



**RECIFE DATA
SATURDAY**

APOIO

- **LS Treinamentos**
- **Consulta BD**
- **Big Data Navigators**
- **Faculdade Senac Pernambuco**



QUEM SOU EU?



- + 12 anos na área de TI/Dados.
- Azure Data Engineer (Pleno) na Dataside
- 4 Certificações Microsoft.
- Admin e escritora do blog Big Data Navigators



Rayza Wilma





Transformando dados em Insights com Synapse Analytics Serveless e Power BI

Big Data



- O termo começou a ser apresentado quando os métodos tradicionais para armazenamento e processamento começaram a não ser tão eficientes.
- Inicialmente definido pelos **3 V's**
 - Volume
 - Variedade
 - Velocidade

Dica: Lá no blog **Big Data Navigators** já tem um post falando bastante sobre o termo. 😊



Você conhece o Azure?



É uma plataforma destinada à execução de aplicativos e serviços baseada nos conceitos da **computação em nuvem** da Microsoft.

- Máquinas virtuais
- Bancos de dados
- Serviços de aplicativos
- Serviços para automações
- Data Lakes
- IoT



Azure Data Lake Storage



- De forma simplificada, podemos dizer que é um repositório de dados centralizado que pode armazenar grandes volumes de dados, de diferentes tipos de dados e arquivos.
- Tipos de Dados
 - Estruturados – Bancos de dados Relacionais
 - Semi Estruturados – Logs, XML e JSON
 - Não Estruturados – E-mails, Áudio e Vídeo
- Tipos de Arquivos
 - Arquivos em formato PARQUET, efetuando excelente compressão nos dados
 - Arquivos em formato DELTA (Atomicidade, Consistência, Isolamento e Durabilidade)
- Controle de ciclo de vida do arquivo (HOT, COOL, ARCHIVE)

Engenharia de Dados



A engenharia de dados é uma área que se concentra na **coleta, organização, processamento e preparação de dados** para **análise**.

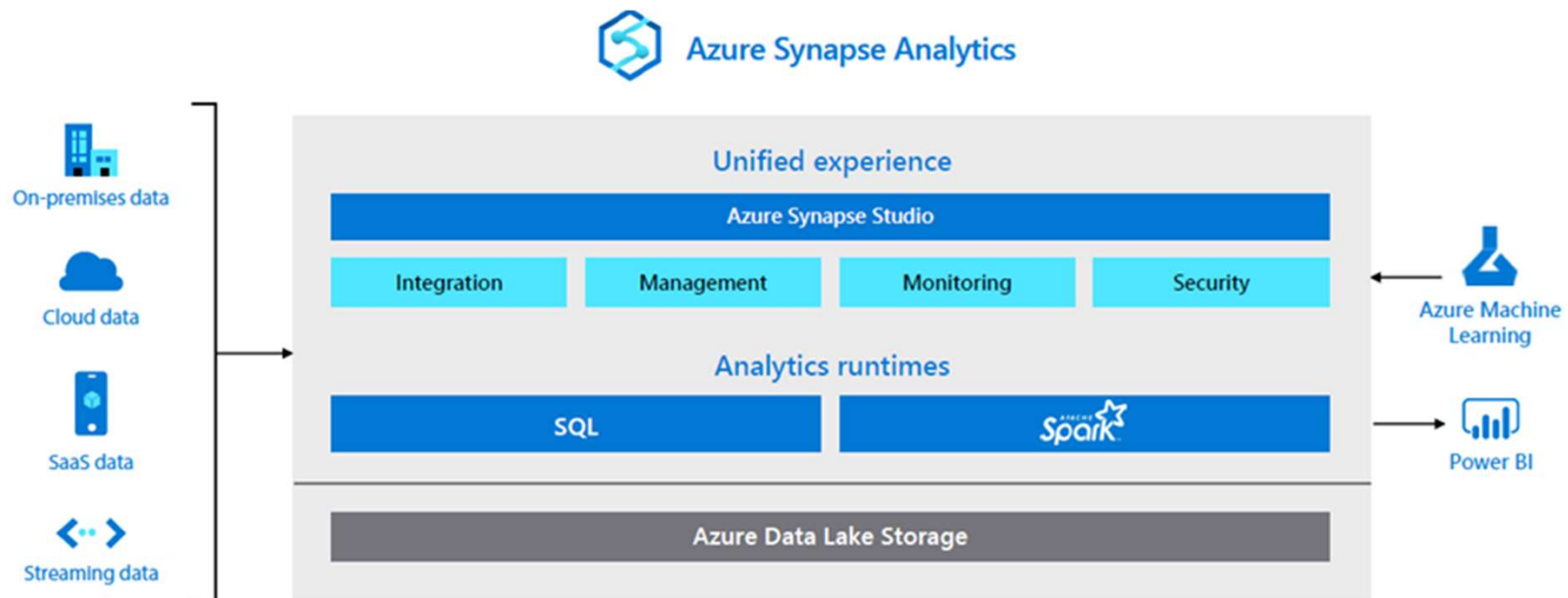
Os engenheiros de dados projetam e mantêm sistemas que permitem que as empresas obtenham informações significativas a partir de seus dados brutos, sendo fundamental para a ciência de dados e da análise de dados.



Azure Synapse Analytics



O Azure Synapse Analytics é um ambiente voltado para análise de Big Data, que oferece recursos em escala sem servidor ou dedicado.

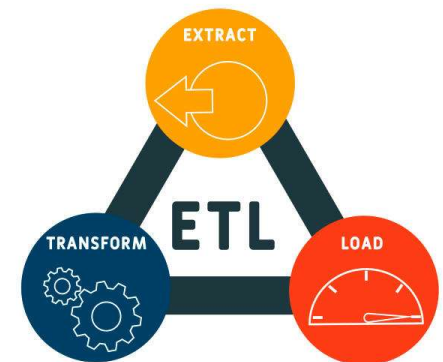


O que é um pipeline ETL?



É um conjunto de processos que inclui a extração de dados de uma variedade de fontes, onde eles passam por determinadas transformações e subsequentemente são carregados em um destino, que pode ser por exemplo um DW.

- **ETL** - Extrair, Transformar e Carregar.
- **ELT** - Extrair, Carregar, Transformar



Componentes necessários para um pipeline



- **IR** – O Integration Runtime é uma infraestrutura computacional usada para conectar o ADF ou Synapse com os serviços do Azure, serviços externos ou executar pacotes do Integration Services. (Azure, Self-Hosted, Azure SSIS)
- **Linked Service** – São conectores que permitem estabelecer conexão com os armazenamentos de dados.

Componentes necessários para um pipeline



- **Dataset** – Representam estruturas de dados dentro das fontes de dados. Ex: tabela em um banco de dados, arquivo no data lake
- **Activity** – Representa a ação que pode ser realizada sobre os dados, como copiar os dados, executar um notebook, obter metadados de um arquivo

Power BI



De maneira simplificada, é uma ferramenta utilizada para a construção de relatórios e painéis (dashboards).

É muito utilizada por analistas de dados para aplicação de **Business Intelligence (BI)**.



Power BI <> BI



Power BI == Ferramenta

BI == Conjunto de técnicas, conceitos e ferramentas que permite a organização e análise das informações para o suporte a tomada de decisão.



Hands-On



**O futuro depende
do que você
escolher fazer hoje.**

Mahatma Gandhi

Mídias



[rayzawilma](#)



[bigdatanavigators](#)



[bigdatanavigators](#)



<https://bigdatanavigators.com/>



OBRIGADA PELA PRESENÇA!