



PUC Minas

Conceitos Inicias

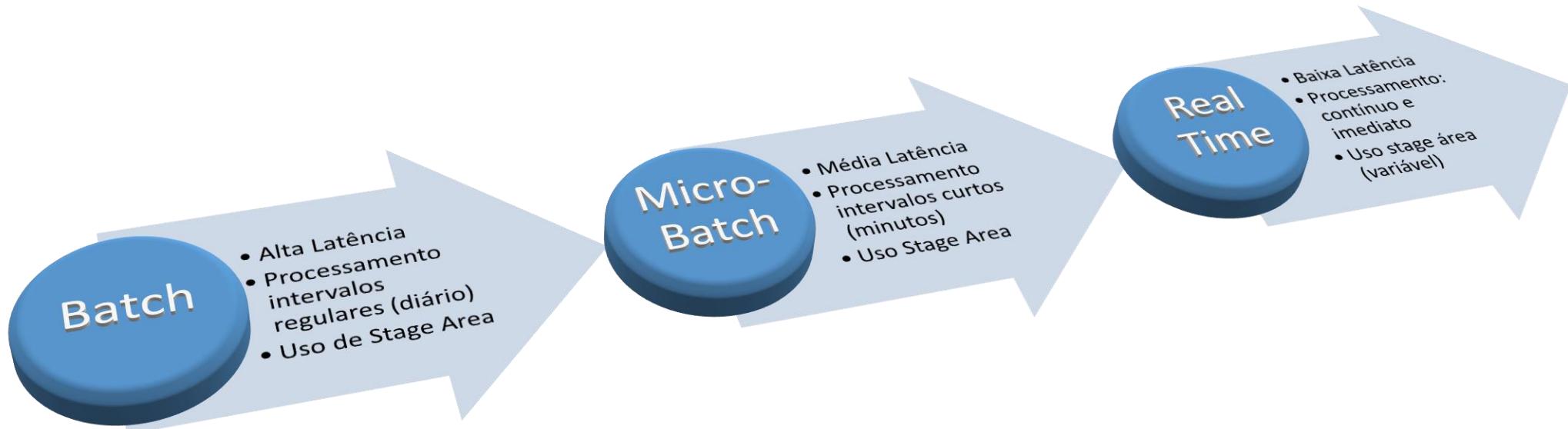
Sheila Dias

Ingestão de Dados

INGESTÃO DE DADOS

- A ingestão de dados é o processo inicial de **COLETAR** e **LEVAR** dados de uma fonte qualquer para um novo destino.
- Trata-se da movimentação de massas dados com rapidez e eficiência

FREQUÊNCIA INGESTÃO



Elaborado pelo autor

ARQUITETURA INGESTÃO DE DADOS



Elaborado pelo autor

TIPOS DE INGESTÃO

- **Empurrada (*push*)**- A aplicação geradora do dado “empurra” o dado para o fluxo de dados.
- **Puxada (*pull*)**- O dado é buscado pelo fluxo de dados e conduzido para o repositório/análises



<https://www.shutterstock.com/g/Vitaliy belozerov>



PUC Minas



PUC Minas

ETL x ELT

Sheila Dias



PUC Minas

DATA LAKE

VISÃO

Um lago de dados é como um reservatório.

Primeiro você cria a estrutura (um cluster) e depois enche de água (dados).

Depois que o lago estiver pronto, você começa a usar a água (dados) para várias finalidades, como geração de energia e consumo (análises eventuais, análises históricas, análises preditivas, etc.).



<https://www.shutterstock.com>

CONCEITO

- É um termo atribuído ao CTO (*Chief Technical Officer*) do *Pentaho*, James Dixon
- A ideia é ter um único repositório dentro da empresa, para que todos os dados brutos estejam disponíveis a qualquer pessoa que precise fazer análise sobre eles.



<https://www.shutterstock.com>

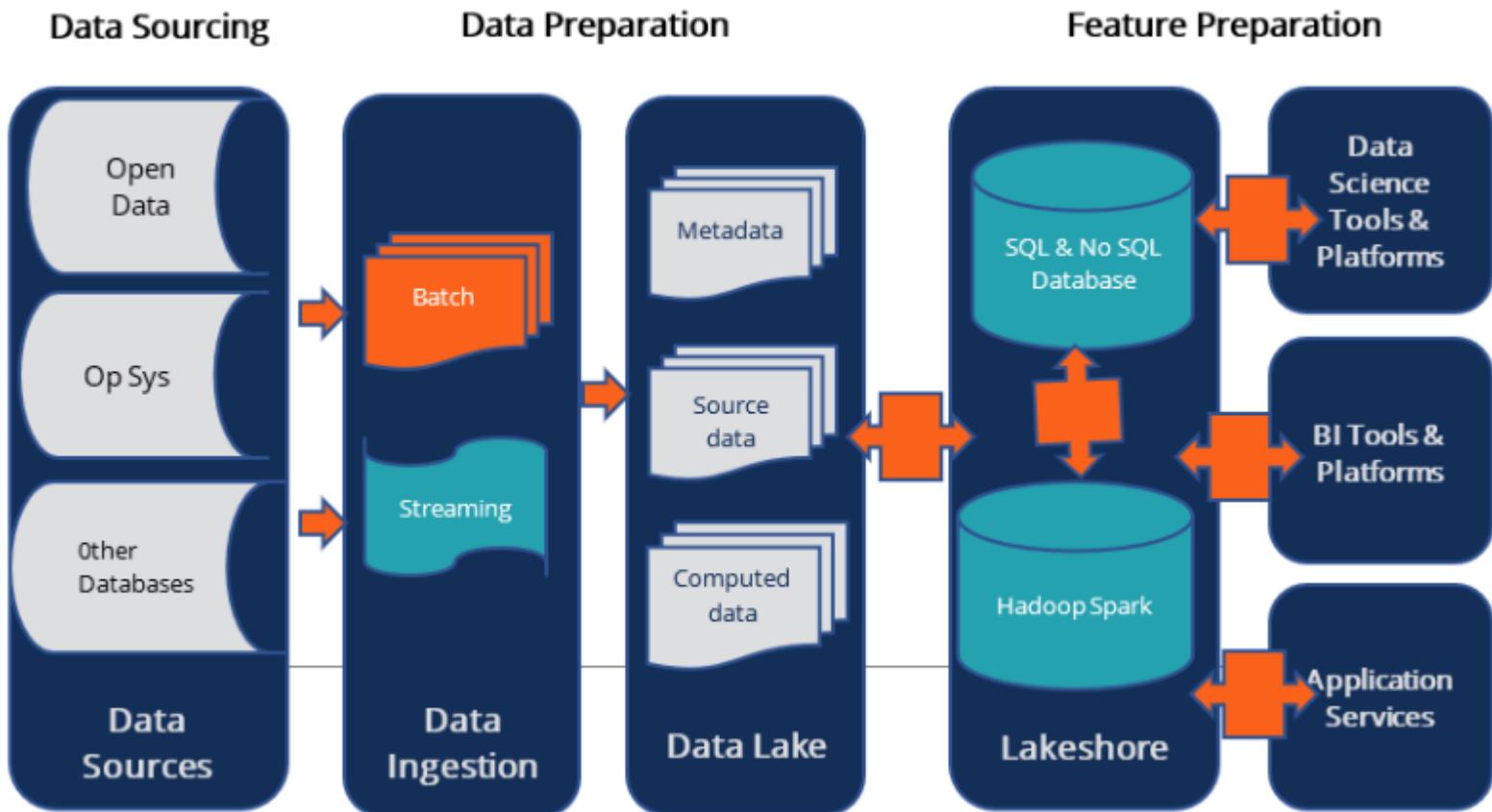
ETL x ELT

BIG DATA



<https://www.linkedin.com/pulse/big-data-vs-lake-mustafa-qizilbash>
Acessado em agosto/2024

ARQUITETURA SIMPLIFICADA DATA LAKE



<https://corporatefinanceinstitute.com/resources/data-science/data-lake/>
Acessado em agosto/2024



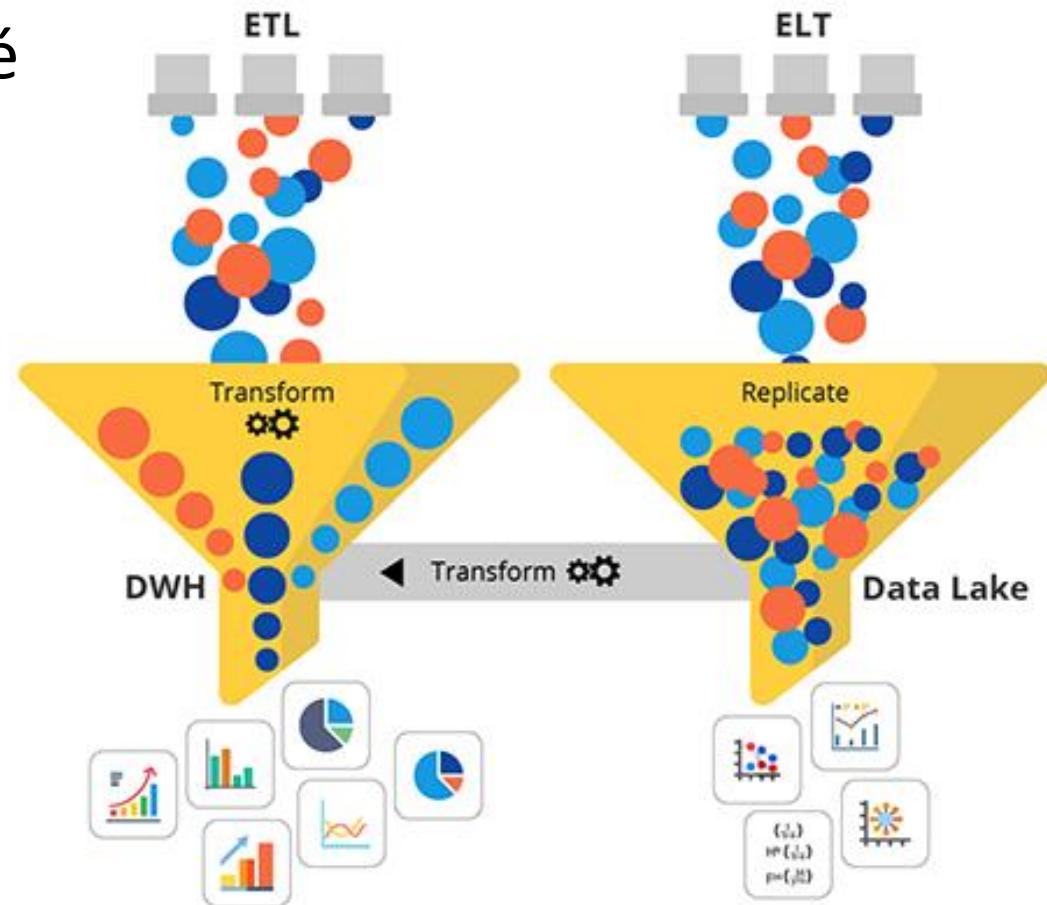
PUC Minas

ETL x ELT

ETL x ELT

- **ETL->** O dado é armazenado após um tratamento de REGRA.

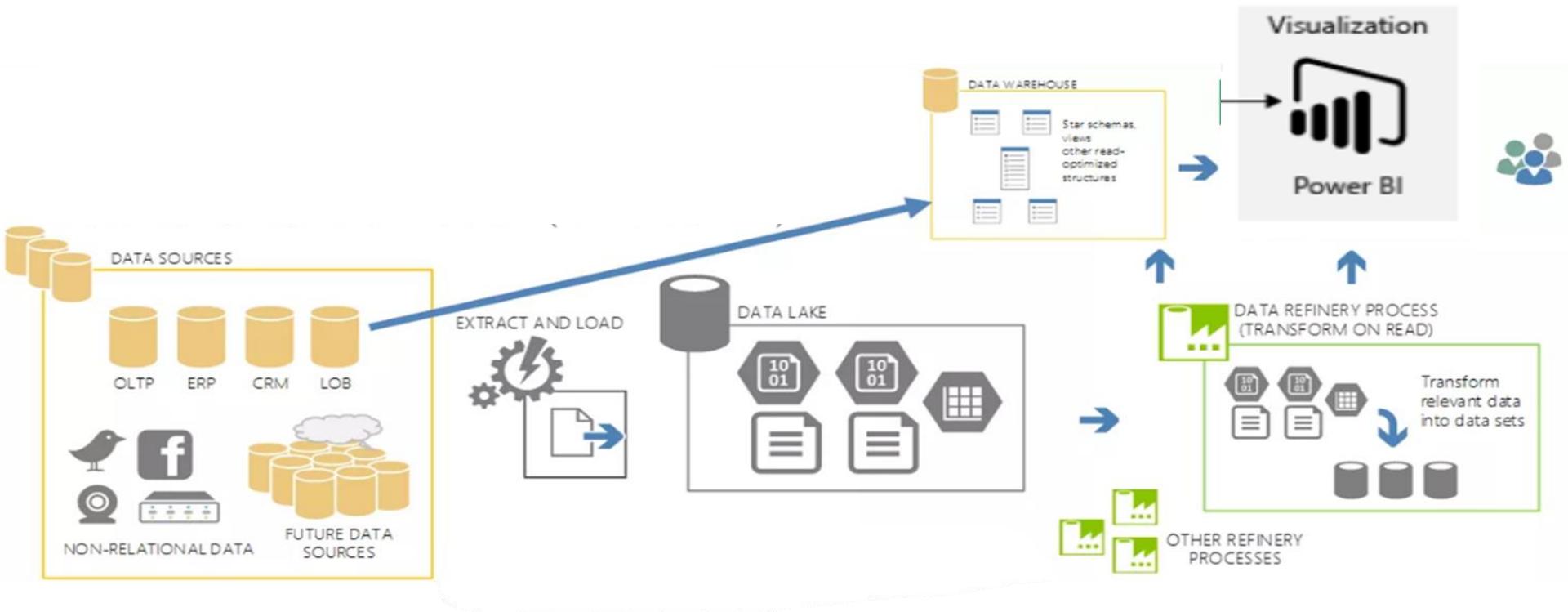
PRONTO para uso.



- **ELT ->** A transformação (limpeza, análise, tratamentos em geral) vêm depois do dado armazenado.

Uso APÓS Tratamento.

ARQUITETURA ETL x ELT





PUC Minas

Ferramentas/Serviços Integração Dados

Sheila Dias

FERRAMENTAS/SERVIÇOS

São soluções que permitem combinar dados provenientes de diferentes fontes e formatos em um sistema unificado e coerente.

BENEFÍCIOS

- Extração Rápida e entrega dos dados
- Acurácia
- Capacidade de Escalar
- Colaboração
- Soluções Econômicas
- Interface Amigável



<https://www.shutterstock.com/g/Bogoiavlenskaia>

EXEMPLOS

- *Informática - PowerCenter*
- *Oracle – Oracle Data Integration*
- *IBM – Data Stage*
- *Talend*
- *SAP- Data Services*
- *Microsoft- SQL Server Integration Services (SSIS)*
- *Hitachi - Pentaho Data Integration (PDI)*

Figure 1: Magic Quadrant for Data Integration Tools



<https://www.informatica.com/data-integration-magic-quadrant.html> Acessado Agosto/2024

MAIS FERRAMENTAS / SERVIÇOS

■ Transformação de dados

- ✓ Apache Nifi
- ✓ SQL
- ✓ Python (Pandas, NumPy)
- ✓ R
- ✓ **AWS Glue : <https://aws.amazon.com/pt/glue/>**

■ Ingestão de dados

- ✓ AWS Glue
- ✓ AWS Kinesis
- ✓ Apache Nifi,
- ✓ Apache Kafka
- ✓ Google Cloud Dataflow
- ✓ Airbyte



PUC Minas