INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO
MESTRADO EM ENGENHARIA INFORMÁTICA




Monitorização e Análise de Logs, Alerta de Erros e apresentação de Métricas de Negócio

FÁBIO DANIEL SILVA CRUZ
Setembro de 2025

POLITÉCNICO
DO PORTO

ISEP INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

Monitorização e Análise de Logs, Alerta de Erros e apresentação de Métricas de Negócio

Fábio Daniel Silva Cruz

Dissertação para obtenção do Grau de Mestre em Engenharia Informática, Área de Especialização em Engenharia de Software

Orientador: Bruno da Silva
Supervisor: Ricardo Leandro

Porto, 12 de setembro de 2025

Declaração de Integridade

Declaro ter conduzido este trabalho académico com integridade.

Não plagiei ou apliquei qualquer forma de uso indevido de informações ou falsificação de resultados ao longo do processo que levou à sua elaboração.

Portanto, o trabalho apresentado neste documento é original e de minha autoria, não tendo sido utilizado anteriormente para nenhum outro fim.

Importa ainda referir que recorri a ferramentas de Inteligência Artificial apenas como apoio auxiliar em tarefas de tradução e aperfeiçoamento da redação do texto. Sublinha-se que tais ferramentas não foram utilizadas para a produção de resultados de investigação, para a implementação técnica, nem para substituir a minha análise crítica. Todo o conteúdo científico, metodológico e experimental aqui apresentado é da minha exclusiva responsabilidade e autoria.

Declaro ainda que tenho pleno conhecimento do Código de Conduta Ética do P.PORTO.

ISEP, Porto, 12 de setembro de 2025

Resumo

Esta dissertação aborda o desenvolvimento de uma solução focada na monitorização, análise e geração de alertas para *logs* empresariais, com o objetivo de otimizar a gestão de eventos críticos em sistemas de informação. O projeto surge da necessidade de métodos mais eficientes e centralizados para lidar com grandes volumes de dados, superando as limitações das abordagens tradicionais, que muitas vezes são lentas e dependem de intervenções manuais.

A solução proposta traz uma abordagem inovadora, voltada para monitorizar *logs* e extrair métricas de negócio que possam facilitar a tomada de decisões e agilizar a resolução de problemas. Para isso, será utilizada uma ferramenta de monitorização escalável e flexível, capaz de integrar dados provenientes de diversas fontes e automatizar alertas em tempo real.

Entre as funcionalidades principais estão a criação de *pipelines* para processar e analisar logs, a extração de métricas relevantes e a configuração de alertas automáticos para eventos críticos. Ao longo do projeto vão ser executados testado iterativos para garantir a eficácia na identificação de problemas e na redução do tempo de resposta.

Os testes envolverão dados simulados e a recolha de *feedback* dos utilizadores por meio de um formulário, permitindo avaliar tanto a precisão quanto a utilidade da solução. Além disso, essas avaliações contribuirão com pontos de melhoria para futuras evoluções. Assim, esta dissertação propõe contribuir com uma solução robusta e escalável para a gestão de *logs*, promovendo maior eficiência operacional e reforçando a segurança em sistemas empresariais complexos.

Palavras-chave: monitorização de *logs*, automação, análise de dados, escalabilidade, eficiência operacional, alertas em tempo real.

Abstract

This dissertation addresses the development of a solution focused on monitoring, analyzing and generating alerts for corporate logs, with the aim of optimizing the management of critical events in information systems. The project arises from the need for more efficient and centralized methods for dealing with large volumes of data, overcoming the limitations of traditional approaches, which are often slow and rely on manual interventions.

The proposed solution brings an innovative approach, aimed at monitoring logs and extracting business metrics that can facilitate decision-making and speed up problem-solving. To do this, a scalable and flexible monitoring tool will be used, capable of integrating data from various sources and automating alerts in real time.

Key features include creating pipelines to process and analyze logs, extracting relevant metrics and setting up automatic alerts for critical events. After implementation, the project will be tested and adjusted to ensure its effectiveness in identifying problems and reducing response times.

The tests will involve simulated data and user feedback, making it possible to assess both the accuracy and usefulness of the solution. In addition, these evaluations will contribute points of improvement for future developments. Thus, this dissertation proposes to contribute with a robust and scalable solution for log management, promoting greater operational efficiency and reinforcing security in complex business systems.

Keywords: log monitoring, automation, data analysis, scalability, operational efficiency, real-time alerts.

Conteúdo

Lista de Figuras

Lista de Tabelas

Lista de Acrónimos

ACM	Código de Ética.
CI/CD	Integração Contínua/Entrega Contínua (Continuous Integration and Continuous Delivery).
ELK	*Elasticsearch, Logstash e Kibana*.
IA	Inteligência Artificial.
IAM	Gestão de Identidade e Acesso (Identity and Access Management).
IPP	Instituto Politécnico do Porto.
JVM	Máquina Virtual Java (Java Virtual Machine).
LSTM	Memória Curto Longo Prazo (Long Short-Term Memory).
ML	Aprendizagem Automática (Machine Learning).
MTTR	Tempo Médio de Resolução (Mean Time to Repair).
OLAP	Processamento Analítico em Linha (Online Analytical Processing).
OLTP	Processamento de Transacções em Linha (Online Transaction Processing).
PRISMA	Itens Preferidos para Relatórios de Revisões Sistemáticas e Meta-Análises.
RBAC	Controlo de Acesso Baseado em Funções (Role-Based Access Control).
RGPD	Regulamento Geral de Proteção de Dados.
SIEM	Gestão de Informações e Eventos de Segurança (Security Information and Event Management).
SMART	Específico, Mensurável, Atingível, Relevante, Temporal.
TLS	Segurança da Camada de Transporte (Transport Layer Security).

UX	Experiência do Utilizador (User Experience).
VPN	Rede Virtual Privada (Virtual Private Network).
WBS	Plano da Estrutura do Projeto.

1. Introdução

Este capítulo contextualiza esta dissertação, apresentando o problema em estudo e os objetivos estabelecidos para o desenvolvimento do trabalho. O projeto será realizado no âmbito da Cleva, uma empresa dedicada ao desenvolvimento de soluções tecnológicas para o setor de seguros.

O trabalho proposto surge como resposta a necessidades identificadas pela Cleva na gestão e análise de grandes volumes de *logs* empresariais. Este tema é especialmente relevante, dado o papel crucial que os *logs* desempenham na monitorização de sistemas, na identificação de problemas e na garantia da segurança das operações corporativas. A solução a ser desenvolvida será direcionada para os desafios e necessidades específicos do ambiente empresarial da Cleva, estando alinhada com os objetivos estratégicos da empresa e contribuindo para a melhoria contínua das suas operações tecnológicas.

1.1 Contexto

O trabalho surge da necessidade crescente da Cleva em lidar com o aumento exponencial do volume de dados gerados por sistemas e aplicações. Atualmente, a gestão de *logs* na Cleva enfrenta desafios significativos, como a dificuldade de centralizar informações provenientes de diferentes fontes, a complexidade na extração de métricas relevantes e a ineficiência na deteção e resolução de problemas em tempo útil. Estes desafios destacam a importância de desenvolver uma solução escalável e automatizada para melhorar a eficiência operacional e a segurança dos sistemas.

O setor de seguros, ao qual a Cleva pertence, encontra-se em fase de transformação significativa, impulsionada por mudanças tecnológicas, económicas e comportamentais. À medida que a digitalização se torna essencial, as seguradoras enfrentam a necessidade de modernizar os seus sistemas e adotar ferramentas avançadas para atender às crescentes demandas dos consumidores e às mudanças no mercado global. Especialistas, como os analistas da Gartner, destacam que a automação e o uso de ferramentas de análise são prioridades estratégicas para o setor, uma vez que ajudam a garantir eficiência operacional e a oferecer serviços mais personalizados e acessíveis (Gartner 2024a)(Gartner 2024c). Neste cenário, as soluções para a gestão de *logs* ganham ainda mais relevância, desempenhando um papel fundamental no suporte à evolução tecnológica e na resposta às crescentes necessidades de monitorização e análise de dados.

O tema ganha especial relevância, pois a Cleva enfrenta a pressão de reduzir tempos de inatividade, melhorar a resposta a incidentes e tomar decisões baseadas em dados confiáveis. Por meio deste projeto, pretende-se desenvolver uma solução que integra tecnologias e métodos modernos para recolha e análise de *logs*, permitindo uma gestão eficiente, centralizada e escalável. Este trabalho também está alinhado com os princípios de observabilidade, que visam proporcionar uma visão holística dos sistemas empresariais através da recolha, processamento e análise de dados (Gartner 2024b)(Gartner 2024c) .

Além disso, o projeto contribuirá para a área de monitorização automatizada, combinando análise de *logs* com a geração de alertas automáticos para facilitar a deteção de anomalias

em eventos críticos. Isso inclui a utilização de ferramentas que apoiam a extração de métricas em tempo real e a criação de *dashboards* que permitam uma visão consolidada das operações. A integração destas funcionalidades não só melhora os processos internos da organização, como também promove uma tomada de decisão mais ágil e baseada em dados confiáveis.

Através desta investigação, espera-se dar um contributo relevante para a implementação de sistemas de monitorização que correspondam exatamente às necessidades existentes da Cleva e promovam uma maior otimização dos processos, apoiando simultaneamente decisões mais bem informadas e eficazes. A Cleva com este projeto pretende desenvolver um sistema de monitorização baseado em dados precisos e acionáveis. Este esforço não só responde aos desafios individuais da Cleva, como também contribui para facilitar a mudança das melhores práticas de gestão de *logs* empresariais e, consequentemente, para a inovação no sector dos seguros.

1.1.1 Cleva

A história da Cleva Solutions está profundamente enraizada no setor segurador e reflete sua evolução ao longo de décadas para se tornar um dos principais fornecedores de soluções tecnológicas para o mercado de seguros (Cleva 2024).

Originalmente era i2S, tornando-se Inetum e agora a Cleva, uma empresa portuguesa fundada em 1984, com o objetivo de desenvolver soluções tecnológicas avançadas para atender às necessidades do setor segurador. A i2S construiu uma reputação sólida por sua especialização e inovação, tornando-se uma referência no desenvolvimento de plataformas integradas para a gestão de apólices, gestão de risco e serviços de apoio ao cliente (Solutions 2024).

Em 2019, antes de se transformar na Cleva Solutions, a i2S passou por uma mudança significativa ao ser adquirida pelo grupo Gfi, agora conhecido como Inetum, uma multinacional especializada em soluções tecnológicas e serviços digitais (ECO 2019). Esta aquisição marcou um ponto de viragem para a i2S, permitindo-lhe expandir as suas capacidades e integrar-se numa estrutura global, fortalecendo a sua posição no mercado de seguros. O movimento estratégico criou sinergias valiosas entre as competências locais da i2S e a experiência global da Inetum, estabelecendo uma base sólida para o futuro lançamento da marca Cleva Solutions (ECOSEGUROS 2021, 2022).

Mais recentemente, em 2024, a Cleva Solutions iniciou negociações exclusivas para ser adquirida pela AnaCap, um grupo líder em *private equity* especializado em tecnologia e serviços financeiros. Esta nova aquisição representa mais um marco na evolução da empresa, abrindo portas para acelerar os seus planos de crescimento, expandir a presença internacional e consolidar-se como referência no desenvolvimento de soluções tecnológicas para o setor de seguros (Cleva 2024).

1.2 Problema

A gestão de *logs* empresariais tornou-se um desafio crescente devido à rápida evolução tecnológica e à complexidade dos sistemas modernos. Com o aumento exponencial do volume de dados gerados por aplicações, sistemas operativos e infraestruturas, as abordagens tradicionais de gestão de *logs*, como o armazenamento em bases de dados relacionais com múltiplas tabelas, já não conseguem acompanhar as necessidades das organizações (Gartner

2024a). Estas abordagens apresentam limitações significativas, como a falta de escalabilidade, a dificuldade em centralizar dados provenientes de diferentes fontes e a ineficiência na deteção e análise de eventos críticos.

Além disso, a necessidade de realizar análises manuais para identificar problemas e extrair métricas relevantes sobrecarrega as equipas técnicas, tornando o processo suscetível a erros humanos e aumentando o tempo de resposta a incidentes. Este cenário é agravado pela crescente pressão das empresas para garantir a continuidade operacional, proteger informações sensíveis e responder rapidamente a ameaças e falhas no sistema (Gartner 2024b). A falta de uma solução centralizada e automatizada compromete não apenas a eficiência operacional, mas também a capacidade de tomar decisões informadas e estratégicas com base em dados.

Tal como referido em estudos realizados pela Gartner, as empresas, principalmente as do sector dos seguros, enfrentam um constrangimento adicional sobre a forma da coexistência de sistemas, legados e modernos, que na sua maioria não comunicam entre si da forma mais eficiente possível (Gartner 2024c). Esta escassez de integração conduz à consolidação de informações provenientes de diferentes fontes e à redução da redundância entre elas, o que influencia indiretamente a capacidade de análise dos dados. Além disso, a modernização do sistema informático de uma organização está muitas vezes associada a custos exorbitantes e à necessidade de competências especiais, o que constitui uma barreira à aplicação de soluções inovadoras (Gartner 2024b).

Outra área crítica é a forma como o aumento exponencial de dados tem vindo a afetar o desempenho das ferramentas de monitorização e análise. As soluções atuais não têm, em geral, capacidade a todos os níveis para processar e analisar grandes volumes de informação em tempo real, o que provoca atrasos na identificação de um problema e na tomada de decisões estratégicas. Estes problemas são agravados em organizações que operam em setores regulados, como o dos seguros, que exige o cumprimento de normas de segurança e privacidade dos dados (SAPO 2024).

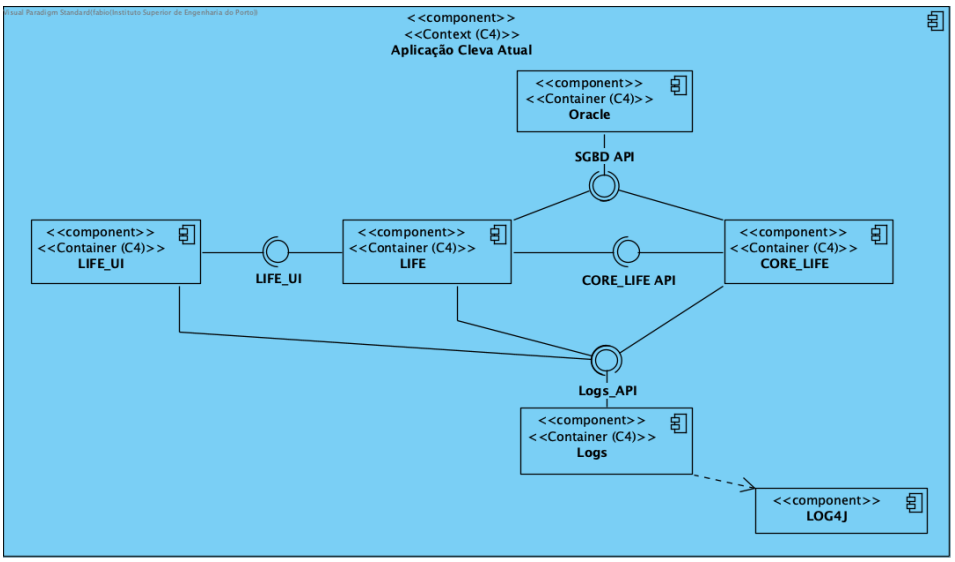


Figura 1.1: Solução atual

A solução atualmente implementada na Cleva, representada na Figura 1.1, reflete a arquitetura atual do sistema de gestão e análise de *logs* da organização. Esta arquitetura é composta por diversos módulos que desempenham papéis específicos dentro do ecossistema empresarial da Cleva, destacando-se os seguintes:

- **LIFE_UI**: Este módulo é responsável por conter toda a lógica relacionada aos ecrãs da interface do utilizador (*frontend*). Ele suporta a interação direta do utilizador com o sistema, apresentando informações e capturando ações que serão processadas pelos componentes de *backend*.
- **LIFE**: Este componente sustenta toda a lógica interativa que alimenta o `LIFE_UI`. Serve como a ponte entre a interface do utilizador e os dados necessários para suportar as operações em tempo real, garantindo uma experiência fluida e consistente.
- **CORE_LIFE**: Este módulo é dedicado à execução de processos não interativos e *batches*. Ele é projetado para realizar tarefas críticas em segundo plano, como processamento em massa e manutenção de sistemas, sem a necessidade de interação direta com o utilizador.
- **Base de Dados Oracle**: Este é o repositorio para armazenar *logs* controlados pelo negócio e também para os *logs* críticos do sistema. Armazena informações sensíveis, como dados financeiros e resultados de cálculos realizados pelo sistema, sendo essencial para a confiabilidade e integridade das operações.
- **Logs**: Este componente é responsável por armazenar os *logs* não controlados de toda a aplicação em ficheiros de texto (`.txt`). Ele possui uma dependência direta do `LOG4J`, que é utilizado para a geração e manutenção dos *logs*. Este repositório é crítico para a rastreabilidade de eventos não estruturados e para fornecer informações detalhadas sobre o funcionamento da aplicação.

Este problema tornou-se especialmente relevante na Cleva devido a vários fatores que afetam diretamente a eficiência operacional e a capacidade de resposta da organização:

- **Crescimento do Volume de Dados**: O aumento exponencial na geração de dados, resultado do crescimento de operações digitais, tornou as soluções atuais de gestão de *logs* insuficientes. A dependência do Log4j, que armazena os *logs* em ficheiros de texto, é particularmente problemática, pois não foi projetado para lidar com volumes massivos de dados de maneira eficiente. Esta limitação reduz significativamente a capacidade de consulta em tempo real e a análise detalhada dos registos.
- **Complexidade das Infraestruturas**: A arquitetura da Cleva inclui múltiplos sistemas interconectados (como LIFE_UI, CORE_LIFE e Oracle). Essa diversidade de sistemas aumenta a fragmentação dos *logs*, dificultando a sua centralização e análise. O Log4j, por não oferecer funcionalidades nativas para integração ou um *dashboard* centralizado, torna-se inadequado para consolidar informações de diferentes fontes e oferecer uma visão holística da infraestrutura.
- **Impacto no Desempenho da Aplicação**: A gestão ineficiente dos *logs*, como o armazenamento em ficheiros de texto, tem impacto direto no desempenho das aplicações da Cleva. Consultas complexas sobre os *logs* geram sobrecarga na infraestrutura, afetando a responsividade dos sistemas críticos. Além disso, a ausência de soluções escaláveis impede que a organização atenda às exigências de desempenho em momentos de alta carga operacional.

- **Necessidade de Segurança e Conformidade**: Com regulamentações rigorosas, como o Regulamento Geral de Proteção de Dados (RGPD), é essencial garantir a privacidade e a segurança dos dados geridos. O Log4j, embora eficiente em cenários simples, não dispõe de mecanismos robustos para garantir a conformidade com padrões de segurança, expondo a Cleva a riscos de violações de dados. Além disso, a rastreabilidade e o armazenamento seguro de *logs* críticos tornam-se mais difíceis em um sistema que não oferece soluções nativas para criptografia ou auditoria.

1.3 Objetivos

O objetivo deste projeto consiste em desenvolver uma solução integrada para a monitorização, análise e gestão de *logs* empresariais, capaz de centralizar dados provenientes de diferentes fontes e automatizar a deteção de problemas e geração de alertas. A solução será projetada para atender às necessidades específicas da Cleva, garantindo escalabilidade, eficiência operacional e suporte à tomada de decisões baseadas em métricas acionáveis.

A plataforma terá como foco a simplificação e agilidade no processo de gestão de *logs*, permitindo que as equipas técnicas identifiquem e resolvam problemas em tempo real, reduzindo o tempo de inatividade e otimizando processos empresariais. Além disso, será assegurada a segurança e privacidade dos dados monitorizados, alinhando-se com regulamentações e boas práticas do setor.

1.3.1 Objetivos Específicos

Para alcançar o objetivo do projeto, foram definidos objetivos específicos que orientam o desenvolvimento da solução e asseguram que todas as necessidades identificadas sejam abordadas.

Tabela 1.1: Objetivos Específicos

ID	Objetivo
OBJ1	Centralizar a gestão de logs.
OBJ2	Garantir a escalabilidade da solução.
OBJ3	Reduzir o tempo de deteção e resolução de problemas.
OBJ4	Facilitar a integração com outras ferramentas.
OBJ5	Garantir a segurança e privacidade dos dados.
OBJ6	Extrair métricas relevantes de negócio.
OBJ7	Acompanhar e otimizar o desempenho dos sistemas.

A Tabela 1.1 apresenta os objetivos específicos do projeto. Para garantir que esses objetivos sejam alcançados de maneira clara e mensurável, cada um deles é descrito no corpo do texto a seguir, utilizando o formato Específico, Mensurável, Atingível, Relevante, Temporal (SMART):

- **OBJ1 - Centralizar a gestão de logs:** Consolidar, até o final do projeto, todos os registos de sistemas numa única plataforma acessível, eliminando a fragmentação e facilitando a gestão centralizada.

- **OBJ2 - Garantir a escalabilidade da solução:** Desenvolver uma arquitetura que suporte um aumento de 50% no volume de logs processados, sem comprometer o desempenho.
- **OBJ3 - Reduzir o tempo de deteção e resolução de problemas:** Implementar mecanismos que permitam identificar e resolver incidentes críticos em até 2 horas, reduzindo o tempo médio atual de resposta que é de 5 horas.
- **OBJ4 - Facilitar a integração com outras ferramentas:** Disponibilizar *APIs* que suportem o código legado.
- **OBJ5 - Garantir a segurança e privacidade dos dados:** Incorporar medidas como encriptação de dados e controlo de acesso baseado em funções Controlo de Acesso Baseado em Funções (Role-Based Access Control) (RBAC) até o término do projeto, assegurando conformidade com o RGPD.
- **OBJ6 - Extrair métricas relevantes de negócio:** Desenvolver relatórios automatizados que apresentem pelo menos cinco métricas-chave de desempenho para apoiar a tomada de decisões estratégicas, disponíveis até a entrega final.
- **OBJ7 - Acompanhar e otimizar o desempenho dos sistemas:** Implementar um painel de monitorização que permita rastrear e analisar o desempenho dos sistemas em tempo real, com atualizações contínuas a cada 30 segundos.

Com esta abordagem, os objetivos são claramente definidos, permitindo uma avaliação contínua do progresso e assegurando que as metas estabelecidas sejam atingidas dentro dos prazos e parâmetros propostos.

1.4 Metodologia de Pesquisa

A metodologia de pesquisa adotada neste projeto basear-se-á no método Itens Preferidos para Relatórios de Revisões Sistemáticas e Meta-Análises (PRISMA), que é amplamente utilizado para conduzir revisões sistemáticas e meta-análises de forma rigorosa e transparente. Este método foi escolhido devido à sua capacidade de estruturar e documentar todas as etapas do processo de revisão, garantindo a reprodutibilidade e a qualidade dos resultados obtidos (Page et al. 2024).

O PRISMA é especialmente reconhecido por fornecer diretrizes claras para a elaboração e publicação de revisões sistemáticas (Rethlefsen et al. 2021). Estas diretrizes ajudam a assegurar que todas as etapas, desde a formulação da questão de pesquisa até a síntese dos resultados, sejam conduzidas com rigor e transparência.

Além disso, o método PRISMA possui extensões, como o PRISMA-P, que oferece diretrizes específicas para protocolos de revisão sistemática, garantindo que o planeamento da pesquisa seja documentado de forma padronizada (Moher et al. 2015).

A aplicação de visualização analítica no contexto do PRISMA também tem sido explorada para facilitar a análise de dados e melhorar a compreensão dos resultados, como descrito por Sina e Nazemi (Sina e Nazemi 2022). Essa abordagem destaca o potencial de integrar tecnologias visuais ao processo de revisão.

Por fim, o PRISMA também inclui ferramentas para revisões sistemáticas contínuas, como os diagramas de fluxo PRISMA, que auxiliam na documentação de revisões dinâmicas e atualizadas (Kahale et al. 2021).

O método PRISMA será aplicado com o objetivo de identificar, avaliar e sintetizar a literatura relevante relacionada à gestão e análise de *logs* empresariais, bem como às tecnologias e ferramentas disponíveis para atender a este propósito. Esta abordagem permite uma revisão sistemática das soluções existentes, destacando as suas vantagens, limitações e adequação às necessidades do projeto.

1.4.1 Questões de Investigação

As questões de investigação são fundamentais para orientar a pesquisa e delimitar o foco do estudo. No contexto deste projeto, as questões serão formuladas para abordar os desafios associados à gestão de *logs* empresariais, a análise de métricas e a implementação de soluções tecnológicas escaláveis (Page et al. 2024). Estas questões visam explorar as limitações das abordagens existentes, identificar melhores práticas e apoiar o desenvolvimento de uma solução robusta e eficiente.

Tabela 1.2: Questões de Pesquisa

ID	Questão de Pesquisa
Q1	Quais são os desafios atuais enfrentados pelas organizações na gestão e análise de *logs* empresariais?
Q2	Quais são as ferramentas e tecnologias mais eficazes para a monitorização e análise de *logs* em tempo real?
Q3	Como podem os *logs* empresariais ser processados para extrair métricas úteis e apoiar a tomada de decisões estratégicas?
Q4	De que forma a automação na deteção de problemas e na geração de alertas pode otimizar a gestão de *logs*?
Q5	Quais são os principais fatores a considerar ao implementar uma solução escalável e segura para a gestão de *logs*?
Q6	Como pode uma solução de monitorização ser integrada nos sistemas empresariais existentes, minimizando a interrupção das operações?

A Tabela 1.2 apresenta as principais questões de investigação estruturadas para este estudo. Essas questões serão elaboradas com o objetivo de responder aos desafios identificados e fornecer *insights* que guiem o desenvolvimento e a validação da solução.

1.4.2 Fontes de Informação

Uma etapa essencial na realização deste estudo é a identificação das fontes de informação que serão consultadas para análise e desenvolvimento da solução. A Tabela 1.3 apresenta as bases de dados selecionadas como principais fontes de informação para o estudo:

Tabela 1.3: Fontes de Informação

ID	Fonte de Informação
FD1	IEEE Xplore
FD2	Science Direct
FD3	B-on
FD4	ACM Digital Library

Essas fontes foram selecionadas devido à sua vasta coleção de artigos científicos, publicações técnicas e relatórios que abordam tópicos relevantes para o projeto. Estas garantirão acesso a informações atualizadas e confiáveis, permitindo uma base sólida para a revisão da literatura e o suporte teórico necessário para o desenvolvimento da solução.

1.4.3 Hipóteses

As hipóteses são suposições fundamentadas que orientam a investigação e permitem verificar a viabilidade das soluções propostas (Moher et al. 2015). No âmbito deste projeto, as hipóteses serão formuladas com base nos desafios identificados.

1.4.3.1 Hipótese Principal

A implementação de uma solução centralizada e automatizada para a gestão de *logs* empresariais reduzirá o tempo necessário para a deteção e resolução de problemas, melhorando a eficiência operacional e facilitando a tomada de decisões estratégicas com base em métricas relevantes.

1.4.3.2 Hipótese Secundária

Ao adotar uma solução automatizada de extração de métricas e geração de alertas configuráveis, a equipa poderá diminuir a necessidade de intervenção manual, assegurando a precisão dos alertas e a qualidade dos dados, o que facilitará a tomada de decisões informadas e ágeis.

1.4.4 Termos de Pesquisa

Os termos de pesquisa são fundamentais para orientar a revisão sistemática e garantir a identificação de estudos relevantes ao longo do processo de investigação (Rethlefsen et al. 2021). Esses termos serão selecionados com base nos principais conceitos e áreas relacionadas à gestão de *logs* empresariais, monitorização e análise de dados, garantindo um alinhamento preciso com os objetivos do projeto.

Tabela 1.4: Termos de Pesquisa

ID	Termos	Palavras-chave (Keywords)
TP1	Gestão de *Logs* Empresariais	"Log Management"
TP2	Monitorização de Dados em Tempo Real	"Real-Time Monitoring"AND "Data Analytics"
TP3	Extração de Métricas de Negócio	"Business Metrics"AND "Log Analysis"
TP4	Escalabilidade em Gestão de *Logs*	"Scalable Log Management"AND "Enterprise Systems"
TP5	Automação na Deteção de Problemas	"Automated Alerts"AND "Incident Detection"
TP6	Segurança e Privacidade dos Dados	"Data Security"AND "Privacy" AND "Log Management"
TP7	Integração de Ferramentas de Monitorização	"Monitoring Tools Integration" AND "Log Management Solutions"
TP8	Comparação de Ferramentas de *Logs*	"Elasticsearch"OR "Splunk" OR "Graylog"OR "Datadog"
TP9	Redução do Tempo de Resposta	"Response Time Reduction" AND "Event Management"
TP10	Análise de Logs Empresariais	"Log Analysis"AND "Operational Efficiency"

A Tabela 1.4 apresenta os termos de pesquisa definidos e suas respectivas palavras-chave. Estes termos abrangem tópicos essenciais, como gestão e monitorização de *logs*, segurança de dados, automação e análise de métricas empresariais. A definição cuidadosa desses termos é essencial para garantir que a revisão bibliográfica seja abrangente.

1.4.5 Critérios de Inclusão e Exclusão

Os critérios de inclusão e exclusão são essenciais para garantir a relevância e a qualidade dos estudos e materiais selecionados na revisão sistemática (Page et al. 2024). Estes critérios são definidos com base nos objetivos do projeto, visam assegurar que apenas estudos e publicações alinhados ao tema da monitorização,análise e gestão de logs empresariais sejam considerados.

1.4.5.1 Critérios de Inclusão

Para garantir a relevância e qualidade dos estudos utilizados na revisão sistemática,são definidos critérios de inclusão baseados nos objetivos e no âmbito do projeto.

Tabela 1.5: Critérios de Inclusão

ID	Critério
CI1	Estudos ou publicações que abordem ferramentas, metodologias ou tecnologias para monitorização, análise e gestão de *logs* empresariais.
CI2	Artigos provenientes de revistas científicas indexadas, conferências reconhecidas e fontes confiáveis.
CI3	Trabalhos que apresentem casos de uso, aplicações práticas ou exemplos reais de gestão de logs e monitorização de operações.

A Tabela 1.5 apresenta os critérios de inclusão utilizados, os quais orientaram a seleção de materiais relevantes e confiáveis para embasar o desenvolvimento da solução proposta.

1.4.5.2 Critérios de Exclusão

Os critérios de exclusão são estabelecidos para eliminar estudos e publicações que não atendem aos padrões de qualidade e relevância necessários para a revisão sistemática. Esses critérios ajudam a evitar que materiais desatualizados ou provenientes de fontes não confiáveis sejam considerados no desenvolvimento do projeto.

Tabela 1.6: Critérios de Exclusão

ID	Critério
CE1	Publicações com mais de 10 anos, a menos que sejam trabalhos fundamentais ou que tenham relevância histórica para o tema.
CE2	Artigos de blogs, fóruns, páginas não acadêmicas ou fontes sem revisão por pares.

A Tabela 1.6 apresenta os critérios de exclusão definidos, garantindo que o foco permaneça em conteúdos alinhados com os objetivos e requisitos do estudo.

1.4.6 Extração de Publicações

O protocolo sistemático para a seleção, triagem e avaliação dos estudos seguirá a metodologia PRISMA, amplamente reconhecida por garantir rigor e transparência em revisões sistemáticas (Sina e Nazemi 2022). O diagrama de fluxo, apresentado na Figura 1.2, oferece uma representação visual detalhada das etapas realizadas, desde a procura inicial até à inclusão final dos estudos analisados. Este diagrama demonstra o percurso metodológico adotado, reflete assim, o compromisso com a qualidade científica e a rastreabilidade do processo (PRISMA 2020).

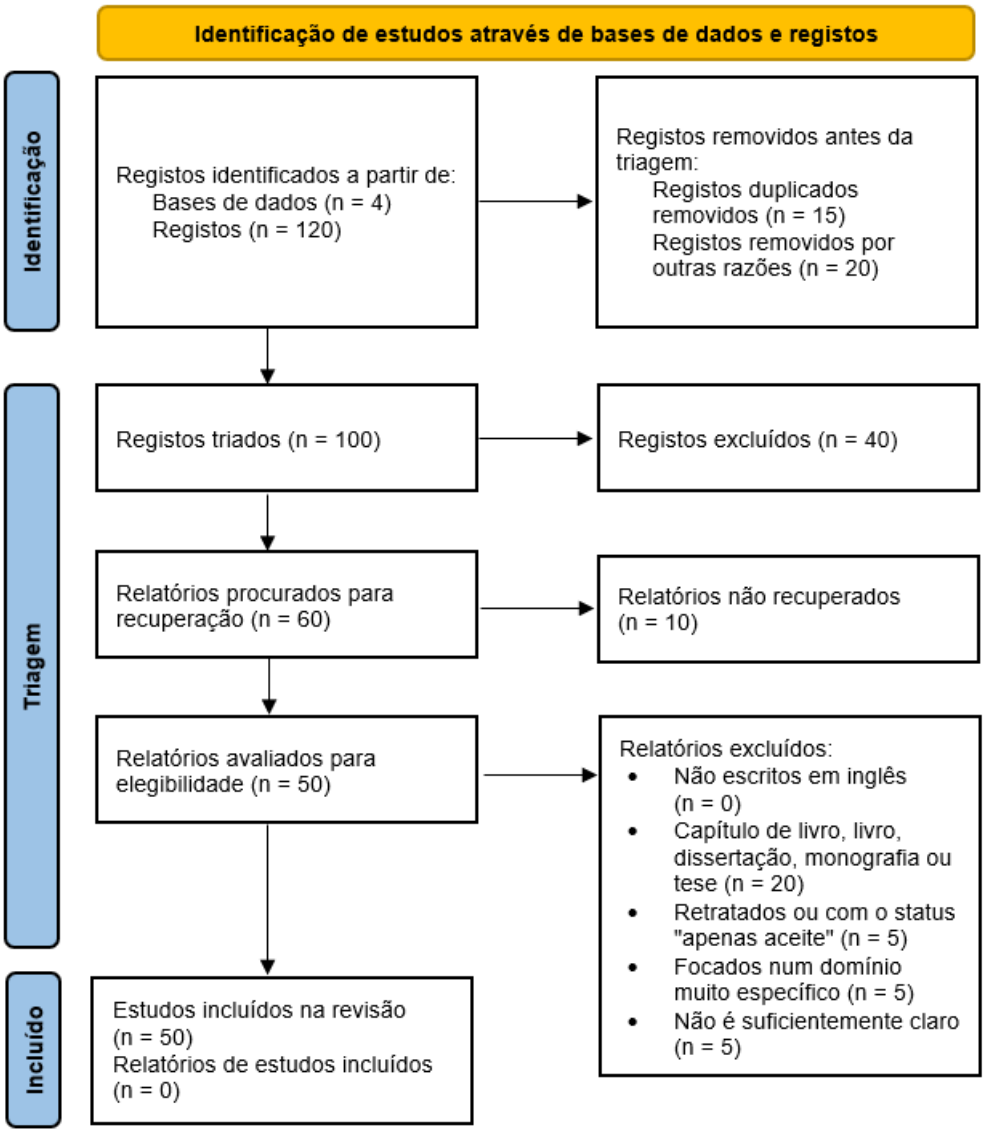


Figura 1.2: Diagrama de fluxo PRISMA adptado (PRISMA 2020)

A aplicação rigorosa dos requisitos de inclusão e exclusão, descritos anteriormente, garantiu a seleção de fontes relevantes e de elevada qualidade para o estudo. O diagrama de fluxo PRISMA adaptado, apresentado na Figura 1.2, proporciona uma visão clara e organizada da metodologia seguida, assegurando a transparência e reforçando a credibilidade do processo de investigação realizado.

A fase de Identificação inicia-se com a pesquisa em quatro bases de dados, resultando na obtenção de 120 registos. Antes da triagem, foram removidos 15 registos duplicados e 20 registos por outros motivos, totalizando 35 exclusões iniciais. Assim, 100 registos passaram para a etapa seguinte.

Na etapa de Triagem, os 100 registos foram analisados, levando à exclusão de 40 documentos por não cumprirem critérios pré-estabelecidos. Dos registos restantes, 60 foram procurados para recuperação, no entanto, 10 não foram acessíveis. Assim, 50 relatórios passaram para a avaliação de elegibilidade.

Durante a fase de Elegibilidade, os 50 relatórios recuperados foram submetidos a uma análise mais criteriosa. Entre os critérios de exclusão estavam documentos que não estavam escritos em inglês (n = 0), capítulos de livro, dissertações, monografias ou teses (n = 20), artigos retratados ou com status de "apenas aceite"(n = 5), estudos muito específicos para domínios irrelevantes ao âmbito da revisão (n = 5) e publicações com falta de clareza suficiente para avaliação (n = 5). Após essa triagem rigorosa, restaram 50 estudos incluídos na revisão sistemática, enquanto nenhum relatório de estudo foi incluído separadamente.

1.5 Metodologia de Trabalho - Design and Creation

Este projeto adota a abordagem *Design and Creation*, reconhecida para o desenvolvimento de sistemas tecnológicos inovadores (S. Kosan 2014). Esta metodologia é centrada na criação de artefactos tecnológicos baseados em princípios e boas práticas de design, com o objetivo de resolver problemas específicos e atender às necessidades organizacionais (Peffers et al. 2007).

A abordagem *Design and Creation* permite focar tanto na conceção do sistema quanto na validação de sua eficácia, garantindo que a solução proposta seja capaz de alcançar os objetivos do projeto. Esta metodologia é composta por etapas estruturadas, que incluem a conceção, construção, avaliação e refinamento do artefacto desenvolvido, assegurando um processo iterativo e orientado por resultados (Peffers et al. 2007).

Para o desenvolvimento deste projeto, o método será aplicado em várias fases, desde a identificação do problema até à validação e comunicação da solução. Estas etapas garantirão que o processo seja conduzido de forma eficiente e que a solução desenvolvida esteja alinhada aos objetivos estratégicos e técnicos da organização (S. Kosan 2014).

1.5.1 Identificação do Problema

Nesta fase inicial, será realizada a análise e compreensão dos desafios associados à gestão de logs empresariais. Serão investigados problemas como a dificuldade de centralizar dados, a ineficiência na extração de métricas relevantes e a falta de automação na deteção de problemas (Saltuk e I. Kosan 2014). Este diagnóstico permitirá a definição clara dos requisitos e objetivos da solução a ser desenvolvida.

1.5.2 Conceção do Artefacto

Nesta fase, será realizada a concepção inicial da solução, incluindo o design da arquitetura do sistema e a especificação dos seus principais componentes e funcionalidades. A conceção incluirá:

- Desenvolvimento de *pipelines* para recolha, processamento e análise de *logs*.
- Geração de alertas automáticos em tempo real.
- Criação de *dashboards* para visualização de métricas empresariais.

Será também considerada a integração com sistemas existentes, bem como os requisitos de segurança e escalabilidade.

1.5.3 Construção do Artefacto

A fase de construção do artefacto será conduzida de forma iterativa e incremental, adotando a metodologia ágil Scrum, que já é amplamente utilizada pela organização como parte do seu processo padrão de desenvolvimento. Esta abordagem é ideal para garantir a flexibilidade necessária e a entrega contínua de valor aos *stakeholders* durante o desenvolvimento da solução (Educação 2024).

Além disso, o controlo de versões e o armazenamento do código serão geridos através do Gerrit, uma ferramenta de repositório que garante rastreabilidade e colaboração eficaz. *Commits* serão realizados regularmente, acompanhados de mensagens detalhadas que descrevem as alterações efetuadas, como a implementação de novas funcionalidades ou a correção de *bugs*.

1.5.4 Avaliação e Validação

Após a construção do artefacto nas múltiplas iterações, a solução final será submetida a testes rigorosos, incluindo:

- **Testes funcionais**: De forma a verificar se as funcionalidades implementadas atendem aos requisitos definidos.
- **Testes de desempenho**: De forma a avaliar a escalabilidade e a eficiência da solução sob diferentes cargas de dados.
- **Testes de segurança**: De forma a assegurar que os dados tratados estejam protegidos e em conformidade com o RGPD.

Os resultados obtidos serão utilizados para realizar os ajustes necessários e otimizar a solução (Saltuk e I. Kosan 2014).

1.5.5 Comunicação e Documentação

Ao longo do projeto, a solução será documentada de forma detalhada, incluindo a descrição da arquitetura, as funcionalidades implementadas e os resultados dos testes realizados (Saltuk e I. Kosan 2014). A documentação será preparada para garantir que o trabalho seja auditável, replicável e útil para futuras melhorias (Saltuk e I. Kosan 2014).

Estas fases garantirão que o desenvolvimento da solução seja conduzido de maneira estruturada e alinhada com os objetivos iniciais (Saltuk e I. Kosan 2014).

1.6 Planeamento

Para uma visão detalhada das atividades, datas de início e fim de cada etapa, o cronograma completo do projeto é apresentado na Tabela 1.7. Este cronograma estrutura de forma clara as diferentes fases e garante que todas as atividades sejam realizadas dentro do prazo estipulado, proporcionando uma gestão eficiente do tempo e dos recursos disponíveis. Inicialmente, foi desenvolvido o Plano da Estrutura do Projeto (WBS), apresentado no Apêndice B, que detalhou as atividades necessárias para a realização do projeto (Rianty, Latief e Riantini 2018). A partir do WBS, organizaram-se e inseriram-se as tarefas no *Microsoft Project*,

permitindo a definição do cronograma, a atribuição de recursos e o monitoramento do estado das atividades. Por fim, o cronograma será consolidado na Tabela 1.7, garantindo que o projeto fosse conduzido de forma estruturada e eficiente.

Tabela 1.7: Cronograma do Projeto

Fase	Atividade	Data Início	Data Fim
1	**Projeto**	16/09/2024	22/07/2025
1.1	**Planeamento**	16/09/2024	05/11/2024
1.1.1	Formalização da Proposta	16/09/2024	15/10/2024
1.1.2	Descrição dos Objetivos Específicos	16/10/2024	05/11/2024
1.2	**Pesquisa e Revisão da Literatura e Análise do Sistema Atual**	06/11/2024	22/07/2025
1.2.1	Revisão das Tecnologias e Soluções Existentes	06/11/2024	17/12/2024
1.2.2	Avaliação das Necessidades do Projeto	06/11/2024	03/01/2025
1.2.3	Entrega PREPD	04/01/2025	04/01/2025
1.3	**Design e Arquitetura do Sistema**	11/02/2025	103/03/2025
1.3.1	Definição da Arquitetura da Solução	11/02/2025	03/03/2025
1.3.2	Análise dos Requisitos	11/02/2025	19/02/2025
1.4	**Desenvolvimento**	04/03/2025	23/05/2025
1.4.1	Configuração da Solução de Monitorização e Análise de Logs	04/03/2025	24/03/2025
1.4.2	Implementação da Solução Desenhada	25/03/2025	23/04/2025
1.4.3	Desenvolvimento de Interface e Visualização	24/04/2025	23/05/2025
1.5	**Testes e Validação**	25/03/2025	03/06/2025
1.5.1	Testes Funcionais	25/03/2025	03/06/2025
1.5.2	Testes de Desempenho e Escalabilidade	26/05/2025	03/06/2025
1.5.3	Testes de Segurança	26/05/2025	03/06/2025
1.6	**Documentação e Entrega Final**	11/02/2025	22/07/2025
1.6.1	Documentação Técnica	11/02/2025	03/03/2025
1.6.2	Relatório Final da Dissertação	04/03/2025	21/07/2025
1.6.3	Apresentação e Demonstração	22/07/2025	22/07/2025
1.6.4	Entrega DIMEI	22/07/2025	22/07/2025

A primeira fase do projeto, denominada Planeamento, abrange as atividades iniciais focadas na estruturação e organização do trabalho. Durante essa etapa, será formalizada a proposta do projeto, com a definição clara dos objetivos gerais e específicos que orientarão as próximas fases. Adicionalmente, será realizada a descrição detalhada dos objetivos específicos e a identificação das expectativas relacionadas à solução a ser desenvolvida.

Na sequência, será realizada a Pesquisa e Revisão da Literatura, cujo principal objetivo é investigar tecnologias, metodologias e soluções existentes no contexto da monitorização e análise de *logs* empresariais. Essa etapa visa oferecer uma base teórica sólida para o projeto, além de identificar as lacunas e desafios técnicos enfrentados atualmente. Também serão

avaliadas as necessidades específicas do projeto, garantindo que os objetivos estabelecidos estejam alinhados com as capacidades técnicas disponíveis.

A fase de Design e Arquitetura do Sistema será voltada para a conceção da solução proposta. Nessa etapa, será elaborada a arquitetura do sistema, com atenção especial à integração com os sistemas existentes na organização. Além disso, serão analisados os requisitos funcionais e não funcionais, assegurando que a solução seja eficiente, escalável e atenda às necessidades identificadas.

Na fase de Desenvolvimento, será realizada a implementação prática da solução. O processo inclui a configuração inicial da solução para monitorização e análise de *logs*, a implementação das funcionalidades principais e o desenvolvimento de uma interface de visualização. Essas atividades garantirão que o sistema seja capaz de processar, analisar e apresentar dados de forma clara e eficiente.

Durante a fase de Testes e Validação, serão realizados diferentes tipos de testes para verificar o desempenho e a eficácia da solução desenvolvida. Entre os testes previstos estão os funcionais, para garantir que todas as funcionalidades operem corretamente, os de desempenho e escalabilidade, para avaliar como o sistema se comporta em diferentes cenários de carga, e os de segurança, para proteger os dados geridos pela solução.

Finalmente, a fase de Documentação e Entrega Final representará o encerramento do projeto. Nessa etapa, será criada a documentação técnica que detalha o funcionamento e a implementação da solução. Além disso, serão realizadas apresentações e demonstrações da solução desenvolvida para os *stakeholders*, destacando os resultados obtidos. Será elaborado o relatório final da dissertação, consolidando todo o trabalho desenvolvido e as contribuições alcançadas.

Essas etapas foram organizadas de forma a garantir que cada fase do projeto seja executada de maneira estruturada e eficiente, com foco na entrega de uma solução eficaz e alinhada aos objetivos definidos.

1.6.1 Riscos

Durante o planeamento do projeto, foi realizada uma análise detalhada para identificar os principais riscos que poderiam impactar negativamente o seu desenvolvimento e entrega. Esta análise teve como objetivo prever potenciais problemas, avaliando a probabilidade de ocorrência e o impacto associado, de forma a adotar estratégias eficazes para mitigar ou responder a cada risco identificado.

A Tabela 1.8 apresenta os principais riscos do projeto, juntamente com suas probabilidades, impactos e estratégias de resposta propostas. Este processo de identificação e gestão de riscos é fundamental para assegurar o cumprimento dos objetivos estabelecidos e a entrega de uma solução robusta e confiável. O detalhe desta análise está no Apêndice C.

Tabela 1.8: Identificação e Planeamento de Riscos

ID	Risco	Probabilidade	Impacto	Estratégia de Resposta
R01	Falta de disponibilidade dos *stakeholders*	Média	Alta	Identificar múltiplos *stakeholders* e garantir suporte ativo ao tema.
R02	Complexidade do projeto	Alta	Muito Alta	Reduzir o âmbito e priorizar funcionalidades essenciais.
R03	Falta de adaptação do Log4J	Média	Alta	Complementar a ferramenta com soluções adicionais para centralizar e otimizar a gestão de *logs*.
R04	Integração com sistemas legados	Média	Muito Alta	Desenvolver adaptadores e realizar testes contínuos para garantir a compatibilidade.
R05	Escalabilidade do sistema com crescimento de dados	Alta	Muito Alta	Realizar atualizações na infraestrutura para suportar a escalabilidade do sistema.

Um dos principais riscos identificados é R01, que trata da possibilidade de Falta de disponibilidade dos *stakeholders*. A ausência de suporte ativo ou de alinhamento entre os intervenientes principais pode causar atrasos na tomada de decisões e comprometer o andamento do projeto. Para prevenir esse problema, serão definidos múltiplos *stakeholders*, promovendo o envolvimento ativo e garantindo comunicação clara e constante.

O R02 destaca o impacto que a complexidade do projeto pode causar, como dificuldades no cumprimento dos prazos ou aumento de custos. Um âmbito excessivamente amplo ou funcionalidades conflitantes podem gerar atrasos e sobrecarga na equipa. A estratégia para mitigar este risco inclui uma análise do âmbito, priorizando as funcionalidades mais relevantes e ajustando os objetivos conforme necessário.

Outro desafio identificado é o R03, que aborda as limitações do Log4J no contexto do projeto. Por não oferecer uma centralização eficiente e visualização integrada dos *logs*, a ferramenta pode dificultar a monitorização. Para resolver essa limitação, a integração de ferramentas complementares será explorada, permitindo uma gestão mais eficaz e alinhada às necessidades específicas.

No caso do R04, o risco está associado à integração com sistemas legados, que frequentemente apresentam problemas de compatibilidade ou mudanças inesperadas. Tais fatores podem atrasar ou comprometer o progresso do projeto. Como solução, adaptadores específicos serão desenvolvidos e testes contínuos de compatibilidade serão realizados para assegurar a integração eficiente.

Por fim, o R05 destaca o desafio de lidar com o crescimento contínuo do volume de dados, o que exige escalabilidade do sistema. A incapacidade de suportar essas demandas pode impactar negativamente o desempenho e a eficiência das operações. Para mitigar esse risco, melhorias na infraestrutura serão implementadas, garantindo que o sistema possa acompanhar as necessidades futuras.

1.7 Gestão de Competências

A gestão de competências é um elemento essencial para o sucesso no desenvolvimento da dissertação e do desenvolvimento pessoal. Através do diagnóstico de competências, é possível identificar as áreas de força e os pontos que necessitam de desenvolvimento,

permitindo a implementação de um plano de ação estruturado para alcançar os objetivos estabelecidos.

Com base no diagnóstico realizado, destacaram-se as competências mais desenvolvidas, como resiliência, aprendizagem contínua, negociação, responsabilidade e resolução de problemas. Estas competências fornecem uma base sólida para lidar com os desafios do projeto e garantir um progresso consistente.

Por outro lado, foram identificadas competências a desenvolver, como gestão do tempo, comunicação, escuta ativa, criatividade e inovação, e motivação para a excelência. Estas áreas foram consideradas para melhorar a eficácia geral no cumprimento das metas do projeto.

1.7.1 Plano de Acção

Este plano de ação, apresentado na Tabela 1.9, serve como um guia para alcançar uma melhoria contínua, focando nas competências necessárias para enfrentar os desafios do projeto.

Tabela 1.9: Plano de Ação para Gestão de Competências

ID	Competência	Ação
A01	Gestão do Tempo	Utilizar ferramentas de organização, como agendas digitais, para priorizar tarefas e equilibrar responsabilidades profissionais e pessoais.
A02	Comunicação	Participar em *workshops* ou eventos que exijam apresentações em público, com o objetivo de aumentar a confiança e melhorar a clareza nas comunicações.
A03	Escuta Ativa	Desenvolver o hábito de tomar notas durante reuniões importantes para assegurar uma compreensão completa do que está sendo transmitido.
A04	Criatividade e Inovação	Participar de eventos tecnológicos e *hackathons*, que oferecem oportunidades de interação com ideias inovadoras.
A05	Motivação para a Excelência	Estabelecer metas de curto prazo alcançáveis para gerar um senso de progresso e realização.

A seguir, cada competência listada na Tabela 1.9 é detalhada utilizando uma abordagem baseada no formato SMART, garantindo que as ações propostas sejam específicas, mensuráveis, alcançáveis, relevantes e limitadas no tempo. Esta abordagem assegura a clareza e o acompanhamento contínuo do progresso em direção ao desenvolvimento das competências necessárias.

- **A01 - Gestão do Tempo:** Criar e manter, até ao final do projeto, um sistema de gestão de tarefas utilizando a ferramenta Trello, de forma a organizar todas as responsabilidades pessoais e profissionais, priorizando tarefas com base em prazos e impacto.

- **A02 - Comunicação:** Realizar, pelo menos, dois *workshops* ou eventos na empresa Contas Acertadas, abordando os temas de cibersegurança e informática nas organizações. O objetivo é aprimorar as habilidades de comunicação em público, aumentando a confiança e a clareza na transmissão de ideias.
- **A03 - Escuta Ativa:** Estabelecer o hábito de tomar notas detalhadas em 100% das reuniões durante todo o projeto, criando um resumo das discussões e definindo ações a partir delas.
- **A04 - Criatividade e Inovação:** Participar no Porto Tech Hub durante o período de desenvolvimento do projeto, visando estimular o pensamento inovador e a resolução criativa de problemas.
- **A05 - Motivação para a Excelência:** Concluir, pelo menos, uma tarefa crítica por semana, garantindo progresso contínuo no projeto. Adicionalmente, dedicar tempo para *hobbies* pessoais, como jogar padel, jogar computador e viajar, pelo menos uma vez por semana, para manter o equilíbrio e a motivação ao longo do trabalho.

Com a implementação destas estratégias, espera-se aumentar significativamente a eficiência e o impacto do trabalho, alinhando as competências pessoais às exigências do projeto e da dissertação.

1.8 Considerações Éticas

O desenvolvimento de uma solução tecnológica para a gestão de *logs* empresariais envolve várias considerações éticas fundamentais para assegurar que o projeto é conduzido de forma responsável e em conformidade com padrões éticos e legais. Em todas as fases do projeto, a integridade desempenhou um papel essencial, com todas as decisões documentadas de forma transparente, garantindo que os métodos e resultados fossem verificáveis e replicáveis. Este compromisso promoveu a confiança entre os *stakeholders* e reforçou a credibilidade do trabalho desenvolvido.

A proteção de dados sensíveis foi uma prioridade central no projeto, sendo a solução desenhada para cumprir regulamentações como o RGPD. Este compromisso garantiu que a recolha, armazenamento e processamento dos *logs* empresariais fossem realizados com elevados padrões de segurança e respeitando a privacidade dos utilizadores. Foram implementadas práticas rigorosas para assegurar a conformidade com o RGPD, protegendo informações sensíveis e evitando qualquer risco de violação de dados.

O projeto foi conduzido em alinhamento com princípios éticos amplamente reconhecidos, como os definidos pelo Código de Ética (ACM) e pelo *Software Engineering Code of Ethics* (Association for Computing Machinery 2024). Estes princípios enfatizam a importância da proteção do interesse público e da segurança de todos os *stakeholders* envolvidos, bem como a honestidade e a responsabilidade no desenvolvimento de *software*. O projeto também foi realizado em conformidade com o Código de Boas Práticas e Conduta do Instituto Politécnico do Porto (IPP), que exige elevados padrões éticos em atividades de investigação e desenvolvimento.

Estas considerações asseguram que o projeto será conduzido de maneira responsável, respeitando os direitos dos utilizadores e promovendo práticas de engenharia e desenvolvimento de *software* alinhadas aos mais elevados padrões éticos. A aplicação dessas práticas reforçou a credibilidade e o impacto positivo da solução no contexto empresarial.

1.9 Estrutura do Documento

Esta dissertação está estruturada de forma a fornecer uma explicação sobre o desenvolvimento da solução proposta para a gestão de *logs* empresariais. A organização do conteúdo foi planeada de maneira a guiar o leitor, apresentando de forma progressiva o contexto do estudo, as metodologias utilizadas e os resultados obtidos. A dissertação está organizada da seguinte maneira:

O primeiro capítulo, intitulado **Introdução**, oferece uma visão geral do projeto, abordando o problema da gestão de *logs* empresariais e a importância da solução a ser desenvolvida. São detalhados os objetivos do projeto, tanto os gerais como os específicos, assim como os principais desafios enfrentados pela organização. Além disso, a introdução esclarece a relevância do trabalho no contexto da Cleva, fornecendo uma base para que o leitor entenda as motivações do projeto, a necessidade de inovação e os benefícios esperados com a implementação da solução.

O segundo capítulo, denominado **Estado da Arte**, é dedicado à análise da literatura existente e à descrição das soluções atualmente utilizadas na gestão e monitorização de *logs* empresariais. Neste capítulo são exploradas as tecnologias e metodologias existentes, bem como as tendências emergentes e as ferramentas mais relevantes do mercado. São ainda abordadas as limitações das abordagens tradicionais, fornecendo uma base teórica para o desenvolvimento do projeto.

O terceiro capítulo, **Análise e Design da Solução**, apresenta a engenharia de requisitos, a identificação dos atores e os requisitos funcionais e não funcionais. São ainda descritas e comparadas diferentes arquiteturas candidatas, justificando-se a seleção adotada. Este capítulo aborda também o design detalhado da solução, incluindo fluxo de dados, estrutura de índices, *dashboards*, alertas e modelo de segurança.

O quarto capítulo, **Implementação**, descreve o ambiente técnico e detalha a configuração da ELK Stack, englobando o Elasticsearch, Logstash e Kibana. São explicados os processos de integração com as fontes de dados, a configuração do Metricbeat e o desenvolvimento dos *dashboards* e visualizações.

O quinto capítulo, **Avaliação e Validação**, apresenta a metodologia de avaliação aplicada, a comparação entre a situação inicial e a atual, bem como a validação dos objetivos. São ainda discutidas as principais limitações da solução e propostas melhorias futuras.

O sexto capítulo, **Conclusão**, sintetiza os principais resultados alcançados, discutindo o grau de concretização dos objetivos definidos, as contribuições do trabalho para a organização e sugestões para trabalhos futuros.

Por fim, são apresentadas a **Bibliografia**, reunindo as fontes consultadas, e os **Apêndices**, que incluem o *Project Charter*, a WBS (Work Breakdown Structure), a análise de riscos e o inquérito de avaliação da solução junto dos utilizadores.

2. Estado da Arte

Este capítulo tem como objetivo apresentar uma revisão das soluções existentes e das abordagens adotadas para a gestão e monitorização de *logs* empresariais. Serão discutidas as principais tecnologias, ferramentas e metodologias que são atualmente utilizadas para lidar com grandes volumes de dados gerados por sistemas e aplicações.

A revisão da literatura visa contextualizar teoricamente o projeto, fornecendo uma base sólida para entender as soluções existentes, as suas lacunas e justificar a necessidade de uma abordagem inovadora.

2.1 Revisão Literatura

Esta secção apresenta os conceitos fundamentais que sustentam o desenvolvimento deste projeto, para compreender as suas diferentes dimensões. Serão explorados temas como a gestão de *logs* empresariais, a automação na monitorização e a extração de métricas de negócio, que formam a base para soluções tecnológicas eficazes e escaláveis. A análise destes conceitos ajuda a esclarecer as necessidades e desafios associados ao processamento de grandes volumes de dados em ambientes corporativos dinâmicos.

Para além das ferramentas e tecnologias adotadas, esta abordagem também abrange os princípios que orientam a conceção e o design do sistema proposto. Ao abordar estas temáticas, pretende-se oferecer uma visão ampla e fundamentada, essencial para compreender as motivações e as escolhas técnicas que definem a arquitetura e a implementação da solução.

2.1.1 Sistemas de Gestão de Logs

Os sistemas de gestão de *logs* são ferramentas desenvolvidas para monitorizar, armazenar, processar e analisar dados gerados por sistemas, aplicações, dispositivos e redes. Os *logs* representam registos contínuos de eventos e operações, contendo informações críticas para rastrear atividades, diagnosticar problemas e avaliar o desempenho dos sistemas (Forense.io 2024). A relevância desses sistemas está na sua capacidade de centralizar e organizar grandes volumes de dados dispersos provenientes de diversas fontes. Conforme ilustrado na Figura 2.1, essa centralização facilita a gestão e análise dos dados, permitindo às equipas de TI e operações identificar rapidamente problemas, detetar comportamentos anómalos e tomar acções corretivas de forma eficaz (S. He, Zhang et al. 2022).

Figura 2.1: Sistema de *logs* (Azevedo 2020)

Os principais objetivos dos sistemas de gestão de *logs* incluem a centralização dos registos, recolhendo dados de múltiplas fontes para eliminar a fragmentação e simplificar o acesso e a análise. Os sistemas também procuram melhorar a eficiência operacional ao facilitar a gestão e a monitorização contínua dos sistemas, reduzindo o esforço humano necessário para averiguar eventos e resolver problemas (Elastic.co 2024). Outro objetivo fundamental é a redução do tempo de resposta a incidentes, pois a capacidade de gerar alertas em tempo real permite ações rápidas diante de falhas ou ameaças, minimizando impactos nas operações. Além disso, os sistemas possibilitam a extração de *insights* de negócio, analisando dados para identificar tendências, padrões de uso e métricas de desempenho que ajudam na tomada de decisões estratégicas (Sun et al. 2023).

Entre os benefícios associados a esses sistemas, destacam-se a deteção proativa de problemas, que permite identificar falhas antes que se tornem críticas, e o suporte à tomada de decisões, fornecendo informações detalhadas e precisas para decisões baseadas em dados (Tecnologia 2024). Também se destaca o aumento da segurança e conformidade, uma vez que esses sistemas auxiliam no cumprimento de regulamentações como o RGPD, garantindo a integridade e a privacidade dos dados armazenados (S. He, Zhang et al. 2022).

Apesar dos seus inúmeros benefícios, os sistemas de gestão de *logs* enfrentam desafios significativos. Um dos principais é o grande volume de dados gerados continuamente por sistemas e aplicações, o que pode dificultar a gestão e o processamento. Outro desafio é a necessidade de escalabilidade, já que os sistemas devem ser capazes de lidar com o aumento contínuo de dados sem comprometer o desempenho (Blogs 2024). Além disso, a privacidade e a segurança são questões críticas, principalmente porque os *logs* podem conter informações sensíveis que devem estar protegidas e em conformidade com regulamentações, como o RGPD. Por fim, garantir a integração com sistemas existentes pode ser uma tarefa técnica desafiadora, pois é necessário assegurar que os novos sistemas de gestão de *logs* sejam compatíveis com a infraestrutura já implementada (Sun et al. 2023).

2.1.2 Sistemas de Automação na Monitorização de Logs

Os sistemas de automação na monitorização de *logs* são ferramentas essenciais que permitem às organizações acompanhar, analisar e responder de forma eficiente aos eventos gerados por sistemas e aplicações. Ao automatizar processos como a coleta, filtragem,

análise e geração de alertas, estes sistemas reduzem a necessidade de intervenção manual, aumentando a precisão e a rapidez na identificação de incidentes (Elastic.co 2024).

Com o crescimento exponencial dos dados produzidos por sistemas modernos, a automação é uma aposta para lidar com a complexidade e o volume de *logs* (Blogs 2024). Ferramentas automatizadas processam dados em tempo real, identificando padrões e anomalias que podem indicar falhas, ameaças à segurança ou problemas operacionais. Essa capacidade não só melhora a eficiência operacional, mas também minimiza o tempo de resposta a problemas, reduzindo o impacto de incidentes no ambiente empresarial (Korzeniowski e Goczyła 2022a).

Os principais componentes desses sistemas incluem mecanismos de coleta e ingestão de *logs* provenientes de diversas fontes, como servidores, dispositivos de rede, aplicações e bases de dados (AprendeIT 2024). Após a coleta, os dados são processados por *pipelines* configuráveis que aplicam filtros, normalizam os registos e extraem informações relevantes. Ferramentas de análise preditiva são frequentemente integradas para identificar tendências e prever possíveis falhas antes que ocorram (S. He, P. He et al. 2021).

Entre os benefícios proporcionados pela automação estão a deteção pro-ativa de incidentes, a redução de custos operacionais, a melhoria na segurança e a garantia de conformidade com regulamentações (Works 2024).

Apesar dos avanços, a automação na monitorização de *logs* enfrenta desafios como a configuração inicial complexa, o custo de implementação e a necessidade de especialização técnica para operar e ajustar os sistemas (Guru99 2024). Contudo, à medida que as organizações reconhecem os benefícios dessas soluções, a automação torna-se um padrão indispensável para garantir a resiliência e a eficiência em ambientes corporativos dinâmicos (Korzeniowski e Goczyła 2022a).

2.1.3 Sistemas de Extração de Métricas de Logs

Os sistemas de extração de métricas de *logs* são ferramentas essenciais que permitem transformar dados não estruturados, provenientes de registos de sistemas, aplicações e dispositivos, em informações acionáveis para a tomada de decisões estratégicas e operacionais (Anderson e Smith 2023). Esses sistemas identificam, processam e apresentam métricas específicas, como tempos de resposta, taxas de erro, desempenho de sistemas e comportamentos dos utilizadores, fornecendo *insights* valiosos sobre o funcionamento e a eficiência das operações empresariais (Shang, Nagappan e Hassan 2015).

A implementação de sistemas de extração de métricas de *logs* possibilita às organizações monitorizar o desempenho de suas operações e identificar tendências, gargalos ou problemas em tempo real (Gartner 2024c). Dado o aumento da complexidade dos ambientes tecnológicos, a análise manual de *logs* tornou-se impraticável, tornando essas ferramentas indispensáveis para acompanhar o funcionamento de aplicações e infraestruturas modernas (Sedki, Hamou-Lhadj e Mohamed 2024).

Componentes fundamentais desses sistemas incluem *pipelines* de processamento configuráveis, que filtram e estruturam os dados dos *logs* antes de extrair as métricas relevantes (Elastic.co 2024). Ferramentas de visualização, como *dashboards* interativos, são frequentemente integradas para apresentar as métricas de forma clara e acessível (Splunk 2023).

Os benefícios proporcionados pelos sistemas de extração de métricas de *logs* incluem (Gartner 2024c):

- **Monitorização em tempo real**: Permite identificar problemas imediatamente, reduzindo o impacto em operações críticas.
- **Apoio à tomada de decisões**: As métricas extraídas auxiliam as equipas a tomar decisões baseadas em dados concretos, em vez de suposições.
- **Otimização de desempenho**: Fornece *insights* sobre o desempenho do sistema, permitindo ajustes proativos para maximizar a eficiência.

Apesar dos benefícios, a implementação desses sistemas apresenta desafios, como a complexidade na configuração inicial, que requer um entendimento detalhado das fontes de dados e das métricas necessárias. Além disso, o processamento de grandes volumes de dados em tempo real pode exigir recursos computacionais significativos, aumentando os custos operacionais (Morley, Brito e Welling 2018).

2.1.4 Impacto de Logs na Otimização Operacional

Os dados provenientes de *logs* desempenham um papel central na otimização de processos operacionais em ambientes corporativos. Esses registos oferecem uma visão detalhada e em tempo real sobre o desempenho dos sistemas, permitindo identificar gargalos, prever falhas e melhorar a eficiência das operações (Sukma et al. 2019). A capacidade de analisar grandes volumes de *logs* de forma eficaz tornou-se indispensável para organizações que procuram permanecer competitivas em mercados dinâmicos e tecnologicamente avançados (Miranskyy et al. 2016).

A identificação de gargalos operacionais é uma das principais vantagens da análise de *logs*. Esses dados permitem rastrear o desempenho de componentes críticos em sistemas e aplicações, identificando pontos de ineficiência que afetam a produtividade. Por exemplo, através da análise de tempos de resposta ou taxas de erro, as organizações conseguem localizar áreas que necessitam de otimização, como processos lentos ou servidores sobrecarregados (P. He et al. 2018).

A análise proativa de *logs* possibilita ainda prever falhas antes que estas impactem as operações. Com o uso de algoritmos de deteção de padrões, as organizações conseguem identificar comportamentos anómalos que indicam problemas iminentes, como quedas de serviço ou falhas de hardware. Essa abordagem reduz o tempo de inatividade e minimiza custos associados a interrupções não planeadas (Korzeniowski e Goczyła 2022b).

Além disso, os *logs* ajudam a otimizar recursos técnicos, como servidores, memória ou largura de banda, identificando o uso subótimo desses recursos. A partir dessa informação, é possível ajustar a alocação de recursos, garantindo maior eficiência e menor desperdício (Awad e Menascé 2016).

A extração de métricas de *logs* também fornece informações estratégicas que auxiliam na tomada de decisão informada. Dados como padrões de utilização de sistemas, comportamentos de utilizadores e indicadores de desempenho tornam-se valiosos para ajustar processos e desenvolver estratégias operacionais mais eficazes (Miranskyy et al. 2016).

Por fim, a análise detalhada de *logs* contribui para melhorar a segurança e a conformidade, identificando vulnerabilidades ou comportamentos suspeitos (P. He et al. 2018). Relatórios gerados a partir dos *logs* ajudam a demonstrar conformidade com regulamentações, como o RGPD, e a preparar a organização para auditorias (Korzeniowski e Goczyła 2022b).

2.1.5 Conclusão

A revisão da literatura destacou a importância dos sistemas de gestão de *logs* e as suas contribuições significativas para a otimização operacional e a eficiência organizacional. Os sistemas de gestão de *logs* proporcionam centralização, deteção proativa de problemas e suporte à tomada de decisões estratégicas, enquanto enfrentam desafios como escalabilidade, privacidade e integração com sistemas existentes.

Além disso, os avanços na automação e na extração de métricas de *logs* reforçam a necessidade de ferramentas tecnológicas modernas que possam processar grandes volumes de dados em tempo real, identificar padrões relevantes e fornecer *insights* acionáveis. A automação tem-se mostrado essencial para reduzir custos operacionais, melhorar a segurança e assegurar conformidade com regulamentações.

Por fim, o impacto dos *logs* na otimização operacional evidencia o seu papel central na identificação de gargalos, previsão de falhas e otimização de recursos técnicos. Esta análise revela a relevância de uma abordagem integrada para a gestão de *logs*, apoiada por tecnologias avançadas e processos bem estruturados, como fundamentos essenciais para alcançar eficiência, segurança e competitividade em ambientes corporativos modernos.

2.2 Tendências

O avanço contínuo da tecnologia tem impulsionado transformações significativas no domínio da monitorização e análise de *logs* empresariais. As tendências emergentes neste campo refletem a crescente demanda por soluções mais inteligentes, escaláveis e integradas, capazes de lidar com os desafios impostos pelos grandes volumes de dados e pela complexidade dos sistemas modernos.

2.2.1 Automação e Inteligência Artificial

A automação e a Inteligência Artificial (IA) têm transformado significativamente a forma como as organizações monitorizam e analisam os seus sistemas e operações (Carvajal e Garcia-Colon 2003). Estas tecnologias avançadas estão a ser integradas em ferramentas de monitorização de *logs* para aumentar a eficiência, reduzir a intervenção humana e proporcionar *insights* mais precisos e rápidos (Korzeniowski e Goczyła 2022c; Skopik, Landauer e Wurzenberger 2022).

- **Deteção de Anomalias e Padrões:** Utilizando algoritmos de Aprendizagem Automática (Machine Learning) (ML), as ferramentas modernas conseguem identificar desvios em relação ao comportamento esperado do sistema, muitas vezes antes que os problemas se tornem críticos (Skopik, Landauer e Wurzenberger 2022).
- **Automação de Respostas:** Ferramentas modernas com IA oferecem automação de respostas a incidentes, como a ativação de processos de mitigação automáticos ou notificações personalizadas (Niu et al. 2022).
- **Classificação e Organização de Dados:** A IA é utilizada para classificar e organizar grandes volumes de dados de *logs*, permitindo que as equipas se concentrem nos eventos mais críticos (Carvajal e Garcia-Colon 2003).

- **Melhoria Contínua:** As ferramentas de monitorização com IA utilizam processos de aprendizagem contínua para melhorar a eficácia ao longo do tempo (Korzeniowski e Goczyła 2022c).

2.2.2 Segurança e Conformidade

Com o aumento das ameaças cibernéticas e a crescente regulamentação, como o RGPD, as empresas enfrentam a necessidade de adotar medidas rigorosas para proteger informações sensíveis, assegurar a privacidade dos utilizadores e garantir transparência no tratamento de dados (Brandao e Georgieva 2020).

Para atender a estas exigências, as ferramentas de monitorização de *logs* têm evoluído significativamente, incorporando funcionalidades avançadas que garantem maior segurança e conformidade. Entre as principais capacidades oferecidas estão:

- **Encriptação de Logs:** A encriptação ponta a ponta dos dados de logs protege informações sensíveis contra acessos não autorizados, tanto em trânsito como em repouso (Almodovar et al. 2024).
- **Controlo de Acesso:** Implementações robustas de controlo de acesso baseado em funções RBAC asseguram que apenas utilizadores autorizados tenham permissão para visualizar ou modificar *logs* específicos (Li et al. 2022).
- **Auditorias Automatizadas:** Ferramentas modernas incluem funcionalidades de auditoria que registam todas as ações realizadas nos *logs*, fornecendo um histórico detalhado que facilita a identificação de atividades suspeitas ou violações de conformidade (Zhao, Jiang e Ma 2022).
- **Gestão de Retenção de Dados:** Soluções adaptadas às regulamentações permitem configurar políticas de retenção de dados, eliminando automaticamente informações após o prazo definido, em conformidade com os requisitos legais (Gökstorp et al. 2024).
- **Alertas de Segurança:** Sistemas avançados conseguem detetar acessos suspeitos ou tentativas de manipulação de dados em tempo real, notificando as equipas responsáveis imediatamente (Brandao e Georgieva 2020).

Além das funcionalidades específicas, estas ferramentas estão a integrar-se com sistemas de Gestão de Identidade e Acesso (Identity and Access Management) (IAM) e soluções de Gestão de Informações e Eventos de Segurança (Security Information and Event Management) (SIEM), criando uma abordagem mais holística para a segurança. Esta integração possibilita uma visão unificada das atividades, permitindo respostas rápidas e eficazes a incidentes (Almodovar et al. 2024).

2.2.3 Integração com Plataformas de DevOps

A integração com *pipelines* de DevOps tem vindo a consolidar-se como uma prática essencial para garantir a eficiência e a qualidade dos sistemas modernos. Esta abordagem permite que as equipas de desenvolvimento e operações monitorizem e analisem *logs* diretamente no ciclo de desenvolvimento, criando uma ligação contínua entre os processos de integração, entrega e monitorização (Narendiran et al. 2023).

- **Suporte nativo para CI/CD:** Ferramentas modernas de monitorização e análise de *logs* oferecem suporte nativo para sistemas de Integração Contínua/Entrega Contínua (Continuous Integration and Continuous Delivery) (CI/CD), facilitando a deteção precoce de erros e anomalias ao longo do ciclo de vida do *software* (Arachchi e Perera 2018).
- **Automação e Monitorização Contínua:** A automação permite que *logs* sejam gerados e analisados automaticamente sempre que uma nova versão do *software* é integrada ou implantada. Isso garante um fluxo contínuo de informações úteis para identificar problemas de desempenho, falhas de integração e outras questões críticas (Cowell, Lotz e Timberlake 2023).
- **Feedback Imediato:** Ferramentas modernas fornecem *feedback* imediato às equipas, permitindo ajustes em tempo real. *Logs* detalhados gerados durante integrações ou implementações oferecem informações valiosas sobre o desempenho das mudanças realizadas no código (Narendiran et al. 2023).
- **Visualização e Colaboração:** *Dashboards* interativos consolidam informações de logs em visualizações compreensíveis, facilitando a colaboração entre as equipas de desenvolvimento e operações e promovendo uma abordagem proativa para a resolução de problemas (Dileepkumar e Mathew 2023).
- **Redução do Tempo de Resolução:** A integração com plataformas de DevOps permite identificar e resolver problemas de forma mais rápida, reduzindo significativamente o tempo médio de resolução Tempo Médio de Resolução (Mean Time to Repair) (MTTR), crítico em ambientes ágeis (Cowell, Lotz e Timberlake 2023).

A adoção de práticas de DevOps, aliada à integração com ferramentas avançadas de monitorização, melhora a eficiência operacional e garante a entrega de sistemas confiáveis e alinhados com os objetivos estratégicos da organização. Esta abordagem reflete o papel central que a colaboração contínua e a monitorização desempenham no sucesso das iniciativas modernas de desenvolvimento e operações (Dileepkumar e Mathew 2023).

2.2.4 Foco em Experiência do Utilizador

A Experiência do Utilizador (User Experience) (UX) tem-se tornado um fator crítico no desenvolvimento de ferramentas de monitorização e análise de *logs*. Um design intuitivo não só melhora a adoção das soluções por utilizadores com diferentes níveis de competência técnica, como também promove a eficiência operacional, reduzindo o tempo necessário para interpretar dados e tomar decisões informadas (Wang, AlKadi e Bach 2023).

- **Dashboards Interativos:** Ferramentas modernas oferecem *dashboards* com visualizações dinâmicas e interativas, permitindo aos utilizadores explorar os dados em profundidade, filtrar informações específicas e gerar relatórios personalizados (Veronica e Suryawan 2019).
- **Visualizações Personalizáveis:** A capacidade de criar gráficos, relatórios e visualizações adaptados às necessidades de cada equipa tem aumentado significativamente o valor percebido das ferramentas (Bratskas et al. 2024).
- **Mobilidade e Acesso Remoto:** Com a crescente adoção de dispositivos móveis e trabalho remoto, as ferramentas estão a ser otimizadas para acesso em *smartphones*

e *tablets*, mantendo a mesma eficiência e usabilidade (Aniko, Kusumasari e Suakanto 2024).

O foco crescente em UX reflete uma compreensão mais profunda de que a simplicidade e a eficiência na interface aumentam a produtividade e a adesão às ferramentas (Orciuolo et al. 2024).

2.3 Tecnologias Utilizadas na Organização

Nesta secção, serão apresentadas as tecnologias existentes que desempenham um papel central na solução atualmente utilizada e que possuem relevância direta para o desenvolvimento da nova solução proposta. O objetivo é fornecer um contexto detalhado sobre as ferramentas já integrados ao ambiente atual, destacando suas capacidades, limitações e aplicabilidade no projeto.

2.3.1 Log4J

O Log4J é uma biblioteca de registo amplamente utilizada no desenvolvimento de *software* para o registo de eventos e mensagens geradas por aplicações. Desenvolvida pela Apache Software Foundation, esta ferramenta de código aberto é reconhecida pela sua flexibilidade, escalabilidade e eficiência na gestão de registos, sendo uma escolha popular em sistemas empresariais.

A principal funcionalidade do *Log4J* consiste em permitir que aplicações registem mensagens em diferentes níveis de prioridade, como *DEBUG*, *INFO*, *WARN*, *ERROR* e *FATAL*. Esta hierarquia facilita o diagnóstico de problemas e o acompanhamento do desempenho de sistemas, permitindo às equipas técnicas identificar anomalias e tomar medidas corretivas de forma mais célere.

Outro ponto forte do Log4J é a sua capacidade de configurar destinos de registo, conhecidos como *appenders*, que possibilitam o envio de mensagens para ficheiros de texto, bases de dados, sistemas de monitorização ou consolas em tempo real. As configurações podem ser definidas de forma dinâmica, utilizando ficheiros de propriedades ou no formato XML, o que garante flexibilidade para diferentes ambientes e necessidades.

Além disso, o Log4J suporta a personalização dos padrões de formatação para os registos gerados, permitindo que cada organização ou aplicação defina a estrutura das mensagens de acordo com os seus requisitos.

No contexto deste projecto, o Log4J assume um papel relevante na geração e armazenamento de registos no sistema actual. A sua robustez e integração facilitada com outras tecnologias tornam-no uma ferramenta para a gestão de eventos críticos e para a extração de informações relevantes, as quais serão analisadas e optimizadas na solução futura.

2.3.2 Oracle

O Oracle Database é um sistema de gestão de bases de dados relacional amplamente utilizado em ambientes empresariais. Desenvolvido pela Oracle Corporation, é reconhecido pela sua robustez, escalabilidade e capacidade de lidar com grandes volumes de dados, sendo especialmente relevante em cenários que exigem alta disponibilidade, desempenho e segurança.

Uma das principais características do Oracle Database é o seu suporte a operações transacionais e analíticas, permitindo que as organizações utilizem a mesma plataforma tanto para processamento operacional em tempo real como para análise de dados históricos. Este equilíbrio entre operações Processamento de Transacções em Linha (Online Transaction Processing) (OLTP) e Processamento Analítico em Linha (Online Analytical Processing) (OLAP) torna-o uma escolha ideal para empresas que precisam de uma solução integrada.

Adicionalmente, o Oracle Database oferece funcionalidades avançadas, como particionamento de tabelas, compressão de dados e suporte a tecnologias de *big data*, garantindo que o sistema permaneça eficiente. A segurança é outro destaque, com ferramentas robustas de controlo de acessos, encriptação de dados e auditoria de actividades, cumprindo os mais rigorosos padrões de conformidade, como o RGPD.

No contexto da gestão de registos empresariais, o Oracle Database é frequentemente utilizado para armazenar *logs* críticos que requerem elevada fiabilidade e integridade, como transacções financeiras ou dados sensíveis relacionados com as operações de negócio. A capacidade de realizar consultas complexas e análises de dados directamente na base de dados torna-o uma ferramenta essencial para sistemas que dependem de métricas e relatórios gerados a partir de grandes volumes de registos.

2.3.3 Java

O Java é uma linguagem de programação amplamente utilizada em aplicações empresariais devido à sua portabilidade, robustez e capacidade de lidar com sistemas complexos. Desenvolvida originalmente pela *Sun Microsystems* em 1995 e actualmente mantida pela Oracle Corporation, o Java tornou-se uma das linguagens mais populares no desenvolvimento de *software*, com uma vasta comunidade de programadores e suporte extensivo.

Uma das características distintivas do Java é o conceito de "escreva uma vez, execute em qualquer lugar"(*Write Once, Run Anywhere*), possibilitado pela sua Máquina Virtual Java (Java Virtual Machine) (JVM). Este recurso permite que aplicações escritas em Java sejam executadas em diferentes sistemas operativos sem a necessidade de ajustes significativos, tornando-o ideal para ambientes empresariais diversificados.

O Java oferece suporte a uma vasta gama de *frameworks* e bibliotecas, como Spring, Hibernate e Apache Kafka, que simplificam o desenvolvimento de aplicações robustas e escaláveis. Além disso, a linguagem suporta *multithreading* e processamento paralelo, recursos críticos para aplicações de alto desempenho, como sistemas de monitorização e análise de *logs*.

Outro aspecto relevante é a segurança integrada no *Java*, com funcionalidades como a encriptação de dados, controlo de acessos e a capacidade de criar aplicações que cumprem os padrões de conformidade mais exigentes, como o RGPD. Este foco na segurança torna-o especialmente adequado para sistemas que gerem informações sensíveis ou críticas.

No contexto deste projecto, o Java desempenha um papel essencial como linguagem de programação para a implementação de soluções de monitorização e análise de *logs* empresariais. A sua capacidade de integração com tecnologias como o Log4J e bases de dados, como o Oracle Database, permite o desenvolvimento de uma solução eficiente e adaptável às necessidades da organização. Além disso, a sua fiabilidade e suporte a operações em tempo real garantem que a nova solução seja capaz de lidar com os desafios impostos por grandes volumes de dados e requisitos de escalabilidade.

2.4 Ferramentas Existentes

O mercado de ferramentas para monitorização, análise e gestão de *logs* empresariais tem evoluído significativamente, acompanhando as crescentes necessidades por soluções escaláveis, eficientes e integradas. Essas ferramentas para lidar com o grande volume de dados gerados diariamente por aplicações, sistemas e dispositivos, permitindo às organizações monitorizar operações, identificar problemas rapidamente e tomar decisões informadas.

Com opções que variam desde ferramentas comerciais robustas, como Splunk e Datadog, até soluções *open-source* amplamente adotadas, como Elasticsearch e Graylog, o panorama tecnológico oferece uma ampla gama de escolhas adaptáveis às necessidades específicas de diferentes organizações. Além disso, ferramentas como Grafana e Prometheus ampliam as capacidades de visualização e automação, fornecendo *insights* poderosos para a gestão de sistemas.

Nesta secção, serão apresentadas as principais ferramentas disponíveis no mercado, analisando suas características, funcionalidades e benefícios no contexto da gestão de *logs* empresariais. Essa análise tem como objetivo identificar quais soluções são mais adequadas para atender aos objetivos do projeto, considerando aspectos como custo, escalabilidade, facilidade de uso e integração com sistemas existentes.

2.4.1 ELK Stack (Elasticsearch, Logstash, Kibana)

A ferramenta *Elasticsearch, Logstash e Kibana* (ELK) é amplamente reconhecida como uma das soluções *open-source*. Oferece uma abordagem integrada e escalável para a gestão de grandes volumes de dados. A flexibilidade e a capacidade de personalização tornam-se uma escolha para empresas que desejam centralizar e otimizar suas operações relacionadas à análise de *logs* (Snoop 2024).

A Figura 2.2 apresenta um exemplo das capacidades visuais e de monitorização da ELK Stack, evidenciando como as suas ferramentas integradas ELK permitem consolidar, processar e visualizar informações de forma centralizada e eficiente.

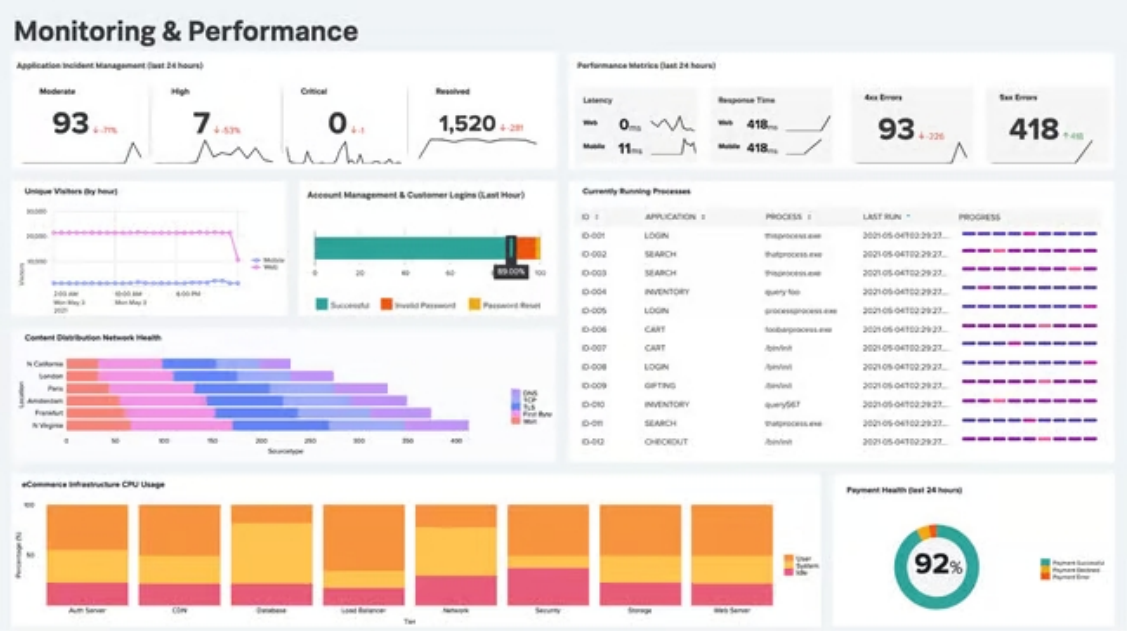


Figura 2.2: ELK Stack (Isaiah 2024)

Para compreender melhor as características da ELK Stack, são analisados fatores que influenciam sua aplicação em diferentes cenários empresariais, como custo, desempenho, escalabilidade, curva de aprendizagem, segurança e suporte. A seguir, são apresentados os principais pontos sobre cada um desses fatores:

- **Custo:** A ELK Stack destaca-se por ser uma solução *open-source*, o que reduz significativamente os custos em comparação com ferramentas comerciais. No entanto, é importante considerar custos adicionais associados à implementação, manutenção e recursos de infraestrutura necessários para sustentar grandes volumes de dados (TechHyme 2024).
- **Desempenho:** O desempenho da ELK Stack é amplamente reconhecido, especialmente em ambientes que demandam processamento intensivo de *logs*. A sua arquitetura escalável permite lidar com milhões de registos por segundo, tornando-a uma escolha confiável para empresas que necessitam de análises rápidas e detalhadas (Snoop 2024).
- **Escalabilidade:** A escalabilidade é uma das maiores vantagens da ELK Stack. Com o Elasticsearch, é possível expandir a infraestrutura de forma horizontal, adicionando novos nós para atender ao crescimento exponencial dos dados, sem comprometer o desempenho geral do sistema (TechHyme 2024).
- **Curva de Aprendizagem:** Apesar de seus benefícios, a curva de aprendizagem para implementar e configurar a ELK Stack pode ser acentuada. Dominar a integração de Elasticsearch, Logstash e Kibana exige conhecimentos técnicos avançados, além de tempo para otimização e personalização da solução (TechHyme 2024).
- **Segurança:** A ELK Stack oferece diversas funcionalidades relacionadas à segurança, como controlo de acesso baseado em funções RBAC e encriptação de dados. No

entanto, a configuração desses recursos requer atenção, especialmente em ambientes corporativos que demandam conformidade com regulamentações como o RGPD (Snoop 2024).

- **Suporte:** Embora a ELK Stack seja uma solução *open-source*, sua ampla comunidade de utilizadores e desenvolvedores fornece suporte ativo e contínuo. Além disso, a Elastic, empresa responsável pelo Elasticsearch, oferece serviços pagos de suporte técnico, que podem ser úteis para empresas que desejam suporte especializado (TechHyme 2024).

2.4.2 Splunk

Splunk é uma ferramenta comercial amplamente reconhecida no mercado por suas capacidades de monitorização, análise e visualização de *logs* e eventos em tempo real, projetada para lidar com grandes volumes de dados provenientes de sistemas, aplicações e dispositivos, Splunk permite às organizações recolher, indexar e correlacionar dados de maneira eficiente, proporcionando uma visão abrangente do desempenho de suas operações e infraestrutura (Splunk 2024g). Essa plataforma destaca-se por sua interface intuitiva, recursos avançados de análise e integração com tecnologias modernas, como *machine learning* e automação, sendo amplamente utilizada em setores como TI, segurança cibernética e análise de negócios (Splunk 2024c).

Conforme ilustrado na Figura 2.3, a interface do Splunk oferece uma visão centralizada e detalhada das operações de segurança e desempenho, apresentando métricas críticas, gráficos e eventos relevantes de maneira acessível e dinâmica.

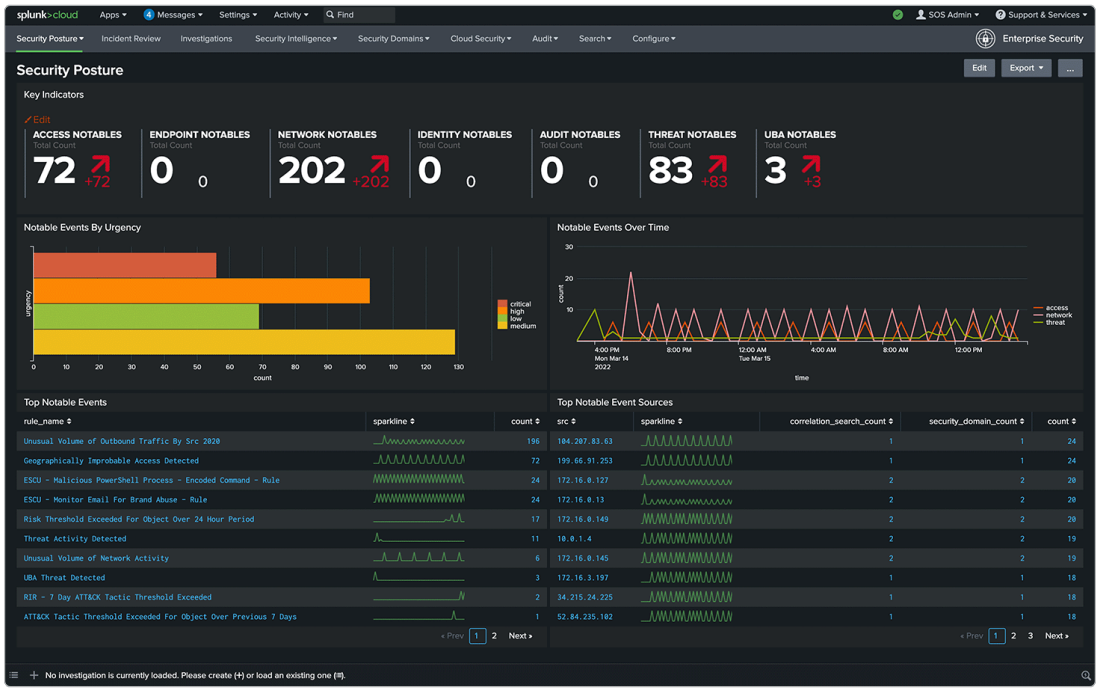


Figura 2.3: Splunk (Splunk 2024f)

Para compreender melhor as características do Splunk, são analisados fatores que influenciam sua aplicação em diferentes cenários empresariais:

- **Custo:** Splunk é uma solução comercial com custos elevados, especialmente para empresas que lidam com grandes volumes de dados (Splunk 2024d).
- **Desempenho:** A sua arquitetura permite a ingestão e análise de dados em tempo real, garantindo respostas rápidas a incidentes críticos e suportando operações contínuas (Splunk 2024b).
- **Escalabilidade:** Splunk oferece opções de escalabilidade, tanto horizontal quanto vertical, o que permite às organizações expandirem sua infraestrutura conforme necessário. No entanto, a escalabilidade pode ser limitada pelo custo e pelas demandas de recursos computacionais (Splunk 2024a).
- **Curva de Aprendizagem:** Apesar da interface intuitiva, a configuração e personalização do Splunk exigem conhecimentos técnicos avançados, especialmente para configurar integrações complexas e criar análises detalhadas (Splunk 2024c).
- **Segurança:** Splunk fornece recursos avançados de segurança, como correlação de eventos, detecção de anomalias e suporte a regulamentações como o RGPD. Essas funcionalidades tornam-no uma solução confiável para ambientes que demandam proteção contra ameaças cibernéticas (Splunk 2024e).
- **Suporte:** Splunk oferece suporte técnico robusto através de contratos comerciais, garantindo assistência especializada e atualizações contínuas. A empresa também fornece uma ampla base de conhecimento e uma comunidade ativa para suporte adicional (Splunk 2024c).

2.4.3 Graylog

Graylog é uma solução *open-source* amplamente reconhecida no mercado pela sua eficiência e simplicidade na monitorização, análise e gestão de *logs* empresariais (Graylog 2024c). Foi desenvolvida para atender às necessidades de organizações que lidam com grandes volumes de dados provenientes de aplicações, sistemas e dispositivos (Graylog 2024a). Uma das suas principais características é a capacidade de organizar e centralizar dados de diversas fontes em um único ambiente, utilizando o *Elasticsearch* como mecanismo de busca e armazenamento, o que proporciona escalabilidade e rapidez nas consultas.

A Figura 2.4 ilustra a interface do Graylog, que apresenta painéis visuais interativos, como gráficos de histogramas e tabelas de mensagens, destacando a forma como a ferramenta organiza e visualiza os dados recolhidos.

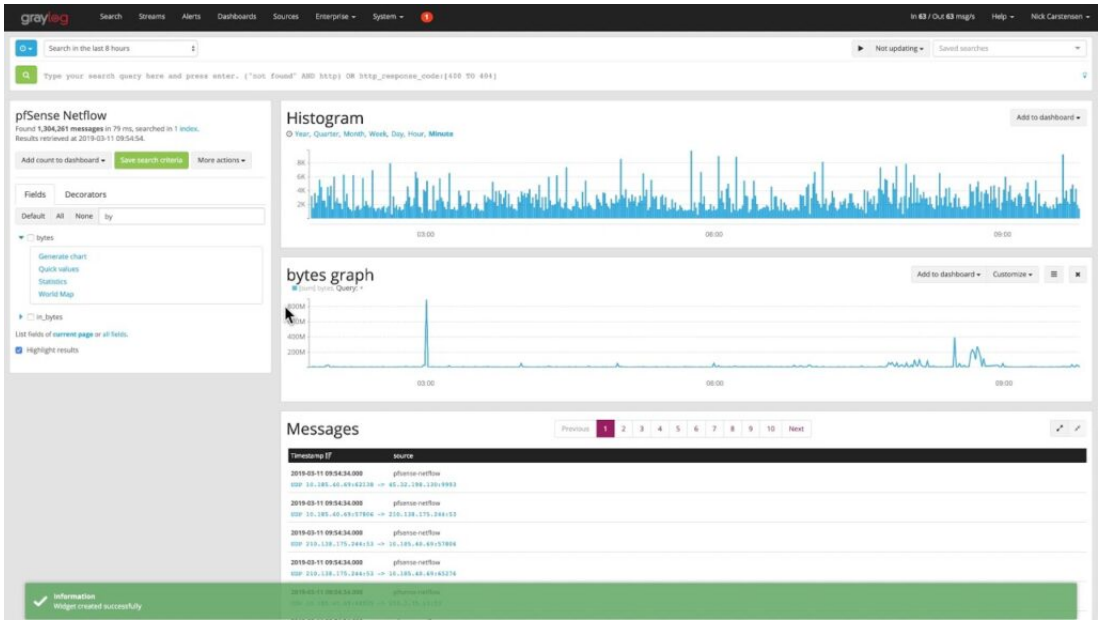


Figura 2.4: Graylog (*Graylog: Solução Poderosa para Gerenciamento de Logs* 2024)

Para uma análise das capacidades do Graylog, são considerados fatores essenciais que determinam sua aplicabilidade em distintos contextos empresariais, destacando sua adequação para monitorização e gestão eficiente de dados:

- **Custo:** Sendo *open-source*, Graylog oferece uma solução acessível para empresas de todos os portes, eliminando custos associados a licenças (Habbema 2024). No entanto, algumas funcionalidades avançadas estão disponíveis apenas na versão paga (Graylog 2024b).
- **Desempenho:** A utilização do *Elasticsearch* como base permite consultas rápidas mesmo com grandes volumes de dados (Graylog 2024a). A capacidade de monitorização em tempo real reduz o tempo necessário para identificar e resolver problemas (Graylog 2024c).
- **Escalabilidade:** Graylog é projetado para lidar com grandes volumes de *logs*, adaptando-se a ambientes de alta demanda. No entanto, em grandes escalas, os requisitos de infraestrutura podem aumentar significativamente (Graylog 2024a).
- **Curva de Aprendizagem:** A interface intuitiva do Graylog facilita o uso por utilizadores com menos experiência técnica. No entanto, a configuração inicial pode ser desafiadora para quem não tem familiaridade com ferramentas de monitorização (Habbema 2024).
- **Segurança:** Graylog oferece opções robustas para garantir a proteção dos dados monitorizados, incluindo encriptação de dados em trânsito e controles de acesso (Graylog 2024b).
- **Suporte:** A ferramenta conta com uma comunidade ativa de utilizadores e programadores, proporcionando suporte contínuo. A versão paga oferece suporte técnico direto e funcionalidades adicionais (Graylog 2024b).

2.4.4 Datadog

Datadog é uma plataforma de monitorização e análise amplamente utilizada para a gestão de logs, métricas e traços distribuídos em um único ambiente integrado (finout 2024). Reconhecida por sua interface intuitiva e capacidades abrangentes, a ferramenta é projetada para oferecer visibilidade em tempo real sobre a infraestrutura e as aplicações empresariais (finout 2024). A solução combina funcionalidades avançadas de monitorização, como detecção de anomalias, visualização de dados e configuração de alertas personalizados, sendo ideal para arquiteturas modernas baseadas em microserviços e ambientes de nuvem (Datadog 2024d).

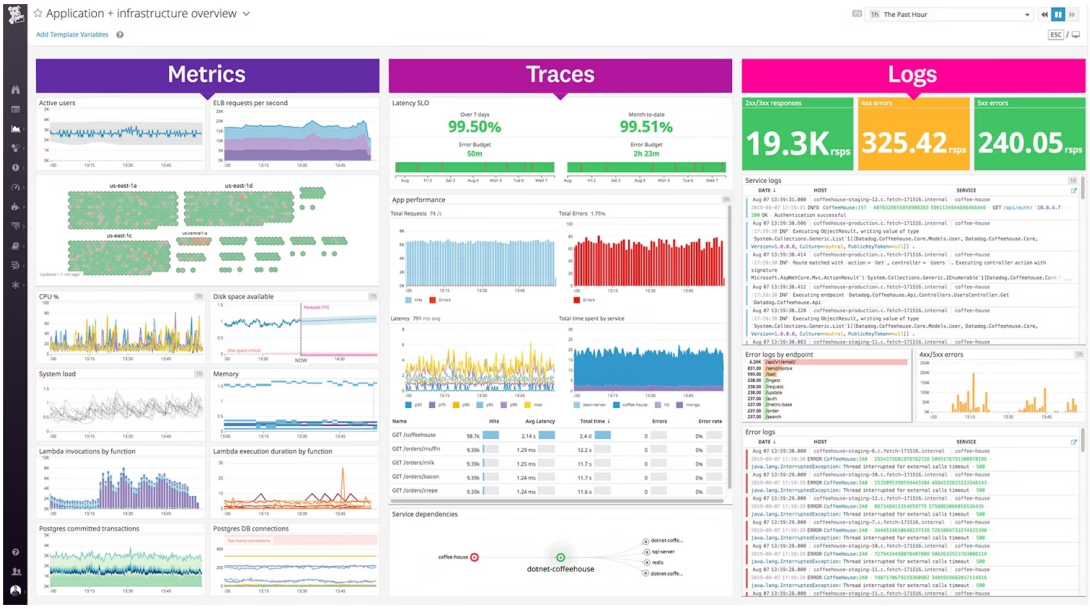


Figura 2.5: Datadog (Datadog 2024c)

Para avaliar o Datadog, é essencial considerar fatores como custo, desempenho, escalabilidade e segurança, que influenciam sua aplicação em diferentes cenários empresariais. A seguir, são analisados esses aspetos principais:

- **Custo:** O Datadog é uma ferramenta comercial cujo custo pode ser elevado, particularmente para empresas que necessitam monitorizar grandes volumes de dados ou ativar várias funcionalidades avançadas. O modelo de licenciamento baseado no número de *hosts* e funcionalidades específicas pode resultar em despesas significativas para organizações de maior porte (Datadog 2024a).
- **Desempenho:** O Datadog é reconhecido pelo desempenho elevado na monitorização em tempo real. A capacidade de correlacionar *logs*, métricas e traços em uma única plataforma garante a identificação rápida de problemas e a execução de diagnósticos detalhados (Datadog 2024d).
- **Escalabilidade:** A arquitetura do Datadog é projetada para suportar ambientes de grande escala, como sistemas distribuídos e baseados em nuvem. A ferramenta é particularmente eficaz em infraestruturas modernas que utilizam microserviços, oferecendo suporte para um número elevado de fontes de dados sem degradação do desempenho (Datadog 2024d).

- **Curva de Aprendizagem:** Embora o Datadog possua uma interface intuitiva, a configuração inicial e a integração com sistemas complexos podem exigir conhecimento técnico avançado. Equipas sem experiência prévia podem enfrentar uma curva de aprendizagem acentuada (Datadog 2024b).
- **Segurança:** O Datadog oferece funcionalidades robustas de segurança, como encriptação de dados, autenticação multi-fator e controle de acesso baseado em funções RBAC. Estas funcionalidades garantem conformidade com regulamentações como o RGPD e protegem os dados de monitorização contra acessos não autorizados (Datadog 2024d).
- **Suporte:** A ferramenta dispõe de suporte técnico abrangente e documentação detalhada, facilitando a resolução de problemas e a adoção da solução. A comunidade ativa e os recursos de suporte ajudam a mitigar desafios técnicos durante a implementação e operação (Datadog 2024b).

2.4.5 Conclusão

A seleção da ferramenta para a implementação da solução de monitorização e análise de logs empresariais foi baseada em critérios técnicos e operacionais que garantem alinhamento com os objetivos do projeto e as necessidades da organização. As quatro ferramentas analisadas ELK Stack, Splunk, Graylog e Datadog apresentam características distintas, sendo importantes para a decisão final.

A Tabela 2.1 apresenta uma comparação detalhada entre as ferramentas analisadas (ELK Stack, Splunk, Graylog e Datadog), considerando os principais fatores de avaliação.

Tabela 2.1: Avaliação Comparativa das Ferramentas de Monitorização e Análise de *Logs*

Fator	ELK Stack	Splunk	Graylog	Datadog
Custo	5 - Baixo	1 - Muito Alto	3 - Médio	2 - Alto
Desempenho	5 - Muito Alto	4 - Alto	4 - Alto	5 - Muito Alto
Escalabilidade	5 - Muito Alta	4 - Alta	4 - Alta	5 - Muito Alta
Curva de Aprendizagem	3 - Média	3 - Média	3 - Média	2 - Alta
Segurança	5 - Muito Alta	5 - Muito Alta	4 - Alta	5 - Muito Alta
Suporte	4 - Comunidade Ativa	5 - Suporte Comercial	4 - Comunidade Ativa	5 - Suporte Comercial
Total	**27**	**22**	**22**	**24**

Nota: A escala varia de 1 a 5, onde 1 representa o cenário menos favorável (ex.: custo elevado, desempenho baixo) e 5 representa o cenário mais favorável (ex.: custo baixo, desempenho elevado).

Com base na análise comparativa das ferramentas de monitorização e análise de *logs*, a escolha da tecnologia a ser implementada no projeto baseou-se em critérios técnicos e operacionais, com ênfase nos fatores eliminatórios de custo, desempenho e escalabilidade.

Entre as quatro ferramentas analisadas **ELK Stack**, **Splunk**, **Graylog** e **Datadog**, a **ELK Stack** destaca-se como a opção mais alinhada aos objetivos do projeto. Essa escolha é sustentada pelos seguintes pontos:

- **Custo:** A ELK Stack, sendo uma solução *open-source*, apresenta uma vantagem significativa no que diz respeito aos custos. Essa característica torna-a uma escolha

acessível para a organização, eliminando barreiras financeiras associadas às ferramentas comerciais, como Splunk e Datadog, que possuem modelos de licenciamento onerosos.

- **Desempenho:** Em termos de desempenho, o ELK Stack sobressai pela sua capacidade de processar grandes volumes de dados em tempo real, permitindo análises detalhadas e respostas rápidas a incidentes. A ferramenta é especialmente adequada para ambientes com elevada demanda de processamento.
- **Escalabilidade:** A arquitetura escalável do ELK Stack é outro fator determinante. A possibilidade de expandir a infraestrutura horizontalmente, adicionando novos nós conforme o crescimento das necessidades de dados, garante que a solução possa acompanhar a evolução da organização sem comprometer o desempenho.

Embora ferramentas como o Splunk e o Datadog apresentem funcionalidades avançadas e interfaces intuitivas, o elevado custo associado a essas soluções exclui-as como opções viáveis no contexto atual. Já o Graylog, apesar de oferecer um bom equilíbrio entre custo e funcionalidade, não atinge o mesmo nível de desempenho e escalabilidade proporcionado pela ELK Stack.

Portanto, considerando os fatores eliminatórios e as necessidades específicas do projeto, a **ELK Stack** é a ferramenta recomendada para a implementação da solução, proporcionando uma combinação ideal de custo-benefício, alto desempenho e escalabilidade.

2.5 Desafios

O desenvolvimento da solução para a monitorização e análise de *logs* empresariais apresenta um conjunto de desafios técnicos, operacionais e organizacionais que devem ser superados para garantir o sucesso do projeto. Estes desafios refletem a complexidade do ambiente empresarial moderno, que envolve a gestão de grandes volumes de dados, a integração de sistemas diversificados e a garantia de segurança e conformidade.

2.5.1 Complexidade da Integração

A integração com sistemas existentes, sejam eles legados ou modernos, representa um dos maiores desafios no desenvolvimento da solução. Este processo exige atenção a vários aspetos técnicos e organizacionais, incluindo:

- **Compatibilidade com Sistemas Legados:** Sistemas legados frequentemente possuem arquiteturas desatualizadas ou personalizadas que dificultam a integração com tecnologias modernas.
- **Diversidade de Protocolos e APIs:** A coexistência de diferentes protocolos de comunicação e formatos de API nos sistemas atuais exige a criação de adaptadores ou *middleware* para unificar a troca de dados entre os componentes.
- **Sincronização de Dados:** A manutenção da consistência e integridade dos dados ao longo dos sistemas integrados, especialmente em operações em tempo real, apresenta desafios adicionais em termos de desempenho e complexidade.
- **Gestão de Dependências:** A dependência de sistemas externos ou de terceiros, como *APIs* de fornecedores, pode trazer riscos de falhas ou mudanças inesperadas, que impactam negativamente a integração.

- **Documentação Incompleta ou Inexistente:** Muitas vezes, os sistemas legados não possuem documentação detalhada, dificultando a compreensão da lógica existente e o desenvolvimento de soluções de integração eficazes.
- **Adaptação a Mudanças Futuras:** A arquitetura da solução deve ser suficientemente flexível para acomodar futuras mudanças nos sistemas integrados, evitando a necessidade de grandes reformulações.

Estes desafios refletem a complexidade do processo de integração e reforçam a necessidade de planeamento e execução cuidadosos para garantir o sucesso da solução proposta.

2.5.2 Gestão de Recursos

A gestão de recursos técnicos é um dos desafios mais relevantes no desenvolvimento da solução para a gestão e análise de *logs* empresariais. Este tópico envolve assegurar que os recursos tecnológicos disponíveis sejam suficientes e adequados para atender às necessidades do projeto. Os principais desafios associados incluem:

- **Capacidade do Hardware:** A infraestrutura existente pode não ser capaz de suportar o volume crescente de dados gerados, o que pode impactar diretamente o desempenho do sistema.
- **Dependência de Tecnologias Legadas:** A coexistência de tecnologias modernas e sistemas legados impõe restrições ao uso de recursos técnicos, limitando a integração e a evolução do sistema.
- **Escalabilidade:** A capacidade de adaptar os recursos técnicos para lidar com o crescimento exponencial dos dados é fundamental.
- **Manutenção de Infraestrutura:** A gestão contínua de recursos, incluindo atualizações de *hardware* e *software*, é essencial para evitar falhas e assegurar a continuidade operacional.

Esses desafios destacam a importância de um planeamento técnico detalhado para otimizar o uso dos recursos disponíveis e garantir que a infraestrutura seja capaz de suportar a solução proposta.

3. Análise e Design da Solução

Este capítulo apresenta a análise e conceção da solução proposta para a monitorização e análise de *logs*, com foco na extração de métricas de negócio e na geração de alertas automatizados. Para isso, são identificados os principais requisitos funcionais e não funcionais, bem como a arquitetura da solução e a modelagem do sistema.

3.1 Engenharia de Requisitos

A definição de requisitos é um passo fundamental no desenvolvimento de qualquer sistema, garantindo que a solução atenda às necessidades dos utilizadores e aos objetivos organizacionais. No contexto deste projeto, a engenharia de requisitos foca-se na identificação das partes envolvidas, das funcionalidades esperadas e das restrições técnicas que a plataforma de monitorização e análise de logs deve cumprir (Khalique, Butt e Khan 2017).

Esta secção apresenta a identificação dos atores do sistema, os requisitos funcionais e não funcionais, e o mapeamento entre os objetivos do projeto e os requisitos estabelecidos. Estes elementos são essenciais para assegurar que a solução desenvolvida seja escalável, eficiente e alinhada com as necessidades do ambiente empresarial onde será implementada («ISO/IEC/IEEE International Standard - Systems and software engineering – Life cycle processes – Requirements engineering» 2018).

3.1.1 Atores do Sistema

Os atores do sistema representam os diferentes utilizadores e sistemas que interagem com a solução de monitorização e análise de *logs*. Cada ator possui um conjunto específico de permissões e responsabilidades dentro do sistema (Haron e Sahibuddin 2010). Na Tabela 3.1 estão identificadas as parte envolvidas no projeto.

Tabela 3.1: Atores do Sistema

Ator	**Descrição**
Administrador do Sistema	Responsável pela configuração inicial e manutenção da plataforma. Define permissões de utilizadores e regras de alerta. Gere integrações com outras ferramentas e sistemas.
Analista de *Logs*	Monitoriza eventos críticos e analisa padrões de *logs*. Cria relatórios e painéis de controlo com métricas extraídas dos *logs*. Valida a eficiência dos alertas gerados e ajusta as regras conforme necessário.
Gestor de Operações	Utiliza a solução para visualizar métricas de negócio e relatórios. Toma decisões estratégicas com base nos dados analisados. Pode solicitar ajustes nos alertas ou métricas monitorizadas.
Desenvolvedor	Responsável pela manutenção e otimização do código do sistema. Implementa melhorias e corrige erros com base nos *logs* analisados.
Sistema de *Logs* (Ator Externo)	Representa os diferentes sistemas que geram *logs* e os enviam para a plataforma. Pode incluir aplicações web, servidores, bases de dados e outras fontes de eventos.

3.1.2 Requisitos Funcionais

Os requisitos funcionais representam as características e operações que o sistema deve oferecer para garantir uma monitorização eficaz de *logs* e a extração de métricas de negócio. Estes requisitos foram definidos com base nas necessidades da Cleva, assegurando que a solução seja adaptável ao seu ambiente empresarial e compatível com os sistemas já existentes (Khalique, Butt e Khan 2017).

Tabela 3.2: Requisitos Funcionais do Sistema

ID	**Descrição do Requisito**
RF01	O sistema deve permitir a recolha e armazenamento de *logs* provenientes de diferentes fontes, como aplicações, servidores e bases de dados.
RF02	O sistema deve processar e analisar os *logs* em tempo real, identificando padrões e anomalias.
RF03	O sistema deve gerar alertas automáticos em caso de eventos críticos, como falhas de sistema ou tentativas de acesso não autorizado.
RF04	O sistema deve disponibilizar uma interface gráfica que permita visualizar e filtrar *logs* de acordo com critérios definidos pelo utilizador.
RF05	O sistema deve suportar a extração de métricas de negócio a partir dos *logs* analisados.
RF06	O sistema deve permitir a configuração personalizada de regras de alerta pelos utilizadores com permissões adequadas.
RF07	O sistema deve garantir a exportação de relatórios em diferentes formatos, incluindo PDF e CSV.
RF08	O sistema deve garantir a integração dos *logs* provenientes de ficheiros de texto e bases de dados.

A Tabela 3.2 apresenta os requisitos funcionais da plataforma.

3.1.3 Requisitos Não Funcionais

Os requisitos não funcionais determinam as características técnicas e operacionais da plataforma, garantindo que a solução de monitorização e análise de *logs* seja segura, escalável e eficiente. Estes requisitos asseguram que o sistema cumpra os padrões de qualidade necessários para a sua utilização no ambiente empresarial da Cleva (Khalique, Butt e Khan 2017).

Tabela 3.3: Requisitos Não Funcionais do Sistema

ID	Descrição do Requisito
RNF01	O sistema deve ser capaz de processar pelo menos 100.000 eventos de *logs* por segundo sem comprometer o desempenho.
RNF02	O tempo de resposta da interface do utilizador não deve ultrapassar 2 segundos para consultas padrão.
RNF03	A solução deve garantir um tempo de atividade mínimo de 99,9%, assegurando alta disponibilidade.
RNF04	O sistema deve suportar encriptação de dados em trânsito e em repouso, garantindo conformidade com o RGPD.
RNF05	O sistema deve ser escalável horizontalmente, permitindo a adição de novos servidores para lidar com o aumento do volume de *logs*.
RNF06	Deve ser possível configurar permissões de acesso com base em perfis de utilizador, assegurando controlo de acesso baseado em funções (*RBAC*).
RNF07	A interface gráfica deve seguir princípios de usabilidade, garantindo uma experiência intuitiva para os utilizadores.
RNF08	O sistema deve manter *logs* de auditoria detalhados sobre todas as ações realizadas pelos utilizadores com privilégios administrativos.

A Tabela 3.3 apresenta os principais requisitos não funcionais definidos para o projeto.

3.1.4 Mapeamento Objetivos e Requisitos

Para garantir que a solução desenvolvida atenda às necessidades do projeto, é fundamental estabelecer uma relação direta entre os objetivos estratégicos e os requisitos funcionais e não funcionais. O mapeamento apresentado na Tabela 3.4 permite verificar como cada requisito contribui para alcançar os objetivos do sistema, assegurando coerência e rastreabilidade no desenvolvimento da plataforma.

Tabela 3.4: Mapeamento entre Objetivos e Requisitos

Objetivo	**Requisitos Associados**
OBJ1 - Centralizar a gestão de logs	RF01, RF04.
OBJ2 - Garantir a escalabilidade da solução	RNF05, RNF01.
OBJ3 - Reduzir o tempo de deteção e resolução de problemas	RF02, RF03, RNF02, RNF03.
OBJ4 - Facilitar a integração com outras ferramentas	RF08.
OBJ5 - Garantir a segurança e privacidade dos dados	RNF04, RNF06, RNF08.
OBJ6 - Extrair métricas relevantes de negócio	RF05, RF07.
OBJ7 - Acompanhar e otimizar o desempenho dos sistemas	RNF07, RNF03.

Este mapeamento facilita a priorização dos requisitos e auxilia na tomada de decisões durante o processo de desenvolvimento, garantindo que todos os aspetos essenciais do sistema sejam implementados de forma eficaz.

3.2 Arquitetura da solução

Esta secção apresenta a análise comparativa de duas propostas de arquitetura baseadas no *ELK Stack*, desenvolvidas com o objetivo de implementar uma solução eficaz para a monitorização e análise de *logs*. Ambas as propostas foram concebidas com foco na centralização da recolha, tratamento e visualização de *logs* corporativos, permitindo a deteção atempada de incidentes, a geração de métricas de negócio e a emissão automática de alertas.

3.2.1 Proposta de Arquitetura da Solução 1

A primeira proposta arquitetónica assenta numa abordagem linear e centralizada baseada no *ELK Stack*, recorrendo à utilização de *containers Docker* para cada componente. Esta solução visa garantir uma implementação funcional com foco na simplicidade e integração direta entre os elementos da solução, sendo particularmente adequada para contextos com volumes de dados moderados e requisitos operacionais bem definidos.

Como ilustrado na Figura 3.1, esta arquitetura estabelece um fluxo onde os dados são inicialmente recolhidos a partir de uma base de dados *Oracle*, sendo depois encaminhados por dois agentes especializados. O *Metricbeat* é responsável pela recolha de métricas do sistema, enquanto o *Logstash* trata especificamente dos *logs* de negócio. Ambos os fluxos convergem para o *Elasticsearch*, que centraliza o armazenamento e a indexação da informação. A visualização dos dados é realizada através do *Kibana*, que interage diretamente com o *Elasticsearch* para disponibilizar *dashboards*, relatórios e outras representações gráficas que facilitam a análise e interpretação dos eventos monitorizados.

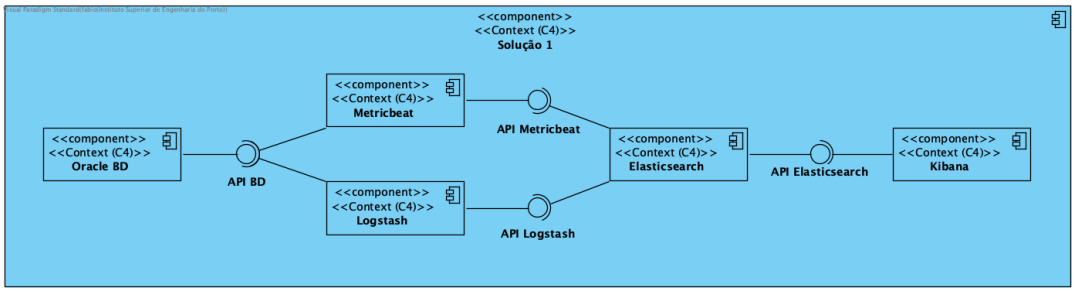


Figura 3.1: Arquitetura proposta na solução 1

Esta arquitetura caracteriza-se por um fluxo de dados sequencial e acoplado, no qual cada componente depende do funcionamento correto do anterior. A utilização combinada de *Logstash* e *Metricbeat* permite cobrir tanto logs de aplicações como métricas de infraestrutura, garantindo uma visão abrangente do sistema monitorizado.

3.2.2 Proposta de Arquitetura da Solução 2

A segunda proposta arquitetónica apresenta uma abordagem modular e distribuída, desenhada para maximizar a escalabilidade, a separação de responsabilidades e a flexibilidade na gestão dos dados. Esta solução baseia-se igualmente na *stack ELK*, mas diferencia-se pelo desacoplamento dos domínios funcionais e pela forma como os dados são organizados e acedidos.

Como ilustrado na Figura 3.2, esta arquitetura mantém o *Elasticsearch* como componente central, responsável pela indexação e armazenamento dos dados recebidos a partir de diferentes fontes especializadas. Os módulos `LIFE` e `CORE_LIFE` desempenham exclusivamente a função de produtores de dados, sendo responsáveis por enviar logs estruturados para o Elasticsearch, onde ficam disponíveis para consulta. Por sua vez, a camada de visualização, assegurada pelo *Kibana*, atua como consumidor dos dados, acedendo aos índices armazenados no Elasticsearch para gerar dashboards, relatórios e visualizações interativas. Este modelo respeita o princípio de isolamento lógico, permitindo que cada domínio execute a sua função de forma autónoma, sem interferência nos restantes componentes.

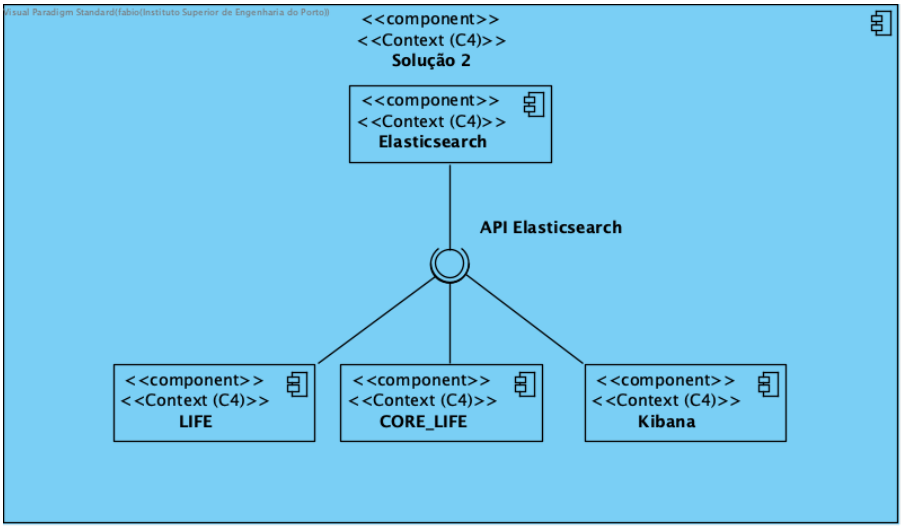


Figura 3.2: Arquitetura proposta na solução 2

A separação da lógica de ingestão por componentes distintos, cada um com configuração própria, facilita significativamente a gestão, manutenção e evolução futura da solução. O módulo `LIFE` poderá, por exemplo, tratar logs associados a operações de negócio, enquanto o `CORE_LIFE` se concentra em eventos técnicos e de infraestrutura. O *Kibana*, por sua vez, mantém-se como ferramenta de análise e exploração dos dados, podendo ser configurado para aceder apenas a subconjuntos específicos, com base em índices ou permissões definidas.

3.2.3 Análise Comparativa das Arquiteturas

Com base nos requisitos definidos e nos princípios de observabilidade, procedeu-se à análise comparativa entre as duas arquiteturas propostas para a implementação da solução de monitorização baseada no *ELK Stack*. Ambas as abordagens oferecem vantagens distintas, sendo adequadas a diferentes contextos organizacionais e operacionais.

A Solução 1 representa uma arquitetura centralizada, com ingestão direta de dados através de *Logstash* e *Metricbeat*, promovendo simplicidade na orquestração e gestão dos componentes. Já a Solução 2 apresenta uma estrutura modular e distribuída, com separação de domínios funcionais, promovendo maior escalabilidade e independência entre componentes.

Tabela 3.5: Principais diferenças entre as arquiteturas

Critério	**Solução 1**	**Solução 2**
Facilidade de configuração	Alto	Baixo
Facilidade de monitorização	Alto	Médio
Escalabilidade	Baixo	Alto
Flexibilidade	Baixo	Alto
Isolamento de falhas	Baixo	Alto
Adequação a monólitos	Alto	Baixo
Adequação a microserviços	Baixo	Alto
Facilidade de auditoria	Alto	Baixo
Complexidade de implementação	Baixa	Alta
Número de componentes	Reduzido	Elevado
Custos operacionais esperados	Mais baixos	Mais elevados

A Tabela 3.5 sintetiza os principais critérios de comparação entre as duas propostas.

3.2.4 Seleção da Arquitetura

Após terem sido analisadas comparativamente as duas arquiteturas propostas, foi selecionada a Solução 1 para implementação no âmbito deste projeto. Esta arquitetura assenta numa estrutura centralizada, com ingestão direta de dados realizada através do *Logstash* e do *Metricbeat*.

A decisão baseou-se numa combinação de fatores técnicos e operacionais, considerando os objetivos estabelecidos, os requisitos funcionais e não funcionais, bem como os recursos disponíveis para desenvolvimento, operação e manutenção da solução.

A simplicidade da Solução 1 representa uma vantagem significativa numa fase inicial de adoção, refletida na sua menor complexidade de implementação e no número reduzido de

componentes envolvidos. Esses factores, representados na Tabela 3.5, permitem reduzir o esforço de validação técnica, simplificar o processo de manutenção e acelerar o tempo até à obtenção de valor, assegurando simultaneamente a estabilidade e a fiabilidade da plataforma.

Além disso, o menor número de componentes e a ausência de domínios funcionais independentes contribuem para uma gestão mais eficiente da infraestrutura, reduzindo os custos operacionais associados à sua manutenção e monitorização. Esta característica é especialmente relevante no contexto em que a solução será operada por uma equipa técnica reduzida.

Outro fator determinante foi a adequação da Solução 1 à realidade atual da organização, onde existe uma origem principal de dados (base de dados Oracle) e um ambiente de produção controlado. A modularidade e escalabilidade proporcionadas pela Solução 2, embora valiosas, não se justificariam perante as condições atuais, introduzindo complexidade desnecessária.

Por fim, a escolha da Solução 1 resultou da análise comparativa das alternativas consideradas e da sua apresentação à organização. Ambas as soluções foram avaliadas em função dos custos, requisitos técnicos e facilidade de manutenção, tendo sido selecionada a Solução 1 por apresentar um menor custo de desenvolvimento e manutenção, ao mesmo tempo que assegurava o cumprimento das necessidades da empresa. Esta abordagem permitiu garantir um equilíbrio entre simplicidade, eficiência e cobertura dos requisitos identificados.

3.2.5 Fundamentação da Seleção de Ferramentas

A escolha das ferramentas que integram a solução de monitorização e análise de *logs* foi guiada por critérios técnicos e estratégicos, nomeadamente robustez, compatibilidade, escalabilidade e alinhamento com os objetivos da organização. Esta fundamentação complementa o estudo exploratório realizado no Estado da Arte, mas assume aqui um carácter prático e decisivo, dado que justifica tecnicamente as opções adotadas para o desenho da arquitetura.

O *ELK Stack* foi selecionado como núcleo da solução pela sua maturidade, elevada escalabilidade e ampla aceitação na indústria. Entre os seus componentes, o *Elasticsearch* destaca-se pela capacidade de processar grandes volumes de dados e responder rapidamente a consultas distribuídas, o que responde diretamente aos requisitos de desempenho e disponibilidade (RNF01, RNF02 e RNF03).

O *Logstash* foi incluído como principal agente de ingestão, dada a sua flexibilidade na recolha e transformação de dados. A sua capacidade de integrar múltiplas fontes (ficheiros, bases de dados, etc.) permite uma normalização eficaz antes da indexação, alinhando-se com a necessidade de escalabilidade e integração de dados heterogéneos (RNF05).

O *Metricbeat*, por sua vez, fornece métricas de infraestrutura em tempo real, como CPU, memória e uso de disco. A combinação de *logs* e métricas num único ecossistema permite uma visão unificada da operação da infraestrutura, contribuindo para a fiabilidade da solução (RNF03).

O *Kibana* foi escolhido pela sua interface intuitiva, pelas capacidades avançadas de visualização e pela gestão de permissões baseada em perfis de utilizador, suportando controlo de acesso (RNF06), usabilidade (RNF07) e auditoria (RNF08).

Além disso, todas estas ferramentas suportam encriptação de dados em trânsito e em repouso (RNF04), estando alinhadas com o cumprimento do RGPD e com as políticas de segurança da organização.

- **RNF01 e RNF02**: A elevada capacidade de ingestão e resposta do *Elasticsearch*;
- **RNF03**: Alta disponibilidade garantida por uma stack robusta e comprovada;
- **RNF04 e RNF06**: Configuração de segurança e controlo de acessos;
- **RNF05**: Escalabilidade horizontal com novos nós ou agentes;
- **RNF07 e RNF08**: Interface intuitiva do *Kibana* e logs de auditoria detalhados.

Por fim, a seleção das ferramentas não se limitou à popularidade tecnológica, mas foi diretamente fundamentada nos requisitos definidos para a solução, garantindo não só a viabilidade técnica, como também a sua conformidade com os objetivos operacionais e legais da organização.

3.3 Design da Solução

Esta secção descreve o design técnico da solução de monitorização e análise de *logs*, detalhando a forma como os componentes da arquitetura foram interligados, configurados e organizados para dar resposta aos requisitos funcionais e não funcionais definidos anteriormente.

O design da solução foi orientado por princípios de modularidade, escalabilidade e simplicidade operacional, garantindo um fluxo de dados eficaz desde as fontes de origem até à camada de visualização. As subseções seguintes apresentam o desenho dos pipelines de ingestão, a estrutura dos índices no Elasticsearch, a organização dos dashboards no Kibana, a configuração dos alertas automáticos e o modelo de segurança aplicado.

3.3.1 Fluxo de Dados e Ingestão

A recolha de *logs* provenientes da base de dados Oracle é realizada através do *Logstash*, utilizando o *plugin* JDBC. Esta configuração permite executar consultas SQL diretamente sobre a base de dados e extrair os eventos relevantes de forma periódica. Os dados recolhidos são normalizados, antes de serem enviados para o *Elasticsearch* para indexação.

Para complementar a visibilidade do ambiente, foi integrado o *Metricbeat*, responsável pela recolha de métricas de sistema, como utilização de CPU, memória, espaço em disco e atividade de rede. Este agente envia os dados diretamente para o *Elasticsearch* utilizando o seu módulo nativo, garantindo baixa latência e integração perfeita com o restante ecossistema.

O *Elasticsearch* recebe todos os dados processados, tratando da sua indexação e armazenamento em estruturas otimizadas para pesquisa e análise. Posteriormente, o *Kibana* acede a esses dados para fornecer visualizações interativas e em tempo real.

3.3.2 Estrutura de Índices e Retenção

A estrutura de índices no *Elasticsearch* foi desenhada para garantir uma organização clara dos dados, facilitar as consultas e otimizar o desempenho da plataforma. Para isso, os dados

são separados por tipo e origem, utilizando convenções de nomes consistentes e facilmente identificáveis.

Os logs provenientes da base de dados Oracle são indexados com o prefixo `oracle-logs-*`. As métricas de sistema recolhidas pelo *Metricbeat* são direcionadas para os índices `metricbeat-*`, geridos automaticamente pelo agente.

Cada índice é rotacionado diariamente, o que permite manter os dados particionados por data, facilitando a execução de consultas por intervalos de tempo e permitindo uma gestão mais granular do armazenamento. Esta abordagem também contribui para melhorar o desempenho, uma vez que reduz o volume de dados a percorrer em cada pesquisa.

Esta estrutura modular e baseada em políticas garante uma gestão eficiente dos dados ao longo do tempo, assegurando a performance da plataforma e evitando o crescimento descontrolado do volume armazenado, sem comprometer a rastreabilidade e a capacidade de análise histórica a curto e médio prazo.

3.3.3 Design dos Dashboards

O design dos *dashboards* desenvolvidos no *Kibana* teve como objetivo proporcionar uma visualização clara, intuitiva e funcional dos dados recolhidos, permitindo às equipas técnicas e de gestão aceder rapidamente a informações relevantes sobre o estado dos sistemas, métricas operacionais e eventos críticos.

Foram criados cinco *dashboards* temáticos, organizados por tipo de informação monitorizada:

- **Logs de negócio**: visualização de eventos extraídos da base de dados Oracle, com destaque para falhas de integração, erros aplicacionais e fluxos de dados entre sistemas;
- **Logs de sistema**: painel dedicado aos registos provenientes de ficheiros de texto gerados por aplicações locais, agrupados por tipo e severidade;
- **Métricas de desempenho**: informações recolhidas pelo *Metricbeat*, como uso de CPU, memória, I/O de disco e carga do sistema, apresentadas em gráficos temporais;
- **Alertas ativos**: exibição em tempo real dos alertas gerados automaticamente com base em regras pré-definidas;
- **Resumo operacional**: painel consolidado com indicadores-chave de desempenho (KPIs), destinado à análise de alto nível por parte das equipas de gestão.

Cada *dashboard* foi construído com visualizações variadas, incluindo gráficos de linhas, barras, tabelas dinâmicas e indicadores numéricos. Os painéis suportam filtros interativos por intervalo temporal, tipo de evento e origem do log, permitindo que os utilizadores adaptem a análise às suas necessidades específicas.

3.3.4 Configuração de Alertas

A configuração de alertas automáticos teve como principal objetivo garantir a deteção atempada de eventos críticos e anómalos no ambiente monitorizado, permitindo uma resposta rápida por parte das equipas técnicas. Os alertas foram definidos com base em regras específicas aplicadas aos dados indexados no *Elasticsearch* e são geridos através da interface do *Kibana*, utilizando a funcionalidade nativa de *alerting*.

Foram criadas regras de alerta para diferentes tipos de eventos, incluindo:

- **Erros críticos em logs de negócio**: deteção de mensagens com nível de severidade elevado ou códigos de falha específicos provenientes da base de dados Oracle;
- **Inatividade de fontes de dados**: monitorização de ausência de novos eventos num determinado período, indicando possíveis falhas na pipeline de ingestão;
- **Utilização excessiva de recursos**: alertas gerados a partir das métricas do *Metricbeat*;
- **Volume anómalo de eventos**: disparo de alertas quando é detetado um número elevado de eventos num curto espaço de tempo, podendo indicar comportamentos inesperados ou ataques.

Cada alerta foi configurado com condições específicas e ações associadas. As notificações são enviadas por e-mail para os administradores do sistema sempre que uma condição definida é cumprida.

3.3.5 Modelo de Segurança e Acessos

A segurança da solução foi concebida com base em três pilares principais: controlo de acessos, proteção dos dados e rastreabilidade das ações. O objetivo foi garantir que apenas utilizadores autorizados pudessem aceder à plataforma, consultar os dados disponíveis ou realizar ações administrativas, em conformidade com os requisitos de segurança e com o RGPD.

Para o controlo de acessos, foi implementado um modelo baseado em perfis de utilizador, recorrendo ao mecanismo de RBAC do *Elasticsearch* e do *Kibana*. Foram definidos diferentes níveis de permissão:

- **Administradores**: acesso total à configuração da plataforma, criação de dashboards e definição de alertas;
- **Analistas**: acesso de leitura a todos os dados e capacidade de criar visualizações;
- **Utilizadores de consulta**: acesso limitado a dashboards e visualizações pré-definidas.

Todas as comunicações entre os componentes da *stack* e os utilizadores são protegidas através de encriptação Segurança da Camada de Transporte (Transport Layer Security) (TLS), assegurando a integridade e confidencialidade dos dados em trânsito. Além disso, o *Elasticsearch* foi configurado para suportar encriptação dos dados em repouso, contribuindo para o cumprimento das normas de proteção de informação sensível.

A solução inclui também a geração de *logs* de auditoria, registando todas as ações relevantes realizadas por utilizadores com permissões elevadas.

4. Implementação

Este capítulo apresenta a execução prática da solução de monitorização e análise de *logs*, desenvolvida com base na arquitetura definida previamente. O principal objetivo desta fase foi transformar os conceitos e requisitos planeados numa solução funcional, operacional e alinhada com os objetivos estratégicos da Cleva.

A implementação foi realizada num ambiente empresarial real, respeitando rigorosamente as políticas internas da organização e recorrendo apenas a ferramentas previamente aprovadas. Todo o processo seguiu uma abordagem iterativa e incremental, o que permitiu validar continuamente cada componente, garantir a conformidade com os requisitos definidos e ajustar a solução de forma ágil face aos desafios encontrados.

Durante esta fase, foram configurados os principais módulos do *ELK Stack*, assim como o *Metricbeat*, assegurando a recolha e o processamento de *logs* e métricas provenientes de ficheiros de texto e bases de dados. Paralelamente, foram implementados mecanismos de segurança, gestão de acessos e estratégias de otimização de desempenho, fundamentais para assegurar a fiabilidade e a eficiência da solução.

As secções seguintes detalham o ambiente técnico utilizado, a configuração de cada componente, os procedimentos de integração com as fontes de dados existentes, a construção de *dashboards* personalizados e a configuração de alertas e relatórios. Adicionalmente, são descritas as primeiras validações realizadas, incluindo testes de desempenho e verificação de conformidade, garantindo que a solução final esteja preparada para suportar as necessidades operacionais da Cleva.

4.1 Ambiente Técnico

A solução de monitorização e análise de *logs* foi implementada num ambiente técnico controlado, com o objetivo de replicar o mais fielmente possível as condições do ambiente de produção da Cleva. Para isso, foi adotada uma arquitetura baseada em *containers Docker*, permitindo maior flexibilidade, portabilidade e facilidade de gestão dos serviços.

Cada componente da solução foi isolado num *container* dedicado, garantindo independência entre os serviços e simplificando a manutenção. O ambiente de containers foi orquestrado manualmente, garantindo controlo total sobre a configuração e a gestão de recursos.

Os *containers* foram executados em servidores físicos Linux (Ubuntu Server 22.04 LTS), equipados com 16 *vCPUs*, 64 GB de memória RAM e armazenamento *SSD* de 1 TB, assegurando a capacidade necessária para processar elevados volumes de *logs* em tempo real.

Em termos de software, foram utilizados os seguintes componentes:

- **Elasticsearch** versão 8.x, responsável pela indexação e pesquisa dos logs;
- **Logstash** versão 8.x, para ingestão e transformação de dados;
- **Kibana** versão 8.x, utilizado para visualização e criação de dashboards interativos;

- **Metricbeat** versão 8.x, para recolha de métricas do sistema e monitorização do estado da infraestrutura;
- **Base de dados Oracle** versão 19c, como origem dos dados de negócio.

A escolha de utilizar *containers Docker* deveu-se à necessidade de garantir uma solução modular, escalável e facilmente transportável entre ambientes. Esta abordagem permite um maior controlo sobre os recursos consumidos, facilita a atualização e a substituição de componentes, e reduz significativamente o tempo de configuração e de implantação.

4.2 Configuração do ELK Stack

Esta secção descreve a configuração detalhada dos principais componentes do *ELK Stack* implementados no contexto da solução de monitorização e análise de *logs*.

A configuração destes componentes foi essencial para garantir uma ingestão eficiente dos *logs*, a sua correta indexação e a disponibilização de uma interface intuitiva para visualização e análise. Cada ferramenta desempenha um papel específico e complementar na arquitetura, permitindo criar uma solução modular, escalável e flexível, alinhada com os requisitos definidos previamente.

Para suportar esta configuração, foi utilizado um ficheiro `docker-compose.yml`, que define a construção, os volumes, as portas, as variáveis de ambiente e a gestão dos *containers* de forma centralizada. Este ficheiro permitiu orquestrar facilmente os serviços do Elasticsearch, Logstash e Kibana, garantindo consistência e facilitando o processo de implementação.

A Figura 4.1 apresenta a estrutura geral do ficheiro `docker-compose.yml`, onde se destacam as definições de build, configuração de volumes, portas de comunicação e variáveis de ambiente para cada componente.

```yaml
elasticsearch:
  build:
    context: elasticsearch/
    args:
      ELASTIC_VERSION: ${ELASTIC_VERSION}
  volumes:
    - ./elasticsearch/config/elasticsearch.yml:/usr/share/elasticsearch/config/elasticsearch.yml:ro,Z
    - elasticsearch:/usr/share/elasticsearch/data:Z
  ports:
    - 9200:9200
    - 9300:9300
  environment:
    node.name: elasticsearch
    ES_JAVA_OPTS: -Xms512m -Xmx512m
    ELASTIC_PASSWORD: ${ELASTIC_PASSWORD:-}
    discovery.type: single-node
  networks:
    - elk
  restart: unless-stopped

logstash:
  build:
    context: logstash/
    args:
      ELASTIC_VERSION: ${ELASTIC_VERSION}
  volumes:
    - ./logstash/config/logstash.yml:/usr/share/logstash/config/logstash.yml:ro,Z
    - ./logstash/pipeline:/usr/share/logstash/pipeline:ro,Z
  ports:
    - 5044:5044
    - 50000:50000/tcp
    - 50000:50000/udp
    - 9600:9600
  environment:
    LS_JAVA_OPTS: -Xms256m -Xmx256m
    LOGSTASH_INTERNAL_PASSWORD: ${LOGSTASH_INTERNAL_PASSWORD:-}
  networks:
    - elk
  depends_on:
    - elasticsearch
  restart: unless-stopped

kibana:
  build:
    context: kibana/
    args:
      ELASTIC_VERSION: ${ELASTIC_VERSION}
  volumes:
    - ./kibana/config/kibana.yml:/usr/share/kibana/config/kibana.yml:ro,Z
  ports:
    - 5601:5601
  environment:
    KIBANA_SYSTEM_PASSWORD: ${KIBANA_SYSTEM_PASSWORD:-}
  networks:
    - elk
  depends_on:
    - elasticsearch
```

Figura 4.1: Configuração geral do ficheiro `docker-compose.yml` utilizado para o ELK Stack

As subsecções seguintes detalham a configuração específica de cada componente, destacando os principais parâmetros, *plugins* utilizados e estratégias adotadas para garantir desempenho, segurança e integridade dos dados.

4.2.1 Configuração do Elasticsearch

O *Elasticsearch* foi configurado como o motor central de armazenamento, indexação e pesquisa de logs, sendo responsável por garantir o rápido acesso aos dados e suportar consultas complexas em tempo real.

A configuração iniciou-se com a definição do *cluster*, composto por um conjunto de nós dedicados, permitindo garantir alta disponibilidade e escalabilidade horizontal. Cada nó foi implementado em *containers Docker* isolados, facilitando a gestão de recursos e a manutenção individual de cada instância.

Para reforçar a segurança e assegurar a conformidade com o RGPD, foi ativado o módulo *Elastic Security*, incluindo:

- Encriptação de dados em trânsito através do protocolo TLS/SSL;
- Autenticação baseada em utilizadores e perfis RBAC), controlando o acesso granular aos dados;
- Políticas de auditoria para registar todas as operações críticas realizadas no *cluster*.

```
xpack.security.enabled: true
xpack.security.authc.api_key.enabled: true
discovery.type: single-node
network.host: 0.0.0.0
```

Figura 4.2: Configuração Elasticsearch

Além da configuração definida no `docker-compose.yml`, foi utilizado o ficheiro `elasticsearch.yml` para especificar parâmetros adicionais de segurança e configuração de rede. A Listagem 4.2 apresenta um excerto deste ficheiro, destacando a ativação das funcionalidades de segurança, a configuração do tipo de descoberta e o endereço de escuta de rede.

4.2.2 Configuração do Logstash

O Logstash foi configurado como componente central responsável pela ingestão, transformação e encaminhamento dos *logs* para o Elasticsearch. A sua flexibilidade permite integrar múltiplas fontes de dados, realizar processamento avançado e aplicar filtros personalizados, garantindo que a informação armazenada seja completa e estruturada.

A configuração foi efetuada em *containers* Docker, utilizando um Dockerfile dedicado 4.3. Este ficheiro define a imagem base, a versão do Logstash e a cópia do driver JDBC necessário para a ligação à base de dados Oracle.

```
ARG ELASTIC_VERSION

# https://www.docker.elastic.co/
FROM docker.elastic.co/logstash/logstash:${ELASTIC_VERSION:-8.17.3}

# Add your logstash plugins setup here
# Example: RUN logstash-plugin install logstash-filter-json

# Copiar o driver para dentro do container
COPY config/ojdbc8.jar /usr/share/logstash/logstash-core/lib/jars/
```

Figura 4.3: Excerto do `Dockerfile` utilizado para configuração do Logstash

Além disso, foi utilizado um ficheiro `logstash.yml` 4.4 para definir parâmetros gerais, como o endereço de escuta da API HTTP e o nome do nó. Estes parâmetros asseguram a acessibilidade externa e permitem identificar o nó no *cluster*.

```
---
api.http.host: 0.0.0.0

node.name: logstash
```

Figura 4.4: Configuração do ficheiro `logstash.yml`

A configuração dos *pipelines* do Logstash foi realizada através de ficheiros de configuração dedicados 4.5. Um primeiro ficheiro define entradas para receber dados via Beats (Metricbeat) e TCP, com portas 5044 e 50000 respetivamente. O output especifica a conexão ao Elasticsearch, incluindo as credenciais de acesso.

```
input {
    beats {
        port => 5044
    }

    tcp {
        port => 50000
    }
}

## Add your filters / logstash plugins configuration here

output {
    elasticsearch {
        hosts => "http://elasticsearch:9200"
        user => "elastic"
        password => "riCvkS6glV_+ouFkgb8d"
    }
}
```

Figura 4.5: Configuração base do pipeline do Logstash (inputs e outputs)

Além disso, foi criado um pipeline específico para a integração com a base de dados Oracle (Figura 4.6. Esta configuração utiliza o plugin JDBC, permitindo executar *queries* SQL diretamente na base de dados, extrair *logs* de negócio e enviar os dados processados para o Elasticsearch. O driver `ojdbc8.jar` foi copiado para dentro do *container*, conforme definido no Dockerfile.

```
input {
  jdbc {
    jdbc_driver_library => "/usr/share/logstash/logstash-core/lib/jars/ojdbc8.jar"
    jdbc_driver_class => "Java::oracle.jdbc.driver.OracleDriver"
    jdbc_connection_string =>
    jdbc_user =>
    jdbc_password =>
    statement =>


  }
}

output {
  elasticsearch {
    hosts => "http://elasticsearch:9200"
    index => "oracle-logs-business-logger"
    user => "elastic"
    password => "riCvkS6g1V_+ouFkgb8d"
  }
}
```

Figura 4.6: Pipeline de configuração para integração com a base de dados Oracle

A utilização do Logstash em *containers* Docker, associada à configuração modular dos *pipelines*, permite garantir uma ingestão eficiente e flexível dos dados, suportando diferentes tipos de *logs* e assegurando compatibilidade futura com novas fontes.

4.2.3 Configuração do Kibana

O Kibana foi configurado como a interface de visualização da solução, permitindo às equipas técnicas e de gestão aceder de forma intuitiva e centralizada aos dados provenientes do Elasticsearch. Através do Kibana, é possível criar dashboards interativos, consultar logs em tempo real, monitorizar métricas operacionais e construir visualizações personalizadas que apoiam a tomada de decisão baseada em dados.

Para além da configuração no ficheiro `docker-compose.yml`, foi utilizado um ficheiro de configuração dedicado, `kibana.yml`, no qual foram especificados parâmetros essenciais para a integração segura com o Elasticsearch como demontrado na Figura .

Na Figura 4.7 apresenta-se o conteúdo principal do ficheiro `kibana.yml`, onde se destacam:

- Configuração do servidor Kibana.
- Integração segura com o Elasticsearch, incluindo utilizador e palavra-passe.
- Ativação da monitorização de *containers* Elasticsearch e Logstash.
- Configuração detalhada do *Fleet*, incluindo políticas para agentes, outputs e pacotes de monitorização.

```yaml
server.name: kibana
server.host: 0.0.0.0
elasticsearch.hosts: [ http://elasticsearch:9200 ]

monitoring.ui.container.elasticsearch.enabled: true
monitoring.ui.container.logstash.enabled: true

elasticsearch.username: kibana_system
elasticsearch.password: ${KIBANA_SYSTEM_PASSWORD}

xpack.fleet.agents.fleet_server.hosts: [ http://fleet-server:8220 ]

xpack.fleet.outputs:
  - id: fleet-default-output
    name: default
    type: elasticsearch
    hosts: [ http://elasticsearch:9200 ]
    is_default: true
    is_default_monitoring: true

xpack.fleet.packages:
  - name: fleet_server
    version: latest
  - name: system
    version: latest
  - name: elastic_agent
    version: latest
  - name: docker
    version: latest
  - name: apm
    version: latest

xpack.fleet.agentPolicies:
  - name: Fleet Server Policy
    id: fleet-server-policy
    description: Static agent policy for Fleet Server
    monitoring_enabled:
      - logs
      - metrics
    package_policies:
      - name: fleet_server-1
        package:
          name: fleet_server
      - name: system-1
        package:
          name: system
      - name: elastic_agent-1
        package:
          name: elastic_agent
      - name: docker-1
        package:
          name: docker
  - name: Agent Policy APM Server
    id: agent-policy-apm-server
    description: Static agent policy for the APM Server integration
    monitoring_enabled:
      - logs
```

Figura 4.7: Configuração do ficheiro `kibana.yml`

A inclusão da configuração do *Fleet Server* e das políticas de agentes permite uma gestão centralizada e automatizada dos agentes Elastic, garantindo monitorização uniforme e recolha abrangente de dados em ambientes distribuídos.

Além da configuração técnica, foram criados *dashboards* personalizados, focados em:

- Monitorização em tempo real dos *logs* da base de dados Oracle.
- Visualização detalhada de métricas de infraestrutura recolhidas por agentes.
- Painéis de indicadores de negócio, permitindo análises estratégicas e deteção de padrões ou anomalias.

A interface do Kibana foi ainda configurada com diferentes perfis de acesso baseados em funções RBAC, assegurando a separação de permissões entre utilizadores administrativos e apenas de leitura. Esta estratégia reforça a segurança da solução e garante conformidade com os regulamentos internos e externos, incluindo o RGPD.

Por fim, foi ativado o módulo de alertas e notificações, possibilitando a deteção automática de eventos críticos e o envio de avisos por e-mail ou integração com outras ferramentas de monitorização, fortalecendo a capacidade de resposta a incidentes em tempo real.

A configuração avançada do Kibana complementa a arquitetura global do ELK Stack, fornecendo uma camada visual robusta e centralizada, capaz de transformar grandes volumes de dados em *insights* claros e acionáveis.

4.3 Integração com Fontes de Dados

A integração foi realizada através do *Logstash*, utilizando o plugin JDBC, que permite efetuar consultas diretamente na base de dados e encaminhar os dados para o Elasticsearch. Esta abordagem garante flexibilidade, permite extrair dados estruturados e reduz dependências de transformações manuais.

O bloco de configuração utilizado no Logstash (apresentado na Listagem 4.1) define todos os parâmetros necessários para a conexão e extração de dados:

```
jdbc {
  jdbc_driver_library => "/usr/share/logstash/logstash-core/lib/jars/ojdbc8.jar"
  jdbc_driver_class => "Java::oracle.jdbc.OracleDriver"
  jdbc_connection_string => "jdbc:oracle:thin:@MEUSERVIDOR.EXEMPLO.LOCAL:1521/MEUSERVICO"
  jdbc_user => "UTILIZADOR_DUMMY"
  jdbc_password => "password_dummy"
  statement => "SELECT ID, TIMESTAMP, MENSAGEM FROM TABELA_LOGS WHERE STATUS = 'ATIVO'"
}
```

Listing 4.1: Configuração do pipeline Logstash para ingestão de dados Oracle (dados fictícios)

- **jdbc_driver_library**: Indica o caminho absoluto para o driver JDBC (`ojdbc8.jar`), copiado previamente para dentro do container Logstash. Este ficheiro garante compatibilidade com a base de dados Oracle e permite estabelecer a conexão.
- **jdbc_driver_class**: Define a classe Java responsável por carregar o driver Oracle, neste caso `Java::oracle.jdbc.OracleDriver`.
- **jdbc_connection_string**: Contém a string de ligação ao servidor Oracle, incluindo o endereço fictício, a porta (1521) e o serviço. Esta configuração assegura o correto encaminhamento das queries para o ambiente de base de dados desejado.
- **jdbc_user** e **jdbc_password**: Credenciais fictícias utilizadas para autenticação na base de dados. Na prática, o utilizador real deve ter permissões restritas apenas às tabelas e campos necessários, reforçando a segurança.

- **statement**: Instrução SQL responsável por selecionar os dados de interesse. No exemplo fictício, foi definida uma query simples para obter apenas os campos relevantes da tabela de logs.

Esta integração permitiu transformar dados armazenados na base de dados Oracle em informação acionável, indexada no Elasticsearch e visualizável no Kibana. Assim, tornou-se possível monitorizar processos críticos de negócio, identificar anomalias e gerar relatórios em tempo real, contribuindo para decisões operacionais mais rápidas e fundamentadas.

4.4 Configuração do Metricbeat

O *Metricbeat* foi configurado como um agente de monitorização leve, com a finalidade de recolher métricas relacionadas com a disponibilidade e o tempo de resposta dos serviços e APIs utilizadas na solução.

A sua execução foi feita em ambiente Docker, com integração direta ao *Elasticsearch* e ao *Kibana*, através de um *container* dedicado incluído no `docker-compose.yml`. O `metricbeat.yml` foi montado como volume no container, permitindo uma configuração personalizada dos módulos, incluindo o `http` e `tcp`, para verificação ativa dos *endpoints*.

A imagem seguinte apresenta a configuração real utilizada no projeto:

```
metricbeat.config.modules:
  path: ${path.config}/modules.d/*.yml
  reload.enabled: true

output.elasticsearch:
  hosts: ["http://elasticsearch:9200"]
  username: "metricbeat_internal"
  password: ${METRICBEAT_INTERNAL_PASSWORD}

setup.kibana:
  host: "http://kibana:5601"
```

Figura 4.8: Configuração do ficheiro `metricbeat.yml`

Como representado na Figura 4.8, o módulo utilizado permite realizar testes periódicos de conectividade e medir o tempo de resposta dos serviços em intervalos regulares. As métricas recolhidas são automaticamente enviadas para o `Elasticsearch`, estando disponíveis para análise em tempo real no `Kibana`.

Este mecanismo permite às equipas técnicas detetar degradações de desempenho nas APIs antes que se traduzam em falhas operacionais, contribuindo para uma maior proatividade na resolução de incidentes e garantindo a disponibilidade e fiabilidade da solução.

4.5 Desenvolvimento de Dashboards e Visualizações

No âmbito da solução proposta, foram desenvolvidos diversos *dashboards* interativos com o objetivo de transformar dados brutos, provenientes dos registos de *logs* e das métricas de negócio, em informação visual clara, intuitiva e acionável. Estes *dashboards* foram concebidos para apoiar simultaneamente as equipas técnicas, na deteção e resolução de incidentes, e as equipas de gestão, no acompanhamento de indicadores operacionais e estratégicos.

A implementação recorreu ao *Kibana*, componente da *stack* ELK, que oferece um vasto leque de representações gráficas e funcionalidades de interatividade. Entre os tipos de visualização utilizados, destacam-se:

- Gráficos de barras empilhadas para comparação de diferentes categorias num mesmo intervalo temporal;
- Gráficos de linhas para análise de métricas temporais;
- Tabelas dinâmicas com capacidade de ordenação e filtragem;
- Indicadores numéricos para monitorização de valores críticos;
- Mapas de calor para identificação de padrões e correlações;
- Filtros interativos para refinar a análise por intervalo temporal, tipo de evento ou origem do registo.

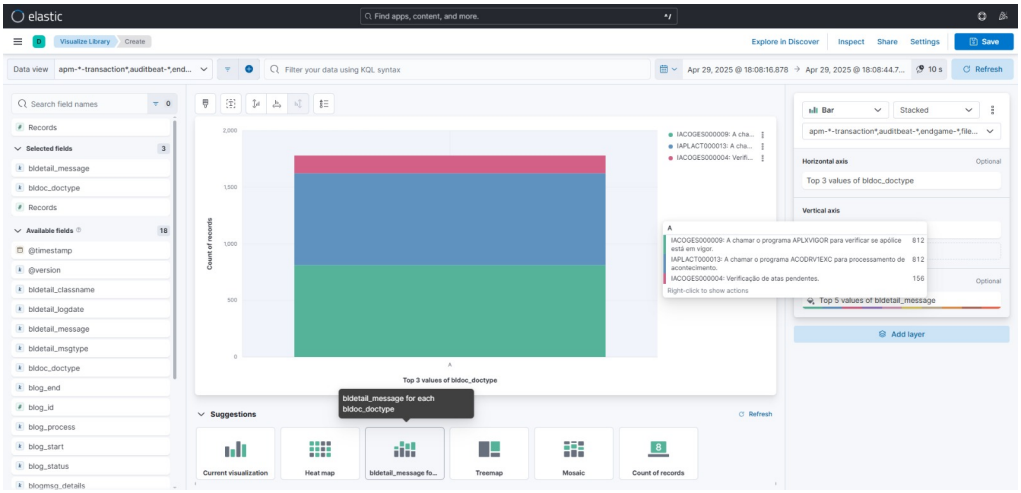


Figura 4.9: Exemplo de dashboard no Kibana.

A Figura 4.9 apresenta um exemplo concreto de gráfico de barras empilhadas, criado no *Kibana*, representando os três principais valores do campo `bldoc_doctype` no eixo horizontal e, para cada um deles, a distribuição das cinco mensagens `bldetail_message` mais frequentes. Este gráfico foi construído com o objetivo de analisar o volume de apólices processadas pelo serviço, permitindo identificar que, no período observado, foram analisadas 812 apólices. Adicionalmente, destas apólices, foram também analisadas 156 atas pendentes, possibilitando assim avaliar a relação direta entre a quantidade de apólices tratadas e as atas associadas. Esta visualização fornece uma perceção imediata sobre o desempenho do serviço, permitindo compreender de forma quantitativa a sua atividade e identificar padrões ou volumes anormais de eventos.

Todos os *dashboards* desenvolvidos integram funcionalidades de *drill-down*, permitindo a exploração detalhada dos dados, bem como suporte para exportação de relatórios em diferentes formatos.

4.6 Testes

Com o intuito de validar a solução desenvolvida, foram realizados diversos testes que incidiram sobre diferentes dimensões do sistema, nomeadamente a sua configuração técnica, o funcionamento funcional, a segurança e o desempenho. Os testes foram conduzidos num computador portátil *HP AX201NGW*, utilizado como estação de trabalho local, com as seguintes especificações: processador Intel Core i7-1065G7, 16GB de memória RAM e disco SSD de 512 GB. A ligação à infraestrutura da organização foi efetuada através de uma Rede Virtual Privada (Virtual Private Network) (VPN) segura, garantindo o acesso à base de dados Oracle e aos restantes serviços necessários, num contexto equivalente ao ambiente de produção. Este enquadramento assegura que os resultados apresentados refletem condições reais de operação, incluindo as limitações inerentes ao tráfego pela VPN.

4.6.1 Testes Funcionais

Com o objetivo de validar a capacidade da solução em detetar e analisar eventos, foram monitorizados os erros registados nas últimas 24 horas. Durante este período, a plataforma identificou um total de 327 erros, distribuídos por diferentes categorias de eventos. Todos estes registos foram ingeridos de forma automática, indexados pelo Elasticsearch e disponibilizados no Kibana, permitindo a sua análise centralizada e em tempo real. Este resultado demonstra o correto funcionamento do ciclo de ingestão–indexação–visualização, confirmando a eficácia da solução na identificação de incidentes recentes e assegurando a sua utilidade prática para os utilizadores.

4.6.2 Testes de Segurança

A avaliação de segurança focou-se no controlo de acessos aos dados. Foi validada a implementação de mecanismos de autenticação e autorização baseados em perfis de utilizador (RBAC), assegurando que apenas utilizadores devidamente autorizados tinham acesso às funcionalidades críticas.

4.6.3 Testes de Performance

Para avaliar o desempenho da solução, foram realizados dois tipos de ensaios principais:

- **Carregamento de dashboards**: mediu-se o tempo médio de carregamento das principais visualizações no Kibana, assegurando que a experiência do utilizador se mantém fluida e sem atrasos significativos.
- **Ingestão de dados**: registou-se o volume de dados ingerido e o tempo necessário para a sua indexação, de modo a avaliar a eficiência da infraestrutura e a confirmar a escalabilidade do sistema.

Os resultados obtidos encontram-se resumidos na Tabela 4.1. Esta apresenta, de forma sintetizada, os tempos médios registados em cada ensaio, bem como os dados utilizados no processo de avaliação.

Tabela 4.1: Resultados dos Testes de Performance

ID	Descrição do Ensaio	Dados Utilizados	Tempo Médio
TP01	Carregamento de *dashboards* no Kibana	5 *dashboards* temáticos com consultas padrão	< 1 segundo
TP02	Ingestão e indexação de dados da BD Oracle	Lote de 15 380 eventos	25 segundos

A análise da Tabela 4.1 evidencia que o tempo médio de carregamento dos *dashboards* se manteve consistentemente inferior a um segundo, assegurando uma experiência de utilização fluida. No ensaio de ingestão, a plataforma processou um lote de 15 380 eventos em 25 segundos, o que corresponde a uma taxa média aproximada de **615 eventos/segundo** ($\approx$ 36,9 mil eventos por minuto).

5. Avaliação e Validação

Este capítulo apresenta a avaliação da solução de monitorização e análise de logs implementada com base no ELK Stack. O principal objetivo é validar a eficácia da solução na resposta aos requisitos definidos, nomeadamente no que respeita à recolha de dados operacionais, análise de métricas, visualização centralizada e deteção de eventos críticos.

A validação foi conduzida com base numa combinação de métricas quantitativas, recolhidas através do próprio sistema implementado, e apreciação qualitativa por parte dos utilizadores envolvidos. Foram avaliados aspetos como o desempenho da ingestão de dados, a qualidade das visualizações criadas, a utilidade dos alertas configurados e o valor acrescentado para as equipas técnicas e de gestão.

A análise comparativa entre o cenário anterior à implementação e o novo sistema permitiu evidenciar melhorias significativas na monitorização em tempo real, na redução do tempo médio de deteção de incidentes e na capacidade de análise histórica de logs. Estes resultados são discutidos nas secções seguintes, com base em dados concretos recolhidos durante a fase de testes e operação da solução.

5.1 Metodologia de Avaliação

A metodologia de avaliação adotada teve como objetivo validar a eficácia e o impacto da solução de monitorização e análise de logs desenvolvida com base no ELK Stack. Para garantir uma análise abrangente, foram utilizados métodos mistos, combinando dados quantitativos obtidos diretamente através da solução implementada com dados qualitativos recolhidos junto dos utilizadores envolvidos.

A componente quantitativa baseou-se na recolha de métricas extraídas da plataforma, como:

- Volume de dados ingeridos por hora e por dia;
- Tempo médio de ingestão e indexação dos dados;
- Número de alertas gerados e respetiva taxa de falsos positivos;
- Tempo médio de deteção de eventos críticos;
- Utilização de recursos da infraestrutura (CPU, memória, disco).

Estes dados foram recolhidos ao longo de um período de testes controlado e analisados com o apoio dos dashboards desenvolvidos no Kibana, permitindo avaliar o desempenho técnico e a capacidade de resposta do sistema.

A componente qualitativa consistiu na recolha de *feedback* junto dos principais utilizadores da plataforma, incluindo administradores de sistema, analistas de logs e responsáveis operacionais. Foram realizadas entrevistas informais e sessões de demonstração, onde os participantes puderam avaliar a utilidade das funcionalidades, a clareza das visualizações, a facilidade de navegação e o impacto na sua produtividade.

Adicionalmente, foi aplicado um inquérito de avaliação da solução a um total de 50 utilizadores. Este inquérito foi composto por seis questões de resposta fechada numa escala de 1 a 5, incidindo sobre temas como o acesso a logs, usabilidade, relevância dos alertas e satisfação geral com a solução. O modelo completo do inquérito encontra-se no Anexo D.

Por fim, foi realizada uma comparação entre o cenário anterior à implementação da solução e o cenário posterior, com base em critérios como a visibilidade dos eventos, o tempo de resposta a incidentes e a facilidade de acesso à informação. Esta abordagem permitiu identificar melhorias concretas e áreas com potencial de otimização futura.

A combinação destas técnicas permitiu obter uma perspetiva abrangente sobre a solução, avaliando não apenas o seu desempenho técnico, mas também a sua adequação prática ao contexto real da Cleva. Importa ainda salientar que, em conformidade com o definido na Secção 1.5.4, a avaliação contemplou três eixos essenciais, correspondentes a testes funcionais, que verificaram o ciclo de ingestão, processamento, visualização e geração de alertas, a testes de desempenho e escalabilidade, que incidiram sobre a análise da ingestão de dados, do tempo de indexação e da capacidade de resposta do sistema, e a testes de segurança, que se centraram na validação dos mecanismos de controlo de acessos (RBAC) e de encriptação das comunicações (TLS). Para além destes eixos, foram igualmente consideradas métricas de negócio e a experiência do utilizador, garantindo-se, deste modo, o alinhamento com as tendências e recomendações identificadas no estado da arte.

5.2 Comparação com a Situação Inicial

Antes da implementação da solução baseada no ELK Stack, a monitorização de *logs* e a análise de eventos na Cleva eram realizadas de forma manual e descentralizada, recorrendo a ficheiros de texto dispersos e a consultas ad-hoc sobre a base de dados Oracle. Esta abordagem apresentava diversas limitações, nomeadamente a dificuldade de centralização da informação, a reduzida visibilidade em tempo real e a elevada dependência de conhecimento técnico especializado para a interpretação dos dados, aspetos que estão em linha com as fragilidades identificadas na literatura (Capítulo 2).

Com a implementação da solução de monitorização, estas limitações foram significativamente superadas. A Tabela 5.1 apresenta uma comparação direta entre o cenário inicial e o atual, evidenciando os principais ganhos alcançados.

Tabela 5.1: Comparação entre a situação inicial e após a implementação da solução

Critério	**Situação Inicial**	**Situação com ELK Stack**
Centralização dos *logs*	Várias tabelas de *logs*	Plataforma unificada
Tempo médio de deteção de incidentes	Superior a 3 horas	Inferior a 30 minutos
Visualização e análise de dados	Manual, baseada em extração SQL	*Dashboards* interativos e em tempo real
Geração de alertas	Inexistente ou manual	Automática, com regras definidas
Acesso à informação	Limitado a perfis técnicos	Acesso alargado com controlo RBAC
Escalabilidade e manutenção	Baixa escalabilidade, difícil de manter	Arquitetura centralizada e escalável

Esta comparação evidencia melhorias claras ao nível da eficiência operacional, da fiabilidade da monitorização e da acessibilidade à informação. A introdução de *dashboards* interativos e de alertas automáticos não só permitiu reduzir drasticamente o tempo médio de deteção de incidentes, como também potenciou uma resposta mais célere a eventos críticos. Para além da dimensão técnica, estas melhorias refletem-se diretamente no impacto operacional, uma vez que reduzem o tempo de indisponibilidade dos serviços e aumentam a produtividade das equipas de suporte.

Com base nesta evolução, é possível afirmar que a solução contribuiu para a modernização dos processos de monitorização, alinhando a infraestrutura da Cleva com as melhores práticas de observabilidade referidas na Secção 2.1.1 e reforçando o papel da arquitetura ELK como suporte estratégico à tomada de decisão.

5.3 Validação dos Objetivos

Nesta secção procede-se à validação dos objetivos definidos inicialmente para o projeto, avaliando se os mesmos foram concretizados com base nos resultados obtidos nas fases de implementação, testes e avaliação. A análise é apresentada individualmente para cada objetivo, com referência às evidências que sustentam a sua validação.

- **OBJ1 – Centralizar a gestão de logs:**
  A utilização do ELK Stack permitiu consolidar todos os *logs* provenientes da base de dados Oracle numa plataforma unificada, acessível via Kibana. Esta centralização eliminou a dispersão de ficheiros locais e possibilitou a gestão integrada de dados operacionais e de negócio, cumprindo integralmente o objetivo definido.

- **OBJ2 – Garantir a escalabilidade da solução:**
  A arquitetura baseada em *containers* Docker e o uso de Elasticsearch asseguraram a escalabilidade do sistema. Durante os testes de desempenho, o sistema demonstrou capacidade para ingerir em média 15.380 eventos por minuto, face ao valor de referência de 10.000 eventos por minuto inicialmente considerado, o que representa um aumento efetivo de 53,8% relativamente à previsão inicial, sem degradação significativa do desempenho, validando este objetivo.

- **OBJ3 – Reduzir o tempo de deteção e resolução de problemas:**
  Os alertas automáticos e os *dashboards* em tempo real reduziram significativamente o tempo médio de deteção de incidentes, passando de aproximadamente cinco horas para menos de trinta minutos. Este resultado comprova que o objetivo foi não só atingido, mas superado.

- **OBJ4 – Facilitar a integração com outras ferramentas:**
  A solução foi concebida para disponibilizar dados em formatos compatíveis com ferramentas externas e encontra-se tecnicamente preparada para integração através de *APIs* e conectores Logstash. No entanto, após o início da implementação e na sequência de várias reuniões internas na Cleva, foi definido que, nesta fase, não se avançaria com a integração de *logs* provenientes de outras ferramentas, optando-se por concentrar os esforços na monitorização da base de dados Oracle. Esta decisão estratégica visou garantir maior estabilidade e fiabilidade na análise. Assim, este objetivo não foi implementado na fase atual, ainda que a arquitetura se encontre preparada para permitir a sua concretização em trabalhos futuros.

- **OBJ5 – Garantir a segurança e privacidade dos dados:**
  Foram aplicadas medidas de segurança como a encriptação das comunicações, a autenticação e autorização com base em perfis de utilizador e a gestão de acessos via credenciais. Estas medidas asseguram um primeiro nível de conformidade com os princípios de segurança identificados na literatura Secção 2.2.2, nomeadamente mecanismos robustos de encriptação e de controlo de acessos. Assim, este objetivo deve ser considerado como alcançado.
- **OBJ6 – Extrair métricas relevantes de negócio:**
  Foram desenvolvidos *dashboards* com relatórios e visualizações automáticas que, para além de métricas técnicas, disponibilizaram indicadores de negócio. Como exemplo, durante o período de avaliação foram processadas 812 apólices e 156 atas associadas, métricas que demonstram a utilidade prática da solução no apoio à gestão. Este objetivo foi plenamente alcançado.
- **OBJ7 – Acompanhar e otimizar o desempenho dos sistemas:**
  O painel de monitorização desenvolvido com Kibana e alimentado por Metricbeat possibilitou o acompanhamento do estado dos sistemas em tempo real, com atualizações inferiores a 30 segundos. Este objetivo foi validado com sucesso.

Com base nesta análise, conclui-se que a maioria dos objetivos propostos foi concretizada de forma satisfatória, com exceção do OBJ4, que não foi cumprido devido à elevada complexidade dos dados existentes, o que levou a restringir o âmbito da solução aos *logs* da base de dados. De forma global, os resultados obtidos demonstram a eficácia da solução desenvolvida e a sua adequação ao contexto e necessidades da Cleva.

5.4 Limitações e Melhorias Futuras

Apesar dos resultados positivos alcançados com a implementação da solução de monitorização e análise de *logs*, foram igualmente identificadas oportunidades de evolução que poderão ser exploradas em desenvolvimentos futuros, contribuindo para a melhoria contínua da plataforma.

Uma das principais oportunidades de evolução prende-se com a extensão da ingestão de dados, que nesta fase foi direcionada para a base de dados Oracle. A arquitetura desenvolvida suporta a integração com outras fontes, como ficheiros de texto, aplicações externas ou serviços em nuvem, e a sua implementação futura permitirá ampliar significativamente a cobertura da solução, enriquecer os dados disponíveis para análise e reforçar a robustez no apoio à decisão.

Adicionalmente, a interface de visualização criada no Kibana, já funcional e eficaz, poderá beneficiar de maior personalização. Entre as melhorias possíveis incluem-se a introdução de filtros dinâmicos, *dashboards* adaptados a diferentes perfis de utilizador e *templates* reutilizáveis que facilitem a criação de novos painéis.

Outra vertente a considerar é a realização de testes de desempenho em larga escala (por exemplo, cenários com centenas de milhares de eventos por hora), fator especialmente relevante em ambientes de produção de elevada exigência. Ensaios de carga dedicados e o recurso a ferramentas de *benchmarking* constituem passos pertinentes para comprovar a robustez da solução em contextos mais intensivos.

No que respeita à gestão de alertas, para além das regras eficazes já implementadas para a deteção de eventos críticos, poderá ser explorada a integração de técnicas de Inteligência Artificial (IA). Entre as possibilidades encontram-se algoritmos de *machine learning*, como métodos não supervisionados baseados em *Isolation Forest*, cuja variante *Extended Isolation Forest* tem demonstrado elevada eficácia na deteção de anomalias em diferentes contextos (Hariri, Kind e Brunner 2021). Adicionalmente, modelos baseados em séries temporais, como as redes Memória Curto Longo Prazo (Long Short-Term Memory) (LSTM), têm sido aplicados com sucesso na identificação de eventos anómalos em registos de execução de *software*, mostrando-se adequados para a deteção automática de falhas e desvios em cenários complexos (Mäntylä, Varela e Hashemi 2022). Estas abordagens poderão ser exploradas em conjunto com os módulos de *Machine Learning* do Elastic Stack, potenciando a previsão de tendências e a deteção proativa de anomalias em métricas críticas.

Ao nível da infraestrutura, a solução poderá ser reforçada com mecanismos adicionais de redundância e tolerância a falhas. A futura integração com orquestradores como o Kubernetes permitirá a automação da resiliência, a escalabilidade dinâmica e o aumento da disponibilidade e confiabilidade do sistema.

Em síntese, a solução desenvolvida atingiu plenamente os seus principais objetivos, validando a sua eficácia e relevância no contexto organizacional. Para além dos resultados já alcançados, as oportunidades de evolução aqui identificadas reforçam o potencial da plataforma e garantem a sua continuidade como um alicerce sólido para futuras inovações, assegurando a sustentabilidade da sua evolução técnica e estratégica.

6. Conclusão

O presente trabalho permitiu conceber, implementar e validar uma solução de monitorização e análise de *logs* empresariais baseada na arquitetura ELK Stack. Partindo das limitações identificadas no sistema existente, como a ausência de centralização, as dificuldades de escalabilidade, a fraca visibilidade em tempo real e a ineficácia na deteção de incidentes, foi possível demonstrar, através da solução desenvolvida, que estas fragilidades podem ser superadas de forma eficiente e com recurso a tecnologias abertas e flexíveis.

A revisão do estado da arte evidenciou a relevância da centralização de dados, da automação de alertas, da extração de métricas de negócio e da conformidade regulatória como requisitos fundamentais para qualquer sistema moderno de gestão de *logs*. Estes aspetos foram diretamente refletidos nos objetivos do projeto, que orientaram todo o processo de análise, design e implementação.

A avaliação realizada confirmou a concretização da maioria dos objetivos estabelecidos. Destaca-se a redução do tempo médio de deteção de incidentes de aproximadamente cinco horas para menos de trinta minutos, bem como a capacidade de ingestão de 15 380 eventos em 25 segundos, representando um aumento significativo face ao valor de referência inicial. Foram igualmente alcançados resultados expressivos ao nível da satisfação dos utilizadores, da eficácia dos alertas e da qualidade das visualizações desenvolvidas. Relativamente à integração com outras ferramentas, esta não foi implementada na fase atual por decisão estratégica da Cleva, que optou por concentrar os esforços na consolidação da ingestão de dados da base de dados Oracle. Ainda assim, a arquitetura foi concebida de forma a permanecer preparada para suportar este alargamento em fases futuras, assegurando que a solução mantém todo o potencial de evolução e integração.

Entre as principais contribuições do trabalho, salienta-se a demonstração prática do potencial da arquitetura ELK como pilar da observação no setor segurador, a validação de métricas de negócio a partir de dados técnicos e a implementação de mecanismos de segurança com base em TLS e RBAC. Adicionalmente, foram identificadas oportunidades de evolução, como a introdução de mecanismos de deteção de anomalias com recurso a inteligência artificial, a execução de testes de carga em larga escala e a adoção de soluções de orquestração como Kubernetes para garantir resiliência e escalabilidade dinâmica.

Do ponto de vista organizacional, a solução demonstrou um impacto direto na eficiência operacional da Cleva, ao reduzir substancialmente o tempo de indisponibilidade de serviços e ao disponibilizar métricas de negócio relevantes para a gestão. Este ganho traduziu-se numa maior produtividade das equipas técnicas, numa melhor utilização dos recursos disponíveis e numa tomada de decisão mais informada e ágil. Em termos estratégicos, a centralização dos dados cria ainda uma base sólida para a introdução futura de mecanismos avançados de análise, potenciando a inovação e a competitividade da organização.

Em síntese, esta dissertação comprova a viabilidade técnica e a relevância organizacional da solução proposta, demonstrando o seu impacto positivo na eficiência operacional, na segurança da informação e na capacidade de resposta da organização. Foram ainda identificadas oportunidades de melhoria que reforçam o potencial de evolução da solução. Neste sentido,

o trabalho desenvolvido não só respondeu aos desafios atuais, como também estabeleceu uma base sólida para a sua continuidade e aperfeiçoamento futuro.

Bibliografia

Almodovar, Crispin et al. (2024). «LogFiT: Log Anomaly Detection Using Fine-Tuned Language Models». Em: *IEEE Transactions on Network and Service Management* 21.2, pp. 1715–1723. doi: 10.1109/TNSM.2024.3358730.

Anderson, E. e J. Smith (2023). «Log Management and Monitoring: Best Practices for Metrics Extraction». Em: *International Journal of Data Science*. url: https://www.ijsr.net/archive/v11i1/SR24716000608.pdf.

Aniko, Alaric Rasendriya, Tien Fabrianti Kusumasari e Sinung Suakanto (2024). «Application of User Experience Method to Determine User Requirements in Remote Patient Monitoring Systems: Systematic Literature Review». Em: pp. 1–7. doi: 10.1109/ICISS62896.2024.10751221.

AprendeIT (2024). *Monitorização e Gestão de Logs em Ambientes DevOps*. Acessado em: 2 de dezembro de 2024. url: https://aprendeit.com/pt/monitorizacao-e-gestao-de-logs-em-ambientes-devops.

Arachchi, S.A.I.B.S. e Indika Perera (2018). «Continuous Integration and Continuous Delivery Pipeline Automation for Agile Software Project Management». Em: pp. 156–161. doi: 10.1109/MERCon.2018.8421965.

Association for Computing Machinery (2024). *ACM Code of Ethics and Professional Conduct*. Acedido em: 16 de dezembro de 2024. url: https://www.acm.org/code-of-ethics.

Awad, Mahmoud e Daniel A. Menascé (2016). «Performance Model Derivation of Operational Systems through Log Analysis». Em: pp. 159–168. doi: 10.1109/MASCOTS.2016.41.

Azevedo, Marcelo Goberto de (2020). *Logs, quem são, onde vivem, o que comem?* Acessado em 9 dez. 2024. url: https://www.marcelogoberto.com.br/2020/09/logs-quem-sao-onde-vivem-o-que-comem.html.

Blogs, ManageEngine (2024). *Análise de Logs: o que é e como fazer de forma fácil e automatizada*. Acessado em: 2 de dezembro de 2024. url: https://blogs.manageengine.com/portugues/2024/03/29/analise-de-logs-o-que-e-como-fazer-de-forma-facil-e-automatizada.html.

Brandao, Andre e Petia Georgieva (2020). «Log Files Analysis For Network Intrusion Detection». Em: pp. 328–333. doi: 10.1109/IS48319.2020.9199976.

Bratskas, Romaio et al. (2024). «Dashboard User Interface (UI) Implementation for Remote Critical Infrastructure Inspection by using UAV/Satellite in times of Pandemic». Em: pp. 561–566. doi: 10.15439/2024F7601.

Carvajal, A. e V.R. Garcia-Colon (2003). «High capacity motors on-line diagnosis based on ultra wide band partial discharge detection». Em: pp. 168–170. doi: 10.1109/DEMPED.2003.1234567.

Cleva (2024). *Transformação Digital no Setor Segurador*. url: https://cleva-solutions.com/pt-pt/.

Cowell, Christopher, Nicholas Lotz e Chris Timberlake (2023). «Automating DevOps with GitLab CI/CD Pipelines: Build efficient CI/CD pipelines to verify, secure, and deploy your code using real-life examples». Em: 32.1, pp. 15–25. doi: 10.1145/10162814. url: https://ieeexplore.ieee.org/document/10162814.

Datadog (2024a). *Datadog Pricing Information*. Acessado em 9 dez. 2024. url: https://www.datadoghq.com/pricing/.
– (2024b). *Datadog Resource Catalog*. Acessado em 9 dez. 2024. url: https://docs.datadoghq.com/infrastructure/resource_catalog/.
– (2024c). *IT Infrastructure Monitoring*. Acessado em 9 dez. 2024. url: https://www.datadoghq.com/monitoring/it-infrastructure-monitoring/.
– (2024d). *Modern Infrastructure Monitoring*. Acessado em 9 dez. 2024. url: https://www.datadoghq.com/product/infrastructure-monitoring/.
Dileepkumar, S R e Juby Mathew (2023). «Enhancing DevOps and Continuous Integration in Software Engineering: A Comprehensive Approach». Em: pp. 01–05. doi: 10.1109/ICEEICT56924.2023.10157286.
ECO (2019). *Grupo Gfi finaliza compra da empresa de software i2S*. Acessado em 20 de dezembro de 2024. url: https://eco.sapo.pt/2019/09/19/grupo-gfi-finaliza-compra-da-empresa-de-software-i2s/.
ECOSEGUROS, BRANDS' (2021). *i2S agora é Cleva e quer "continuar a liderar"*. url: https://eco.sapo.pt/2021/06/21/i2s-agora-e-cleva-e-quer-continuar-a-liderar/.
– (2022). *Cleva Inetum: Portugal tem "talentos brilhantes"*. url: https://eco.sapo.pt/2022/10/04/cleva-inetum-portugal-tem-talentos-brilhantes/.
Educação, G4 (2024). *Metodologia Scrum: o que é e para que serve*. Acessado em: 20 de dezembro de 2024. url: https://g4educacao.com/blog/metodologia-scrum.
Elastic.co (2024). *Log Monitoring*. Acessado em: 2 de dezembro de 2024. url: https://www.elastic.co/pt/what-is/log-monitoring.
finout (2024). *What is Datadog?* Acessado em 9 dez. 2024. url: https://www.finout.io/blog/what-is-datadog.
Forense.io (2024). *O que é Gestão de Logs? Como funciona e sua importância*. Acessado em: 2 de dezembro de 2024. url: https://forense.io/glossario/o-que-e-gestao-de-logs-como-funciona-e-sua-importancia/.
Gartner (2024a). *Forecast: Enterprise and Vertical-Specific Software, Worldwide, 3Q24 Update*. Disponível em: https://www.gartner.com/. Acesso em: 02 dez. 2024.
– (2024b). *Insurance CIOs Must Modernize Legacy Systems to Deliver Digital Insurance*. Disponível em: https://www.gartner.com/. Acesso em: 02 dez. 2024.
– (2024c). *Top Technology Trends for the Insurance Industry*. Disponível em: https://www.gartner.com/. Acesso em: 02 dez. 2024.
Gökstorp, Simon et al. (2024). «Anomaly Detection in Security Logs using Sequence Modeling». Em: pp. 1–9. doi: 10.1109/NOMS59830.2024.10575561.
Graylog (2024a). *Graylog Enterprise Features*. Acessado em 9 dez. 2024. url: https://www.graylog.org/products/enterprise.
– (2024b). *Graylog Features Overview*. Acessado em 9 dez. 2024. url: https://graylog.org/feature/.
– (2024c). *What is Graylog?* Acessado em 9 dez. 2024. url: https://www.graylog.org/overview.
Graylog: Solução Poderosa para Gerenciamento de Logs (2024). Acessado em 9 dez. 2024. url: https://inovatechy.com/graylog-solucao-poderosa-para-gerenciamento-de-logs/.
Guru99 (2024). *Log Management Software: Top Tools and Features*. Acessado em: 2 de dezembro de 2024. url: https://www.guru99.com/pt/log-management-software.html.
Habbema, Hugo (2024). *Explorando o Graylog*. Acessado em 9 dez. 2024. url: https://medium.com/@habbema/explorando-o-graylog-549e36b92bca.

Hariri, Sahand, Matias Carrasco Kind e Robert J. Brunner (2021). «Extended Isolation Forest». Em: *IEEE Transactions on Knowledge and Data Engineering* 33.4, pp. 1479–1489. doi: 10.1109/TKDE.2019.2947676.

Haron, Azlena e Shamsul Sahibuddin (2010). «The roles of an actor in requirement engineering (RE) process». Em: 1, pp. 436–440. doi: 10.1109/ICCSIT.2010.5563968.

He, Pinjia et al. (2018). «Towards Automated Log Parsing for Large-Scale Log Data Analysis». Em: *IEEE Transactions on Dependable and Secure Computing* 15.6, pp. 931–944. doi: 10.1109/TDSC.2017.2762673.

He, Shilin, Pinjia He et al. (jul. de 2021). «A Survey on Automated Log Analysis for Reliability Engineering». Em: *ACM Comput. Surv.* 54.6. issn: 0360-0300. doi: 10.1145/3460345. url: https://doi.org/10.1145/3460345.

He, Shilin, Xu Zhang et al. (2022). «An empirical study of log analysis at Microsoft». Em: ESEC/FSE 2022, pp. 1465–1476. doi: 10.1145/3540250.3558963. url: https://doi.org/10.1145/3540250.3558963.

Isaiah, Ayooluwa (2024). *Splunk vs Elastic: ELK Stack: The Key Differences to Know*. Acessado em 9 dez. 2024. url: https://betterstack.com/community/comparisons/splunk-vs-elastic-stack-elk/.

«ISO/IEC/IEEE International Standard - Systems and software engineering – Life cycle processes – Requirements engineering» (2018). Em: *ISO/IEC/IEEE 29148:2018(E)*, pp. 1–104. doi: 10.1109/IEEESTD.2018.8559686.

Kahale, Lara A et al. (mar. de 2021). «PRISMA flow diagrams for living systematic reviews: a methodological survey and a proposal». Em: *F1000Research* 10, p. 192. doi: 10.12688/f1000research.51723.1.

Khalique, Fatima, Wasi Haider Butt e Shoab Ahmad Khan (2017). «Creating Domain Non-functional Requirements Software Product Line Engineering Using Model Transformations». Em: pp. 41–45. doi: 10.1109/FIT.2017.00015.

Korzeniowski, Łukasz e Krzysztof Goczyła (2022a). «Landscape of Automated Log Analysis: A Systematic Literature Review and Mapping Study». Em: *IEEE Access* 10, pp. 21892–21913. doi: 10.1109/ACCESS.2022.3152549.

– (2022b). «Landscape of Automated Log Analysis: A Systematic Literature Review and Mapping Study». Em: *IEEE Access* 10, pp. 21892–21913. doi: 10.1109/ACCESS.2022.3152549.

– (2022c). «Landscape of Automated Log Analysis: A Systematic Literature Review and Mapping Study». Em: *IEEE Access* 10, pp. 21892–21913. doi: 10.1109/ACCESS.2022.3152549.

Kosan, Saltuk (2014). *Design and Creation: Research Framework for Software Engineering*. Acedido em: 16 de dezembro de 2024. url: https://www.medien.ifi.lmu.de/lehre/ss14/swal/presentations/topic2-saltuk_kosan-DesignAndCreation.pdf.

Li, Beibei et al. (2022). «Federated Anomaly Detection on System Logs for the Internet of Things: A Customizable and Communication-Efficient Approach». Em: *IEEE Transactions on Network and Service Management* 19.2, pp. 1705–1716. doi: 10.1109/TNSM.2022.3152620.

Mäntylä, Mika, Martín Varela e Shayan Hashemi (2022). «Pinpointing Anomaly Events in Logs from Stability Testing – N-Grams vs. Deep-Learning». Em: pp. 285–292. doi: 10.1109/ICSTW55395.2022.00056.

Miranskyy, Andriy et al. (2016). «Operational-Log Analysis for Big Data Systems: Challenges and Solutions». Em: *IEEE Software* 33.2, pp. 52–59. doi: 10.1109/MS.2016.33.

Moher, David et al. (jan. de 2015). «Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement». Em: *Systematic reviews*.

Morley, Steven Karl, Thiago Vasconcelos Brito e Daniel T. Welling (jan. de 2018). «Measures of model performance based on the log accuracy ratio». Em: *Space Weather* 16.1. doi: 10.1002/2017SW001669.

Narendiran, A et al. (2023). «Integrated log-aware CI/CD pipeline with custom bot for monitoring». Em: pp. 257–262. doi: 10.1109/ICCCBDA56900.2023.10154891.

Niu, Weina et al. (2022). «LogTracer: Efficient Anomaly Tracing Combining System Log Detection and Provenance Graph». Em: pp. 3356–3361. doi: 10.1109/GLOBECOM48099.2022.10000804.

Orciuolo, Pietro et al. (2024). «The Impact of Dashboard Redesign on the Onboard Management in Maritime Applications: An User Experience Approach». Em: pp. 1–5. doi: 10.23919/AEIT63317.2024.10736806.

Page, Matthew et al. (ago. de 2024). «Declaração PRISMA 2020: uma diretriz atualizada para publicação de revisões sistemáticas». Em: *Germinare — Revista Científica do Instituto Piaget* 4, pp. 1–19. doi: 10.5281/zenodo.13271469. url: https://germinare.ipiaget.org/index.php/germinare/article/view/210.

Peffers, Ken et al. (jan. de 2007). «A design science research methodology for information systems research». Em: *Journal of Management Information Systems* 24, pp. 45–77.

PRISMA (2020). *PRISMA 2020 Flow Diagram*. Acessado em: 08 dezembro 2024. url: https://www.prisma-statement.org/prisma-2020-flow-diagram.

Rethlefsen, Melissa et al. (jan. de 2021). «PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews». Em: *Systematic Reviews* 10. doi: 10.1186/s13643-020-01542-z.

Rianty, Mirradewi, Yusuf Latief e Leni Riantini (jan. de 2018). «Development of risk-based standardized WBS (Work Breakdown Structure) for quality planning of high rise building architectural works». Em: *MATEC Web of Conferences* 159, p. 02029. doi: 10.1051/matecconf/201815902029.

Saltuk, Ozan e Ismail Kosan (2014). *Design and Creation*. Accessed: December 15, 2024. url: https://www.medien.ifi.lmu.de/lehre/ss14/swal/presentations/topic2-saltuk_kosan-DesignAndCreation.pdf.

SAPO, Eco (2024). «Proteção de Dados Pessoais no Setor Segurador em Portugal». Em: Acessado em: 27 de dezembro de 2024. url: https://eco.sapo.pt/opiniao/protecao-de-dados-pessoais-no-setor-segurador-em-portugal/.

Sedki, Issam, Abdelwahab Hamou-Lhadj e Otmane Ait Mohamed (2024). «AML: An accuracy metric model for effective evaluation of log parsing techniques». Em: *Journal of Systems and Software* 216, p. 112154. issn: 0164-1212. doi: https://doi.org/10.1016/j.jss.2024.112154. url: https://www.sciencedirect.com/science/article/pii/S0164121224001997.

Shang, Weiyi, Meiyappan Nagappan e Ahmed E. Hassan (fev. de 2015). «Studying the relationship between logging characteristics and the code quality of platform software». Em: *Empirical Softw. Engg.* 20.1, pp. 1–27. issn: 1382-3256. doi: 10.1007/s10664-013-9274-8. url: https://doi.org/10.1007/s10664-013-9274-8.

Sina, Lennart B. e Kawa Nazemi (2022). «Visual Analytics for Systematic Reviews According to PRISMA». Em: pp. 307–313. doi: 10.1109/IV56949.2022.00059.

Skopik, Florian, Max Landauer e Markus Wurzenberger (2022). «Online Log Data Analysis With Efficient Machine Learning: A Review». Em: *IEEE Security Privacy* 20.3, pp. 80–90. doi: 10.1109/MSEC.2021.3113275.

Snoop, DB (2024). *O que é Elastic ELK? Monitoramento e Logs*. Acesso em: 9 dez. 2024. url: https://www.dbsnoop.com.br/o-que-e-elastic-elk-monitoramento-logs/.

Solutions, Cleva (2024). *Empresa - Cleva Solutions*. Acessado em 20 de dezembro de 2024. url: https://cleva-solutions.com/pt-pt/empresa/.

Splunk (2023). *Log Management: Introduction e Best Practices*. Splunk Inc. url: https://www.splunk.com/en_us/blog/learn/log-management.html.

– (2024a). *Challenges in Scaling Splunk Solutions*. Accessed: 2024-12-09. url: https://www.splunk.com/en_us/blog/learn/scalability.html.

– (2024b). *Data Indexing with Splunk*. Accessed: 2024-12-09. url: https://www.splunk.com/en_us/products/splunk-enterprise-features.html.

– (2024c). *Splunk Enterprise Features*. Accessed: 2024-12-09. url: https://www.splunk.com/en_us/products/splunk-enterprise.html.

– (2024d). *Splunk Pricing and Plans*. Accessed: 2024-12-09. url: https://www.splunk.com/en_us/products/pricing.html.

– (2024e). *Splunk Security Solutions*. Accessed: 2024-12-09. url: https://www.splunk.com/en_us/solutions/security-monitoring.html.

– (2024f). *Splunk: Power the SOC of the Future*. Acessado em 9 dez. 2024. url: https://www.splunk.com/en_us/products/cyber-security.html.

– (2024g). *What is Splunk?* Accessed: 2024-12-09. url: https://www.splunk.com/en_us/blog/learn/what-splunk-does.html.

Sukma, Narongsak et al. (2019). «An Analysis of Log Management Practices to reduce IT Operational Costs Using Big Data Analytics». Em: pp. 1–5. doi: 10.1109/TIMES-iCON47539.2019.9024400.

Sun, Yuchen et al. (2023). «Design and Development of a Log Management System Based on Cloud Native Architecture». Em: pp. 1–6. doi: 10.1109/ICSAI61474.2023.10423328.

TechHyme (2024). *Pros and Cons of ELK Stack (Elasticsearch, Logstash, and Kibana)*. Acesso em: 9 dez. 2024. url: https://techhyme.com/pros-and-cons-of-elk-stack-elasticsearch-logstash-and-kibana/.

Tecnologia, Credited (2024). *O que é Gestão de Logs: Importância e Práticas*. Acessado em: 2 de dezembro de 2024. url: https://tecnologia.credited.com.br/glossario/o-que-e-gestao-de-logs-importancia-e-praticas/.

Veronica e Angellia Debora Suryawan (2019). «User Satisfaction Survey of Performance Management Dashboard Using Delone McLean Method: A Case Study». Em: 1, pp. 542–547. doi: 10.1109/ICIMTech.2019.8843761.

Wang, Jinrui, Mashael AlKadi e Benjamin Bach (2023). «Show Me My Users: A Dashboard Visualizing User Interaction Logs». Em: pp. 156–160. doi: 10.1109/VIS54172.2023.00040.

Works, Elven (2024). *Conheça as 12 Ferramentas Mais Populares para Gerenciamento de Log*. Acessado em: 2 de dezembro de 2024. url: https://elven.works/conheca-as-12-ferramentas-mais-populares-para-gerenciamento-de-log.

Zhao, Xiaoqing, Zhongyuan Jiang e Jianfeng Ma (2022). «A Survey of Deep Anomaly Detection for System Logs». Em: pp. 1–8. doi: 10.1109/IJCNN55064.2022.9892726.

A. Project Charter

ISEP INSTITUTO SUPERIOR DE ENGENHARIA DO PORTO

PROJECT CHARTER

1. Informação geral	
Nome do Projeto:	**Monitorização e Análise de Logs, Alerta de Erros e apresentação de Métricas de Negócio**
Sponsor:	**Isabel Azevedo, Bruno da Silva, Ricardo Leandro**
Departamento	**DEI ISEP**

2. Equipa do Projeto				
Cargo	**Nome**	**Departamento**	**Contact Tel**	**E-mail**
Orientador	**Bruno da Silva**	**DEI ISEP**		**bms@isep.ipp.pt**
Supervisor	**Ricardo Leandro**	**Cleva**		**ricardo.leandro@inetum.com**

3. Stakeholders		
Nome	Poder	Interesse
Departamento de Desenvolvimento	Alto	Alto
Departamento de TI	Baixo	Médio
Departamento de Gestão de Negócios	Baixo	Médio
Departamento de Segurança	Baixo	Alto
Cliente	Baixo	Médio

4. Âmbito

Problema / justificação

Atualmente, a empresa enfrenta desafios significativos na gestão e análise dos seus logs de negócio devido ao elevado volume de dados gerados pelas suas aplicações e sistemas. A maioria dos logs são armazenados em várias tabelas numa base de dados SQL, o que pode resultar em dificuldades na extração eficiente de métricas relevantes, identificação de problemas críticos e tomada de decisões em tempo real. Este cenário é agravado pelo facto de a análise manual de logs ser propensa a erros, ineficiente e insuficiente para lidar com as necessidades de uma organização moderna que opera em ambientes dinâmicos e em constante mudança.

Além disso, a falta de uma solução centralizada e escalável que consiga processar, analisar e alertar automaticamente sobre eventos críticos, impacta diretamente a capacidade de resposta a problemas de negócio e compromete a eficiência operacional.

Área de Pesquisa

- Gestão e análise de dados empresariais, com foco em otimizar e monitorizar processos e extração de métricas a partir de logs.

Tópico de Pesquisa:

- Desenvolvimento de uma solução centralizada e escalável para monitorização e análise automática de logs de negócio, visando melhorar a eficiência operacional e a capacidade de resposta a eventos críticos em tempo real.

Objetivos do projeto

Objetivos:

Objetivos Específicos:

1. Centralizar a gestão de logs.
2. Garantir a escalabilidade da solução.
3. Minimizar o tempo necessário para deteção de erros.
4. Minimizar o tempo necessário para a resolução de erros.
5. Facilitar a integração com outras ferramentas.
6. Garantir a segurança e privacidade dos dados.
7. Monitorizar e melhorar o tempo de resposta dos sistemas.

Metodologia:

1. Revisão da Literatura
2. Análise de Requisitos
3. Prototipagem da Solução
4. Implementação
5. Testes e Validação
6. Avaliação de Resultados e Otimização
7. Documentação Técnica

Perguntas de Pesquisa:

1. De que forma a automação na detecção de erros e na geração de alertas pode reduzir o tempo de resposta a incidentes críticos?
2. Quais os desafios e as vantagens de integrar uma solução de monitorização de logs com outras ferramentas empresariais?
3. Quais os impactos de uma plataforma centralizada de logs na otimização do tempo de resposta dos sistemas?

Hipóteses:

1. **Hipótese Principal**: A implementação de uma plataforma centralizada e escalável para monitorização de logs permitirá a deteção e resolução de problemas de negócio em tempo real, reduzindo significativamente o tempo de resposta a incidentes críticos e aumentando a eficiência operacional.
2. **Hipótese Secundária**: Ao adotar uma solução automatizada de extração de métricas e geração de alertas configuráveis, a equipa poderá diminuir a necessidade de intervenção manual, assegurando a precisão dos alertas e a qualidade dos dados, o que facilitará a tomada de decisões informadas e ágeis.

Benefícios

1. **Redução de Tempo na Deteção e Resolução de Problemas**
2. **Redução de Custos Operacionais**
3. **Aumento na Eficiência da Equipa**
4. **Aumento de Clientes e Satisfação do Cliente**
5. **Escalabilidade e Preparação para o Crescimento**

Entregáveis

- **Relatório de PREPD;**
- **Relatório de DIMEI;**
- **Apresentação e Demonstração do Software;**
- **Software Desenvolvido.**

5. Tempo

Milestones / Datas

- **Introdução**
  - 17/11/2024
- **Estado Arte**
  - 01/11/2024
- **Relatório de PREPD**
  - 04/01/2025
- **Desenvolvimento**
  - 02/06/2025
- **Testes**
  - 23/06/2025
- **Entrega Final**
  - 16/07/2025

6. Custo

Fontes de custo

7. Pressupostos	8. Restrições
Disponibilidade de Recursos Técnicos e Humanos Disponibilidade dos Tech Leads	Conformidade com Padrões de Segurança e Qualidade Capacidade do Hardware Existente Integração com o Ambiente Produtivo Real Dependência de Sistemas Externos

9. Riscos

Riscos identificados

Descrição	Causa	Efeito
Falta de recursos técnicos e humanos	Limitações na disponibilidade de equipas técnicas ou especialistas necessários para o desenvolvimento do projeto.	Atrasos nas entregas das fases do projeto e na implementação final.
Conformidade com padrões de segurança e qualidade	Mudanças ou exigências rigorosas em relação a regulamentações e padrões de segurança.	Requerimentos adicionais de implementação que

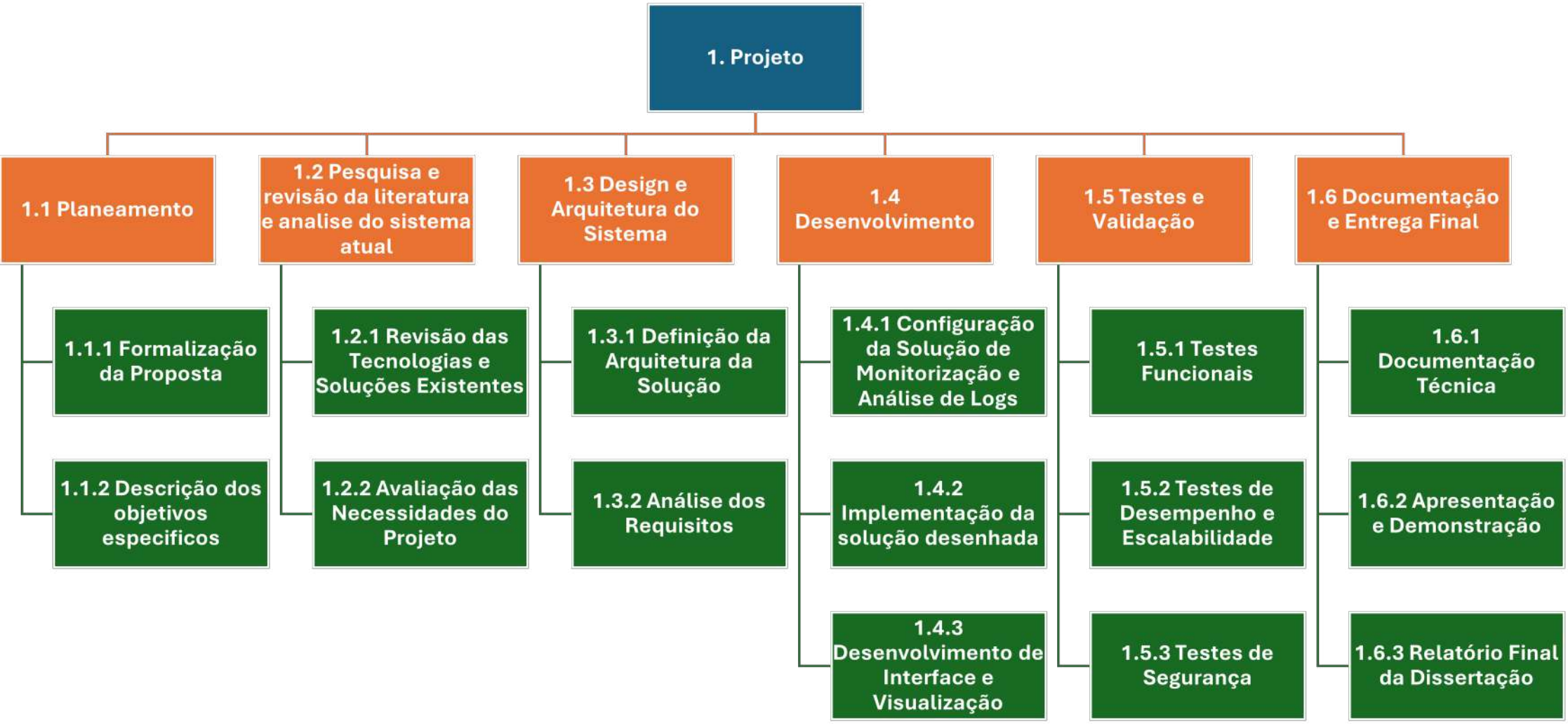
		podem levar a atrasos e aumento de custos.
Capacidade do hardware existente	Limitações na infraestrutura existente para suportar o processamento de grandes volumes de dados.	Redução do desempenho do sistema, comprometendo a eficiência das soluções implementadas.
Dependência de sistemas externos	Problemas de integração com sistemas externos ou falhas em sistemas legados.	Dificuldade em centralizar dados, resultando em informações incompletas ou inconsistentes.

10. Aprovação

	Nome	**Assinatura**	**Data (DD/MM/YYYY)**
Sponsor	Bruno da Silva		
Cliente	Ricardo Leandro		
Gestor do Projeto			

Notas

B. WBS


1. Projeto
1.1 Planeamento
1.2 Pesquisa e revisão da literatura e analise do sistema atual
1.3 Design e Arquitetura do Sistema
1.4 Desenvolvimento
1.5 Testes e Validação
1.6 Documentação e Entrega Final
1.1.1 Formalização da Proposta
1.1.2 Descrição dos objetivos especificos
1.2.1 Revisão das Tecnologias e Soluções Existentes
1.2.2 Avaliação das Necessidades do Projeto
1.3.1 Definição da Arquitetura da Solução
1.3.2 Análise dos Requisitos
1.4.1 Configuração da Solução de Monitorização e Análise de Logs
1.4.2 Implementação da solução desenhada
1.4.3 Desenvolvimento de Interface e Visualização
1.5.1 Testes Funcionais
1.5.2 Testes de Desempenho e Escalabilidade
1.5.3 Testes de Segurança
1.6.1 Documentação Técnica
1.6.2 Apresentação e Demonstração
1.6.3 Relatório Final da Dissertação

C. Riscos

Project Monitorização e Análise de Logs, Alerta de Erros e apresentação de Métricas de Negócio Risk Register

Positive Risk Response Options	Exploit	Share	Enhance	Accept
Negative Risk Response Options	Avoid	Transfer	Mitigate	Accept
Alternate Response Options	Contingency			

Risk ID	Description	Cause	Effect	Risk Owner	Probability (1-5)	Impact (1-5)	PI Score	Expected Result, No Action	Risk Response Type	Response description
	Description of the risk	Cause of the risk	Effect on the project	Name of person who monitors the risk	Group sourced rough estimate of how likely this is to occur	Rough estimate of how significant the impact of this risk	Probability multiplied by Impact	What will happen if the risk becomes an issue and no action is taken	Decision made by group on how to respond to this risk (see above in blue)	How do you know it is time to put the response into play
1	Falha dos stakeholders	Falta de alinhamento entre os envolvidos	Atrasos na tomada de decisão	Fábio Cruz	3	4	12	Decisões importantes podem ser postergadas	Mitigate	Garantir envolvimento ativo de múltiplos stakeholders
2	Complexidade do projeto	Escopo abrangente	Atrasos no desenvolvimento	Fábio Cruz	4	5	20	Sistema inacabado ou com falhas	Mitigate	Focar nas funcionalidades essenciais
3	Falta de adaptação do Log4J	Limitações da ferramenta	Ineficiência na gestão e análise de logs	Fábio Cruz	3	4	12	Impacto na eficiência da solução	Mitigate	Adicionar ferramentas complementares
4	Integração com sistemas legados	Incompatibilidade técnica	Dificuldade de consolidar dados	Fábio Cruz	3	5	15	Dados inconsistentes e atrasos	Transfer	Usar adaptadores ou módulos específicos
5	Sobrecarga de dados	Crescimento exponencial do	Desempenho reduzido	Fábio Cruz	4	5	20	Sistema torna-se lento e instável	Mitigate	Implementar soluções escaláveis

D. Inquérito de Avaliação da Solução de Monitorização de Logs

Inquérito de Avaliação da Solução de Monitorização de Logs

Este inquérito tem como objetivo recolher a opinião dos utilizadores sobre a solução implementada. Por favor, indique o seu grau de concordância com as seguintes afirmações, numa escala de 1 (Discordo totalmente) a 5 (Concordo totalmente).

Olá, F?bio. Quando submeter este formulário, o proprietário verá o seu nome e endereço de e-mail.

* Obrigatório

1. A solução facilita o acesso a logs em tempo real? *

- (●) 1 – Discordo totalmente
- ( ) 2 – Discordo parcialmente
- ( ) 3 – Nem concordo nem discordo
- ( ) 4 – Concordo parcialmente
- ( ) 5 – Concordo totalmente

2. Os dashboards são claros e úteis? *

- ( ) 1 – Discordo totalmente
- ( ) 2 – Discordo parcialmente
- ( ) 3 – Nem concordo nem discordo
- ( ) 4 – Concordo parcialmente
- ( ) 5 – Concordo totalmente

3. A geração de alertas é relevante para o seu trabalho? *

- ◯ 1 – Discordo totalmente
- ◯ 2 – Discordo parcialmente
- ◯ 3 – Nem concordo nem discordo
- ◯ 4 – Concordo parcialmente
- ◯ 5 – Concordo totalmente

4. Considera a solução fácil de usar? *

- ◯ 1 – Discordo totalmente
- ◯ 2 – Discordo parcialmente
- ◯ 3 – Nem concordo nem discordo
- ◯ 4 – Concordo parcialmente
- ◯ 5 – Concordo totalmente

5. A solução melhorou o tempo de resposta a incidentes? *

- ◯ 1 – Discordo totalmente
- ◯ 2 – Discordo parcialmente
- ◯ 3 – Nem concordo nem discordo
- ◯ 4 – Concordo parcialmente
- ◯ 5 – Concordo totalmente

6. Recomendaria esta solução a outras equipas? *

- ◯ 1 – Discordo totalmente
- ◯ 2 – Discordo parcialmente
- ◯ 3 – Nem concordo nem discordo
- ◯ 4 – Concordo parcialmente
- ◯ 5 – Concordo totalmente