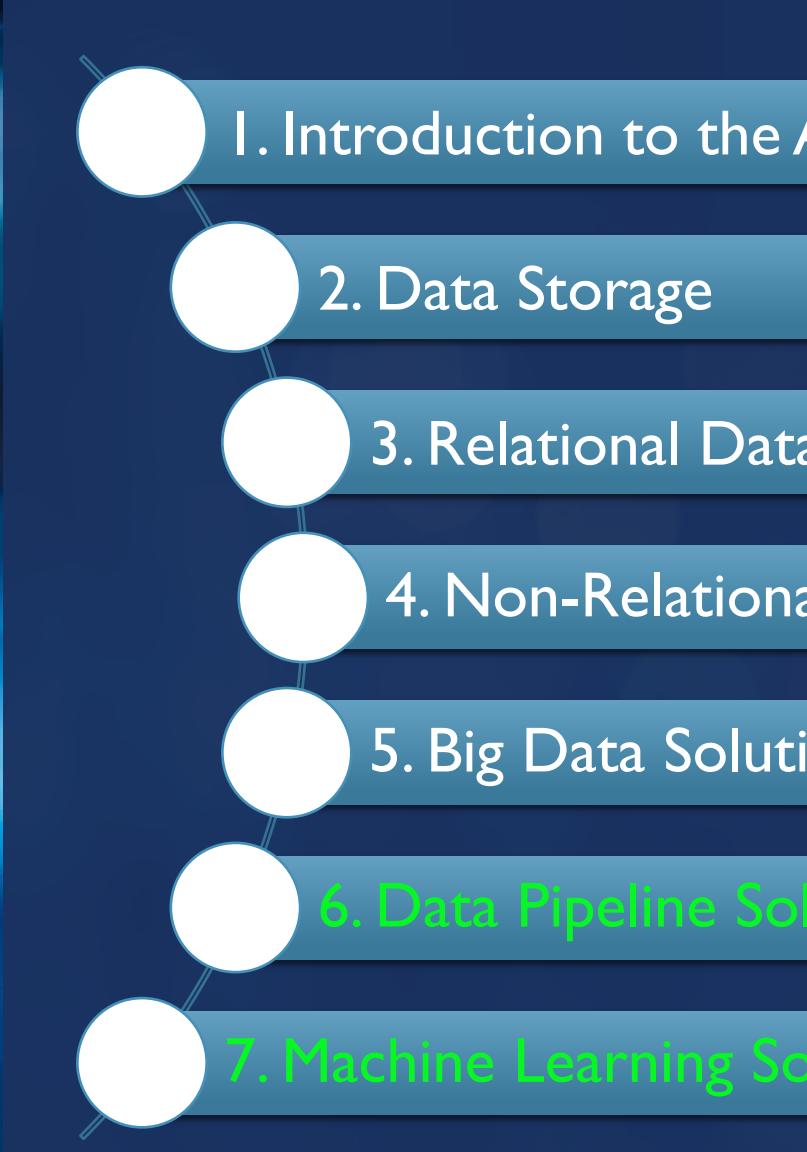




CLOUD – DATA SOLUTIONS - BIG DATA AND
MACHINE LEARNING

- 
- 
- I. 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

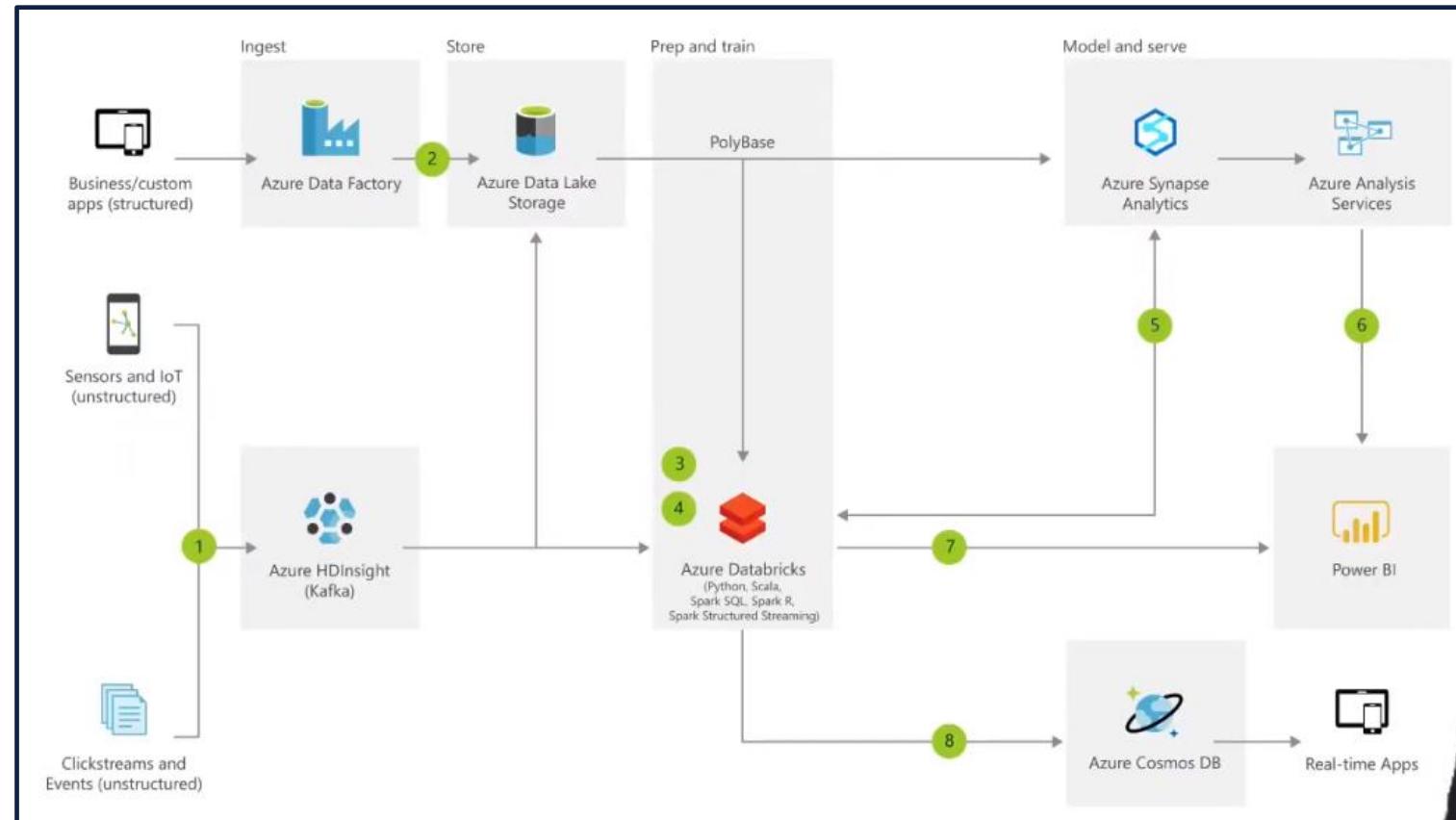
- 7.1 Overview do Azure Machine Learning

6. SOLUÇÕES PARA PIPELINE DE DADOS

6.1. INTRODUÇÃO AO AZURE DATA FACTORY

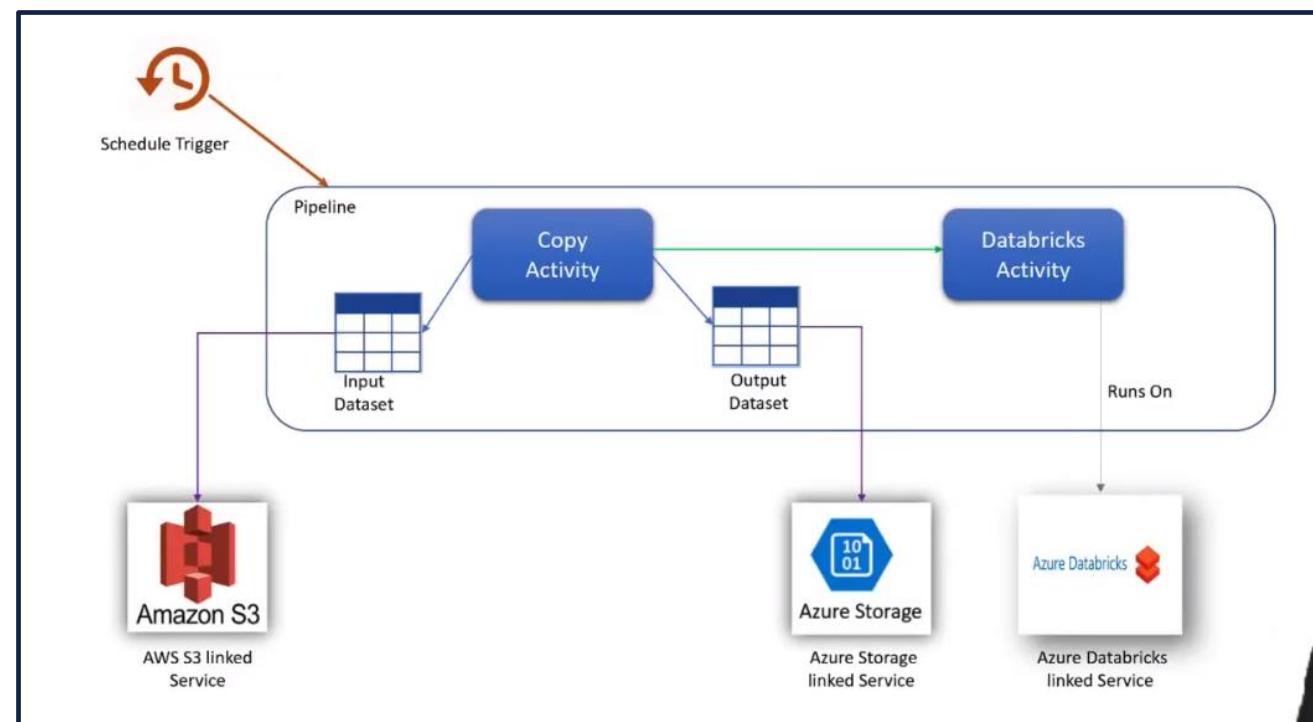
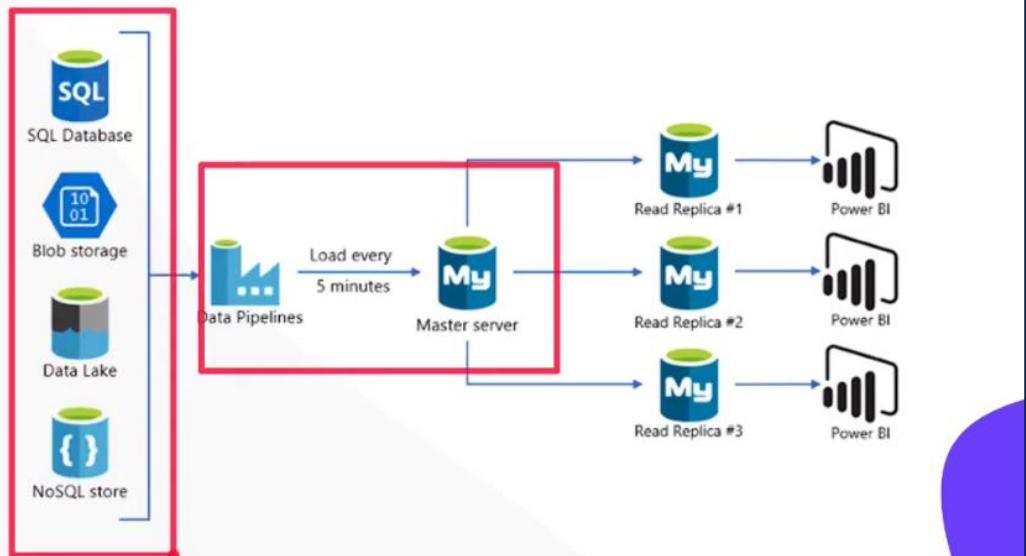
Overview do Azure Data Factory

- 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.



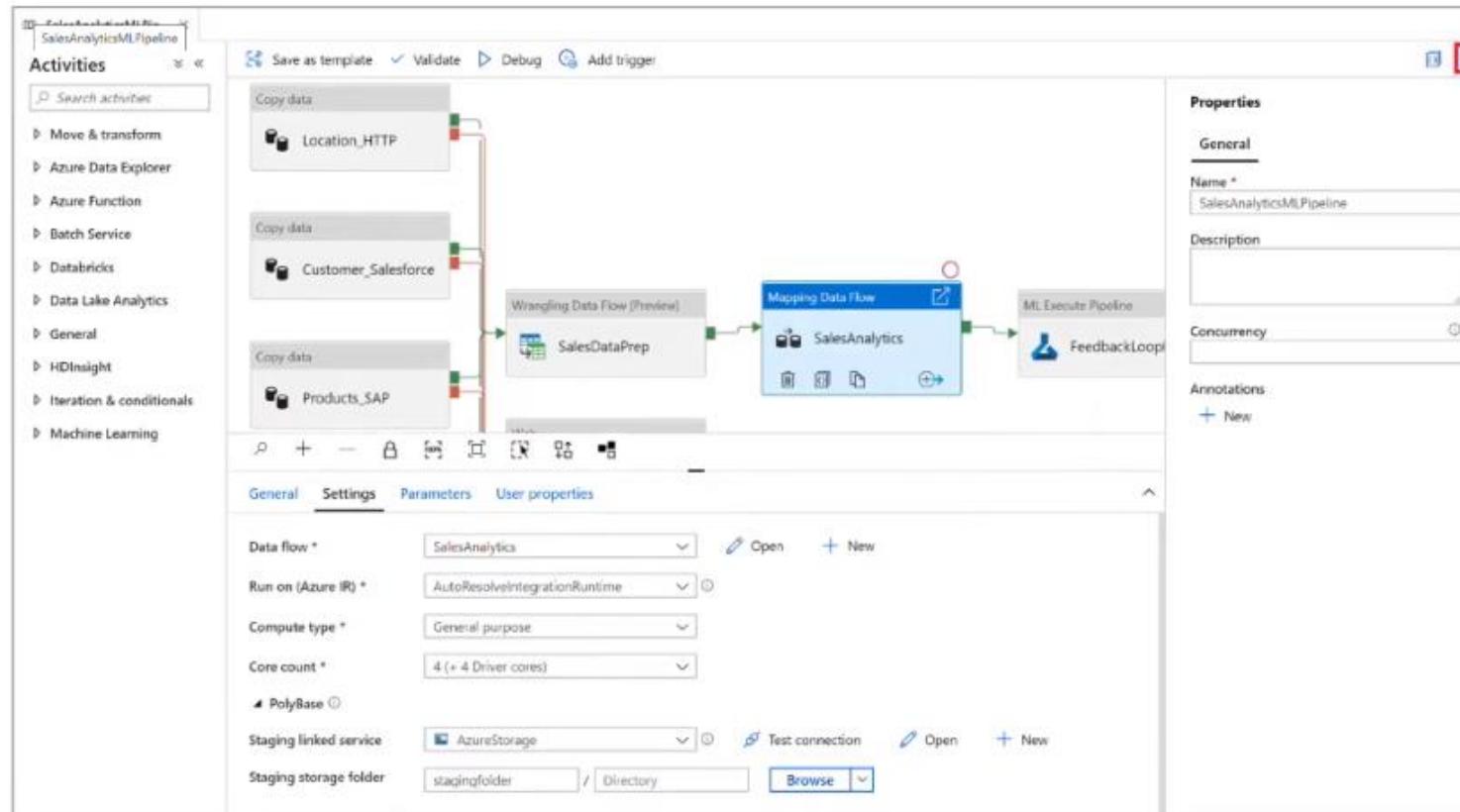
OVERVIEW DO AZURE DATA FACTORY

- 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)

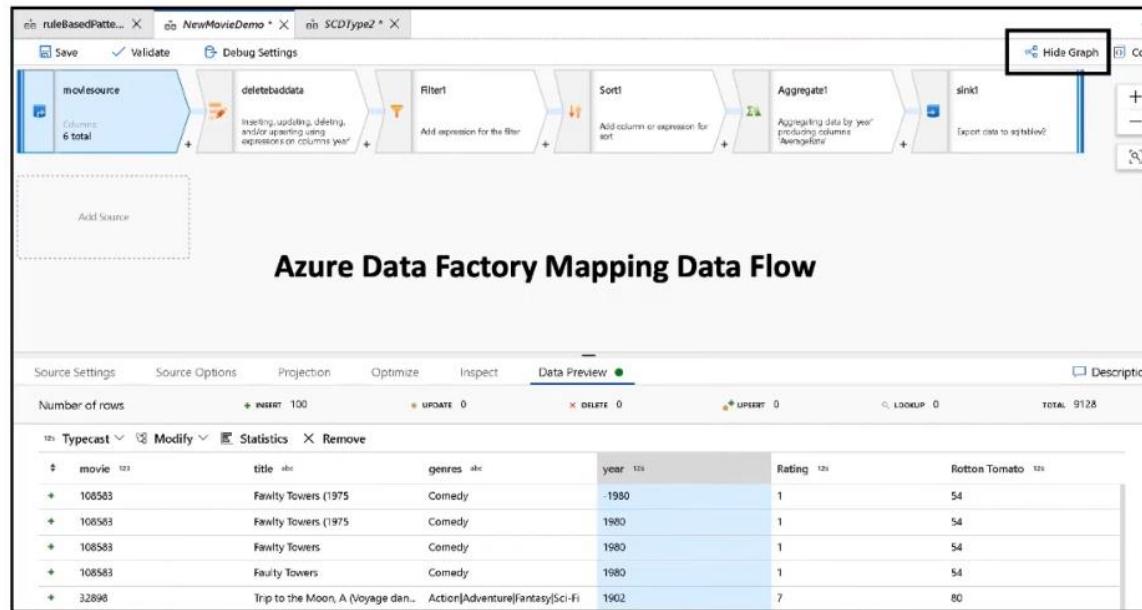
- São transformações de dados visualmente projetadas no Azure Data Factory;
- Permitem que os engenheiros de dados desenvolvam lógicas de transformação de dados sem escrever código;
- O Data Factory executará a lógica em um cluster Spark, autogerenciado pelo Azure, que será ativado e desativado quando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

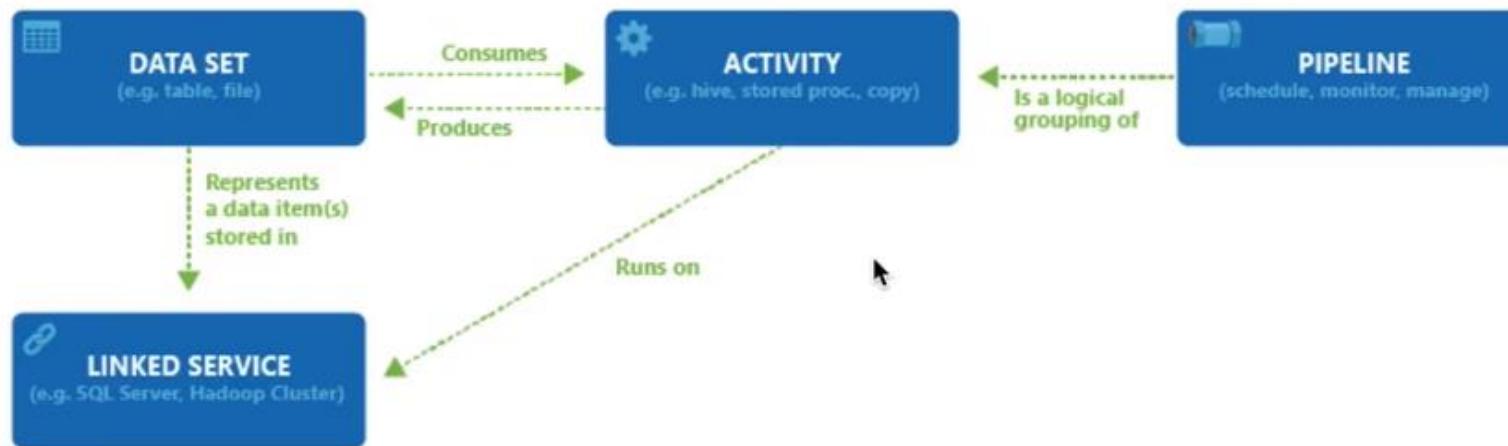
Data Factory



- **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

New

I° Criar o workspace:

The screenshot shows the Azure Marketplace interface. On the left, there's a sidebar with categories like 'Get started', 'Recently created', 'AI + Machine Learning', 'Analytics' (which is highlighted with a red box), 'Blockchain', 'Compute', 'Containers', 'Databases', 'Developer Tools', 'DevOps', 'Identity', 'Integration', 'Internet of Things', 'IT & Management Tools', 'Media', 'Migration', 'Mixed Reality', 'Monitoring & Diagnostics', 'Networking', 'Security', 'Software as a Service (SaaS)', 'Storage', and 'Web'. At the bottom right, there's a 'Data Factory' item with a red box around it.

2° ir para o ambiente de criação dos pipelines:

The screenshot shows the 'datafactorybtcclc2021' Data factory (V2) overview page. It displays basic information such as Resource group (change), Status, Location, Subscription (change), and Subscription ID. Below this, there are sections for 'Documentation' and 'Author & Monitor'. A red circle highlights the 'Author & Monitor' button.

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

The screenshot shows the 'Connections' page under the 'datafactorybtcclc2021' Data Factory. It lists various services like 'Monitor', 'Manage', 'Integration runtimes', 'Azure Purview (Preview)', 'Data control', 'Git configuration', 'ARM template', 'Parameterization template', 'Author', 'Triggers', 'Global parameters', 'Security', 'Customer managed key', and 'Managed private endpoints'. A red box highlights the 'Manage' button. At the bottom, there's a 'Create linked service' button with a red circle around it.

A screenshot of a table titled 'Linked services' showing two entries: 'AzureBlobStoragebtc' (Type: Azure Blob Storage) and 'AzureSqlDatabasebtc' (Type: Azure SQL Database). Red arrows point from the 'Manage' button on the previous screen to each of these entries in the table.

Name	Type
AzureBlobStoragebtc	Azure Blob Storage
AzureSqlDatabasebtc	Azure SQL Database

CRIANDO UM PIPELINE DE DADOS COM O AZURE DATA FACTORY

4º Author: onde são criados os pipelines:

Criar o dataset que irá apontar para os serviços vinculados.

The screenshot shows the Microsoft Azure Data Factory 'Author' interface. On the left, there's a sidebar with 'Data Factory', 'Author' (highlighted with a red box), 'Monitor', and 'Manage'. The main area is titled 'Factory Resources' with a search bar. It lists 'Pipelines', 'Datasets' (highlighted with a red box), and 'Data flows'. Below these are 'Pipeline from template', 'Dataset' (highlighted with a red box and has a cursor icon over it), 'Data flow', and 'Copy data'. A large blue arrow points downwards from this interface to the 'Copy Data tool' interface shown in the bottom right.

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

Name	Type
AzureBlobStoragebtc	Azure Blob Storage
AzureSqlDatabasebtc	Azure SQL Database

5º Criar pipelines:

The screenshot shows the Microsoft Azure Data Factory 'Pipelines' interface. On the left, there's a sidebar with 'Data Factory', 'Author' (highlighted with a red box), 'Monitor', and 'Manage'. The main area is titled 'Factory Resources' with a search bar. It lists 'Pipelines' (highlighted with a red box), 'Datasets', 'Data flows', and 'Power Query (Preview)'. A 'New pipeline' button is visible next to the Pipelines section. A large blue arrow points from the 'Author' interface above to this 'Pipelines' interface.

6º Exemplo de
Cópia de dados:

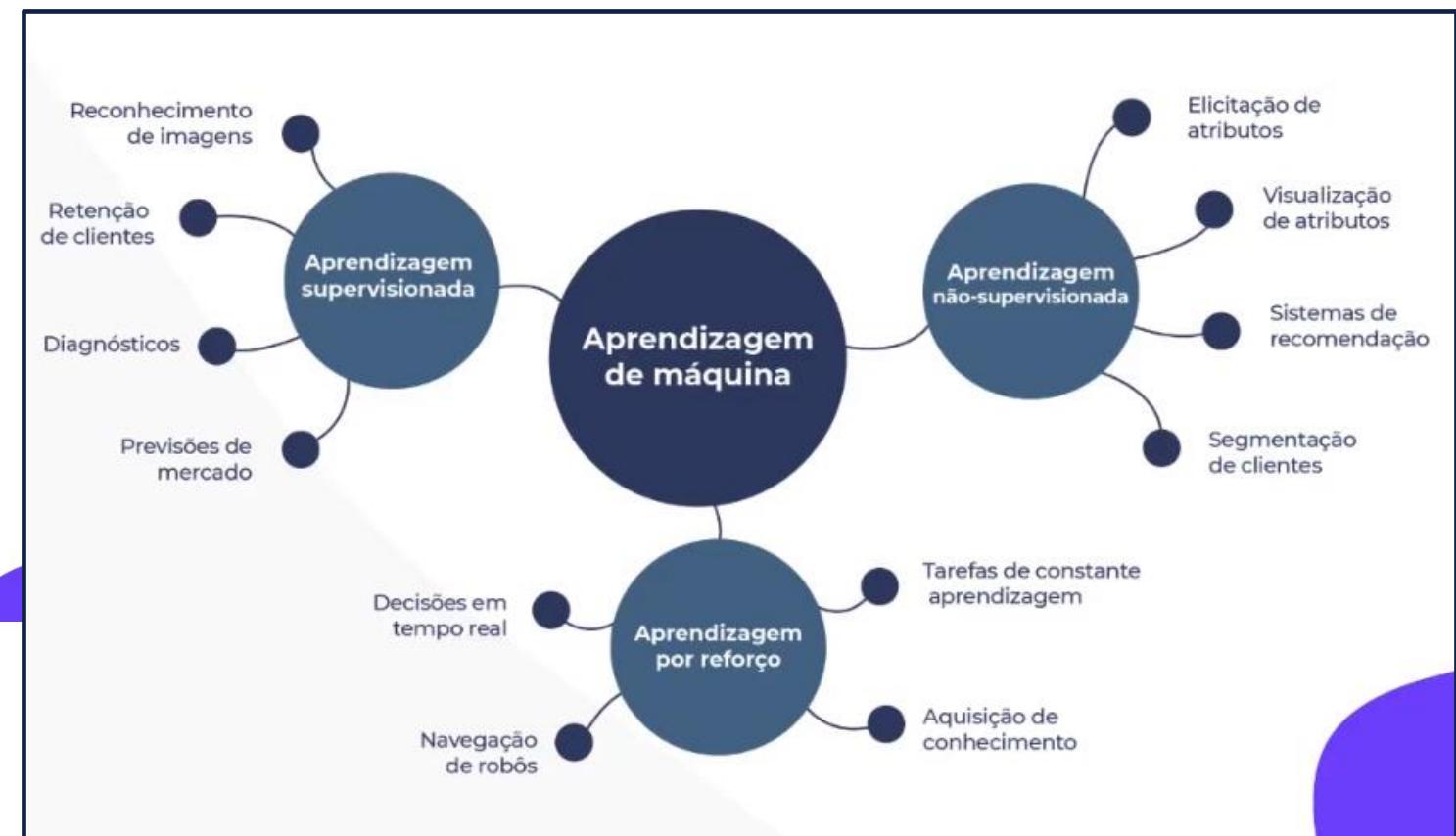
The screenshot shows the 'Copy Data tool' interface. It displays a flow diagram from 'Azure Blob Storage' to 'Azure SQL Database'. The steps are: Properties, Source, Destination, Settings, Summary, and Deployment. The status for each step is 'Deployment complete' with a green checkmark. Below the steps, it says 'Deployment complete'. At the bottom, there are buttons for 'Finish', 'Edit pipeline', and 'Monitor'.

Existem
Templates
prontos de
pipeline no
Azure
Data
Factory!!

7. SOLUÇÕES DE MACHINE LEARNING

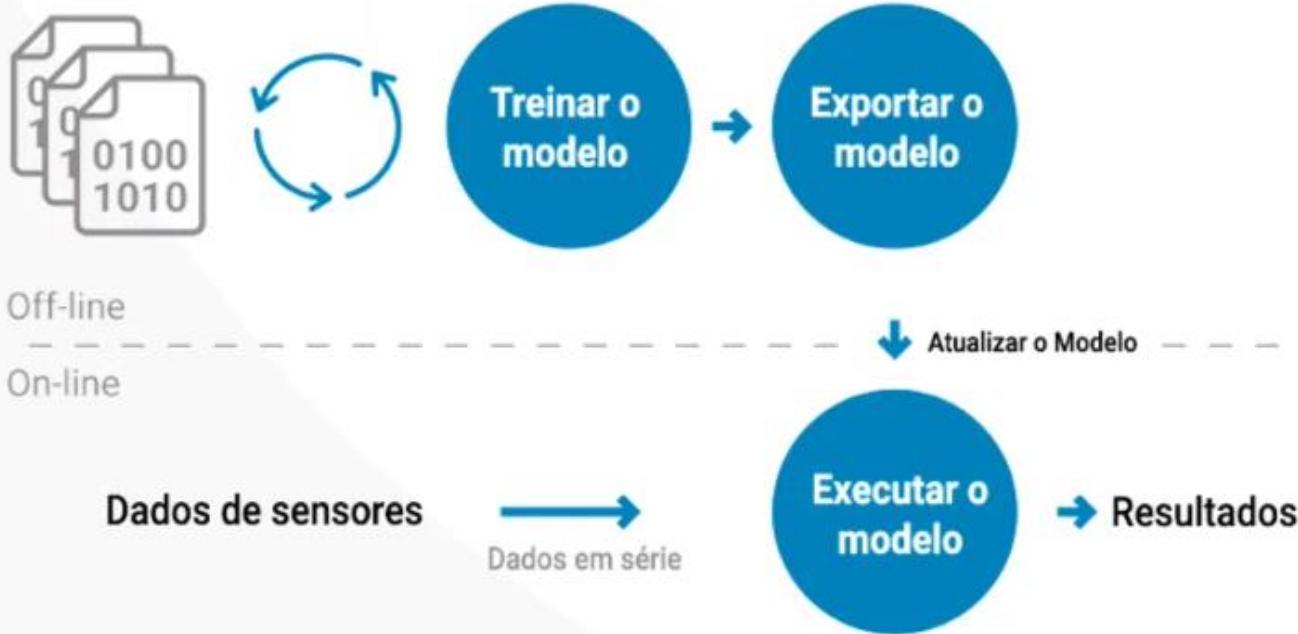
7.1 OVERVIEW DO AZURE MACHINE LEARNING

- 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

Dados paralelos



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.



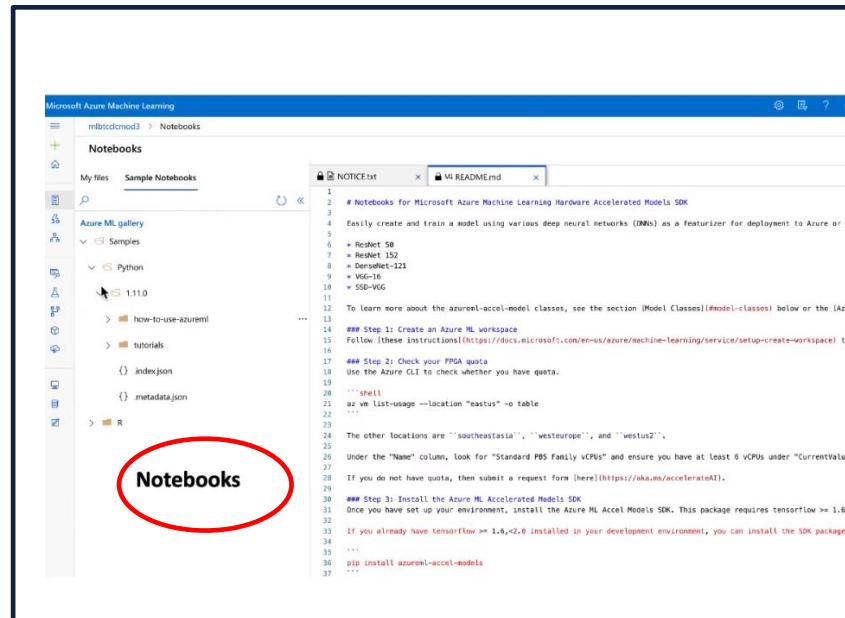
OVERVIEW DO AZURE MACHINE LEARNING

1º Criar um workspace:

The screenshot shows the Microsoft Azure portal interface. On the left, there's a sidebar with various service icons like Home, Machine Learning, Compute, Datasets, etc. The main area is titled 'Workspace' and shows details for 'mlbtcclmod3'. It includes sections for Overview, Activity log, Access control (IAM), Tags, and Diagnose and solve problems. A large central box is titled 'Azure Machine Learning studio' with a sub-section 'Getting Started' containing links to 'View Documentation', 'View more samples at GitHub', and 'Learn about Enterprise Edition (preview)'. A large blue arrow points from this workspace screen to the right.

The screenshot shows the 'Azure Machine Learning Studio' home page. The top navigation bar has 'mlbtcclmod3 > Home'. The main header says 'Welcome to the studio!'. On the left, there's a sidebar with 'New' (selected), 'Home' (disabled), 'Author', 'Notebooks', 'Automated ML (preview)', 'Designer (preview)', 'Assets', 'Datasets', 'Experiments', 'Pipelines', 'Models', 'Endpoints', 'Manage', 'Compute', 'Datastores', and 'Data Labeling'. The main content area features four cards: 'Notebooks' (with a 'Create new' button and 'Start now' button), 'Automated ML (preview)' (with a 'Start now' button and 'Learn more' button), 'Designer (preview)' (with a 'Learn more' button), and 'Tutorials' (with six items: 'What is Azure Machine Learning?', 'Train your first ML model with Notebook', 'Create, explore and deploy Automated ML experiments.', 'What is Azure Machine Learning designer?', 'What are compute targets in Azure Machine Learning?', and 'Deploy models with Azure Machine Learning'). At the bottom, there are 'Links' for 'Blog' and 'Documentation'.

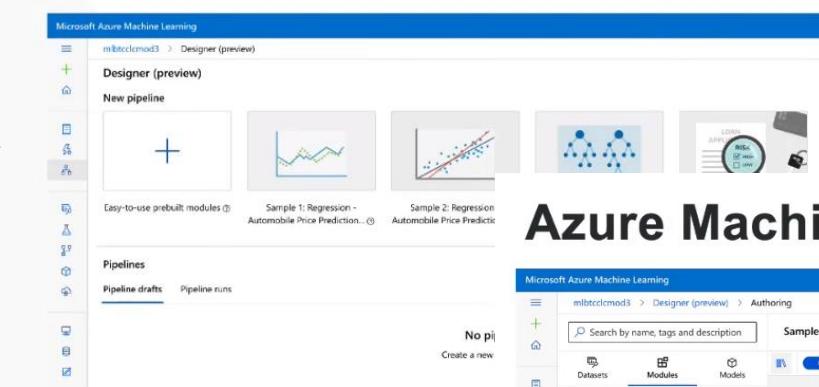
OVERVIEW DO AZURE MACHINE LEARNING



A screenshot of the Microsoft Azure Machine Learning studio interface. On the left, there's a sidebar with 'My files' and 'Samples' sections. The 'Notebooks' section is highlighted with a red oval. A large blue arrow points from this section towards the center of the screen.

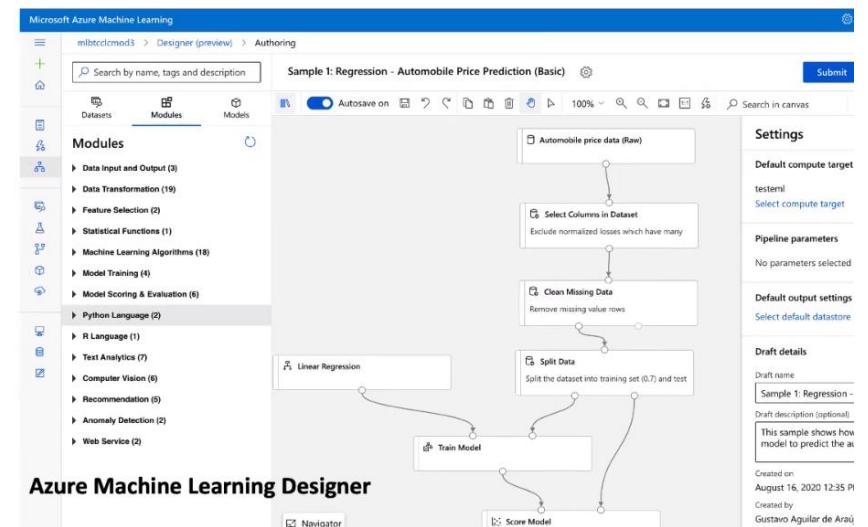
AZURE MACHINE LEARNING DESIGNER

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



A screenshot of the Azure Machine Learning Designer interface. It shows a 'Designer (preview)' window with a 'New pipeline' section. There are three sample pipelines: 'Easy-to-use prebuilt modules', 'Sample 1: Regression - Automobile Price Prediction...', and 'Sample 2: Regression Automobile Price Predict...'. Below this is a 'Pipelines' section with 'Pipeline drafts' and 'Pipeline runs' tabs. A 'Modules' catalog on the left lists various machine learning and data processing modules.

Azure Machine Learning Studio



A screenshot of the Azure Machine Learning Studio interface. It shows a detailed pipeline for a 'Regression - Automobile Price Prediction' task. The pipeline starts with 'Automobile price data (Raw)', followed by 'Select Columns in Dataset', 'Clean Missing Data', 'Linear Regression', 'Train Model', 'Score Model', and 'Split Data'. A 'Python Language' module is also visible in the catalog. The right side of the screen displays pipeline parameters, settings, and details for the current step.

Azure Machine Learning Designer

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

Ele escolhe o melhor modelo!

MACHINE LEARNING AUTOMATIZADO

Select Open Dataset

Create dataset from Open Datasets

San Francisco Safety Data

Sample: Diabetes

US National Employment Hours and Earnings

NOAA Global Forecast System (GFS)

US Labor Force Statistics

US Consumer Price Index



THANKS