



Instituto Superior de Engenharia

Politécnico de Coimbra

DEPARTAMENTO DE INFORMÁTICA E SISTEMAS

Metodologias de ETL: resumo e comparação Sistemas de Informação II

Relatório de Licenciatura

Autores

Ana Rita Conceição Pessoa – 2023112690

David José Gomes de Matos – 2023143725



INSTITUTO POLITÉCNICO DE
COIMBRA

INSTITUTO SUPERIOR
DE ENGENHARIA
DE COIMBRA

Coimbra, outubro de 2025

1 ÍNDICE

1.1 Índice de texto

1	Índice.....	1
1.1	Índice de texto	1
1.2	Índice de figuras	1
2	Lista de siglas, acrónimos e símbolos.....	2
2.1	Lista de siglas e acrónimos	2
3	Resumo	3
4	Introdução	4
5	Contextualização e objetivos	6
6	Metodologia da revisão e critérios de seleção	8
7	Estado da arte sobre o tema	10
8	Análise e discussão.....	12
9	Limitações e trabalho futuro	16
10	Conclusões.....	19
11	Referências.....	22

1.2 Índice de figuras, quadros e afins

Figura 1:	Esquema geral do processo ETL.....	4
Figura 2:	Arquitetura de um Data Warehouse (modelo Kimball).....	6
Figura 3:	Esquema em estrela de um Data Warehouse.	12
Figura 4:	Exemplo de modelo físico de Data Mart com integração via ETL.....	14
Figura 5:	Pipeline ELT em ambiente cloud. <i>Microsoft (2025)</i> ,.....	17
Tabela 1:	Comparação entre as abordagens de integração de dados Lazy e Eager.	13
Tabela 2:	Comparação entre as metodologias ETL e ELT.....	20

2 LISTA DE SIGLAS, ACRÓNIMOS E SÍMBOLOS

2.1 Lista de siglas e acrónimos

ETL	Extract, Transform, Load
ELT	Extract, Load, Transform
DW	Data Warehouse
DM	Data Mart
BI	Business Intelligence
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
ERP	Enterprise Resource Planning
CRM	Customer Relationship Management
SCM	Supply Chain Management
SQL	Structured Query Language
SSIS	SQL Server Integration Services
SAS	Statistical Analysis System
ML	Machine Learning
RGPD	Regulamento Geral de Proteção de Dados
CIF	Corporate Information Factory
JSON	JavaScript Object Notation
XML	Extensible Markup Language

3 RESUMO

O processo de *Extract, Transform, Load* (ETL) constitui um elemento central na construção e funcionamento de um *Data Warehouse* (DW), garantindo a integração, consistência e qualidade dos dados utilizados para apoio à decisão¹. De acordo com os conteúdos teóricos, um DW é um repositório de dados integrado, orientado por temas, não volátil e variante no tempo, criado para consolidar informação proveniente de múltiplas fontes e apoiar a gestão estratégica².

O processo ETL é descrito como responsável por extrair dados de diferentes origens, transformá-los através de processos de limpeza, padronização e agregação, e carregá-los no DW de forma eficiente e fiável¹. Nos materiais das aulas, este processo é enquadrado na área de *Data Staging*, considerada a ponte entre os sistemas transacionais e o ambiente informacional¹.

São também abordadas duas metodologias de integração de dados:

- a abordagem *lazy* (*integração conduzida pelas interrogações*), que realiza a integração apenas quando os dados são solicitados, permitindo acesso direto às fontes²;
- e a abordagem *eager* (*integração conduzida pelo DW*), que prepara e armazena antecipadamente os dados para análise, otimizando o desempenho e a consistência²

O ETL assume, assim, um papel fundamental na transformação de dados brutos em informação estruturada e fiável, assegurando que os sistemas analíticos e de *Business Intelligence* disponham de bases consistentes para a tomada de decisão^{3,4}. As metodologias de ETL evoluíram acompanhando a necessidade de processar grandes volumes de dados e integrar múltiplas fontes em ambientes distribuídos e em nuvem, surgindo variações como o ELT e abordagens híbridas^{5,6}.

4 INTRODUÇÃO

A gestão eficiente de dados é um fator determinante para o sucesso das organizações modernas, que dependem de informação precisa e atual para apoiar a tomada de decisões estratégicas. Com o crescimento exponencial do volume e diversidade dos dados, tornou-se essencial adotar processos que garantam a sua integração, consistência e disponibilidade em tempo útil. Nesse contexto, surge o processo de *Extract, Transform, Load* (ETL), cuja principal função é extrair dados de múltiplas fontes, transformá-los em formatos adequados e carregá-los em repositórios analíticos, como o *Data Warehouse* (DW), onde poderão ser utilizados para análises e relatórios de apoio à decisão^{3,7}.

O *Data Warehouse* é definido como uma infraestrutura informacional que armazena dados integrados, orientados por temas, não voláteis e variantes no tempo, sendo o elemento central dos sistemas de apoio à decisão nas organizações.² A sua construção exige um processo sistemático que assegure a qualidade e coerência dos dados provenientes de sistemas operacionais heterogêneos, função desempenhada pelo ETL, o qual atua como uma ponte entre o ambiente transacional e o ambiente informacional.¹ Esse processo permite consolidar grandes volumes de informação em bases unificadas, promovendo a criação de conhecimento e o suporte a ferramentas de *Business Intelligence* (BI) e *Online Analytical Processing* (OLAP)⁸

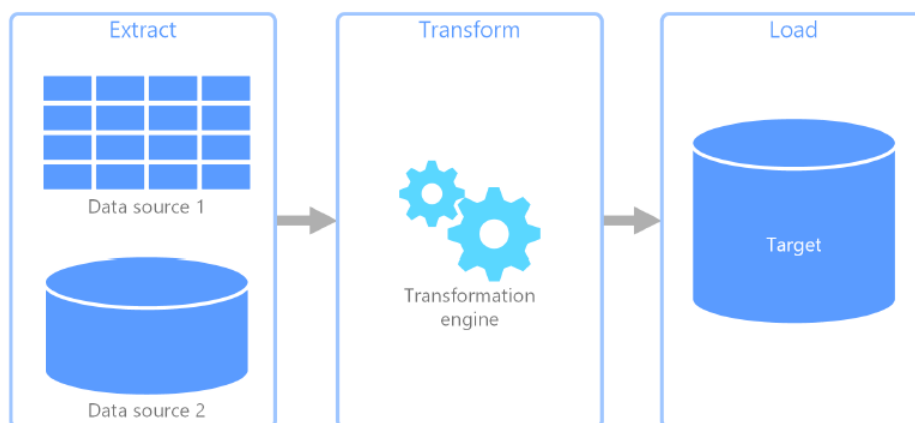


Figura 1: Esquema geral do processo ETL.

De Microsoft (2025). “*Extrair, transformar, carregar (ETL) – Azure Architecture Center*”, p.1.

De forma geral, o ETL é composto por três etapas fundamentais: **extração**, responsável por recolher dados de diferentes origens; **transformação**, que aplica regras de limpeza, padronização e enriquecimento dos dados; e **carga**, que insere os dados tratados no destino final^{9,10}. Durante a transformação, são realizadas operações que visam eliminar redundâncias, corrigir inconsistências e adaptar os dados aos modelos analíticos do DW¹¹. O processo é geralmente automatizado

através de ferramentas especializadas, que aumentam a eficiência e reduzem a probabilidade de erro humano¹².

A metodologia ETL evoluiu ao longo do tempo para acompanhar a crescente complexidade dos sistemas empresariais. Inicialmente, o processo era implementado em ambientes locais e executado em lotes (*batch processing*), mas, com a expansão da computação em nuvem, surgiram novas abordagens como o *Extract, Load, Transform* (ELT), que transfere parte das transformações para o ambiente do destino final^{5,6}. Esta evolução reflete a necessidade de escalabilidade, processamento distribuído e integração em tempo real, características essenciais em ecossistemas modernos de dados¹³

Para além de otimizar o desempenho analítico, o ETL é crucial para assegurar a governança e qualidade dos dados, permitindo que as organizações construam relatórios e indicadores de desempenho baseados em informação fidedigna. No contexto dos Sistemas de Informação, o domínio de metodologias de ETL é considerado uma competência estruturante, integrando o conjunto de tecnologias e práticas que sustentam o *Data Warehousing* e o *Business Intelligence*¹⁴. Assim, compreender e comparar diferentes metodologias de ETL constitui um passo essencial para projetar sistemas de dados eficazes e alinhados com as necessidades de negócio.

5 CONTEXTUALIZAÇÃO E OBJETIVOS

A crescente digitalização das organizações e o aumento do volume de dados gerados diariamente transformaram a forma como as empresas armazenam, processam e utilizam a informação. Os sistemas de informação modernos produzem grandes quantidades de dados operacionais, provenientes de fontes internas e externas, que necessitam de ser consolidados para gerar valor estratégico ². Nesse cenário, surge o conceito de *Data Warehouse* (DW) como uma infraestrutura essencial para reunir, organizar e disponibilizar informação relevante para a tomada de decisão ¹.

O *Data Warehouse* funciona como um repositório de dados integrado e centralizado, alimentado por processos de *Extract, Transform, Load* (ETL), que garantem a coerência e qualidade dos dados analisados. O ETL permite a recolha de dados provenientes de múltiplos sistemas, a sua transformação em formatos consistentes e a respetiva carga no DW, formando a base dos sistemas de apoio à decisão ^{3,7}. Ao automatizar estas etapas, as organizações conseguem reduzir o tempo de preparação de dados e melhorar a fiabilidade da informação utilizada nas análises ¹².

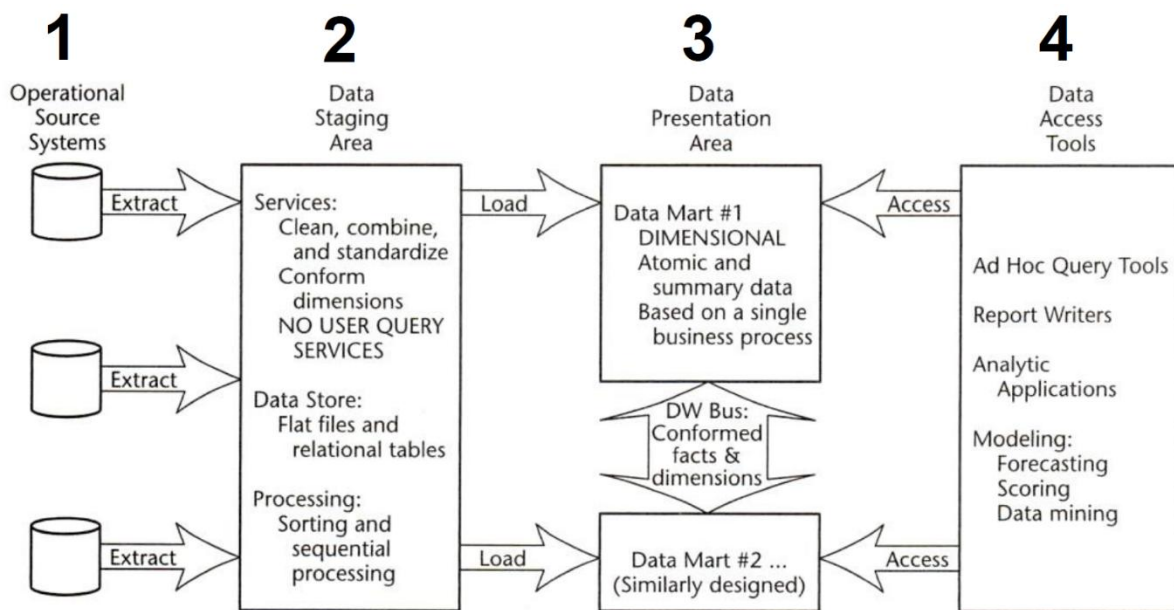


Figura 2: Arquitetura de um Data Warehouse (modelo Kimball).

Adaptado de Filipe Sá (2025), "SI – Parte 3 – DW – BI", diap. 8, com base em Kimball, R. (The Data Warehouse Toolkit).

Com a evolução tecnológica e a disseminação da computação em nuvem, o ETL passou a ser adaptado para novos contextos de armazenamento e processamento. As metodologias tradicionais de execução local deram origem a abordagens híbridas e distribuídas, que utilizam a capacidade escalável de plataformas *cloud* para processar grandes volumes de dados ^{6,13}. Paralelamente, surgiu o modelo *Extract, Load, Transform* (ELT), no qual a transformação ocorre após o carregamento dos dados no

destino final, aproveitando a potência computacional das bases de dados modernas⁵. Esta evolução reflete uma transição de metodologias centradas no desempenho local para soluções orientadas à agilidade e flexibilidade¹¹

A importância do ETL ultrapassa o domínio técnico: trata-se de um processo crítico para garantir a integridade, qualidade e rastreabilidade dos dados corporativos. A sua aplicação direta é visível em ambientes de *Business Intelligence* (BI) e *Online Analytical Processing* (OLAP), onde os dados consolidados são transformados em relatórios, indicadores e *dashboards* de apoio à decisão⁸. De acordo com o material teórico, o sucesso das plataformas de BI depende da fiabilidade do processo ETL, uma vez que dele depende a consistência dos dados que alimentam os sistemas analíticos¹.

No âmbito da unidade curricular Sistemas de Informação II, o estudo das metodologias de ETL permite compreender as diferenças entre os modelos clássicos e modernos de integração de dados, bem como a sua aplicação prática na construção de *Data Warehouses*. Assim, o presente trabalho tem como objetivo geral apresentar uma análise comparativa das principais metodologias de ETL, explorando as suas características, vantagens, limitações e adequação a diferentes contextos tecnológicos.

De forma mais específica, pretende-se:

1. Descrever as etapas fundamentais do processo ETL (extração, transformação e carga) e a sua importância para o funcionamento de um DW^{9,10};
2. Comparar metodologias de ETL quanto à arquitetura, desempenho e flexibilidade, incluindo a distinção entre ETL e ELT^{5,13};
3. Analisar a evolução das práticas de ETL face às novas exigências de integração de dados em ambientes *cloud* e *big data*^{3,13}
4. Identificar boas práticas e desafios comuns na implementação de processos ETL, tendo em vista a eficiência, qualidade e governança da informação^{12,15}.

Em síntese, este estudo contextualiza o ETL como componente estratégico do ecossistema de *Data Warehousing* e *Business Intelligence*, salientando a sua evolução metodológica e a relevância da sua correta implementação para a fiabilidade dos sistemas de apoio à decisão^{1,14}.

6 METODOLOGIA DA REVISÃO E CRITÉRIOS DE SELEÇÃO

A elaboração deste trabalho baseou-se numa revisão documental e comparativa, realizada a partir de fontes teóricas e técnicas fornecidas no âmbito da unidade curricular Sistemas de Informação II. Todo o processo de recolha e análise de informação seguiu uma abordagem sistemática, visando garantir a relevância, coerência e fiabilidade dos conteúdos utilizados. Para tal, foram considerados exclusivamente os materiais disponibilizados pelo docente e os documentos complementares sobre ETL anexados nesta investigação, assegurando a consistência metodológica e a fidelidade ao contexto académico do curso ¹⁴

A seleção das fontes obedeceu a critérios de pertinência temática, atualidade e aplicabilidade ao tema “*Metodologias de ETL: resumo e comparação*”. Assim, priorizaram-se os slides teóricos do professor Filipe Sá, que definem as bases conceptuais do *Data Warehousing*, *Business Intelligence* e *OLAP*, e contextualizam o papel do processo ETL na integração e consolidação de dados^{1,2,8}. Estes materiais permitiram compreender a estrutura e os fundamentos do ETL enquanto componente essencial do *Data Warehouse*, bem como as abordagens *lazy* e *eager* de integração de dados².

Paralelamente, foram incluídos documentos técnicos e institucionais de entidades de referência na área das tecnologias de informação, como IBM, Google Cloud, Microsoft Azure, AWS, SAS e Astera, que apresentam explicações atualizadas sobre as etapas e metodologias de ETL^{3,6,10,11,13,16}. Estes documentos foram selecionados por descreverem práticas amplamente adotadas no mercado, detalhando as fases de extração, transformação e carregamento de dados, bem como a sua aplicação em ambientes *cloud* e *big data*.

O processo de análise baseou-se em critérios qualitativos, centrados na comparação e convergência de conceitos entre as diferentes fontes. As informações recolhidas foram organizadas em função dos seguintes eixos:

1. Conceitos fundamentais de *Data Warehouse* e ETL: obtidos dos materiais teóricos do professor ^{1,2};
2. Definições e componentes técnicos do processo ETL: analisados a partir das publicações técnicas ^{7,9,12};
3. Comparação entre metodologias ETL e ELT, e suas adaptações a contextos de *cloud computing*: estudadas através das fontes da AWS, Microsoft e Google ^{3,5,6};
4. Aplicações práticas e ferramentas associadas ao ETL: sustentadas por exemplos de mercado e abordagens académicas ^{10,15}.

Cada fonte foi lida integralmente e analisada em função da sua relevância para o enquadramento teórico e comparativo do tema. As citações e conceitos comuns entre os documentos foram cruzados, permitindo identificar padrões metodológicos e diferenças na implementação de ETL entre contextos tradicionais e modernos.

Em termos metodológicos, a revisão privilegiou fontes primárias e oficiais, garantindo a credibilidade da informação. Não foram utilizadas fontes secundárias, redes sociais ou artigos de opinião, de forma a preservar o rigor académico e técnico da investigação. Assim, todo o conteúdo deste relatório reflete a síntese dos conceitos presentes nos materiais do docente Filipe Sá e nos documentos técnicos oficiais das plataformas tecnológicas analisadas ^{3,6,11,13}.

Em suma, a metodologia adotada baseou-se numa revisão documental comparativa, que integrou perspetivas académicas e industriais sobre o processo ETL. Esta abordagem permitiu construir uma visão abrangente das metodologias de extração, transformação e carregamento de dados, fornecendo a base teórica necessária para as análises desenvolvidas nos capítulos seguintes ^{1,4}.

7 ESTADO DA ARTE SOBRE O TEMA

O processo de *Extract, Transform, Load* (ETL) consolidou-se como um dos pilares das arquiteturas de *Data Warehousing* e *Business Intelligence*, desempenhando um papel fundamental na integração de dados provenientes de fontes heterogêneas. A literatura técnica e académica demonstra que o ETL é responsável por extrair dados brutos de diferentes sistemas, transformá-los em formatos consistentes e carregá-los em repositórios analíticos, garantindo qualidade, coerência e rastreabilidade da informação ^{9,11}.

De acordo com o enquadramento teórico apresentado pelo professor Filipe Sá, o Data Warehouse (DW) é um ambiente informacional estruturado que integra dados históricos e operacionais com o objetivo de apoiar a tomada de decisão ². As suas principais características (integração, orientação por temas, não volatilidade e variação temporal) definem o contexto em que o ETL atua, permitindo transformar dados dispersos em informação consolidada ¹. O ETL é, assim, a espinha dorsal do processo de construção do DW, assegurando que os dados extraídos de múltiplas fontes são submetidos a um ciclo de preparação antes de serem armazenados e explorados por ferramentas de análise.

Segundo as principais definições técnicas, o ETL é composto por três fases interdependentes:

1. Extração (Extract): recolha dos dados de diferentes sistemas de origem, como bases de dados, aplicações corporativas e ficheiros externos ^{3,7};
2. Transformação (Transform): tratamento, limpeza, padronização e enriquecimento dos dados, eliminando inconsistências e duplicações ^{12,16};
3. Carga (Load): inserção dos dados processados no destino final, que pode ser um *Data Warehouse*, *Data Mart* ou outro sistema analítico ^{1,10}.

Os materiais da Microsoft Azure e da AWS evidenciam que as metodologias de ETL evoluíram para responder às necessidades de escalabilidade e desempenho em ambientes *cloud* e *big data*. Enquanto o ETL clássico executa as transformações antes da carga, o modelo ELT (Extract, Load, Transform) inverte esta ordem, permitindo que as transformações sejam realizadas diretamente na infraestrutura de destino, aproveitando o poder computacional das plataformas modernas ^{5,6,13}. Esta mudança metodológica permite reduzir a latência, otimizar recursos e integrar dados em tempo quase real.

A documentação técnica da IBM e da Google Cloud reforça que o ETL moderno incorpora princípios de automação, orquestração e monitorização contínua, com mecanismos que asseguram a rastreabilidade e o controlo de qualidade em todo o ciclo de vida dos dados ^{3,11}. Estes sistemas utilizam pipelines configuráveis, capazes de lidar com fluxos massivos de dados estruturados e não estruturados, além de oferecerem suporte a formatos variados como JSON, XML e CSV. Tais

características tornam o ETL um componente central na governança de dados e na implementação de estratégias de integração entre sistemas empresariais.

Nos conteúdos lecionados em Sistemas de Informação II, o processo ETL é ainda associado à área de *Data Staging*, onde ocorre a manipulação e preparação dos dados antes do carregamento no DW ¹. Este estágio intermedeia a comunicação entre os sistemas operacionais e o ambiente analítico, garantindo que as operações de extração e transformação não impactem o desempenho das bases de dados transacionais. O material também destaca que um ETL eficaz é essencial para a fiabilidade dos sistemas de *Business Intelligence* e *OLAP*, que dependem diretamente da consistência e da atualidade dos dados ⁸.

A análise dos documentos técnicos revela ainda a diversificação das ferramentas de ETL disponíveis no mercado, tais como SAS, Talend, Pentaho e *SQL Server Integration Services (SSIS)*, que suportam processos automatizados de integração e carregamento ^{10,15}. Estas ferramentas diferem na arquitetura e na metodologia de execução, podendo operar em modo *batch*, incremental ou em tempo real.

O estado da arte atual demonstra que o ETL deixou de ser um processo meramente técnico para se tornar uma metodologia estratégica que suporta a tomada de decisões empresariais baseadas em dados. A sua evolução acompanha as transformações digitais e o advento da computação distribuída, sendo hoje essencial em ambientes de *Data Warehousing*, *Data Lakes* e *Business Intelligence* ^{3,13}. Assim, compreender e comparar as diferentes metodologias de ETL (clássicas e modernas) é indispensável para projetar soluções eficazes de integração de dados e garantir a qualidade e o valor da informação corporativa ^{2,4}.

A revisão dos documentos técnicos e dos materiais teóricos demonstra uma convergência entre as abordagens académicas e as práticas do mercado. Enquanto os conteúdos do professor Filipe Sá descrevem a estrutura funcional do *Data Warehouse* e o papel do ETL no seu ciclo de vida ¹, as fontes técnicas apresentam as adaptações das metodologias às novas realidades tecnológicas. Esta articulação entre teoria e prática constitui a base para a análise comparativa das metodologias de ETL, permitindo avaliar como os modelos clássicos e modernos respondem aos desafios atuais da integração e qualidade dos dados ^{3,11}.

8 ANÁLISE E DISCUSSÃO

O processo de *Extract, Transform, Load* (ETL) representa o elo entre os sistemas operacionais e as plataformas analíticas das organizações, garantindo que os dados utilizados na tomada de decisão sejam completos, consistentes e atualizados. A sua aplicação é essencial no contexto dos *Data Warehouses*, cuja principal finalidade é consolidar e disponibilizar informação proveniente de diferentes fontes para apoiar a análise e o planeamento estratégico ^{1,2}.

Do ponto de vista técnico, o ETL é um processo estruturado em três fases principais: extração, transformação e carregamento. Na fase de extração, os dados são recolhidos a partir de múltiplas fontes (bases de dados relacionais, sistemas ERP e CRM, ficheiros planos ou plataformas *cloud*) assegurando que a recolha não comprometa o desempenho das aplicações de origem ^{3,13}. Durante a transformação, ocorre a limpeza, normalização e integração dos dados, incluindo a conversão de tipos, eliminação de duplicados e padronização de formatos ^{11,12}. Por fim, na etapa de carregamento, os dados transformados são armazenados em sistemas analíticos como *Data Warehouses* ou *Data Marts*, tornando-se disponíveis para consultas e análises ^{1,10}.

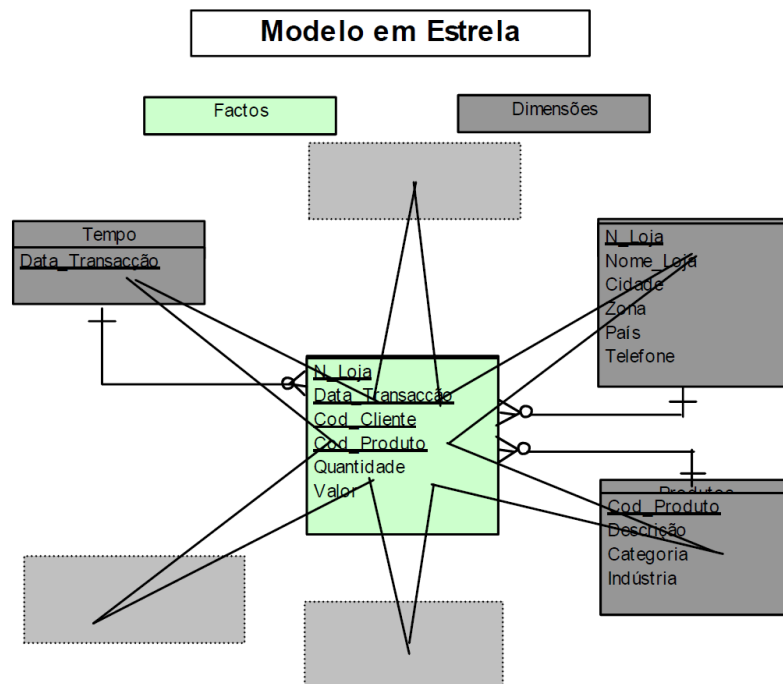


Figura 3: Esquema em estrela de um Data Warehouse.

Filipe Sá (2025), “SI – Parte 4 – DW – OLAP”, diap. 30–33, com base em Santos & Ramos (2017).

As metodologias de ETL podem variar de acordo com o momento e o local em que as transformações são executadas. O modelo ETL clássico realiza as transformações antes do carregamento, garantindo que apenas dados limpos e formatados cheguem ao repositório final ^{7,9}. Este método é eficiente em ambientes locais e sistemas de médio porte, mas apresenta limitações quanto à escalabilidade e ao processamento em tempo real. Em contrapartida, o modelo ELT (Extract, Load, Transform), amplamente difundido em infraestruturas *cloud*, realiza a carga antes das transformações, explorando o poder computacional das bases de dados modernas para otimizar o desempenho ^{5,6}. Essa abordagem é mais flexível e adequada a volumes elevados de dados, permitindo análises quase em tempo real ¹³.

Nos materiais teóricos da unidade curricular, são ainda referidas duas abordagens conceituais de integração de dados (*lazy* e *eager*) que influenciam diretamente as metodologias de ETL². A abordagem *lazy*, ou integração conduzida pelas interrogações, realiza a integração apenas quando os dados são solicitados, resultando em maior atualidade da informação, mas menor desempenho. Já a abordagem *eager*, ou integração conduzida pelo *Data Warehouse*, antecipa o processamento e armazenamento dos dados, assegurando consistência e rapidez na execução de consultas. Estas duas perspectivas encontram correspondência direta nos modelos ETL (pré-processamento) e ELT (processamento pós-carga), demonstrando a evolução natural das práticas de integração de dados.

Tabela 1: Comparação entre as abordagens de integração de dados Lazy e Eager.
Baseado em Filipe Sá (2025), “SI – Parte 2 – DW”, diáp. 19–20.

Abordagem	Descrição	Vantagens	Desvantagens
Lazy (Integração conduzida pelas interrogações)	Integra os dados apenas quando são solicitados, acedendo diretamente às fontes.	Dados sempre atuais; Sem duplicação de dados.	Processamento lento; Dependência das fontes; Perda de dados históricos.
Eager (Integração conduzida pelo Data Warehouse)	Dados preparados e integrados antecipadamente no DW.	Alto desempenho; Não interfere com as fontes; Permite consultas complexas no DW.	Duplicação de dados; Ligeiros atrasos na atualização.

As ferramentas de ETL desempenham um papel determinante na eficiência do processo. Segundo as fontes técnicas, aplicações como SAS Data Integration, SQL Server Integration Services (SSIS) e Talend permitem automatizar as tarefas de extração, transformação e carga, integrando fluxos de dados complexos de forma escalável e segura^{10,15}. Estas ferramentas também incluem mecanismos de monitorização e controlo de erros, fundamentais para garantir a rastreabilidade e a qualidade dos dados carregados no DW¹⁶.

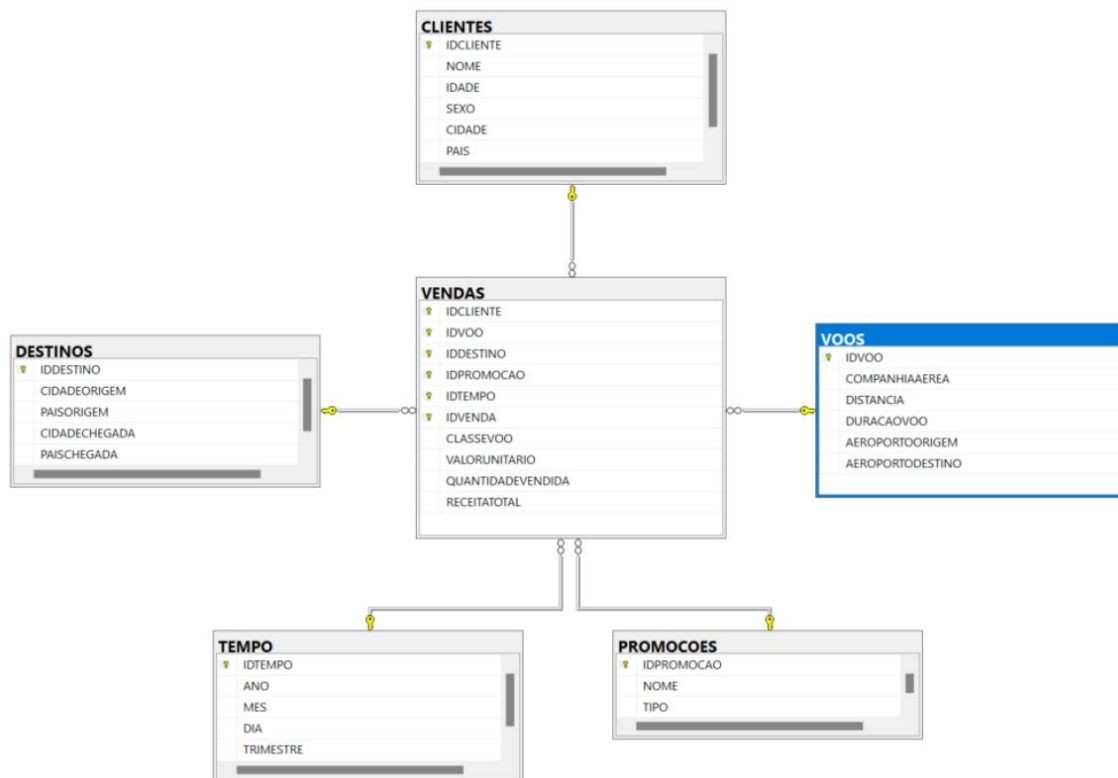


Figura 4: Exemplo de modelo físico de Data Mart com integração via ETL.
Filipe Sá (2025), “SI – Parte 5 – Exemplo Completo DW”, diáp. 8.

Do ponto de vista comparativo, o ETL tradicional destaca-se pela sua robustez, previsibilidade e controlo da qualidade dos dados, sendo ideal para processos periódicos e estruturados⁴. No entanto, o seu processamento em lote (*batch*) pode limitar a atualização em tempo real. Já o ELT, ao tirar partido da computação distribuída, apresenta maior escalabilidade e flexibilidade, adaptando-se melhor a arquiteturas *cloud-native*⁵. Contudo, essa abordagem exige maior capacidade de processamento e mecanismos adicionais de segurança e governança de³.

A análise das fontes também evidencia que as metodologias de ETL estão em convergência com as práticas modernas de integração contínua de dados, aproximando-se de conceitos como *DataOps* e *Streaming ETL*. Estas variações incorporam fluxos em tempo real, pipelines automatizados e integrações diretas com

ferramentas de análise, tornando o processo mais dinâmico e responsivo às necessidades do negócio ^{6,11}

Em termos de impacto organizacional, o ETL é um elemento crítico para a qualidade e governança da informação. Um processo mal definido pode introduzir redundâncias, atrasos e inconsistências que comprometem a confiabilidade das análises ^{1,12}. Assim, a seleção da metodologia adequada deve considerar fatores como volume de dados, frequência de atualização, infraestrutura disponível e requisitos de desempenho.

As ferramentas de ETL diferem não apenas nas suas funcionalidades, mas também nas metodologias que implementam. Soluções consolidadas, como o SQL Server Integration Services (SSIS) e o SAS Data Integration, seguem predominantemente o modelo ETL clássico, no qual as transformações são processadas antes da carga final ^{10,15}. Por outro lado, plataformas modernas como Google Cloud Dataflow e AWS Glue aplicam princípios de ELT, permitindo o processamento distribuído e escalável dentro das próprias infraestruturas *cloud* ^{3,13}. Esta diversidade de ferramentas evidencia que a escolha da metodologia depende não apenas do tipo de dados, mas também da infraestrutura tecnológica e dos objetivos de negócio de cada organização.

Em síntese, a análise comparativa demonstra que tanto o ETL tradicional como o ELT moderno são indispensáveis em contextos específicos. O primeiro é mais adequado a ambientes controlados e processos periódicos; o segundo, a sistemas distribuídos e integrações de larga escala. Ambos convergem no objetivo comum de transformar dados brutos em informação de qualidade, promovendo a tomada de decisões informadas e o alinhamento estratégico das organizações ^{3,8}.

9 LIMITAÇÕES E TRABALHO FUTURO

Embora o processo de *Extract, Transform, Load* (ETL) seja amplamente reconhecido como essencial para a integração e gestão de dados empresariais, a análise das fontes evidencia diversas limitações técnicas e operacionais que podem comprometer a sua eficiência e adaptabilidade. Uma das principais limitações está relacionada com o desempenho e escalabilidade dos processos tradicionais de ETL, especialmente em ambientes com grandes volumes de dados e múltiplas fontes heterogêneas ^{4,11}. Os processos executados em modo *batch* podem gerar atrasos no carregamento e dificultar a atualização em tempo real, tornando-se menos adequados para arquiteturas dinâmicas e distribuídas ³.

Nos materiais teóricos, o ETL é apresentado como um processo estruturado e sequencial, cuja eficácia depende da integridade dos dados de origem e do correto desenho do *Data Warehouse* ¹. No entanto, alterações nas fontes de dados, falta de normalização e erros de transformação podem comprometer a consistência do resultado final ². Além disso, a manutenção de pipelines complexos requer supervisão constante, aumentando os custos operacionais e a dependência de equipas especializadas ^{10,12}.

Outro desafio identificado nas fontes técnicas prende-se com a governança e a segurança dos dados. O aumento da integração de dados provenientes de sistemas *cloud* e externos expõe as organizações a riscos acrescidos de conformidade e privacidade ^{3,13}. A ausência de mecanismos robustos de auditoria e rastreabilidade pode dificultar a identificação da origem de erros e a reconstrução do histórico de transformações ¹⁶. Assim, a implementação de políticas de controlo de acesso e de encriptação, bem como a documentação detalhada das regras de transformação, são medidas apontadas como indispensáveis para mitigar estes riscos ¹¹.

Em termos metodológicos, a transição entre o modelo ETL clássico e o modelo ELT moderno também apresenta limitações. Apesar de o ELT oferecer maior flexibilidade e desempenho em ambientes *cloud*, ele depende fortemente da capacidade computacional e da arquitetura da base de dados de destino ^{5,6}. Além disso, as transformações complexas realizadas diretamente nas plataformas *cloud* podem aumentar os custos de processamento e dificultar a otimização de recursos ¹³.

As fontes teóricas e práticas analisadas sugerem que as principais oportunidades de evolução do ETL concentram-se na sua integração com tecnologias emergentes. O avanço das plataformas *cloud-native* e dos ambientes *big data* impulsiona o desenvolvimento de ETL automatizado e inteligente, baseado em princípios de *Machine Learning* e *DataOps* ^{3,11}. Estas abordagens prometem pipelines mais resilientes e autoadaptáveis, capazes de identificar anomalias e otimizar o desempenho em tempo real ⁶.

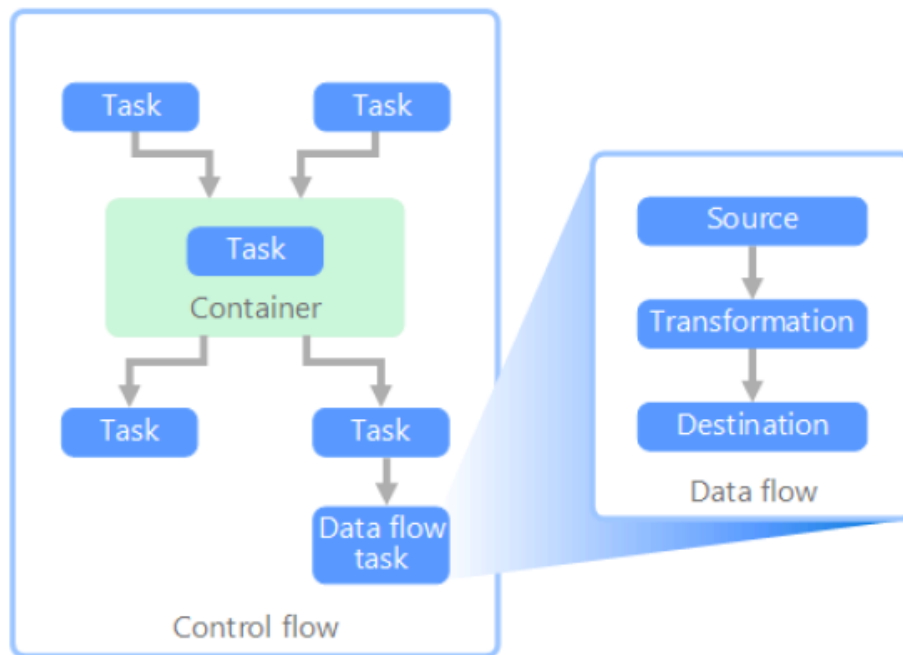


Figura 5: Pipeline ELT em ambiente cloud.

Microsoft (2025), “*Extract, transformar, carregar (ETL) – Azure Architecture Center*”, p.4.

Adicionalmente, o uso de ETL em tempo real (*Streaming ETL*) é apontado como uma tendência relevante, especialmente em contextos que exigem análise imediata, como comércio eletrônico, detecção de fraudes e monitorização de redes ^{3,13}. Esta evolução implica a reformulação das metodologias tradicionais, substituindo os ciclos de carga periódicos por fluxos contínuos de ingestão e transformação de dados.

Do ponto de vista acadêmico e profissional, é necessário aprofundar a comparação entre metodologias ETL e ELT em diferentes cenários de implementação, avaliando o impacto de fatores como volume de dados, tipo de processamento, complexidade das transformações e custo operacional ^{5,12}. Também se sugere o desenvolvimento de estudos aplicados sobre ferramentas open-source e soluções híbridas que combinem o melhor dos dois paradigmas, possibilitando maior flexibilidade e interoperabilidade entre sistemas ^{10,15}.

Por fim, os trabalhos futuros poderão explorar a integração do ETL com técnicas de automação e monitorização contínua, visando aumentar a eficiência e reduzir o esforço de manutenção. A adoção de pipelines inteligentes, sustentados por métricas de desempenho e alertas automatizados, constitui um passo natural na modernização dos processos de *Data Warehousing* e *Business Intelligence* ^{3,11}.

Em síntese, as limitações identificadas (relacionadas com desempenho, segurança, custo e complexidade) reforçam a necessidade de inovação contínua nas metodologias de ETL. O futuro aponta para soluções mais ágeis, inteligentes e

integradas, capazes de responder às exigências de ambientes de dados cada vez mais distribuídos e em tempo real ^{1,3,6}.

Um outro aspeto relevante para trabalhos futuros consiste em aprofundar a relação entre as metodologias de ETL e a governança de dados, especialmente no que respeita à conformidade com o Regulamento Geral de Proteção de Dados (RGPD) e às políticas de rastreabilidade e auditoria. A crescente integração de dados pessoais e sensíveis em sistemas *cloud* torna essencial o desenvolvimento de pipelines que garantam segurança, privacidade e transparência em todo o ciclo de vida da informação ^{3,11}. Este enfoque ético e regulatório representa uma evolução natural do processo ETL, alinhando as práticas técnicas às exigências legais e sociais de proteção da informação.

10 CONCLUSÕES

A análise das metodologias de *Extract, Transform, Load* (ETL) permitiu compreender o papel central que este processo desempenha na integração e preparação de dados para os sistemas de *Data Warehousing* e *Business Intelligence*. Através das fontes teóricas e técnicas estudadas, conclui-se que o ETL constitui a base sobre a qual se constrói a confiabilidade da informação utilizada nas organizações, garantindo que os dados recolhidos de diferentes fontes sejam transformados em conhecimento útil para apoio à decisão ^{1,2}

O ETL tradicional, ao realizar as transformações antes do carregamento, destaca-se pela sua robustez, controlo e fiabilidade, sendo adequado a ambientes com grande necessidade de validação e limpeza de dados ^{7,9}. No entanto, este modelo apresenta limitações em termos de desempenho e escalabilidade, especialmente em contextos de grandes volumes de dados e em infraestruturas *cloud* ^{3,11}.

A evolução tecnológica levou ao surgimento de novas metodologias, como o ELT (*Extract, Load, Transform*), que inverte a ordem das etapas clássicas. Neste modelo, as transformações são executadas diretamente nas bases de dados de destino, aproveitando a sua capacidade computacional e o paralelismo disponível em plataformas *cloud* ^{5,6}. O ELT apresenta-se, assim, como uma solução mais flexível e escalável, adequada a ambientes distribuídos e à integração de dados em tempo real ¹³.

A comparação entre ETL e ELT demonstra que não existe uma metodologia universalmente superior, mas sim abordagens adequadas a diferentes cenários. O ETL mantém relevância em sistemas que exigem processos de validação rigorosos e integração estruturada; já o ELT é mais vantajoso em contextos que privilegiam a agilidade, a escalabilidade e a análise imediata dos dados ^{3,12}. Ambas as metodologias convergem no mesmo objetivo: assegurar que os dados corporativos sejam limpos, coerentes e prontos para análise, fortalecendo a tomada de decisão e a estratégia organizacional ⁸.

Tabela 2: Comparação entre as metodologias ETL e ELT.
Baseado em “O que são ETL e ELT e quais suas diferenças” (Coodesh, 2025).

Aspeto	ETL (Extract, Transform, Load)	ELT (Extract, Load, Transform)
Ordem das etapas	A transformação dos dados ocorre antes do carregamento para o destino final.	A transformação ocorre após o carregamento, já dentro do sistema de destino (por exemplo, um <i>Data Warehouse</i> ou <i>Data Lake</i>).
Velocidade de processamento	Tende a ser mais lenta, pois a transformação é feita num servidor intermédio antes da carga.	É mais rápida, aproveitando o poder computacional das plataformas <i>cloud</i> e das bases de dados modernas.
Manutenção e infraestrutura	Exige mais recursos e manutenção, devido à existência de uma camada de processamento dedicada.	Requer menos manutenção, pois a transformação é executada na infraestrutura de destino.
Maturidade tecnológica	É uma metodologia consolidada e amplamente utilizada em sistemas empresariais tradicionais.	É uma abordagem mais recente e em expansão, adaptada a ambientes distribuídos e <i>cloud-native</i> .
Tipo de dados suportado	Trabalha melhor com dados estruturados e integrados em fontes relacionais.	Suporta dados estruturados, semi-estruturados e não estruturados, como JSON, XML ou logs.
Volume e escalabilidade	Adequado para volumes de dados moderados, em ambientes locais ou híbridos.	Altamente escalável, ideal para grandes volumes de dados e arquiteturas <i>big data</i> .

As fontes técnicas analisadas indicam ainda que o futuro do ETL está fortemente associado à automação, inteligência artificial e integração contínua de dados ^{3,11}. As novas abordagens, como *Streaming ETL* e *ETL em tempo real*, permitem que os dados sejam processados e carregados de forma contínua, reduzindo a latência e aumentando a capacidade de resposta das organizações ^{6,13}.

Além disso, os materiais do professor Filipe Sá reforçam que o ETL é uma competência estruturante no domínio dos Sistemas de Informação, sendo fundamental para o desenvolvimento e manutenção de *Data Warehouses* eficazes ^{14,15}. A compreensão das suas metodologias, ferramentas e boas práticas é essencial para garantir a qualidade dos dados e o sucesso dos sistemas de apoio à decisão.

Em síntese, o estudo realizado demonstra que as metodologias de ETL evoluíram de processos estáticos para pipelines inteligentes e distribuídos, acompanhando as exigências de um mundo orientado por dados. O ETL clássico continua relevante pela sua estabilidade e rigor, enquanto o ELT e as suas variações modernas refletem a adaptação necessária à era da computação em nuvem e da análise em tempo real

^{3,13}. A tendência atual aponta para soluções híbridas, capazes de combinar o melhor de ambas as abordagens, promovendo um equilíbrio entre desempenho, qualidade e flexibilidade.

Assim, compreender e comparar as diferentes metodologias de ETL é fundamental não apenas do ponto de vista técnico, mas também estratégico, pois constitui o alicerce da qualidade da informação e da eficácia dos sistemas de *Business Intelligence* e *Data Warehousing* ^{1,4}.

11 REFERÊNCIAS

1. Sá, F. *SI - Parte 3 - DW - BI*.
2. Sá, F. & Pt, F. S. *SI - Parte 2 - DW*.
3. Google Cloud. O que é ETL? | Google Cloud. <https://cloud.google.com/learn/what-is-etl>.
4. Thiago Coutinho. *ETL: Entenda o Que é, Sua Importância, Etapas e Aplicações*. <https://voitto.com.br/blog/artigo/etl-extracao-transformacao-e-carga> (2023).
5. Felipe Rastelli. O que são ETL e ELT e quais suas diferenças? <https://coodesh.com/blog/candidates/o-que-sao-etl-e-elt-e-quais-suas-diferencas/>.
6. Microsoft Learn. *Extrair, Transformar, Carregar (ETL) – Azure Architecture Center*. <https://learn.microsoft.com/pt-pt/azure/architecture/data-guide/relational-data/etl>.
7. Gabriel Sacramento. *Entenda o Que é ETL e Sua Finalidade Na Transformação de Dados*. <https://blog.somostera.com/data-science/etl-processo>.
8. Sá, F. *SI - Parte 4 - DW - OLAP*.
9. Benedito Gaia. O Básico sobre ETL. *Digital Innovation One (DIO)* <https://www.dio.me/articles/o-basico-sobre-etl> (2023).
10. SAS Institute Inc. Ferramentas e processos ETL | SAS. https://www.sas.com/pt_pt/insights/data-management/what-is-etl.html.
11. IBM Corporation. *O Que é ETL (Extract, Transform, Load)?* <https://www.ibm.com/br-pt/think/topics/etl>.
12. Diego Dias. *ETL: A Importância Da Preparação de Dados*. <https://www.preditiva.ai/blog/etl-a-importancia-da-preparacao-de-dados>.
13. Amazon Web Services (AWS). *O Que é ETL? – Explicação Sobre Extrair, Transformar e Carregar*. <https://aws.amazon.com/pt/what-is/etl/>.
14. Sá, F. & Pt, F. S. *SI - Parte 0 - Apresentação*.
15. Sá, F. *SI - Parte 5 - Exemplo Completo DW*.
16. Equipa de Marketing da Astera. *O Que é ETL? - Extração, Transformação e Carregamento Explicados*. <https://www.astera.com/pt/type/blog/etl/> (2025).



**Instituto Superior
de Engenharia**

Politécnico de Coimbra