

# CURSO SUPERIOR DE ADS

## Tecnologias de BI: OLAP e ETL



Prof. Fernando Marlon Soares Figueiredo

Disciplina: Ciência de Dados e Bigdata



# Introdução

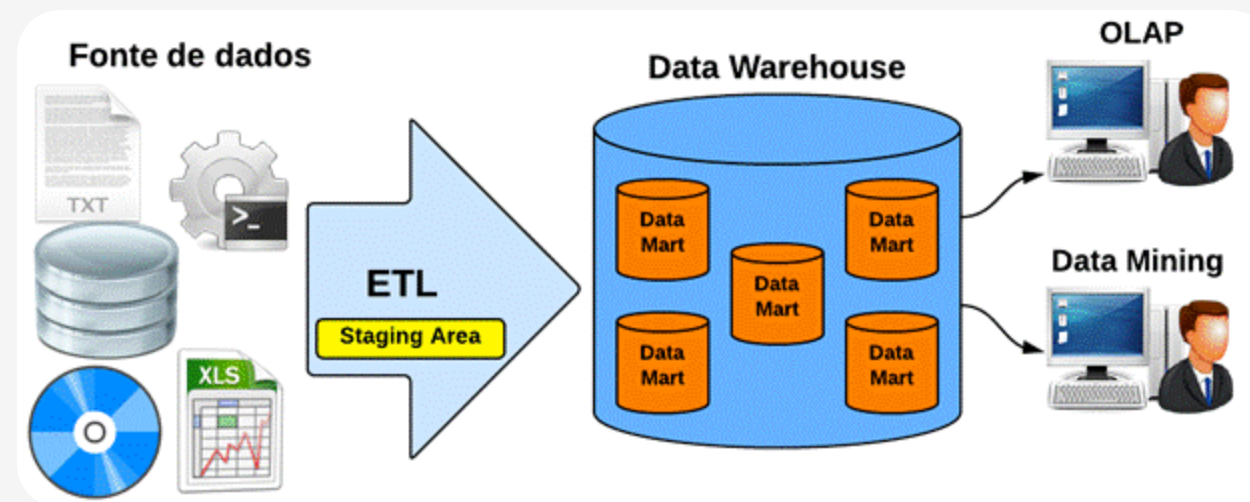
O Business Intelligence depende de um conjunto de **tecnologias que tornam possível transformar dados brutos em informações úteis**.

Duas delas são fundamentais: **ETL (Extract, Transform, Load)** e **OLAP (Online Analytical Processing)**.

O ETL é o processo responsável por **extrair, transformar e carregar dados** para o ambiente analítico.

Já o OLAP permite **analisar esses dados de forma multidimensional**, criando uma visão mais detalhada e dinâmica dos resultados.

Compreender essas tecnologias é essencial para entender como os dados fluem dentro de um sistema de BI e como se tornam disponíveis para as ferramentas de visualização.



# Contextualização

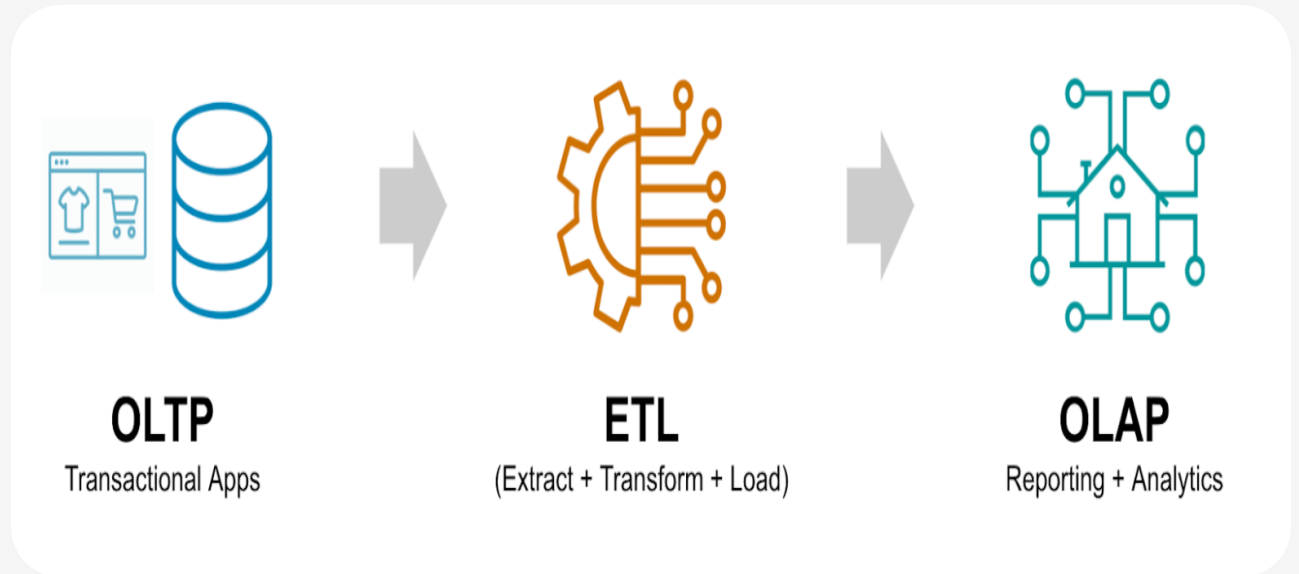
Nos últimos anos, a quantidade de dados gerados pelas empresas aumentou drasticamente.

Sem processos adequados, esses dados permaneceriam desorganizados, em múltiplas fontes, e sem valor analítico.

O **ETL** e o **OLAP** surgem como soluções complementares:

- O ETL **prepara e organiza** os dados.
- O OLAP **permite analisar** esses dados de maneira eficiente e interativa.

Essas tecnologias são a base do que chamamos de **pipeline de BI**, o caminho que os dados percorrem desde sua origem até a tomada de decisão.



Transactional Apps

OLTP

(Extract + Transform + Load)

ETL

Reporting + Analytics

OLAP

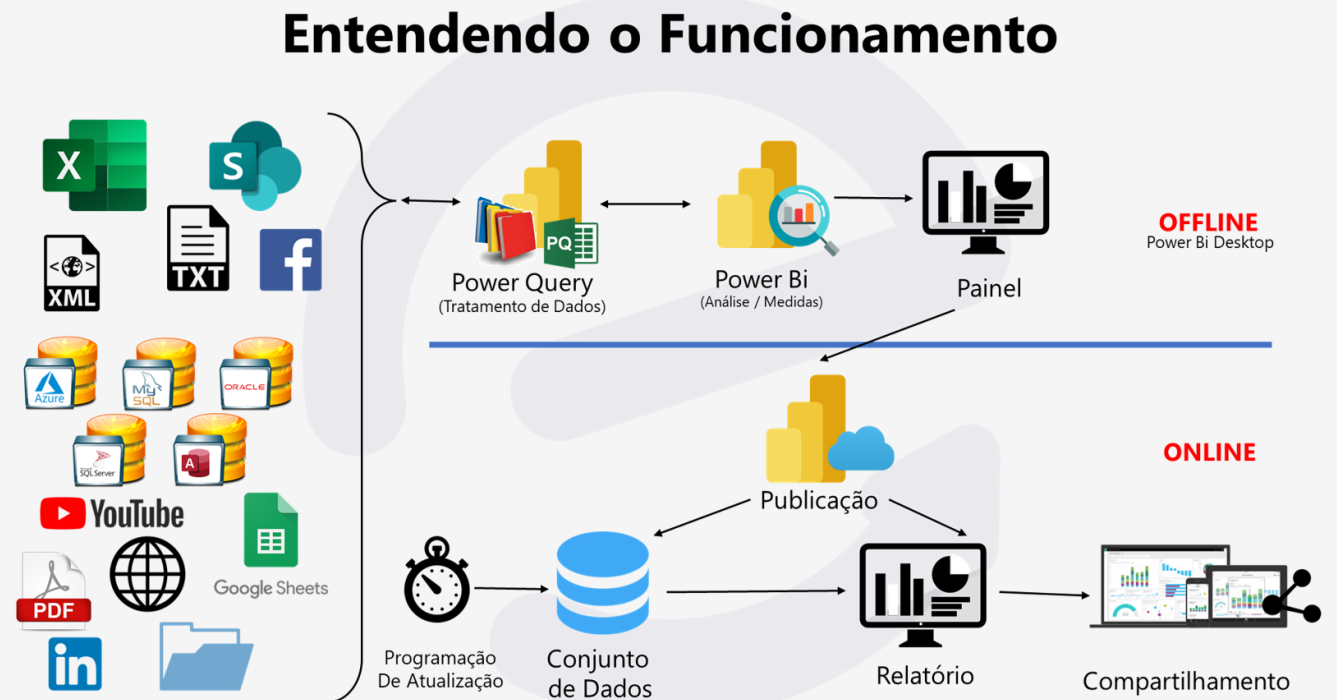
# Revisando o Fluxo do BI

Para entender o papel do ETL e do OLAP, é importante visualizar o fluxo completo de um sistema de BI:

Fontes de Dados → ETL → Data Warehouse → OLAP → Visualização (Power BI / Tableau)

Cada etapa cumpre uma função específica:

- **Fontes de Dados:** sistemas transacionais (ERP, CRM, planilhas, APIs).
- **ETL:** coleta, limpa e transforma os dados.
- **Data Warehouse:** repositório central organizado.
- **OLAP:** estrutura os dados em cubos e dimensões.
- **Visualização:** apresenta os resultados em gráficos e dashboards.



# O que é ETL (Extract, Transform, Load)

O processo de ETL é responsável por preparar os dados para análise.

Sem ele, os dados permaneceriam inconsistentes, duplicados ou incompatíveis entre sistemas.

O ETL é dividido em três etapas principais:

## 1. Extract (Extração):

É o processo de coletar dados de diversas fontes planilhas, bancos de dados, APIs, arquivos CSV, entre outros.

O desafio aqui é garantir que os dados sejam extraídos de forma íntegra e sem perdas.

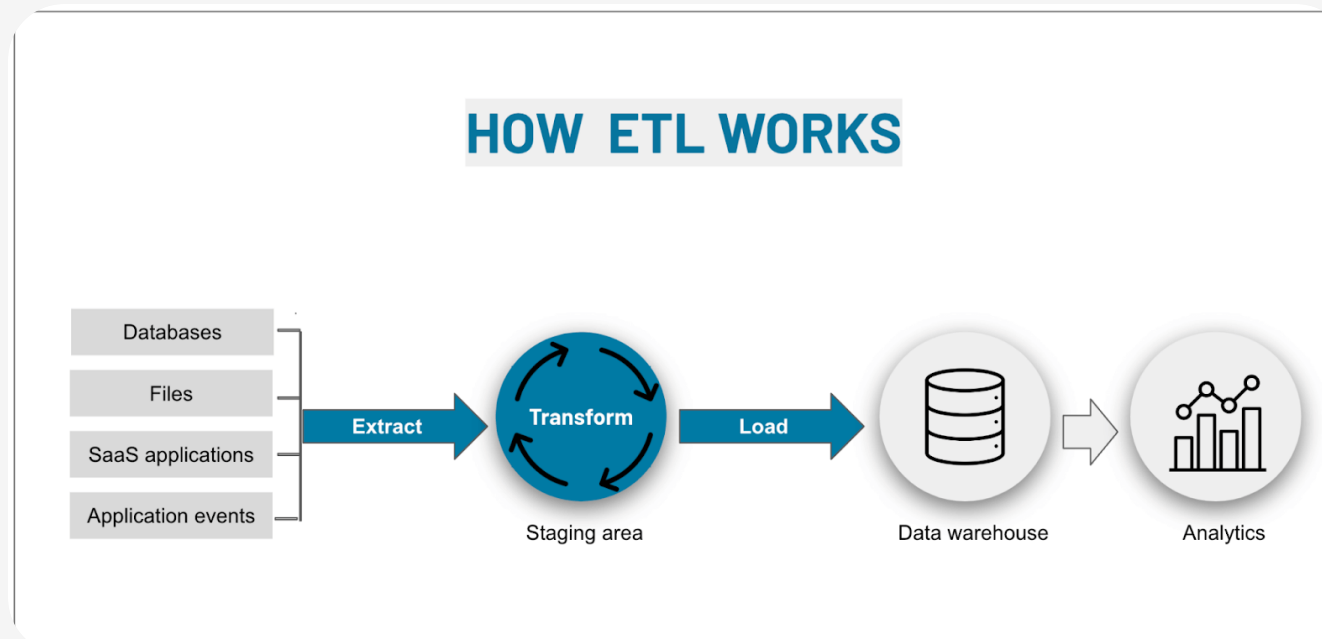
## 2. Transform (Transformação):

Nesta etapa, os dados são **limpos, convertidos e padronizados**.

Isso inclui remover valores nulos, corrigir erros, alterar formatos e criar novas colunas derivadas (como totais, médias ou classificações).

## 3. Load (Carregamento):

Finalmente, os dados transformados são **carregados para o Data Warehouse**, onde serão organizados em tabelas e dimensões para análise posterior.



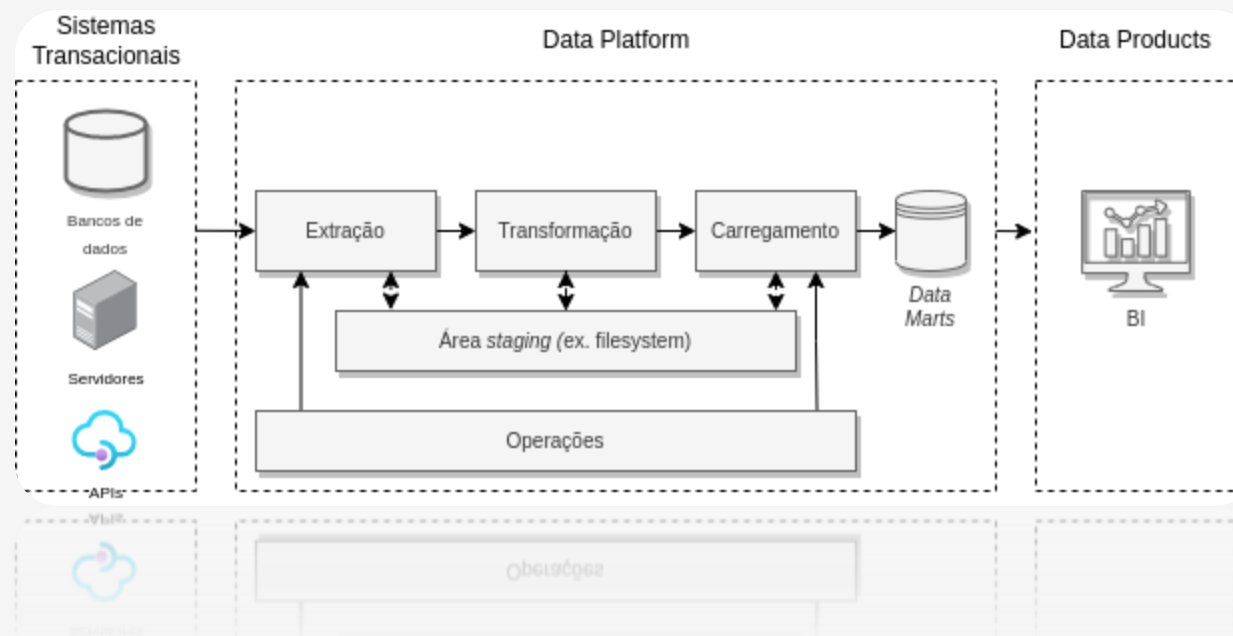
# Exemplo de Processo ETL

Imagine uma empresa que possui dados de vendas em diferentes planilhas (por região).

O processo de ETL poderia funcionar assim:

- **Extração:** importar as planilhas "Vendas\_Nordeste.xlsx", "Vendas\_Sul.xlsx" e "Vendas\_Sudeste.xlsx".
- **Transformação:** padronizar os nomes das colunas, converter moedas, eliminar duplicatas e calcular o total por produto.
- **Carga:** enviar o resultado final para a tabela "Vendas\_Gerais" dentro do Data Warehouse.

Esse processo garante que o BI trabalhe com **dados consolidados e confiáveis**, prontos para serem usados em dashboards e relatórios.



# Ferramentas de ETL

Existem diversas ferramentas que automatizam o processo de ETL, desde soluções corporativas até versões open-source.

## Exemplos populares:

**Pentaho Data Integration (Kettle):** ferramenta open-source amplamente usada em ambientes educacionais e corporativos.

**Talend:** solução completa de integração de dados, com interface gráfica para montar fluxos ETL.

**Informatica PowerCenter:** ferramenta corporativa robusta, muito utilizada em grandes empresas.

**Microsoft SQL Server Integration Services (SSIS):** parte do ecossistema Microsoft, ideal para integração com o Power BI.

**Apache NiFi:** solução open-source voltada para fluxo de dados em tempo real.

Essas ferramentas permitem que analistas construam pipelines de dados visuais, sem precisar programar manualmente cada etapa.



# O que é OLAP (Online Analytical Processing)

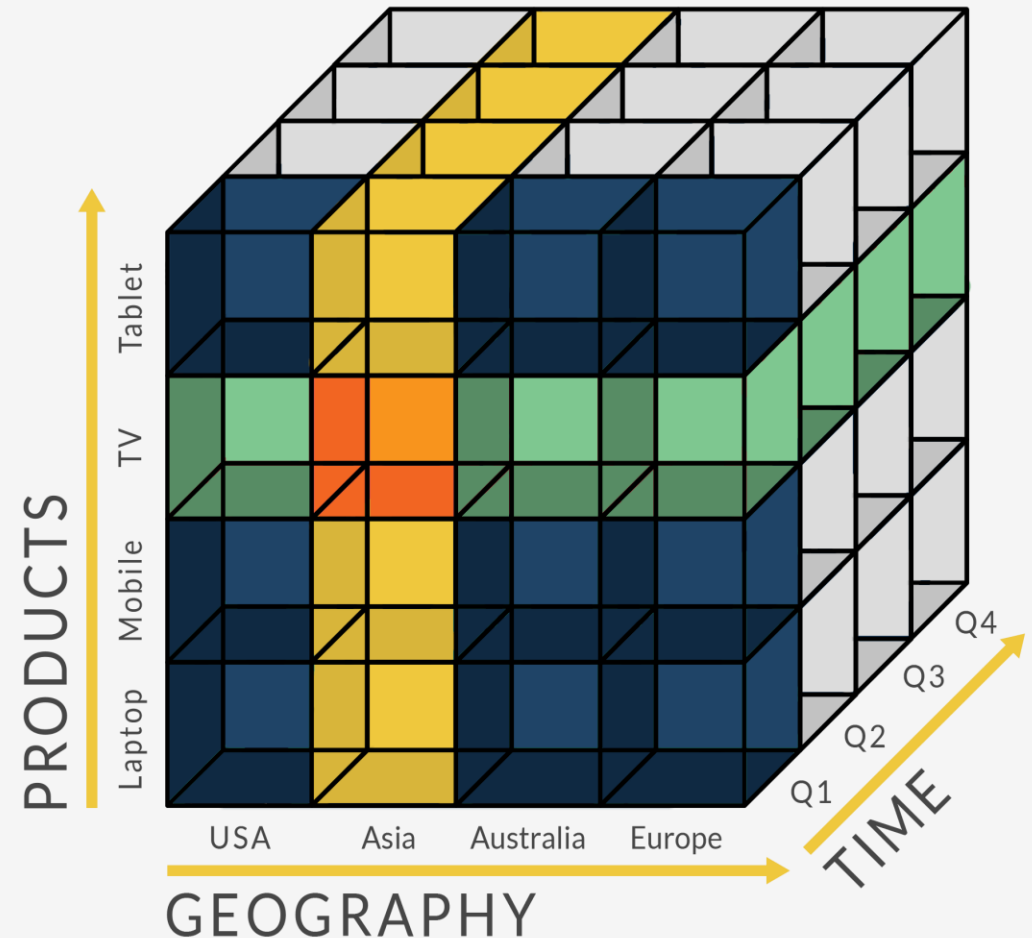
O OLAP é a tecnologia responsável por permitir **análises complexas e rápidas** sobre grandes volumes de dados.

Enquanto bancos de dados tradicionais (OLTP) são otimizados para **transações**, o OLAP é otimizado para **consultas e análises**.

Ele organiza os dados em **estruturas multidimensionais chamadas cubos**, que permitem examinar informações sob diferentes perspectivas como tempo, produto, região ou cliente.

Exemplo:

Um gerente pode visualizar o total de vendas por ano, depois "abrir" o dado por trimestre, depois por região e produto tudo dinamicamente, sem precisar criar novas consultas SQL.



# OLTP x OLAP

O OLAP complementa o OLTP, oferecendo um ambiente voltado à **tomada de decisão**, não à operação.

Característica	OLTP (Transacional)	OLAP (Analítico)
Finalidade	Registrar transações diárias	Analisar e resumir dados
Tipo de Dado	Atual, detalhado	Histórico, agregado
Exemplo de Uso	Sistema de vendas, ERP	Dashboards de BI
Estrutura	Normalizada (muitas tabelas)	Desnormalizada (fato e dimensão)
Consultas	Curtas e frequentes	Longas e analíticas
Performance	Otimizada para escrita	Otimizada para leitura

# Cubos OLAP

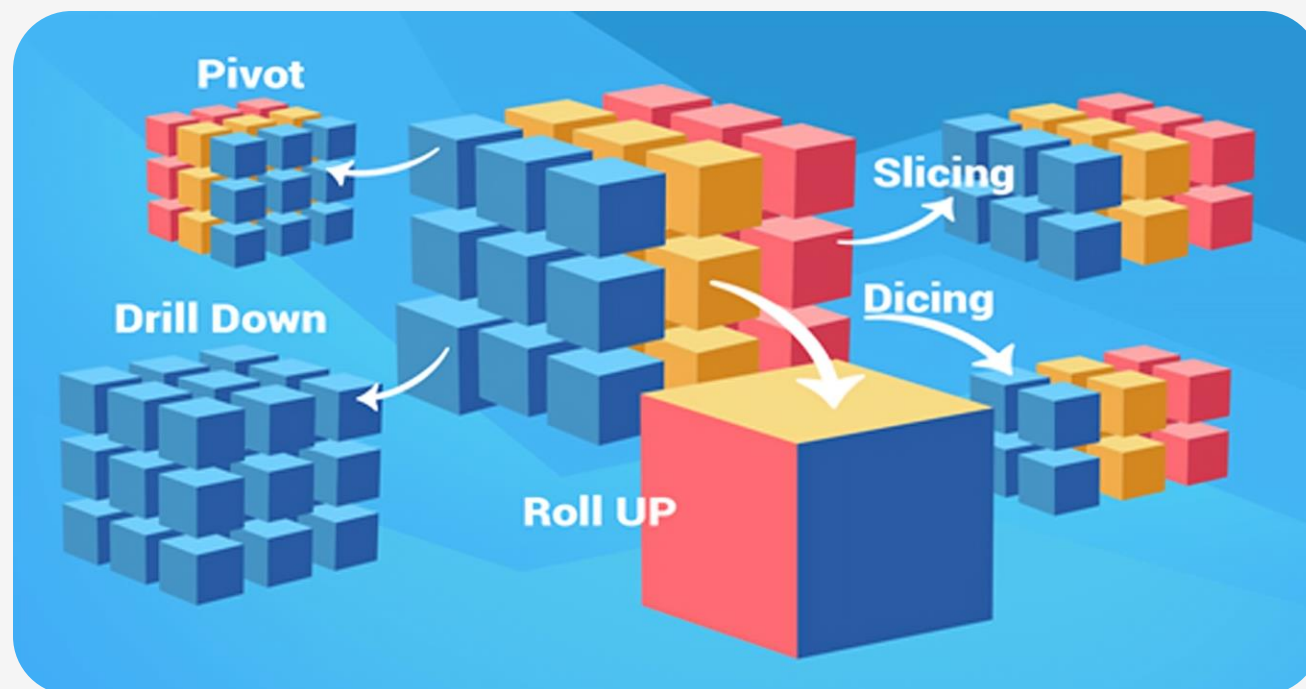
Os **cubos OLAP** são estruturas que organizam os dados em múltiplas dimensões, permitindo **navegação hierárquica** entre níveis de detalhe.

## Principais conceitos:

- **Dimensão:** perspectiva de análise (ex.: tempo, produto, região).
- **Medida:** valor numérico a ser analisado (ex.: vendas, lucro, custo).
- **Hierarquia:** níveis dentro de uma dimensão (ex.: Ano → Trimestre → Mês → Dia).

Essa estrutura facilita a execução de operações como:

- **Drill-down:** aprofundar nos detalhes (de ano para mês).
- **Roll-up:** resumir informações (de mês para ano).
- **Slice:** filtrar por uma dimensão específica (ex.: apenas Nordeste).
- **Dice:** combinar múltiplos filtros.



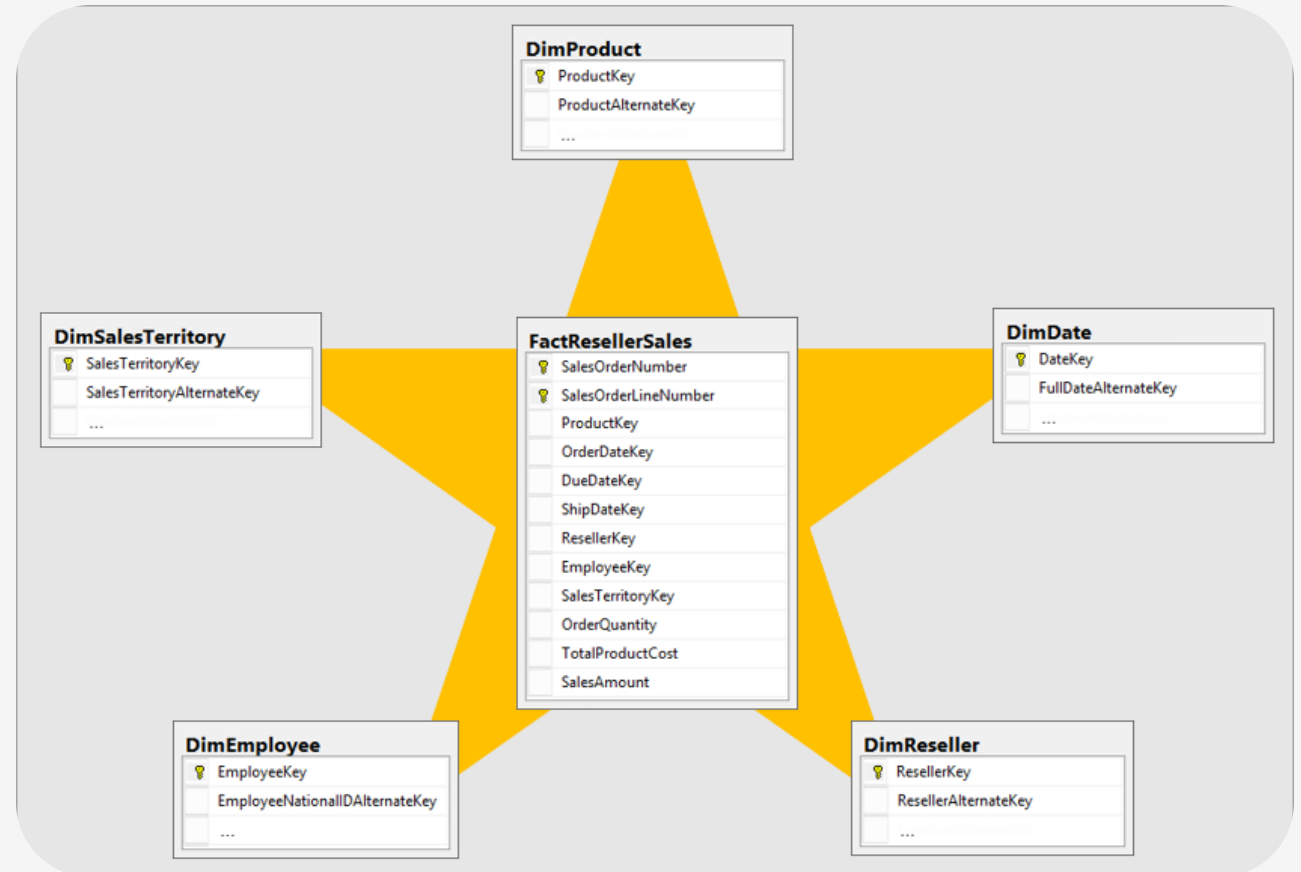
# Modelagem Dimensional

O OLAP depende da **modelagem dimensional**, que organiza os dados em **tabelas fato e dimensão**.

- **Tabela Fato:** contém os valores numéricos de análise (ex.: total de vendas, quantidade).
- **Tabelas Dimensão:** armazenam informações descritivas (ex.: produto, cliente, tempo, região).

Um modelo típico segue o formato **Star Schema (Esquema em Estrela)**, em que uma tabela fato central se conecta a várias dimensões.

Esse modelo facilita consultas analíticas e otimiza a performance das análises no Power BI.



# Tipos de OLAP

## 1. MOLAP (Multidimensional OLAP):

Dados armazenados em estruturas multidimensionais otimizadas para análise.

Alta performance, mas requer pré-processamento.

## 2. ROLAP (Relational OLAP):

Usa bancos de dados relacionais (SQL) para gerar as análises dinamicamente.

Mais flexível, mas pode ser mais lento.

## 3. HOLAP (Hybrid OLAP):

Combina as duas abordagens, armazenando dados resumidos no cubo e detalhes no banco relacional.

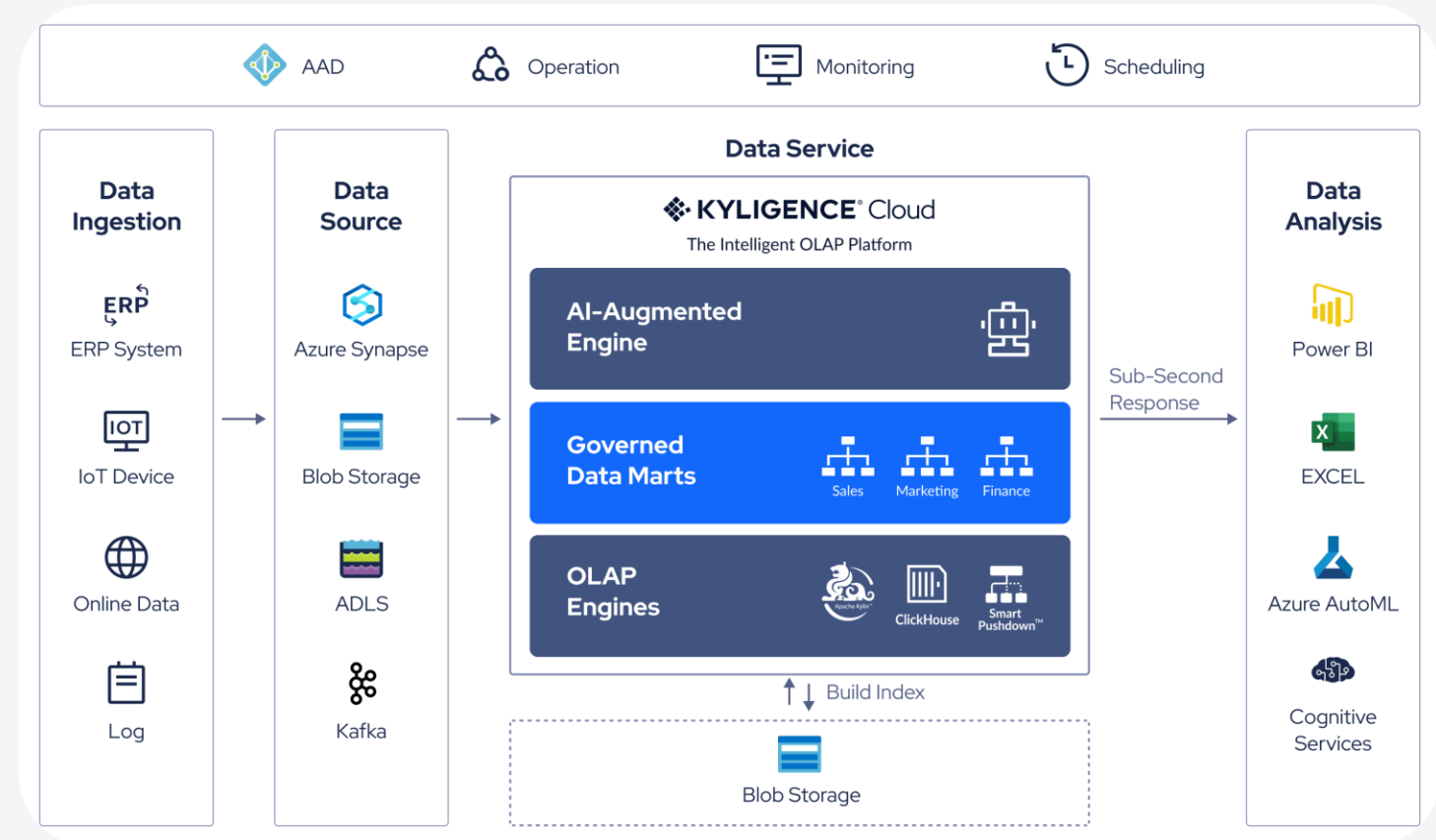


# OLAP e Power BI

O Power BI utiliza conceitos de OLAP internamente para estruturar dados em **modelos relacionais otimizados**.

Quando o aluno cria relacionamentos entre tabelas (ex.: FatoVendas e DimProduto), o Power BI monta automaticamente **estruturas dimensionais**, permitindo operações como drill-down e filtros hierárquicos.

Por isso, entender OLAP e ETL ajuda o estudante a construir **dashboards mais consistentes e rápidos**.



# Encerramento e Reflexão

O **ETL** e o **OLAP** são as engrenagens invisíveis do **BI** eles garantem que os dados cheguem organizados e analisáveis às ferramentas de visualização.

Sem o ETL, o BI ficaria sobrecarregado de dados despadronizados.

Sem o OLAP, as análises seriam lentas e limitadas.

Compreender essas tecnologias prepara o aluno para pensar **de forma sistêmica**, entendendo o que acontece “por trás” do Power BI.

