“Cada um escolhe sua guerra”

“O muro das lamentações fica lá em Israel”

Sumário

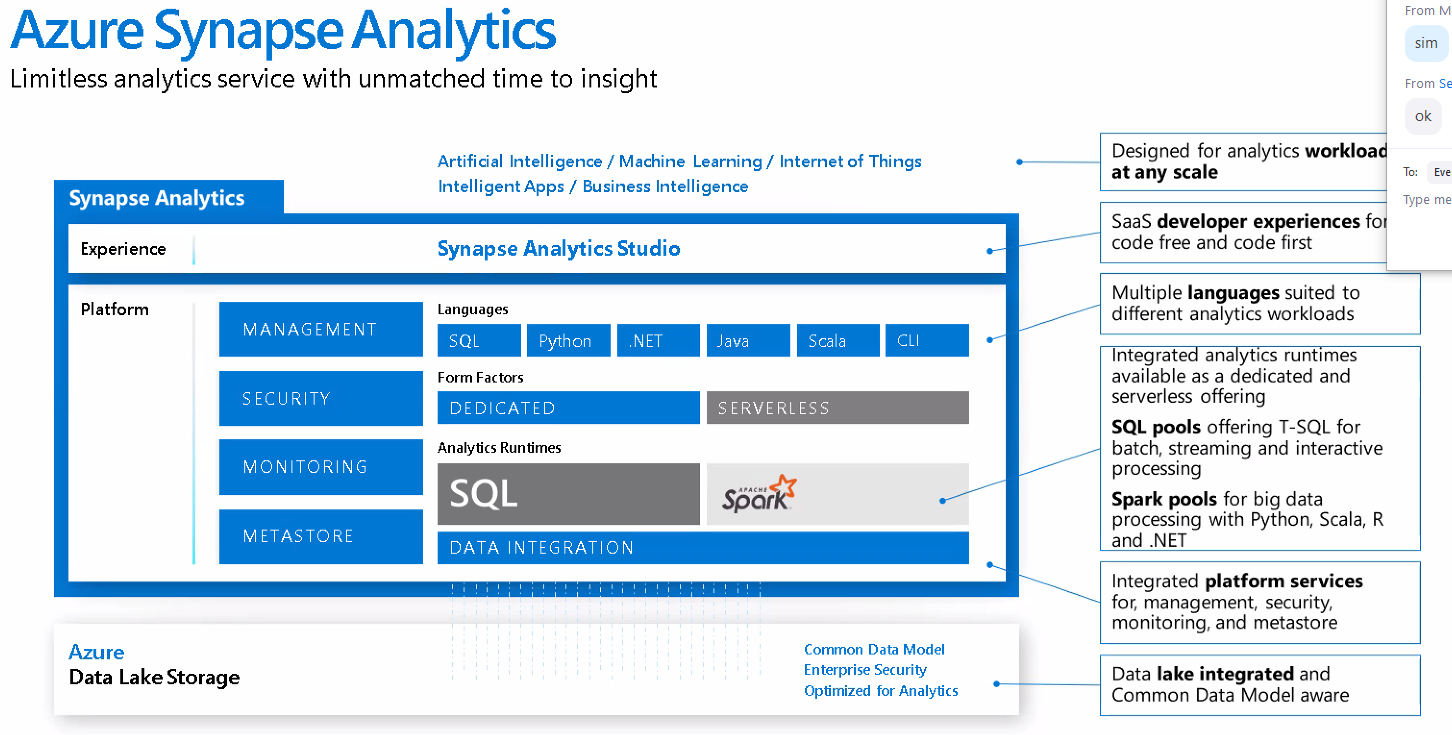
[Aula 1 1](#_Toc79516507)

[Aula 2 9](#_Toc79516508)

# Aula 1

Trabalho de Engenharia de Dados -> Trabalhar com os dados para os demais cargos.

Azure Synapse Analytics -> Tem de dentro dele. Processamento pode executar em SQL ou SPARK.



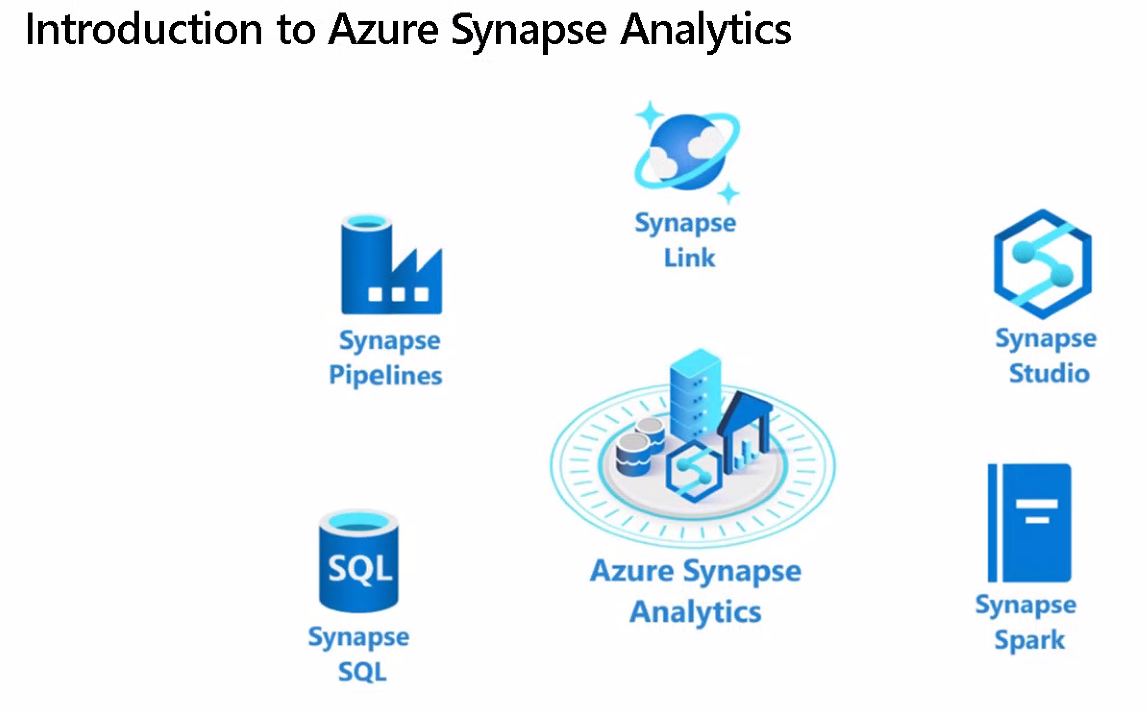
Synapse = DW + BI. O foco não é Ciência de Dados.

DataBricks = Big Data + Ciência de Dados. O foco do DataBrincks não é o DW.

Os dois podem fazer a mesma coisa, porém cada um tem focos diferentes.

Big Data = Pelo menos ter os três Velocidade, Volume e Variedade .

Azure Synapse Analytics Studio



Synapse é um conjunto de:

Synapse Pipelines (é a mesma coisa do Data Factory, ETL),

Synapse Studio (Gerenciamento).

Synapse SQL -> É uma solução que não é barata. Tem um Pool de 60 servidores. Quando clica para executar, ele liga 60 máquinas. Aqui Giga o pouco. É de Tera para cima.

Limites -> Dados não estruturados

Synapse Spark -> Criar um DataFrame, com vários tipos de dados.

Synapse Link -> Quando o modelo estiver pronto, liga um cabo para conectar as partes, sem precisar do ETL. Os dados já vem prontos.

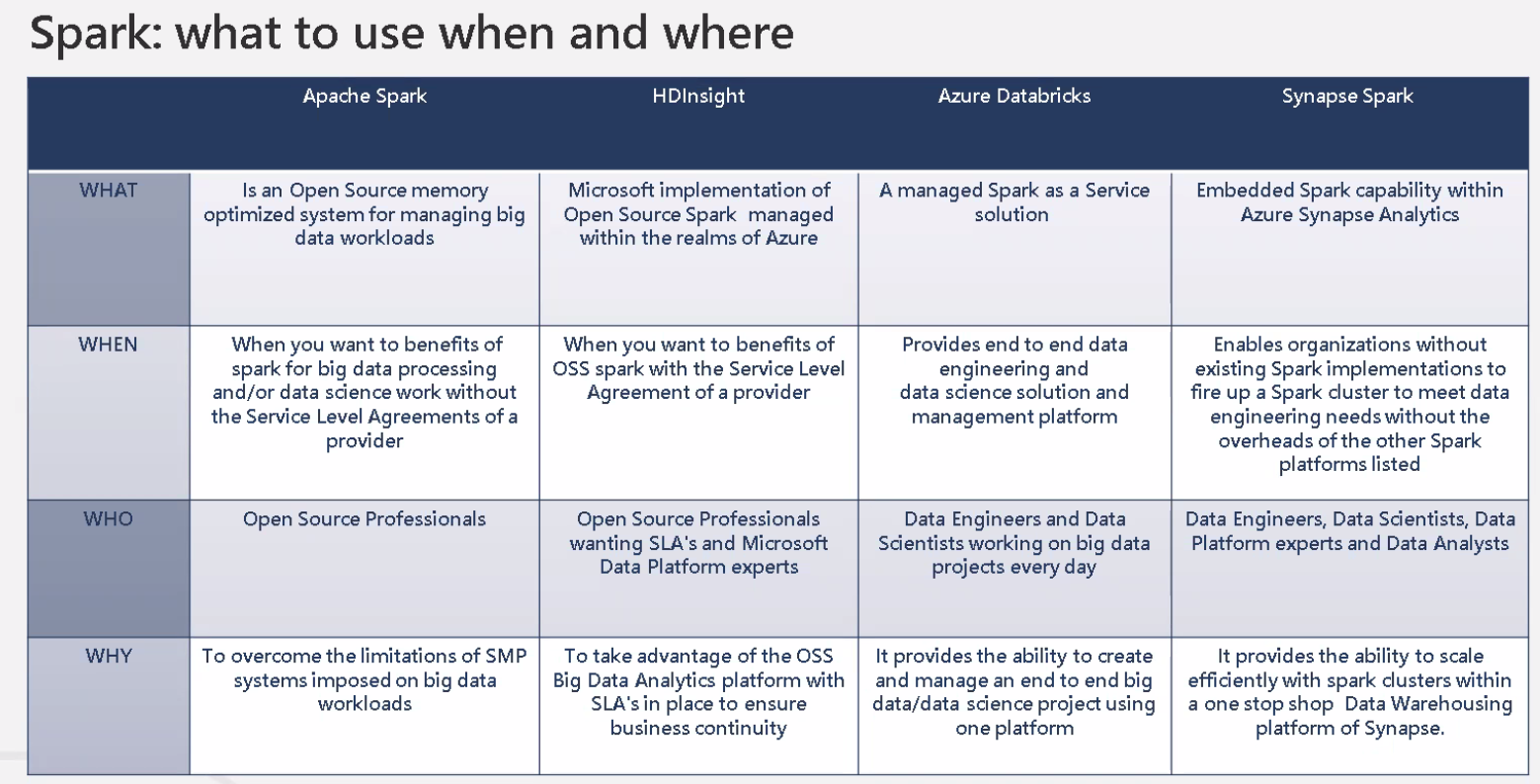
DataWareHouse Moderno -> Trabalhar com dados sem ser estruturados.

Apache Spark -> Serviço em memória. Open Source. Usar em SLA, não tem poder de escabilidade.

HDInsight ->

Azure DataBrincks ->

Synapse Spark ->



Lesson 1 -> Azure DataBrincks

É baseado no Spark, processamento feio em memória.

Possui bibliotecas de SQL, DataFrames e etc...

Apache Spark

Processamento paralelo, em memória, em Big Data em ambiente local.

Componentes:

Spark, DataFrames, Streaming, Mlib, GraphX (Processamento de grafos), SpartCore API.

Segurança de Dados (Enterprise Security)

Integração da Segurança dos Dados.

Usa o Azure Active Directory. Acesso por meio do papel do cliente na organização.

Se integra com Data Lake Store, CosmosDB, Data Factory, Azure Synapse Analytics, Blob Store, Cognitive Services.

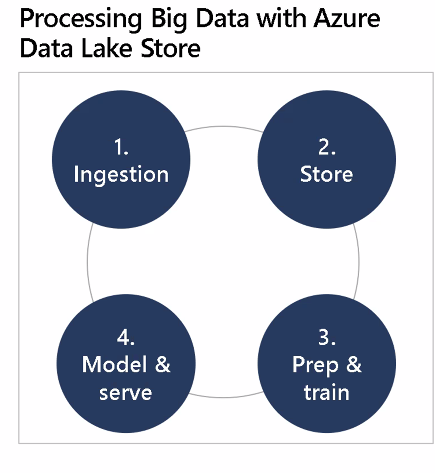
Lesson 1 -> Introduction Azure Data Lake Storage

Local onde grava o dado bruto da origem.

Serve para tirar os dados do transacional, para poder ter segurança.

Aqui não se processa, apenas armazena.

O persistir escolhe se é no Cosmo ou outro lugar.



Para o Big Data

Compatível com o Hadoop.

Blob vs Data Lake

Blob é mais lento que o Data Lake.

Blob é mais para armazenar dados que não precisaram ser processadas.

No Blob não tem suporte ao Spark.

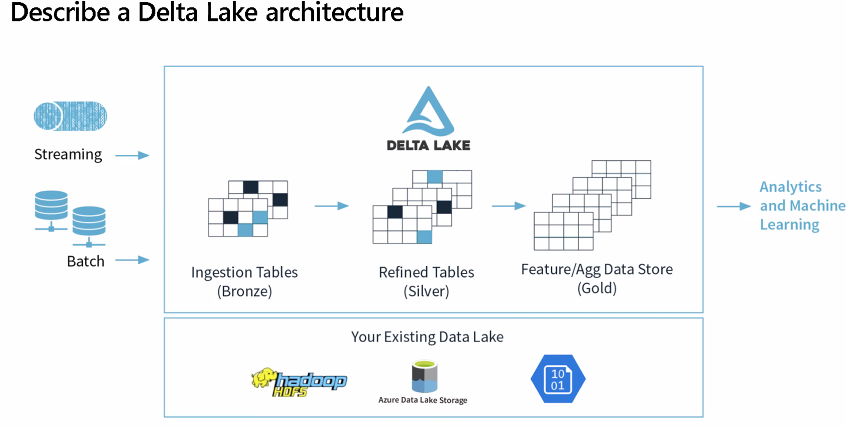
Data Lake se encaixa em um Modern DataWareHouse, Análise de Dados mais Avançados e Processamento de Real Time.

Arquitetura Delta Lake

Implementa uma camada acid.

Não é um serviço, é apenas uma arquitetura.

É uma arquitetura Batch e Streaming processam em conjuntos.



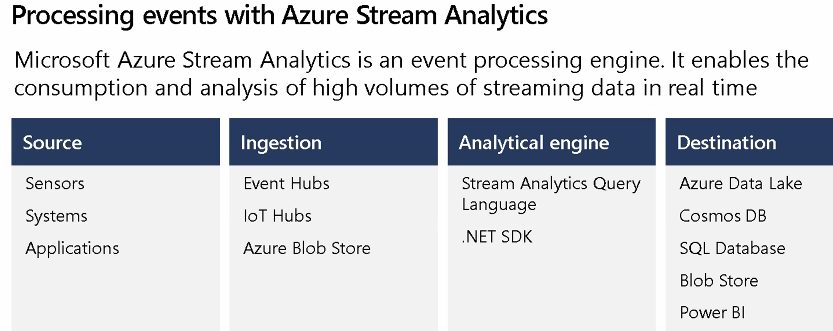
Quando recebe um dado, ele vai para a camada bronze, de forma bruta. Depois para a camada prata, que refinados dados. Após isso, passa a ser aparentado ao Cientista dos Dados, por exemplo, faz o Ouro.

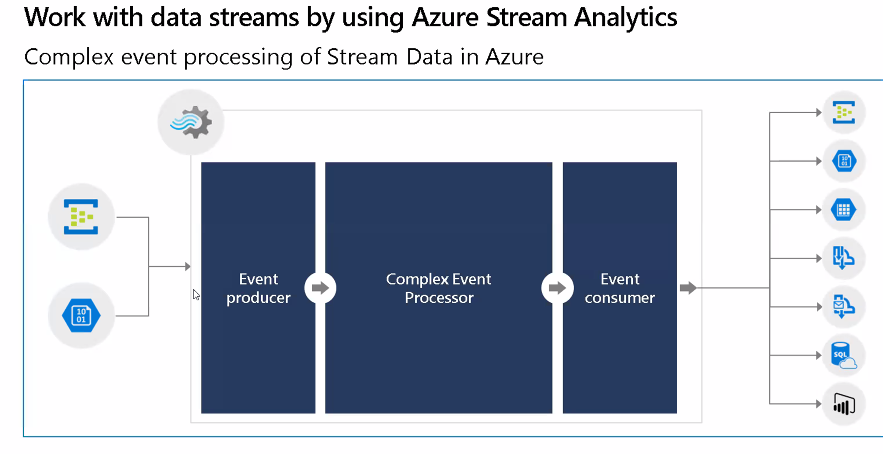
Lesson 1 - > Work with data streams by using Azure Stream Analytics

Data Streams -> Não existe agendamento de carga. Os dados são continuamente. Ex: Eventos de IOT. OS dados ficam circulando continuamente.

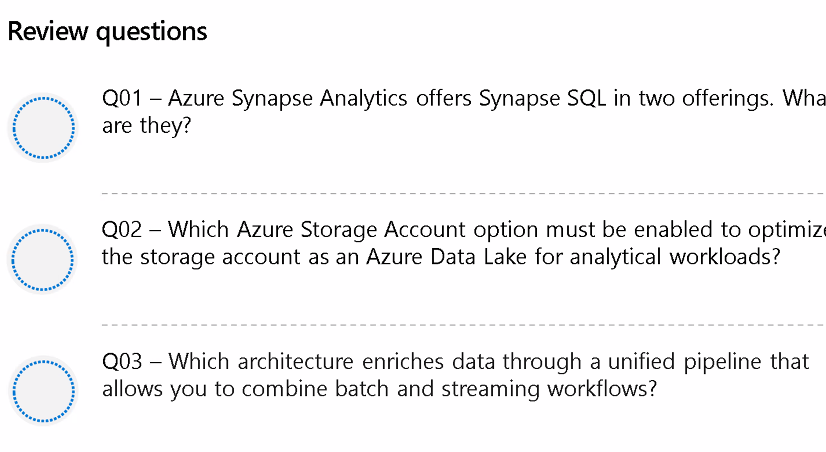
Isso permite analises de dados contínuos.

O que é contínuo é a entrada e o processamento, porque na saída já não é mais tarefa do Azure Stream Analytics.

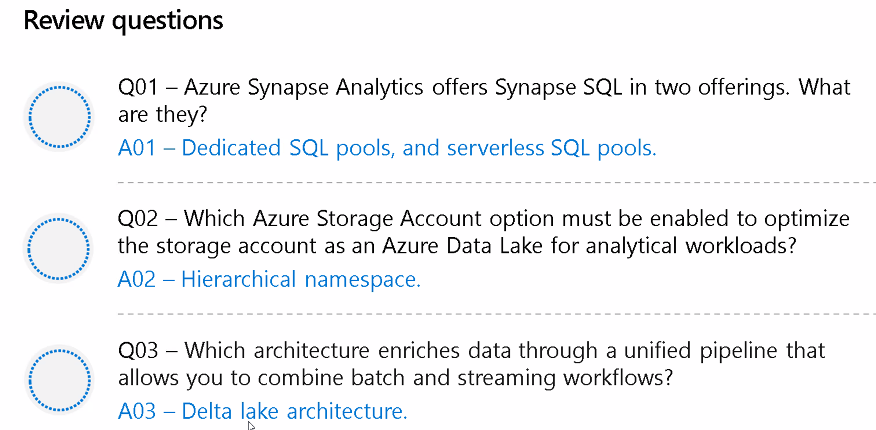




O dado vem de um Hub ou Blob, faz todo o processamento e insere em algum lugar. O dado não pode parar porque senão dá erro.



Q01 -> Uma das ofertas é trabalhar com Servelees (Processar os dados mas sem armazenar. Apenas pega de um lugar e passa para outro). Ou de Provisionamento (Gravar os Dados).



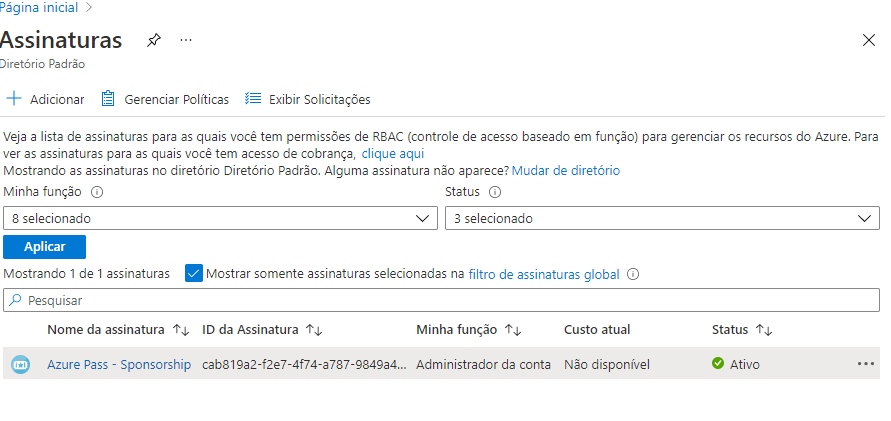
AZURE PASS

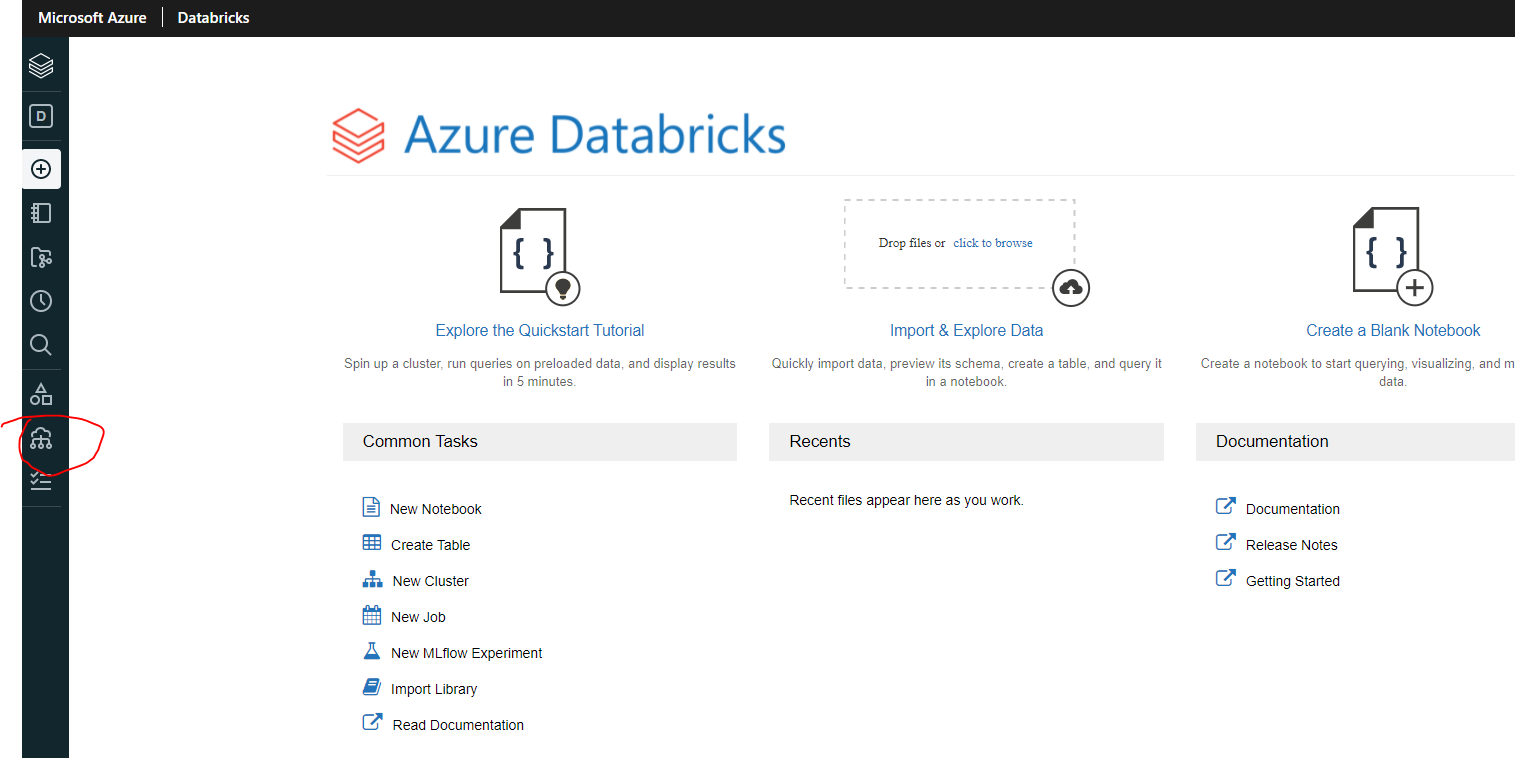
<https://www.microsoftazurepass.com/> -> Adicionar um PASS

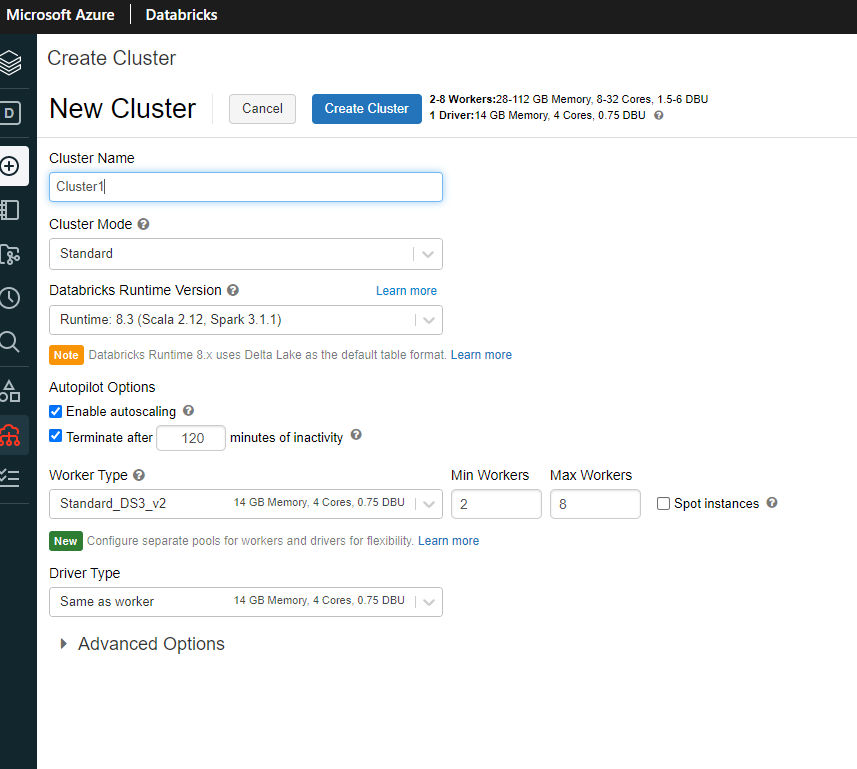
<https://www.microsoftazuresponsorships.com/> -> Verificar o que tem de dinheiro

<https://github.com/solliancenet/microsoft-data-engineering-ilt-deploy/blob/main/setup/01/lab-01-setup.md>

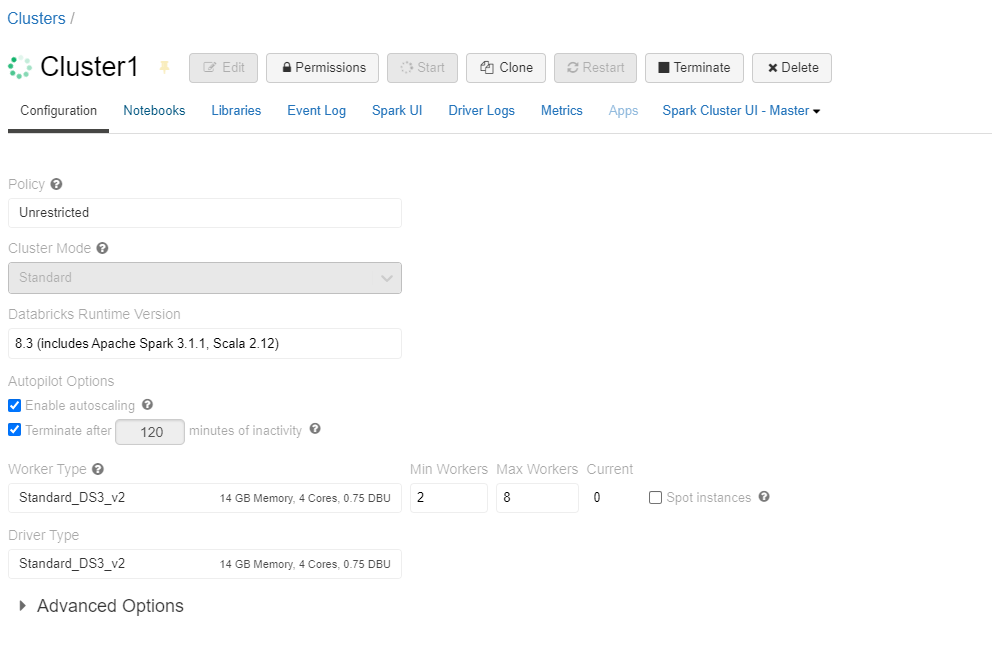
Mudar o nome da Assinatura

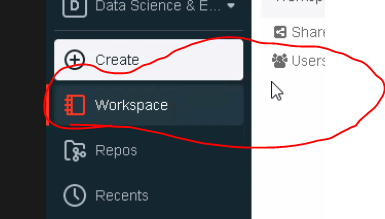






Clicar em Start





Vai aparecer meu e-mail. Clicar em Import.

Clicar em URL

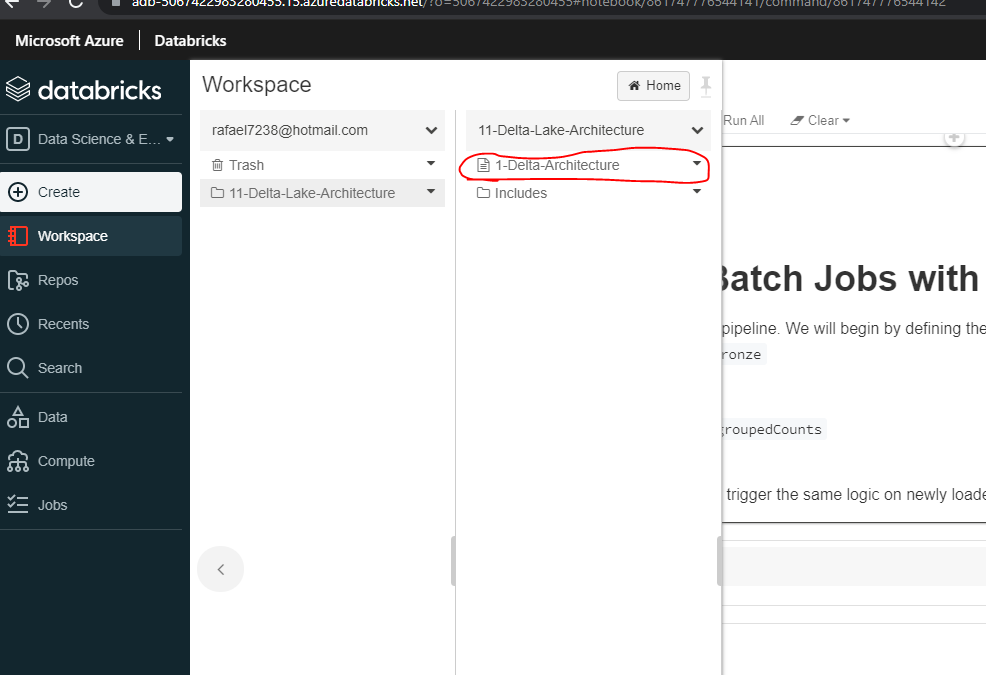
<https://github.com/solliancenet/microsoft-learning-paths-databricks-notebooks/blob/master/data-engineering/DBC/11-Delta-Lake-Architecture.dbc?raw=true>

Depois disso, clicar em TERMINATE

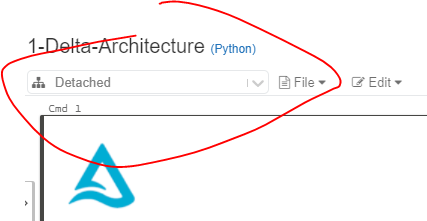
Para toda assinatura do Azure, ele não libera todos os serviços automaticamente. Tem que ir na assinatura e Provedores de Recursos.

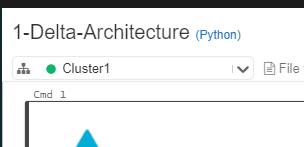
# Aula 2

Abrir um Notebook;



Carregará um Jupyter Notebook;





Executar o Notebook passo a passo.

. Criar uma pasta com alguns dados;

<https://alh.learnondemand.net/User/Login?ReturnUrl=%2fUser%2fCreate%2f>

Máquina Virtual que fica na Internet

MÓDULO 2 – Run Querys using Synaps Analytics serverless Sql Pools.

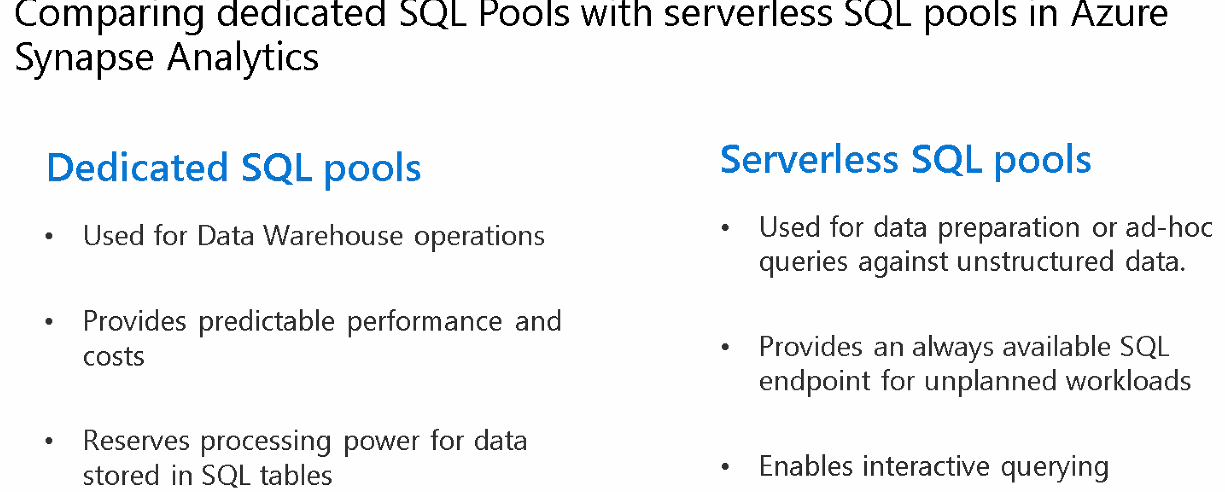
Todo o Azure Synapse vem com um Workspace.

Diferença entre Dedicated Sql Pools e Serverless Sql Pools.

Nunca usar o Synapse Analytics para dados pequenos.

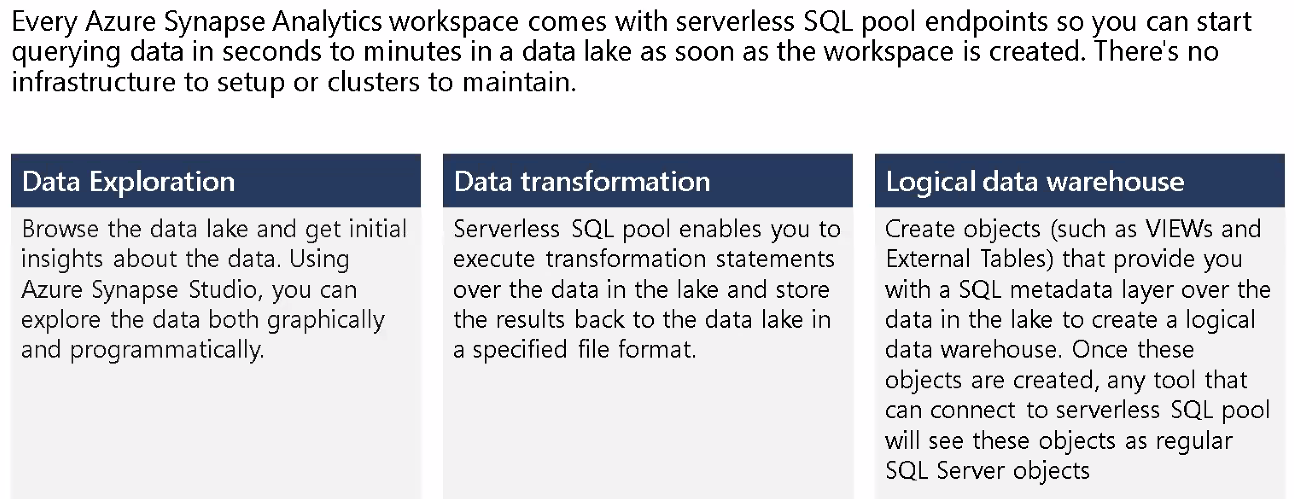
Por exemplo: E-commerce, transações de banco.

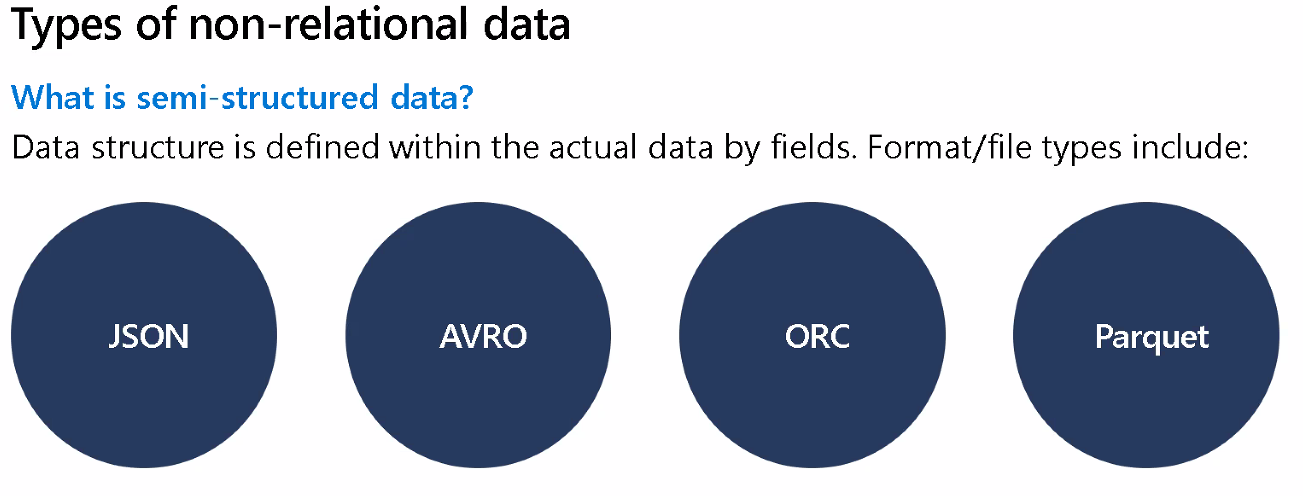
Serverless só executa e vai embora. Dedicated armazena os dados.



Quando usar o serveless:

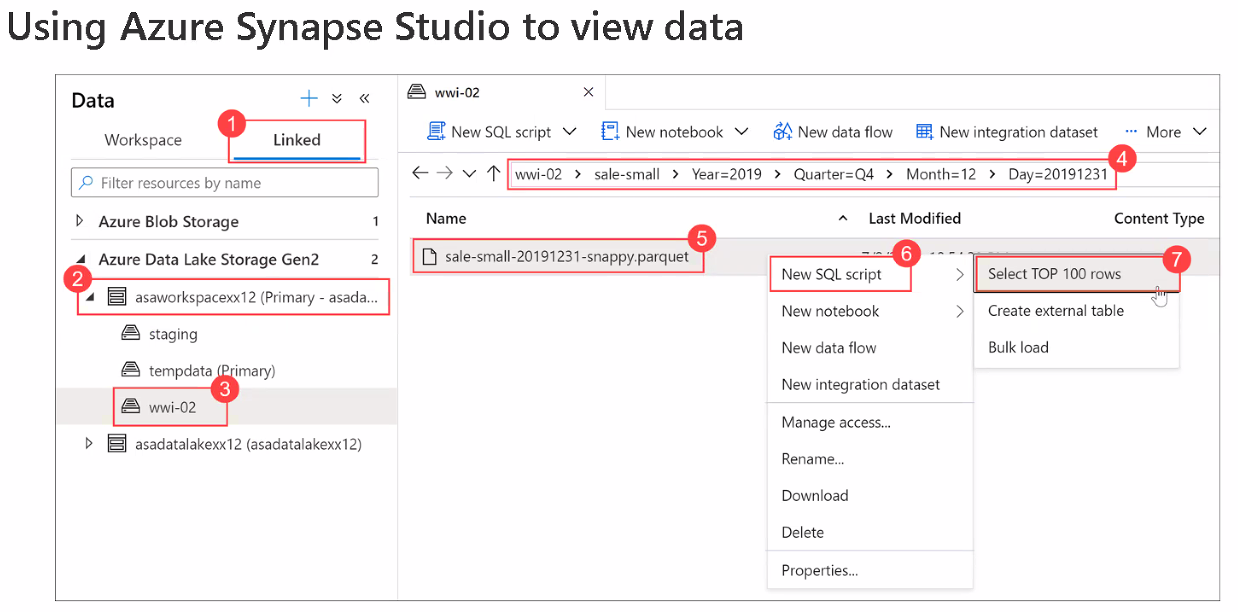
Não precisa criar a infraestrutura.



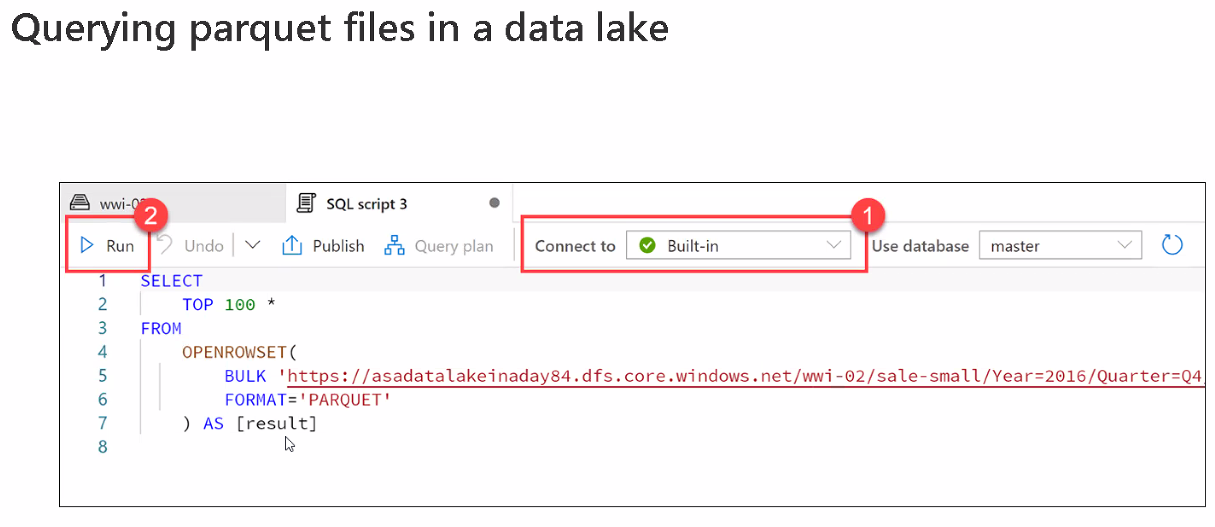


Lesson 1 – Utilizar Querys no DataLake





Built-in é o Serveless



Lesson 1- > Create Objetoss no Azure Serverless SQL Pools

Os objetos são virtuais.

Vamos Criar um banco Lógico no Serverless

Criar uma Credencial

Criar um External Data Source

Criar um External File Format

