



*Arquitetura Conceitual de Sistema para
Interoperabilidade entre Business Intelligence e
Sistemas Legados*

Robson Carlos Bosse

Novembro / 2024

Dissertação de Mestrado em Ciência da Computação

Arquitetura Conceitual de Sistema para Interoperabilidade entre *Business Intelligence* e Sistemas Legados

Esse documento corresponde à Dissertação apresentada à Banca Examinadora no curso de Mestrado em Ciência da Computação da UNIFACCAMP - Centro Universitário Campo Limpo Paulista.

Campo Limpo Paulista, 2 de dezembro de 2024.

Robson Carlos Bosse

Ferruccio de Franco Rosa (Orientador)

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

Ficha catalográfica elaborada pela
Biblioteca Central da Unifaccamp

B757a

Bosse, Robson Carlos

Arquitetura conceitual de sistema para interoperabilidade entre *business intelligence* e sistemas legados / Robson Carlos Bosse. Campo Limpo Paulista, SP: Unifaccamp, 2024.

66 p.: il.

Orientador: Prof. Dr. Ferrucio de Franco Rosa

Dissertação (Programa de Mestrado Profissional em Ciência da Computação) – Centro Universitário Campo Limpo Paulista – Unifaccamp.

1. Qualidade de dados. 2. Análise de dados. 3. Sistemas legados. 4. Arquitetura de sistemas. 5. *Business intelligence*. 6. Interoperabilidade. I. Rosa, Ferrucio de Franco. II. Centro Universitário Campo Limpo Paulista. III. Título.

CDD – 005.75

Agradecimentos

Primeiramente, agradeço a Deus por nos conceder a capacidade de realizar nossos sonhos.

À minha esposa Cássia e à minha filha Fernanda, que compartilharam esse sonho comigo, me apoiaram e compreenderam minhas longas ausências enquanto me dedicava aos estudos para tornar este trabalho possível.

Ao professor Ferrucio de Franco Rosa, por seu incentivo, apoio e fé inabalável na viabilidade deste projeto. Agradeço também pela paciência e parceria em todos os momentos.

Aos professores Amândio Ferreira Balcão Filho e Rodrigo Bonacin, pelas valiosas contribuições durante a qualificação. Suas observações foram fundamentais para aprimorar este trabalho e alcançar o melhor resultado possível.

Ao amigo e colega Milton França Filho, pelo valioso apoio e assistência nos ajustes finais deste trabalho.

À Associação Planalto Central, pelo apoio e pela oportunidade de realizar este trabalho.

Aos professores e colaboradores da UNIFACCAMP, pela dedicação e compromisso em garantir o bom funcionamento do programa de Mestrado.

Resumo. *Business Intelligence (BI) é amplamente utilizado por empresas e organizações com o objetivo de fornecer dados críticos de maneira organizada e precisa, possibilitando a tomada de decisão assertiva fundamentada em dados e informações. Entretanto, há muitos sistemas legados, caracterizados por complexidade e dificuldades de manutenção e integração, que constituem um desafio significativo. Sistemas legados frequentemente sofrem com lentidão e interrupção ao executar requisições de BI em suas bases de dados. Nesta dissertação, uma arquitetura conceitual de sistema é proposta (Conceptual System Architecture for Interoperability between Business Intelligence and Legacy Systems - CSA-IBILS) com o objetivo de promover a interoperabilidade entre sistemas de BI e legados. CSA-IBILS é uma arquitetura em camadas que visa a superar limitações de interoperabilidade identificadas na literatura e em ambiente operacional, proporcionando integração adequada e minimizando os impactos adversos nas operações de negócios. Para validação da proposta, um estudo de caso foi conduzido no ambiente real de uma holding educacional e os resultados foram discutidos. CSA-IBILS destina-se a ser útil para pesquisadores e administradores de sistemas de infraestrutura computacional.*

Palavras-chaves: *Qualidade de Dados, Análise de Dados, Sistemas Legados, Arquitetura de Sistemas, Business Intelligence, Interoperabilidade.*

Abstract. Companies and organizations widely use Business Intelligence (BI) to provide critical data in an organized and accurate manner, enabling assertive decision-making based on data and information. However, many legacy systems, characterized by complexity and difficulties in maintenance and integration, constitute a significant challenge. Legacy systems often suffer from slowness and interruption when executing BI requests in their databases. We propose the Conceptual System Architecture for Interoperability between Business Intelligence and Legacy Systems (CSA-IBILS). CSA-IBILS is a layered architecture that aims to overcome interoperability limitations identified in the literature and an operational environment, providing seamless integration and minimizing adverse impacts on business operations. To validate the proposal, a case study was carried out in the real environment of an educational holding company and the results were discussed. CSA-IBILS is intended to be useful for researchers and administrators of computing infrastructure systems.

Keywords: Data Quality, Data Analysis, Legacy System, System Architecture, Business Intelligence, Interoperability.

Sumário

1	Introdução	13
1.1	Questão de Pesquisa, Objetivos e Contribuições	14
1.2	Estrutura da Dissertação	16
2	Referencial Teórico	17
2.1	Arquitetura Conceitual	17
2.2	Interoperabilidade	17
2.3	Análise de Dados	18
2.4	Qualidade de Dados	18
2.5	Microserviços	19
2.6	Arquitetura em Camadas	19
2.7	Sistemas Legados	21
2.8	<i>Business Intelligence</i>	22
3	Revisão da Literatura e Trabalhos Relacionados	23
3.1	Metodologia	23
3.2	Outras Revisões de Literatura	24
3.3	Qualidade e Análise de Dados	26
3.3.1	Trabalhos que Abordam Questões Relacionadas à Qualidade de Dados	27
3.3.2	Trabalhos que Abordam Questões Relacionadas à Análise de Dados	28
3.4	Discussão Sobre os Resultado da Revisão da Literatura	29
3.5	Trabalhos Relacionados	30
4	Proposta Conceitual	32

4.1	Bases de Conhecimento	33
4.2	Componentes de Software	35
4.3	Processos	37
4.4	Serviços	38
4.5	Segurança	39
4.6	Papéis	40
5	Estudo de Caso	41
5.1	Bases de Conhecimento	43
5.2	Componentes de Software	44
5.3	Processos	46
5.4	Serviços	47
5.5	Segurança	47
5.6	Papéis	48
6	Discussão e Análise Sobre os Resultados do Estudo de Caso	50
7	Conclusão	53
7.1	Limitações e Trabalhos Futuros	53
7.2	Resultados da Pesquisa	54
A	<i>Digital Twin - Configuração de Réplica de Banco de Dados com Atualização Contínua Utilizando SQL Server 2016</i>	62
B	<i>Integração com Atualização Contínua entre SQL Server e MongoDB utili- zando Apache NiFi</i>	64
C	<i>BI Dashboards</i>	66

Glossário

AD	<i>Active Directory</i>
API	<i>Application Programming Interface</i>
BI	<i>Business Intelligence</i>
CSA-IBILS	<i>Conceptual System Architecture for Interoperability between Business Intelligence and Legacy Systems</i>
DBMS	<i>Data Base Management System</i>
ETL	<i>Extract, Transform and Load</i>
HTTPS	<i>Hypertext Transfer Protocol Secure</i>
IES	Instituições de Ensino Superior
INPI	Instituto Nacional da Propriedade Industrial
IoT	<i>Internet of Things</i>
ITNG	<i>International Conference on Information Technology - New Generations</i>
LGPD	Lei Geral de Proteção de Dados
NoSQL	<i>Not Only SQL</i>
SIMS	<i>Strategic Information Monitoring System</i>
SL	Sistema Legado
SOA	<i>Service-Oriented Architecture</i>
SQL	<i>Structured Query Language</i>
VM	<i>Virtual Machine</i>
VPN	<i>Virtual Private Network</i>
WSFC	<i>Windows Server Failover Clustering</i>

Lista de Tabelas

1	Critérios de inclusão e exclusão	24
2	Síntese dos trabalhos de revisão de literatura	26
3	Síntese dos trabalhos analisados	27
4	Resumo dos trabalhos relacionados	31

Lista de Figuras

1	Visão Geral de CSA-IBILS	33
2	Camada Bases de Conhecimento	33
3	Camada Componentes de Software	35
4	Camada Processos	37
5	Camada Serviços	39
6	Camada Segurança	39
7	Camada Papéis	40
8	Diagrama de Rede	41
9	Cronologia de Falhas no Banco de Dados Legado e Consultas de BI . . .	51
10	<i>Dashboard</i> Matrículas	66
11	<i>Dashboard</i> Alvo Financeiro	66
12	<i>Dashboard</i> Ticket Médio	67
13	<i>Dashboard</i> Controle de Bolsas	67
14	<i>Dashboard</i> Alunos por Sala	68

1. Introdução

Sistemas de *Business Intelligence* (BI) têm ganhado uma importância crescente nas empresas e organizações, pois permitem a coleta, análise e visualização de dados para a tomada de decisões mais assertivas (Caserio & Trucco; 2018a). No entanto, muitas empresas ainda utilizam sistemas legados (SL) para gerenciar seus dados, geralmente resultando em problemas de desempenho e escalabilidade (Sandborn & Prabhakar; 2015). SLs são sistemas considerados antigos, complexos, difíceis de atualizar e integrar.

Esses problemas prejudicam a experiência do usuário e dificultam a tomada de decisão baseada em dados, uma vez que não temos dados de qualidade tratados corretamente. Isso ocorre porque os SLs foram projetados para atender a requisitos de desempenho distintos daqueles necessários para sistemas de BI (Jha et al.; 2017). Requisitos esses que foram projetados para uma época com pouca consulta direta em base de dados, onde aplicações trabalhavam com instruções de manipulação de dados (e.g., *Insert*, *Update* e *Delete*). Neste contexto, não era demandado da base de dados legada dados de qualidade para atender relatórios de BI com atualizações contínuas, juntamente com as operações cotidianas.

Como SLs geralmente não estão projetados para suportar grandes cargas de consultas aos dados armazenados, acontece um problema relacionado principalmente à indexação de dados, que faz com que consultas de BI e os serviços do SL para o usuário final sejam interrompidos. Neste caso, é necessário reorganizar os índices para que o SL volte a operar.

A integração entre SLs e sistemas de BI também enfrenta obstáculos significativos no que diz respeito à interoperabilidade. Isso acontece porque os SLs foram construídos com tecnologias e padrões muitas vezes incompatíveis com as ferramentas modernas de análise de dados. Essa incompatibilidade acaba dificultando a extração de dados com atualizações contínuas, uma demanda cada vez mais frequente para operações de BI, prejudicando a capacidade de gerar *insights* valiosos com base em dados atualizados (Arthur & Mukherjee; 2023). No contexto deste trabalho, *insights* representam a capacidade de obter uma compreensão intuitiva, precisa e profunda de uma pessoa, objeto ou fenômeno, baseada na análise dos dados disponíveis.

A interoperabilidade, nesse contexto, representa a habilidade de diferentes sistemas,

inclusive os legados, de se comunicarem e operarem de maneira integrada e eficiente, sem a necessidade de adaptações significativas. No entanto, muitas vezes essa interoperabilidade enfrenta limitações, especialmente quando os SLs são submetidos a grandes volumes de consultas simultâneas, como ocorre nos sistemas de BI. Esse tipo de sobrecarga pode levar o SL a falhas ou até à inoperância, uma vez que sua arquitetura desatualizada não é projetada para suportar o elevado fluxo de dados exigido pelas ferramentas de BI modernas (Arthur & Mukherjee; 2023).

Além disso, a modernização dos SLs para atender às necessidades das soluções de BI traz à tona desafios relacionados ao alto custo e à complexidade do processo. Muitas organizações relutam em investir grandes quantias na atualização de seus sistemas, optando por medidas paliativas que, muitas vezes, não resolvem os problemas de integração de forma adequada. Isso pode impactar negativamente a eficiência operacional e a competitividade da empresa, já que a utilização de dados atualizados se torna cada vez mais crítica para a agilidade e precisão das decisões estratégicas (Sandborn & Prabhakar; 2015).

1.1. Questão de Pesquisa, Objetivos e Contribuições

Neste trabalho, buscamos responder a seguinte **Questão de Pesquisa**: *Como resolver os problemas de interoperabilidade entre SLs e sistemas de BI, focando na qualidade dos dados, desde o tratamento até a análise e visualização, de modo a possibilitar a tomada de decisões?*

O **Objetivo Principal** deste trabalho é propor uma arquitetura conceitual de sistema que solucione o problema de interoperabilidade entre SLs e BI. A proposta visa a disponibilizar dados gerados e processados com foco na qualidade, facilitando sua análise e utilização. Para isso, uma arquitetura em camadas é apresentada e validada por meio de um estudo de caso em uma *holding* escolar. Arquiteturas em camadas e microsserviços oferecem vantagens distintas para a integração de SLs com BI. A arquitetura em camadas fornece uma clara separação de preocupações e simplifica a interoperabilidade, especialmente considerando SLs rígidos que possuem interfaces limitadas. Portanto, para este caso específico, a simplicidade inerente e a compatibilidade das arquiteturas em camadas tornam-na uma escolha mais adequada. Essa abordagem garante uma integração e fluxo

de dados contínuos entre SLs heterogêneos e sistemas de BI. Na validação da proposta conceitual, uma base de dados real em produção foi utilizada em um contexto de inoperabilidade entre SL e BI, onde consultas que exigem uma grande quantidade de dados para BI são realizadas.

Como **Objetivos Específicos**, apresentamos os seguintes:

1. Realizar uma revisão sistemática da literatura com o objetivo de responder a seguinte questão de busca: “*Quais abordagens recentes abordam qualidade e análise de dados para apoio à tomada de decisão no contexto de BI?*”
2. Propor uma arquitetura conceitual de sistema que busque solucionar os problemas de inoperabilidade entre SLs e BI.
3. Validar a arquitetura proposta através de um estudo de caso em uma instituição de ensino, utilizando uma base de dados real em produção.
4. Demonstrar a eficácia da arquitetura em um contexto onde consultas de BI exigem grandes quantidades de dados sem interromper os serviços do SL.

Podem ser citadas como **Contribuições** técnico-científicas deste trabalho as seguintes:

- Arquitetura conceitual em camadas que aborda os problemas de interoperabilidade entre SLs e sistemas de BI, buscando assegurar a continuidade dos serviços dos SLs, mesmo sob alta demanda por consultas de BI. Com foco na avaliação, qualidade e análise de dados, a arquitetura proposta permite aprimorar o processo de tomada de decisão. Nesse estudo, foi incorporado o conceito de *Digital Twin*, que surge como um elemento estratégico da arquitetura conceitual. Conforme o levantamento bibliográfico realizado, não foi identificada a adoção de tal elemento conceitual pelos trabalhos analisados.
- Um estudo de caso prático que valida a aplicabilidade da arquitetura em ambiente operacional. A arquitetura se mostrou eficaz na redução de falhas operacionais resultantes de defeitos de interoperabilidade na infraestrutura computacional.

1.2. Estrutura da Dissertação

A dissertação está estruturada da seguinte forma:

- **Capítulo 2** - Apresenta o referencial teórico, introduzindo conceitos essenciais à discussão dos temas abordados nos capítulos subsequentes.
- **Capítulo 3** - Traz a revisão da literatura e os trabalhos relacionados, destacando os principais desafios e oportunidades associados à qualidade de dados em SLs, além de explorar as principais abordagens e estratégias voltadas para o aprimoramento da qualidade de dados por meio de ferramentas de BI.
- **Capítulo 4** - Descreve a arquitetura conceitual do sistema (*Conceptual System Architecture for Interoperability between Business Intelligence and Legacy Systems - CSA-IBILS*). CSA-IBILS é organizada em seis camadas, que são detalhadas neste capítulo juntamente com seus componentes principais.
- **Capítulo 5** - Apresenta um estudo de caso da aplicação da arquitetura conceitual em um contexto do mundo real.
- **Capítulo 6** - Os resultados obtidos a partir da aplicação de CSA-IBILS no estudo de caso são discutidos.
- **Capítulo 7** - Apresenta as conclusões do trabalho, incluindo limitações, propostas para trabalhos futuros e os resultados da pesquisa.

2. Referencial Teórico

Esta seção tem como objetivo explorar conceitos fundamentais e as principais abordagens associadas a arquiteturas conceituais de sistema que promovem a interoperabilidade entre BI e SLs. Inicialmente, serão discutidos os conceitos de interoperabilidade e suas características, com ênfase nas formas de interação entre diferentes tecnologias. A interoperabilidade em sistemas de BI é particularmente relevante, pois envolve a integração de dados provenientes de várias fontes, a harmonização de diferentes formatos de informação e a implementação de mecanismos de comunicação eficiente entre sistemas antigos e novas plataformas (Kerber & Schweitzer; 2017).

Além disso, serão abordadas as arquiteturas de sistemas que suportam essa interoperabilidade, incluindo a arquitetura orientada a microsserviços (Costa et al.; 2020), bem como a arquitetura em camadas, que tem se mostrado eficaz nesse contexto. A escolha de uma arquitetura adequada é essencial para o sucesso de uma integração eficiente, garantindo flexibilidade, escalabilidade e manutenção ao longo do tempo. Também serão abordados os conceitos relacionados a cada uma das camadas da arquitetura proposta.

2.1. Arquitetura Conceitual

O termo arquitetura conceitual de sistema, no contexto deste trabalho, refere-se à representação em alto nível de sua estrutura, componentes principais e interações. Essa visão global, geralmente apresentada por meio de diagramas, facilita o entendimento das partes interessadas sobre o *design* do sistema, sem a necessidade de detalhamento técnico profundo. Dessa forma, a arquitetura conceitual atua como um instrumento de comunicação entre os envolvidos, promovendo uma compreensão clara e compartilhada dos principais aspectos do sistema (Maier & Rechtin; 2009).

2.2. Interoperabilidade

A interoperabilidade entre sistemas é um aspecto crucial para a integração de tecnologias modernas com infraestruturas existentes, garantindo a continuidade de operações e facilitando a tomada de decisões estratégicas (Abukwaik & Rombach; 2017).

A interoperabilidade emergiu como um tema cada vez mais crítico em ambientes acadêmicos e industriais, impulsionada pela proliferação de sistemas de software dis-

tribuídos. Isso é particularmente evidente em paradigmas modernos, como microsserviços, Internet das Coisas (IoT), *Big Data* e Indústria 4.0. Consequentemente, a interoperabilidade tornou-se um desafio central no *design*, desenvolvimento, teste, implantação e evolução dos sistemas, muitas vezes envolvendo a integração de componentes de sistemas existentes (Valle et al.; 2019).

Interoperabilidade é a capacidade de sistemas distribuídos e colaborativos de se comunicarem e trocarem informações de forma eficaz, superando barreiras fundamentais que possam dificultar essa interação. Essa capacidade é essencial para a construção e gestão desses sistemas, e envolve a satisfação de requisitos críticos que garantem a compatibilidade e a colaboração entre diferentes plataformas e tecnologias. A compreensão desses requisitos e dos desafios associados é crucial para a engenharia de software de sistemas interoperáveis, permitindo a implementação de melhores práticas que promovam uma integração mais fluida e eficiente (Sadeghi et al.; 2023).

2.3. Análise de Dados

A análise de dados é um processo fundamental para transformar grandes volumes de dados brutos em informações úteis e significativas. Muitas vezes, os dados, por si só, não revelam de imediato seu valor, pois as informações estão ocultas em meio à sua complexidade. Para extrair essas informações, é necessário submeter os dados a técnicas de processamento e análise adequadas. Esse processo permite compreender a estrutura dos dados e, quando necessário, modificá-los para torná-los mais acessíveis e interpretáveis (Uzhga-Rebrov & Grabusts; 2021).

2.4. Qualidade de Dados

Qualidade de dados refere-se ao grau em que os dados coletados, categorizados e armazenados são adequados para análise e tomada de decisões. Em um ambiente digital onde o volume de dados é massivo, a baixa qualidade de dados pode gerar desafios significativos para as empresas, resultando em análises imprecisas e decisões equivocadas. Além disso, as soluções de qualidade de dados podem ser demoradas, especialmente em contextos de grandes volumes de dados, impactando negativamente o processo decisório e gerando potenciais perdas. A qualidade dos dados tem um impacto direto na qualidade da

exploração de dados em processos decisórios, onde um modelo que considere a adequação da qualidade dos dados torna-se crucial (Gyulgyulyan et al.; 2018).

2.5. Microsserviços

A arquitetura de microsserviços é amplamente adotada em sistemas distribuídos por oferecer flexibilidade, escalabilidade e independência no desenvolvimento de componentes. Cada microsserviço é projetado para executar uma funcionalidade específica e se comunica com os demais através de APIs bem definidas. Seus benefícios incluem alta disponibilidade do sistema e agilidade em todo o ciclo de vida do serviço. No entanto, a natureza distribuída dos microsserviços também introduz desafios em áreas como modelagem, segurança e infraestrutura (Costa et al.; 2020).

2.6. Arquitetura em Camadas

A arquitetura em camadas organiza sistemas e processos em níveis técnicos, semânticos e organizacionais, facilitando a interoperabilidade entre diferentes entidades. A camada técnica lida com a troca de dados, a semântica assegura um entendimento compartilhado, e a organizacional alinha processos e estruturas. Essa abordagem permite a construção de sistemas escaláveis e flexíveis, sendo essencial para ambientes complexos (Ralf et al.; 2011).

Arquiteturas em camadas são amplamente empregadas em sistemas de *Big Data* devido à organização de funcionalidades relacionadas em níveis distintos. Essa abordagem modular abrange aspectos como segurança, gerenciamento, comunicação entre módulos e interação com sistemas externos. Comumente encontradas em propostas de arquitetura de referência, o modelo em camadas facilita a compreensão da estrutura da solução, dos módulos componentes e suas interações (Sena et al.; 2017). A aplicabilidade dessa arquitetura em um contexto como a integração de SLs com soluções de BI demonstra sua eficácia em lidar com desafios como a interoperabilidade.

Em nossa proposta, para garantir uma melhor organização e modularidade, a arquitetura em camadas foi adotada. Esse modelo, inspirado na arquitetura tradicional de software, permite separar e distribuir responsabilidades entre camadas, facilitando tanto o desenvolvimento quanto a manutenção do sistema (Newman; 2021). Como neste trabalho

propõe-se uma arquitetura conceitual em camadas, cabe discutir os conceitos associados a camadas requeridas ou esperadas, conforme apresentado nos parágrafos seguintes.

A primeira camada esperada de uma arquitetura de sistema é a relacionada às Bases de Conhecimento, que são estruturas que organizam e armazenam o conhecimento em forma de dados, fatos e regras, sendo frequentemente utilizadas em sistemas baseados em conhecimento para resolver problemas complexos. Essas bases compreendem plataformas capazes de centralizar dados e informações, armazenando-os de forma organizada para garantir sua disponibilidade para acesso. Essas plataformas servem como repositórios centralizados, permitindo a fácil recuperação e utilização dos dados contidos nelas. Essas bases não apenas armazenam informações, mas também as estrutura de modo que a máquina possa compreender e utilizar essas informações para a tomada de decisão, enriquecendo a interação entre o usuário e o sistema (Gutierrez & Sequeda; 2021).

Uma das tecnologias utilizadas neste trabalho foi o *Digital Twin*, que serviu como uma das bases de conhecimento. Segundo Fuller et al. (2020), o *Digital Twin* é uma réplica digital de um ativo físico, sistema ou processo. Ele assume um papel crucial ao armazenar, integrar e processar grandes volumes de dados provenientes tanto do sistema físico quanto da réplica digital. Esses dados permitem a criação de modelos que aprimoram a precisão das previsões, automatizam diagnósticos e viabilizam ajustes operacionais, promovendo a manutenção preditiva e a otimização constante do desempenho do sistema.

Outra camada requerida é a de Componentes de Software, que são definidos por Sommerville (2015) como módulos independentes capazes de serem combinados para formar sistemas complexos. Esses componentes possuem interfaces bem definidas e encapsulam tanto os dados quanto o comportamento necessário para executar uma função específica, promovendo assim a separação de preocupações dentro do sistema.

Segundo Omer et al. (2019), componentes de software são unidades modulares e autônomas que desempenham funções específicas dentro de uma arquitetura conceitual de sistema. Estes componentes são projetados para encapsular funcionalidades, com o objetivo de promover a modularidade, reutilização e manutenibilidade do sistema como um todo. Cada componente opera de maneira independente, interagindo com outros componentes por meio de interfaces claramente definidas.

A camada de Processos representa as entidades dinâmicas responsáveis pela execução de atividades e pela manipulação de dados. Esses processos desempenham um papel crucial no gerenciamento e coordenação das operações dentro do sistema, garantindo a obtenção eficiente dos objetivos estabelecidos. Cada processo é caracterizado por uma sequência lógica de passos, interações e operações que contribuem para a funcionalidade global de arquitetura de sistema (Pressman & Maxim; 2019).

Na camada de Serviços estão os componentes que realizam tarefas específicas dentro de um sistema, oferecendo funcionalidades que podem ser facilmente acessadas e utilizadas por outros elementos da arquitetura. Eles atuam como intermediários que facilitam a interação e a colaboração entre diferentes partes do sistema, promovendo a modularidade e a reutilização de código (Sommerville; 2015).

A camada de Segurança, no contexto de sistemas, refere-se à capacidade de um sistema de proteger dados e operações contra acessos não autorizados, garantindo que as funções do sistema sejam executadas de acordo com as políticas de segurança estabelecidas (McGraw; 2006). Além disso, a segurança deve possibilitar a proteção dos dados sensíveis, mantendo a confidencialidade das informações e impedindo que dados críticos sejam divulgados a partes não autorizadas, o que é fundamental para a integridade e a confiança.

Na camada Papéis são estabelecidos os perfis de usuários, que são cruciais para estruturar a coordenação dinâmica entre os diversos componentes, processos e serviços, fornecendo uma visão clara da distribuição de tarefas e autoridades dentro do ambiente do sistema (Sommerville; 2015). Papéis representam entidades (ou perfis) designadas para desempenhar funções específicas e definir responsabilidades dentro do contexto operacional da arquitetura.

2.7. Sistemas Legados

SLs são sistemas de software antigos ainda utilizados por uma organização, apesar da disponibilidade de tecnologias mais recentes. Esses sistemas frequentemente desempenham papéis críticos nas operações de negócios, o que pode complicar sua substituição ou atualização. A manutenção de SL é desafiadora devido a problemas como a falta de suporte para tecnologias obsoletas, a dificuldade em encontrar profissionais com o conhecimento necessário e a dificuldade de integração com novas tecnologias. No entanto,

sua continuidade é muitas vezes necessária para garantir a estabilidade e a continuidade dos serviços (Sommerville; 2015).

2.8. *Business Intelligence*

BI compreende um conjunto de tecnologias, processos e práticas projetados para a coleta, análise e apresentação de informações empresariais com o objetivo de apoiar a tomada de decisões estratégicas. BI utiliza uma variedade de ferramentas e métodos para transformar dados brutos em *insights* valiosos, facilitando a identificação de tendências e padrões que auxiliam as organizações na conquista de seus objetivos. No contexto do BI, *dashboards* desempenham um papel fundamental ao servir como interfaces visuais que consolidam e apresentam dados de forma clara e intuitiva. Esses *dashboards* permitem que os usuários monitorem as informações de maneira eficaz e tomem decisões baseadas em dados (Turban et al.; 2018).

3. Revisão da Literatura e Trabalhos Relacionados

Uma revisão sistemática de literatura foi conduzida (Bosse et al.; 2024b) abrangendo a análise criteriosa de 24 artigos. Esta revisão de literatura teve como objetivo principal responder a seguinte questão de busca: “*Quais abordagens recentes abordam qualidade e análise de dados para apoio à tomada de decisão no contexto de BI?*”

3.1. Metodologia

A metodologia utilizada na revisão da literatura foi baseada nas diretrizes propostas por Kitchenham (2004). Esta revisão busca destacar os principais desafios e oportunidades associados à qualidade de dados em SLs, além de explorar as principais abordagens e estratégias voltadas para o aprimoramento da qualidade de dados através de ferramentas de BI. Foram realizadas buscas em três bases de dados científicas bem conhecidas da área de Computação (IEEE Xplore¹, ACM Digital Library² e Springer Link³).

A escolha dessas bases se justifica pelos seguintes motivos:

1. **Reputação e Credibilidade:** Essas plataformas são reconhecidas globalmente por sua contribuição significativa à pesquisa em Computação e áreas relacionadas. A qualidade dos artigos publicados nelas é rigorosamente avaliada, o que aumenta a confiabilidade das informações obtidas.
2. **Diversidade de Conteúdo:** Cada uma dessas bases de dados abriga uma variedade de publicações, incluindo artigos de conferências, periódicos e relatórios técnicos. Essa diversidade permite uma exploração abrangente dos desafios e soluções propostas na literatura sobre a questão pesquisada.
3. **Foco em Inovação e Tendências:** As três bases são conhecidas por incluir pesquisas de ponta que abordam as últimas inovações e tendências em tecnologia e métodos. Isso é fundamental para identificar não apenas os problemas existentes, mas também as novas abordagens e práticas recomendadas.

A combinação dessas três bases de dados possibilita uma revisão da literatura robusta e bem fundamentada.

¹<https://ieeexplore.ieee.org/Xplore/home.jsp>

²<https://dl.acm.org/>

³<https://link.springer.com/>

O período de buscas estabelecido foi de 11 anos (2013 a 2023) e a seguinte string de busca geral foi utilizada, ajustando a sintaxe a cada base de dados: (*"Data Analysis" OR "Data Quality" OR "Database"*) AND (*"Business Analytics" OR "Business Intelligence" OR "Business Intelligence Performance"*) AND (*"Evaluation" OR "Assessment"*) AND (*"Legacy System" OR "Brownfield System"*).

A execução da busca considerou todos os artigos retornados. A busca inicial obteve um total de 112 artigos, sendo 43 artigos de IEEE Xplore, 33 de ACM Digital Library e 36 de Springer Link.

Os critérios de inclusão e exclusão (Tabela 1) foram estabelecidos durante a pesquisa exploratória por meio de um processo iterativo de análise dos artigos relevantes ao tema proposto. A avaliação inicial considerou título, resumo e palavras-chave, resultando na seleção de 24 artigos. Destes, cinco são revisões de literatura com foco em qualidade ou análise de dados, sendo abordados na Seção 3.2. A Seção 3.3 apresenta trabalhos que investigam especificamente qualidade e análise de dados. Os resultados da revisão de literatura que contribuem para a tomada de decisão baseada em dados são discutidos na Seção 3.4. Por fim, três trabalhos relacionados à interoperabilidade de sistemas são analisados na Seção 3.5.

Tabela 1. Critérios de inclusão e exclusão

Critérios de Inclusão	Critérios de Exclusão
(i.1) Artigos de periódicos ou anais de eventos científicos, com texto completo disponível	(ii.1) Área de pesquisa que não seja Ciência da Computação
(i.2) Publicações realizadas de 2013 a 2023	(ii.2) Trabalhos que não se relacionam com a pergunta norteadora
(i.3) Trabalhos publicados na língua inglesa	(ii.3) Artigos curtos e resumos

3.2. Outras Revisões de Literatura

Nesta seção, apresentamos trabalhos que apresentam revisões de literatura sobre qualidade e análise de dados para tomada de decisão utilizando BI. As descrições das

revisões de literatura analisadas são apresentadas a seguir.

Santi & Putra (2018) apresentam uma revisão sistemática da literatura com o objetivo de identificar a aplicação de BI no setor educacional, especialmente em termos de tecnologia, contribuição e aplicação. Segundo os autores, a aplicação de BI no setor educacional tem o potencial de melhorar a qualidade da educação, além de ajudar a melhorar a eficiência, eficácia e equidade do setor educacional.

Egas et al. (2022) conduziram uma revisão sistemática da literatura com o objetivo de apresentar a situação atual da gestão estratégica nas Instituições de Ensino Superior (IES), baseada em *Data Analytics* e *Business Analytics*. A revisão da literatura identificou que a gestão estratégica baseada em *Data Analytics* e *Business Analytics* é uma ferramenta importante para as IES tomarem decisões mais assertivas. Os dados coletados podem ser utilizados para identificar tendências, oportunidades e riscos, o que pode ajudar as IES a melhorar seus processos, produtos e serviços. Segundo os autores, a gestão estratégica baseada em *Data Analytics* e *Business Analytics* é uma tendência crescente nas IES e instituições que adotarem essa abordagem terão uma vantagem competitiva no mercado.

Lakshen et al. (2016) conduziram uma revisão de literatura sobre qualidade de dados no contexto de *Big Data*. Neste trabalho, os autores enfatizam a importância da qualidade dos dados para a confiabilidade das análises.

Wilson et al. (2017) apresentaram uma revisão de literatura sobre o uso de *Big Data* na área da saúde com o objetivo de analisar o impacto dos dados para a tomada de decisões mais precisas para o cuidado da saúde de pacientes.

Jozef & Osmólski (2023) apresentaram uma revisão de literatura sobre a aplicação de *Big Data* em soluções de logística com o objetivo de identificar desafios, métodos e melhores práticas.

Na Tabela 2, os trabalhos de revisão de literatura foram classificados de acordo com seus objetivos e domínios de aplicação. Na Tabela 2, consideramos Objetivos: *Framework* (F); Revisão da Literatura (RI); Sistema (S); Abordagem (Ab); Avaliação (Av); e Domínios de Aplicação: (1) *Business Intelligence*; (2) *Big Data*; (3) Gerenciamento de Dados; (4) Análise de Dados; (5) *Data Warehouse*; (6) Qualidade de Dados. Os objetivos presentes

na tabela incluem todos aqueles encontrados nos trabalhos analisados. No entanto, para essa etapa da análise, foram considerados apenas os trabalhos que realizam revisões de literatura.

Tabela 2. Síntese dos trabalhos de revisão de literatura

Autores	Objetivos					Aplicação					
	F	RI	S	Ab	Av	1	2	3	4	5	6
Santi & Putra (2018)		X				X					
Egas et al. (2022)		X							X		
Lakshen et al. (2016)		X					X				X
Wilson et al. (2017)		X					X				
Jozef & Osmólski (2023)		X					X				
<i>Nossa Revisão</i>		X				X	X	X	X	X	X

3.3. Qualidade e Análise de Dados

Esta seção apresenta uma análise sistemática de 16 trabalhos de pesquisa, categorizados em dois grupos distintos: i) trabalhos que priorizam a análise de dados e ii) trabalhos que enfatizam a qualidade dos dados como parâmetro principal. É importante ressaltar que a qualidade dos dados e a análise de dados são intrinsecamente interligadas e se complementam, sendo ambas cruciais para a pesquisa rigorosa. A Subseção 3.3.1 apresenta os trabalhos que se concentram na qualidade dos dados, enquanto a Subseção 3.3.2 explora os trabalhos focados na análise de dados.

A Tabela 3 apresenta uma síntese dos estudos analisados e está estruturada da seguinte forma: Objetivos: *Framework* (F); Sistema (S); Abordagem (Ab); Avaliação (Av) e Domínios de Aplicação: (1) *Business Intelligence*; (2) *Big Data*; (3) Gerenciamento de Dados; (4) Análise de Dados; (5) *Data Warehouse*; (6) Qualidade de Dados.

Tabela 3. Síntese dos trabalhos analisados

Autores	Objetivos				Aplicação					
	F	S	Ab	Av	1	2	3	4	5	6
Fernández et al. (2017)				X	X					X
Anwar & Handayani (2022)				X	X					
Sianipar et al. (2019)				X	X					X
Bouchana & Idrissi (2015)				X	X					
Büyüközkan et al. (2019)				X	X					
Peng et al. (2017)			X		X			X		
Verma & Voids (2016)				X	X					
Klisarova-Belcheva et al. (2019)			X		X					
Malaka & Brown (2015)			X			X		X		
Sumera & Miskon (2020)				X	X					
Caserio & Trucco (2018a)			X		X					
Mario & Fiduccia (2022)		X						X	X	
Song et al. (2018)				X	X					
Wilfried & Rinderle-Ma (2015)			X				X			
Prashneel & Kumar (2023)	X				X				X	
Lopes et al. (2022)			X		X					X

3.3.1. Trabalhos que Abordam Questões Relacionadas à Qualidade de Dados

Fernández et al. (2017) propõem um modelo de qualidade para soluções de BI acadêmicas. O modelo é baseado na série de normas ISO/IEC 25000, que definem requisitos de qualidade para sistemas de informação.

Anwar & Handayani (2022) conduziram uma avaliação do uso de BI em uma empresa de energia com o objetivo de apresentar fatores essenciais para uso contínuo de BI.

Sianipar et al. (2019) apresentam uma avaliação com o objetivo de identificar e descrever os fatores críticos de sucesso na implementação de BI, a saber: (1) Visão de

negócios e plano estratégico claros; (2) Adequação das soluções de BI às expectativas dos usuários; e (3) Qualidade e integridade dos dados. Segundo os autores, é importante que os dados sejam de alta qualidade e estejam integrados.

3.3.2. Trabalhos que Abordam Questões Relacionadas à Análise de Dados

Bouchana & Idrissi (2015) apresentam um modelo de avaliação de sistemas de BI orientado a produto, que avalia dois aspectos principais em BI: i) a satisfação do usuário final e ii) a qualidade dos dados. O modelo é baseado em uma lista de dimensões que avaliam a relevância, a facilidade de uso e a utilidade da informação, bem como a precisão, a completude e a relevância dos dados.

Büyüközkan et al. (2019) apresentam uma avaliação para a seleção de sistemas de BI. O objetivo é selecionar os sistemas de BI mais adequados para uma determinada tarefa complexa, pois é necessário considerar vários critérios, tais como custos, funcionalidades, desempenho e suporte.

Peng et al. (2017) apresentam uma abordagem para aplicação de BI em instituições de ensino superior com o objetivo de extrair dados de aprendizagem de estudantes e verificar a operação dessas instituições a partir dos dados.

Verma & Voidsa (2016) apresentam uma avaliação da utilização de BI em uma organização de recursos humanos com o objetivo de identificar os pontos necessários para o desenvolvimento de uma cultura de dados.

Klisarova-Belcheva et al. (2019) apresentam uma abordagem para avaliar o uso de BI no processo de tomada de decisão em pequenas e médias empresas na Bulgária. O objetivo principal foi identificar problemas enfrentados e apresentar recomendações para o uso eficiente de BI.

Malaka & Brown (2015) apresentam uma abordagem para aplicação de *Big Data Analytics* em organizações de telecomunicações da África do Sul.

Sumera & Miskon (2020) conduziram uma avaliação para identificar o impacto do uso de BI na indústria têxtil e de vestuário.

Caserio & Trucco (2018a) apresentam uma abordagem para identificar um conjunto

de fatores críticos de sucesso que permitem uma implementação eficaz de um sistema de BI que possa satisfazer as necessidades da empresa.

Mario & Fiduccia (2022) apresentam o Sistema de Monitoramento de Informações Estratégicas (SIMS) do Grupo Italiano Ferrovie dello Stato. O objetivo do SIMS é prover ferramentas de análise de dados através de um *Data Warehouse* para apoiar o processo de tomada de decisão estratégica da companhia.

Song et al. (2018) apresentam uma avaliação para entender o comportamento do uso de sistemas de BI em larga escala nas organizações indígenas chinesas. O objetivo é fornecer um conjunto de propostas para fundamentar pesquisas futuras sobre BI na China.

Wilfried & Rinderle-Ma (2015) apresentam uma abordagem para provisionamento de dados, desde coleta e extração até transformação e armazenamento. O objetivo do artigo é apresentar os principais conceitos e métodos envolvidos no processo de provisionamento, com foco nos dados transacionais, que são os dados gerados pelas operações cotidianas de uma empresa.

Prashneel & Kumar (2023) apresentam um *framework* que usa BI, algoritmos de mineração de dados e *Data Warehouse* com o objetivo de analisar o histórico de passageiros e compreender as reservas de passagens nas companhias aéreas, possibilitando a tomada de decisões estratégicas.

Lopes et al. (2022) apresentam uma abordagem para identificar fatores críticos de sucesso tecnológico (i.e., integração de sistemas e qualidade da informação) para a implementação bem-sucedida de sistemas de BI. Os resultados foram obtidos através de um estudo de caso realizado em uma instituição de ensino superior.

3.4. Discussão Sobre os Resultado da Revisão da Literatura

Esta revisão de literatura apresentou abordagens voltadas ao aprimoramento da qualidade e análise de dados para apoio à tomada de decisão, no contexto de sistemas BI. A revisão identificou uma série de técnicas e métodos que podem ser utilizados para melhorar a qualidade dos dados, considerando os requisitos específicos de tomada de decisão. Os resultados da revisão indicam que o aprimoramento da qualidade dos dados em sistemas é um desafio complexo, que requer uma abordagem multidimensional.

É fundamental reconhecer que a qualidade dos dados não é apenas uma questão técnica, mas também uma questão estratégica. Organizações que desejam melhorar a qualidade de seus dados devem adotar uma abordagem holística que englobe aspectos técnicos, processuais e culturais. As tendências emergentes em automação e inteligência artificial oferecem muitas oportunidades para melhorar a qualidade dos dados de forma eficiente. No entanto, é importante abordar essas tecnologias com cuidado, garantindo a transparência, a ética e a segurança no tratamento dos dados.

Os resultados desta revisão contribuem para uma compreensão mais abrangente das estratégias disponíveis para aprimorar a qualidade dos dados. Esperamos que este trabalho inspire organizações a investirem na melhoria da qualidade de seus dados, reconhecendo o valor fundamental que dados confiáveis têm no processo de tomada de decisão.

A partir da análise dos resultados da revisão, identifica-se a necessidade uma arquitetura conceitual de sistema que promova a qualidade de dados no contexto de BI. Essa arquitetura deve enfatizar a qualidade dos dados, abordando aspectos como desempenho, integridade e utilidade.

3.5. Trabalhos Relacionados

Como trabalhos relacionados que passaram pela *string* de busca e pelos critérios de inclusão e exclusão, destacamos 3 estudos que de alguma forma abordam falhas e questões relacionadas à interoperabilidade em sistemas de BI.

Akhir et al. (2018) apresentam requisitos para a criação de um módulo de relatório de visualização de dados, visando otimizar as informações apresentadas por um sistema judiciário. O objetivo principal foi propor uma abordagem eficiente de gestão de dados para facilitar a tomada de decisão.

Caserio & Trucco (2018b) apresentam uma avaliação das possíveis relações entre os sistemas de ERP (*Enterprise Resource Planning*) e BI e a sobrecarga (i.e., gerar uma grande quantidade de dados) ou sub-carga (i.e., dificultar o acesso a dados relevantes) de informação.

Fraihat et al. (2021) apresentam um *framework* de BI para o mercado imobiliário jordaniano. Segundo os autores, o *framework* proposto oferece análises de dados e suporte à

tomada de decisão direcionados a potenciais investidores, sendo especificamente projetado para ser utilizado por aqueles com pouco ou nenhum conhecimento prévio do mercado.

A Tabela 4 apresenta uma comparação entre os trabalhos relacionados e este trabalho com foco na arquitetura conceitual. Ressaltamos que nenhum dos trabalhos analisados na revisão apresentam como contribuição uma arquitetura de sistema. Na Tabela 4, consideramos Objetivos: *Framework* (F); Sistema (S); Avaliação (Av); Arquitetura (Ar); e Domínios de Aplicação: (1) *Business Intelligence*; (2) Análise de Dados; (3) Qualidade de Dados.

Tabela 4. Resumo dos trabalhos relacionados

Autores	Objetivos				Aplicação		
	F	S	Av	Ar	1	2	3
Akhir et al. (2018)		X				X	
Caserio & Trucco (2018b)			X		X		
Fraihat et al. (2021)	X				X		X
<i>Este Trabalho</i>				X	X	X	X

4. Proposta Conceitual

Uma arquitetura conceitual bem construída é crucial na construção de sistemas robustos e eficientes, ao fornecer uma estrutura de alto nível que guia o desenvolvimento subsequente do sistema (Benedictis et al.; 2023).

Para lidar com os desafios relacionados a disponibilidade, escalabilidade, confiabilidade e interoperabilidade no acesso a dados de SLs para sistemas de BI, propomos CSA-IBILS (*Conceptual System Architecture for Interoperability between Business Intelligence and Legacy Systems*) (Bosse et al.; 2024a).

A arquitetura conceitual de sistema proposta está organizada em seis camadas: Bases de Conhecimento, Componentes de Software, Processos, Serviços, Segurança e Papéis. Detalhamos cada camada e seus componentes para uma compreensão aprofundada.

A Figura 1 apresenta uma visão geral de CSA-IBILS. As figuras da arquitetura proposta, assim como os demais diagramas que representam as camadas dessa arquitetura, estão em inglês devido ao fato de a figura com a visão geral da arquitetura conceitual de sistemas, bem como as figuras que detalham cada camada e seus componentes, serem reproduções do conteúdo do artigo principal desta dissertação que foi submetido para uma conferência internacional.

Ao propor CSA-IBILS, buscamos mitigar os riscos e enfrentar os desafios inerentes à integração de dados provenientes de SL, promovendo a eficiência operacional e a confiabilidade de BI. Nossa proposta busca superar as limitações de desempenho, além de oferecer uma estrutura escalável que atenda às demandas crescentes dos ambientes computacionais complexos.

CSA-IBILS foi desenvolvida utilizando a notação Archimate Modelling (Archimate Tool; 2024). Usamos esta notação devido a sua capacidade de representar com clareza elementos e relações associados à arquitetura de sistemas, facilitando assim a compreensão e a comunicação.

Cada camada de CSA-IBILS foi planejada com o objetivo de estruturar e tratar os dados, lidando com possíveis questões de interoperabilidade. Nas subseções seguintes, apresentamos em detalhe cada camada, iniciando pela camada inferior (Bases de Conheci-

mento) e finalizando com a camada de mais alto nível de abstração (Papéis). Na descrição das camadas, destacamos as contribuições de cada camada para a eficácia da arquitetura conceitual.

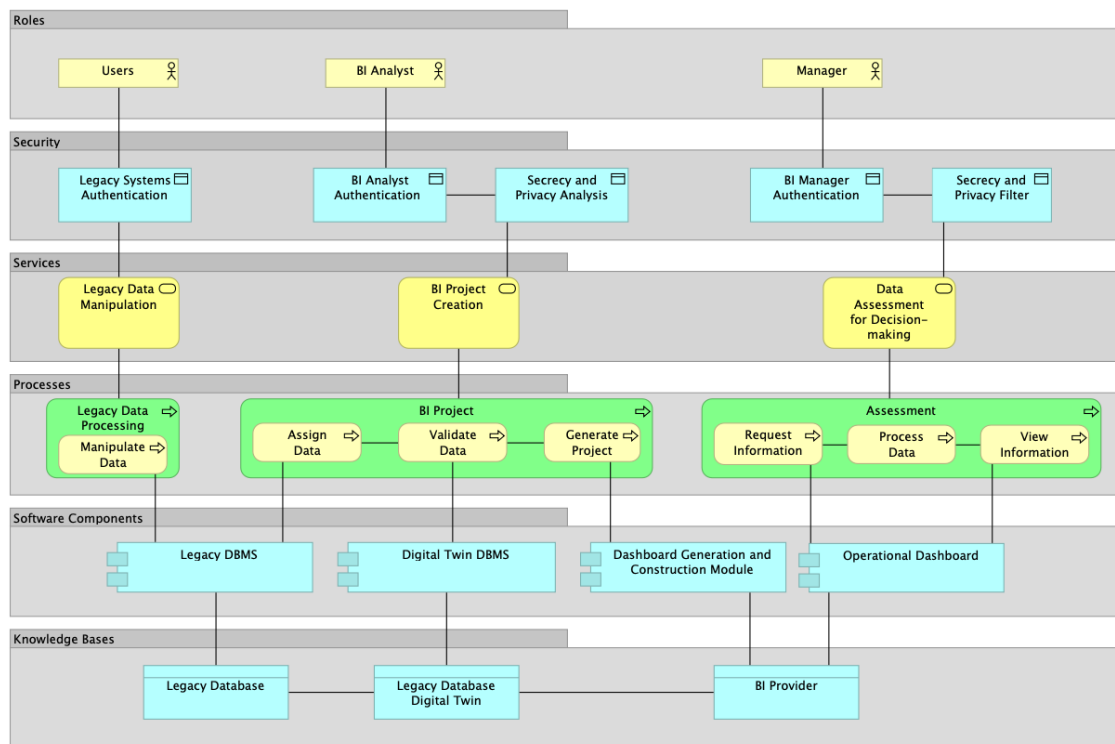


Figura 1. Visão Geral de CSA-IBILS

4.1. Bases de Conhecimento

CSA-IBILS emprega três bases de conhecimento distintas, a saber: *Legacy Database*, *Legacy Database Digital Twin* e *BI Provider*. A Figura 2 apresenta a camada Bases de Conhecimento, separada das demais para facilitar a análise de sua estrutura e composição.

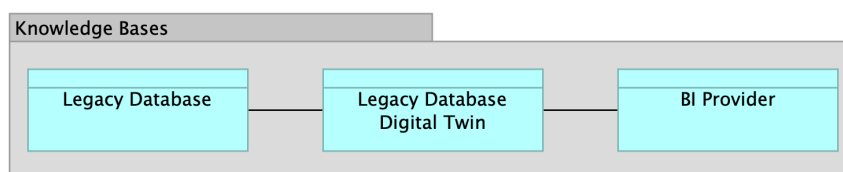


Figura 2. Camada Bases de Conhecimento

Legacy Database refere-se a uma infraestrutura mais antiga para armazenamento de dados que tem resistido ao teste do tempo. Geralmente, esse tipo de banco de dados é marcado por tecnologia obsoleta e, em muitos casos, por uma arquitetura desatualizada. A principal característica dessa base de dados legada é a sua natureza relacional, o que significa que ela é baseada em modelos de dados estruturados que utilizam entidades (tabelas) para organizar informações.

Legacy Database Digital Twin refere-se a uma réplica virtual completa e dinâmica de um banco de dados real. *Digital Twin* é definido como uma representação virtual de um sistema físico que permanece sincronizado com sua contraparte do mundo real por meio de troca contínua de dados. Essa conexão permite monitoramento, simulação e otimização de processos com atualização contínua, possibilitando melhor tomada de decisão e análise preditiva (Grönman et al.; 2023). Essa base de conhecimento é atualizada constantemente, mantendo-se como uma cópia fiel de *Legacy Database*.

BI Provider é um repositório consolidado de dados que já passaram pelo processo de ETL (*Extract, Transform and Load*) (Zineb & Rachid; 2023). Nesta Base de Conhecimento, os dados essenciais são organizados e armazenados de forma a atender às necessidades específicas do software de BI. No contexto de CSA-IBILS, *BI Provider* é uma base de conhecimento dedicada exclusivamente ao BI, e seu abastecimento é proveniente de *Legacy Database Digital Twin*. *BI Provider* adota uma abordagem NoSQL, escolhida estrategicamente para conferir dinamicidade à base de conhecimento, visando atender a requisitos gerais de desempenho do BI. Entendemos que a opção por um modelo NoSQL promove a flexibilidade necessária para lidar com dados complexos e variados, permitindo uma adaptação ágil às demandas em constante evolução do ambiente de negócios.

Um aspecto crítico a ser considerado na interligação e comunicação entre estas três bases de dados é a infraestrutura da rede lógica, a qual deve ser adaptada para se adequar às especificidades de cada contexto durante a implementação de CSA-IBILS em diferentes ambientes operacionais. Para otimizar a eficiência do sistema, é crucial estabelecer uma integração fluida entre os bancos de dados e os servidores virtualizados, permitindo uma comunicação eficaz e sem entraves. Além disso, a sincronização de dados entre essas bases deve ser cuidadosamente planejada, levando em consideração a consistência e a

integridade das informações em todo o ambiente.

A adaptação da infraestrutura de rede lógica é crucial no processo de implementação, pois as características e demandas de conectividade podem variar significativamente entre os ambientes operacionais. É imperativo realizar uma análise detalhada das necessidades específicas de cada contexto, garantindo a configuração adequada da rede para suportar a eficiente transferência de dados e a comunicação entre os servidores virtualizados e os bancos de dados. CSA-IBILS oferece uma solução robusta para o ambiente legado e a flexibilidade necessária para se adaptar a ambientes operacionais diversos. A capacidade de ajustar a infraestrutura de rede lógica conforme a necessidade específica de cada implementação assegura a eficácia desta arquitetura em diferentes contextos, promovendo uma gestão eficiente e segura dos dados.

4.2. Componentes de Software

CSA-IBILS possui quatro componentes de software com o objetivo de trabalhar com as bases de conhecimento de forma independente para que o melhor de cada base seja coletado e transformado em dados para tomada de decisão. A seguir, apresentamos cada componente de CSA-IBILS. A Figura 3 apresenta a camada Componentes de Software.

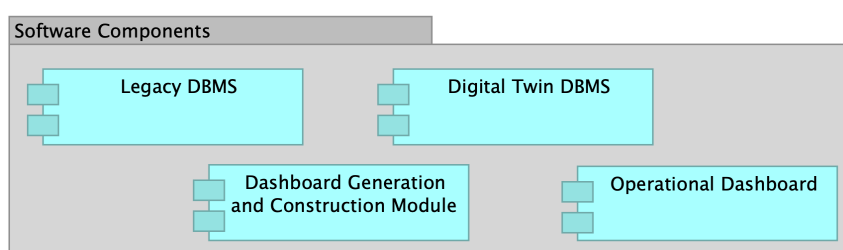


Figura 3. Camada Componentes de Software

Legacy DBMS é responsável por facilitar o armazenamento, recuperação e manipulação de dados em um ambiente organizado, sendo um componente fundamental na infraestrutura de tecnologia da informação. DBMS (*Data Base Management System*) desempenham um papel fundamental na gestão eficaz de grandes volumes de informações, independentemente de ser um sistema legado ou não (Grigoriev; 2015). *Legacy DBMS* fornece uma interface para usuários e aplicativos interagirem com o banco de dados. No contexto de CSA-IBILS, este componente viabiliza a manipulação da base de dados legada.

Digital Twin DBMS é o componente que manipula e acessa a base de dados *Digital Twin* da aplicação legada, possuindo a mesma estrutura de *Legacy DBMS*. A inclusão na proposta de um gêmeo digital da base de dados legada visa a desobrigar o DBMS legado de responder consultas que consomem muito processamento, mantendo sua estabilidade.

Dashboard Generation and Construction Module é uma ferramenta especializada, voltada para a criação, customização e apresentação de *dashboards* interativos. *Dashboards* representam interfaces gráficas que oferecem visualizações consolidadas e intuitivas de dados, facilitando a análise e a tomada de decisões (Muppidi et al.; 2022). A função principal deste componente é simplificar o processo de criação de *dashboards*, possibilitando aos usuários a seleção de diversas fontes de dados, a personalização de *layouts*, e a incorporação de diversos elementos visuais, tais como gráficos, tabelas e indicadores de desempenho. Este componente é proposto com o objetivo de potencializar a eficiência na apresentação e interpretação de informações, provendo uma abordagem mais eficaz na análise de dados para suportar a tomada de decisão.

Operational Dashboard representa uma interface visual que oferece uma representação gráfica e resumida de dados operacionais, proporcionando informações relevantes para apoiar a tomada de decisão. *Dashboards* são projetados para fornecer uma visão instantânea e compreensível do desempenho de processos, atividades ou sistemas em uma organização. Esse componente destaca-se por apresentar de forma clara e concisa dados importantes para tomada de decisão, usando gráficos, tabelas e outros elementos visuais (Kristyanti et al.; 2020). *Dashboards* operacionais são usados em ambientes empresariais para monitorar com dados atualizados, métricas cruciais, tais como indicadores de desempenho, metas de produção e eficiência operacional. Uma das características deste componente é sua capacidade de apresentar informações de maneira acessível, permitindo a rápida identificação de tendências e padrões, habilitando usuários a tomar decisões de maneira mais assertiva (Freitas & Alturas; 2018). O componente proposto pode ser personalizado conforme as necessidades específicas de uma organização, setor ou processo, proporcionando uma visão geral e intuitiva do ambiente operacional. Ao proporcionar uma representação visual dinâmica, *dashboards* contribuem para uma compreensão abrangente do estado operacional e, conseqüentemente, para a otimização contínua dos processos.

4.3. Processos

Nesta camada, os dados são tratados e manipulados, seguindo o fluxo estabelecido pela arquitetura conceitual. Cada processo é delineado por uma sequência lógica de passos, interações e operações que contribuem para a funcionalidade geral. A Figura 4 apresenta a camada Processos.

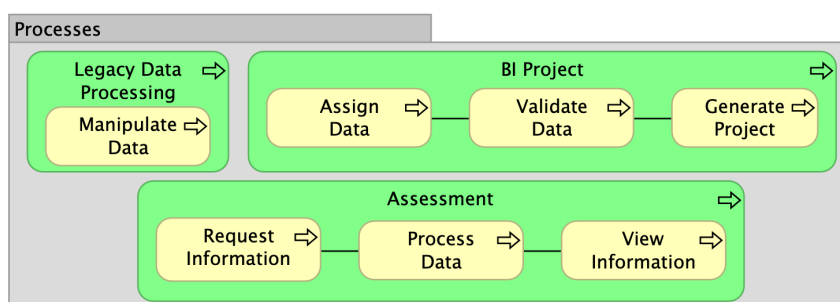


Figura 4. Camada Processos

Manipulate Data diz respeito às operações e processos envolvidos na gestão, modificação e organização dos dados no âmbito de um sistema ou ambiente de tecnologia da informação.

BI Project visa à coleta, processamento e análise dos dados organizacionais, com o objetivo de fornecer *insights* para a tomada de decisão. Isso é alcançado por meio da implementação de ferramentas, tecnologias e processos específicos, destinados a transformar dados brutos em informações significativas. A arquitetura do projeto inclui a integração de fontes de dados, modelagem de dados, e o desenvolvimento de *dashboards* interativos e relatórios. Este processo é conduzido utilizando-se um sistema legado e uma base de dados legada como ponto de partida. Três atividades essenciais ocorrem neste ponto do processo: i) *Assign Data*, que consiste em assegurar a associação correta das informações aos elementos relevantes, fornecendo uma base sólida para análise e tomada de decisão; ii) *Validate Data*, que consiste no procedimento para verificar a precisão, consistência e integridade dos dados utilizados pelo BI. Este estágio visa a garantir que as informações empregadas em análises ou relatórios sejam confiáveis e livres de erros; e iii) *Generate Project*, que consiste em garantir uma implementação eficiente e bem-sucedida das soluções de BI, atendendo às necessidades organizacionais, e disponibilizando as

informações e dados desejados em um *dashboard* para visualização.

No processo de *Assessment* acontece a análise e mensuração do desempenho, eficácia e eficiência das análises implementadas nos relatórios de BI. Este processo visa a fornecer *insights* acionáveis, identificar oportunidades de melhoria e garantir que os objetivos de negócios sejam alcançados de maneira otimizada. As métricas de avaliação podem abranger desde a precisão e relevância dos dados até a usabilidade das interfaces de *dashboard*, contribuindo assim para a tomada de decisão. No processo de *Assessment* temos as seguintes atividades: i) *Request Information*, que se refere ao processo de solicitar informações específicas ou conjuntos de dados relevantes para a tomada de decisão. Isso envolve a formulação de consultas ou pedidos direcionados a fontes de dados; ii) *Process Data*, que envolve a manipulação e transformação de conjuntos de dados brutos em informações significativas e úteis; e iii) *View Information*, que é à representação visual de dados e *insights* obtidos por meio do processo de BI. Isso inclui a criação de gráficos, *dashboards* interativos, relatórios visuais e outras representações visuais que facilitam a compreensão rápida e eficaz das tendências, padrões e informações relevantes derivadas dos dados.

4.4. Serviços

Na camada Serviços, a atividade *Legacy Data Manipulation* aborda o gerenciamento, transformação e manipulação de informações que estão armazenadas em base de dados de SL. *BI Project Creation* diz respeito ao projeto e implementação de um conjunto de atividades e processos destinados a coletar, organizar, analisar e apresentar dados relevantes para suportar a tomada de decisão em uma organização. *Data Assessment for Decision-making* busca fazer com que as decisões organizacionais sejam informadas e alinhadas aos objetivos estratégicos de empresas e organizações, para aproveitar ao máximo suas informações e a enfrentar desafios de maneira mais assertiva utilizando dados. A Figura 5 apresenta a camada Serviços.

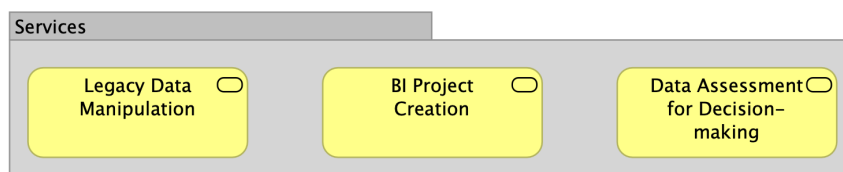


Figura 5. Camada Serviços

4.5. Segurança

Atualmente, segurança é um elemento fundamental em diversas áreas, especialmente no âmbito da proteção de dados e privacidade. Embora as regulamentações e diretrizes variem entre países, existe um interesse geral em assegurar níveis adequados de segurança para proteger usuários, empresas e entidades governamentais. Em CSA-IBILS todos os acessos aos dados devem passar por uma camada de segurança que busca maximizar as exigências para o acesso aos dados. Nossa proposta é que os perfis de uso acessem informações e serviços somente após autenticação, análise de sigilo e privacidade, além de filtros de informações privadas. Na Figura 6, apresentamos a camada Segurança.

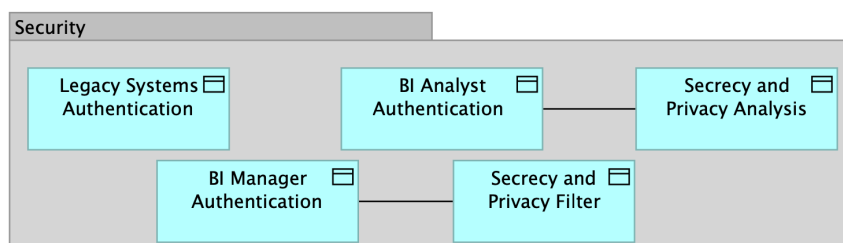


Figura 6. Camada Segurança

Embora exista uma camada dedicada à segurança, toda a arquitetura foi projetada com essa característica em mente, uma vez que todas as camadas lidam com dados confidenciais e sensíveis. Portanto, é essencial destacar que todos os aspectos relacionados à LGPD (Lei Geral de Proteção de Dados) são tratados. Desde o uso de VPNs (*Virtual Private Network*), que estabelecem conexões privadas virtuais criptografadas para proteger os dados em trânsito entre as partes envolvidas, até a adoção de protocolos HTTPS (*Hypertext Transfer Protocol Secure*), utilizados em módulos Web para assegurar a transferência segura de dados.

4.6. Papéis

O perfil *Users* representa pessoas, organizações ou outras entidades que utilizam o sistema para realizar tarefas, obter informações ou alcançar objetivos específicos. Dentre os tipos definidos de usuário do perfil *User*, existe um usuário administrador, que é responsável pela manutenção da infraestrutura computacional. É esse usuário que realiza a instalação e a configuração das bases de conhecimento, por exemplo. Ele também é responsável pela infraestrutura de rede local e pela VPN, essenciais para a comunicação de dados entre as partes interessadas. Além disso, esse usuário cuida da aplicação e da gestão dos certificados gerados para o protocolo HTTPS, utilizado nos módulos web, e disponibiliza o licenciamento adequado para que o *BI Analyst* possa utilizar os sistemas de BI, assim como possibilita o acesso, via autenticação, aos *dashboards* criados e disponibilizados para o *Manager*.

BI Analyst desempenha um papel crucial na implementação e manutenção de soluções de inteligência de negócios, sendo responsável por coletar, analisar e transformar dados brutos em informações significativas para apoiar a tomada de decisão, sendo este o responsável por, analisar, projetar, desenvolver e disponibilizar os *dashboards*.

Manager representa o responsável por supervisionar operações, coordenar ou operar componentes de gestão no sistema, além de tomar decisões estratégicas na organização. Na Figura 7, apresentamos a camada Papéis.

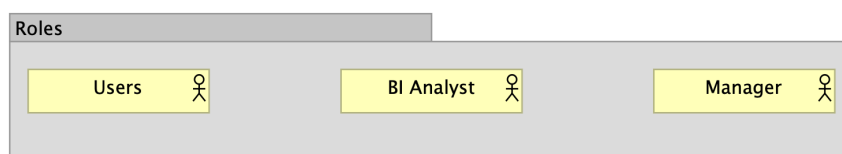


Figura 7. Camada Papéis

5. Estudo de Caso

CSA-IBILS surge como uma solução para enfrentar os desafios intrínsecos à interoperabilidade entre SLs e sistemas de BI. Esperamos, na aplicação da arquitetura conceitual, proporcionar uma experiência aprimorada no uso de *dashboards*, possibilitando que gestores e analistas de BI usufruam plenamente das capacidades de análise sem se preocupar com eventuais impactos negativos no desempenho do sistema legado.

Para verificar a aplicabilidade da proposta conceitual, um estudo de caso foi conduzido, considerando um contexto de mundo real onde opera um SL de uma *holding* de escolas que abrange desde o ensino infantil até o ensino médio. Atualmente, a *holding* possui 8 unidades escolares, contando com mais de 8 mil alunos. O SL foi adaptado e ajustado ao longo do tempo para atender às necessidades evolutivas da *holding* escolar. A Figura 8 apresenta um diagrama que ilustra a comunicação essencial entre as unidades escolares e a sede administrativa, com o objetivo de esclarecer o funcionamento desse processo.

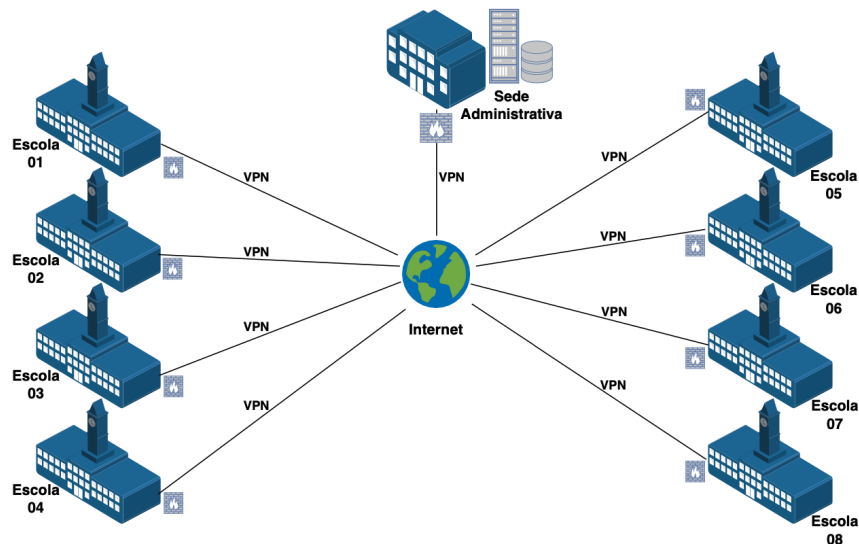


Figura 8. Diagrama de Rede

As 8 unidades escolares estão distribuídas em dois estados diferentes, e a sede administrativa abriga o *Legacy Database*, que é acessado pelas unidades através de uma VPN, possibilitando a utilização dos sistemas pelos usuários. No entanto, nem sempre todos os módulos da base de dados estiveram centralizados na sede administrativa. No início da utilização desse sistema, o módulo responsável pela secretaria ficava armazenado

localmente em uma máquina da escola, e os dados eram enviados ao módulo financeiro, que ficava em uma base de dados central, por meio de uma atualização diária.

Inicialmente, existiam apenas esses dois módulos, e surgiam diversos problemas nessa estratégia de interoperabilidade. O primeiro estava relacionado à segurança dos dados dos clientes, que eram armazenados localmente nas escolas, o que exigia um cuidado especial com a máquina que hospedava a base de dados, expondo-a a sérias vulnerabilidades, uma vez que não havia um local seguro para seu armazenamento. Além disso, na época, o professor precisava preencher manualmente as notas e as faltas dos alunos em uma ficha, que era posteriormente entregue à secretaria da escola para que os dados fossem inseridos no sistema, aluno por aluno.

Com o tempo, foi adicionado o módulo professor, baseado na Web, que permite aos professores inserirem diretamente os dados dos alunos no sistema. O *Legacy Database* passou por várias atualizações de sistema operacional e de versão de banco de dados até chegar à sua versão atual. Essas mudanças exigiram diversas adaptações na base de dados, resultando no acúmulo de inconsistências que comprometem sua eficiência. No entanto, era essa mesma base que precisava fornecer dados de qualidade por meio de *dashboards* de BI, com o objetivo de apoiar a tomada de decisões.

Para solucionar a problemática da interoperabilidade entre o SL e BI, foram implementadas duas bases de conhecimento fundamentais: *Legacy Database Digital Twin* e *BI Provider*. A base *Legacy Database Digital Twin* foi desenvolvida com o objetivo de replicar e sincronizar de forma contínua os dados do SL, possibilitando consistência, disponibilidade e qualidade para análises avançadas. Já a base *BI Provider* foi projetada para centralizar e otimizar o acesso a esses dados, fornecendo uma interface eficiente e escalável para o BI. Ambas as bases foram estruturadas para garantir a compatibilidade entre diferentes tecnologias, resolvendo assim o desafio da interoperabilidade e viabilizando a integração contínua entre o SL e BI.

Neste estudo de caso, para encurtar, chamaremos o Sistema de Gestão Escolar de *Legacy*. *Legacy* possui dois módulos *stand-alone* (Secretaria e Financeiro), além de um módulo Web destinado a professores para o registro de avaliações e faltas dos alunos. Chamaremos o Sistema de Consultas e Visualização de *Business Intelligence* de *BI Dashboard*.

A base de dados de *Legacy* possui uma estrutura estabelecida há aproximadamente 20 anos e ao longo desse período foram implementadas diversas modificações para suportar as crescentes demandas e as atualizações de versões do DBMS.

Na aplicação de CSA-IBILS no contexto proposto, iniciamos pela camada base da arquitetura, progredindo de forma ascendente pelas camadas subsequentes até alcançar a camada superior.

5.1. Bases de Conhecimento

Legacy Database utiliza o MS SQL Server 2016 (Microsoft Corporation; 2016b) como sistema gerenciador de banco de dados (DBMS). O DBMS opera em um ambiente de produção virtualizado utilizado XenServer (Citrix Systems; 2018) como repositório de máquinas virtuais (VM). *Legacy Database* possui 284 entidades e 52 *views*, estando estruturado para uso em todo o sistema *Legacy* e para apoio às diversas integrações acumuladas ao longo dos anos.

Legacy Database Digital Twin utiliza o SQL Server e está hospedada em uma VM separada, configurada com características idênticas à do *Legacy Database*. Essa configuração proporciona um ambiente isolado para a manipulação de forma contínua dos dados provenientes de *Legacy Database*, que consideramos como a fonte primária de dados. A *Legacy Database Digital Twin* atua como uma réplica com dados atualizados da base de dados primária, recebendo atualizações constantes através da rede local que conecta os servidores virtuais. Esta abordagem permite uma sincronização contínua entre as bases de dados, assegurando consistência e precisão das informações sem sobrecarregar a base de dados primária. A configuração utilizada, bem como os passos realizados para implementá-la, estão detalhados no Apêndice A (Microsoft Corporation; 2016a).

A base de dados *BI Provider* foi implementada em um ambiente de servidor virtualizado para atender às demandas do *BI Dashboard*. Esta base de conhecimento utiliza MongoDB (MongoDB Inc.; 2024) como DBMS, proporcionando uma estrutura flexível e escalável para armazenar os dados tratados provenientes de *Legacy Database Digital Twin*. A escolha do MongoDB como o DBMS para a base de dados *BI Provider* é estratégica, pois oferece a capacidade de lidar com grandes volumes de dados de forma eficiente e oferece suporte a esquemas dinâmicos. Dessa forma, *Legacy Database Digital Twin* atua

como uma ponte entre *Legacy Database* e *BI Provider*, possibilitando a entrega dos dados relevantes e adequados para análise no contexto de BI.

Para integrar os dados atualizados continuamente entre *Legacy Database Digital Twin* (SQL Server) e *BI Provider* (MongoDB), foi utilizado o Apache NiFi (Apache Software Foundation; 2024) como ferramenta central de ETL. CSA-IBILS aproveita a interface visual do NiFi para orquestrar fluxos de dados, extraindo dados do SQL Server através do componente QueryDatabaseTable. Este componente estabelece a conexão com o SQL Server e executa consultas SQL para extrair os dados desejados, permitindo também a configuração da frequência de execução das consultas para possibilitar a atualização constante no MongoDB. A transformação dos dados é realizada com o auxílio do JoltTransformJSON, que possibilita a realização de transformações complexas nos dados extraídos, como renomear campos, adicionar ou remover atributos.

Por fim, a gravação dos dados no MongoDB é realizada utilizando o PutMongo, que se conecta ao MongoDB e insere os dados transformados. Este processo possibilita a atualização contínua dos dados, com recursos de monitoramento e tratamento de erros que demonstram a confiabilidade da integração. A configuração, o passo a passo para a configuração do NiFi (responsável pela integração) e o processo de ETL entre SQL Server e MongoDB são detalhados no Apêndice B (Apache Software Foundation; 2024).

No MongoDB, foram criadas *Collections* específicas para cada *BI Dashboard*: i) Matrículas, ii) Alvo Financeiro, iii) Ticket Médio, iv) Controle de Bolsas, v) Alunos por Sala. Cada *Collection* armazena os dados relevantes para o respectivo *BI Dashboard*.

5.2. Componentes de Software

Os quatro componentes de software propostos operam de maneira independente, visando extrair o melhor de cada base de conhecimento.

Legacy DBMS é o mecanismo responsável pela manipulação de dados e informações na base de dados legada. *Digital Twin DBMS* é encarregado de manipular os dados relacionados à *Legacy Database Digital Twin*, fornecendo os dados selecionados e tratados por meio do processo ETL para o *BI Provider*.

A base de dados *BI Provider* fornece os dados necessários para o componente

Dashboard Generation and Construction Module operar. Este componente foi projetado com o objetivo de simplificar a criação de relatórios e *dashboards* destinados aos gestores, transformando dados em informações acessíveis para apoiar a tomada de decisão. Na implementação dos *dashboards*, foi utilizada a ferramenta MS Power BI (Microsoft Corporation; 2024), que oferece vários recursos, desde a integração de dados até a visualização avançada.

Operational Dashboard provê interfaces de visualização para os gestores, assegurando um acesso autorizado e seguro. As telas de visualização fornecem uma representação gráfica e interativa dos dados essenciais, permitindo aos gestores obterem *insights* precisos e atualizados. Existem atualmente cinco *dashboards*, a saber: i) Matrículas, que apresenta o número total de alunos matriculados, considerando todas as unidades, a quantidade de matrículas por unidade escolar e uma comparação com as matrículas do ano anterior; ii) Alvo Financeiro, que demonstra os valores estabelecidos pela instituição educacional como metas para atingir os orçamentos previstos para o ano letivo em andamento; iii) Ticket Médio, que apresenta os valores relacionados à média das mensalidades, de todas as escolas em conjunto e separadamente, por unidade escolar. Essa informação é relevante, uma vez que os valores variam de uma região para outra; iv) Controle de Bolsas, que apresenta informações sobre descontos concedidos aos estudantes, sendo fundamental para a gestão financeira da instituição educacional. Existe uma quantidade máxima de bolsas que a instituição consegue conceder sem afetar a operação do negócio. Essas bolsas são oferecidas por meio de um processo seletivo para tornar o procedimento transparente e atender às necessidades de quem realmente precisa desse auxílio; e v) Alunos por Sala, que monitora a capacidade de ocupação das salas de aula para garantir que o número de vagas oferecidas não exceda a disponibilidade das salas de aula.

No Apêndice C, são apresentadas as telas dos cinco *dashboards*, sendo os principais o de matrícula e o financeiro, exibidos no início do referido apêndice. Destacamos que esses dois *dashboards* são os que mais demandavam recursos do *Legacy*, devido ao grande volume de dados necessário para sua visualização. Vale ressaltar que informações sensíveis, como valores e nomes de unidades escolares, foram intencionalmente borradas ou alteradas, de modo a garantir o sigilo e impedir a identificação das escolas envolvidas.

5.3. Processos

Nesta seção, apresentamos a descrição dos processos integrados de CSA-IBILS. No âmbito de *Legacy Data Processing*, na atividade *Manipulate Data* ocorre a operação dos SL, que utilizam a base de dados legada. Neste ponto, ocorrem as interações dos usuários com o sistema no cotidiano, i.e., cadastros são realizados, consultas e relatórios são gerados, compreendendo a alimentação e manipulação dos dados na base de dados legada.

Legacy compreende três módulos com processos específicos: i) Secretaria, responsável por matrículas e rematrículas de alunos, cadastros de alunos, responsáveis, funcionários e professores, além do gerenciamento de disciplinas e turmas; ii) Financeiro, que abrange contas a receber e controle de inadimplência; e iii) Professor, responsável pelo gerenciamento de notas e faltas.

No processo subsequente, *BI Project* envolve tratamentos específicos e a criação de relatórios em *dashboard*, contendo informações relevantes para a tomada de decisão gerencial e análise de comportamento. Para execução de *BI Project*, é essencial atribuir os dados (*Assign Data*) necessários à geração do *BI Dashboard*. O processo se inicia com a atribuição dos dados ainda na base de dados legada, consultando *Legacy DBMS* para acessar os dados no *Legacy Database*. O próximo passo no processo é validar os dados a serem utilizados.

Na etapa de validação (*Validate Data*), os dados são consultados no *Legacy Database Digital Twin* por meio do componente de software *Digital Twin DBMS*. No último passo do processo, temos a geração efetiva do projeto de consulta ao *BI Provider*, a base de dados que alimenta as consultas de *BI Dashboard*.

Por fim, há o processo *Assessment*, no qual o gestor, munido de acesso ao *BI Dashboard*, realiza a requisição da informação necessária no momento. Para isso, os dados são processados e buscados em *BI Provider*. A informação é então visualizada por meio de um *dashboard* disponível no perfil do gestor.

5.4. Serviços

A camada Serviços é delineada por três elementos críticos para otimização do fluxo de informações e para tomada de decisão: *Legacy Data Manipulation*, *BI Project Creation* e *Data Assessment for Decision-making*.

Em *Legacy Data Manipulation*, os usuários se deparam com as operações cotidianas que frequentemente revelam os desafios inerentes aos SLs. Foi neste serviço que os usuários puderam perceber que o sistema apresentava atrasos incomuns no atendimento de requisições, independentemente de estas serem geradas diretamente por consultas de relatórios de BI ou por outras fontes. Atrasos e instabilidade eram causados por vários fatores, desde congestionamentos de solicitações até processos específicos, como a visualização de boletins escolares. Isso ilustra a complexidade envolvida no trabalho com sistemas que dependem de uma base de dados legada. Outros exemplos de serviços providos ilustram a complexidade dos módulos de *Legacy*: i) Secretaria: validação de documentos, geração de contratos, controle de vagas, geração de grade horária e histórico escolar; ii) Financeiro: emissão de boletos e relatórios de inadimplência; iii) Professor: lançamento de notas e faltas dos alunos por disciplina.

BI Project Creation é o ponto de partida para a elaboração de *dashboards* voltados a auxiliar os gestores na análise dos dados. Neste serviço, relatórios operacionais online são projetados e elaborados para permitir a análise e tomada de decisão por parte dos gestores. A ferramenta utilizada na criação dos projetos de BI foi o MS Power BI.

Em *Data Assessment for Decision-making*, o *BI Dashboard* está operacional e os dados foram devidamente tratados, proporcionando ao gestor uma ferramenta adequada para apoiar a tomada de decisão. Isso possibilita ao gestor dispor de informações precisas para tomar decisões informadas ou para analisar aspectos importantes de sua gestão com maior profundidade.

5.5. Segurança

A camada Segurança é implementada para atender às políticas de segurança e privacidade dos dados, garantindo o acesso restrito aos dados. Neste contexto, *Legacy Systems Authentication* é empregada como modelo de autenticação, amplamente utilizada em SL e atuais, incluindo aplicações Web e sistemas *desktop*, devido à sua facilidade

de gerenciamento de usuários. Este modelo é implementado por meio de uma tabela de usuários em *Legacy Database*, que armazena informações de logins e senhas criptografadas. O controle do nível de acesso de cada usuário é gerenciado pela tabela 'userauthorization'.

No contexto específico deste estudo de caso, trata-se de um sistema *desktop* executado exclusivamente em sistemas operacionais Windows nos computadores dos usuários. Além da autenticação, o acesso ao sistema educacional é restrito à base de dados legada, acessível apenas dentro da rede local ou, no caso de escolas, por meio de uma conexão VPN estabelecida entre a sede administrativa (local onde os servidores ficam alocados) e as unidades escolares.

O acesso do *BI Analyst* é realizado por meio de um MS Active Directory (AD) (Binduf et al.; 2018), utilizando autenticação de dois fatores. Dessa forma, o analista acessa os recursos do PowerBI Desktop, permitindo a criação, edição e publicação de *dashboards*. Ressaltamos que o acesso às informações da base de dados legada requer permissões específicas dentro do PowerBI no servidor local onde a base de dados está hospedada. Além disso, é necessário ter credencial com autorização para acessar os dados na base de dados legada, a qual possui apenas permissão de leitura.

Por fim, para que o *Manager* possa acessar os *dashboards* criados pelo *BI Analyst*, é necessário realizar uma autenticação no AD utilizando autenticação de dois fatores. Após a autenticação bem-sucedida, o *Manager* pode então acessar a aplicação do PowerBI por meio da página Web, onde terá acesso aos *dashboards* personalizados para sua visualização.

5.6. Papéis

Em CSA-IBILS, os papéis desempenhados são distintos e estrategicamente definidos. *User* representa os envolvidos nas operações diárias de *Legacy*. Esses usuários manipulam continuamente o *Legacy*, sendo responsáveis por inclusão, exclusão e edição de dados. Cada módulo de sistema possui perfis de usuários específicos. No módulo de Secretaria, apenas os funcionários da administração escolar (secretária e seus auxiliares) têm permissão para manipular dados cadastrais, gerar históricos e boletins de notas. Diretores, coordenadores e orientadores também acessam este módulo, mas com permissões de somente leitura. Da mesma forma, o módulo Financeiro é restrito aos funcionários da tesouraria escolar e o módulo Professor está disponível para todos os 335 professores das

8 unidades, com acesso gerenciado pelo departamento de secretaria de cada unidade.

BI Analyst foi responsável pela criação dos *BI Dashboards*, Matrículas, Alvo Financeiro, Ticket Médio, Controle de Bolsas e Alunos por Sala. Esses *dashboards* foram desenvolvidos com o propósito de apoiar o *Manager* no processo de tomada de decisão. *BI Analyst* criou um ambiente visual que apresenta métricas, indicadores e informações relevantes de maneira acessível, facilitando a análise de dados e possibilitando decisões mais assertivas por parte do *Manager*. Este profissional é capaz de interpretar as necessidades do *Manager* e traduzi-las em *BI Dashboards*.

Manager demanda uma visão abrangente e estratégica para embasar suas decisões, com acesso a informações precisas e dados relevantes e sem comprometer a operação eficiente do *Legacy*. O perfil *Manager* representa dois grupos de usuários das unidades escolares: i) *Administration Manager* (Financeiro e Gestor Educacional); e ii) *General Manager* (Diretores e Tesoureiros).

6. Discussão e Análise Sobre os Resultados do Estudo de Caso

Os resultados obtidos a partir da aplicação do CSA-IBILS no estudo de caso validam sua aplicabilidade em um contexto de mundo real.

Antes da aplicação de CSA-IBILS, observamos que problemas de travamento e lentidão de *Legacy* eram frequentes, ocorrendo todos os dias, entre 9 e 11 horas da manhã e entre 14 e 16 horas da tarde, durante o período de matrículas escolares (4 meses do ano). Nestes meses e horários, era programado que as consultas na base de dados legada entregassem as informações atualizadas das matrículas escolares de cada unidade no *dashboard*.

Após a aplicação de CSA-IBILS, consultas que eram realizadas diretamente em *Legacy Database* cessaram. A eliminação de interrupções e a redução da lentidão percebida para o usuário final do sistema legado durante as consultas de BI destacam a contribuição de CSA-IBILS para ambientes operacionais que envolvem interoperabilidade entre SL e sistemas de BI atuais.

Na Figura 9 apresentamos 2 gráficos, ilustrando (antes e depois) a interação entre o SL e BI. O primeiro gráfico mostra o número de falhas no SL ao longo de 24 meses (de 02/05/2022 a 31/05/2024). O segundo gráfico mostra a quantidade de consultas de BI ao banco de dados legado no mesmo período. Apesar do aumento das consultas de BI (Gráfico 2), podemos identificar no Gráfico 1 uma diminuição de falhas a partir do 18º mês (10/2023), acompanhando a implementação do CSA-IBILS no ambiente operacional.

A implementação bem-sucedida de CSA-IBILS contribuiu para a redução dos custos operacionais, uma vez que a manutenção e a integração dos sistemas se tornaram mais eficientes. Por exemplo, não é mais necessário deslocar um profissional para resolver os chamados das unidades escolares, o que possibilita direcionar o funcionário para atender a outras demandas. A interoperabilidade alcançada não apenas atende às demandas atuais, mas também prepara as organizações para enfrentar os desafios futuros no contexto da constante evolução da análise de dados.

As considerações práticas e técnicas incorporadas em CSA-IBILS fornecem uma base sólida para pesquisas e implementações adicionais no campo da integração entre BI e

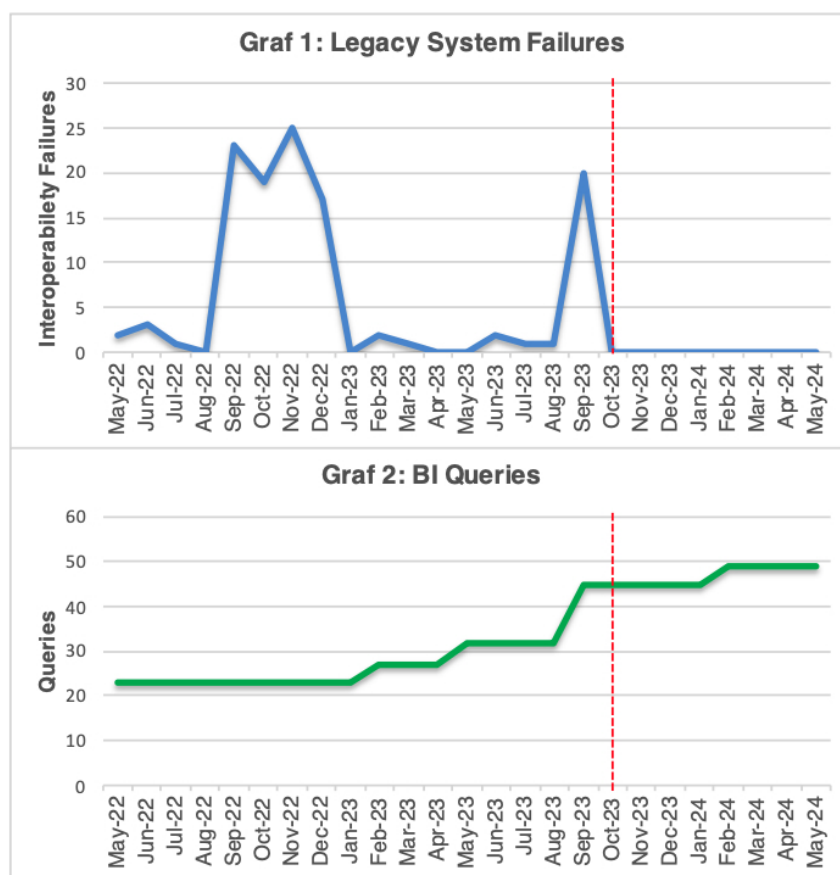


Figura 9. Cronologia de Falhas no Banco de Dados Legado e Consultas de BI

SL. Ressaltamos que a complexidade das infraestruturas empresariais pode variar, exigindo adaptações específicas. Por exemplo, a infraestrutura utilizada para validar em produção a arquitetura conceitual proposta empregou um ambiente virtual de servidores; no entanto, tal abordagem não é uma regra, e ambientes sem virtualização também podem se beneficiar de CSA-IBILS. Portanto, incentivamos pesquisas adicionais e experimentações práticas para validar e estender os resultados obtidos neste estudo.

Embora o estudo de caso tenha possibilitado a validação da aplicabilidade de CSA-IBILS, não observamos um ganho significativo de velocidade nas consultas para os *dashboards* operacionais testados. O tempo de resposta permaneceu praticamente inalterado. Neste estudo, não estávamos focados em analisar o desempenho dos sistemas. A principal preocupação era prover interoperabilidade com números reduzidos de falhas, e, nesse aspecto, CSA-IBILS apresentou resultados positivos, uma vez que as constantes reclamações dos usuários finais não estão sendo mais apresentadas. Eram abertos cerca de

3 chamados diários com reclamações de lentidão ou interrupção do sistema.

Como eventos de lentidão excessiva e interrupções pararam de acontecer, a melhora na qualidade do sistema foi percebida pelos usuários finais. Usuários finais apenas operam o sistema (e.g., realizam matrículas) e não utilizam *dashboards* de BI, representando a grande maioria dos usuários. É para esse público que o sistema não pode parar. Embora os *dashboards* de BI sejam importantes para os gestores, é imprescindível que o sistema esteja disponível e funcional para atender aos clientes, que são os pais e responsáveis dos estudantes. Todos os perfis de usuário dos três módulos de *Legacy* foram impactados positivamente em suas atividades diárias após a implementação de CSA-IBILS, pois as falhas de sistema foram reduzidas e o desempenho se manteve estável.

Em resumo, antes da implementação da arquitetura conceitual, especialmente durante períodos de grande volume de manipulação dos dados legados (e.g. durante o período de matrícula dos alunos), ocorriam interrupções diárias nos sistemas. Isso acontecia quando os relatórios de BI estavam em processo de consulta para fornecer as informações desejadas pelos gestores. Depois da aplicação de CSA-IBILS, os usuários não precisam mais esperar a restauração dos serviços após uma consulta de BI ao banco de dados do sistema legado. Essa melhoria substancial na experiência do usuário é o principal benefício sentido à partir da arquitetura conceitual implementada.

7. Conclusão

Diante dos desafios enfrentados ao integrar BI e SLs, nossa proposta de arquitetura conceitual de sistema surge como uma solução promissora. A necessidade de extrair dados críticos de SLs de maneira eficiente e sem interrupções motivaram este trabalho.

A arquitetura em camadas proposta apresenta um caminho inovador para superar os obstáculos enfrentados por muitas organizações. Ao fornecer uma estrutura que facilita a comunicação efetiva entre sistemas de BI e legados, a interoperabilidade é alcançada de maneira mais eficiente. CSA-IBILS visa não apenas melhorar a eficiência na extração de dados, mas também aprimorar o processo de manutenção e integração desses sistemas, eliminando barreiras que tradicionalmente causam lentidão e interrupções em SLs.

A implementação bem-sucedida de CSA-IBILS para interoperabilidade pode resultar em benefícios significativos, tais como, maior agilidade operacional do sistema legado e uma infraestrutura de informática mais resiliente. CSA-IBILS aborda lacunas significativas na integração entre BI e SLs, fornecendo uma base sólida para aprimorar a eficiência por meio da entrega de informações relevantes via *dashboards*. Essa capacidade de tomada de decisão, mesmo em ambientes que utilizam SLs, é de extrema importância para as organizações em um contexto cada vez mais orientado por dados.

7.1. Limitações e Trabalhos Futuros

A necessidade de alocação duplicada de recursos computacionais (hardware e software) pode ser considerada uma limitação para aplicação de CSA-IBILS quando implementada em ambientes que utilizam software proprietário. Além do servidor com *Legacy Database*, já provisionado no orçamento da organização, a aplicação requereu um segundo servidor para o gêmeo digital (*Legacy Database Digital Twin*) com recursos computacionais equivalentes. No estudo de caso, utilizamos Windows Server e MS SQL Server, além de recursos de hardware para as máquinas virtualizadoras. Para um desempenho ideal, dois ambientes idênticos (hardware e software) são recomendados, mas as organizações podem não possuir os recursos.

Outra limitação encontrada é a falta de experimentação da arquitetura conceitual para tratar outras implementações de bases de conhecimento, onde pode ocorrer a utilização

de bases de dados NoSQL já como *Legacy Database*. Nesse caso, todas as três bases de conhecimento seriam NoSQL. É necessário experimentar essa arquitetura em um número maior de contextos possíveis para obtermos uma noção mais precisa de sua aplicabilidade.

Como trabalhos futuros, planejamos evoluir CSA-IBILS incorporando elementos de *Cloud Computing*, incluindo *Fog and Edge Computing*. Por exemplo, implementar *Legacy Database Digital Twin* e *BI Provider* da camada Bases de Conhecimento diretamente na nuvem, visando a otimizar os recursos da rede local e eliminar a necessidade de alocar recursos de máquina para hospedar essas duas máquinas virtuais localmente. Dessa forma, as consultas de BI seriam realizadas diretamente na nuvem. Esse cenário permitirá avaliar CSA-IBILS em uma condição diferente, ampliando as possibilidades e o alcance dessa arquitetura conceitual.

Entre os desafios para empregar *Cloud Computing*, implementando *Legacy Database Digital Twin* e *BI Provider* na nuvem, destaca-se a infraestrutura necessária. Seria imprescindível trabalhar com uma conexão VPN para garantir o tráfego seguro dos dados. A vantagem desse modelo é que não seria necessário dispor de recursos de hardware localmente, tampouco adquirir novos servidores ou realizar atualizações nos equipamentos. No entanto, seria necessário um estudo aprofundado quanto aos custos, já que, dependendo da quantidade de dados trafegados, o custo de manter um servidor virtualizado na nuvem pode se tornar elevado. Por isso, há a necessidade de testar a arquitetura conceitual em outros contextos, de modo a explorar todas as possibilidades que possam surgir.

7.2. Resultados da Pesquisa

Como produção científica e tecnológica resultante da pesquisa, destacam-se três contribuições: i) publicação já realizada em uma conferência internacional, que apresenta a revisão de literatura; ii) manuscrito submetido a uma conferência internacional, que contém a proposta da arquitetura conceitual desenvolvida e implementada; e iii) registro de um programa de computador junto ao Instituto Nacional da Propriedade Industrial (INPI). A seguir, são apresentados maiores detalhes dessas contribuições:

1. **Artigo da Revisão da Literatura** (Bosse et al.; 2024b) - Publicado na trilha de Ciência de Dados da *21st International Conference on Information Technology - New Generations (ITNG 2024)*, classificada como Qualis CC B1. O *abstract* do

artigo é apresentado a seguir:

Abstract - Data quality is an increasingly important concern for organizations as their dependence on data to make decisions is growing. Business Intelligence methods and techniques are used by companies and institutions to collect, treat, organize, analyze, and view critical information in a structured and objective way. We present a literature review to identify approaches aimed at improving data quality to support decision-making in the context of Business Intelligence. Approaches involving data transformation, integration, and maintenance were identified, considering the specific decision-making requirements. Our review offers a comprehensive view of the latest strategies, presenting works that address issues regarding data quality and analysis.

- 2. Artigo da Arquitetura Conceitual** (Bosse et al.; 2024a) - Submetido para a 22nd International Conference on Information Technology - New Generations (ITNG 2025), que será realizada entre os dias 13 e 16 de abril de 2025, em Las Vegas, Nevada, Estados Unidos. Segue o *abstract* do artigo submetido:

Abstract - Business Intelligence is widely employed by companies and organizations to deliver critical data in an organized and accurate manner, facilitating data-driven decision-making. However, numerous legacy systems, characterized by complexity and challenges in maintenance and integration, pose a significant obstacle. Legacy systems often experience slowdowns and disruptions when executing BI queries on their databases. We present a Conceptual System Architecture aimed at fostering interoperability between BI and legacy systems. The architecture seeks to overcome existing limitations, providing seamless integration and minimizing adverse impacts on business operations.

- 3. Registro de Programa de Computador submetido ao INPI. O número do processo é BR1020240189752.**

Referências

- Abukwaik, H. & Rombach, D. (2017). Software interoperability analysis in practice: A survey, *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering*, EASE '17, Association for Computing Machinery, New York, NY, USA, p. 12–20.
URL: <https://doi.org/10.1145/3084226.3084255>
- Akhir, E. A. P., Aziz, N. & Roslin, A. F. (2018). Data visualization for evaluation form management system, pp. 117–122.
- Anwar, M. C. R. & Handayani, P. W. (2022). Continuous use evaluation of business intelligence implementation in energy company, pp. 247–252.
- Apache Software Foundation (2024). Apache nifi.
URL: <https://nifi.apache.org/>
- ArchiMate Tool (2024). Archi - archimate modelling.
URL: <https://www.archimatetool.com/>
- Arthur, A. L. & Mukherjee (2023). *Transforming Legacy Systems to Data Platforms*, Springer International Publishing, pp. 243–275.
URL: https://doi.org/10.1007/978-3-031-26401-6_11
- Benedictis, A. D., Flammini, F., Mazzocca, N., Somma, A. & Vitale, F. (2023). Digital twins for anomaly detection in the industrial internet of things: Conceptual architecture and proof-of-concept, *IEEE Transactions on Industrial Informatics* **19**: 11553–11563.
- Binduf, A., Alamoudi, H. O., Balahmar, H., Alshamrani, S., Al-Omar, H. & Nagy, N. (2018). Active directory and related aspects of security, pp. 4474–4479.
- Bosse, R. C., de Franco Rosa, F. & Jino, M. (2024a). A conceptual system architecture for interoperability between business intelligence and legacy systems - submitted, *ITNG 2025: 22nd International Conference on Information Technology-New Generations*, Springer Nature Switzerland.
- Bosse, R. C., de Franco Rosa, F. & Jino, M. (2024b). A study on data quality and analysis in business intelligence, in S. Latifi (ed.), *ITNG 2024: 21st International Conference on Information Technology-New Generations*, Springer Nature Switzerland, pp. 249–253.
- Bouchana, S. & Idrissi, M. A. J. (2015). Towards an assessment model of end user satisfaction and data quality in business intelligence systems, pp. 1–6.

- Büyüközkan, G., Mukul, E. & Güler, M. (2019). Business intelligence system selection with hesitant fuzzy linguistic mcdm methods, pp. 141–146.
- Caserio, C. & Trucco, S. (2018a). *Business Intelligence Systems*, Springer International Publishing, pp. 43–73.
URL: https://doi.org/10.1007/978-3-319-77679-8_3
- Caserio, C. & Trucco, S. (2018b). Erp and bi as tools to improve information quality in the italian setting: The research design, *Enterprise Resource Planning and Business Intelligence Systems for Information Quality: An Empirical Analysis in the Italian Setting* pp. 75–104.
- Citrix Systems (2018). XenServer.
URL: <https://www.citrix.com/pt-br/products/xenserver/>
- Costa, D. I. C., e Silva Filho, E. P., da Silva, R. F., de C. Quaresma Gama, T. D. & Cortés, M. I. (2020). Microservice architecture: A tertiary study, *Proceedings of the 14th Brazilian Symposium on Software Components, Architectures, and Reuse*, Association for Computing Machinery, pp. 61–70.
URL: <https://doi.org/10.1145/3425269.3425277>
- Egas, P. F. B., Saltos, M. A. G., Proaño, I. C., Recalde, H. & León, G. (2022). Strategic management for heis based on data analytics: literature review : Case study uisrael, pp. 1–6.
- Fernández, M., Dávila, A. & Angeleri, P. (2017). Data quality applied to an academic business intelligence solution: Lesson learned, pp. 1–6.
- Fraihat, S., Salameh, W. A., Elhassan, A., Tahoun, B. A. & Asasfeh, M. (2021). Business intelligence framework design and implementation: A real-estate market case study, *J. Data and Information Quality* **13**.
URL: <https://doi.org/10.1145/3422669>
- Freitas, F. & Alturas, B. (2018). Operational dashboard implementation for a family taxi business company, pp. 1–6.
- Fuller, A., Fan, Z., Day, C. & Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research, *IEEE Access* **8**: 108952–108971.
- Grigoriev, E. (2015). Dbms for business systems, Vol. 2, pp. 1–6.
- Grönman, J., Rantanen, P. & Saari, M. (2023). Categorization of digital twins: A literature review of iot and industry, pp. 27–32.

- Gutierrez, C. & Sequeda, J. F. (2021). Knowledge graphs, *Commun. ACM* **64**(3): 96–104.
URL: <https://doi.org/10.1145/3418294>
- Gyulgyulyan, E., Ravat, F., Astsatryan, H. & Aligon, J. (2018). Data quality impact in business intelligence, *2018 Ivannikov Memorial Workshop (IVMEM)*, pp. 47–51.
- Jha, S., Jha, M., O'Brien, L. & Wells, M. (2017). Supporting decision making with big data: Integrating legacy systems and data, pp. 120–128.
- Jozef, W. F. & Osmólski (2023). *Application of Big Data Analytics in Modern Logistics Solutions*, Springer International Publishing, pp. 37–47.
URL: https://doi.org/10.1007/978-3-031-15412-6_4
- Kerber, W. & Schweitzer, H. (2017). Interoperability in the digital economy, *Technical Report 12-2017*, MAGKS Joint Discussion Paper Series in Economics, Marburg.
URL: <https://www.uni-marburg.de/en/fb02/research-groups/economics/macroeconomics/research/magks-joint-discussion-paper-series>
- Kitchenham, B. (2004). Procedures for performing systematic reviews, *Keele, UK, Keele University* **33**: 1–26.
- Klisarova-Belcheva, S., Yankova, T. & Ilieva, G. (2019). Bi or no bi? presence and future of bi in small and medium-sized enterprises from south central bulgaria, Association for Computing Machinery.
URL: <https://doi.org/10.1145/3351556.3351575>
- Kristyanti, S. F., Kusumasari, T. F. & Alam, E. N. (2020). Operational dashboard development as a data quality monitoring tools using data deduplication profiling result, Vol. 1, pp. 1–6.
- Lakshen, G. A., Vraneš, S. & Janev, V. (2016). Big data and quality: A literature review, pp. 1–4.
- Lopes, F. C., Morais, P. & Lopes, A. C. (2022). Critical success factors for bi implementation in a portuguese higher education institution, Springer Nature Singapore, pp. 171–179.
- Maier, M. W. & Rechtin, E. (2009). *The Art of Systems Architecting*, 3 edn, CRC Press.
- Malaka, I. & Brown, I. (2015). Challenges to the organisational adoption of big data analytics: A case study in the south african telecommunications industry, Association for Computing Machinery.
URL: <https://doi.org/10.1145/2815782.2815793>

- Mario, A. T. & Fiduccia (2022). Geo-business intelligence and spatial data warehousing: A railway company case study, Springer International Publishing, pp. 141–155.
- McGraw, G. (2006). *Software Security: Building Security In*, Addison-Wesley.
- Microsoft Corporation (2016a). Alwayson availability groups sql server 2016.
URL: <https://learn.microsoft.com/en-us/sql/database-engine/availability-groups/windows/overview-of-always-on-availability-groups-sql-server?view=sql-server-ver16>
- Microsoft Corporation (2016b). MS SQL Server 2016.
URL: <https://www.microsoft.com/en-us/sql-server/sql-server-2016>
- Microsoft Corporation (2024). Microsoft power bi.
URL: <https://powerbi.microsoft.com/>
- MongoDB Inc. (2024). MongoDB.
URL: <https://www.mongodb.com/>
- Muppidi, A., Hashim, A. S. B. & Hasan, M. H. B. (2022). Proposed user-experience model for the design and development of bi dashboards, pp. 23–28.
- Newman, S. (2021). *Building microservices*, "O'Reilly Media, Inc."
- Omer, N., Jha, S. K. & Khatri, S. K. (2019). Maintaining reusable software components, pp. 1350–1352.
- Peng, M. Y.-P., Tuan, S.-H. & Liu, F.-C. (2017). Establishment of business intelligence and big data analysis for higher education, Association for Computing Machinery, pp. 121–125.
URL: <https://doi.org/10.1145/3134271.3134296>
- Prashneel, K. G. & Kumar (2023). Perceiving airline passenger booking lifecycle with the utilization of data warehouse, Springer Nature Singapore, pp. 241–258.
- Pressman, R. & Maxim, B. (2019). *Software Engineering: A Practitioner's Approach 9th Edition*.
- Ralf, Herbert, S. H. J. K. & Cimander (2011). *Interoperability in Government*, Springer Berlin Heidelberg, pp. 17–34.
URL: https://doi.org/10.1007/978-3-642-22502-4_2
- Sadeghi, M., Carenini, A., Corcho, O., Rossi, M., Santoro, R. & Vogelsang, A. (2023). Interoperability of heterogeneous systems of systems: Review of challenges, emerging requirements and options, *Proceedings of the 38th ACM/SIGAPP Symposium on Applied*

Computing, SAC '23, Association for Computing Machinery, New York, NY, USA, p. 741–750.

URL: <https://doi.org/10.1145/3555776.3577692>

Sandborn, P. A. & Prabhakar, V. J. (2015). The forecasting and impact of the loss of critical human skills necessary for supporting legacy systems, *IEEE Transactions on Engineering Management* **62**: 361–371.

Santi, R. P. & Putra, H. (2018). A systematic literature review of business intelligence technology, contribution and application for higher education, pp. 404–409.

Sena, B., Allian, A. P. & Nakagawa, E. Y. (2017). Characterizing big data software architectures: a systematic mapping study, *Proceedings of the 11th Brazilian Symposium on Software Components, Architectures, and Reuse*, Association for Computing Machinery.

URL: <https://doi.org/10.1145/3132498.3132510>

Sianipar, K. C., Wicaksana, S., Parikenan, B. & Hidayanto, A. N. (2019). Business intelligence critical success factors evaluation using analytical hierarchy process, pp. 1–6.

Sommerville, I. (2015). *Software Engineering*, 10th edn, Pearson.

Song, Y., Arnott, D. & Gao, S. (2018). *Business Intelligence System Use in Chinese Organizations*, Springer International Publishing, pp. 79–94.

URL: https://doi.org/10.1007/978-3-319-58097-5_7

Sumera, S. A. & Miskon (2020). The adoption of business intelligence systems in textile and apparel industry: Case studies, Springer International Publishing, pp. 12–23.

Turban, E., Sharda, R. & Delen, D. (2018). *Business Intelligence: A Managerial Approach*, Pearson.

Uzhga-Rebrov, O. & Grabusts, P. (2021). Comparative evaluation of four methods for exploratory data analysis, *2021 62nd International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)*, pp. 1–5.

Valle, P. H. D., Garcés, L. & Nakagawa, E. Y. (2019). A typology of architectural strategies for interoperability, *Proceedings of the XIII Brazilian Symposium on Software Components, Architectures, and Reuse*, Association for Computing Machinery, pp. 3–12.

URL: <https://doi.org/10.1145/3357141.3357144>

Verma, N. & Voids, A. (2016). On being actionable: Mythologies of business intelligence

and disconnects in drill downs, Association for Computing Machinery, pp. 325–334.

URL: <https://doi.org/10.1145/2957276.2957283>

Wilfried, S. G. & Rinderle-Ma (2015). *Data Provisioning*, Springer Berlin Heidelberg, pp. 87–118.

URL: https://doi.org/10.1007/978-3-662-46531-8_3

Wilson, M. L., Weaver, C. A., Procter, P. M. & Beene, M. S. (2017). *Big Data in Healthcare: A Wide Look at a Broad Subject*, Springer International Publishing, pp. 11–31.

URL: https://doi.org/10.1007/978-3-319-53300-1_2

Zineb, L. & Rachid, F. (2023). Etl technologies for big data: A comparative study, 2023 *IEEE International Conference on Advances in Data-Driven Analytics And Intelligent Systems (ADACIS)*, pp. 1–6.

A. *Digital Twin* - Configuração de Réplica de Banco de Dados com Atualização Contínua Utilizando SQL Server 2016

Este apêndice descreve o processo detalhado de configuração de uma réplica de banco de dados alimentada com atualização contínua utilizando o conceito de *Digital Twin* entre dois servidores SQL Server 2016. O recurso utilizado para essa configuração é o *AlwaysOn Availability Groups*.

Pré-requisitos:

- Dois servidores com SQL Server 2016 instalados.
- Certifique-se de que os dois servidores estejam no mesmo domínio.
- Ambos os servidores devem ter o SQL Server 2016 Enterprise Edition.
- Certifique-se de que os servidores possam se comunicar entre si.
- Configurar o Windows Server Failover Clustering.

Configuração do Windows Server Failover Clustering (WSFC):

- Instalar a funcionalidade de Failover Clustering nos servidores.
- Validar a configuração do cluster.
- Criar o Cluster.
- Adicionar um Nome de Rede para o Cluster.

Configuração do SQL Server AlwaysOn Availability Groups:

Habilitar o AlwaysOn Availability Groups em cada instância do SQL Server:

- Acessar o SQL Server Configuration Manager.
- Clicar em 'SQL Server Services'.
- Clicar com o botão direito no 'SQL Server (INSTANCENAME)' e selecione 'Properties'.
- Acessar a aba 'AlwaysOn High Availability' e marque 'Enable AlwaysOn Availability Groups'.
- Reiniciar o serviço SQL Server.

Criar um Availability Group: No SQL Server Management Studio (SSMS), conecte-se ao servidor primário. Acessar 'Object Explorer', expandir 'AlwaysOn High Availability', clicar com o botão direito em 'Availability Groups' e selecionar 'New Availability Group Wizard'. Seguir o assistente para criar um novo Availability Group:

- Dar um nome ao grupo.
- Adicionar o banco de dado que deseja replicar.
- Especificar o servidor secundário.
- Configurar o modo de failover e a sincronização de dados.
- Configurar os listeners e as propriedades adicionais conforme necessário.
- Verificar a configuração e finalizar a criação do grupo.

Adicionar Réplicas ao Availability Group: Na mesma tela de criação do grupo, é possível adicionar réplicas secundárias. Configurar a réplica para sincronização automática ou manual.

Verificação e Monitoramento: No SSMS, acessa 'Object Explorer', expandir 'AlwaysOn High Availability', expandir 'Availability Groups', selecionar o grupo e verificar o status das réplica.

Monitorar a replicação: Usar o *Dashboard* do 'AlwaysOn' para monitorar o estado da réplica e a integridade do grupo.

B. Integração com Atualização Contínua entre SQL Server e MongoDB utilizando Apache NiFi

Este apêndice descreve o processo de integração com atualização contínua entre o SQL Server e o MongoDB utilizando o Apache NiFi como ferramenta de ETL.

Configuração da Conexão com o SQL Server:

- Abra o Apache NiFi e acesse a interface de criação de fluxo de dados.
- Arraste o componente QueryDatabaseTable para o canvas.
- Configure a conexão com o SQL Server:
- Propriedade Database Connection Pooling Service: Configure um serviço de conexão de banco de dados para o SQL Server.
- Propriedade SQL Query: Defina a consulta SQL para extrair os dados desejados.
- Propriedade Max Rows Per Flow File: Defina o número máximo de linhas a serem extraídas por execução.
- Propriedade Query Interval: Especifique o intervalo de tempo para que a consulta seja executada continuamente e os dados sejam atualizados.

Transformação dos Dados:

- Arraste o componente JoltTransformJSON para o canvas e conecte-o ao QueryDatabaseTable.
- Configure o JoltTransformJSON:
- Propriedade Jolt Specification: Defina a especificação de transformação JOLT para manipular os dados extraídos, como renomear campos, adicionar ou remover atributos, ou reorganizar a estrutura JSON.

Inserção dos Dados no MongoDB:

- Arraste o componente PutMongo para o canvas e conecte-o ao JoltTransformJSON.
- Configure a conexão com o MongoDB:
- Propriedade Mongo URI: Forneça a URI de conexão do MongoDB.
- Propriedade Database Name: Especifique o nome do banco de dados no MongoDB.

- Propriedade Collection Name: Indique a coleção onde os dados serão inseridos.
- Propriedade Batch Size: Defina o número de documentos a serem inseridos por vez.

Monitoramento e Tratamento de Erros:

- Configure o MonitorActivity para monitorar a atividade do fluxo de dados e detectar possíveis falhas no processo.
- Adicione o componente LogMessage para registrar logs de erro e configurar alertas automáticos em caso de falha na integração.

Este processo possibilita a atualização contínua dos dados do SQL Server para o MongoDB, permitindo maior confiabilidade e eficiência no fluxo de dados.

C. BI Dashboards



Figura 10. Dashboard Matrículas

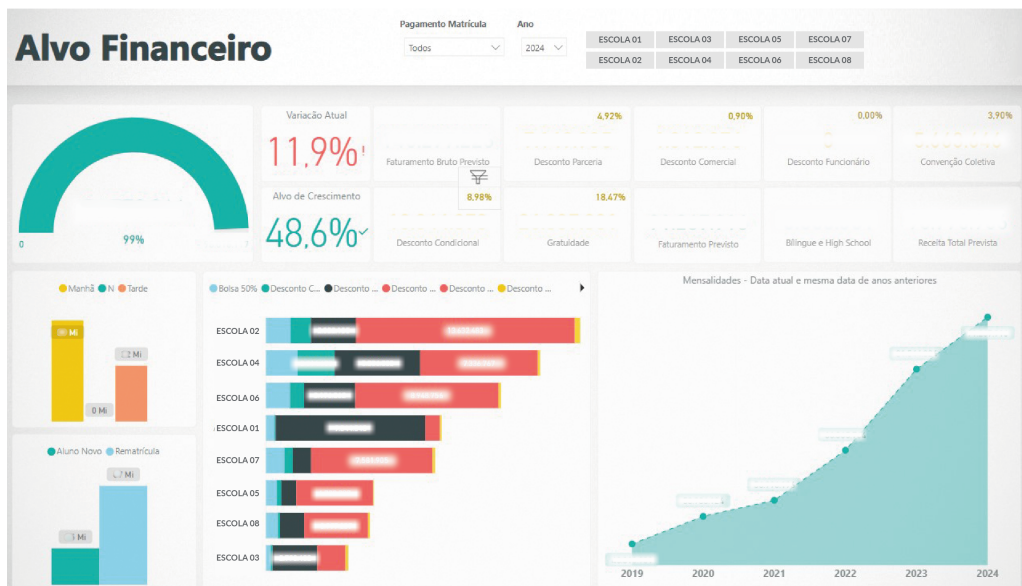


Figura 11. Dashboard Alvo Financeiro

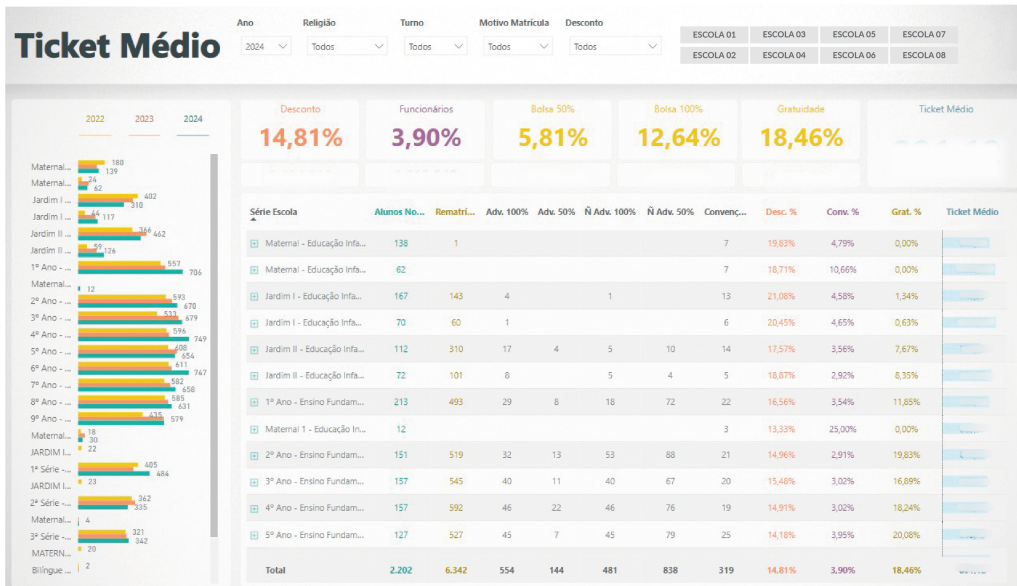


Figura 12. Dashboard Ticket Médio



Figura 13. Dashboard Controle de Bolsas

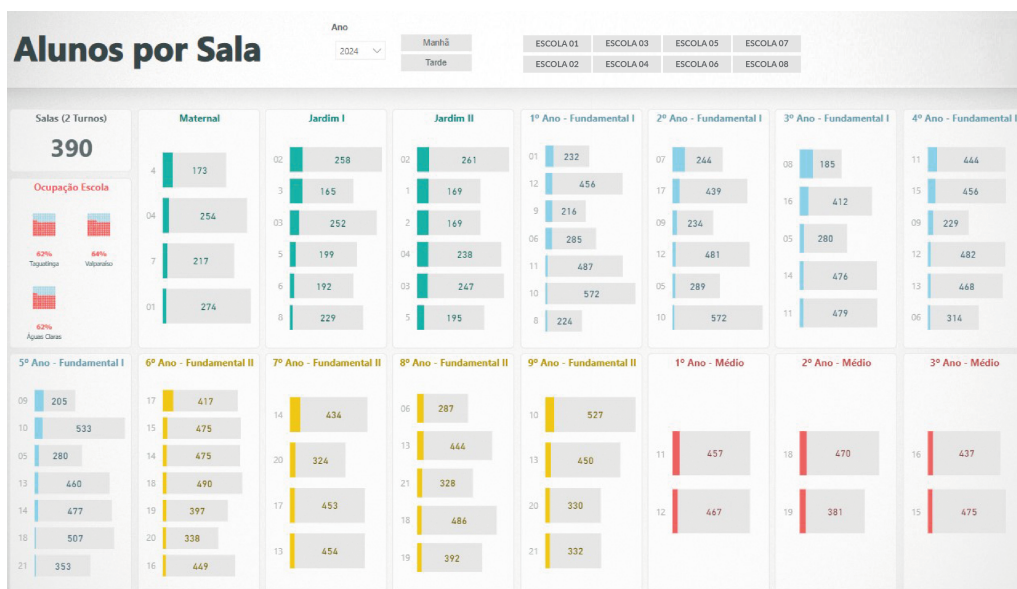


Figura 14. Dashboard Alunos por Sala