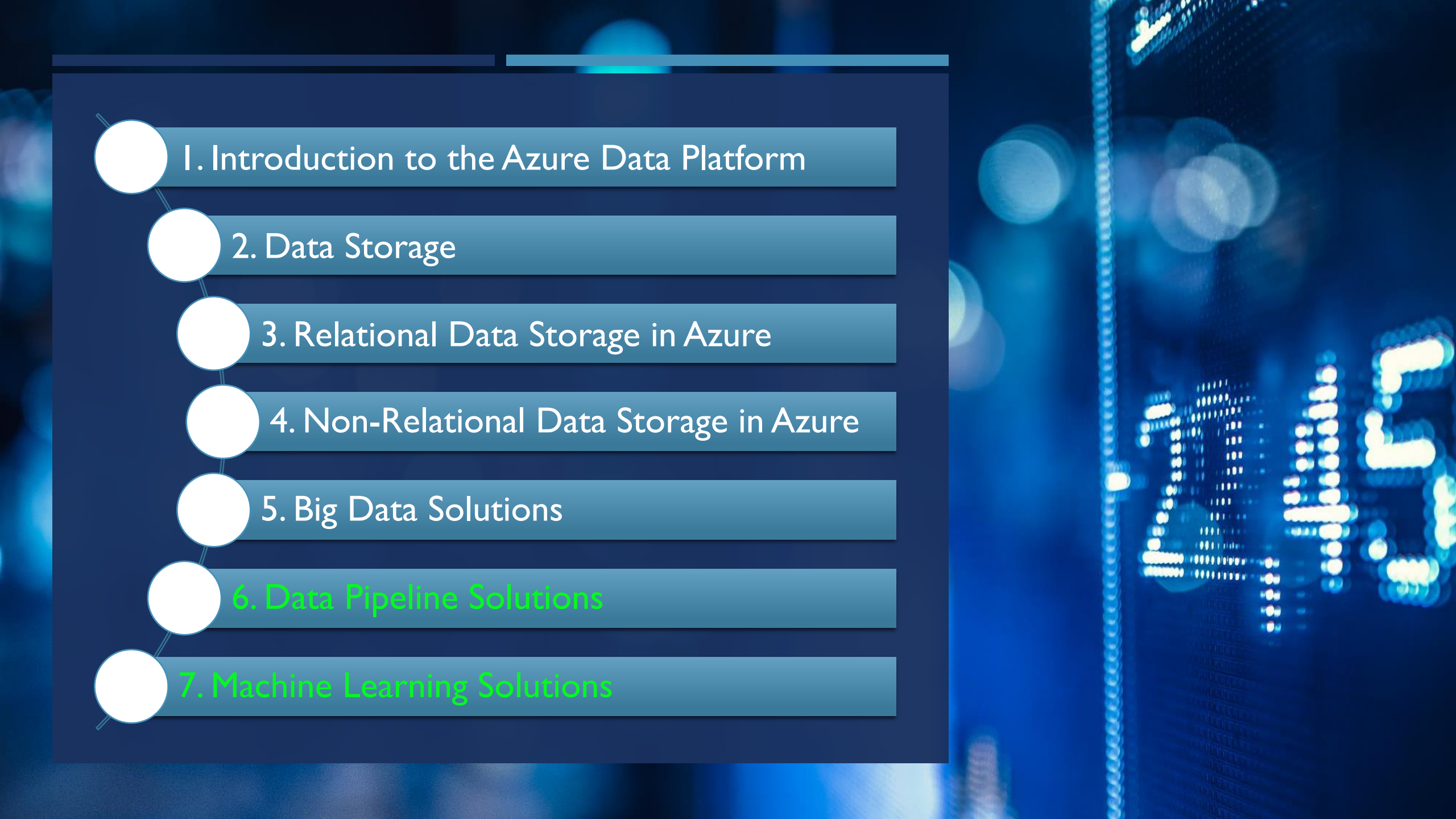
CLOUD – DATA SOLUTIONS - BIG DATA AND MACHINE LEARNING


1. Introduction to the Azure Data Platform
2. Data Storage
3. Relational Data Storage in Azure
4. Non-Relational Data Storage in Azure
5. Big Data Solutions
6. Data Pipeline Solutions
7. Machine Learning Solutions

DESCRIPTION

Capítulo 1. Introdução à Plataforma de Dados do Azure

- 1. Introdução
- 1.1. Modalidades de Serviços
- 1.2. Tipos de Dados
- 1.3. Perfis de Profissionais de Dados
- 1.4. Plataforma de Dados do Azure
- 1.5. Outros Serviços da Plataforma de Dados do Azure

Capítulo 2. Armazenamento de Dados

- 2.1. Storage Account
- 2.2. Criando uma Storage Account
- 2.3. Ingestão de Dados
- 2.4. Criando um Data Lake Storage Gen2

Capítulo 3. Armazenamento de Dados Relacionais no Azure

- 3.1. Bancos de Dados Relacionais em IaaS
- 3.2. Azure SQL Database Managed Instance
- 3.3. Azure SQL Database
- 3.4. Azure Cosmos DB
- 3.5. Bancos de Dados Open Source no Azure
- 3.6. Azure Synapse Analytics

Capítulo 4. Armazenamento de Dados Não Relacionais no Azure

- 4.1 Bancos de Dados Não Relacionais no Azure

Capítulo 5. Soluções de Big Data

- 5.1. Introdução ao Big Data
- 5.2. Introdução ao HDInsight
- 5.3. Aprovisionando um Ambiente do HDInsight
- 5.4. Introdução ao Azure DataBricks
- 5.5. Demonstração do Azure DataBricks

Capítulo 6. Soluções para Pipeline de Dados

- 6.1. Introdução ao Azure Data Factory
- 6.2. Criando um Pipeline de Dados com o Azure Data Factory

Capítulo 7. Soluções de Machine Learning

- 7.1 Overview do Azure Machine Learning

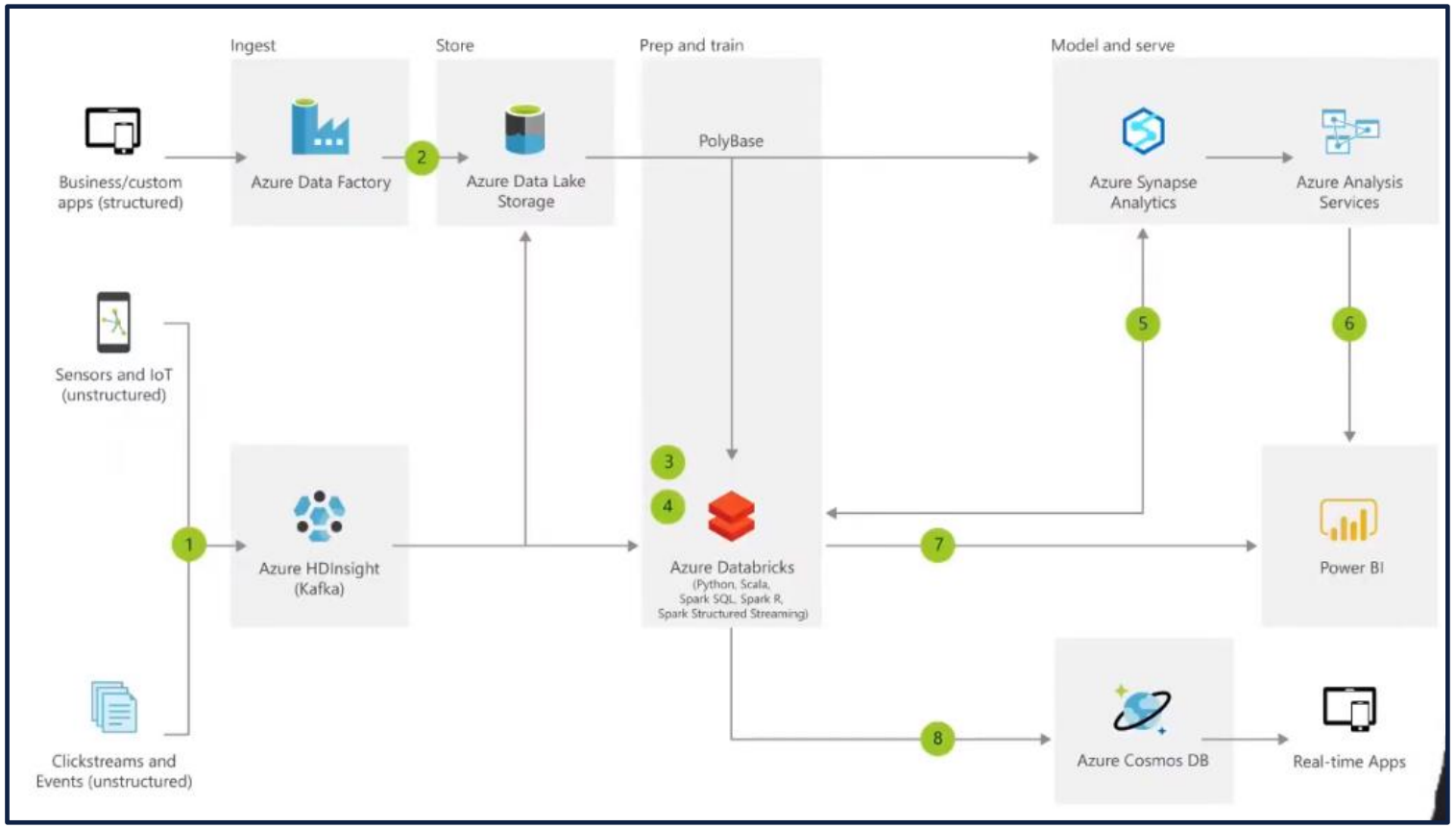
6. SOLUÇÕES PARA PIPELINE DE DADOS

6

6.1. INTRODUÇÃO AO AZURE DATA FACTORY

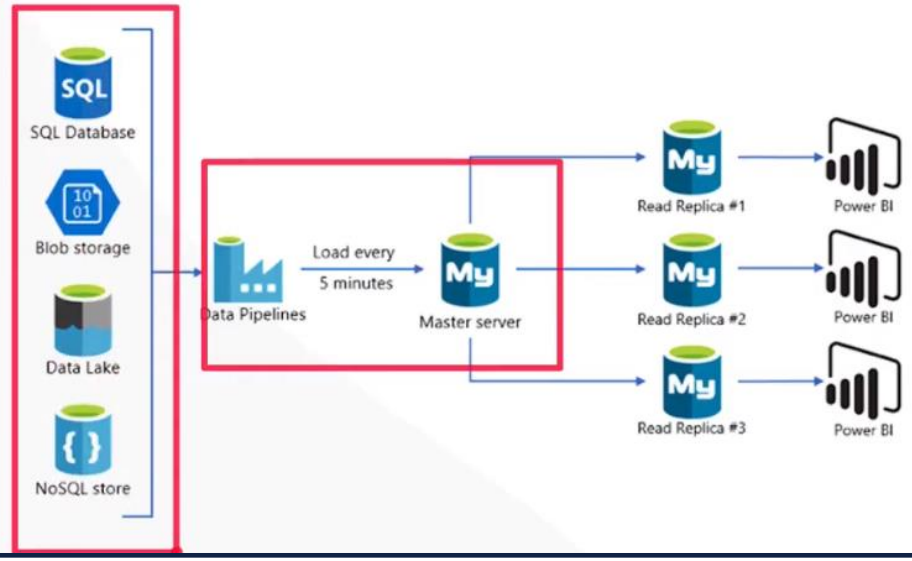
Overview do Azure Data Factory

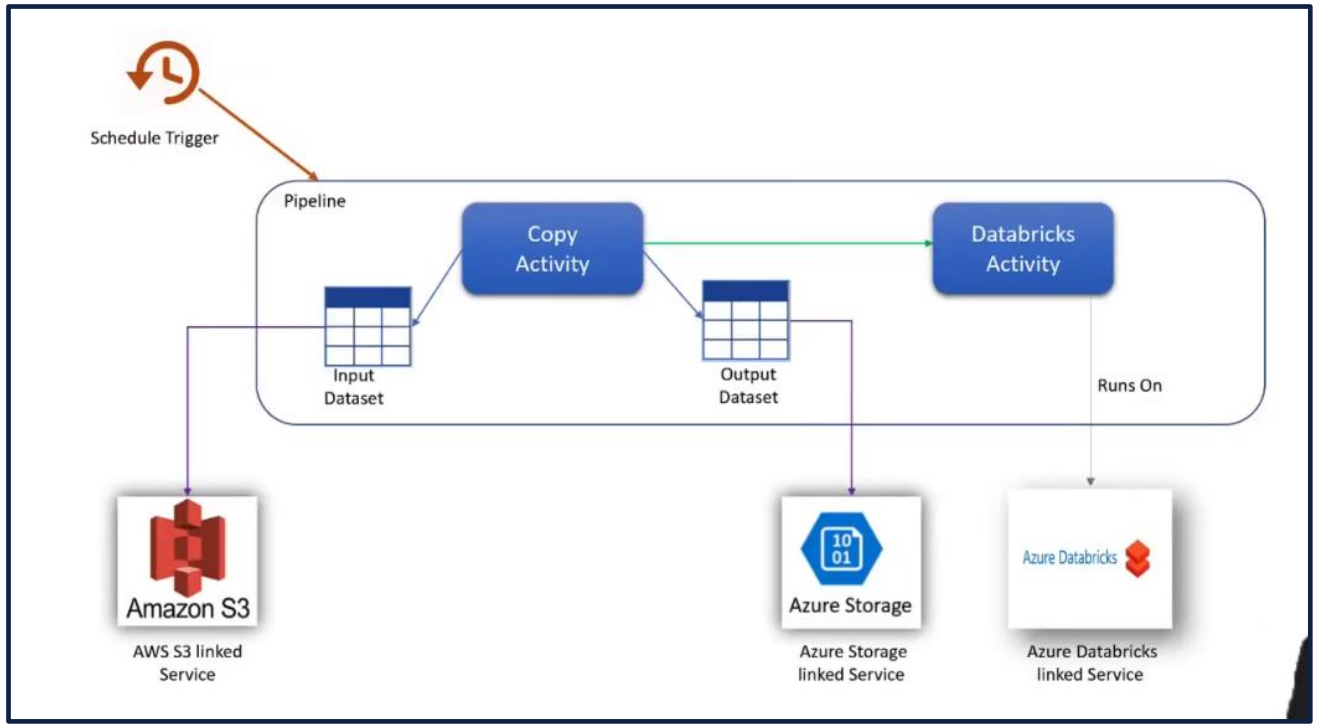
- Serviço de integração de dados e ETL baseado em nuvem;
- Conectores para mais de 90 tipos de fontes de dados diferentes;
- ETLs simples à complexos, com integração com Azure HDInsight, Azure Databricks ou Azure Synapse Analytics.



OVERVIEW DO AZURE DATA FACTORY

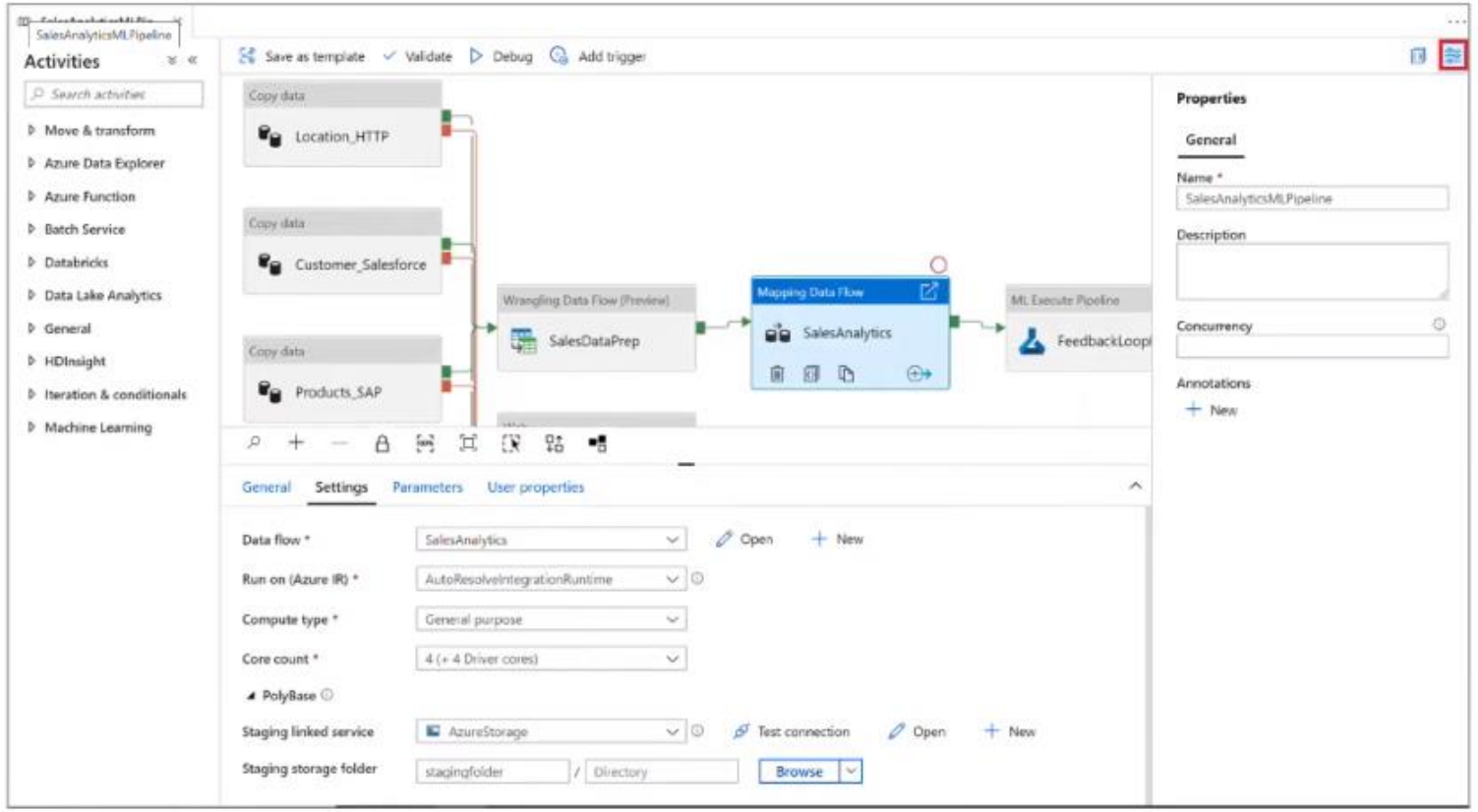
- Permite criar e agendar fluxos de trabalho orientados a dados para orquestrar a movimentação de dados e transformá-los;





OVERVIEW DO AZURE DATA FACTORY

- Possibilita a criação de pipelines de forma gráfica ou via código.



COMPONENTES DO AZURE DATA FACTORY

Componentes do Azure Data Factory

ATIVIDADE

- Representa uma etapa de processamento em um pipeline.
  - Atividade para copiar dados de um repositório de dados para outro;
  - Atividade que executa uma consulta de Hive em um cluster do Azure HDInsight para transformar ou analisar dados;
  - Etc.
- O Data Factory dá suporte a três tipos de atividades:
  - Atividades de movimentação de dados;
  - Atividades de transformação de dados;
  - Atividades de controle.

MAPEAMENTO DE FLUXO DE DADOS (MAPPING DATA FLOW)

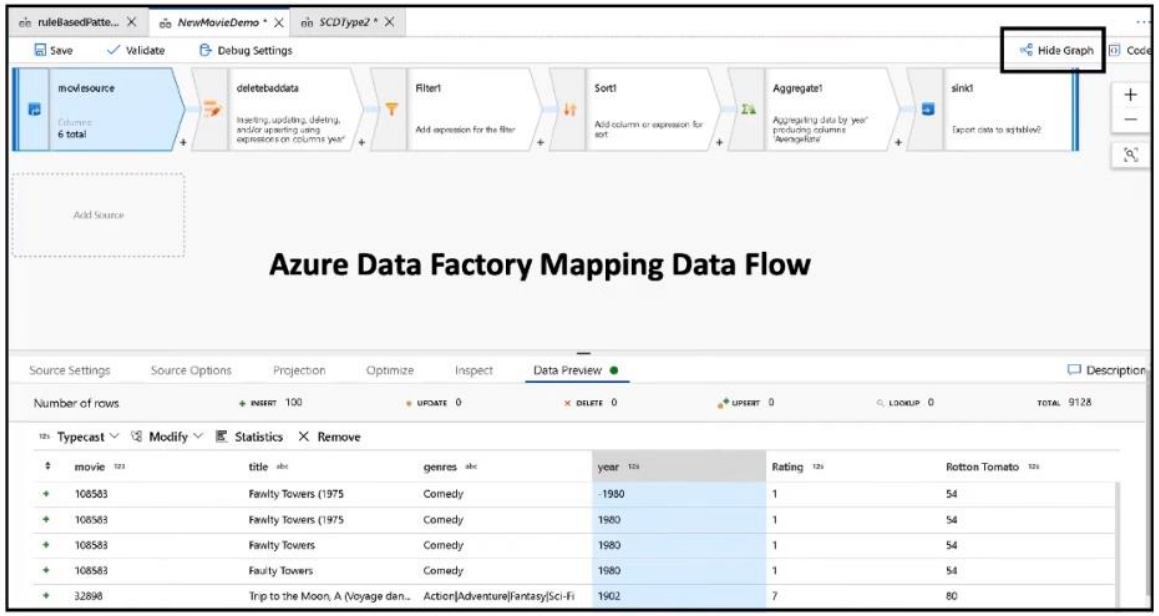
- São transformações de dados visualmente projetadas no Azure Data Factory;
- Permitem que os engenheiros de dados desenvolvam lógicas de transformação de dados sem escrever código;
- O Data Factory executará a lógica em um cluster Spark, autogerenciado pelo Azure, que será ativado e desativado quando necessário.

PIPELINE

- Agrupamento lógico de atividades que realiza uma unidade de trabalho. Juntas, as atividades em um pipeline executam uma tarefa.
- Exemplo: pipeline contém um grupo de atividades que ingere dados provenientes de um blob do Azure e, em seguida, executa uma consulta Hive em um cluster HDInsight para particionar os dados.
- Pipeline permite gerenciar atividades como um conjunto, em vez de gerenciar cada uma individualmente.
- Atividades podem operar de modo sequencial ou de forma independente, em paralelo.

COMPONENTES DO AZURE DATA FACTORY

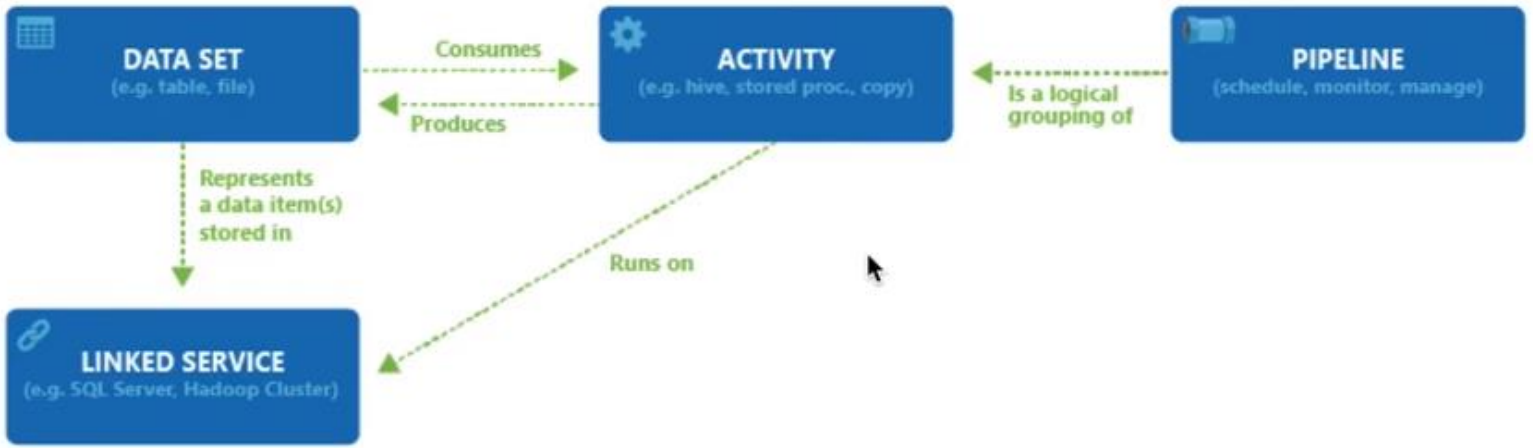
Data Factory



- **CONJUNTO DE DADOS (DATASET):** representam as estruturas de dados nos repositórios de dados, que simplesmente apontam para ou fazem referência aos dados que deseja-se usar em atividades, seja como entrada ou saída.
- **SERVIÇO VINCULADO (LINKED SERVICE):** define as informações de conexão necessárias para que o Data Factory se conecte aos recursos externos. Duas finalidades:
  - Para representar um **armazenamento de dados:** ex. banco SQL / Oracle;
  - Para representar um **recurso de computação** que pode hospedar a execução de uma atividade: ex. um cluster Hadoop do HDInsight, onde a atividade HDInsightHive é executada.

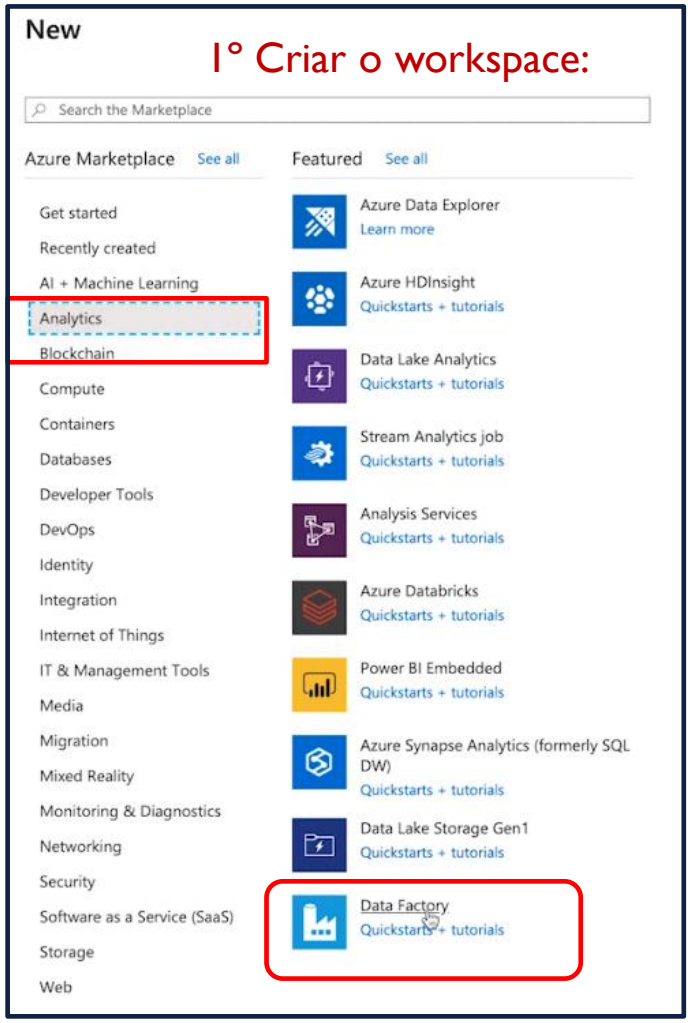
COMPONENTES DO AZURE DATA FACTORY

- Um serviço vinculado define a conexão à fonte de dados e um conjunto de dados representa a estrutura dos dados.
  - Por exemplo, um serviço vinculado de armazenamento do azure especifica a string de conexão para conectar-se à conta de Armazenamento do Azure (Storage Account), e um conjunto de dados de blob do Azure especifica o contêiner de blob e a pasta que contém os dados.

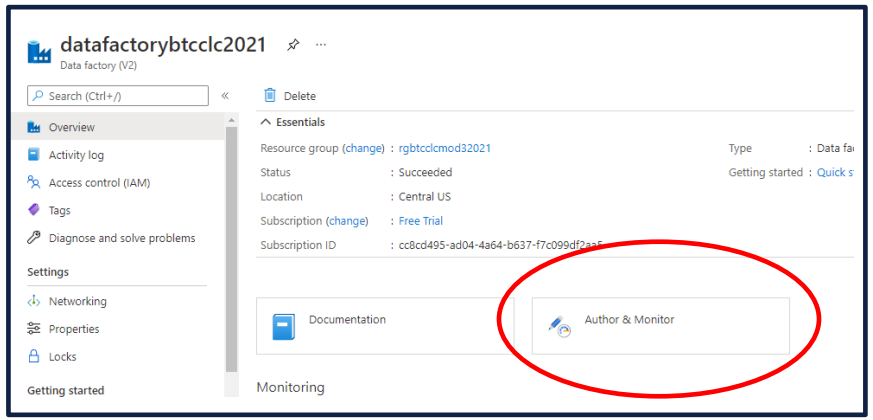


6.2. CRIANDO UM PIPELINE DE DADOS COM O AZURE DATA FACTORY

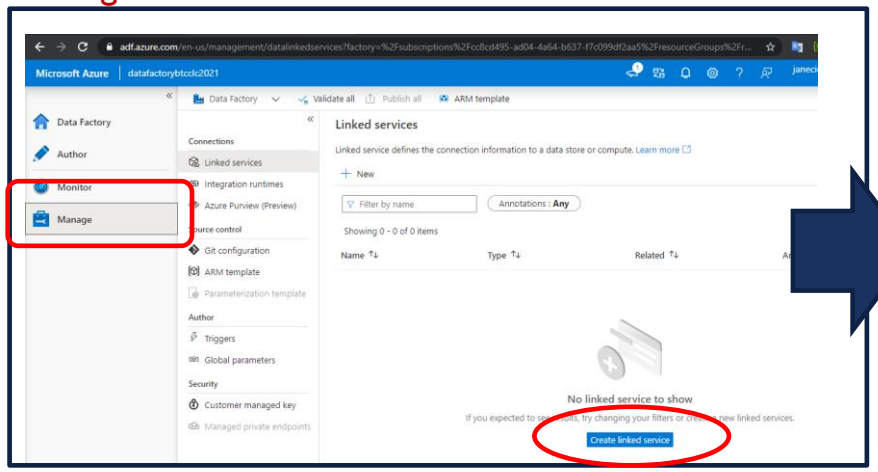
1° Criar o workspace:



2° ir para o ambiente de criação dos pipelines:

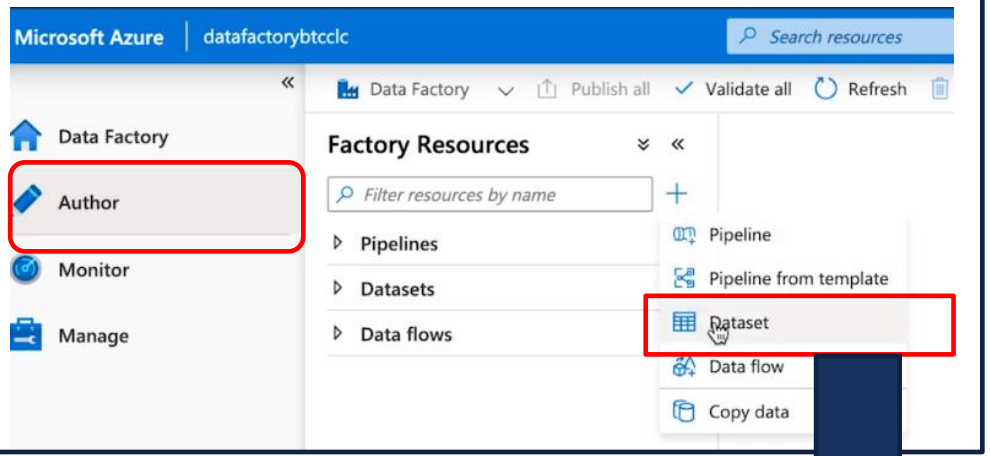


3° Conectores: de onde os dados serão extraídos e onde serão gravados:



CRIANDO UM PIPELINE DE DADOS COM O AZURE DATA FACTORY

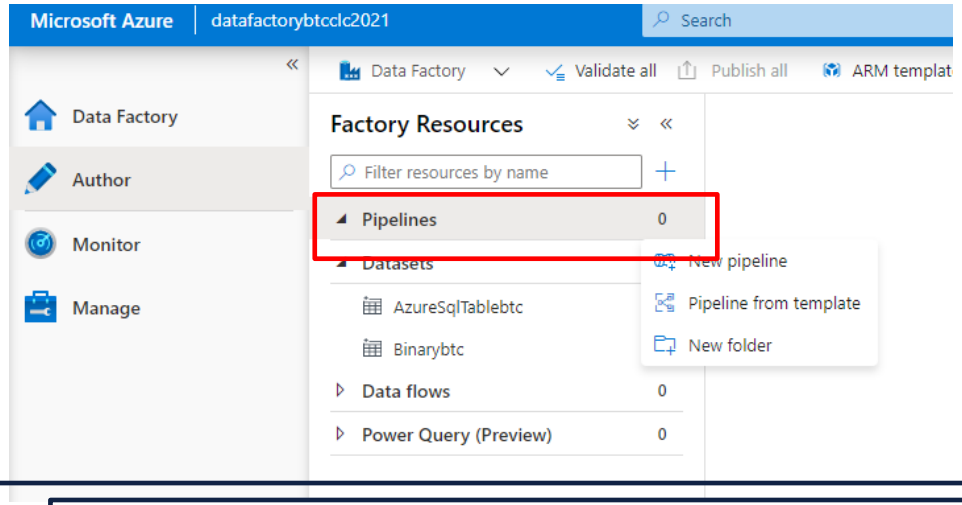
4° Author: onde são criados os pipelines:
Criar o dataset que irá apontar para os serviços vinculados.



Um para o
- BlobStorage e outro para o
- SQL Database

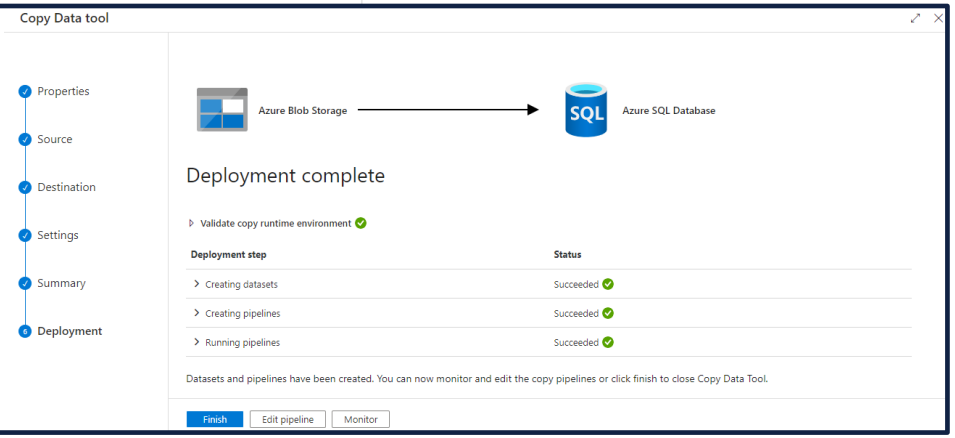
Name	Type
AzureBlobStoragebtc	Azure Blob Storage
AzureSqlDatabasebtc	Azure SQL Database

5° Criar pipelines:



Existem Tamplates prontos de pipeline no Azure Data Factory!!

6° Exemplo de Cópia de dados:

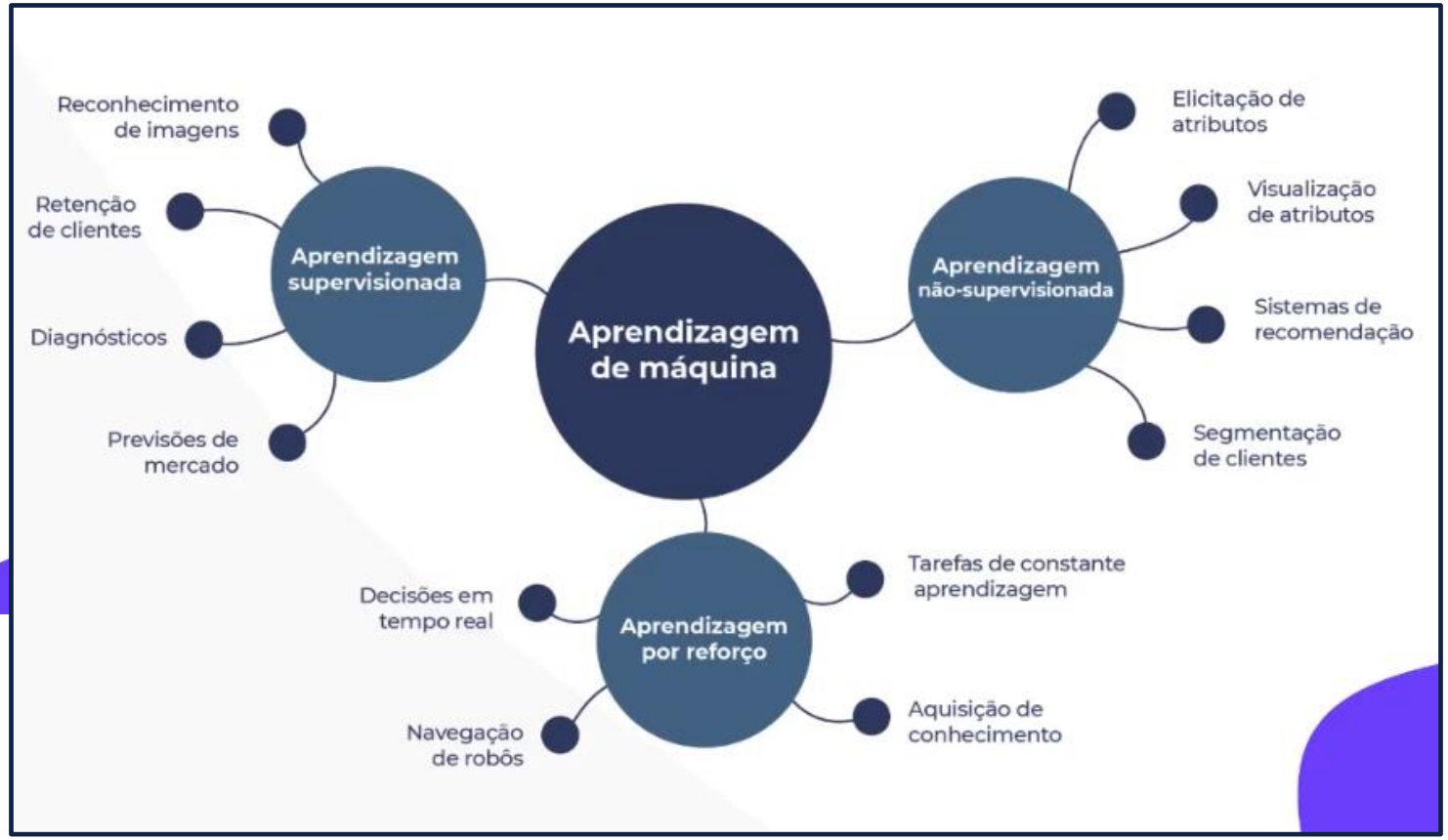


7. SOLUÇÕES DE MACHINE LEARNING

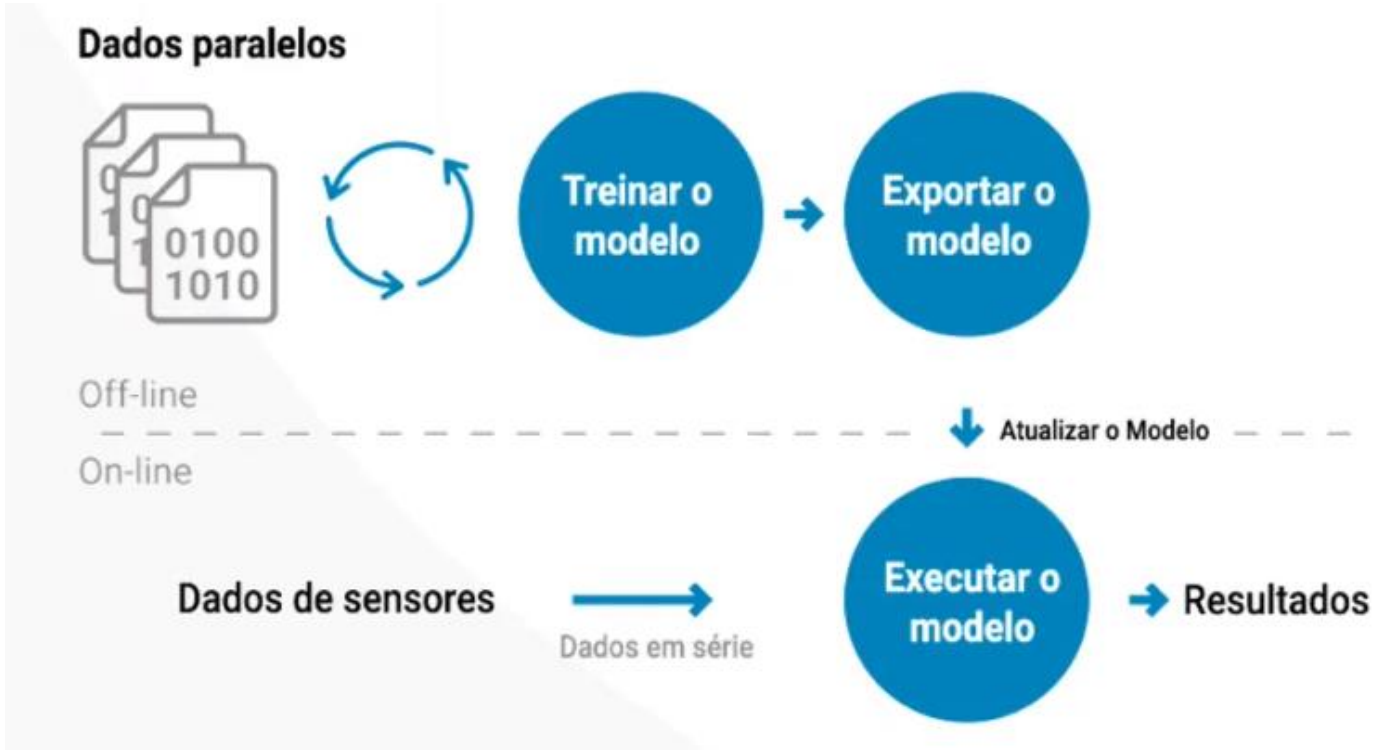
7

7.1 OVERVIEW DO AZURE MACHINE LEARNING

- Aprendizado de máquina (Machine Learning - ML) é uma técnica da ciência de dados que permite que os computadores usem os dados existentes para prever tendências, resultados e comportamentos futuros.
- Usando ML, os computadores têm a capacidade de aprender de acordo com as respostas esperadas por meio das associações de diferentes dados, os quais podem ser imagens, áudio, números, etc.



OVERVIEW DO AZURE MACHINE LEARNING

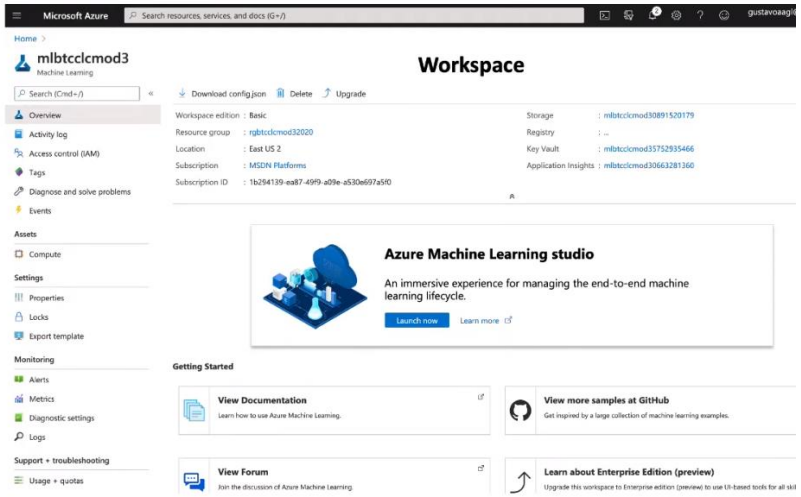


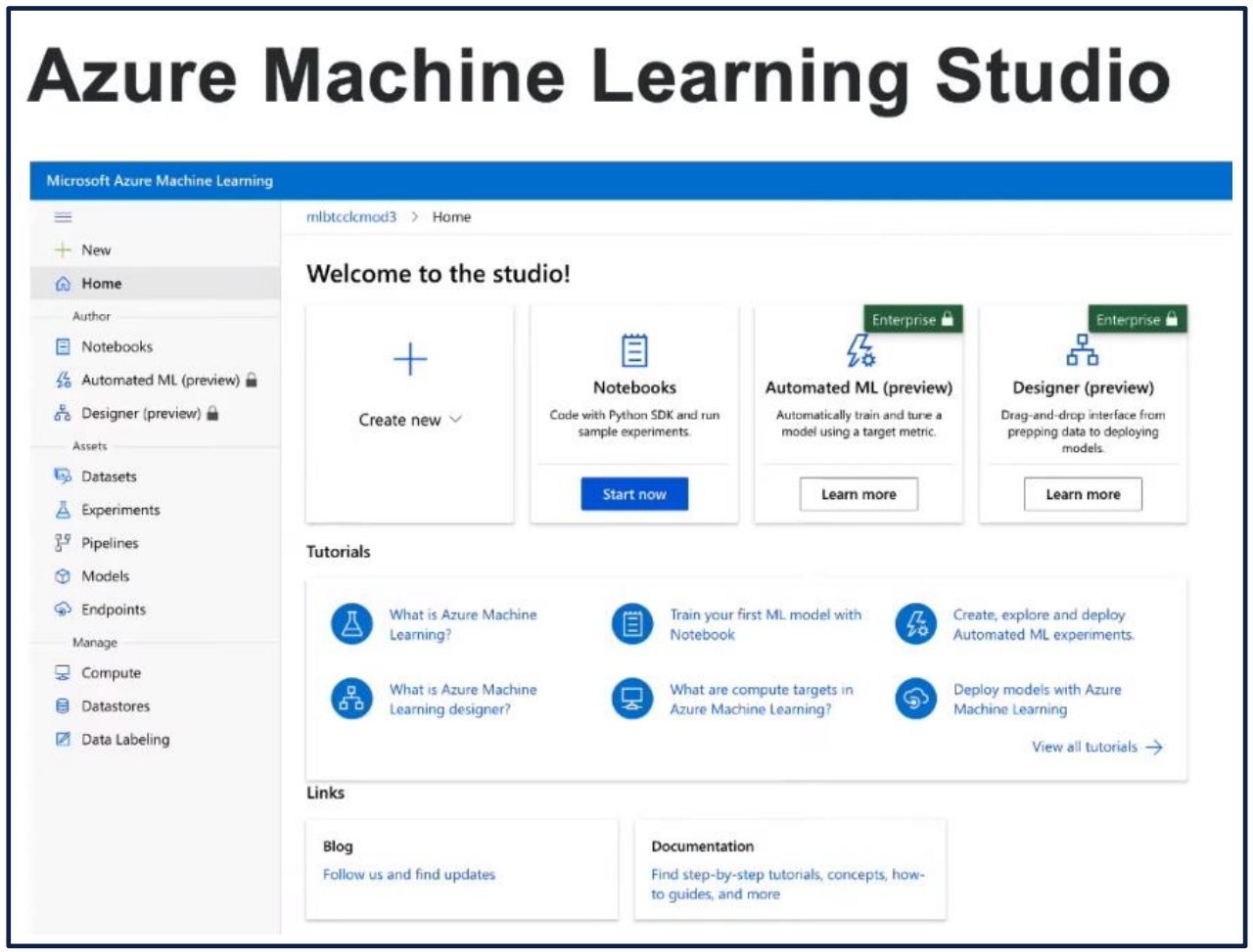
Azure Machine Learning

- Ambiente baseado em nuvem que pode ser usado para treinar, implantar, automatizar, gerenciar e rastrear modelos de ML;
- Pode ser usado para qualquer tipo de aprendizado de máquina, desde ML clássico até aprendizado profundo, aprendizado supervisionado e não supervisionado.

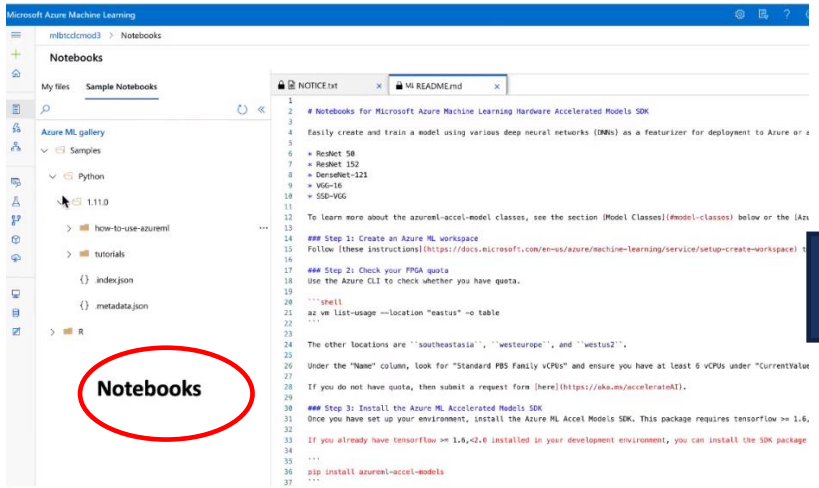
OVERVIEW DO AZURE MACHINE LEARNING

1° Criar um workspace:





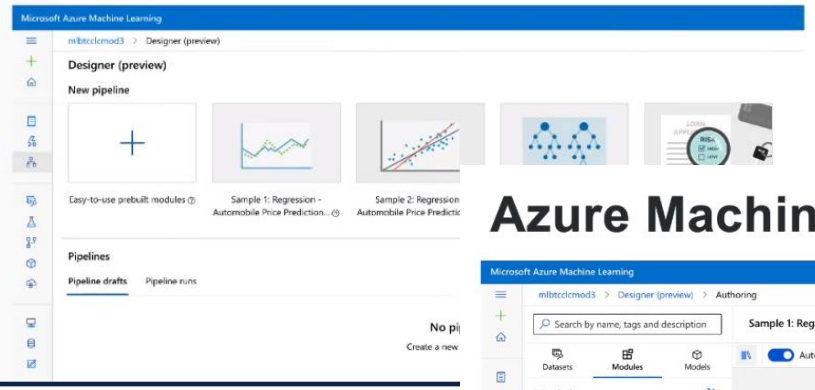
OVERVIEW DO AZURE MACHINE LEARNING



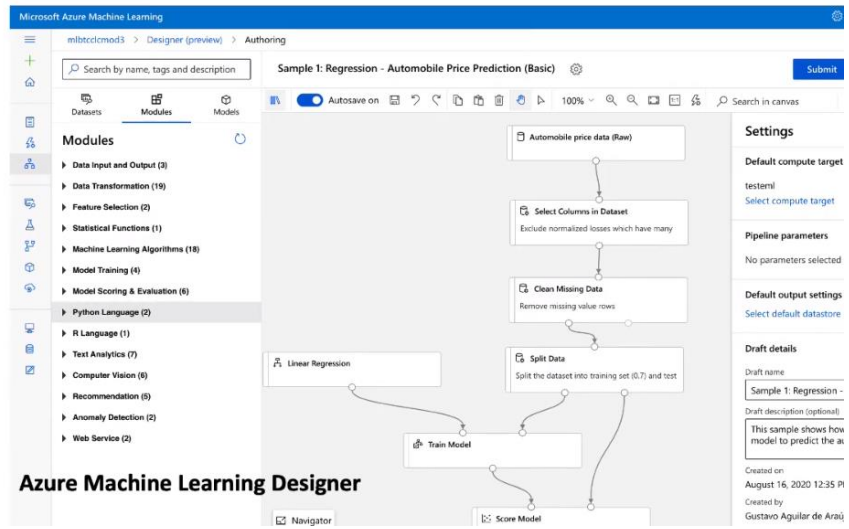
Ou

AZURE MACHINE LEARNING DESIGNER

- Preparar dados, treinar, testar, implantar, gerenciar e rastrear modelos de aprendizado de máquina sem escrever nenhum código.



Azure Machine Learning Studio

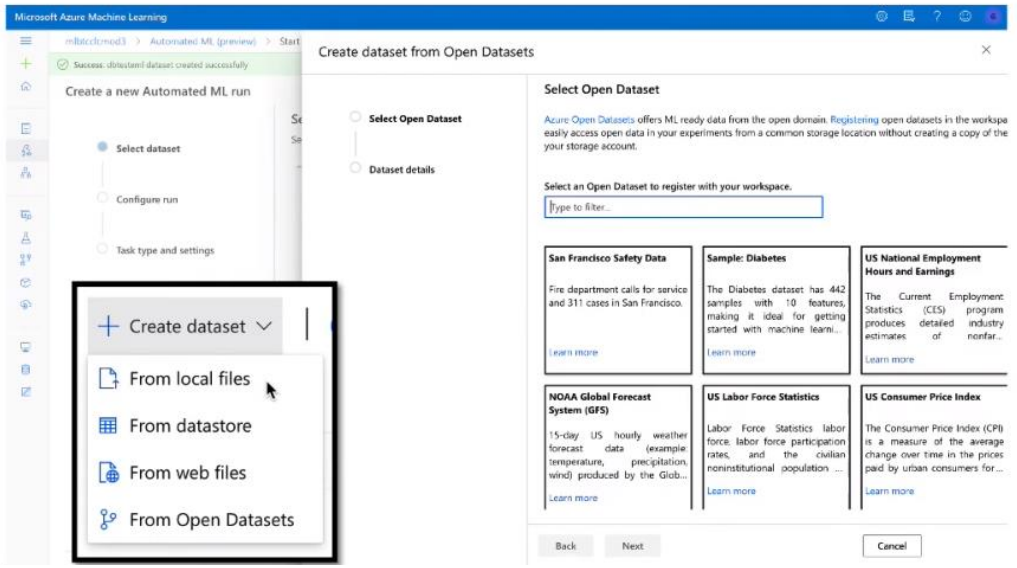


OVERVIEW DO AZURE MACHINE LEARNING

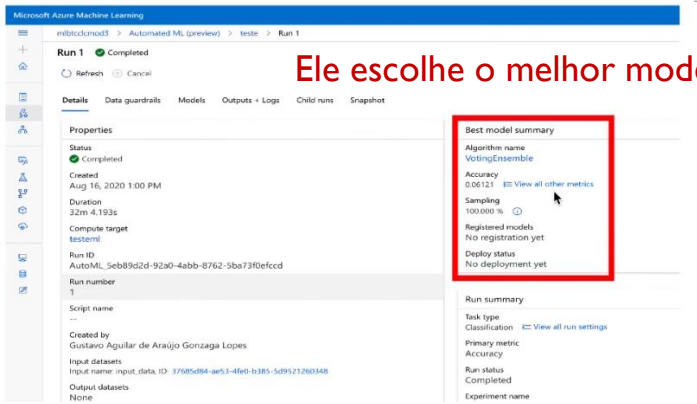
MACHINE LEARNING AUTOMATIZADO

- Automatizar tarefas intensivas e demoradas;
- Construção drag & drop (interface com componentes prontos);
- Realiza a iteração, de forma rápida, entre várias combinações de algoritmos e parâmetros, para ajudar a encontrar o melhor modelo com base em uma métrica selecionada;
- Somente na assinatura **Enterprise,** assim como o Azure Machine Learning Designer.

MACHINE LEARNING AUTOMATIZADO



MACHINE LEARNING AUTOMATIZADO



Ele escolhe o melhor modelo!

THANKS