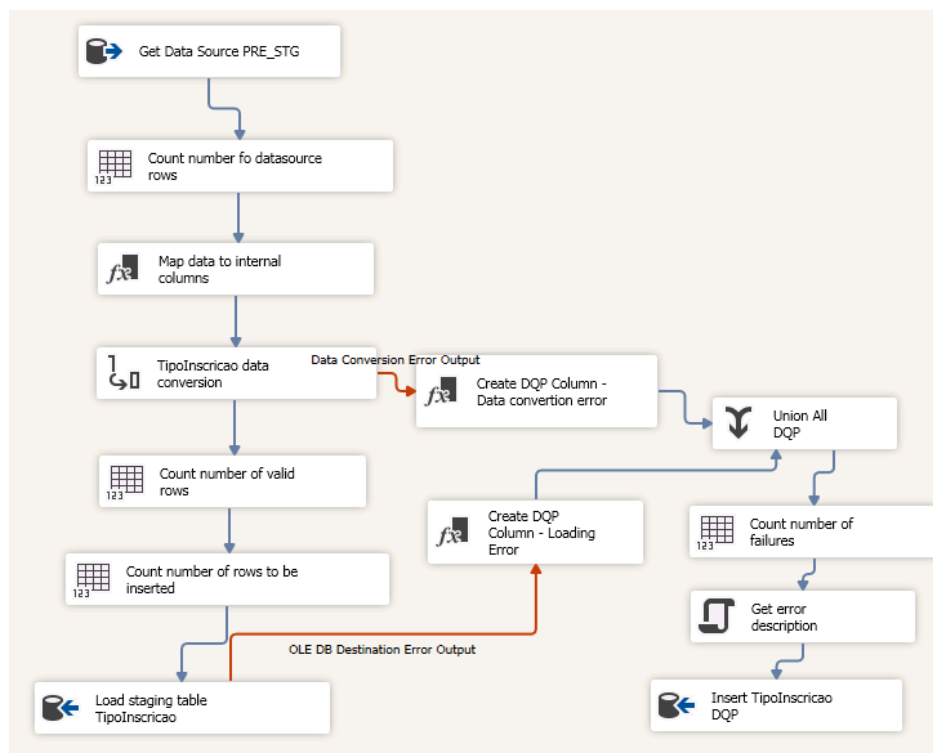


Figura 46 - Exemplo de Processo Extração e Transformação de Dados

De seguida, é feito um mapeamento dos campos existentes na fonte para campos internos do pacote *SSIS* para permitir controlar as dependências dos diferentes nomes dos campos que variam na fonte ao longo do tempo sendo que, desta forma a alteração apenas é feita neste

componente mantendo o restante processo intacto. Para garantir que o tipo de dados está de acordo com o destino, é feita uma conversão dos tipos de dados para o pretendido garantindo assim que o processo não falhe ao inserir os dados na tabela destino. Como já foi anteriormente explicitado, caso nalgum dos componentes *Map data to internal columns* e *TipInscricao Data Conversion* ocorra erro, os registos com problemas são redirecionados para a tabela *DQP* correspondente para permitir um melhor rastreamento do problema.

No último passo, os dados que estão prontos para serem inseridos são contabilizados e são enviados para a tabela destino. Adicionalmente, é feita uma validação. Caso ocorra erro ao inserir na tabela de *staging*, os registos com erro são redirecionados para a tabela *DQP* correspondente.

Para concluir, no fim do processo é inserido um registo na tabela de auditoria nomeadamente o identificador do pacote, nome do pacote, data e tempo a que este foi iniciado, número de registos extraídos da fonte de dados, número de registos inseridos, número de registos válidos que foram inseridos, número de registos com erros e o estado de como terminou o processo.

Não obstante, os outros pacotes de *Staging* presentes no diagrama da Figura 25 seguiram esta mesma metodologia de implementação. É de salientar que o nome do pacote de SSIS, para cada estrutura presente na *Staging Area* corresponde ao tipo de informação que cada uma contém.

### 5.3.2 Armazém de Dados

Neste subcapítulo irá ser apresentado ao pormenor todo o processo de carregamento para as dimensões e tabelas de factos.

Por último, será também apresentado mecanismos de controlo de erro para aumentar a fiabilidade e a disponibilidade dos dados acessíveis aos utilizadores.

#### 5.3.2.1 Dimensões

Na Figura 22 estão presentes as diferentes dimensões do AD, no entanto apenas algumas foram implementadas no âmbito desta dissertação, nomeadamente *Dim\_Aluno*, *Dim\_AnoCurricular*, *Dim\_Date*, *Dim\_Horario* e *Dim\_TipoInscricao*. As outras dimensões presentes foram implementadas pela colega que implementou o *Data Mart* de avaliações.

Como anteriormente explicitado, a importância das dimensões num AD é o contexto que estas oferecem aos factos (inseridos nas tabelas de factos). Cada dimensão armazena informação específica de contextualização, ou seja, no caso da dimensão aluno esta armazena apenas detalhes sobre os alunos da instituição.

O processo de carregamento de informação para as dimensões, ilustrado na Figura 47, é muito semelhante ao de *Staging*, com a exceção no caso de falha no processo, em que é executado o processo de remoção dos dados execução contidos na estrutura em processamento, por forma a garantir que não há inconsistências, como previamente explicitado. Desta forma, as dimensões seguem os seguintes passos:

- Criar registo de auditoria para o pacote que está a executar
- Carregar os dados vindos de *Staging* para a dimensão pretendida
- Finalizar o processo de auditoria com dados relativos ao carregamento efetuado: número de registos selecionados, inseridos, atualizados, em que data ocorreu e o estado. No caso de inconsistência ou de erros, estes são inseridos na tabela auditoria para um melhor controlo e uma maior rapidez na resolução de eventuais problemas.
- No caso de erro durante o processo é adicionalmente executado o processo de remoção de registos inseridos naquela execução.

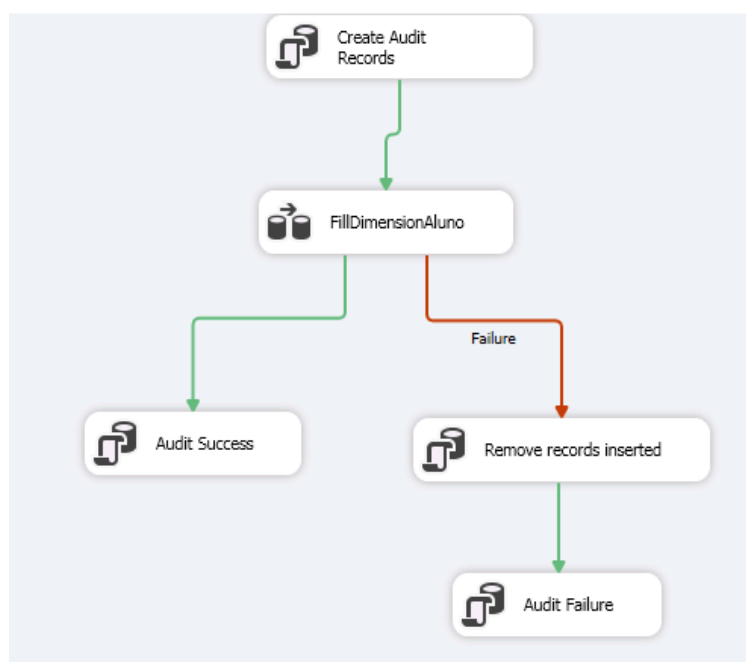


Figura 47 - Processo carregamento das dimensões

Por conseguinte, e de acordo com a Figura 48, é apresentado como exemplo o processo de carregamento da dimensão Aluno. Inicialmente é feita a extração e contabilização dos registos extraídos no pacote de SSIS da estrutura da *staging* correspondente, neste caso é a estrutura da *staging* que contém os dados dos alunos, por forma a auditar o volume de dados importados.

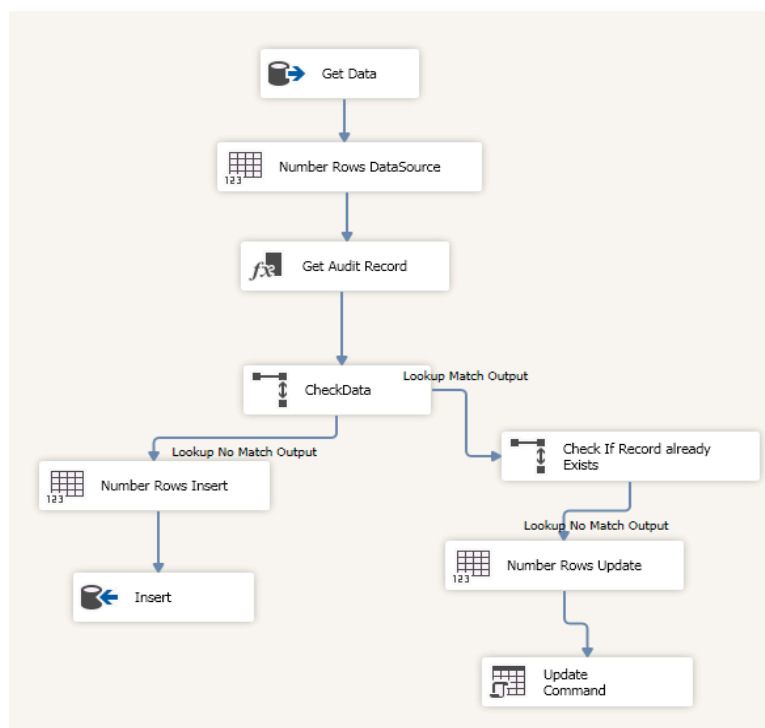


Figura 48 - Processo carregamento da dimensão Aluno

De seguida, é obtido o identificador de auditoria criado para esta execução, com o objetivo de identificar os registos que foram inseridos e atualizados nesta execução. No passo seguinte, por forma a manter a integridade do AD é verificado se cada registo extraído se encontra presente no mesmo e se contém os campos atualizados.

Por fim, o número de registos inseridos e o número de registos atualizados são atualizados consoante a verificação feita no passo anterior, e os registos são inseridos ou atualizados na dimensão.

### 5.3.2.2 Tabelas de Factos

Como já foi anteriormente explicitado, as tabelas de factos contêm os eventos ocorridos no sistema operacional, segundo uma granularidade bem definida, com o intuito de serem medidos para se retirarem conclusões sobre o desempenho da instituição e a área em questão.

No caso desta dissertação, os eventos que se pretendem armazenar são as inscrições dos alunos da LEI-ISEP. Com base nestes eventos serão criadas as métricas pretendidas por forma a serem utilizadas nas análises pretendidas pelo cliente.

Os dados carregados para este tipo de tabelas contêm informação sensível, o que faz com seja necessário ter especial cuidado ao obter as relações entre os factos e as dimensões necessárias, de modo a obter o contexto imprescindível. Também, é importante garantir que as métricas calculadas no processo estão corretas sendo então necessário fazer algumas comparações entre a fonte de dados e o resultado obtido.

Posto isto, e tendo em conta os fatores anteriormente enunciados, foi desenhado um processo de ETL, presente na Figura 49 e na Figura 50 , por forma a garantir que estes são respeitados.

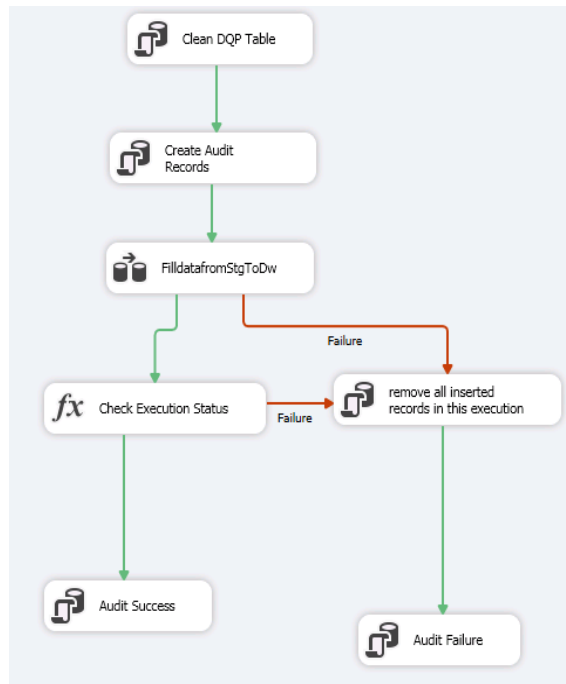


Figura 49 - Processo principal do carregamento da tabela de factos de Inscrições

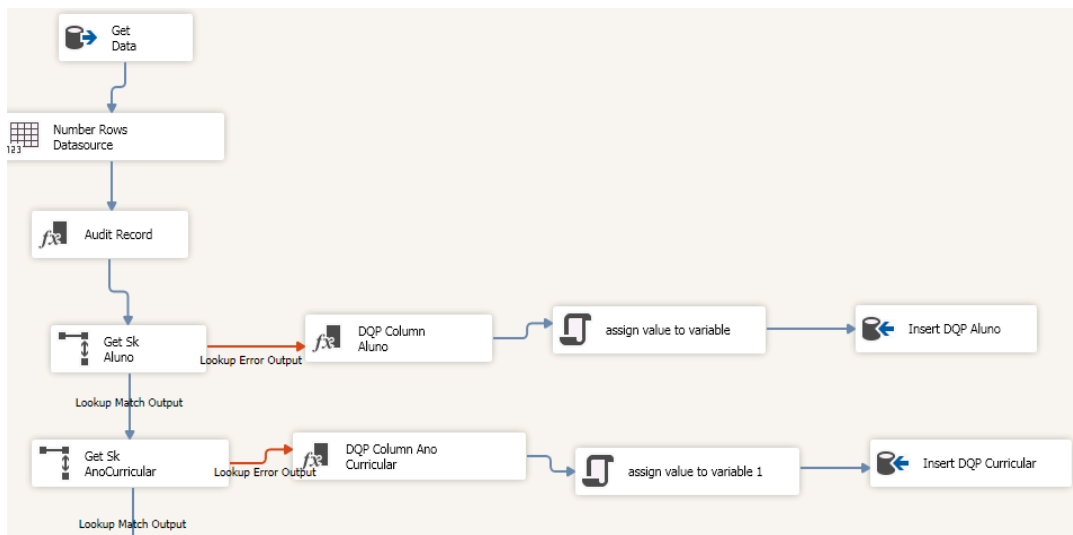


Figura 50 - Processo de carregamento da tabela de factos de inscrições

No início do processo, é criado o registo de auditoria que indica o início da execução do pacote para a ronda de ETL em que se encontra. Desta forma, e como já referido anteriormente, permite guardar informação sobre os dados extraídos de *staging*, como por exemplo o número de registos extraídos, número de registos inseridos e número de registos inválidos ou com erros.

Para além disso, permite indicar quais registos inseridos numa determinada data e por exemplo quanto tempo demorou a executar o processo.

De seguida é feita a extração dos dados do pacote SSIS da estrutura de *staging* correspondente, que no exemplo da Figura 50 é o pacote de inscrições. onde é feita a contabilização do número de registos extraídos. No passo seguinte, dentro da tarefa *FilldatafromStgToDw* (Figura 49), o identificador de auditoria é transformado numa coluna adicional, para ser passível deste ser inserido na tabela final.

Por conseguinte, por forma a garantir a integridade anteriormente referida, são estabelecidas as relações entre os dados extraídos e as diferentes dimensões de modo a que os identificadores respetivos de cada uma possam ser inseridos na tabela de factos, possibilitando assim a contextualização dos factos inseridos. A título de exemplo são apresentados dois componentes de pesquisa nas dimensões Aluno e Ano Curricular. Estes têm como objetivo obter a chave substituta (*Surrogate Key*) necessária para identificar o contexto do aluno e do ano curricular em que se encontra a cada registo extraído.

Por forma a garantir disponibilidade e ao mesmo tempo integridade no processo, caso não haja obtenção de uma chave substituta, o componente gera um erro e este é tratado. O tratamento efetuado é muito semelhante ao da *staging área*, sendo que o registo é impedido de ser inserido na tabela de factos e é redirecionado para uma tabela *DQP* específica para a tabela de factos de inscrições, permitindo assim uma melhor rastreabilidade dos registos com problemas. Ao ser inserido na tabela *DQP* correspondente, são juntamente guardados os campos contextualizadores do possível problema, componente onde ocorreu e o erro específico que o componente produziu.

Por fim, os registos que obtiverem relação com todas as dimensões são contabilizados e são inseridos na tabela de factos.

Ao finalizar, caso o processo liberte um erro anormal e não seja tratado dentro da tarefa *FilldatafromStgToDw* (Figura 49), este é tratado no processo principal que também se encontra presente na Figura 49. Como anteriormente explicitado, a tarefa de *Remove All Inserted Records in this execution* é acionada no caso de falha do processo, removendo assim todos os registos inseridos nesta execução através do registo de auditoria inserido conjuntamente com os dados. Desta forma é mantida a integridade desta tabela ao longo do tempo e das várias execuções do processo de ETL.

Não obstante, é importante referir que no caso da tabela de factos agregada de inscrições, *Fact\_Inscricao\_Aggreg*, o processo é ligeiramente diferente. O processo principal ilustrado na Figura 49 mantém-se, sendo que o que altera é a fonte de dados, passando a ser a tabela de factos de inscrições.

## 5.4 Orquestrador do processo de ETL

Por forma a automatizar todo o processo de execução dos pacotes desenvolvidos para o processo de ETL, foi desenvolvido um orquestrador.

Este consiste em executar todos os pacotes criados numa determinada ordem para garantir que os dados são importados na ordem correta. Para possibilitar o carregamento dos mesmos, estes necessitam de ser classificados de acordo com a área que pertencem (*Staging* ou AD) com o objetivo de permitir realizar ações de intervenção no caso de erro no processo de ETL, atuando apenas na área afetada e respetivas dependências. Para exemplificar melhor como está estruturado o armazenamento das configurações é apresentado na Figura 51 o modelo ER (modelo de entidades e relacionamentos) respetivo.

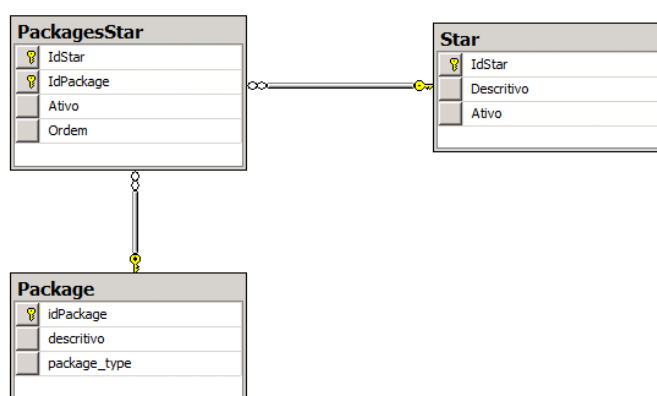


Figura 51 – Modelo ER do Orquestrador

Como é possível verificar, para cada pacote (*idPackage*) é feita uma categorização de acordo com o tipo de pacote (*package\_type*), que tem como objetivo distinguir a que área este pertence. Para indicar a ordem com que cada pacote é carregado foi criado na tabela *PackagesStar* o campo *Ordem*, associado ao identificador do pacote (*idPackage*). Como se pode verificar na mesma tabela são também associados os campos *Ativo* e o campo identificador da estrela associado. O campo *Ativo* define se o pacote está ativo ou não para ser incluído no processo de ETL e, o campo que identifica a estrela a que está associada, como o próprio nome indica, identifica uma estrela do AD nomeadamente uma área de negócio. Desta forma, e como exemplo, é possível executar os vários pacotes de ETL associado à estrela de Inscrições numa determinada ordem definida para que as dimensões e tabelas de factos sejam preenchidas corretamente.

Por fim, para poder carregar todo o AD incluindo também o *Data Mart* de avaliações (configurada pela colega que desenvolveu esta área, Gisela Couto) foi criada também uma estrela que contempla estas duas áreas, de modo a possibilitar também carregar o AD através de uma única configuração.

## 5.5 Cubos

Nesta secção são apresentados os cubos analíticos criados para responder às análises pretendidas pelo cliente de uma forma intuitiva e *self-service*. Devido às suas capacidades analíticas, estes permitem realizar outras análises que o cliente pretenda no momento e obter os resultados em pouco tempo, mediante as métricas disponibilizadas até ao momento.

Nesta dissertação será apresentado o cubo de inscrições que contém apenas informação relativa a inscrições de alunos e o cubo inscrições e avaliações de alunos. Este último é apresentado, pois é necessário para realizar análises que necessitem de relacionar inscrições com avaliações.

### 5.5.1 Cubo de Inscrições

Como o próprio nome indica, este cubo contém apenas dados relativos a inscrições de alunos. Ao criar este cubo, as análises apenas referentes a inscrições serão mais rapidamente obtidas visto este ter um menor volume de dados quando comparado com o cubo de inscrições e avaliações.

Outro motivo para a criação deste cubo deve-se à disponibilidade. Devido ao menor volume de dados, como explicitado anteriormente, este leva um menor tempo para estar novamente disponível quando é atualizado, permitindo assim aos utilizadores atualizarem os valores com um menor tempo de espera quando comparado com o cubo de inscrições e avaliações.

Para construir este cubo analítico foi necessário efetuar os seguintes passos:

- Criar uma ligação à fonte de dados, que neste caso é o AD. Esta ligação disponibiliza todas as dimensões e tabelas de factos existentes no AD.
- Criar a estrutura do cubo, definindo a sua *Data Source View*: identificar as tabelas de factos e dimensões a serem incorporadas, para disponibilizar as métricas e os contextos associados nas análises efetuadas. Esta estrutura encontra-se presente na Figura 52, sendo que as ligações efetuadas entre tabelas de factos e as dimensões são criadas automaticamente através das restrições existentes (chaves estrangeiras)
- Criar por fim o cubo, utilizando a *Data Source View* criada anteriormente e processar para este ficar disponível.

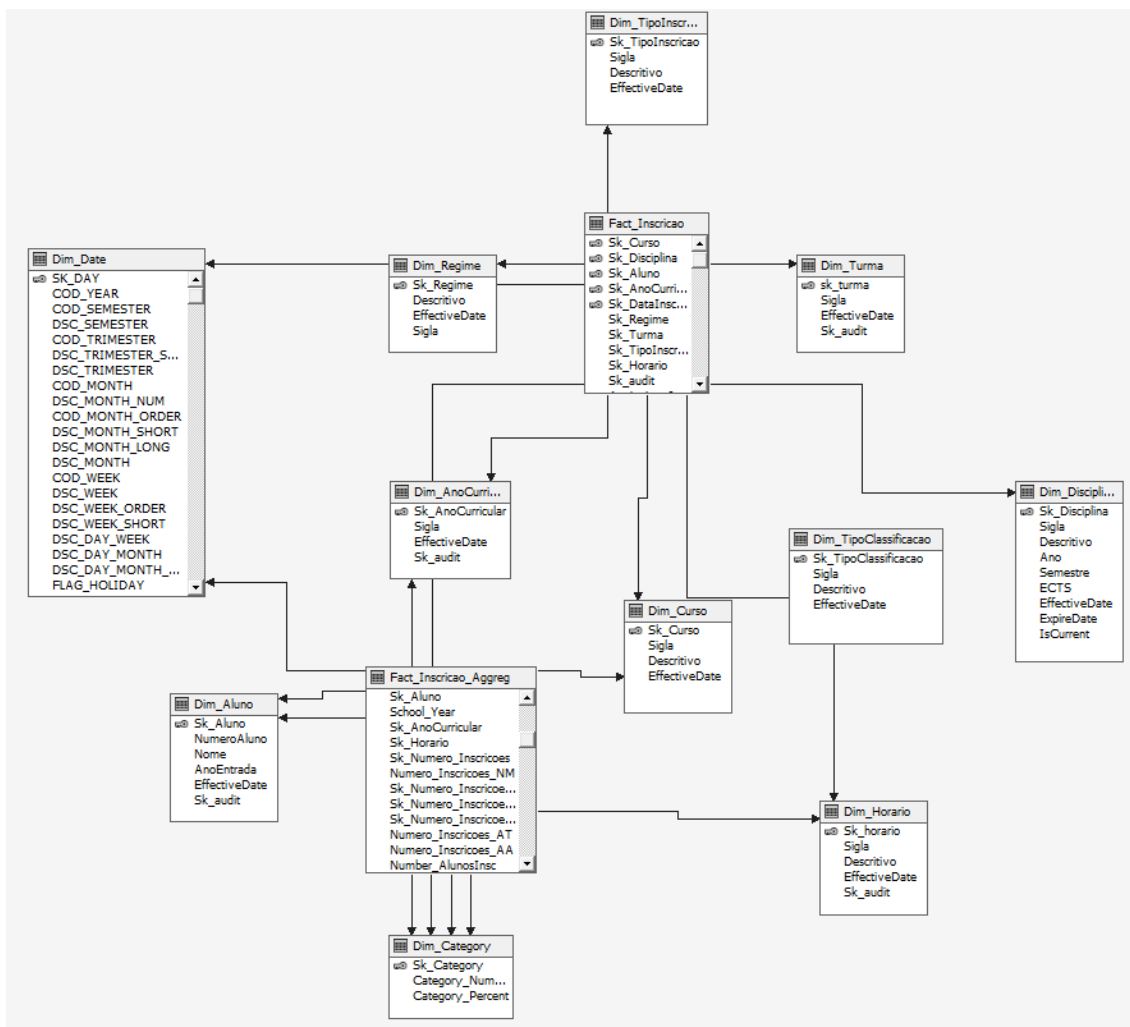


Figura 52 - Estrutura do Cubo Inscrições

Não obstante, é importante realçar que ao criar o cubo pelo assistente disponibilizado pelo *Analysis Services*, este cria automaticamente as dimensões necessárias a serem processadas devido às relações existentes entre as tabelas de factos e dimensões (chaves estrangeiras). Caso seja necessário acrescentar outra dimensão a ser processada terá que ser adicionada manualmente como ilustrado na Figura 53.

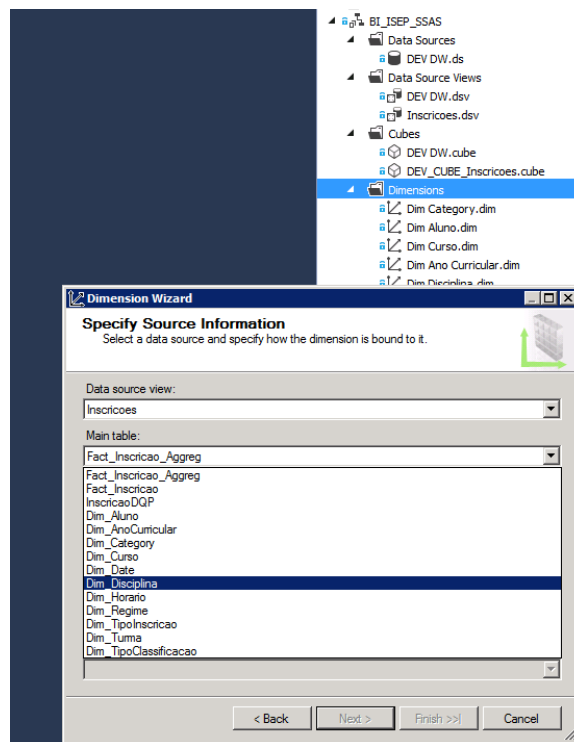


Figura 53 - Adicionar dimensão manualmente no cubo

### 5.5.2 Cubo de Inscrições e Avaliações

No cubo de inscrições e avaliações, como o próprio nome indica, está contida informação relativa a inscrições e avaliações de alunos. É com base neste que são disponibilizadas todas as análises pretendidas pelo cliente até ao momento.

O procedimento de criação é semelhante ao cubo de inscrições contemplando adicionalmente a tabela de factos de avaliações e as respetivas dimensões. Como este possui mais tabelas que não existem na estrutura do cubo de Inscrições, foi necessário criar uma nova *Data Source View*, que está presente na Figura 54.

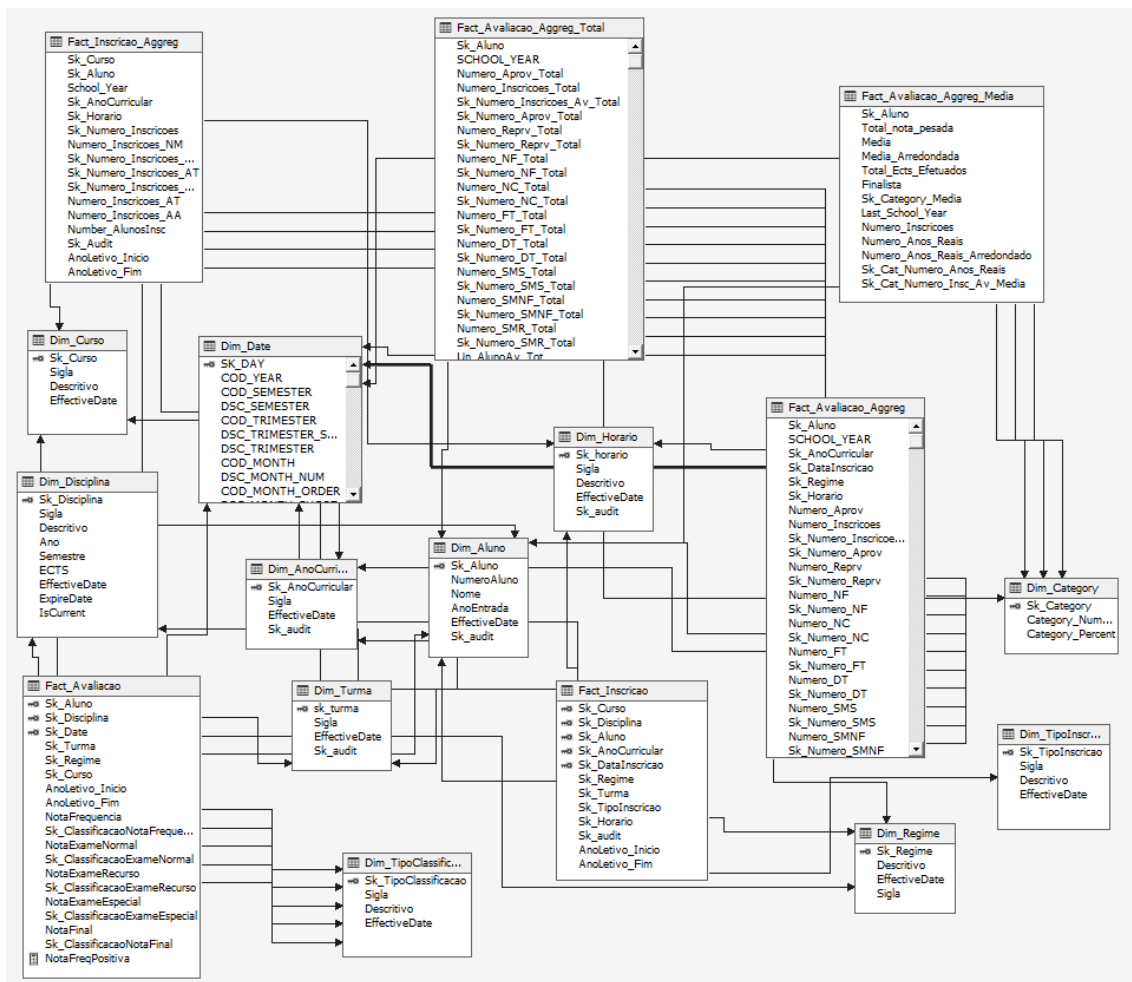


Figura 54 - Esquema do Cubo de Inscrições e Avaliações

Por fim é de salientar que, para atualizar este cubo com toda a informação corrente é necessário correr o processo de ETL para as inscrições e avaliações. Caso não seja executado o processo de ETL para ambas as estrelas, este cubo está parcialmente atualizado podendo causar alguma entropia nas análises realizadas.



## 6 Avaliação de resultados

As análises e/ou Dashboards são a última camada presente numa solução de BI. Com as ferramentas adequadas, são apresentadas análises com base nos dados fornecidos pelo cubo analítico, que por sua vez utiliza como base as tabelas do AD.

Na ferramenta *Excel* serão apresentadas as análises que o cliente já possui permitindo assim retro compatibilidade com o que já existia. As suas capacidades analíticas são um ponto forte, apesar das suas limitações.

Também, serão apresentadas novas análises implementadas numa outra ferramenta analítica e muito recentemente lançada pela Microsoft, o *Power Bi*. O objetivo de apresentar as análises nesta ferramenta é demonstrar a flexibilidade e alternativa à solução que o cliente possui.

Contudo, como ao importar os dados da fonte para o AD é feito um tratamento dos dados há a possibilidade de um desvio aquando a comparação das análises realizadas pelo cliente e as mesmas reproduzidas com base no AD. Desta forma, foi estipulado um desvio máximo de cinco por cento aquando da comparação entre as análises. No caso de haver desvios acima do estipulado, apenas os mais críticos serão justificados perante as análises efetuadas aos dados contidos na fonte, como no AD.

Por forma a comprovar a veracidade das análises efetuadas no AD é feita uma comparação direta a cada análise reproduzida e fornecida pelo cliente. Adicionalmente é também apresentado o respetivo desvio entre ambas tornando mais perceptíveis possíveis diferenças. Para solidificar a veracidade das análises no AD, é feita a comparação de resultados com base em três anos letivos: 2012/2013, 2013/2014 e 2014/2015.

A fórmula utilizada para validar os dados baseia-se no valor absoluto da diferença entre o valor apresentado na análise reproduzida no AD e a análise reproduzida pelo cliente, a dividir pela soma dos dois valores. Desta forma, obtém-se a percentagem de diferença entre ambos os valores. Para tornar mais perceptível as diferenças serão utilizadas cores de acordo com o desvio

obtido: verde no caso de a percentagem de desvio ser até cinco por cento, acima de cinco por cento até vinte por cento é marcado como amarelo, acima de vinte por cento até trinta por cento é marcado como laranja e o mais crítico de todos é acima de trinta por cento e é marcado com a cor vermelha.

## **6.1 Análises no Microsoft Excel**

Para permitir retrocompatibilidade com a solução anterior, são disponibilizadas as análises que o cliente possui no *Microsoft Excel*, mas como fonte de dados o novo AD.

Deste modo, serão apresentadas as análises que o cliente possui com a respetiva análise de comparação entre a reproduzida no AD e a fornecida pelo cliente.

Por forma a comprovar a veracidade dos factos contidos no AD serão apresentadas as análises realizadas no AD, as análises realizadas pelo cliente e a respetiva comparação para os anos letivos 2012/2013, 2013/2014 e 2014/2015.

### **6.1.1 Número de alunos inscritos por unidade curricular e ano curricular, em cada ano formativo**

A primeira análise presente no documento fornecido, foi o número de alunos inscritos por UC, ano curricular, em cada ano formativo. Recorrendo às capacidades analíticas da ferramenta de análise esta foi fielmente reproduzida em aspeto. A análise encontra-se presente na , tendo como base o AD.

SCHOOL YEAR INSC	2012-2013			
Fact Inscricao Count	Rótulos de Coluna			
Rótulos de Linha	1	2	3	Total Geral
⊞ 1	2870	663	233	3766
ALGAN	313	48	9	370
AMATA	316	77	20	413
APROG	298	37	2	337
ESOFT	280	104	32	416
LAPR1	288	13		301
LAPR2	266	77	42	385
MATCP	271	111	50	432
MDISC	266	71	30	367
PProg	280	103	45	428
PRCMP	292	22	3	317
⊞ 2	316	1955	565	2836
ARQCP	19	266	67	352
BDDAD	36	254	18	308
EAPLI	41	153	55	249
ESINF	5	240	118	363
FSIAP	21	254	65	340
LAPR3	1	190	2	193
LAPR4	39	146	38	223
LPROG	55	170	63	288
RCOMP	58	148	56	262
SCOMP	41	134	83	258
⊞ 3	65	295	2040	2400
ALGAV		3	226	229
ARQSI		12	267	279
ASIST		19	212	231
COMPA	22	53	190	265
CORGA	12	55	144	211
GESTA	1	53	182	236
IARTI	20	49	198	267
LAPR5		4	202	206
PESTI	10	38	206	254
SGRAI		9	213	222
<b>Total Geral</b>	<b>3251</b>	<b>2913</b>	<b>2838</b>	<b>9002</b>

Figura 55 - Inscrições por disciplina em cada ano curricular no ano letivo 2012/2013 (AD)

Já na Figura 56 encontra-se a análise fornecida pelo cliente. Como se pode verificar a análise foi fielmente reproduzida com estrutura e apresentação que o cliente possuía.

Inscrições a disciplinas					
Ano	UC	Ano em que o aluno se encontra			Total
		1	2	3	
1º	ALGAN	412	67	13	492
	AMATA	416	99	28	543
	APROG	355	47	3	405
	LAPR1	288	14		302
	PRCMP	320	28	3	351
	ESOFT	416	90	1	507
	LAPR2	339	58	10	407
	MATCP	396	111	28	535
	MDISC	349	48	7	404
	PPROG	384	101	3	488
2º	ARQCP	26	435	119	580
	BDDAD	52	409	27	488
	ESINF	6	298	160	464
	FSIAP	30	431	301	762
	LAPR3	1	189	2	192
	EAPLI	2	254	31	287
	LAPR4		202	19	221
	LPROG	9	365	29	403
	RCOMP	20	299	27	346
	SCOMP	1	230	84	315
3º	ALGAV		3	371	374
	ARQSI		15	391	406
	ASIST		27	324	351
	GESTA	1	66	218	285
	LAPR5		4	195	199
	SGRAI		11	283	294
	COMPA		37	354	391
	CORGA		41	191	232
	IARTI		25	400	425
	PESTI			460	460
P/EST-ERASMUS 10			1	1	
P/EST-ERASMUS 20			2	2	
P/EST-ERASMUS 30			2	2	
Total		3823	4004	4087	11914

Figura 56 - Inscrições por disciplina em cada ano curricular no ano letivo 2012/2013 (Cliente)

No entanto, aquando a comparação entre as análises do AD e a fornecida pelo cliente para o ano letivo 2012/2013, que está presente na Figura 57, esta contém valores muito díspares. Mediante a classificação anteriormente apresentada, os casos mais críticos são nitidamente os acima de trinta por cento.

Disciplina	Ano 1	Ano 2	Ano 3	Total
ALGAN	13,66%	16,52%	18,18%	14,15%
ALGAV	0,00%	0,00%	24,29%	24,05%
AMATA	13,66%	12,50%	16,67%	13,60%
APROG	8,73%	11,90%	20,00%	9,16%
ARQCP	15,56%	24,11%	27,96%	24,46%
ARQSI	0,00%	11,11%	18,84%	18,54%
ASIST	0,00%	17,39%	20,90%	20,62%
BDDAD	18,18%	23,38%	20,00%	22,61%
COMPA	100,00%	17,78%	30,15%	19,21%
CORGA	100,00%	14,58%	14,03%	4,74%
EAPLI	90,70%	24,82%	27,91%	7,09%
ESINF	9,09%	10,78%	15,11%	12,21%
ESOFT	19,54%	7,22%	93,94%	9,86%
FSIAP	17,65%	25,84%	64,48%	38,29%
GESTA	0,00%	10,92%	9,00%	9,40%
IARTI	100,00%	32,43%	33,78%	22,83%
LAPR1	0,00%	3,70%	0,00%	0,17%
LAPR2	12,07%	14,07%	61,54%	2,78%
LAPR3	0,00%	0,26%	0,00%	0,26%
LAPR4	100,00%	16,09%	33,33%	0,45%
LAPR5	0,00%	0,00%	1,76%	1,73%
LPROG	71,88%	36,45%	36,96%	16,64%
MATCP	18,74%	0,00%	28,21%	10,65%
MDISC	13,50%	19,33%	62,16%	4,80%
PESTI	100,00%	100,00%	38,14%	28,85%
PPROG	15,66%	0,98%	87,50%	6,55%
PRCMP	4,58%	12,00%	0,00%	5,09%
RCOMP	48,72%	33,78%	34,94%	13,82%
SCOMP	95,24%	26,37%	0,60%	9,95%
SGRAI	0,00%	10,00%	14,11%	13,95%

Figura 57 - Comparação entre a análise do AD e do Cliente para o ano letivo 2012/2013

Numa primeira fase foram analisados os registos acima dos trinta por cento de diferença por serem os mais críticos. Segundo a lógica de inscrições a disciplinas na LEI-ISEP, os alunos podem efetuar inscrições a disciplinas de anos curriculares superiores caso tenham os créditos necessários para o efetuarem e esta não contenha precedências de UCs que estes ainda não tenham terminado com sucesso. Posto isto, pode-se verificar que os casos que mais se evidenciam são alunos que se encontram inscritos a disciplinas do terceiro ano e que se encontram ainda presentes no primeiro ano curricular. De acordo com a análise do cliente, que está presente na Figura 56, as disciplinas de COMPA, CORGA, IARTI e PESTI são as mais evidentes comparativamente à reproduzida no AD.

Posto isto, para provar que a análise reproduzida no AD contém os valores corretos foi feita uma pesquisa na fonte de dados fornecida pelo cliente por forma a validar se realmente estes casos acontecem. Para analisar os valores foi feito um filtro, na fonte de dados, pelos alunos que se encontram no primeiro ano curricular no ano letivo 2012/2013.

Para o caso da disciplina de PESTI foi ainda adicionado o filtro que indica que há alunos inscritos a esta disciplina e o resultado obtido está presente na Figura 58. É importante relembrar que esta UC é anual e não possui uma turma, sendo marcada a inscrição através do caracter asterisco.

Número	Nome	Regime Freq.	Ano Curr	PESTI	Ano
1040107		Integral	1	*	20122013
1050121		Parcial	1	*	20122013
1090380		Integral	1	*	20122013
1090407		Integral	1	*	20122013
1090461		Integral	1	*	20122013
1090480		Integral	1	*	20122013
1100711		Parcial	1	*	20122013
1100717		Integral	1	*	20122013
1110865		Integral	1	*	20122013
1121218		Integral	1	*	20122013

Figura 58 - Inscrições efetuadas em PESTI por alunos do 1º ano no ano letivo 2012/2013

Como se pode constatar na Figura 58, o número de inscritos à UC curricular de PESTI são exatamente os que são apresentados na respetiva análise com base no AD. Desta forma, comprova-se que a análise do AD está correta no caso desta UC. De seguida, por forma a analisar os valores da unidade curricular IARTI aplicou-se o filtro de alunos inscritos a esta na fonte de dados e o resultado obtido está presente na Figura 59.

Número	Nome	Regime Freq.	Ano Curr	IARTI	Ano
1040107		Integral	1	3NC	20122013
1050119		Integral	1	3DJ	20122013
1050121		Parcial	1	3DJ	20122013
1070269		Parcial	1	3NB	20122013
1090383		Integral	1	3DI	20122013
1090407		Integral	1	3DJ	20122013
1090440		Parcial	1	3DI	20122013
1090461		Integral	1	3DI	20122013
1090480		Integral	1	3NA	20122013
1090488		Parcial	1	3NB	20122013
1090498		Parcial	1	3DE	20122013
1090500		Integral	1	3DG	20122013
1100663		Integral	1	3DH	20122013
1100711		Parcial	1	3NA	20122013
1100717		Integral	1	3DB	20122013
1100731		Integral	1	3DF	20122013
1110865		Integral	1	3DH	20122013
1121211		Integral	1	3DI	20122013
1121218		Integral	1	3NB	20122013
1921853		Integral	1	3NA	20122013

Figura 59 - Inscrições efetuadas em IARTI por alunos do 1º ano no ano letivo 2012/2013

De acordo com os dados extraídos da fonte, que estão presentes na Figura 59, ao contabilizar o número de registos pode-se verificar que está de acordo com o que está presente na análise reproduzida com base no AD e, ao mesmo tempo, não corresponde ao que o cliente apresenta.

De seguida foram efetuadas as consultas à fonte de dados para o caso das disciplinas de COMPA e CORGA. Para cada uma, foram efetuados os filtros relativos a cada UC, para além do primeiro ano curricular e ano letivo, estando os resultados presentes na Figura 60 e Figura 61, respetivamente.

Númer	Nom	Regime Freq.	Data de Insc.	Ano Curr	COMP A	Ano
1040107		Integral		1	3NC	20122013
1050119		Integral		1	3DJ	20122013
1050121		Parcial		1	3DJ	20122013
1070269		Parcial		1	3NB	20122013
1090383		Integral		1	3DI	20122013
1090422		Integral		1	3DE	20122013
1090440		Parcial		1	3DI	20122013
1090461		Integral		1	3DI	20122013
1090480		Integral		1	3NA	20122013
1090488		Parcial		1	3NB	20122013
1090498		Parcial		1	3DE	20122013
1090500		Integral		1	3DG	20122013
1100663		Integral		1	3DH	20122013
1100711		Parcial		1	3NA	20122013
1100717		Integral		1	3DB	20122013
1100731		Integral		1	3DF	20122013
1110865		Integral		1	3DH	20122013
1121211		Integral		1	3NA	20122013
1121217		Integral		1	3NA	20122013
1121218		Integral		1	3NB	20122013
1121219		Integral		1	3NC	20122013
1921853		Integral		1	3NA	20122013

Figura 60 - Alunos Inscritos a COMP A no ano letivo 2012/2013

Número	Nome	Regime Freq.	Ano Curr	D/N	CORGA	Ano
1070269		Parcial	1	N	3NB	20122013
1090422		Integral	1	D	3DE	20122013
1090440		Parcial	1	D	3DC	20122013
1090453		Integral	1	D	3DC	20122013
1090480		Integral	1	D	3NA	20122013
1090488		Parcial	1	D	3NB	20122013
1090500		Integral	1	D	3DG	20122013
1100711		Parcial	1	N	3NA	20122013
1100717		Integral	1	D	3DB	20122013
1100731		Integral	1	D	3DF	20122013
1110865		Integral	1	N	3DH	20122013
1921853		Integral	1	N	3NA	20122013

Figura 61 - Alunos inscritos a CORGA no ano letivo 2012/2013

Mais uma vez, mediante o número de registos obtidos a cada UC pode-se comprovar que os valores estão de acordo com o que está presente na análise reproduzida com base no AD e não representa os valores apresentados na análise do cliente. Também, esta é a causa de todos os outros valores que se desviam mais de cinco por cento.

Por conseguinte foi efetuada uma análise aos dados da mesma análise, mas relativamente ao ano letivo de 2013/2014. Na Figura 62 encontra-se a análise reproduzida com base no AD e na Figura 63 encontra-se a análise efetuada pelo cliente. Como se pode verificar na Figura 64 os valores encontram-se com diferenças pouco significativas no geral.

Fact Inscricao	Rótulos de Coluna			
Rótulos de Lin	1	2	3	Total Geral
ALGAN	325	37	7	369
ALGAV	3	5	258	266
AMATA	332	105	34	471
APROG	316	37		353
ARQCP	6	276	40	322
ARQSI	4	6	218	228
ASIST	2	8	193	203
BDDAD	24	305	32	361
COMPA	2	11	222	235
CORGA	2	13	149	164
EAPLI	5	268	35	308
ESINF	8	278	57	343
ESOFT	368	53	1	422
FSIAP	18	281	142	441
GESTA	1	10	142	153
IARTI	2	16	249	267
LAPR1	322	9		331
LAPR2	355	53	5	413
LAPR3	3	242	8	253
LAPR4		262	34	296
LAPR5		1	178	179
LPROG	4	308	27	339
MATCP	364	87	18	469
MDISC	366	71	7	444
PESTI			239	239
PPROG	378	65		443
PRCMP	320	20	3	343
RCOMP	8	244	16	268
SCOMP	2	184	48	234
SGRAI	4	6	243	253
<b>Total Geral</b>	<b>3544</b>	<b>3261</b>	<b>2605</b>	<b>9410</b>

Figura 62 - Inscrições por disciplina em cada ano curricular no ano letivo 2013/2014 (AD)

Ano	UC	Ano em que o aluno se encontra			Total
		1	2	3	
1º	ALGAN	324	38	7	369
	AMATA	331	106	34	471
	APROG	315	37		352
	LAPR1	321	9		330
	PRCMP	318	20	3	341
	ESOFT	366	53		419
	LAPR2	354	56	5	415
	MATCP	362	87	17	466
	MDISC	364	71	7	442
	PPROG	377	65		442
2º	ARQCP	6	276	40	322
	BDDAD	24	305	32	361
	ESINF	8	278	57	343
	FSIAP	18	281	142	441
	LAPR3	3	242	8	253
	EAPLI	4	268	35	307
	LAPR4		262	35	297
	LPROG	3	308	27	338
	RCOMP	8	244	16	268
	SCOMP	2	184	49	235
3º	ALGAV	3	5	259	267
	ARQSI	4	6	218	228
	ASIST	2	8	193	203
	GESTA	1	10	142	153
	LAPR5		1	179	180
	SGRAI	4	6	243	253
	COMPA	2	11	222	235
	CORGA	2	13	149	164
	IARTI	2	15	248	265
	PESTI			240	240
Total		3528	3265	2607	9400

Figura 63 - Inscrições por disciplina em cada ano curricular no ano letivo 2013/2014 (Cliente)

Disciplina	Ano 1	Ano 2	Ano 3	Total
ALGAN	0,15%	1,33%	0,00%	0,00%
ALGAV	0,00%	0,00%	0,19%	0,19%
AMATA	0,15%	0,47%	0,00%	0,00%
APROG	0,16%	0,00%	0,00%	0,14%
ARQCP	0,00%	0,00%	0,00%	0,00%
ARQSI	0,00%	0,00%	0,00%	0,00%
ASIST	0,00%	0,00%	0,00%	0,00%
BDDAD	0,00%	0,00%	0,00%	0,00%
COMPA	0,00%	0,00%	0,00%	0,00%
CORGA	0,00%	0,00%	0,00%	0,00%
EAPLI	11,11%	0,00%	0,00%	0,16%
ESINF	0,00%	0,00%	0,00%	0,00%
ESOFT	0,27%	0,00%	100,00%	0,36%
FSIAP	0,00%	0,00%	0,00%	0,00%
GESTA	0,00%	0,00%	0,00%	0,00%
IARTI	0,00%	3,23%	0,20%	0,38%
LAPR1	0,16%	0,00%	0,00%	0,15%
LAPR2	0,14%	2,75%	0,00%	0,24%
LAPR3	0,00%	0,00%	0,00%	0,00%
LAPR4	0,00%	0,00%	1,45%	0,17%
LAPR5	0,00%	0,00%	0,28%	0,28%
LPROG	14,29%	0,00%	0,00%	0,15%
MATCP	0,28%	0,00%	2,86%	0,32%
MDISC	0,27%	0,00%	0,00%	0,23%
PESTI	0,00%	0,00%	0,21%	0,21%
PPROG	0,13%	0,00%	0,00%	0,11%
PRCMP	0,31%	0,00%	0,00%	0,29%
RCOMP	0,00%	0,00%	0,00%	0,00%

Figura 64 - Comparação entre a análise do AD e do Cliente para o ano letivo 2013/2014

Como se pode constatar na Figura 64, as diferenças relevantes existem na disciplina de ESOFT, EAPLI e LPROG.

Por forma a validar estes casos foi efetuada uma análise às disciplinas identificadas. De acordo com o tipo de análise efetuada anteriormente para o ano letivo 2012/2013, foi realizada a mesma, mas para o ano letivo 2013/2014 e para alunos do primeiro ano no caso de LPROG e EAPLI e para alunos do terceiro ano no caso da unidade curricular ESOFT.

No caso da unidade curricular de ESOFT, que neste caso é o que contém a maior percentagem de desvio, realizou-se uma pesquisa na fonte e não se obteve nenhum resultado. Posto isto, pode-se afirmar que o resultado obtido na análise do AD está correto de acordo com a fonte fornecida. Posteriormente no caso das disciplinas LPROG e EAPLI a consulta à fonte de dados está presente na Figura 65 e Figura 66 respetivamente.

Númer	Nom	Regime Freq.	Data de Insc.	Ano Curr	LPROG	Ano
1060171		Integral		1	2NC	20132014
1100533		Integral		1	2NC	20132014
1110923		Extraordinário		1	*	20132014
1110948		Integral		1	2NC	20132014

Figura 65 - Alunos Inscritos a LPROG no ano letivo 2013/2014

Númer	Nom	Regime Freq.	Data de Insc.	Ano Curr	EAPLI	Ano
1050140		Integral		1	2DL	20132014
1060171		Integral		1	2DB	20132014
1100593		Integral		1	2DB	20132014
1110948		Integral		1	2DB	20132014
1120971		Integral		1	2DB	20132014

Figura 66 - Alunos Inscritos a EAPLI no ano letivo 2013/2014

Por fim, no caso do ano letivo 2014/2015 verifica-se uma grande discrepância de valores como se pode constatar na Figura 67.

Disciplina	Ano 1	Ano 2	Ano 3	Total
ALGAN	1,73%	16,92%	6,67%	2,93%
ALGAV	100,00%	66,67%	0,77%	0,56%
AMATA	1,38%	2,91%	20,93%	0,32%
APROG	0,32%	23,33%	0,00%	1,73%
ARQCP	0,00%	10,39%	22,33%	11,90%
ARQSI	100,00%	20,00%	10,29%	8,80%
ASIST	100,00%	11,11%	0,00%	0,00%
BDDAD	65,52%	3,48%	0,00%	0,41%
COMPA	100,00%	4,76%	6,53%	5,62%
CORGA	100,00%	25,71%	9,42%	10,38%
EAPLI	11,11%	7,75%	16,67%	5,54%
ESINF	77,78%	12,30%	12,98%	11,37%
ESOFT	3,68%	1,92%	100,00%	3,12%
FSIAP	89,47%	3,93%	5,19%	0,92%
GESTA	100,00%	16,67%	9,27%	9,47%
IARTI	100,00%	50,00%	1,39%	0,95%
LAPR1	2,87%	5,26%	0,00%	2,94%
LAPR2	3,15%	10,89%	0,00%	1,43%
LAPR3	100,00%	9,02%	0,00%	8,17%
LAPR4	100,00%	4,90%	55,56%	0,50%
LAPR5	0,00%	0,00%	11,60%	11,55%
LPROG	33,33%	2,84%	1,82%	3,15%
MATCP	0,82%	7,41%	2,86%	0,54%
MDISC	1,22%	3,65%	0,00%	0,45%
PESTI	0,00%	0,00%	6,98%	6,98%
PPROG	0,00%	28,71%	100,00%	3,27%
PRCMP	2,30%	20,00%	25,00%	3,81%
RCOMP	23,08%	7,22%	3,23%	5,96%

Figura 67 - Comparação entre a análise do AD e do Cliente para o ano letivo 2014/2015

A origem das diferenças encontradas baseia-se nas mesmas diferenças encontradas para o ano letivo 2012/2013. Estas diferenças podem ter sido causadas pelo algoritmo de anonimização de dados executado pelo cliente deturpando assim os resultados. Apesar dos alunos poderem inscrever-se em unidades curriculares em avanço, PESTI é uma que não pode ser feita se o aluno não estiver no terceiro ano. Não obstante, outra possibilidade para ocorrer este problema é terem sido usadas outras fontes de dados para gerar as análises fornecidas pelo cliente, no entanto como se pode comprovar ao longo das várias análises efetuadas os resultados apresentados com base no AD representam fielmente o que se encontra na fonte de dados fornecida pelo cliente.

### 6.1.2 Alunos inscritos a cada unidade curricular em cada ano formativo

A segunda análise apresentada pelo cliente tem como objetivo visualizar o número de alunos por ano curricular de acordo com o número de unidades curriculares a que se inscreveu.

Desta forma, para comprovar os resultados efetuou-se o mesmo tipo de análise utilizado anteriormente. Começando pelo ano letivo 2012/2013 é apresentada na Figura 68 a análise reproduzida no AD e na Figura 69 encontra-se a análise fornecida pelo cliente.

Number Alunos Insc		Rótulos de Coluna														Rótulos de Linha	1
		2	3	4	5	6	7	8	9	10	11	12	13	14	Total Geral		
1		3	3	1	8	23	17	37	49	25	198	10	2		376		
2		1	4	8	11	26	37	35	41	60	69	40	19	1	352		
3		29	22	21	26	26	29	38	40	34	69	39	18	3	2	396	
<b>Total Geral</b>		<b>33</b>	<b>29</b>	<b>30</b>	<b>45</b>	<b>75</b>	<b>83</b>	<b>110</b>	<b>130</b>	<b>119</b>	<b>336</b>	<b>89</b>	<b>39</b>	<b>4</b>	<b>2</b>	<b>1124</b>	

Figura 68 – Análise proveniente do AD do ano letivo 2012/2013

		Nº de UCs a que o aluno está inscrito														
Ano	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	4	2	1	8	21	21	23	32	25	230						367
2		5	3	8	27	40	30	28	51	72	57	28	2			351
3	52	28	26	18	21	29	31	30	38	65	35	22	12	1		408
<b>Total</b>	<b>56</b>	<b>35</b>	<b>30</b>	<b>34</b>	<b>69</b>	<b>90</b>	<b>84</b>	<b>90</b>	<b>114</b>	<b>367</b>	<b>92</b>	<b>50</b>	<b>14</b>	<b>1</b>	<b>0</b>	<b>1126</b>

Figura 69 – Análise fornecida pelo cliente referente ao ano letivo 2012/2013

Mediante os valores apresentados foi aplicada a mesma metodologia, fórmula para analisar os dados utilizada na análise anterior e os resultados encontram-se presentes na Figura 70.

5	6	7	8	9	10	11	12	13	14	15
4,55%	10,53%	23,33%	20,99%	0,00%	7,48%	100,00%	100,00%	0,00%	0,00%	0,00%
1,89%	3,90%	7,69%	18,84%	8,11%	2,13%	17,53%	19,15%	33,33%	0,00%	0,00%
10,64%	0,00%	10,14%	14,29%	5,56%	2,99%	5,41%	10,00%	60,00%	28,40%	0,00%

Figura 70 - Comparação entre análise do AD e do cliente para o ano letivo 2012/2013

Como se pode constatar, existem bastantes diferenças entre o AD e a análise fornecida pelo cliente. Estas diferenças devem-se às mesmas encontradas na análise anterior pois utilizam a mesma fonte de dados de inscrições.

Por conseguinte, será feita a análise para o ano letivo 2013/2014. Na Figura 71 encontra-se novamente a análise com base no AD e na Figura 72 a análise fornecida pelo cliente. De igual forma encontra-se na Figura 73 a comparação efetuada para o ano letivo correspondente.

Number Alunos Insc		Número														
Ano	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total Geral	
1		4	8	13	6	24	22	35	30	28	245	2			417	
2			2		7	25	47	24	56	62	82	44	19	8	1	377
3			39	30	25	28	33	35	32	29	15	57	33	22	12	390
<b>Total Geral</b>		<b>45</b>	<b>38</b>	<b>38</b>	<b>41</b>	<b>82</b>	<b>104</b>	<b>91</b>	<b>115</b>	<b>105</b>	<b>384</b>	<b>79</b>	<b>41</b>	<b>20</b>	<b>1</b>	<b>1184</b>

Figura 71 – Análise proveniente do AD do ano letivo 2013/2014

		Nº de UCs a que o aluno está inscrito														
Ano	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	3	7	15	6	24	22	35	28	30	244	1					415
2	2			7	25	46	26	55	60	82	46	19	8	1		377
3	39	29	26	27	33	35	30	30	14	59	33	22	12			389
<b>Total</b>	<b>44</b>	<b>36</b>	<b>41</b>	<b>40</b>	<b>82</b>	<b>103</b>	<b>91</b>	<b>113</b>	<b>104</b>	<b>385</b>	<b>80</b>	<b>41</b>	<b>20</b>	<b>1</b>	<b>0</b>	<b>1181</b>

Figura 72 - Análise fornecida pelo cliente referente ao ano letivo 2013/2014

Ano	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	14,29%	6,67%	7,14%	0,00%	0,00%	0,00%	0,00%	3,45%	3,45%	0,20%	33,33%	0,00%	0,00%	0,00%	0,00%
2	0,00%	0,00%	0,00%	0,00%	0,00%	1,08%	4,00%	0,90%	1,64%	0,00%	2,22%	0,00%	0,00%	0,00%	0,00%
3	0,00%	1,69%	1,96%	1,82%	0,00%	0,00%	3,23%	1,69%	3,45%	1,72%	0,00%	0,00%	0,00%	0,00%	0,00%

Figura 73 - Comparação entre análise do AD e do cliente para o ano letivo 2013/2014

Mais uma vez pode-se verificar que existem algumas diferenças não tão acentuadas comparativamente ao ano letivo 2012/2013. É importante referir que se verifica o mesmo comportamento quando comparado com a análise da secção anterior para o mesmo ano letivo. Quanto aos desvios apresentados, estes baseiam-se no mesmo problema pois tem a mesma base.

Por fim, para o ano letivo 2014/2015 encontra-se na Figura 74 a análise reproduzida com base no AD, na Figura 75 a análise fornecida pelo cliente e na Figura 76 encontra-se a comparação realizada entre o AD e a análise do cliente respetivamente.

Number Alunos Insc		Número de Ucs a														
Ano	1	2	3	4	5	6	7	8	9	10	11	12	13	14	Total Geral	
1		7	2	6	10	20	22	33	48	18	247				413	
2		1	4	4	6	27	38	40	44	51	132	37	18	12	1	415
3		46	34	26	23	27	23	36	24	16	78	40	24	15	1	413
<b>Total Geral</b>		<b>54</b>	<b>40</b>	<b>36</b>	<b>39</b>	<b>74</b>	<b>83</b>	<b>109</b>	<b>116</b>	<b>85</b>	<b>457</b>	<b>77</b>	<b>42</b>	<b>27</b>	<b>2</b>	<b>1241</b>

Figura 74 - Análise proveniente do AD do ano letivo 2014/2015

Ano	Nº de UCs a que o aluno está inscrito															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	3	7	15	6	24	22	35	28	30	244	1					415
2	2			7	25	46	26	55	60	82	46	19	8	1		377
3	39	29	26	27	33	35	30	30	14	59	33	22	12			389
<b>Total</b>	<b>44</b>	<b>36</b>	<b>41</b>	<b>40</b>	<b>82</b>	<b>103</b>	<b>91</b>	<b>113</b>	<b>104</b>	<b>385</b>	<b>80</b>	<b>41</b>	<b>20</b>	<b>1</b>	<b>0</b>	<b>1181</b>

Figura 75 - Análise fornecida pelo cliente referente ao ano letivo 2013/2014

Ano	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	40,00%	55,56%	42,86%	25,00%	9,09%	0,00%	2,94%	26,32%	25,00%	0,61%	100,00%	0,00%	0,00%	0,00%	0,00%
2	33,33%	100,00%	100,00%	7,69%	3,85%	9,52%	21,21%	11,11%	8,11%	23,36%	10,84%	2,70%	20,00%	0,00%	0,00%
3	8,24%	7,94%	0,00%	8,00%	10,00%	20,69%	9,09%	11,11%	6,67%	13,87%	9,59%	4,35%	11,11%	8,24%	0,00%

Figura 76 - Comparação entre análise do AD e do cliente para o ano letivo 2014/2015

Como se pode constatar existem diferenças consideráveis entre o obtido no AD e a análise fornecida pelo cliente. É de salientar que o mesmo comportamento acontece para este ano letivo na análise apresentada na secção anterior, tendo como base as mesmas justificações para as diferenças encontradas.

### 6.1.3 Inscrições versus o número de cadeiras em atraso

Esta análise, como já foi identificado anteriormente, tem como base inscrições e avaliações de alunos. Visto que esta dissertação apenas se centra no tema de inscrições, todo o conteúdo de avaliações encontra-se explicado na dissertação da colega Gisela Couto.

O objetivo desta análise é mostrar o número de unidades curriculares a que o aluno se encontra inscrito e das respetivas quais terminou com reprovação.

Para o ano letivo 2012/2013, encontra-se presente na Figura 77 a análise reproduzida com base no AD, na Figura 78 a análise fornecida pelo cliente e na Figura 79 a comparação efetuada entre ambas.

Fact Avaliacao	Aggreg Count	Número Ucs Reprovadas						Total Geral
Número Ucs Inscritos	0	1	2	3	4	5	6	Total Geral
1		59	4					63
2		49	11					60
3		37	15	2	1			55
4		57	19	5				81
5		75	41	16	2			134
6		57	36	17	5	3		118
7		42	23	24	5	1		95
8		38	26	19	7	4	1	95
9		31	17	13	4	2	1	68
10		134	55	46	19	9	6	271
11		13	12	5	7	1	1	39
12		4	6	3	3	2		18
13			1	1				2
14		1						1
<b>Total Geral</b>		<b>597</b>	<b>266</b>	<b>151</b>	<b>53</b>	<b>21</b>	<b>10</b>	<b>2 1100</b>

Figura 77 - Análise proveniente do AD do ano letivo 2012/2013

Nº UCs	Nº de UCs em atraso							Total
	0	1	2	3	4	5	6	
1	41	14						55
2	20	10	5					35
3	10	11	8	1				30
4	13	12	7	1	1			34
5	33	16	6	3	5	6		69
6	44	15	11	10	8	2		90
7	29	13	20	18	3	1		84
8	44	12	16	10	7	1		90
9	48	18	13	22	12	1		114
10	325	17	6	8	10		1	367
11	5	45	28	4	9	1		92
12	1	3	17	18	11			50
13				7	7			14
14					1			1
15								0
<b>Total</b>	<b>613</b>	<b>186</b>	<b>137</b>	<b>102</b>	<b>74</b>	<b>12</b>	<b>1</b>	<b>1125</b>

Figura 78 - Análise fornecida pelo cliente referente ao ano letivo 2012/2013

Nº UCs *	0	1	2	3	4	5	6
1	18,00%	55,56%	0,00%	0,00%	0,00%	0,00%	0,00%
2	42,03%	4,76%	100,00%	0,00%	0,00%	0,00%	0,00%
3	57,45%	15,38%	60,00%	0,00%	0,00%	0,00%	0,00%
4	62,86%	22,58%	16,67%	100,00%	100,00%	0,00%	0,00%
5	38,89%	43,86%	45,45%	20,00%	100,00%	100,00%	0,00%
6	12,87%	41,18%	21,43%	33,33%	45,45%	100,00%	0,00%
7	18,31%	27,78%	9,09%	56,52%	100,00%	0,00%	0,00%
8	7,32%	36,84%	8,57%	17,65%	27,27%	0,00%	0,00%
9	21,52%	2,86%	0,00%	69,23%	71,43%	0,00%	0,00%
10	41,61%	52,78%	76,92%	40,74%	5,26%	100,00%	33,33%
11	44,44%	57,89%	69,70%	27,27%	80,00%	0,00%	0,00%
12	60,00%	33,33%	70,00%	71,43%	69,23%	0,00%	0,00%
13	0,00%	100,00%	100,00%	100,00%	100,00%	0,00%	0,00%
14	100,00%	0,00%	0,00%	0,00%	100,00%	0,00%	0,00%
15	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%

Figura 79 - Comparação entre análise do AD e do cliente para o ano letivo 2012/2013

Como se pode constatar através da Figura 79 existem bastantes diferenças entre a análise do AD e a análise fornecida pelo cliente. Os casos mais críticos são os que se encontram a cinza tendo uma percentagem de diferença muito grande. Para comprovar a veracidade dos dados contidos no AD fez-se uma análise detalhada aos alunos inscritos a 14 unidades curriculares e não tendo qualquer reprovação e a alunos inscritos em 13 unidades curriculares e com 1 uma unidade curricular reprovada.

Na primeira análise pesquisou-se no AD quais os alunos que estavam inscritos a 14 unidades curriculares e que tem marcadas como reprovadas 0 UCs para o ano letivo correspondente e obtiveram-se exatamente dois alunos, como anteriormente apresentado na Figura 77. De seguida pesquisou-se na fonte de dados pelo número de aluno dos retornados na pesquisa anterior, tanto na folha de avaliações como de inscrições e obteve-se o resultado presente Figura 80.

Numero	Nome	NFre	Exam	Nota	Disciplin	Epoc	Regime	AprovDisc
1100582		13,2	9,5	11	ALGAV	NM	Integral	Aprovado
1100582		16,8	2,2	SMR	ARQCP	NM	Integral	
1100582			11,6	14	ARQCP	RE	Integral	Aprovado
1100582		17	6,4	SMR	ARQSI	NM	Integral	
1100582			15	16	ARQSI	RE	Integral	Aprovado
1100582		12,2	7,9	10	ASIST	NM	Integral	Aprovado
1100582			14	14	COMP	NM	Integral	Aprovado
1100582		14,1	12	13	CORGA	NM	Integral	Aprovado
1100582			8,9	13	FSIAP	DZ	Integral	Aprovado
1100582		17	FT	NC	FSIAP	NM	Integral	
1100582		14,4	7,9	13	GESTA	NM	Integral	Aprovado
1100582			10,7	11	IARTI	NM	Integral	Aprovado
1100582		18	16	18	LAPR5	NM	Integral	Aprovado
1100582			FT	NC	MATCP	NM	Integral	
1100582			12	12	MATCP	RE	Integral	Aprovado
1100582			17	17	PESTI	DZ	Integral	Aprovado
1100582			FT	NC	PESTI	NM	Integral	
1100582		13,33	13,5	13	SCOMP	NM	Integral	Aprovado
1100582		15	11,6	13	SGRAI	NM	Integral	Aprovado

Figura 80 - Aluno Inscrito a 14 UCs e Reprovado a 0 UCs

Como se pode verificar, houve um aluno que esteve inscrito a 14 disciplinas e foi aprovado a todas, comprovando que a análise do AD está correta em detrimento da análise fornecida pelo cliente.

Na segunda análise utilizou-se a mesma metodologia. Foi feita uma análise sobre o AD para obter o aluno que está representado na análise. De seguida, fez-se uma pesquisa na fonte de dados e obteve-se o resultado presente na Figura 81.

Numero	Nome	NFre	Exam	Nota	Disciplin	Epoc	AprovDisc	
1110816			19		19	ALGAN	NM	Aprovado
1110816		17,4	13,2	15	ALGAV	NM	Aprovado	
1110816		14		14	AMATA	NM	Aprovado	
1110816		12,9	8	10	ARQCP	NM	Aprovado	
1110816		14	12	13	ARQSI	NM	Aprovado	
1110816		14	9,9	12	ASIST	NM	Aprovado	
1110816		12	8,5	10	BDDAD	NM	Aprovado	
1110816		12		12	COMP	NM	Aprovado	
1110816		14,8	12,5	14	CORGA	NM	Aprovado	
1110816		13,1	10,7	12	EAPLI	NM	Aprovado	
1110816		17		14	ESINF	NM	Aprovado	
1110816		12,2	1,6	9	FSIAP	NM	Reprovado	
1110816		10,7		9	FSIAP	NM		
1110816			8,55	10	FSIAP	RE	Aprovado	
1110816		13,3	12	13	GESTA	NM	Aprovado	
1110816		10,5		11	IARTI	NM	Aprovado	
1110816		12		12	LAPR3	NM	Aprovado	
1110816		14		14	LAPR4	NM	Aprovado	
1110816		13	14	13	LAPR5	NM	Aprovado	
1110816		16	10,7	13	LPROG	NM	Aprovado	
1110816			16	16	PESTI	NM	Aprovado	
1110816		15,6	17,33	16	PRCMP	NM	Aprovado	
1110816		16,1	10,9	14	RCOMP	NM	Aprovado	
1110816		13,67	14,8	14	SCOMP	NM	Aprovado	
1110816		14	13,2	13	SGRAI	NM	Aprovado	

Figura 81 - Aluno Inscrito a 13 UCs e Reprovado a 1 UCs

Uma das razões para estas diferenças é que segundo o cliente a sua análise contempla apenas avaliações de épocas normal e de recurso, sendo que a fonte de dados fornecida já contempla também outras épocas, como se pode verificar na Figura 82.

O que isto significa é que no caso da análise do cliente poderão estar a ser consideradas unidades curriculares que o aluno até ao momento se encontra reprovado, mas na época especial ter conseguido realizar com sucesso. Desta forma, no AD como está considerada a época especial os resultados apresentados são naturalmente diferentes. Outra possível razão adicional é que devido ao algoritmo de anonimização executado, este possivelmente adulterou os dados como comprovado anteriormente, e como esta análise tem como base as inscrições realizadas pelos alunos, ter afetado consequentemente os resultados da mesma.

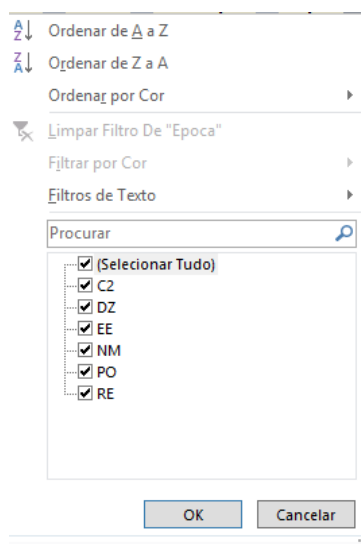


Figura 82 – Épocas existentes na fonte de dados.

Por fim, para os anos letivos 2013/2014 e 2014/2015 o comportamento é semelhante sendo que os desfasamentos visíveis se baseiam na justificação anteriormente referida para o ano letivo 2012/2013. No entanto, estão presentes na Figura 83 e na Figura 84 as análises efetuadas para o ano letivo 2013/2014 e 2014/2015, respetivamente.

Nº UCs *	0	1	2	3	4	5	6
1	3,90%	27,27%	0,00%	0,00%	0,00%	0,00%	0,00%
2	6,90%	28,57%	100,00%	0,00%	0,00%	0,00%	0,00%
3	10,34%	44,44%	33,33%	0,00%	0,00%	0,00%	0,00%
4	19,05%	4,76%	69,23%	33,33%	100,00%	0,00%	0,00%
5	11,36%	4,17%	11,11%	50,00%	50,00%	100,00%	0,00%
6	17,39%	3,33%	6,25%	50,00%	100,00%	100,00%	100,00%
7	12,20%	52,38%	3,70%	36,36%	33,33%	100,00%	0,00%
8	10,20%	26,92%	33,33%	33,33%	46,67%	33,33%	100,00%
9	13,76%	33,33%	8,33%	46,67%	28,57%	0,00%	100,00%
10	22,91%	58,00%	66,67%	45,00%	40,00%	71,43%	100,00%
11	94,59%	25,71%	46,67%	0,00%	0,00%	66,67%	100,00%
12	100,00%	100,00%	21,43%	30,77%	100,00%	0,00%	100,00%
13	100,00%	100,00%	100,00%	55,56%	85,71%	0,00%	0,00%
14	0,00%	0,00%	100,00%	0,00%	100,00%	0,00%	0,00%
15	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%

Figura 83 - Análise comparativa do Ano Letivo 2013/2014

Nº UCs *	0	1	2	3	4	5	6
1	2,63%	27,27%	0,00%	0,00%	0,00%	0,00%	0,00%
2	0,00%	5,88%	100,00%	0,00%	0,00%	0,00%	0,00%
3	1,96%	36,84%	20,00%	100,00%	0,00%	0,00%	0,00%
4	5,56%	9,09%	15,79%	100,00%	0,00%	0,00%	0,00%
5	8,33%	3,85%	53,85%	50,00%	100,00%	100,00%	0,00%
6	1,30%	14,81%	15,38%	71,43%	100,00%	100,00%	100,00%
7	3,37%	62,26%	12,00%	57,89%	0,00%	100,00%	0,00%
8	13,68%	26,92%	46,67%	27,27%	69,23%	33,33%	100,00%
9	27,84%	25,00%	10,34%	11,11%	80,00%	100,00%	0,00%
10	13,42%	66,40%	77,14%	31,25%	0,00%	33,33%	0,00%
11	90,00%	15,79%	7,32%	9,09%	33,33%	100,00%	100,00%
12	100,00%	100,00%	17,24%	25,93%	16,67%	100,00%	0,00%
13	100,00%	100,00%	100,00%	27,27%	85,71%	0,00%	0,00%
14	0,00%	100,00%	100,00%	0,00%	100,00%	0,00%	0,00%
15	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%

Figura 84 - Análise comparativa do Ano Letivo 2014/2015

#### 6.1.4 Unidades curriculares em atraso, por ano curricular e por ano de entrada do aluno

Esta análise consiste em analisar o número de UCs que os alunos reprovaram de acordo com o ano curricular em que se encontram.

No ano letivo 2012/2013 foi reproduzida a análise do cliente com base no AD e esta encontra-se presente na Figura 85.

Fact Avaliacao Aggreg Count		Número Ucs Reprovadas							
Ano Curricular	0	1	2	3	4	5	6	Total Geral	
1		165	89	70	22	11	7	2	366
2		193	88	43	16	5	2		347
3		239	89	38	15	5	1		387
<b>Total Geral</b>		<b>597</b>	<b>266</b>	<b>151</b>	<b>53</b>	<b>21</b>	<b>10</b>	<b>2</b>	<b>1100</b>

Figura 85 - Análise com base no AD para o ano letivo 2012/2013

Por sua vez, na Figura 79 encontra-se a análise fornecida pelo cliente e a respetiva comparação das mesmas na Figura 87.

Ano	Nº de UCs em atraso							Total
	0	1	2	3	4	5	6	
1	367							367
2	107	86	63	52	34	8	1	351
3	139	100	74	50	40	4		407
<b>Total</b>	<b>613</b>	<b>186</b>	<b>137</b>	<b>102</b>	<b>74</b>	<b>12</b>	<b>1</b>	<b>1125</b>

Figura 86 – Análise do cliente para o ano letivo 2012/2013

Nº UCs *	0	1	2	3	4	5	6
1	37,97%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
2	28,67%	1,15%	18,87%	52,94%	74,36%	60,00%	100,00%
3	26,46%	5,82%	32,14%	53,85%	77,78%	60,00%	0,00%

Figura 87 - Comparação Ano Letivo 2012/2013

Como se pode verificar na Figura 87, as discrepâncias entre a análise reproduzida com base no AD e a do cliente continuam a ser grandes, como anteriormente verificado. A diferença em relação à análise do subcapítulo anterior é que em vez de relacionar com o número de inscrições efetuadas a UCs é utilizado o ano curricular em que o aluno se encontra. Sendo que ambos têm a mesma base para realizar a análise, as razões são as mesmas enunciadas na análise da secção 6.1.3.

Para finalizar é importante referir que o mesmo comportamento se verifica para os anos letivos seguintes, sendo apenas apresentado este ano letivo.

### 6.1.5 Inscrições por ano curricular versus o número de cadeiras aprovadas

Por fim, esta é a última análise que contempla inscrições ainda que relacionado com avaliações. Esta consiste em revelar o número de unidades curriculares a que o aluno se encontra inscrito e das respetivas quais terminou com aprovação.

Por forma a comprovar a veracidade dos factos contidos no AD serão apresentadas as análises realizadas no AD, as análises realizadas pelo cliente e a respetiva comparação para os anos letivos 2012/2013, 2013/2014 e 2014/2015.

Analisando o ano letivo 2013/2014 encontra-se na Figura 88 a análise reproduzida com base no AD, na Figura 88 a análise fornecida pelo cliente e na Figura 90 a respetiva comparação.

Fact Avaliacao	Aggreg Count	Número Ucs Aprovadas	1	2	3	4	5	6	7	8	9	10	11	12	13	Total Geral
Número Ucs Inscritos	0															
1	20		24													44
2	14		6	18												38
3	12		7	3	16											38
4	20		4	4	5	7										40
5	26		11	5	13	12	15									82
6	14		9	13	15	15	21	17								104
7	13		8	12	10	11	13	11	14							92
8	12		12	13	13	14	13	11	12	13						113
9	11		7	15	13	12	10	7	8	12	11					106
10	28		16	10	11	21	12	20	31	35	54	145				383
11	4		2	4	2	4	6	8	7	6	12	13	10			78
12			2	3	1	4	6	3	4	5	3	7	1	2		41
13				1			2	1	3	2	2	2	2	2	3	20
14									1							1
<b>Total Geral</b>	<b>174</b>		<b>108</b>	<b>101</b>	<b>99</b>	<b>100</b>	<b>98</b>	<b>78</b>	<b>80</b>	<b>73</b>	<b>82</b>	<b>167</b>	<b>13</b>	<b>4</b>	<b>3</b>	<b>1180</b>

Figura 88 - Análise com base no AD para o ano letivo 2013/2014

Nº UCs *	Número de UCs em que o aluno obteve aprovação														Total	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13		14
1	16	28														44
2	12	2	22													36
3	14	6	5	16												41
4	19	5	4	4	8											40
5	26	10	5	12	13	16										82
6	14	9	12	13	14	21	20									103
7	13	7	12	8	12	11	12	16								91
8	12	12	13	12	10	15	10	14	15							113
9	11	7	15	14	12	9	7	7	10	12						104
10	28	16	10	11	19	13	21	28	37	52	150					385
11	4	2	4	2	5	5	7	6	7	9	11	18				80
12	2	2	3	1	4	5	3	5	1	5	7	3	2			41
13			1			2	1	3	2	2	2	1	2	4		20
14								1								1
15																0
<b>Total</b>	<b>169</b>	<b>106</b>	<b>106</b>	<b>93</b>	<b>97</b>	<b>97</b>	<b>81</b>	<b>80</b>	<b>72</b>	<b>80</b>	<b>170</b>	<b>22</b>	<b>4</b>	<b>4</b>	<b>0</b>	<b>1181</b>

Figura 89 - Análise do cliente para o ano letivo 2013/2014

Nº UCs *	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	11,11%	7,69%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
2	7,69%	50,00%	10,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
3	7,69%	7,69%	25,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
4	2,56%	11,11%	0,00%	11,11%	6,67%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
5	0,00%	4,76%	0,00%	4,00%	4,00%	3,23%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
6	0,00%	0,00%	4,00%	7,14%	3,45%	0,00%	8,11%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
7	0,00%	6,67%	0,00%	11,11%	4,35%	8,33%	4,35%	6,67%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
8	0,00%	0,00%	0,00%	4,00%	16,67%	7,14%	4,76%	7,69%	7,14%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
9	0,00%	0,00%	0,00%	3,70%	0,00%	5,26%	0,00%	6,67%	9,09%	4,35%	0,00%	0,00%	0,00%	0,00%	0,00%
10	0,00%	0,00%	0,00%	0,00%	5,00%	4,00%	2,44%	5,08%	2,78%	1,89%	1,69%	0,00%	0,00%	0,00%	0,00%
11	0,00%	0,00%	0,00%	0,00%	11,11%	9,09%	6,67%	7,69%	7,69%	14,29%	8,33%	28,57%	0,00%	0,00%	0,00%
12	0,00%	0,00%	0,00%	0,00%	0,00%	9,09%	0,00%	11,11%	66,67%	25,00%	0,00%	50,00%	0,00%	0,00%	0,00%
13	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	33,33%	0,00%	14,29%	0,00%
14	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%
15	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%

Figura 90 - Comparação Ano Letivo 2013/2014

Como mais uma vez se pode comprovar, através da Figura 90, existem diferenças entre a análise realizada com base no AD e a fornecida pelo cliente. Por forma a comprovar a veracidade dos

registos do AD foi feita uma análise detalhada aos registos que se encontram a cinzento, pois são os mais críticos de acordo com a escala definida. Inicialmente foi feita uma análise detalhada aos alunos inscritos a 12 unidades curriculares com aprovação a 8 unidades curriculares.

Para esta análise efetuou-se uma pesquisa no AD, mais propriamente na tabela de factos que alimenta esta análise, *Fact\_Avaliacao\_Agreg\_Count*, em que o número de inscrições efetuadas são 12 e número de UCs a que obteve aprovação de 8 UCs. De seguida, foi feita uma pesquisa na fonte de dados de modo a comprovar se estes registos estão corretos ou não.

Como se pode verificar na Figura 91, existem cinco alunos que se inscreveram a doze disciplinas e ficaram aprovados apenas a oito, ao contrário do que é apresentado na análise do cliente.

Numero	Nome	NFre	Exar	Noti	Disciplin	Epoca	AprovDisc	Ano
1100725		12	9	11	ARQSI	NM	Aprovado	20132014
1100725		9,8	9,1	10	ASIST	NM	Aprovado	20132014
1100725			8	12	BDDAD	NM	Aprovado	20132014
1100725			12	12	COMP	RE	Aprovado	20132014
1100725		14,55	12	13	CORGA	NM	Aprovado	20132014
1100725			14,3	11	GESTA	RE	Aprovado	20132014
1100725		16	9	15	LAPR5	NM	Aprovado	20132014
1100725		17	8,3	11	SGRAI	NM	Aprovado	20132014
1110858			12	12	ARQSI	RE	Aprovado	20132014
1110858			10,1	10	ASIST	RE	Aprovado	20132014
1110858			9,1	13	BDDAD	NM	Aprovado	20132014
1110858		11,8	11	11	CORGA	NM	Aprovado	20132014
1110858		11,6	9,9	11	GESTA	NM	Aprovado	20132014
1110858		11		11	LAPR4	NM	Aprovado	20132014
1110858		16	18	16	LAPR5	NM	Aprovado	20132014
1110858		14	8	10	SGRAI	NM	Aprovado	20132014
1110896		14	10	12	ARQSI	NM	Aprovado	20132014
1110896		10,8	10,1	11	ASIST	NM	Aprovado	20132014
1110896			11	12	BDDAD	RE	Aprovado	20132014
1110896			12,8	14	CORGA	RE	Aprovado	20132014
1110896			11,8	12	GESTA	RE	Aprovado	20132014
1110896			10,6	11	IARTI	NM	Aprovado	20132014
1110896		18	17	18	LAPR5	NM	Aprovado	20132014
1110896			15	15	PESTI	NM	Aprovado	20132014
1110936		12,7	16,3	15	ALGAV	NM	Aprovado	20132014
1110936		14	10	12	ARQSI	NM	Aprovado	20132014
1110936		12,8	10,6	12	ASIST	NM	Aprovado	20132014
1110936		10		10	COMP	NM	Aprovado	20132014
1110936		16,2	12,5	14	CORGA	NM	Aprovado	20132014
1110936		10,2	10,5	10	GESTA	NM	Aprovado	20132014
1110936		14	16	14	LAPR5	NM	Aprovado	20132014
1110936		15	8,5	10	SGRAI	NM	Aprovado	20132014
1121202			12,1	12	ARQCP	EE	Aprovado	20132014
1121202		13,0	8,5	11	BDDAD	NM	Aprovado	20132014
1121202			9,1	11	EAPLI	RE	Aprovado	20132014
1121202		10,0		10	LAPR3	NM	Aprovado	20132014
1121202		11,0		11	LAPR4	NM	Aprovado	20132014
1121202			10,5	12	LPROG	EE	Aprovado	20132014
1121202		11,1	8,3	10	MDISC	NM	Aprovado	20132014
1121202		11,1	7,5	10	RCOMP	NM	Aprovado	20132014
1121202			8,8	12	SCOMP	EE	Aprovado	20132014

Figura 91 - Alunos inscritos a 12 UCs e aprovados a 8, no ano letivo 2013/2014

Desta forma, mais uma vez fica provada a veracidade dos dados existentes no AD, permitindo ao cliente obter uma maior precisão nos valores resultantes das análises.

Por fim, para os outros anos letivos a análise foi efetuada e apresentam também diferenças que se baseiam nas mesmas apresentadas até ao momento.

## **6.2 Análises complementares**

Nesta secção serão apresentadas as várias análises adicionais anteriormente definidas recorrendo à ferramenta *Power Bi*.

Também, a cada análise apresentada será descrito o valor acrescentado que esta traz para o cliente.

É importante salientar que depois da análise intensiva efetuada às análises fornecidas pelo cliente e reproduzidas no AD são conhecidas as diferenças e ficou provada a veracidade do AD com base nalguns casos demonstrados.

### **6.2.1 Inscrições a unidades curriculares por tipo de inscrição**

Esta análise consiste em apresentar as unidades curriculares a que os alunos se inscrevem por tipo de inscrição: normal, em atraso ou em avanço.

Esta análise adiciona uma mais valia na medida em que é possível visualizar e analisar o tipo de inscrições que se efetuam nas diferentes UCs, podendo retirar daí conclusões sobre quais não estão a obter resultados positivos.

Na Figura 92 encontra-se um exemplo da estrutura da análise realizada sobre o AD para o ano letivo 2013/2014.

### Inscrições a unidades curriculares por tipo de inscrição

Descritivo	Em atraso	Em avanço	Normal	Total
ALGAN	44		325	369
ALGAV		8	258	266
AMATA	139		332	471
APROG	37		316	353
ARQCP	40	6	276	322
ARQSI		10	218	228
ASIST		10	193	203
BDDAD	32	24	305	361
COMPA		13	222	235
CORGA		15	149	164
EAPLI	35	5	268	308
ESINF	57	8	278	343
ESOFT	54		368	422
FSIAP	142	18	281	441
GESTA		11	142	153
IARTI		18	249	267
LAPR1	9		322	331
LAPR2	58		355	413
LAPR3	8	3	242	253
LAPR4	34		262	296
LAPR5		1	178	179
LPROG	27	4	308	339
MATCP	105		364	469
MDISC	78		366	444
PESTI			239	239
PPROG	65		378	443
PRCMP	23		320	343
RCOMP	16	8	244	268
SCOMP	48	2	184	234
SGRAI		10	243	253
<b>Total</b>	<b>1051</b>	<b>174</b>	<b>8185</b>	<b>9410</b>

Figura 92 – Inscrições a unidades curriculares por tipo de inscrição para o ano letivo 2013/2014

### 6.2.2 Alunos inscritos por regime em cada turma

Permite visualizar o número de alunos em cada turma por regime em que se encontra inscrito. Esta análise adiciona valor porque, o regime a que o aluno se inscreveu pode ter influência na realização da UC. No caso de um aluno que vem de ERASMUS este provavelmente necessitará de mais apoio para se integrar e acompanhar o conteúdo da UC.

De modo a exemplificar esta análise é apresentado um exemplo da mesma na Figura 93 para o ano letivo 2013/2014.

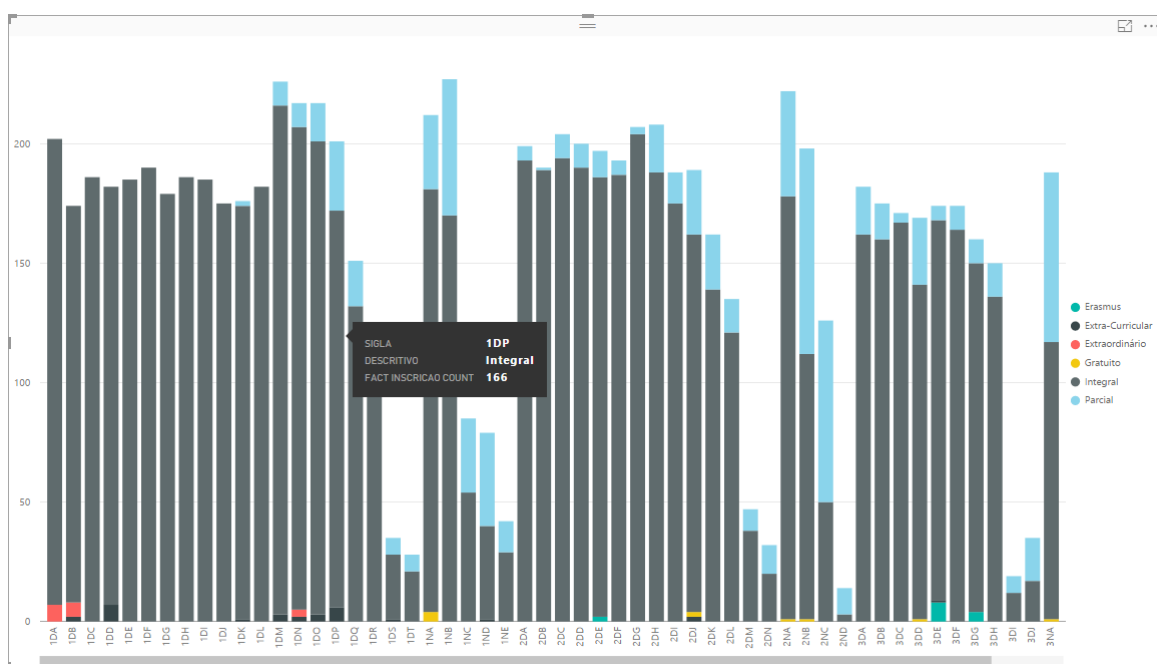


Figura 93 – Alunos inscritos por regime em cada turma, no ano letivo 2013/2014

Como se pode verificar na Figura 93, verifica-se que a maior percentagem de alunos por turma pertence ao regime integral, e uma menor parte ao regime parcial.

### 6.2.3 Alunos inscritos por tipo de horário por cada ano curricular

A presente análise contempla os alunos que se inscreveram por exemplo em Laboral ou Pós-Laboral de acordo com o ano curricular em que se encontra. Desta forma é perceptível visualizar a distribuição de alunos de acordo com o ano curricular em que se encontra e caso se encontra, por exemplo, em Laboral ou Pós-Laboral.

Na Figura 94 encontra-se um exemplo da análise realizada sobre o AD, mais especificamente para o ano letivo 2013/2014. A sigla neste caso tem o significado do ano curricular.

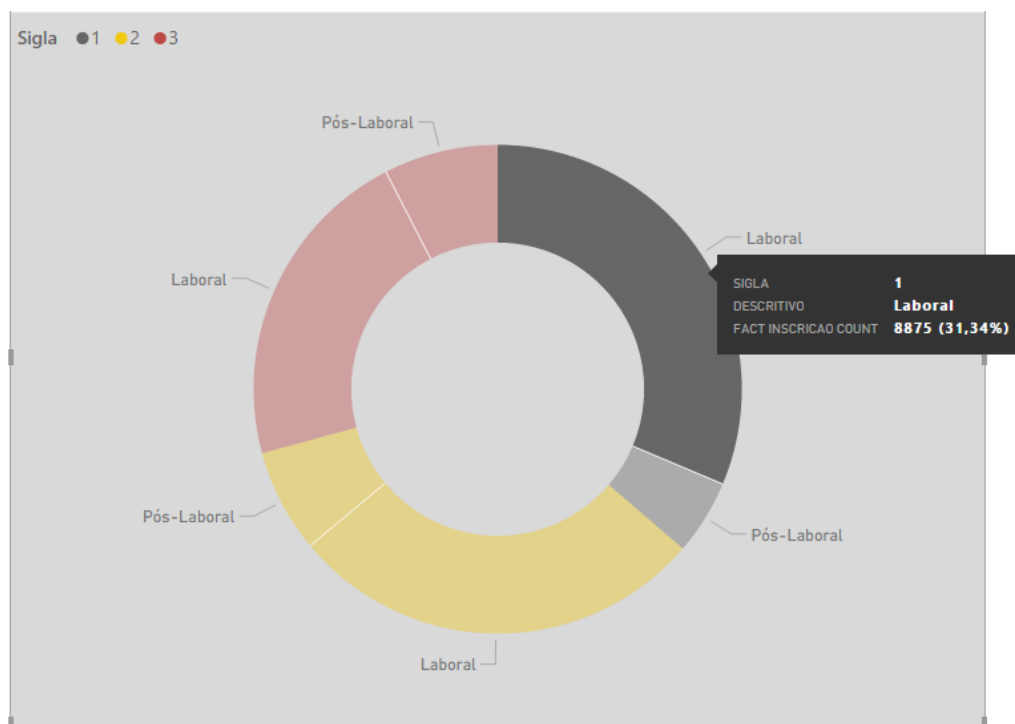


Figura 94 – Alunos inscritos por tipo de horário em cada ano curricular para o ano letivo 2013/2014

Como se pode verificar através da Figura 94 é possível verificar que a distribuição dos alunos diurnos e noturnos é sensivelmente a mesma nos anos letivos presentes.

#### 6.2.4 Alunos inscritos por ano curricular versus alunos inscritos em atraso

Esta análise possibilita visualizar quantos alunos inscritos a cada ano curricular, a um determinado número de unidades curriculares versus o número de unidades curriculares em atraso. Por outras palavras permite ver quantos alunos se inscreveram pelo menos a uma ou mais unidades curriculares e quantos fizeram zero ou mais inscrições em atraso. No caso de ser zero é o caso normal em que se está a inscrever no período regular à unidade curricular, ou seja, não a está a repetir.

Deste modo, esta adiciona valor pois permite perceber de acordo com o número de disciplinas a que um aluno se inscreveu, tem um maior número de alunos inscritos em atraso. Assim permite verificar a distribuição de alunos que se inscrevem a um determinado número de UCs versus as disciplinas que possuem em atraso.

Na Figura 95 encontra-se um exemplo desta análise para o ano letivo 2013/2014.

Número de Inscrições versus número de inscrições efetuadas em atraso

Category Number	0	1	2	3	4	5	6	Total
1	38	7						<b>45</b>
2	29	9						<b>38</b>
3	23	13	2					<b>38</b>
4	17	10	11	3				<b>41</b>
5	39	25	10	3	3	2		<b>82</b>
6	38	31	16	12	5	1	1	<b>104</b>
7	45	10	15	15	2	4		<b>91</b>
8	57	19	12	13	11	2	1	<b>115</b>
9	61	16	14	4	9	1		<b>105</b>
10	340	20	6	11	6	1		<b>384</b>
11	2	43	22	5	2	5		<b>79</b>
12			17	17	7			<b>41</b>
13				7	13			<b>20</b>
14					1			<b>1</b>
<b>Total</b>	<b>689</b>	<b>203</b>	<b>125</b>	<b>90</b>	<b>59</b>	<b>16</b>	<b>2</b>	<b>1184</b>

Figura 95 – Alunos inscritos versus inscrições realizadas em atraso, no ano letivo 2013/2014

De acordo com a análise presente na Figura 95, pode-se verificar que os alunos com zero cadeiras em atraso estão em maioria, seguido dos alunos com apenas uma inscrição em atraso.

### 6.2.5 Alunos inscritos por ano curricular versus alunos inscritos em avanço

Esta análise permite visualizar quantos alunos inscritos a um determinado número de unidades curriculares por ano curricular versus o número de unidades curriculares em avanço. É semelhante à análise anterior, diferindo no caso desta permitir visualizar unidades curriculares em avanço e não em atraso. O exemplo desta análise está presente na Figura 96, para o ano letivo 2014/2015.

Número de inscrições versus o número de inscrições efetuadas em avanço

Category Number	0	1	2	3	4	5	7	Total
1	54							54
2	40							40
3	36							36
4	39							39
5	70	3	1					74
6	72	9	2					83
7	93	7	8	1				109
8	96	7	6	6	1			116
9	80	3	1	1				85
10	456					1		457
11	77							77
12	42							42
13	26						1	27
14	2							2
<b>Total</b>	<b>1183</b>	<b>29</b>	<b>18</b>	<b>8</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1241</b>

Figura 96 - Alunos inscritos versus inscrições realizadas em avanço, no ano letivo 2014/2015

De acordo com a Figura 96 pode-se verificar que são poucos os alunos que efetuam inscrições em avanço.

### 6.2.6 Alunos inscritos em cada regime por tipo de inscrição

Esta análise possibilita visualizar os alunos inscritos às diferentes UCs em cada regime por cada tipo de inscrição (avanço, atraso ou normal). Desta forma é possível verificar a influência do regime em que o aluno está inscrito no tipo de inscrições que efetua às diferentes UCs.

Na Figura 97 é apresentado um exemplo da análise para o ano letivo 2014/2015. Para contextualizar o *popup* apresentado, o primeiro descritivo representa o regime e o segundo descritivo o tipo de inscrição. Por fim, o campo *Fact Inscricao Count* representa o número de inscrições efetuadas às diferentes UCs, neste caso em regime Parcial e com o tipo de inscrição normal.

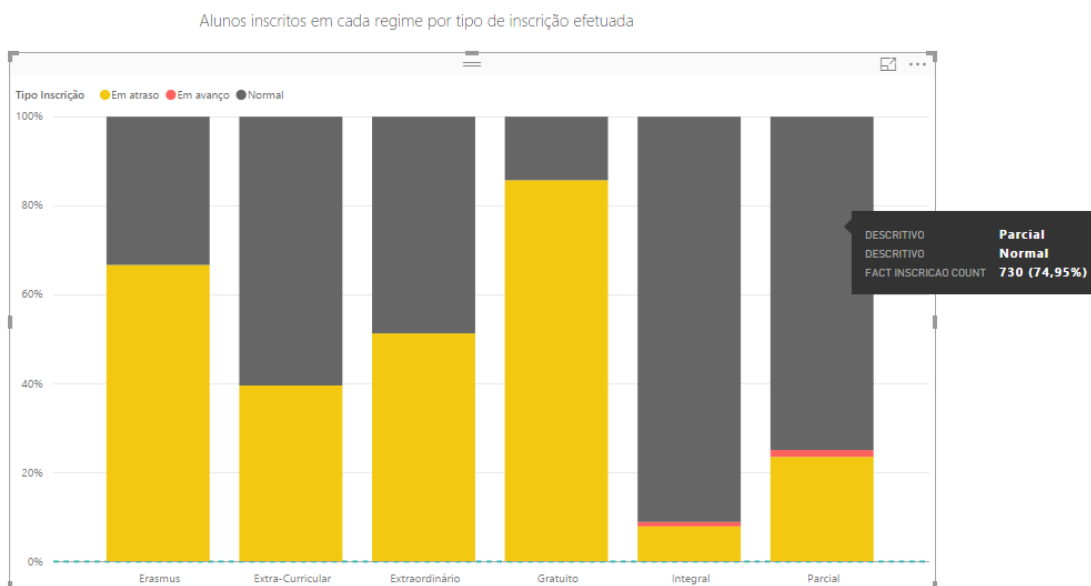


Figura 97 – Alunos inscritos em cada regime por tipo de inscrição, no ano letivo 2014/2015

De acordo com o gráfico presente na Figura 97 verifica-se que nos regimes parcial e integral as inscrições são na sua maioria feitas com o tipo de inscrição normal. No entanto, o paradigma altera-se nos outros regimes sendo que a maioria do tipo de inscrição é feita em atraso.

### 6.2.7 Alunos inscritos por ano de entrada em cada regime

Permite obter uma visão geral do número de alunos inscritos em cada regime, de acordo com o seu ano de entrada. Esta análise acrescenta valor na medida em que é possível obter uma visão geral ao longos dos vários anos, quantos alunos se inscreveram nos diferentes tipos de regime. Na Figura 98 encontra-se um exemplo desta análise para o ano letivo 2014/2015.

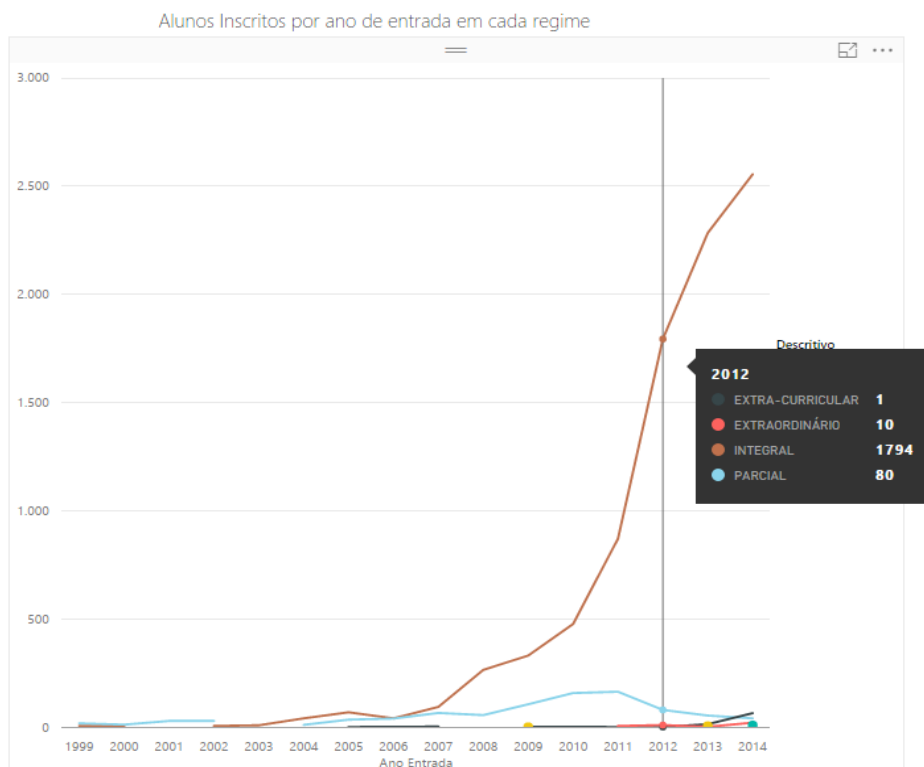


Figura 98 - Alunos inscritos por ano de entrada em cada regime, no ano letivo 2014/2015

De acordo com a Figura 98 verifica-se que o número de alunos em regime integral possui uma diferença notória quando comparado com os outros regimes.

### 6.3 Avaliação do Sistema

Um dos objetivos propostos a atingir era conseguir reduzir o tempo de atualização de dados e obtenção das análises que o cliente possuía na antiga solução.

Visto que o cliente utilizava o *Microsoft Excel* com fonte de dados e tratamento dos mesmos penalizava em muito o tempo de atualização das análises. Outro problema que também possuía era de ter de copiar para novas folhas as análises reproduzidas na tabela dinâmica criada. Segundo o cliente, o tempo de atualização e disponibilização das análises poderia levar até quatro horas.

Com esta solução é possível disponibilizar todos os dados já tratados e atualizados em quinze minutos para um ano letivo, como se pode verificar na Figura 99. Este tempo, apesar de um pouco elevado para o número de registos tratados, deve-se pelo facto de o *Microsoft SSIS* utilizar ao máximo a memória disponível em sistema, sendo que esta máquina é virtual com um processador *dual core* e oito gigabytes de ram para executar o processo. Também existe ainda a agravante dos recursos da máquina serem partilhados, o que afeta diretamente o processamento devido à carga existente no momento.

	audit_id	package_name	start_time	end_time	result	fig_success	number_datasource_rows
361	1102	DW_Update_SSAS_Cube_Insc	2016-08-01 22:28:21.0000000	2016-08-01 22:28:29.9000000	Success	1	NULL
362	1101	DW_Update_SSAS_Cube_InscAv	2016-08-01 22:27:43.0000000	2016-08-01 22:28:19.5400000	Success	1	NULL
363	1100	DW_FactAvaliacaoAggreg	2016-08-01 22:27:32.0000000	2016-08-01 22:27:42.4730000	Success	1	0
364	1099	DW_FactInscricao_AG	2016-08-01 22:27:29.0000000	2016-08-01 22:27:31.8700000	Success	1	1188
365	1098	DW_FactAvaliacao	2016-08-01 22:27:23.0000000	2016-08-01 22:27:28.0930000	Success	1	0
366	1097	DW_FactInscricao	2016-08-01 22:27:03.0000000	2016-08-01 22:27:21.2930000	Success	1	9414
367	1096	DW_DimTurma	2016-08-01 22:27:00.0000000	2016-08-01 22:27:01.3000000	Success	1	60
368	1095	DW_DimTipoInscricao	2016-08-01 22:26:57.0000000	2016-08-01 22:26:59.2270000	Success	1	0
369	1094	DW_DimTipoClassificacao	2016-08-01 22:26:55.0000000	2016-08-01 22:26:56.7700000	Success	1	0
370	1093	DW_DimRegime	2016-08-01 22:26:53.0000000	2016-08-01 22:26:55.0070000	Success	1	0
371	1092	DW_DimHorario	2016-08-01 22:26:52.0000000	2016-08-01 22:26:53.2770000	Success	1	3
372	1091	DW_Disciplina	2016-08-01 22:26:49.0000000	2016-08-01 22:26:51.3500000	Success	1	0
373	1090	DW_DimCurso	2016-08-01 22:26:48.0000000	2016-08-01 22:26:49.3730000	Success	1	0
374	1089	DW_DimAnoCurricular	2016-08-01 22:26:46.0000000	2016-08-01 22:26:47.5730000	Success	1	3
375	1088	DW_DimAluno	2016-08-01 22:26:43.0000000	2016-08-01 22:26:45.5570000	Success	1	1190
376	1087	STG_Turma	2016-08-01 22:26:41.0000000	2016-08-01 22:26:42.4330000	Success	1	60
377	1086	STG_TipoInscricao	2016-08-01 22:26:38.0000000	2016-08-01 22:26:40.5800000	Success	1	3
378	1085	STG_TipoClassificacao	2016-08-01 22:26:36.0000000	2016-08-01 22:26:38.2200000	Success	1	12
379	1084	STG_Regime	2016-08-01 22:26:33.0000000	2016-08-01 22:26:35.4200000	Success	1	9
380	1083	STG_Inscricao	2016-08-01 22:26:29.0000000	2016-08-01 22:26:32.8000000	Success	1	10364
381	1082	STG_Horario	2016-08-01 22:26:27.0000000	2016-08-01 22:26:29.2100000	Success	1	3
382	1081	STG_Disciplina	2016-08-01 22:26:25.0000000	2016-08-01 22:26:27.2800000	Success	1	33
383	1080	STG_Curso	2016-08-01 22:26:22.0000000	2016-08-01 22:26:24.4530000	Success	1	1
384	1079	STG_ANO_CURRICULAR	2016-08-01 22:26:21.0000000	2016-08-01 22:26:21.8470000	Success	1	3
385	1078	STG_Aluno	2016-08-01 22:26:17.0000000	2016-08-01 22:26:20.2730000	Success	1	1190
386	1077	STG_Avaliacao	2016-08-01 22:25:13.0000000	2016-08-01 22:26:15.9230000	Success	1	18946
387	1076	STG_Aluno	2016-08-01 22:18:30.0000000	2016-08-01 22:18:35.9000000	Success	1	1844
388	1075	STG_Avaliacao	2016-08-01 22:13:29.0000000	2016-08-01 22:18:29.3670000	Success	1	32278

Figura 99 - Logs do Processo de ETL e atualização dos Cubos

No entanto, comparativamente ao tempo indicado pelo cliente este processo tem uma vantagem nítida ao disponibilizar os dados e tratá-los sem a necessidade da intervenção humana, levando assim a uma maior confiabilidade nos dados disponibilizados.

Por outro lado, após estar concluído o processo de ETL apenas é necessário atualizar os dados existentes nas análises disponibilizadas e filtrar pelo ano letivo que pretende, não tendo que recriar novamente todas as análises. Deste modo, o cliente poupa imenso tempo de espera para visualizar os dados que pretende.

Por fim, no caso da usabilidade, o mecanismo atual de criação de novas análises é mais intuitivo, visto que apenas é necessário efetuar *drag and drop* dos campos a utilizar na ferramenta de criação das mesmas. É de salientar que estes campos são os disponibilizados pelos cubos analíticos.



# 7 Conclusão

## 7.1 Objetivos Alcançados

Ao longo dos vários anos, as inscrições e avaliações de alunos são dados importantes para classificar o desempenho da instituição e com base nas mesmas tomar decisões importantes para adequar os métodos de ensino praticados.

De acordo com o enunciado na secção Objetivos, foram traçados vários objetivos a serem alcançados. O primeiro objetivo enunciado foi o de desenvolver um AD que modele a informação de acordo com as especificidades da instituição e do curso. Posto isto, e de acordo com o que foi apresentado neste documento, esse objetivo foi cumprido porque as diferentes análises apresentadas foram fielmente reproduzidas com a estrutura que o cliente possui, estando assim familiarizado com a solução.

Como segundo objetivo era pretendido elaborar um processo de extração, de transformação e carregamento de dados adequado, por forma a disponibilizar informação mais precisa, mais rápida e confiável. Ao longo do capítulo de avaliação de resultados foi comprovado que os resultados obtidos no AD correspondem à realidade que está contida na fonte de dados. Também, ao revelar que o problema que origina as diferenças encontradas entre o AD e as análises disponibilizadas pelo cliente não se encontra nesta nova solução, permite demonstrar que os registos contidos no AD são confiáveis. Quanto à rapidez, esta foi claramente conseguida como se pode constatar na secção de Avaliação do Sistema, em que o processo de atualização e disponibilização dos dados acontece em quinze minutos, em vez das quatro horas indicadas pelo cliente.

No terceiro objetivo foi pretendido que esta solução contivesse retrocompatibilidade com a ferramenta de apresentação das análises que o cliente utilizava. Como se pôde observar na secção de Análises no Microsoft Excel, essa retrocompatibilidade foi conseguida

disponibilizando as análises que o cliente possuía na mesma ferramenta e com a mesma estrutura.

Como terceiro objetivo foi pretendido assegurar a qualidade de dados presentes no AD e esse foi assegurado como explicitado no segundo objetivo. Adicionalmente, para obter um maior controlo dos dados são mantidos os registos de auditoria aquando da execução do processo de ETL.

Por fim, o último objetivo pretendido foi o de criar novas análises de uma forma mais *self-service*. Esse objetivo foi cumprido pois apenas é necessário realizar a tarefa de *drag and drop* dos vários campos disponibilizados na ferramenta de construção e apresentação das mesmas.

Não obstante, foram ainda disponibilizadas novas análises recorrendo a outra ferramenta analítica, *Power BI*, permitindo assim colmatar as limitações da ferramenta que o cliente utiliza no momento. Também ao disponibilizar novas análises, é permitido ao cliente enriquecer o seu conhecimento no que toca à evolução do curso, sendo assim uma mais valia para esta nova solução.

## 7.2 Trabalho Futuro

Com o objetivo de melhorar o trabalho desenvolvido nesta dissertação foram identificados alguns pontos a desenvolver, por forma a aumentar a qualidade do AD. A primeira melhoria identificada é substituir os ficheiros Excel como fonte de dados, diretamente pelo Portal do ISEP permitindo assim obter a informação da respetiva fonte e não depender de ficheiros que possam conter erros.

Como segunda melhoria identificada, e tendo como base a melhoria anterior, tem-se o desenvolvimento de um mecanismo incremental de extração de dados por forma a permitir uma integração contínua de dados oriundos da respetiva fonte para o AD.

Por fim, após as melhorias anteriormente identificadas seria também interessante integrar os dados não só do curso da LEI no ISEP, mas sim de todos os outros cursos existentes na instituição por forma a obter um panorama geral do sucesso do ensino praticado.

## 8 Anexos

### 8.1 Análises do Cliente

Ano	UC	Ano em que o aluno se encontra			Total
		1	2	3	
1º	ALGAN	324	38	7	369
	AMATA	331	106	34	471
	APROG	315	37		352
	LAPR1	321	9		330
	PRCMP	318	20	3	341
	ESOFT	366	53		419
	LAPR2	354	56	5	415
	MATCP	362	87	17	466
	MDISC	364	71	7	442
	PPROG	377	65		442
2º	ARQCP	6	276	40	322
	BDDAD	24	305	32	361
	ESINF	8	278	57	343
	FSIAP	18	281	142	441
	LAPR3	3	242	8	253
	EAPLI	4	268	35	307
	LAPR4		262	35	297
	LPROG	3	308	27	338
	RCOMP	8	244	16	268
	SCOMP	2	184	49	235
3º	ALGAV	3	5	259	267
	ARQSI	4	6	218	228
	ASIST	2	8	193	203
	GESTA	1	10	142	153
	LAPR5		1	179	180
	SGRAI	4	6	243	253
	COMPA	2	11	222	235
	CORGA	2	13	149	164
	IARTI	2	15	248	265
	PESTI			240	240
	Total	3528	3265	2607	9400

Figura 100 – Alunos Inscritos a cada disciplina em cada ano curricular (análise 1)

Na Figura 100 está presente a análise disponibilizada pelo cliente no qual se pode visualizar os Alunos Inscritos a cada disciplina em cada ano curricular.

Nº de UCs a que o aluno está inscrito																
Ano	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	3	7	15	6	24	22	35	28	30	244	1					415
2	2			7	25	46	26	55	60	82	46	19	8	1		377
3	39	29	26	27	33	35	30	30	14	59	33	22	12			389
<b>Total</b>	<b>44</b>	<b>36</b>	<b>41</b>	<b>40</b>	<b>82</b>	<b>103</b>	<b>91</b>	<b>113</b>	<b>104</b>	<b>385</b>	<b>80</b>	<b>41</b>	<b>20</b>	<b>1</b>	<b>0</b>	<b>1181</b>

Figura 101 - Número de inscrições efetuadas por alunos em cada ano curricular (análise 2)

Na Figura 101 está presente a análise disponibilizada pelo cliente na qual se pode visualizar o Número de inscrições efetuadas por alunos em cada ano curricular.

Nº de UCs em atraso								
Nº UCs	0	1	2	3	4	5	6	Total
1	37	7						44
2	27	9						36
3	26	13	2					41
4	17	10	11	2				40
5	39	25	10	3	3	2		82
6	38	31	15	12	5	1	1	103
7	46	10	14	15	2	4		91
8	54	19	12	14	11	2	1	113
9	62	15	13	4	9	1		104
10	338	21	8	11	6	1		385
11	1	44	22	5	2	5	1	80
12			17	17	7			41
13				7	13			20
14					1			1
15								0
<b>Total</b>	<b>685</b>	<b>204</b>	<b>124</b>	<b>90</b>	<b>59</b>	<b>16</b>	<b>3</b>	<b>1181</b>

Figura 102 - Inscrições versus Unidades Curriculares Reprovadas (análise 3)

Na Figura 102 está presente a análise disponibilizada pelo cliente na qual se pode visualizar o Número de inscrições efetuadas versus o número de alunos que reprovaram a essas mesmas unidades curriculares a que se inscreveram.

Nº de UCs em atraso								
Ano	0	1	2	3	4	5	6	Total
1	415							415
2	137	92	52	54	29	10	3	377
3	133	112	72	36	30	6		389
<b>Total</b>	<b>685</b>	<b>204</b>	<b>124</b>	<b>90</b>	<b>59</b>	<b>16</b>	<b>3</b>	<b>1181</b>
	<b>58,0%</b>	<b>17,3%</b>	<b>10,5%</b>	<b>7,6%</b>	<b>5,0%</b>	<b>1,4%</b>	<b>0,3%</b>	

Figura 103 - Inscrições por ano letivo versus Unidades Curriculares Reprovadas (análise 4)

Na Figura 103 está presente a análise disponibilizada pelo cliente na qual se pode visualizar o por ano curricular do curso, o número de alunos que reprovaram a unidades curriculares a que se inscreveram.

Nº UCs *	Número de UCs em que o aluno obteve aprovação														Total	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13		14
1	16	28														44
2	12	2	22													36
3	14	6	5	16												41
4	19	5	4	4	8											40
5	26	10	5	12	13	16										82
6	14	9	12	13	14	21	20									103
7	13	7	12	8	12	11	12	16								91
8	12	12	13	12	10	15	10	14	15							113
9	11	7	15	14	12	9	7	7	10	12						104
10	28	16	10	11	19	13	21	28	37	52	150					385
11	4	2	4	2	5	5	7	6	7	9	11	18				80
12		2	3	1	4	5	3	5	1	5	7	3	2			41
13			1			2	1	3	2	2	2	1	2	4		20
14								1								1
15																0
<b>Total</b>	169	106	106	93	97	97	81	80	72	80	170	22	4	4	0	1181

Figura 104 - Inscrições versus unidades curriculares aprovadas (análise 5)

Na Figura 104 Figura 102 está presente a análise disponibilizada pelo cliente na qual se pode visualizar o Número de inscrições efetuadas versus o número de alunos que obtiveram aprovação a essas mesmas unidades curriculares a que se inscreveram.



# Referências

- 1keydata, 2015. MOLAP, ROLAP, and HOLAP [WWW Document]. URL <http://www.1keydata.com/datawarehousing/molap-rolap.html> (accessed 9.20.15).
- Anoop Kumar, 2013. Top 10 Common Transformations in SSIS - Developer.com [WWW Document]. URL <http://www.developer.com/db/top-10-common-transformations-in-ssis.html> (accessed 6.14.16).
- ARSON Group SAC, n.d. Oracle Data Integrator [WWW Document]. ARSON Group SAC. URL <http://www.arsongroup.com/web/productos/oracle-data-integrator/> (accessed 6.22.16).
- Atanazio, J., 2013. Conceituando BI – Parte IV: Diferenças entre OLTP x OLAP. bisaopaulo.com.
- Ballard, C., Farrell, D.M., Gupta, A., Mazuela, C., Stanislav, V., 2006. Dimensional Modeling: In a Business Intelligence Environment, First Edition. ed.
- CaseNex, 2010. DataCation [WWW Document]. URL <http://www.datacation.com/Services/Data%20Warehousing/> (accessed 1.31.16).
- Dean, T., 2015. Gain the Competitive Edge with Business Intelligence Software & Analytics [WWW Document]. URL <http://www.business2community.com/business-intelligence/gain-competitive-edge-business-intelligence-software-analytics-01280045> (accessed 9.15.15).
- Elias, D., 2015. Entendendo a modelagem multidimensional - Business Intelligence [WWW Document]. Canaltech. URL <http://corporate.canaltech.com.br/materia/business-intelligence/entendendo-a-modelagem-multidimensional-19988/> (accessed 9.15.15).
- ETL, D., 2015. ETL Tools - Top 10 ETL Tools Reviews - Database ETL.
- ETL Tools comparison, 2016.
- FenProf, 2012. O sistema de ensino Superior em Portugal.
- Ferreira, E., 2015. Experimentação e avaliação.
- Ferreira, J., Miranda, M., Abelha, A., Machado, J., 2016. O Processo ETL em Sistemas Data Warehouse.
- Filzmoser, M., Vetschera, R., 2008. A Classification of Bargaining Steps and their Impact on Negotiation Outcomes, in: Group Decision and Negotiation. Springer Science, pp. 421–443.
- Fundamentos e Modelagem de Bancos de Dados Multidimensionais [WWW Document], 2015. URL <https://msdn.microsoft.com/pt-br/library/cc518031.aspx> (accessed 9.17.15).
- G2 Crowd, 2016. MicroStrategy vs. Power BI | G2 Crowd [WWW Document]. Comp. MicroStrategy Vs Power BI. URL <https://www.g2crowd.com/compare/microstrategy-vs-microsoft-power-bi> (accessed 2.13.16).
- Henrique, O., 2012. Tabela Fato x Tabela Dimensão - artigos TechNet - Brasil (Português) - TechNet Wiki [WWW Document]. Tabela Fato X Tabela Dimens. URL <http://social.technet.microsoft.com/wiki/pt-br/contents/articles/12577.tabela-fato-x-tabela-dimensao.aspx> (accessed 9.16.15).
- IBM, 2015. IBM InfoSphere Platform – big data, information integration, data warehousing, master data management, lifecycle management and data security [WWW Document]. URL <http://www-01.ibm.com/software/data/infosphere/> (accessed 12.23.15).
- IBM, 2006. School District of Philadelphia improves performance with data-driven decision making 4.
- IBM, 1999. DATA WAREHOUSE OPERATIONAL ARCHITECTURE 14.

- InformationWeek, 2015. Put to the Test: Oracle Data Integrator - InformationWeek [WWW Document]. Put Test Oracle Data Integr. - InformationWeek. URL [http://www.informationweek.com/software/information-management/put-to-the-test-oracle-data-integrator/d/d-id/1054359?page\\_number=1](http://www.informationweek.com/software/information-management/put-to-the-test-oracle-data-integrator/d/d-id/1054359?page_number=1) (accessed 12.23.15).
- Inmon, W.H., 2002. Building the Data Warehouse, 3<sup>o</sup>. ed. John Wiley & Sons, Inc., Canada.
- Kimball, R., Caserta, J., 2004. The Data Warehouse ETL Toolkit. Wiley Publishing, Inc., Canada.
- Kimball, R., Ross, M., 2013. The Data Warehouse Toolkit Third Edition, 3<sup>a</sup>. ed. John Wiley & Sons, Inc., Canada.
- Martins, A., Costa, A., 2010. Mass Customization in Engineering Programs: A Framework for Program Management.
- McBurney, V., 2007. Wiki Wednesday: comparing Talend and Pentaho Kettle open source ETL tools [WWW Document]. URL <http://it.toolbox.com/blogs/infosphere/wiki-wednesday-comparing-talend-and-pentaho-kettle-open-source-etl-tools-16294> (accessed 2.19.16).
- Microsoft, 2016. Power BI - Descrição Geral e Aprendizagem - Power BI [WWW Document]. URL <https://support.office.com/pt-pt/article/Power-BI-Descri%C3%A7%C3%A3o-Geral-e-Aprendizagem-02730e00-5c8c-4fe4-9d77-46b955b71467> (accessed 2.13.16).
- Microsoft, 2015. SQL Server Integration Services [WWW Document]. URL [https://msdn.microsoft.com/en-us/library/ms141026\(v=sql.120\).aspx](https://msdn.microsoft.com/en-us/library/ms141026(v=sql.120).aspx) (accessed 12.26.15).
- MicroStrategy, 2015. About MicroStrategy | MicroStrategy [WWW Document]. Soluções Bus. Intell. Anal. Mob. MicroStrategy. URL <https://www.microstrategy.com/pt/sobre-nos/sobre-nos> (accessed 2.13.16).
- Moreira, E., 2006. Modelo Dimensional para Data Warehouse - [WWW Document]. URL <http://imasters.com.br/artigo/3836/gerencia-de-ti/modelo-dimensional-para-data-warehouse> (accessed 9.17.15).
- Narasimharajan], M., 2011. Integration, Maximize Success on Data.
- New York University, 2014. University Data Warehouse Plus [WWW Document]. URL <https://www.nyu.edu/employees/resources-and-services/administrative-services/university-data-warehouse-plus.html> (accessed 2.2.16).
- Nicola, S., 2016. Análise de valor de negócio.
- Nicola, S., Ferreira, E., Ferreira, J., 2014. A Quantitative Model for Decomposing & Assessing the 110–120.
- OLAP.com, 2016. What is the Definition of OLAP? OLAP Definition [WWW Document]. URL <http://olap.com/olap-definition/> (accessed 9.20.15).
- Oracle, 2007. Data Mart Concepts [WWW Document]. URL [http://docs.oracle.com/html/E10312\\_01/dm\\_concepts.htm](http://docs.oracle.com/html/E10312_01/dm_concepts.htm) (accessed 9.17.15).
- Oracle, 2002. Data Warehousing Concepts [WWW Document]. URL [http://docs.oracle.com/cd/B10500\\_01/server.920/a96520/concept.htm](http://docs.oracle.com/cd/B10500_01/server.920/a96520/concept.htm) (accessed 9.24.15).
- Oracle, O., 2016. Oracle Data Integrator [WWW Document]. Oracle Data Integr. URL <http://www.oracle.com/technetwork/middleware/data-integrator/overview/index.html> (accessed 2.13.16).
- Patusky, C., Botwinik, L., Shelley, M., 2007. The Philadelphia SchoolStat Model. Philadelphia.
- Pentaho, C., 2015. Data Integration | Pentaho Community [WWW Document]. Data Integr. - Kettle. URL <http://community.pentaho.com/projects/data-integration/> (accessed 12.29.15).
- Pereira, N., 2016. Escrita Técnico - científica.
- Phocas, S., 2015. Phocas: Successful Business Intelligence Software and Data Discovery [WWW Document]. URL <http://www.phocassoftware.com/> (accessed 2.19.16).

- Rouse, M., 2015. Mobile BI tools, trends and best practices guide [WWW Document]. TechTarget. URL <http://searchbusinessanalytics.techtarget.com/essentialguide/Mobile-BI-tools-trends-and-best-practices-guide> (accessed 9.15.15).
- Sansu George, 2012. Inmon vs. Kimball: Which approach is suitable for your data warehouse? [WWW Document]. ComputerWeekly. URL <http://www.computerweekly.com/tip/Inmon-vs-Kimball-Which-approach-is-suitable-for-your-data-warehouse> (accessed 6.25.16).
- Saúde, S., Borralho, C., Féria, I., Lopes, S., 2014. A necessária especificidade da avaliação de desempenho das Instituições de Ensino.
- School Buffalo, 2015. EdVantage Data Dashboard [WWW Document]. URL <http://www.buffaloschools.org/informationtech.cfm?subpage=66927> (accessed 1.31.16).
- SchoolCity Inc., 2015. School Data Warehousing | Edvantage™ [WWW Document]. URL <http://www.schoolcity.com/data-warehousing.html> (accessed 1.31.16).