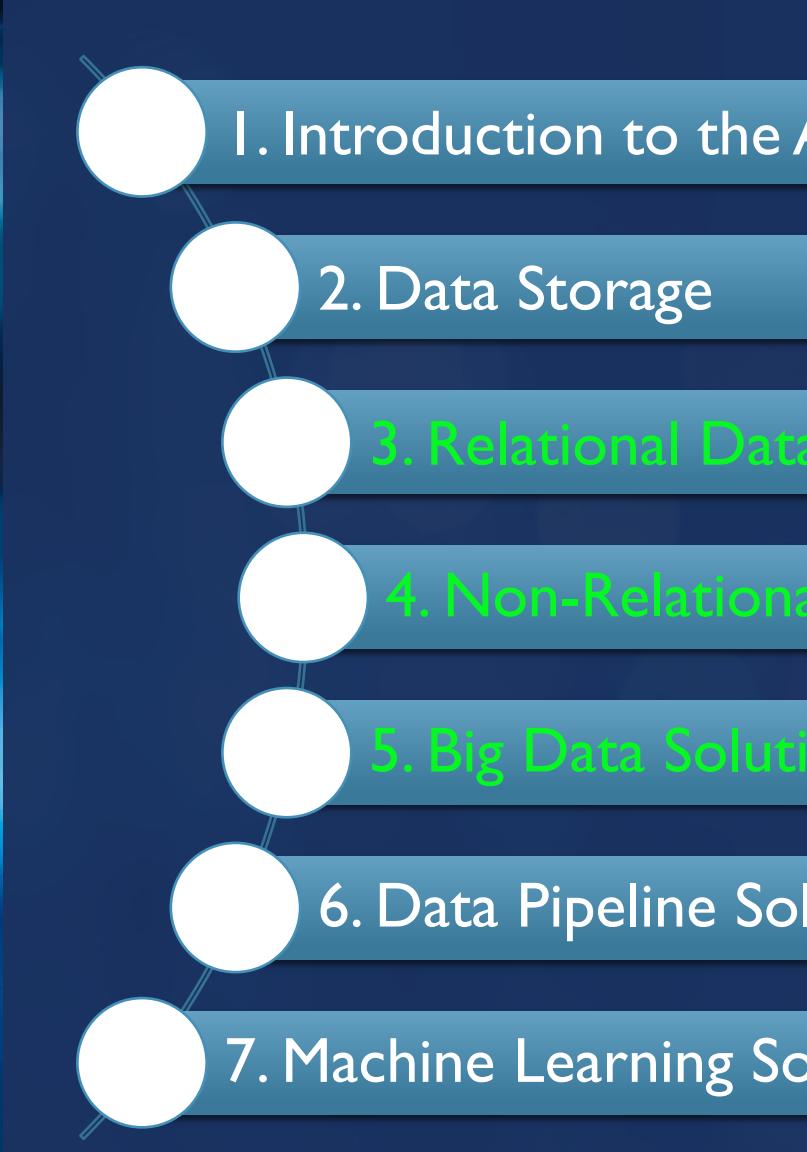




CLOUD – DATA SOLUTIONS - BIG DATA AND
MACHINE LEARNING

- 
- 
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

3. ARMAZENAMENTO DE DADOS RELACIONAIS NO AZURE

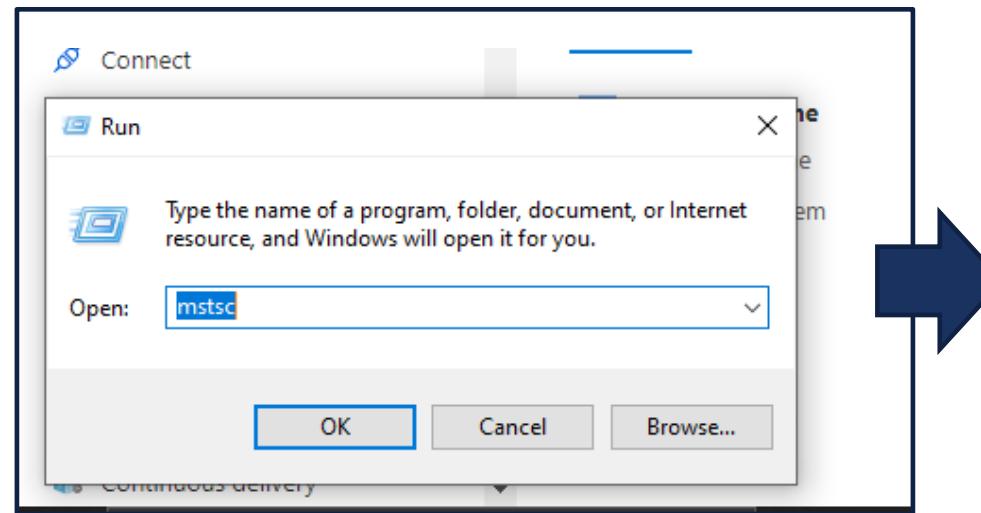
3

3.1. BANCOS DE DADOS RELACIONAIS EM IAAS

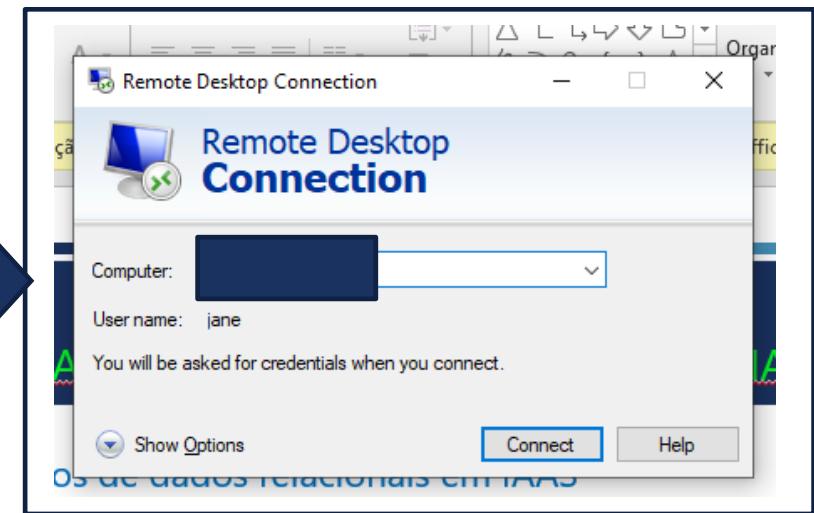
- Usando o serviço IaaS é possível fazer a instalação manual do Banco de Dados.

(Windows e Linux, nesse caso liberar a porta SSH(22) para ter livre acesso a VM)

Conectar à VM:

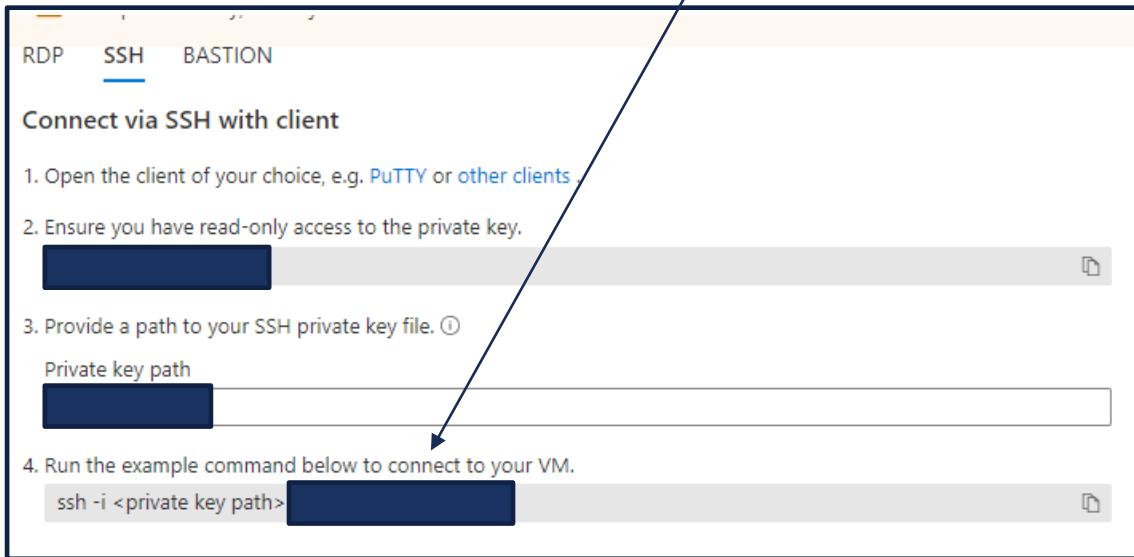


Colar o Public IP address da VM:



BANCOS DE DADOS RELACIONAIS EM IAAS

Conectar à VM Linux: com esse comando

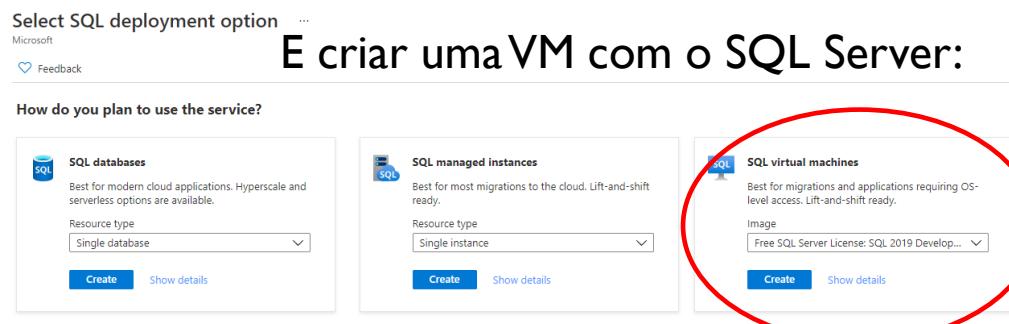


Instalar manualmente:

Instalar o SQL Server:

`sudo apt-get update`

`sudo apt-get install -y mssql-server`



E criar uma VM com o SQL Server:



The screenshot shows the Azure Marketplace search results for 'SQL Server'. It lists several options, with 'Azure SQL' circled in red.

Product	Provider	Rating	Plan	Description
SQL server (logical server)	Microsoft	4.3 (310 ratings)	\$0.005/hour	Virtual Machine
SQL Server 2016 SP1 Web w/ VulnerabilityAssessment	Cognosys Inc.	No rating	\$0.005/hour	Virtual Machine
Azure SQL	Microsoft	No rating	\$0.038/hour	Create and manage SQL Server resources from a single view

Ou pesquisar no Marketplace:

BANCOS DE DADOS RELACIONAIS EM IAAS

SQL Server do Azure Marketplace

The screenshot shows the Microsoft Azure Marketplace interface. At the top, there's a search bar and a navigation bar with 'Microsoft Azure' and 'Search resources, services, and docs (G+/)'. Below that, a breadcrumb trail shows 'Home > New >'. The main title is 'SQL Server 2019 on Windows Server 2019'. A red 'SQL' icon is on the left. The central area has a 'Create' button and a 'Start with a pre-set configuration' button. Below these are dropdown menus for 'Select a plan' (showing 'SQL Server 2019 Web on Windows S...') and 'Deploy with Resource Manager (change to Classic)'. There's also a 'Save for later' button.

Overview Plans

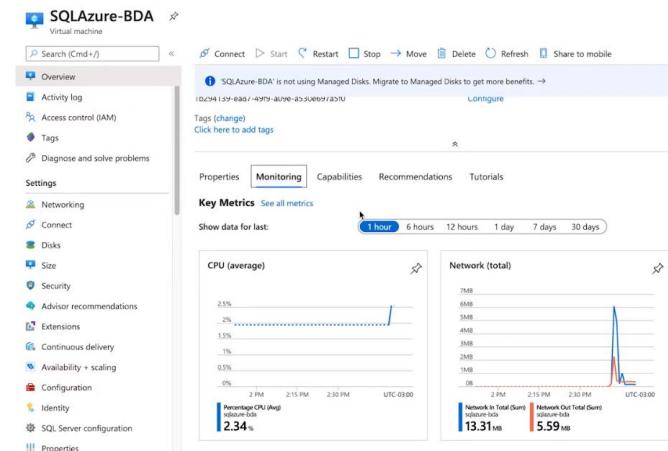
SQL Server 2019 Standard, Enterprise and Developer image on Windows Server 2019

Useful Links
Documentation
SQL Server 2019 information
Support forum
Pricing details

SQL Server do Azure Marketplace

- Quando aprovisiona uma VM do Marketplace com SQL Server: parte do processo instala o **SQL Server IaaS Agent Extension**:
 - Extensões são códigos executados na pós-implantação da VM;
 - Instalação de antivírus ou instalação de um recurso do Windows;
- Três recursos principais que podem reduzir a sobrecarga administrativa:
 - Backup automatizado do SQL Server;
 - Patches automatizados do SQL Server;
 - Integração com Azure Key Vault.
- Informações sobre a configuração e utilização do SQL Server.

SQL Server do Azure Marketplace



3.2. AZURE SQL DATABASE MANAGED INSTANCE

- Ofertas de Instâncias Gerenciada
 - Recursos da Instância Gerenciada
 - Criando uma SQL Managed Instance
 - Usando uma SQL Managed Instance

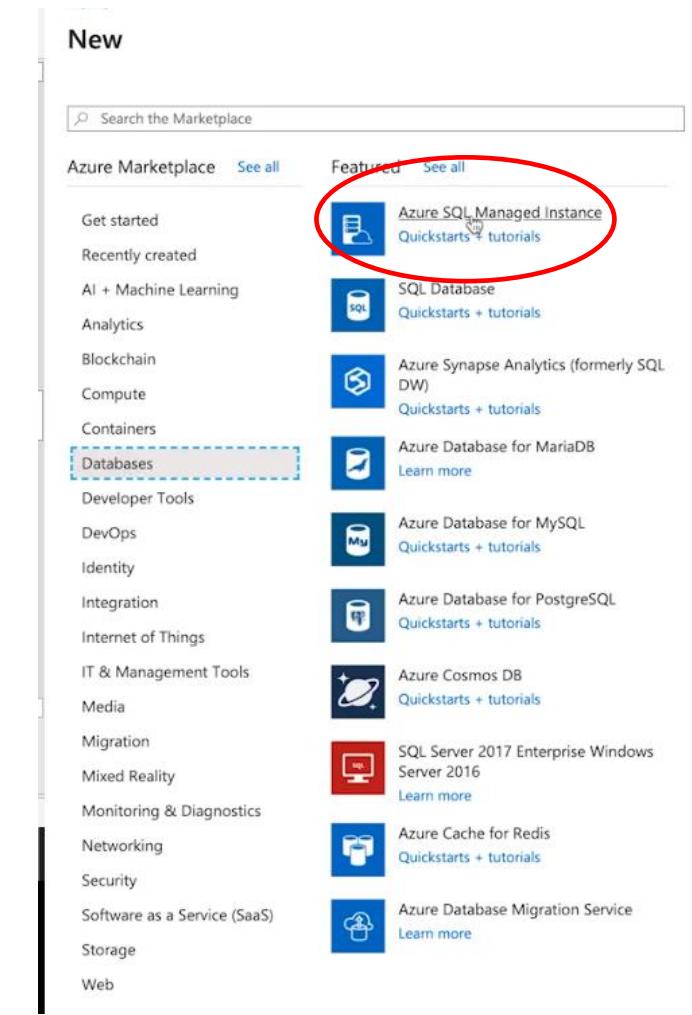
Ofertas de Instâncias Gerenciada

- General Purpose
- Business Critical
 - Maior desempenho e disponibilidade;
 - Suporta recursos In-Memory OLTP do SQL;
 - Leitura em réplicas secundárias;
 - Mais memória RAM por vCore;
 - Usa storage attachado diretamente (sem NAS)
 - Menor latência nas operações de I/O.
 - Uso das licenças de SQL já compradas para reduzir o custo.

Recursos da Instância Gerenciada

- 99.99% de disponibilidade;
- Atualizações automáticas de patches Windows / SQL;
- Backup gerenciado pelo Azure
 - Possível gerar um backup manual "copy only" para o Azure Storage.
- Automatic tuning:
 - Identificar queries onerosas;
 - Forçar o último plano execução bom;
 - Adicionar índices;
 - Remover índices.

Aqui podemos criar os Bancos de Dados!



3.3. AZURE SQL DATABASE

- Possui 2 modelos de implantação;

- **Single Database**

- Banco de dados isolado;
 - Totalmente gerenciado pelo Azure.

- **Elastic Pool**

- Coleção de single databases;
 - Com um conjunto compartilhado de recursos (CPU / RAM).

- Modelos de contratação

- **Baseado em vCore:** permite selecionar a quantidade de CPU, memória RAM e velocidade do storage.
 - **Baseado em DTU (Database Transaction Units):** combinação de recursos de computação, memória e I/O em três camadas de serviço, para suportar cargas de trabalho de banco de dados leves a pesadas.

- **Serverless:**

- Dimensiona automaticamente os recursos com base na demanda da carga de trabalho e cobra pela quantidade de recursos usados por segundo.
 - Pausa automaticamente os bancos de dados durante os períodos inativos → cobra o storage.

- Disponível em 3 camadas de serviço (service tiers);

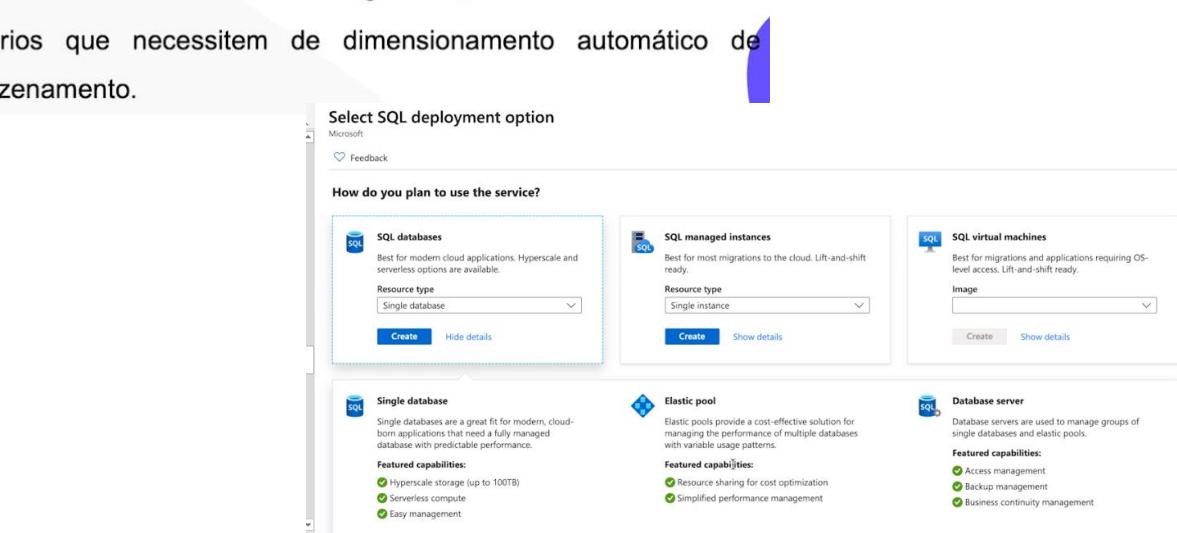
- **General Purpose / Standard:** para cargas de trabalho comuns.

- **Business Critical / Premium:**

- Para aplicações OLTP com alta taxa de transações;
 - Cenários que necessitam de baixa latência de I/O;
 - Oferece a maior resiliência a falhas usando várias réplicas isoladas.

- **Hyperscale:**

- Para bancos de dados OLTP muito grandes;
 - Cenários que necessitem de dimensionamento automático de armazenamento.



3.4. AZURE COSMOS DB

- Pré-requisitos para Criação
- Criando um Azure Cosmos DB (SQL)
- Usando um Azure Cosmos DB (SQL)

Pré-requisitos para Criação

■ Azure Cosmos DB Account

- Recurso que atua como uma entidade organizacional;
- Cada conta está associada a um dos vários modelos de dados aos quais o Azure Cosmos DB oferece suporte;
- Quantas contas precisar.



■ Request Unit (RU)

- Medida da taxa de transferência (throughput) por segundo;
- Reservar o número de RU/s que deseja que o Azure Cosmos DB provisione com antecedência, para que ele possa lidar com a carga estimada e você possa aumentar ou diminuir sua RU/s a qualquer momento.

Item size	Reads/second	Writes/second	Request units
1 KB	500	100	$(500 * 1) + (100 * 5) = 1,000 \text{ RU/s}$
1 KB	500	500	$(500 * 1) + (500 * 5) = 3,000 \text{ RU/s}$
4 KB	500	100	$(500 * 1.3) + (100 * 7) = 1,350 \text{ RU/s}$
4 KB	500	500	$(500 * 1.3) + (500 * 7) = 4,150 \text{ RU/s}$
64 KB	500	100	$(500 * 10) + (100 * 48) = 9,800 \text{ RU/s}$
64 KB	500	500	$(500 * 10) + (500 * 48) = 29,000 \text{ RU/s}$

Home >

New

Search the Marketplace

Azure Marketplace See all

Featured See all

Get started

Azure SQL Managed Instance
Quickstarts + tutorials

Recently created

SQL Database
Quickstarts + tutorials

AI + Machine Learning

Azure Synapse Analytics (formerly SQL DW)
Quickstarts + tutorials

Analytics

Blockchain

Compute

Containers

Databases

Azure Database for MariaDB
Learn more

Developer Tools

Azure Database for MySQL
Quickstarts + tutorials

DevOps

Identity

Integration

Internet of Things

IT & Management Tools

Azure Cosmos DB
Quickstarts + tutorials

Media

Migration

Mixed Reality

Monitoring & Diagnostics

Networking

SQL Server 2017 Enterprise Windows Server 2016
Learn more

Security

Azure Cache for Redis
Quickstarts + tutorials

Software as a Service (SaaS)

Storage

Azure Database Migration Service
Learn more

Web

3.5. BANCOS DE DADOS OPEN SOURCE NO AZURE

- Azure Database for MySQL
- Azure Database for MariaDB
- Azure DataBase for PostgreSQL

- Mecanismos para deploy:

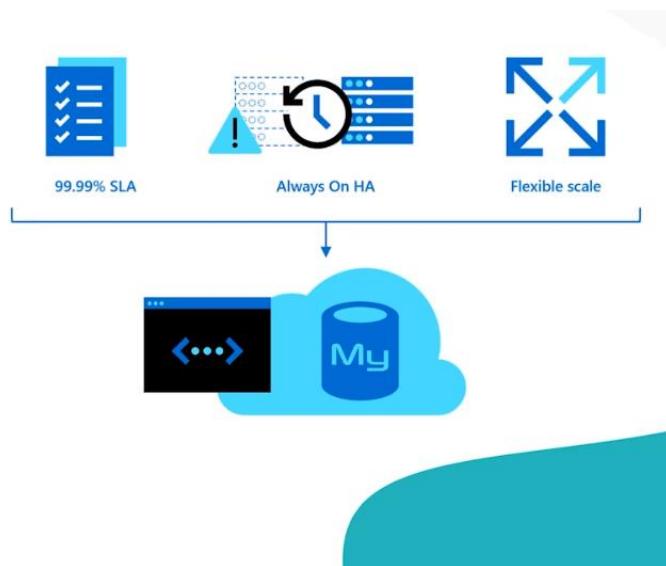
- Azure Portal;
- PowerShell;
- ARM templates;
- Azure CLI.

- Azure Database for MySQL / MariaDB

- Single Database / Replicação.

- Azure Database for PostgreSQL

- Single Database / Hyperscale.



Marketplace

My Saved List
Recently created
Service Providers

Categories

Showing All Results

Get Started

Azure Database for MySQL
Microsoft
Managed MySQL database service for app developers.

MySQL with Replication
Bitnami
Production ready, up-to-date, and secure.

PHP MYSQL and IIS on Windows Server 2016
Cognos Inc.
1 cent/hr in Windows Server 2016

WordPress on Server 2011
PhpMyAdmin + MySQL
Cloud Infrastructure Services
Quickly deploy WordPress on Windows 2011 with built-in MySQL and phpMyAdmin

Cloudera on Azure
Cloudera
Cloudera Enterprise Data Hub

Azure CycleCloud
microsoft-azure-cyclecloud -
Azure CycleCloud is the easiest way to create and operate HPC clusters of any scale in Azure.

My Saved List
Recently created
Service Providers

Categories

Showing All Results

Get Started

Azure Database for MySQL
Microsoft
Managed MySQL database service for app developers.

MySQL with Replication
Bitnami
Production ready, up-to-date, and secure.

PHP MYSQL and IIS on Windows Server 2016
Cognos Inc.
1 cent/hr in Windows Server 2016

WordPress on Server 2011
PhpMyAdmin + MySQL
Cloud Infrastructure Services
Quickly deploy WordPress on Windows 2011 with built-in MySQL and phpMyAdmin

Cloudera on Azure
Cloudera
Cloudera Enterprise Data Hub

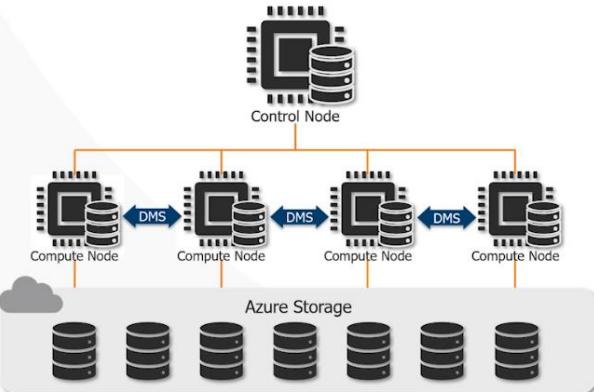
Azure CycleCloud
microsoft-azure-cyclecloud -
Azure CycleCloud is the easiest way to create and operate HPC clusters of any scale in Azure.

3.6. AZURE SYNAPSE ANALYTICS

- SQL Pools
- Workload Groups
- Demo Azure Synapse Analytics

SQL Pools

- Agrupamento de recursos (CPU, RAM e I/O);
- Tamanho é determinado por DWU (Data Warehousing Units).



Workload Groups

- Definir os recursos para isolar e reservar recursos para uso:
 - Reserva de recursos para um grupo de solicitações;
 - Limite da quantidade de recursos que um grupo de solicitações pode consumir;
 - Recursos compartilhados acessados com base no nível de importância do workload;
 - Definição do valor do tempo limite da consulta.

```
CREATE WORKLOAD GROUP group_name  
WITH.....  
(  
    MIN_PERCENTAGE_RESOURCE = value  
    , CAP_PERCENTAGE_RESOURCE = value  
    , REQUEST_MIN_RESOURCE_GRANT_PERCENT = value  
    [ [ . ] REQUEST_MAX_RESOURCE_GRANT_PERCENT = value ]  
    [ [ . ] IMPORTANCE = {LOW | BELOW_NORMAL | NORMAL | ABOVE_NORMAL | HIGH} ] [ [ . ]  
    ] QUERY_EXECUTION_TIMEOUT_SEC = value ]
```

```
CREATE WORKLOAD CLASSIFIER classifier_name  
WITH  
(  
    WORKLOAD_GROUP = 'name'  
    , MEMBERNAME = 'security_account'  
    .....  
)
```

New

A screenshot of the Azure Marketplace interface. At the top, there's a search bar labeled 'Search the Marketplace'. Below it, a grid of service cards. A blue dashed box highlights the 'Databases' category. An orange circle highlights the 'Azure Synapse Analytics (formerly SQL DW)' card, which includes a 'Quickstarts + tutorials' link. Other visible categories include 'Compute', 'Containers', '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'.

4

4. ARMAZENAMENTO DE DADOS NÃO RELACIONAIS NO AZURE

- Banco de Dados Não Relacionais com IaaS
- Banco de Dados Não Relacionais com PaaS

→ Instalar um Banco de Dados manualmente na VM!

Ou criar uma VM com o serviço através do Marketplace:

Azure Database Migration Service

Microsoft

Azure Database Migration Service Save for later

Create

Overview

Azure Database Migration Service (DMS) is designed to streamline the process of migrating on-premises databases to Azure. Get started with step-by-step guidance.

Common scenarios:

- SQL Server → Azure SQL Database
- SQL Server → Azure SQL Database Managed Instance
- MongoDB → Azure Cosmos DB
- MySQL → Azure Database for MySQL
- PostgreSQL → Azure Database for PostgreSQL
- DB2 → Azure SQL Database
- Oracle → Azure Database for PostgreSQL (requires ora2pg)

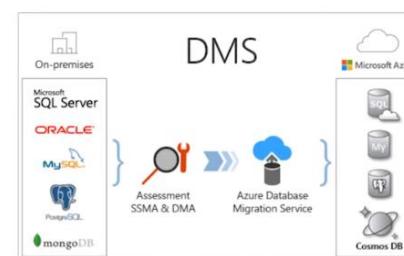
Don't see your migration scenario here? Find it in the [Azure Database Migration Guide](#).

After you've completed the step by step guidance for your migration scenario, proceed with creating a migration service by selecting **Create**.

Additional resources:

Use these tools to assess your database(s) for feature parity and potential compatibility issues.

- SQL Server on-premises database(s): [Data Migration Assistant \(DMA\)](#)
- Migrating from Oracle: [SQL Server Migration Assistant \(SSMA\)](#)



Marketplace

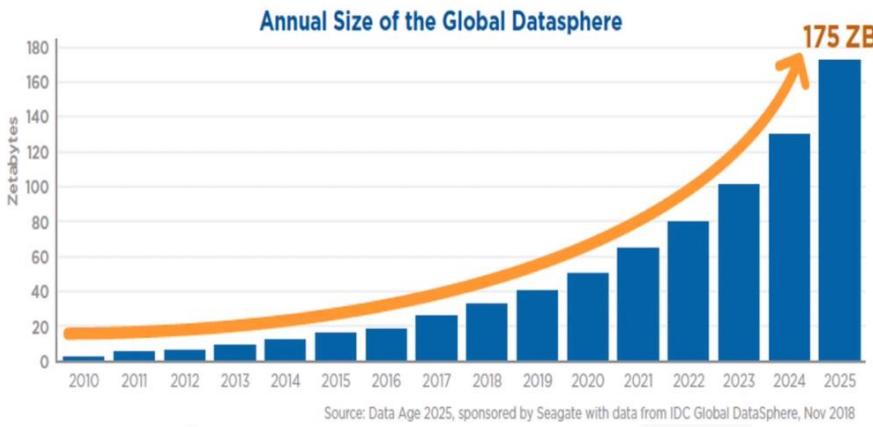
Showing All Results

Pricing: All Operating System: All Publisher: All

Image	Name	Description
	MongoDB with Replication	Binami Production ready, up-to-date, and secure.
	MongoDB Atlas on Azure	MongoDB, Inc. The Best Way to Run MongoDB on Microsoft Azure
	MongoDB Certified by Bitnami	Bitnami Up-to-date, secure, and ready to run.
	MongoDB Sharded Container Image	Bitnami Up-to-date, customizable, and secure.
	MongoDB Exporter Container Image	Bitnami Up-to-date, customizable, and secure.
	MongoDB Container Image	Bitnami Up-to-date, customizable, and secure.
	MongoDB Community Server on CentOS Websoft9	Websoft9 Pre-configured, customizable, secure, one-click to deploy MongoDB Community Server on
	MongoDB Helm Chart	Bitnami Up-to-date, secure, and ready to deploy on Kubernetes.
	MongoDB Community Server on Ubuntu	Websoft9 Pre-configured, customizable, secure, one-click to deploy MongoDB Community Server on
	MongoDB 2.6.12 for LINUX CentOS 7.7	Tidal Media Inc MongoDB 2.6.12 for LINUX CentOS 7.7 Certified by Tidal Media Inc.
	MongoDB Server	Jetware A one-click install solution for MongoDB.
	Percona Server for MongoDB	Jetware A one-click install solution for Percona Server for MongoDB.
	MongoDB on Centos 7.5	Apps4rent LLC An open-source DBMS for storing high volumes of data.
	MongoDB Sharded Helm Chart	Bitnami Up-to-date, secure, and ready to deploy on Kubernetes.
	MongoDB Helm Chart	Bitnami Up-to-date, secure, and ready to deploy on Kubernetes.
	Official MEAN Machine - Mean.io app out-of-the-box	Limnatech Technologies Ltd. MEAN Machine is the fastest way to have a functional MEAN.js app. Boot it up, and enjoy your app!
	CR8 Continuous SQL to Big Data streaming	DBS-H Ltd CR8 streams SQL data changes in real time to Big Data or Cloud Data Lakes
	Let's Chat Certified by Bitnami	Bitnami Up-to-date, secure, and ready to run.
	DataSurprise Data & Database Security	DataSurprise Inc DataSurprise Data & Database Security, Data Audit, Data Masking and Discovery, SOX & GDPR
	Azure Database Migration Service	Microsoft Azure Database Migration Service
	Hystax Live Cloud Migration to Azure	Hystax Inc Hystax Live Cloud Migration to Azure replicates customer workloads from various platforms to Azure

5. SOLUÇÕES DE BIG DATA

5.1. INTRODUÇÃO AO BIG DATA



O Que É Big Data

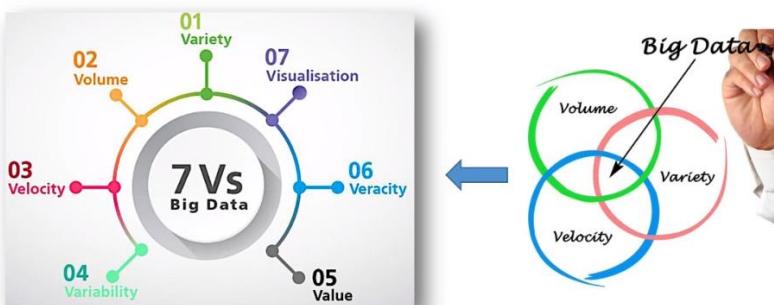
- Campo que trata de maneiras de analisar, extrair sistematicamente informações ou, de outra forma, lidar com conjuntos de dados que são muito grandes ou complexos para serem tratados por softwares de aplicativos de processamento de dados tradicionais.



O Que É Big Data

- Termo adotado pelo mercado para descrever problemas no gerenciamento e processamento de informações extremas, as quais excedem a capacidade das tecnologias de informações tradicionais ao longo de uma ou várias dimensões.
- Big Data está focado principalmente em questões de volume de conjunto de dados extremamente grandes gerados a partir de práticas tecnológicas, tais como mídia social, tecnologias operacionais, acessos à Internet e fontes de informações distribuídas. Big Data é essencialmente uma prática que apresenta novas oportunidades de negócios. (Gartner Group)

Fundamentos de Big Data



O Que É Big Data

- A intensa utilização de redes sociais online, de dispositivos móveis para conexão à Internet, transações e conteúdos digitais, e também o crescente uso de computação em nuvem tem gerado quantidades incalculáveis de dados. O termo Big Data refere-se à este conjunto de dados cujo crescimento é exponencial e cuja dimensão está além da habilidade das ferramentas típicas de capturar, gerenciar e analisar dados. (McKinsey Global Institute)

Conclusão

- Big Data é mais que um produto de software ou hardware;
- É um conjunto de tecnologias, processos e práticas que permitem às empresas analisarem dados que antes não tinham acesso e tomar decisões, ou mesmo gerenciar atividades de forma muito mais eficiente.



5.2. INTRODUÇÃO AO HDINSIGHT

Overview do HDInsight

- Recurso para processamento e análise de Big Data;
- Criação rápida de clusters de Big Data sob demanda, com escalabilidade horizontal e/ou vertical;
- Solução em nuvem de baixo custo para projetos de Big Data;
- Possui os principais softwares livres da Apache para projetos de Big Data:

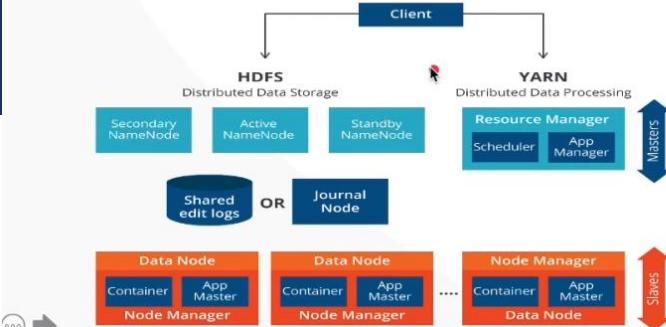


APACHE ML SERVICES

- Cluster para soluções de aprendizagem de máquina;
- Fornece aos cientistas de dados, estatísticos e programadores de R / Python o acesso sob demanda a métodos escalonáveis e distribuídos de análise no HDInsigh;
- Possui um conjunto de modelos e algoritmos de machine learning que podem ser adaptados.

APACHE HADOOP

- Estrutura que usa HDFS, gerenciamento de recursos YARN e um modelo de programação MapReduce simples para processar e analisar, paralelamente, dados em lote.



APACHE HBASE

- Banco de dados NOSQL baseado em Hadoop que fornece acesso aleatório e forte coerência para grandes quantidades de dados sem esquema;
- Construído com base no Google BigTable;
- Os dados são armazenados nas linhas e colunas de uma tabela, e os dados em uma linha são agrupados por família de colunas.

APACHE HIVE LLAP

- O Hive oferece uma interface semelhante à SQL para consulta de dados em diferentes bancos de dados e sistemas de arquivos integrados ao Hadoop;
- Comandos tradicionais de SQL são implementados na API Java (HiveQL) para serem executados em dados distribuídos.
- O Hive LLAP é um recurso para cache de dados em memória para consultas do Hive.

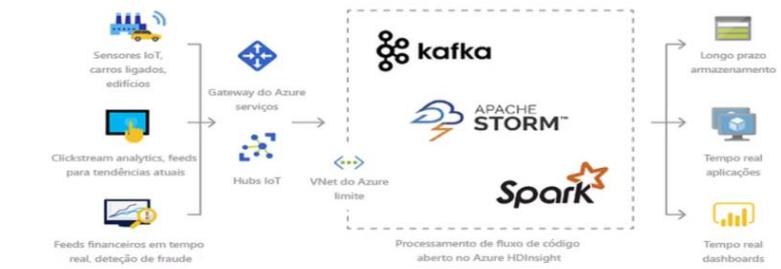
APACHE SPARK

- Estrutura de processamento paralelo, de software livre, que dá suporte ao processamento em memória para melhorar o desempenho dos aplicativos de análise de Big Data.



APACHE KAFKA

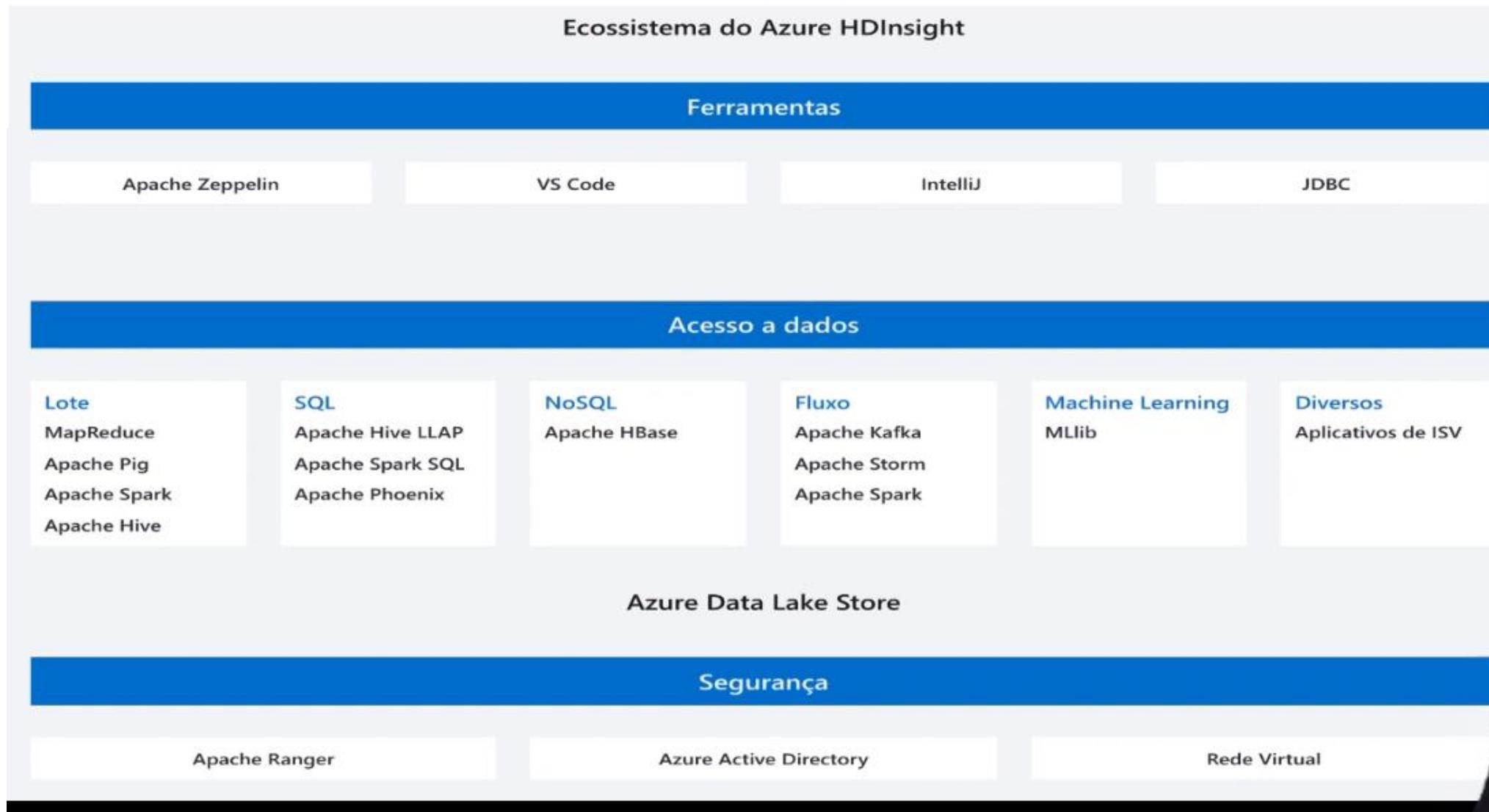
- Plataforma de código fonte aberto usada para criar aplicativos e pipelines de streaming de dados;
- Também fornece funcionalidade de fila de mensagens, o que permite publicar e consumir pipelines de dados.



APACHE STORM

- Sistema de computação distribuído e em tempo real para processar rapidamente grandes fluxos de dados;
- Processa topologias ao invés de trabalhos MapReduce.

Ecossistema do Azure HDInsight



5.3. APROVISIONANDO UM AMBIENTE DO HDINSIGHT

Home >

New

Search the Marketplace

Azure Marketplace See all

Featured See all

Get started

Recently created

AI + Machine Learning

Analytics

Blockchain

Compute

Containers

Databases

Developer Tools

DevOps

Azure Data Explorer Learn more

Azure HDInsight Quickstarts + tutorials

Data Lake Analytics Quickstarts + tutorials

Stream Analytics job Quickstarts + tutorials

Analysis Services

Create HDInsight cluster

New to HDInsight? Get started with our [training resources](#). Create a managed HDInsight cluster. Select from Spark, Kafka, Hadoop, Storm, and more. [Learn more](#)

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * **MSDN Platforms**
Resource group * **rgbtccmod32020**
[Create new](#)

Cluster details

Name your cluster, pick a region, and choose a cluster type and version. [Learn more](#)

Cluster name * **hdinsightbtccl2020**

Region * **East US**

Cluster type * **Hadoop**
[Change](#)

Version * **Hadoop 3.1.0 (HDInsight 4.0)**

Cluster credentials

Enter new credentials that will be used to administer or access the cluster.

Cluster login username * **admin**

Cluster login password * *********

Confirm cluster login password * *********

Secure Shell (SSH) username * **sshuser**

Use cluster login password for SSH

[Review + create](#) [Next: Storage >](#)

Verificar as versões!

O Ambari é a ferramenta gráfica para administrar o cluster!

hdinsightbtccl2020 HDInsight cluster

Search (Cmd+ /) Overview Move Delete Refresh

Resource group (change) : rgbtccmod32020 Status : Running

Location : East US Subscription (change) : MSDN Platforms

Subscription ID : 1b294139-ea87-49f9-a09e-a530e697a5f0 Tags (change) : Click here to add tags

Activity log Access control (IAM) Tags Diagnose and solve problems Quick start Tools

Cluster size Quota limits SSH + Cluster login Data Lake Storage Gen1 Storage accounts Applications Script actions External metastores HDInsight partner Properties Locks

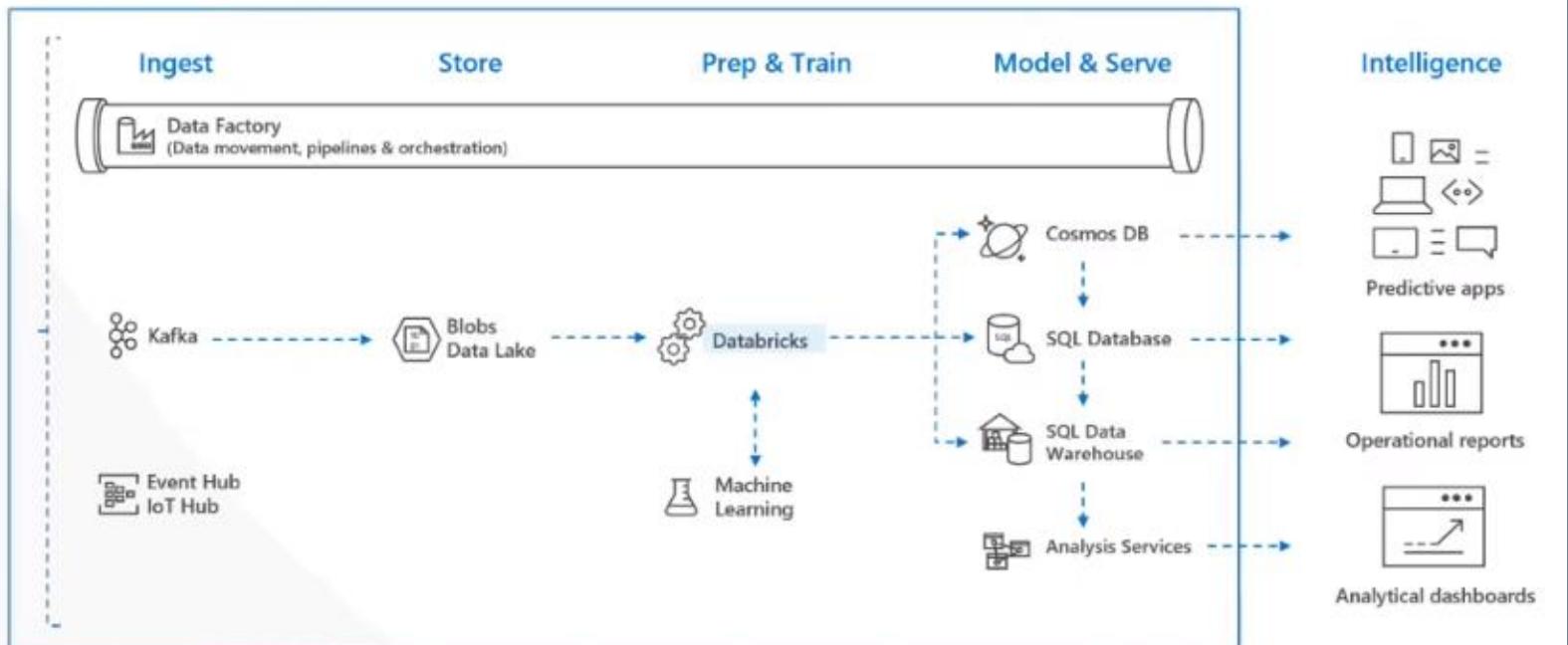
Cluster dashboards Cluster management interfaces Ambari home Ambari views

Cluster size

6 nodes

Type	↑↓	Size	↑↓	Cores	↑↓
Head		E2 V3		4	
Worker		E2 V3		8	

5.4. INTRODUÇÃO AO AZURE DATABRICKS



Overview do Azure Databricks

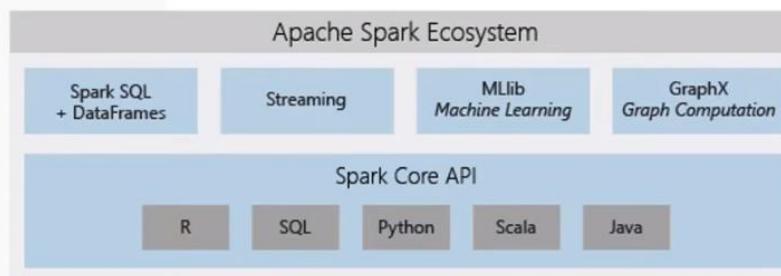
- Plataforma de análise baseada no Apache Spark e otimizada para a plataforma de serviços de nuvem do Microsoft Azure;
- Configuração rápida e automatizada do cluster Spark;
- Fluxos de trabalho simplificados;
- Workspace interativo que permite a colaboração entre os cientistas de dados, os engenheiros de dados e os analistas de dados / negócios.

INTRODUÇÃO AO AZURE DATABRICKS

Componentes do Azure Databricks

SPARK SQL E DATAFRAMES

- Módulo Spark para trabalhar usando dados estruturados;
- DataFrame é uma coleção distribuída de dados organizados em colunas nomeadas (conceitualmente equivalente a uma tabela em um banco de dados relacional ou uma estrutura de dados em R/Python).



STREAMING

- Módulo para processamento de dados em tempo real e análise para aplicativos analíticos e interativos;
- Integra-se com HDFS, Flume e Kafka.

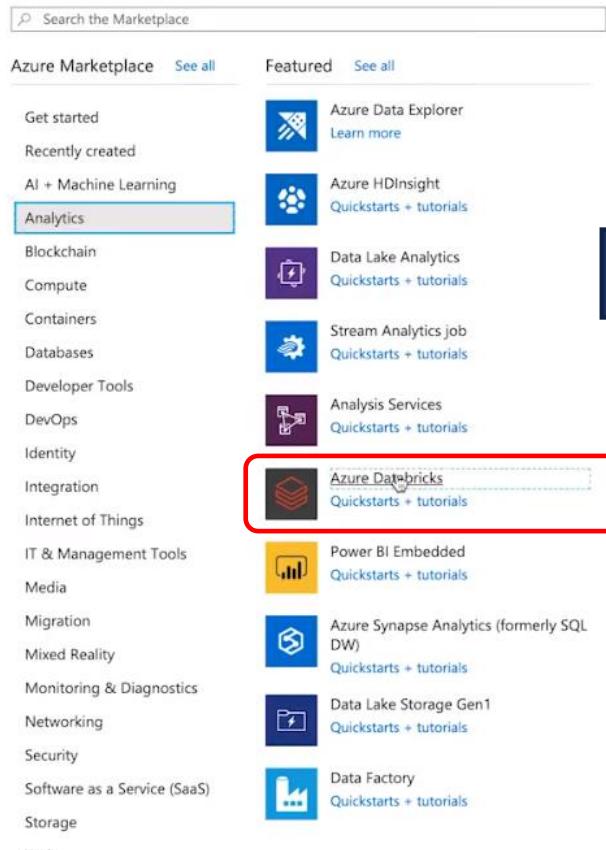
MLIB: biblioteca Machine Learning que consiste em algoritmos e utilitários de aprendizado comuns, incluindo classificação, regressão, clustering, etc.

GRAPHX: módulo para tarefas de análise de gráficos e operações em grafos.

SPARK CORE API: suporte para R, SQL, Python, Scala e Java.

5.5. DEMONSTRAÇÃO DO AZURE DATABRICKS

New



Azure Marketplace See all

Featured See all

- Get started
- Recently created
- AI + Machine Learning
- Analytics**
- 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
- Web

Azure Data Explorer Learn more

Azure HDInsight Quickstarts + tutorials

Data Lake Analytics Quickstarts + tutorials

Stream Analytics job Quickstarts + tutorials

Analysis Services Quickstarts + tutorials

Azure Databricks Quickstarts + tutorials

Power BI Embedded Quickstarts + tutorials

Azure Synapse Analytics (formerly SQL DW) Quickstarts + tutorials

Data Lake Storage Gen1 Quickstarts + tutorials

Data Factory Quickstarts + tutorials



Status : Active

Resource group : rgbtclic

Location : Central US

Subscription : MSDN Platforms

Subscription ID : 1b294139-ea87-49f9-a09e-a530e697a5f0

Tags (change) : Click here to add tags

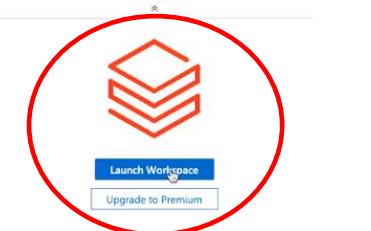
Managed Resource Group : rgbtclic

URL : https://adl-3564829742446865.azuredatabricks.net

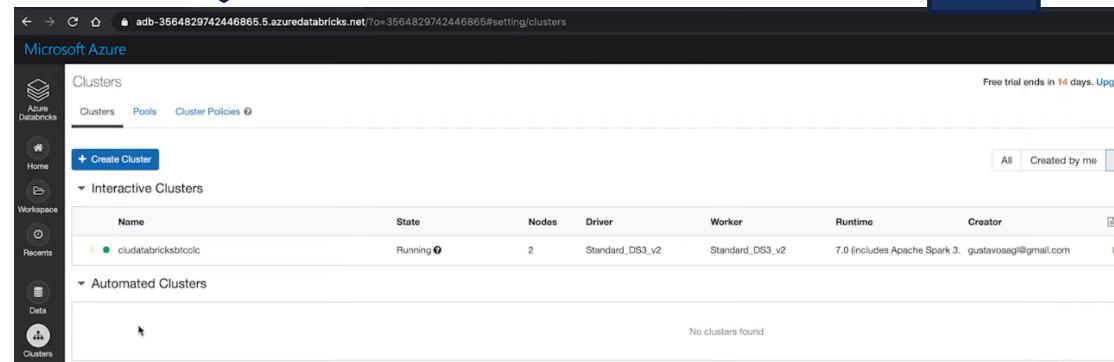
Pricing Tier : Standard

Launch Workspace

Upgrade to Premium



Primeiro criar um Cluster:

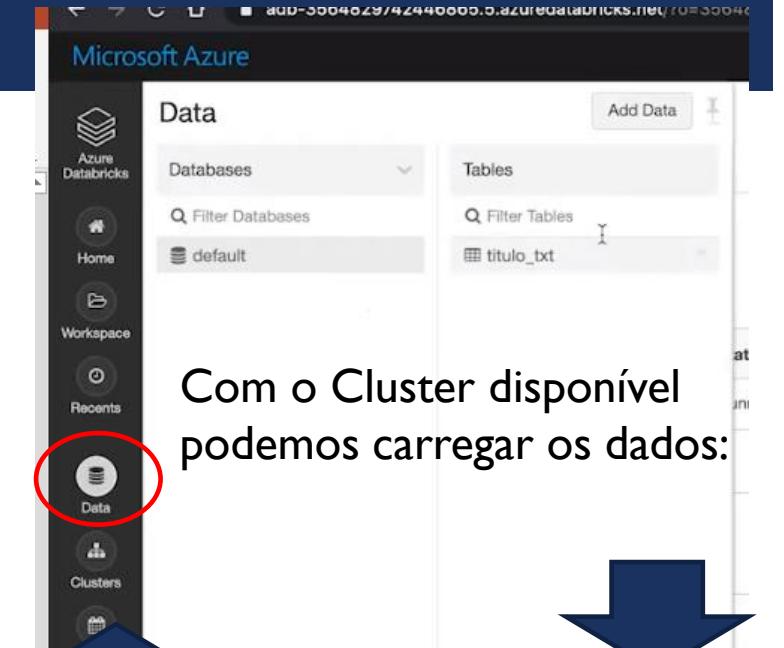


Clusters

+ Create Cluster

Name	State	Nodes	Driver	Worker	Runtime	Creator
cludatabricksbtclc	Running	2	Standard_DS3_v2	Standard_DS3_v2	7.0 (includes Apache Spark 3.0)	gustavosagi@gmail.com

No clusters found



Microsoft Azure

Data

Databases

Tables

Recents

Data

Clusters

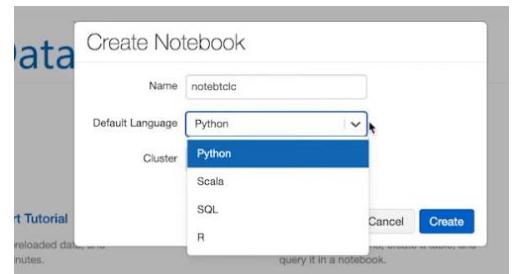
Launch Workspace

Upgrade to Premium

Com o Cluster disponível podemos carregar os dados:



Criando um Notebook:



Create Notebook

Name : notebtclc

Default Language : Python

Cluster : Python

Scala

SQL

R

Create



CONTINUA...