



Serviços de Armazenamento, Banco de Dados e Analytics

Capítulo 1. Introdução à Plataforma de Dados do Azure

PROF. GUSTAVO AGUILAR

Serviços de Armazenamento, Banco de Dados e Analytics

CAPÍTULO 1. AULA 1.0. APRESENTAÇÃO DO PROFESSOR E DA DISCIPLINA

PROF. GUSTAVO AGUILAR

Nesta Aula



- Apresentação do Professor
- Apresentação da Disciplina

Gustavo Aguilar



Formação



- Bacharelado em Ciência da Computação (PUC-Minas)
- Pós-Graduação em Administração de Banco de Dados (Newton)
- Especialização em Docência do Ensino Superior (Newton)
- MBA em Ciência de Dados (IGTI)
- Instrutor Certificado Microsoft (MCT)



Atuação



- Administração de Bancos de Dados
- Ambientes de missão crítica em diversas plataformas
- Modelagem, Arquitetura e Engenharia de Dados
- Persistência e Pesquisa de Dados
- Sistemas Distribuídos e Cloud Computing
- Metodologias Ágeis e DevOps
- Coordenador e Professor de MBA e Bootcamp

Apresentação da Disciplina



- Módulo de apresentação das soluções disponíveis no Azure para:
 - Armazenamento de dados não estruturados;
 - Armazenamento de dados relacionais / não relacionais em bancos de dados;
 - Distribuição de dados;
 - Disaster recovery na camada de dados;
 - Big Data;
 - Análise de dados;
 - Aprendizado de Máquina (Machine Learning).

Apresentação da Disciplina



Espera-se que o aluno consiga, ao final deste módulo:

- ✓ Compreender os papéis dos tipos de profissionais que trabalham com plataforma de dados;
- ✓ Criar soluções de armazenamento de dados usando IaaS ou PaaS no Azure;
- ✓ Utilizar serviços de banco de dados relacional ou não relacional no Azure;
- ✓ Planejar soluções com distribuição de dados usando o CosmosDB;
- ✓ Projetar soluções de preparação de dados usando o Azure Databricks;

Apresentação da Disciplina



Espera-se que o aluno consiga, ao final deste módulo:

- ✓ Planejar soluções de extração, ingestão e transformação de dados com Data Factory;
- ✓ Projetar soluções para o armazenamento e processamento de dados em larga escala usando o Azure HDInsight;
- ✓ Projetar soluções para análise de dados com o Azure Synapse Analytics;
- ✓ Identificar os serviços de Inteligência Artificial (AI) e Aprendizado de Máquina (Machine Learning) do Azure.

Apresentação da Disciplina



- **Material Didático**

- ✓ Apostila
- ✓ Aulas gravadas
- ✓ Slides das aulas gravadas
- ✓ Aulas interativas ao vivo

- **Atividades dos Alunos**

- ✓ Fórum de dúvidas e fórum de debates
- ✓ Trabalho prático
- ✓ Desafio final
- ✓ Reposição da aula interativa *
- ✓ Feedback pontuado

Apresentação da Disciplina

- **Distribuição de pontos**

- ✓ **10 pontos** de participação na 1ª Aula Interativa
 - ✓ **25 pontos** do Trabalho Prático do módulo 2
 - ✓ **10 pontos** de participação na 2ª Aula Interativa
 - ✓ **40 pontos** do Desafio do módulo 2
 - ✓ **10 pontos** de participação no Fórum de Debates do módulo 2
 - ✓ **5 pontos** do Feedback do aluno em relação ao módulo 2
-
- **TOTAL: 100 pontos**

Próxima Aula



- Tipos de Profissionais de Dados.

Introdução à Plataforma de Dados do Azure

CAPÍTULO 1. AULA 1.1. TIPOS DE PROFISSIONAIS DE DADOS

PROF. GUSTAVO AGUILAR

Nesta Aula



- ❑ Overview do Processo de Soluções de Dados
- ❑ Perfis de Profissionais de Dados

Overview do Processo de Soluções de Dados



Projetar a
Solução



Implantar a
Infraestrutura
para a Solução



Implantar o
Pipeline de
Dados



Análise de Dados
Ciência de Dados
Aprendizado de
Máquina
Etc.

Perfis de Profissionais de Dados



- Arquiteto de Soluções
- Arquiteto de Dados
- Administrador de Banco de Dados
- Engenheiro de Dados
- Analista de Dados
- Cientista de Dados
- Engenheiro de Inteligência Artificial

Perfis de Profissionais de Dados



▪ Arquiteto de Soluções (Solutions Architect)

- Visão holística de toda a solução;
- Integração das soluções;
- Recursos e capacidade para a solução;
- Estratégias de backup, monitoramento;
- Disponibilidade e escalabilidade da solução;
- Plano de Continuidade de Negócio (PCN);
- Custo da solução;
- Etc.



Perfis de Profissionais de Dados



▪ Arquiteto de Dados (Data Architect)

- Estruturas de armazenamento dos dados;
- Modelos de dados;
- Administração de dados corporativos;
- Backup de dados de negócio.



Perfis de Profissionais de Dados



▪ Administrador de Banco de Dados (DBA)

- Instalação / aprovisionamento
 - SGBDs
 - Plataformas de armazenamento de dados
 - Bancos de dados
 - Repositórios / quotas
- Aspectos operacionais;
- Tuning e troubleshooting;
- Disponibilidade dos SGBDs / plataformas;
- Segurança de acesso.



Perfis de Profissionais de Dados



▪ Engenheiro de Dados (Data Engineer)

- Projeto do Pipeline (Fluxo) de Dados:
 - Extração de dados
 - Transformação de dados
 - Ingestão (carga) de dados
- Implementação e gerenciamento do fluxo de dados estruturados e não estruturados de diversas origens (fontes de dados).



Perfis de Profissionais de Dados



▪ Analista de Dados (Data Analyst)

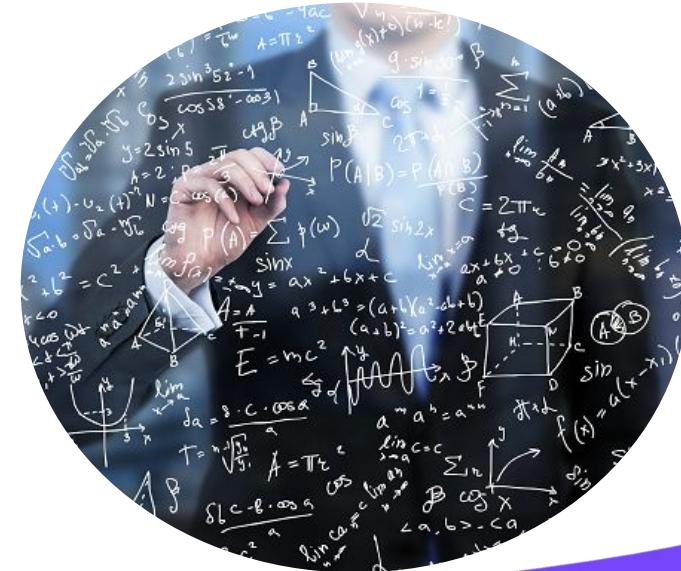
- Projetar e construir modelos de dados analíticos;
- **Análise Descritivas:** ajudar a responder questões acerca do que aconteceu;
- Transformar dados em informações analíticas com valor comercial;
- Planejamento e gerenciamento de dashboards.



Perfis de Profissionais de Dados

▪ Cientista de Dados (Data Scientist)

- Análise avançada para ajudar a gerar valor a partir dos dados;
- **Análises Preditivas:** o que vai acontecer;
- **Análises Prescritivas:** o que precisa ser feito para atingir determinado objetivo;
- Suporte a decisões orientadas a dados (Data-Driven Decision).

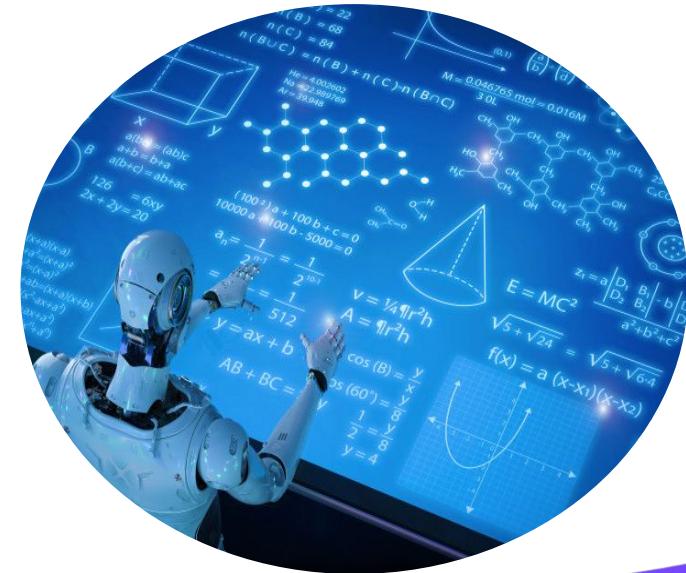


Perfis de Profissionais de Dados



▪ Engenheiro de Inteligência Artificial

- Arquitetar e implementar soluções de IA
 - Serviços Cognitivos
 - Machine Learning
 - Mineração de conhecimento
- Suporte a soluções de processamento de linguagem natural, reconhecimento de voz e imagem;
- Bots e agentes virtuais.



Overview do Processo de Soluções de Dados



Arquiteto de Soluções
Arquiteto de Dados



Adm. Banco de Dados



Engenheiro de Dados



Analista de Dados
Cientista de Dados
Engenheiro de IA

Próxima Aula



- Tipos de Dados.

Introdução à Plataforma de Dados do Azure

CAPÍTULO 1. AULA 1.2. TIPOS DE DADOS

PROF. GUSTAVO AGUILAR

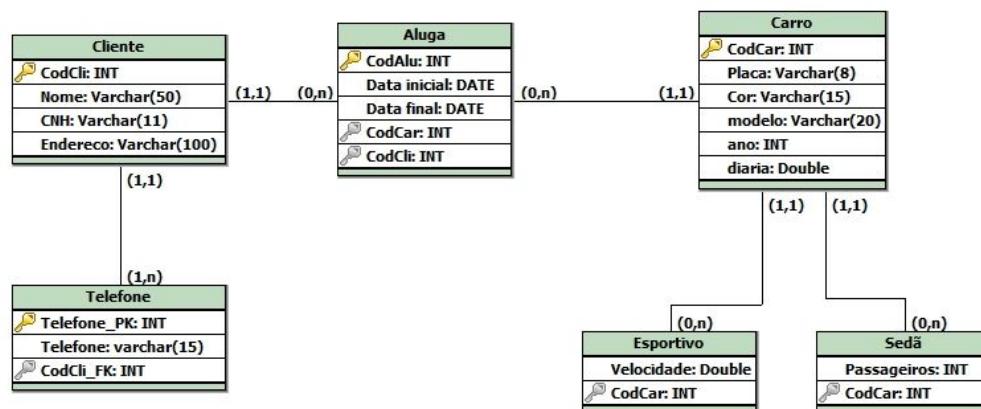
Nesta Aula



- Dados Estruturados
- Dados Semiestruturados
- Dados Não Estruturados
- Dados na Era da Informação

Dados Estruturados

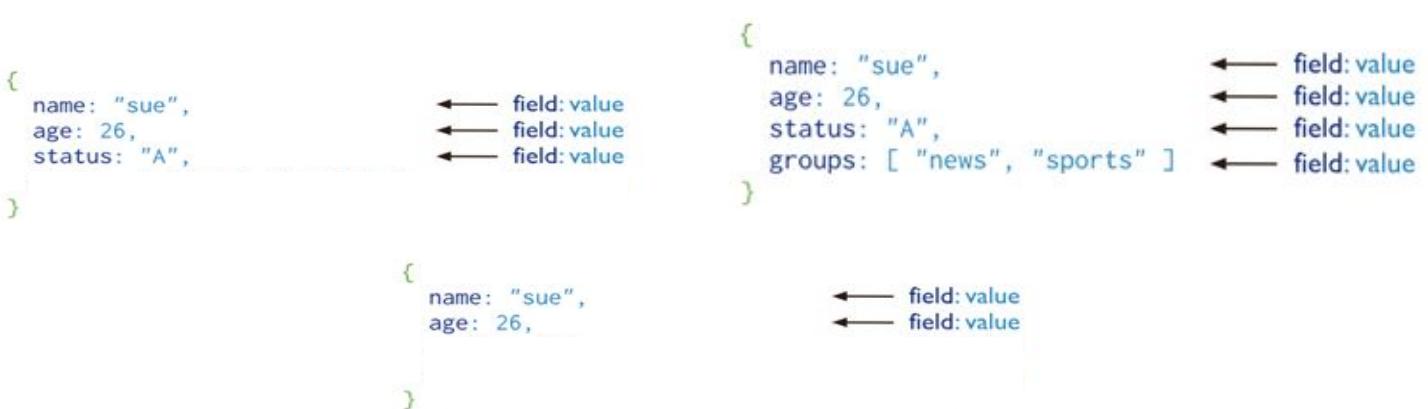
- Organizados e representados com uma estrutura rígida;
 - Aderem a um esquema (schema físico dos dados).
- Definição prévia de um modelo de dados;
 - Todos os dados têm os mesmos campos ou propriedades.
- Em geral, armazenados no formato tabular (em tabelas);
 - Linha + Coluna(s) ☐ tupla ☐ dados relacionais.



Dados Semiestruturados



- Não contém toda a rigidez requerida na definição dos tipos de dados estruturados;
- Procuram manter certa uniformidade no armazenamento das informações;
 - Mantém tags e marcações internas que identificam elementos de dados separados.
- Comumente chamados de dados não relacionais ou NOSQL.



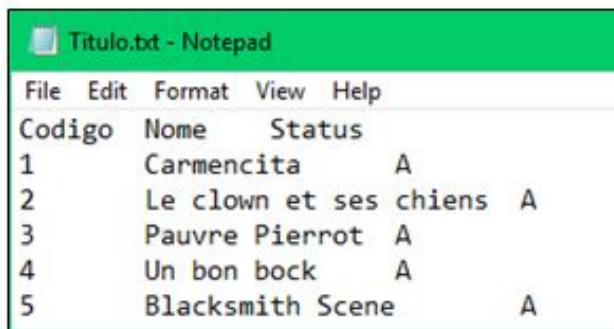
Dados Não Estruturados

- Não são estruturados por meio de modelos ou esquemas de dados predefinidos:
 - Significa que não há restrições quanto aos tipos de dados que podem conter.
 - **Dados binários (Blob)** pode conter:
 - Documento PDF
 - Imagem JPEG
 - Áudio MP3
 - Vídeo MPEG
 - Etc.



Dados na Era da Informação

- “Início da Era da Internet”:
 - Volume de dados não muito significativo;
 - Grande parte dos dados de tipo textual;
 - Quase a totalidade dos sistemas trabalhando com **dados estruturados**.



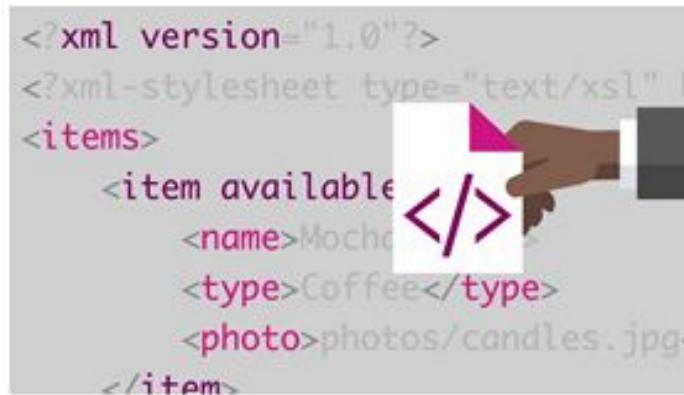
```
Titulo.txt - Notepad
File Edit Format View Help
Codigo Nome Status
1 Carmencita A
2 Le clown et ses chiens A
3 Pauvre Pierrot A
4 Un bon bock A
5 Blacksmith Scene A
```

Codigo	Nome	Status
1	Carmencita	A
2	Le clown et ses chiens	A
3	Pauvre Pierrot	A
4	Un bon bock	A
5	Blacksmith Scene	A

Dados na Era da Informação

IGTI

- “Pré-era da Informação”:
 - Expansão da Internet;
 - Volume de dados iniciando um crescimento exponencial;
 - Sistemas trabalhando também com dados semiestruturados.

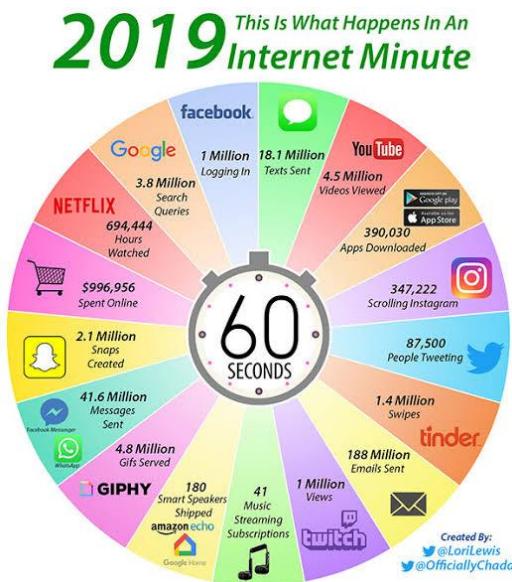


Dados na Era da Informação

IGTI

- Era da Informação:

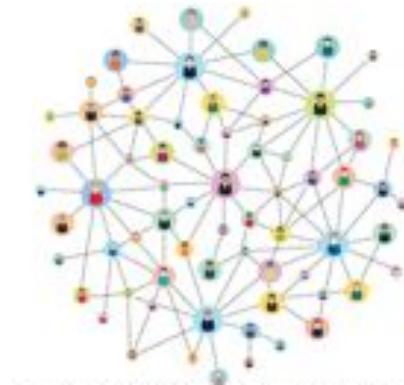
- Uso massivo de Internet banda larga;
- Computação ubíqua, Internet das Coisas (IOT), Big Data, etc.;
- Uso intensivo de redes sociais para divulgação de informações e conteúdo.



Dados na Era da Informação

IGTI

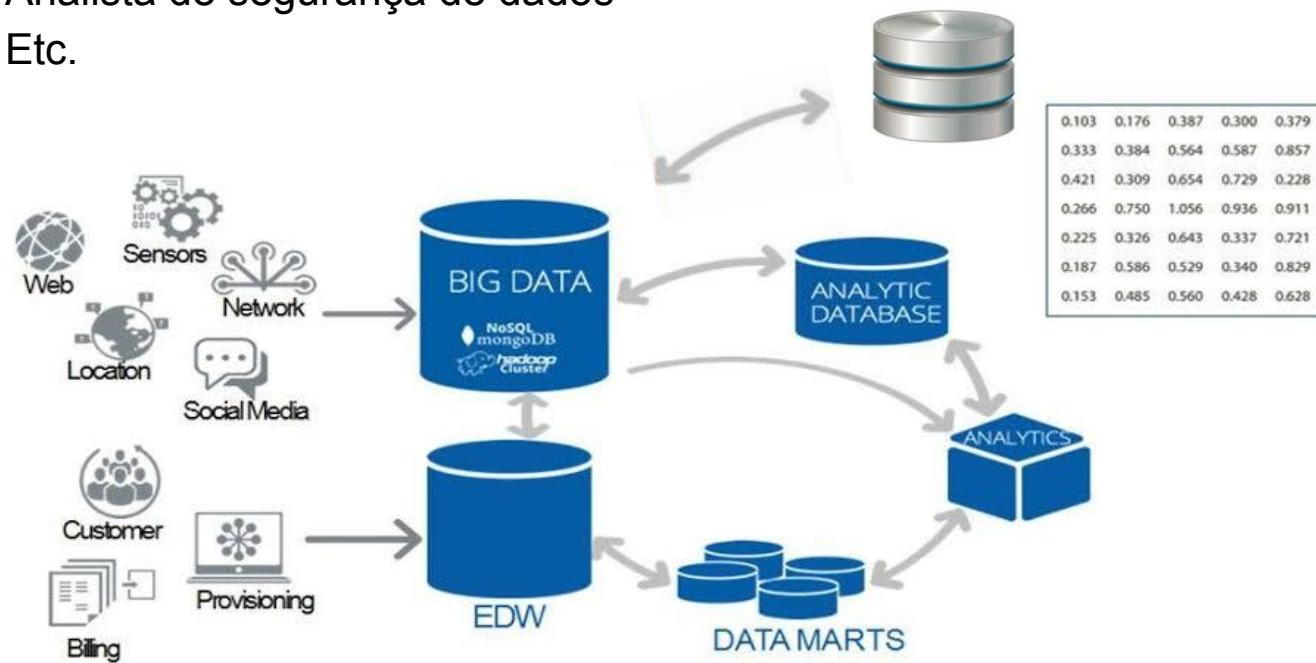
- Era da Informação:
 - Crescimento exponencial do volume de dados, principalmente não estruturados
 - Fotos, vídeos, posts, likes, snaps, etc.;
 - Proliferação de soluções de Big Data para lidar com esse alto volume de dados não estruturados;
 - Coexistência com soluções legadas e/ou baseadas em dados estruturados / semiestruturados.



Dados na Era da Informação

IGTI

- Papel fundamental do **Profissional de Dados**.
 - Arquiteto de dados
 - Engenheiro de dados
 - Analista de dados
 - Analista de segurança de dados
 - Etc.



Próxima Aula



- Modalidades de Serviços para Armazenamento de Dados.

Introdução à Plataforma de Dados do Azure

CAPÍTULO 1. AULA 1.3. MODALIDADES DE SERVIÇOS PARA ARMAZENAMENTO DE DADOS

PROF. GUSTAVO AGUILAR

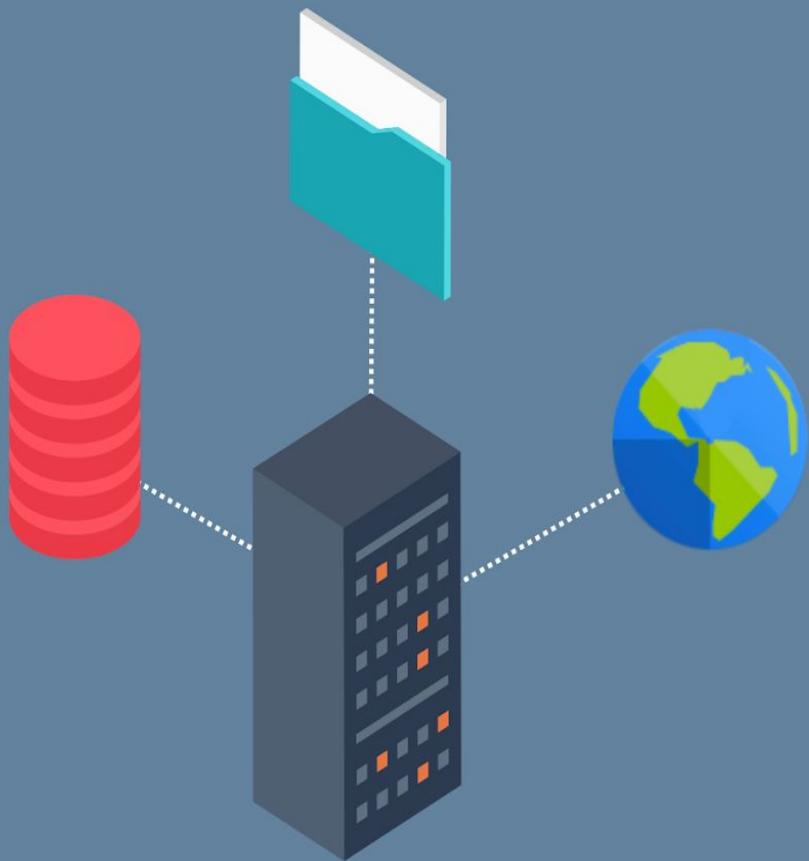
Nesta Aula



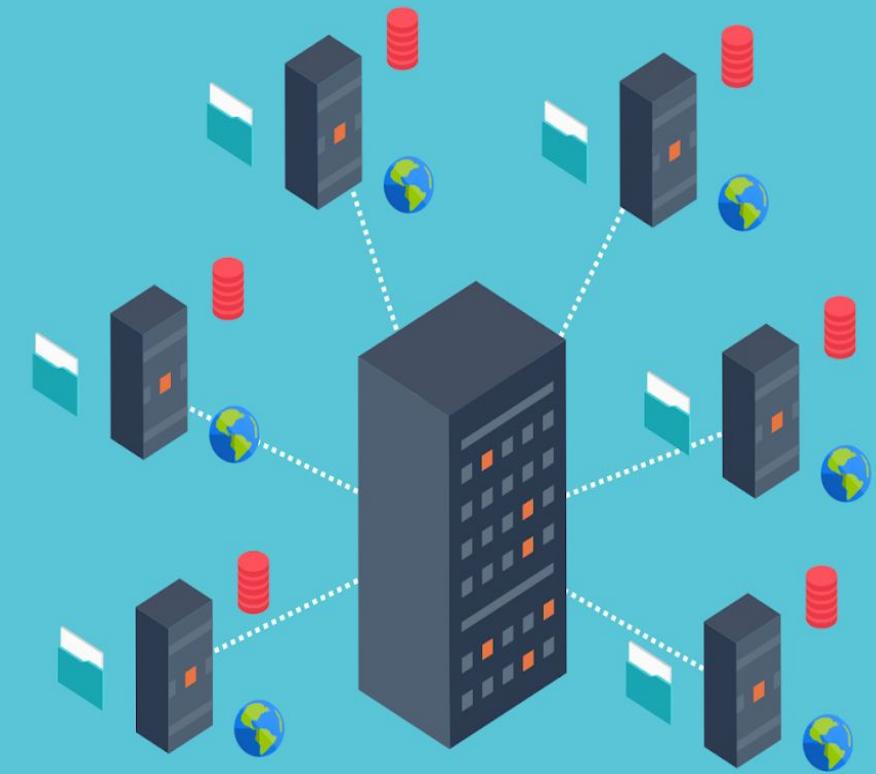
- Servidor Físico x Virtual x Container
- Modalidades de Serviços
- IaaS x PaaS x SaaS
- Quando Usar IaaS

Servidor Físico x Virtual

IGTI

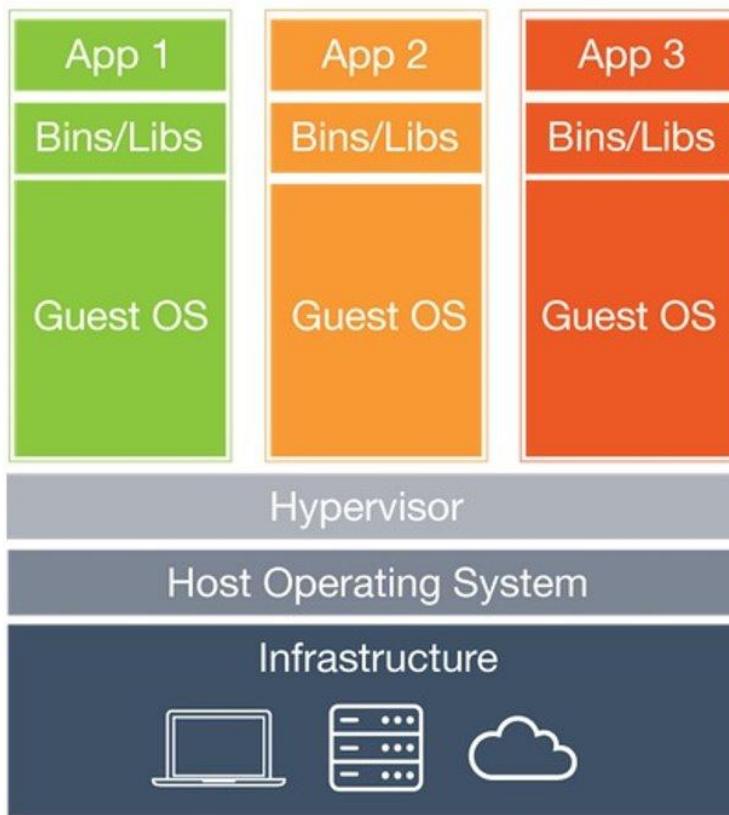


X

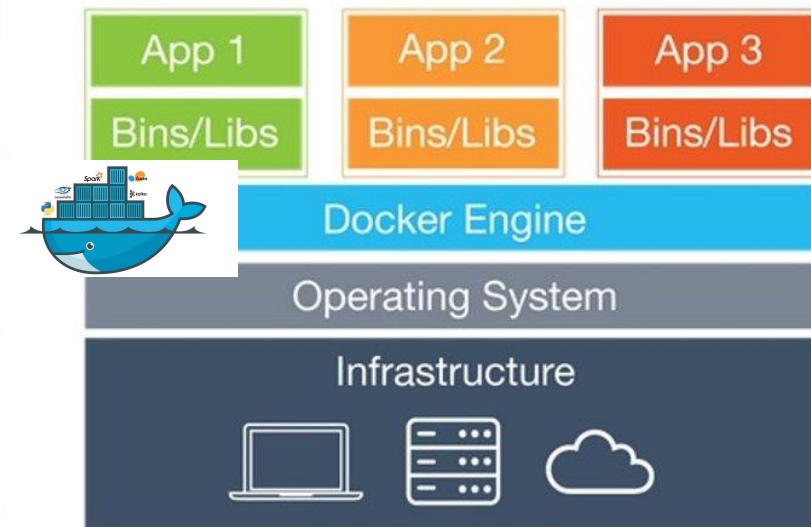


Servidor Virtual x Container

IGTI



Virtual Machines



Containers

Modalidades de Serviços



IaaS

Infraestrutura como serviço

Servidores, armazenamento, segurança, rede
Windows Azure VM, EC2, VPCs, S3, etc.

Definem os diferentes níveis de responsabilidade compartilhada pelos quais um provedor de nuvem e um locatário de nuvem são responsáveis.



PaaS

Plataforma como serviço

Ferramentas Dev., SO, BD

Windows Azure, SQL Azure, etc.



SaaS

Software como Serviço

Apps / aplicativos hospedados

Office 365, Webmails, Redes Sociais, etc.

Modalidades de Serviços

● customer's responsibility ● vendor's responsibility

On-Premises	IaaS	PaaS	SaaS
Servers	Servers	Servers	Servers
Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking
Virtualization	Virtualization	Virtualization	Virtualization
OS	OS	OS	OS
Middleware	Middleware	Middleware	Middleware
Runtime	Runtime	Runtime	Runtime
Apps	Apps	Apps	Apps
Data	Data	Data	Data

Faça tudo

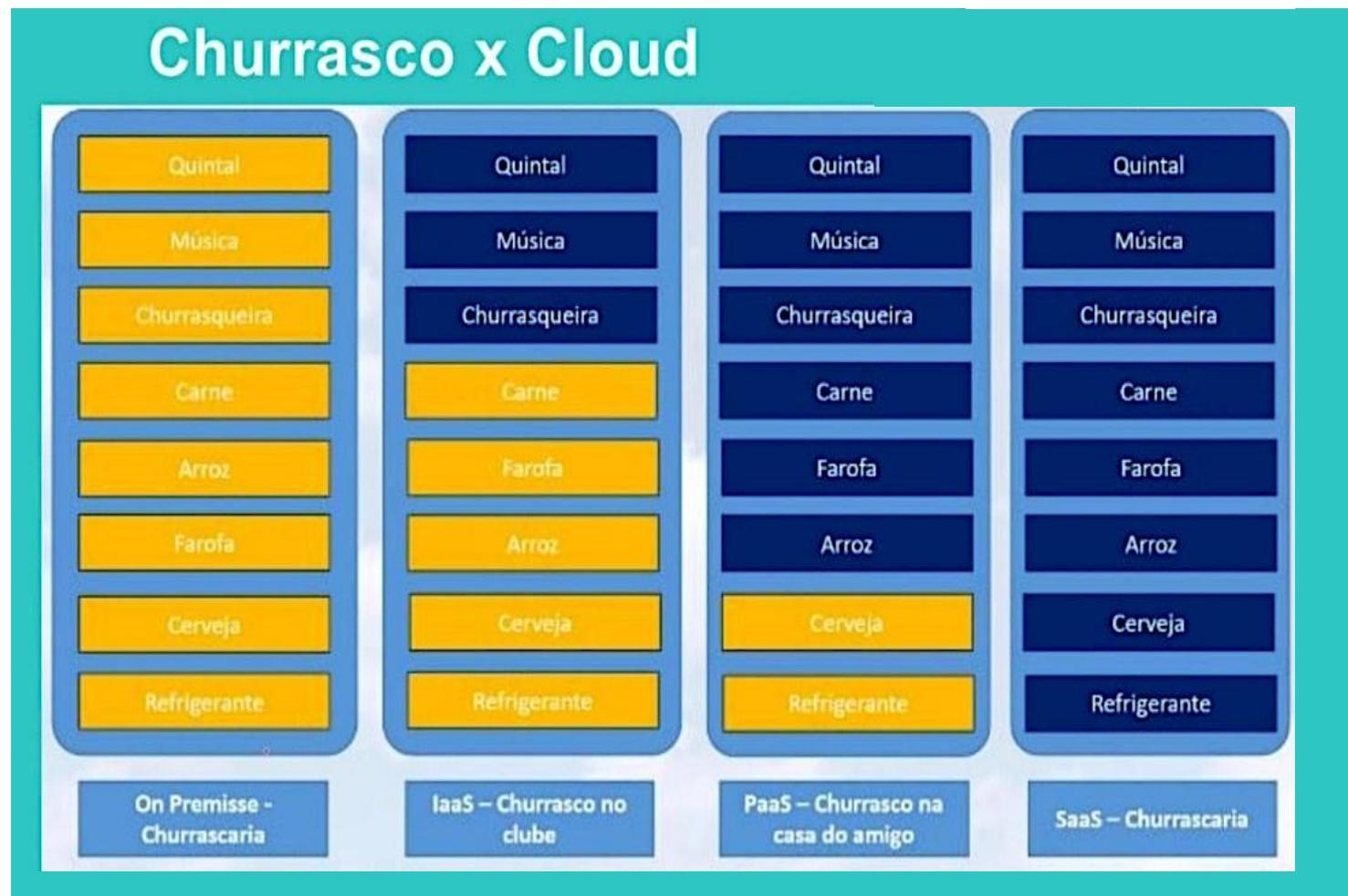
Migre para

Construa com

Use

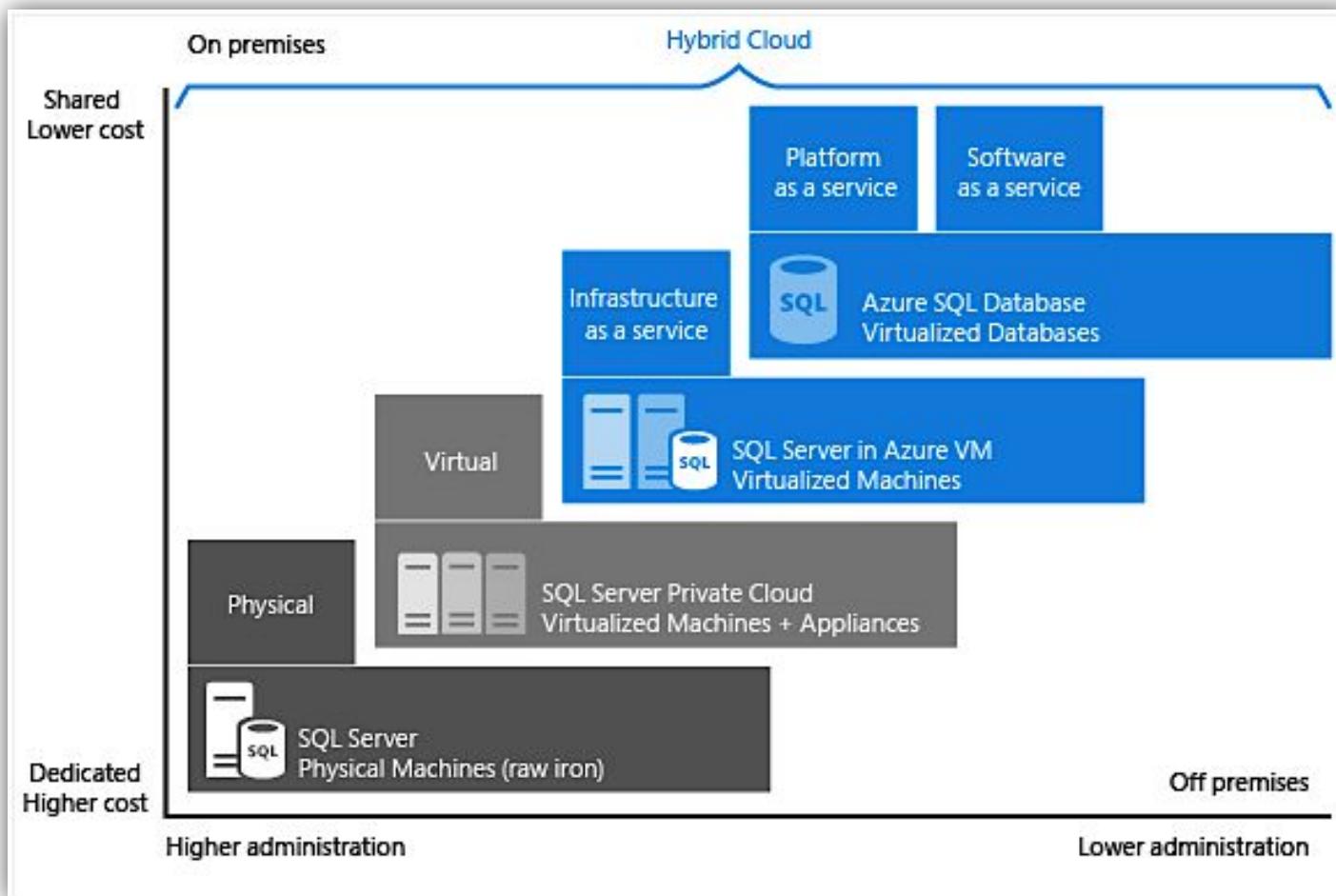
Responsabilidades

Modalidades de Serviços



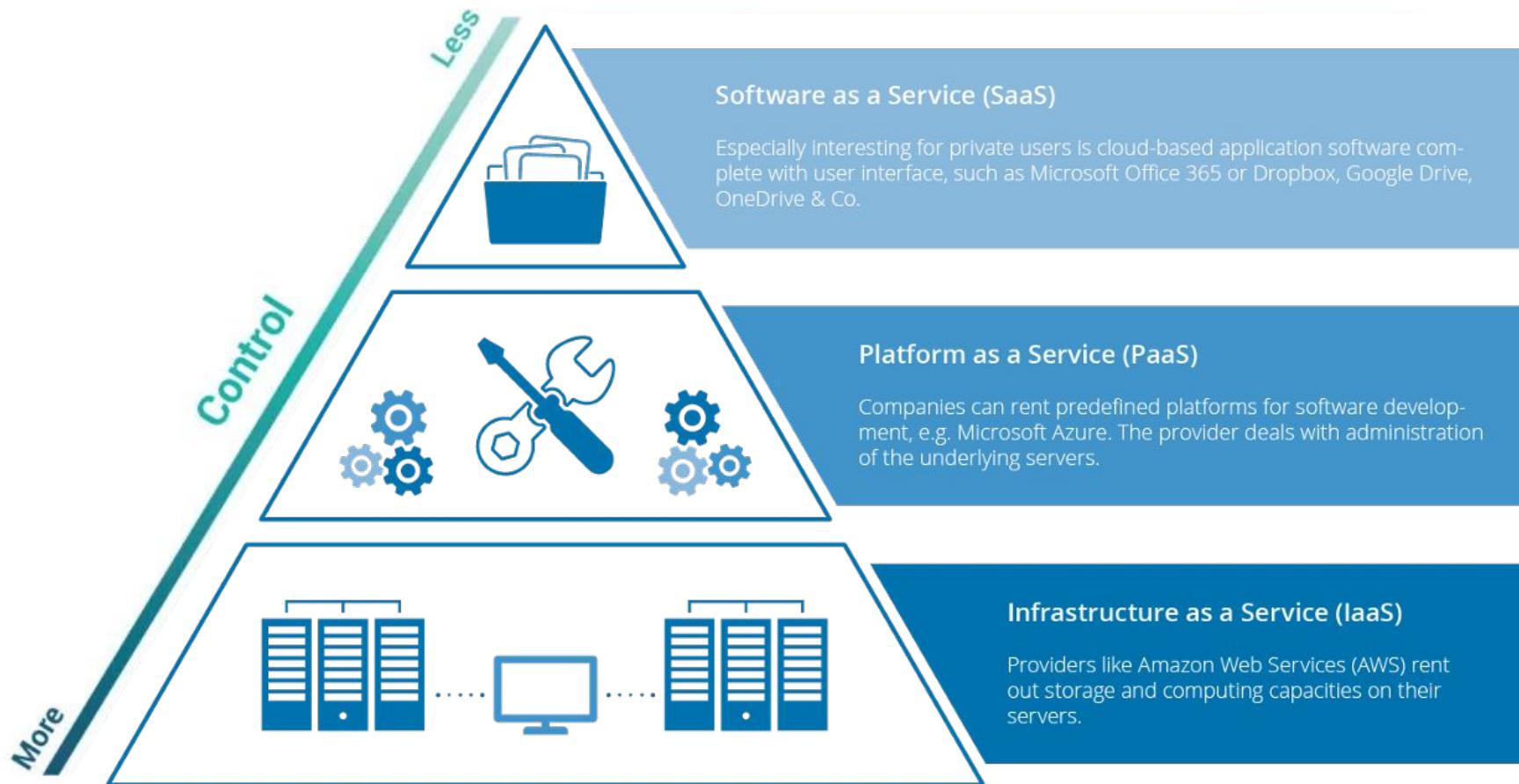
Modalidades de Serviços

IGTI



Modalidades de Serviços

iG Ti



Quando Usar IaaS

- Versões mais antigas de sistema operacional / SGBD;
- Uso de outros serviços do SO / SGBD / SW;
- Necessidades dos aplicativos x Recursos de PaaS;
- Não desejar atualizações automáticas;
- Serviço com menor custo;
- Facilidade de migração: VM on premise ☐ VM na nuvem. (Lift & Shift).

Próxima Aula



- Plataforma de Dados do Azure.

Introdução à Plataforma de Dados do Azure

CAPÍTULO 1. AULA 1.4. PLATAFORMA DE DADOS DO AZURE

PROF. GUSTAVO AGUILAR

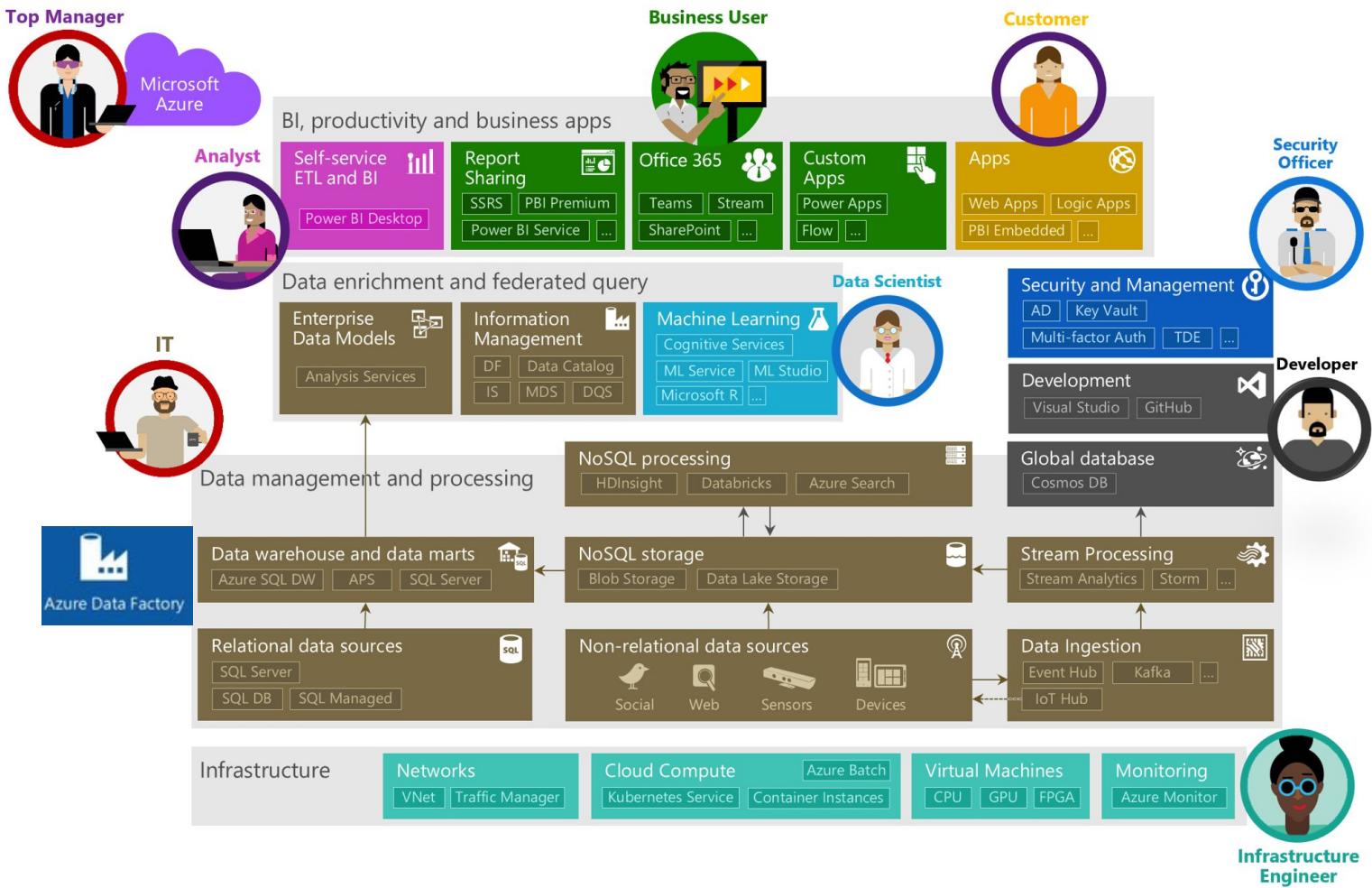
Nesta Aula



- ❑ Overview da Plataforma de Dados do Azure

Plataforma de Dados do Azure

IGTI



Azure Storage

- Tipo de armazenamento básico usado no Microsoft Azure;
- Armazenamento de arquivos, dados textuais / binários / não estruturados e de fila de mensagens;
- Pode ser usado com máquinas virtuais, via SMB ou API;
- Cada serviço é acessado através de uma conta de armazenamento (Storage Account);
- Possuem cinco opções:
 - Azure Blobs;
 - Azure Files;
 - Azure Queues;
 - Azure Tables;

Azure Storage

IGTI

IaaS			PaaS			
Storage	Virtual machines	Networking	Existing frameworks	Web and mobile	Microservices	Serverless Compute
Disks Persistent disks for Azure IaaS VMs Standard Storage Disks: Magnetic disk based, low IOPS, moderate latency Premium Storage Disks: SSD based, high IOPS, low latency Managed Disks	Files Fully Managed File Shares in the Cloud SMB and REST access "Lift and shift" legacy apps	Blobs Highly scalable, REST based cloud object store Block Blobs: Sequential I/O, Hot, Cool and Archive Tiers Page Blobs: Random-write pattern data Append Blobs	Tables Massive auto-scaling NoSQL store Dynamic scaling based on load Scale to PBs of table data Fast key/value lookups	Queues Reliable queues at scale for cloud services Decouple and scale components Message visibility timeout and update message to protect against unreliable dequeuers		

SQL Virtual Machines

IGTI

Microsoft Azure Search resources, services, and docs (G+/-)

Home > Marketplace > Azure SQL >

Select SQL deployment option

Microsoft

Feedback

How do you plan to use the service?

SQL databases
Best for modern cloud applications. Hyperscale and serverless options are available.

Resource type

Create **Show details**

SQL managed instances
Best for most migrations to the cloud. Lift-and-shift ready.

Resource type

Create **Show details**

SQL virtual machines
Best for migrations and applications requiring OS-level access. Lift-and-shift ready.

Image

SQL Server 2019 on Windows Server 2019
Free SQL Server License: SQL 2019 Developer on Windows Server 2019

SQL Server 2019 Enterprise Windows Server 2019

SQL Server 2019 Standard on Windows Server 2019

{BYOL} SQL Server 2019 on Windows Server 2019

{BYOL} SQL Server 2019 Enterprise Windows Server 2019

{BYOL} SQL Server 2019 Standard on Windows Server 2019

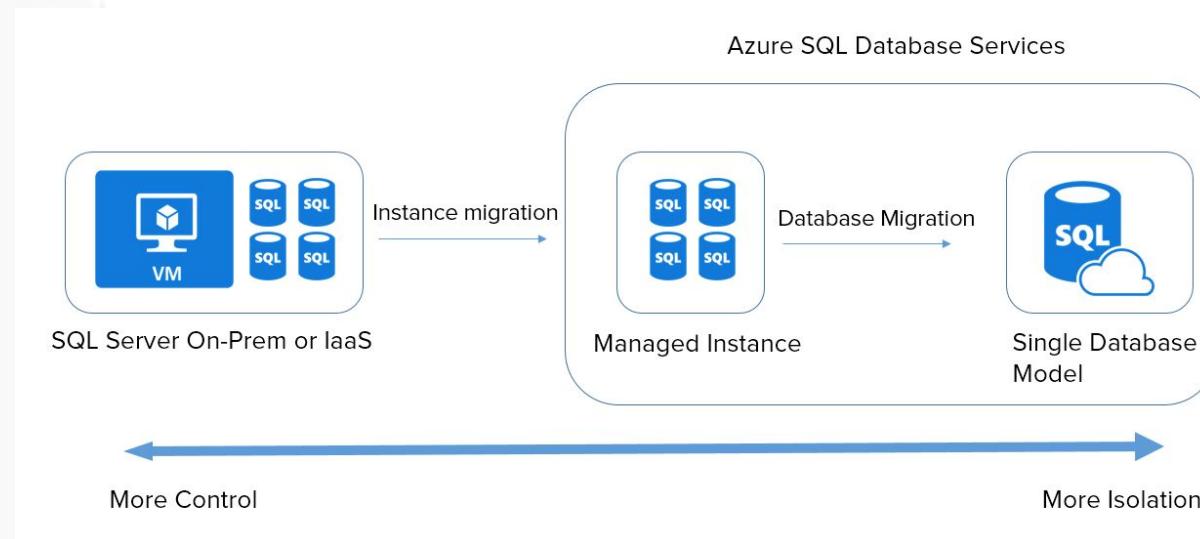
SQL Server 2019 on RHEL74
Free SQL Server License: SQL Server 2019 Developer on Red Hat Enterprise Linux 7.4



"IaaS"
Infrastructure-as-a-Service
host

Azure SQL Managed Instance

- Fornece uma instância inteira do SQL Server □ PaaS
 - Até 100 bancos de dados;
- Fornece outros recursos que não estão disponíveis no Azure SQL Database:
 - Consultas entre bancos de dados;
 - CLR (Common Language Runtime);
 - SQL Agent (jobs e scheduler).



Azure SQL Database

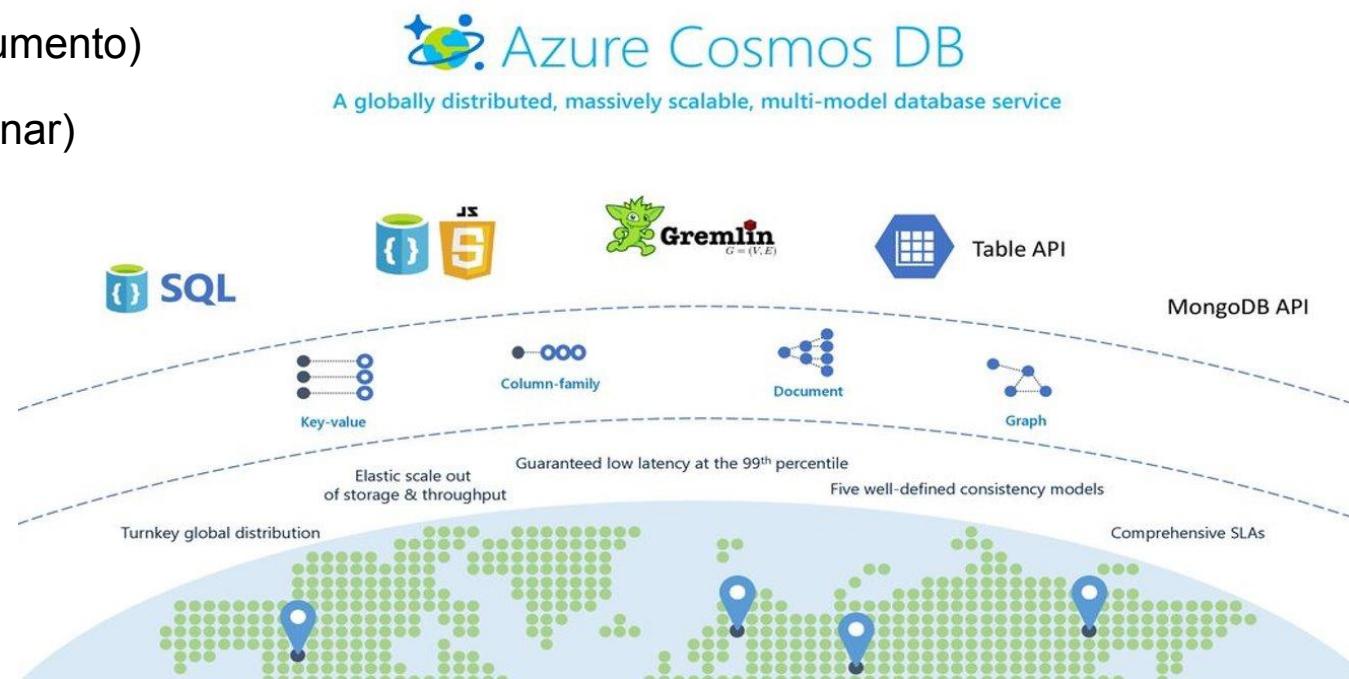


- Serviço de banco de dados relacional gerenciado no Azure;
- Database as a Service □ PaaS;
- Suporta estruturas como dados relacionais e formatos não estruturados, como dados espaciais e XML.



Azure Cosmos DB

- Serviço de banco de dados multimodelo e distribuídos globalmente;
- Database as a Service ☐ PaaS;
- Suporte a 5 APIs (modelos de DB Engine)
 - SQL (relacional);
 - Mongo DB (documento)
 - Cassandra (colunar)
 - Gremlin (grafo)
 - Table



Azure Data Services

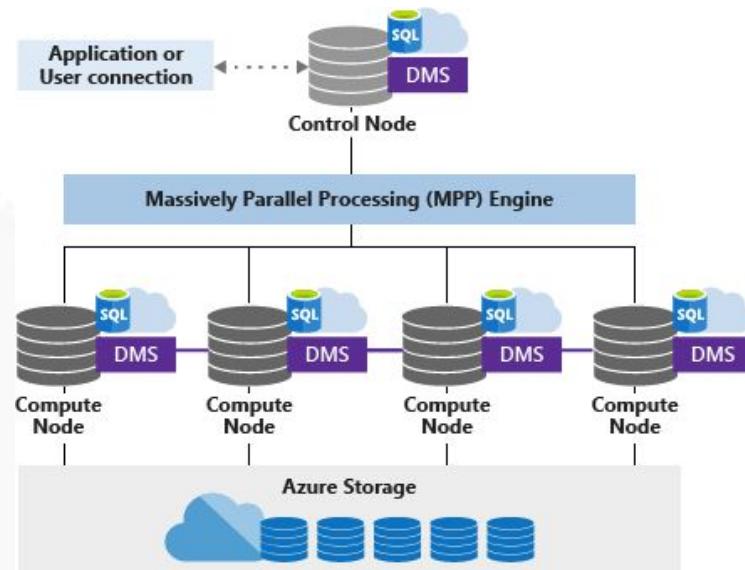
iGTTI



Azure
Redis Cache

Azure Synapse Analytics

- Anteriormente conhecido como SQL DW;
- Fornece um ambiente unificado combinando data warehouse e os recursos de análise de Big Data do Spark;
- MPP □ processamento massivo paralelo.



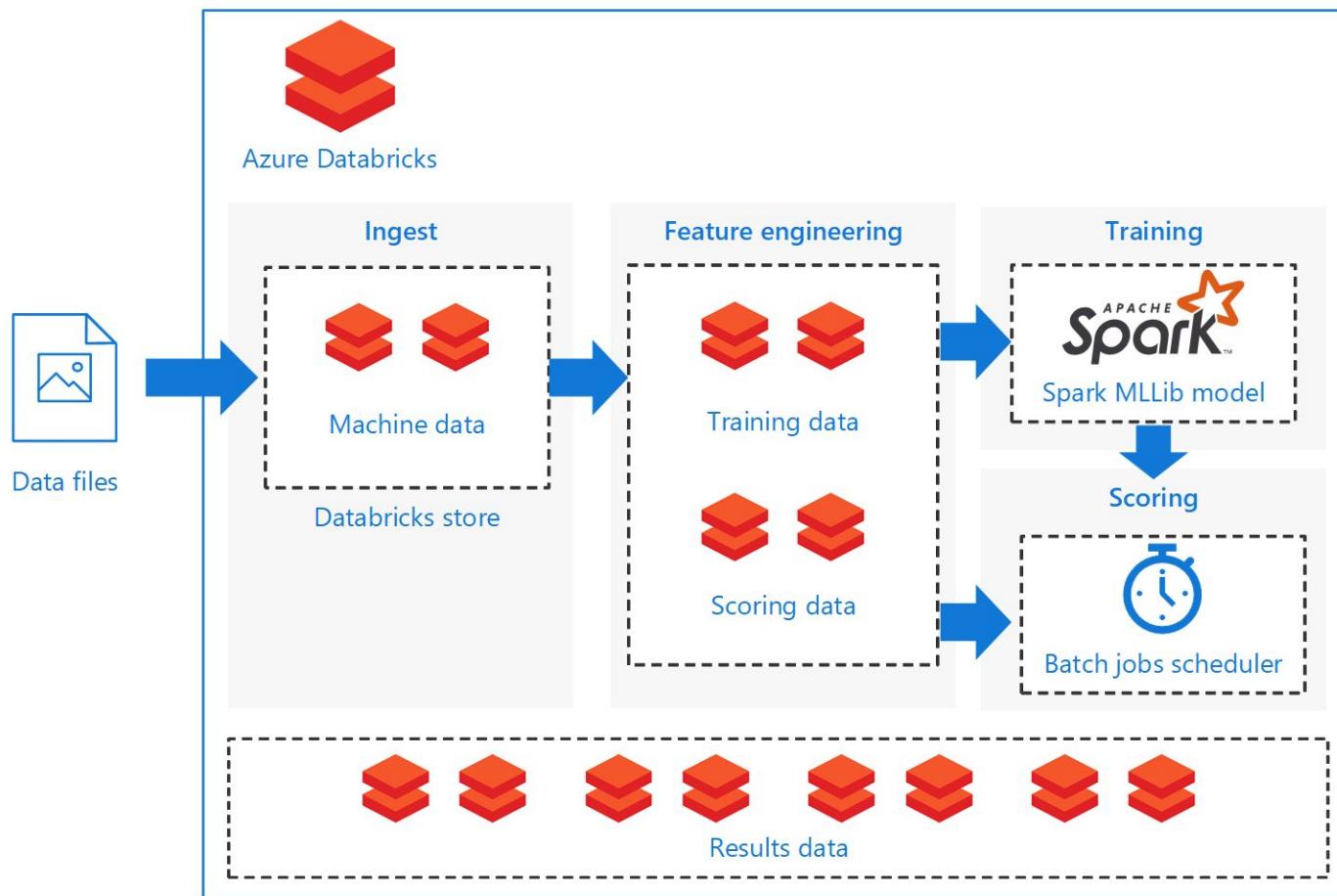
Azure Databricks



- O Databricks é uma versão do popular mecanismo de análise e processamento de dados Apache Spark;
- O Azure Databricks é a versão totalmente gerenciada do Databricks □ PaaS;
- Oferece uma plataforma de Big Data e aprendizado de máquina;
- Suporte a Python, Scala, R, Java e SQL, além de estruturas e bibliotecas de ciência de dados, incluindo TensorFlow e PyTorch.

Azure Databricks

IGTI



Azure HDInsight

- Processamento e análise de big data para dar suporte ao processamento em lote, de data warehousing, IoT e Data Science;
- Solução em nuvem de baixo custo que contém várias tecnologias:



Apache Hadoop



Apache Spark



Apache Kafka



Apache HBase



Apache Hive LLAP



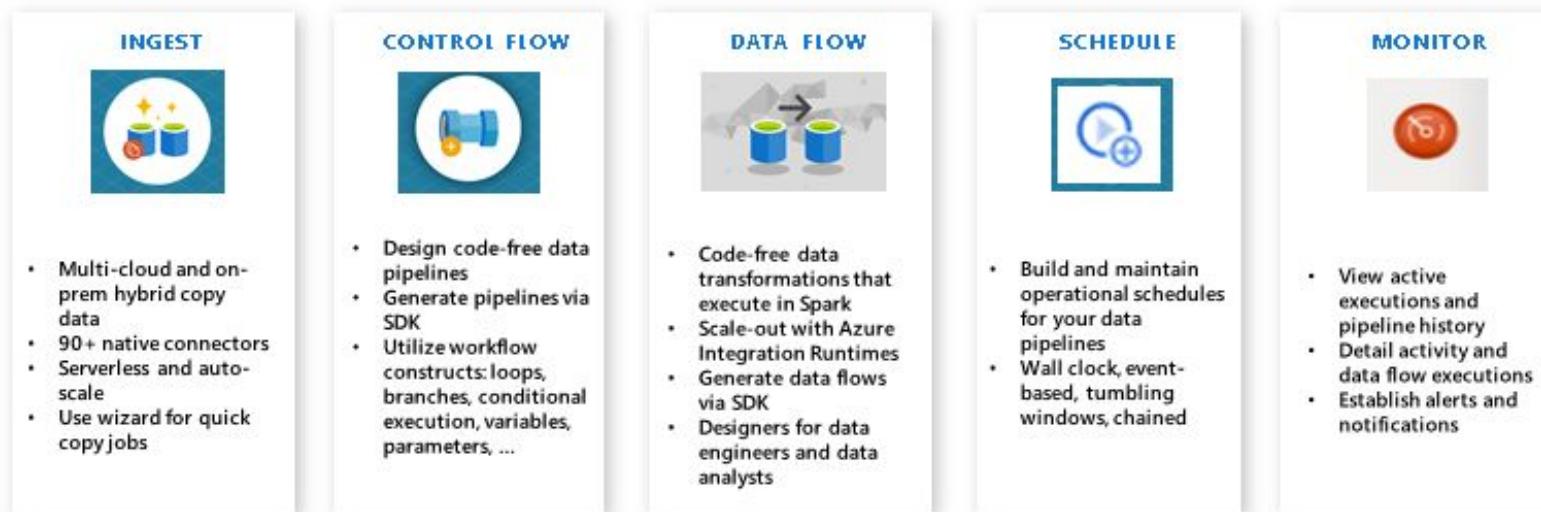
Apache Storm



Machine Learning

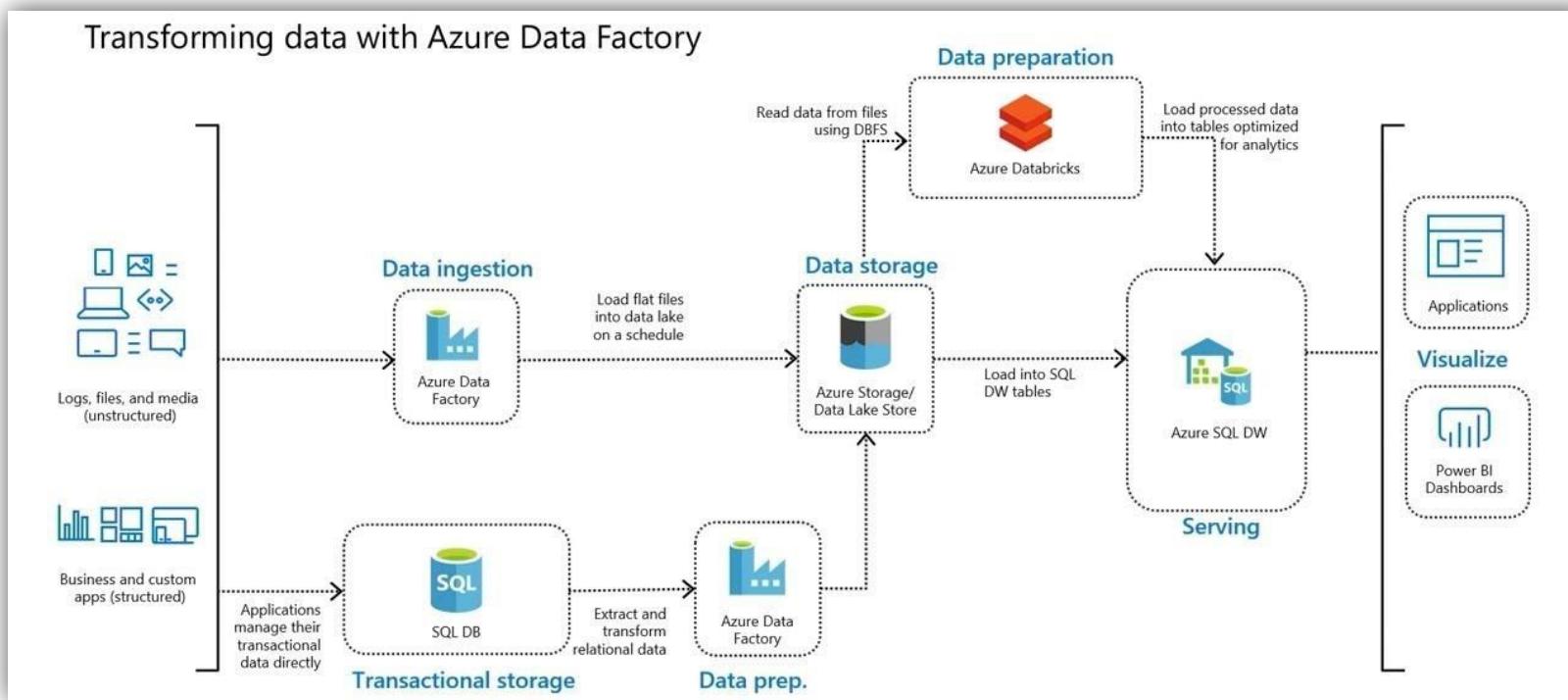
Azure Data Factory

- Serviço de integração de dados que orquestra a movimentação de dados entre várias fontes de dados;
- ETL baseado em nuvem
 - Similar ao que o Integration Services (SSIS) / Pentaho faz no on premises.

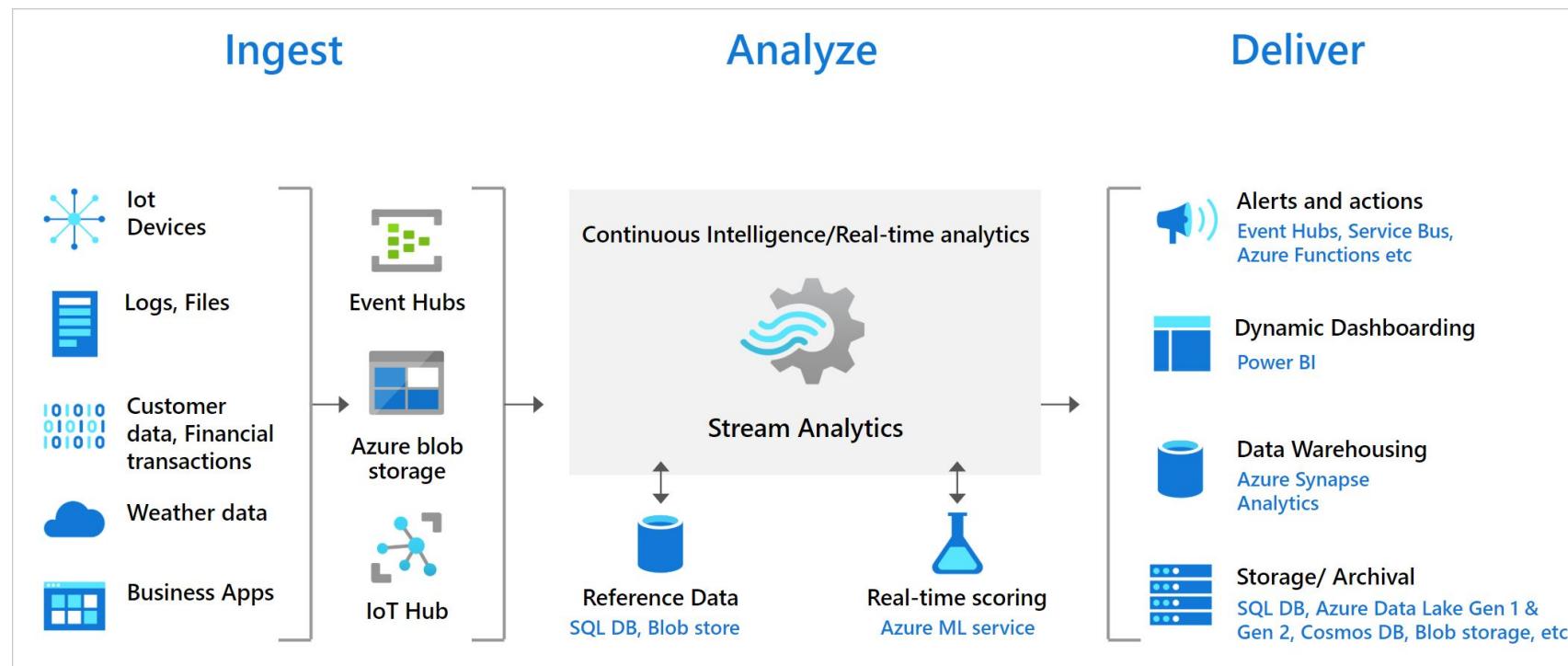


Azure Data Factory

IGTI

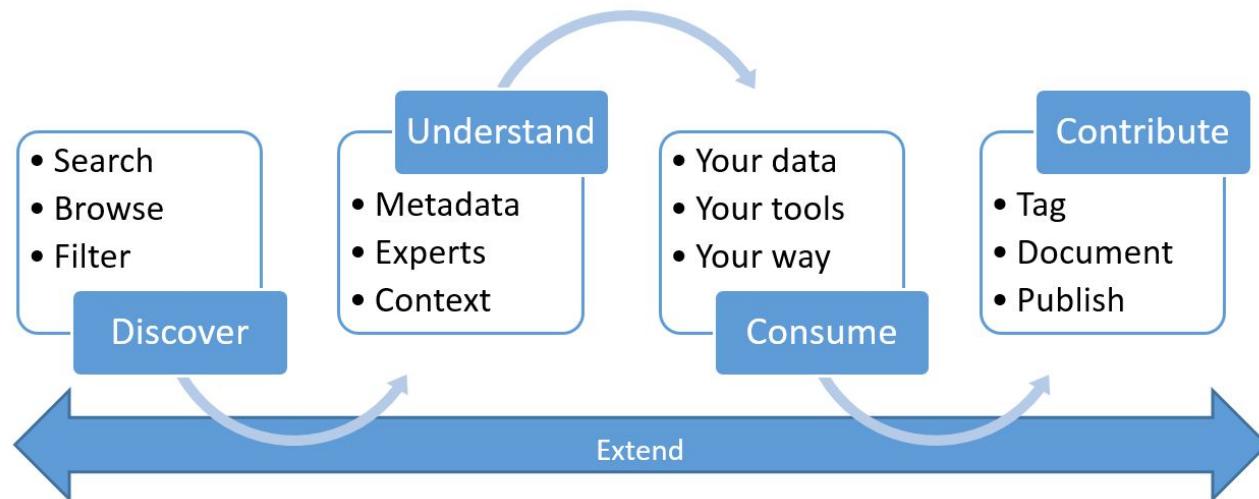


Stream Analytics



Azure Data Catalog

- Serviço para catalogar dados e fontes de dados;
- Inclui um modelo de alimentação de metadados □ *crowdsourcing*;
- Os usuários podem descobrir as fontes de dados de que precisam e entender as fontes de dados que encontram, além de usar o Catálogo de Dados para documentar as informações sobre suas fontes de dados.



Azure Purview

iG Ti

Microsoft Azure | Purview > Adatum Corp

Revenue

contoso@contoso.com MICROSOFT

Sources

Register + New collection Refresh

Showing 5 collections, 1134 sources

Map view

NorthAmericaDataCenter Collection

- OnPremSQLServer-Fina... SQL Server
- Teradata-FinanceData Teradata (Preview)
- HiveMetastore Hive Metastore (Preview)
- FinanceSQLServer SQL Server
- Teradata Teradata (Preview)
- OnPremSQLServer SQL Server

EuropeDataCenter Collection

- SAP-S4HANA-Procurement... SAP S/4Hana (Preview)
- SAP-ECC-SalesData SAP ECC (Preview)
- SAP-S4HANA SAP S/4Hana (Preview)
- SAP-ECC SAP ECC (Preview)

AzureAndBINorthAmerica Collection

- AzureDataLakeStorage... Azure Data Lake Storage Gen2
- AzureBlobStorage Azure Blob Storage
- AzureSQLDB-SalesInvo... Azure SQL Database
- RevenuePBIDashboards Power BI
- WebLogs Azure Files
- AzureSqlManagedInsta... Azure SQL Database Managed Instance

AmazonNorthAmerica Collection

- AmazonS3-HRData Amazon S3
- AWSS3 Amazon S3

AzureEurope Collection

- AzureDataLakeStorage... Azure Data Lake Storage Gen2

Map view

Azure Purview

IGTi

Microsoft Azure | Purview > Adatum Corp

contoso@contoso.com MICROSOFT

Adatum Corp

13,020 users 1,134 sources 76,392,971 assets 312 terms

Revenue

Search suggestions

- revenue customer
- revenue finance
- revenue glossary
- revenue information

Asset suggestions

- TaxRevenue
teradata://adatumteradata.retail.com/DB/TaxRevenue
- CustomerRevenue_{N}
https://adatumcostorage.blob.core.windows.net/CUST/CustomerRevenue_{N}.csv
- Revenue
https://adatumco.dfs.core.windows.net/incr/raw/revenue
- Revenue2020.tsv
https://adatumco.dfs.core.windows.net/sales/Revenue2020.tsv

View search results

Recently accessed My items

Name	Type	Last updated	Actions
Order	Asset	10 days ago	View details
Percent Sales	Asset	1 month ago	View details
SalesOrderHeader	Asset	1 month ago	View details
ProductCategory	Asset	1 month ago	View details
SSNNumber	Asset	1 month ago	View details
SalesLT.SalesOrderHeader.csv	Asset	1 month ago	View details

View all recently accessed

Knowledge center
Discover learning and tutorials.

/view insights
Get insights on your data.

Próxima Aula



- Escalabilidade na Nuvem.

Introdução à Plataforma de Dados do Azure

CAPÍTULO 1. AULA 1.5. ESCALABILIDADE NA NUVEM

PROF. GUSTAVO AGUILAR

Nesta Aula



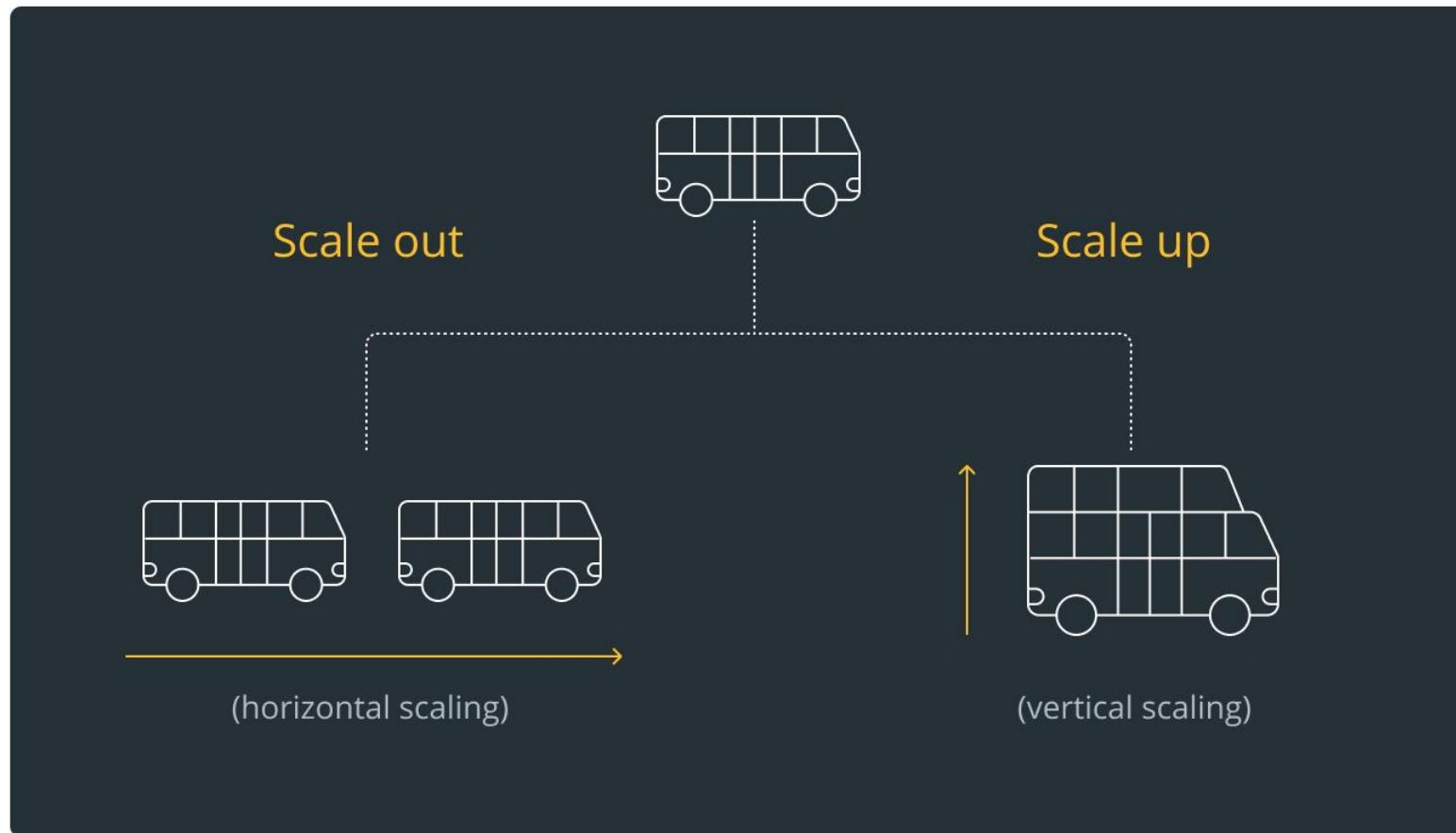
- O Que é Escalabilidade?
- Escalabilidade Vertical
- Escalabilidade Horizontal
- Elasticidade

O Que É Escalabilidade?

- Capacidade de escalar o ambiente, ou seja, de adicionar mais recursos computacionais ao ambiente;
 - ✓ Poder de processamento (CPU)
 - ✓ Cache (memória RAM)
 - ✓ Capacidade de armazenamento (disco)
 - ✓ Velocidade de transmissão (network)
 - ✓ Nós de processamento (worker node)
 - ✓ Nós de armazenamento (data node)
- Pode ser feita de forma transparente ou não;
- Pode ser vertical ou horizontal.

Horizontal x Vertical

iG Ti



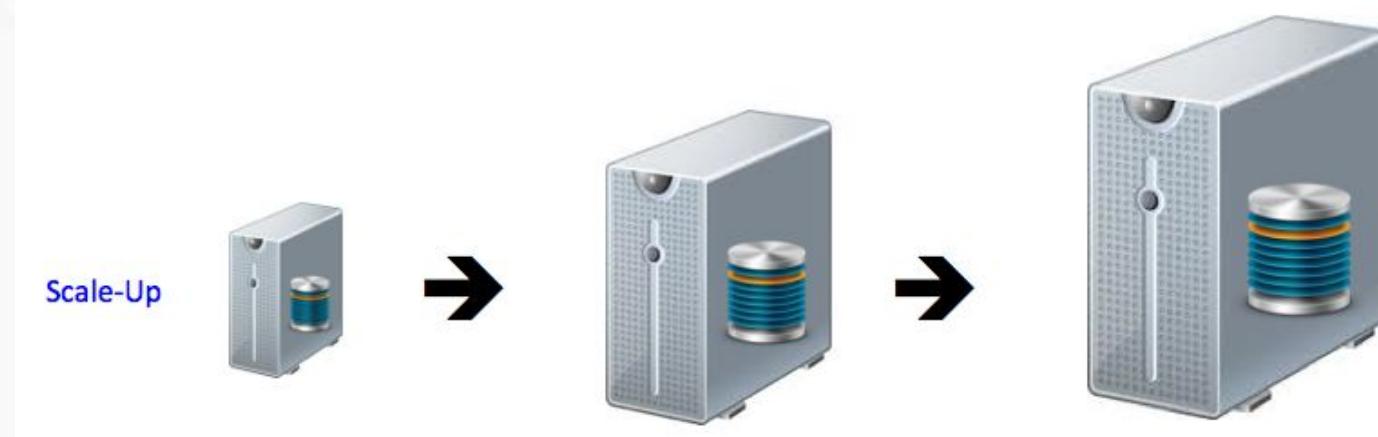
Escalabilidade Vertical

IGTI



Escalabilidade Vertical

- Adicionar mais recursos (CPU, RAM, etc.) aos integrantes do ambiente;



- **Não se altera a quantidade de integrantes do ambiente.**

Escalabilidade Horizontal

IGTI



Escalabilidade Horizontal

- Adicionar novos integrantes ao ambiente;



- Altera-se a quantidade de integrantes do ambiente.

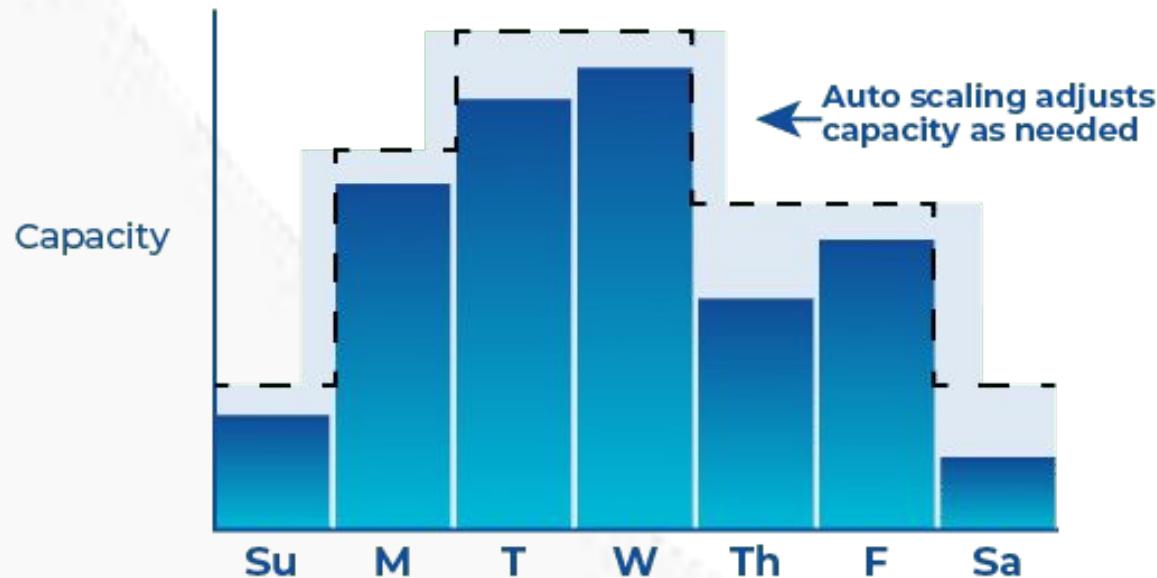
Elasticidade

- Capacidade de aumentar e reduzir rapidamente os recursos computacionais de um ambiente;
- Requer escalabilidade vertical / horizontal.



Computação Elástica

- Para atender às exigências dinâmicas, sem se preocupar com o planejamento de capacidade e picos de utilização.



Próxima Aula



- Cargas de Trabalho.

Introdução à Plataforma de Dados do Azure

CAPÍTULO 1. AULA 1.6. CARGAS DE TRABALHO

PROF. GUSTAVO AGUILAR

Nesta Aula



- Carga de Trabalho
- Tipos de Workloads
- Workloads x Escalabilidade

Carga de Trabalho

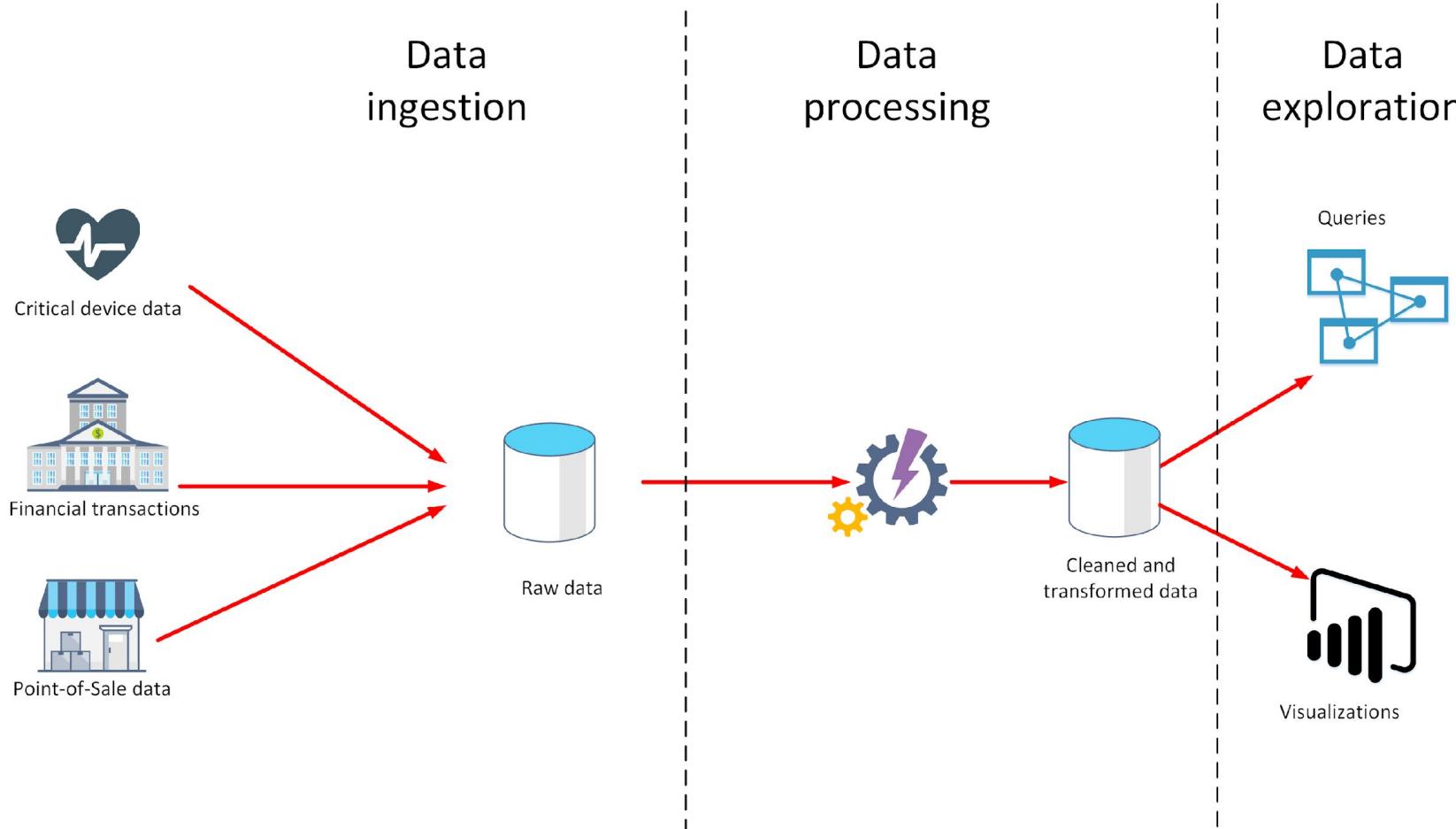
- **Workload:**
- Quantidade de processamento que determinados recursos computacionais recebem para ser executado em um determinado momento;
- Definida por **três propriedades:**
 - **Tempo de Resposta:** o tempo entre uma solicitação e uma resposta;
 - **Taxa de Transferência:** quanto trabalho é realizado durante um período de tempo;
 - **Tipo de Workload:** forma e frequência como a carga de trabalho é executada.

Tipos de Workloads

- Existem várias classificações, mas na área de dados, a mais comum é:
 - **TRANSACIONAL**: baseado em transações (unidade de trabalho pequena e discreta) e também conhecido como **OLTP (Online Transactional Processing)**;
 - **ANALÍTICO**: voltado para cenários de análises de dados oriundos de diversas fontes e também conhecido **OLAP (Online Analytical Processing)**.

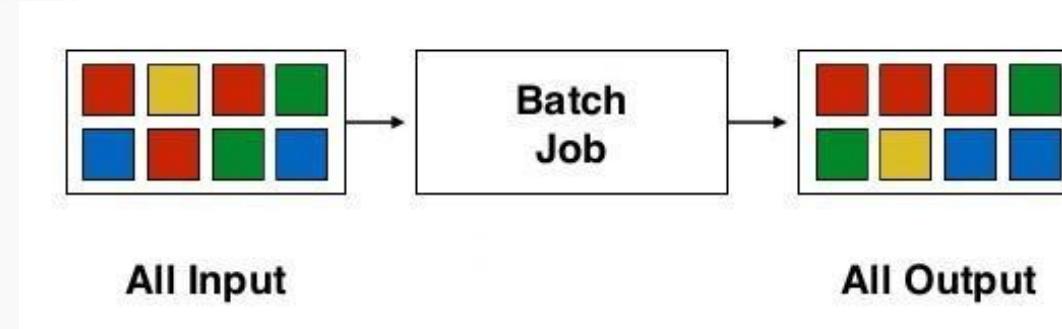
Workload OLAP

IGTI



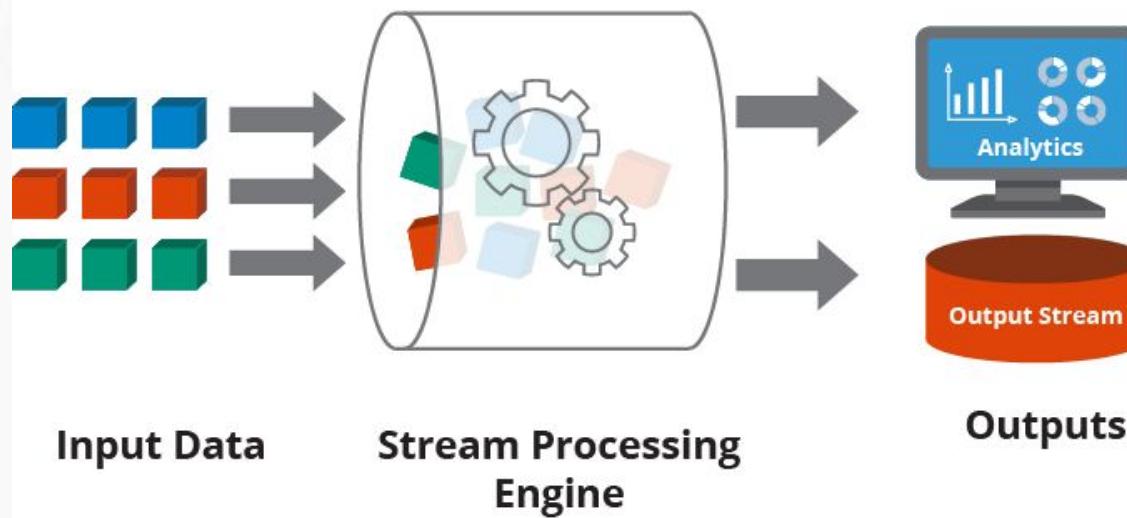
Workload OLAP

- Pode acontecer de duas formas: **BATCH** e **STREAMING**.
 - **BATCH**: dados são coletados em um grupo, e todo o grupo é então processado em um momento futuro como um lote.



Workload OLAP

- **STREAMING:** cada novo dado é processado quando chega.



Batch x Streaming

- **Escopo dos dados:**

- Batch □ pode processar todos os dados do conjunto de dados.
- Streaming □ normalmente só tem acesso aos dados mais recentes recebidos, ou dentro de uma janela de tempo (os últimos 30 segundos, por exemplo).

- **Tamanho dos dados:**

- Batch □ lida com grandes conjuntos de dados de forma eficiente.
- Streaming □ destinado a registros individuais ou micro lotes compostos de poucos registros.

Batch x Streaming

- **Desempenho:**

- Batch □ a latência é normalmente de algumas horas.
- Streaming □ o processamento normalmente ocorre imediatamente, com latência da ordem de segundos ou milissegundos.

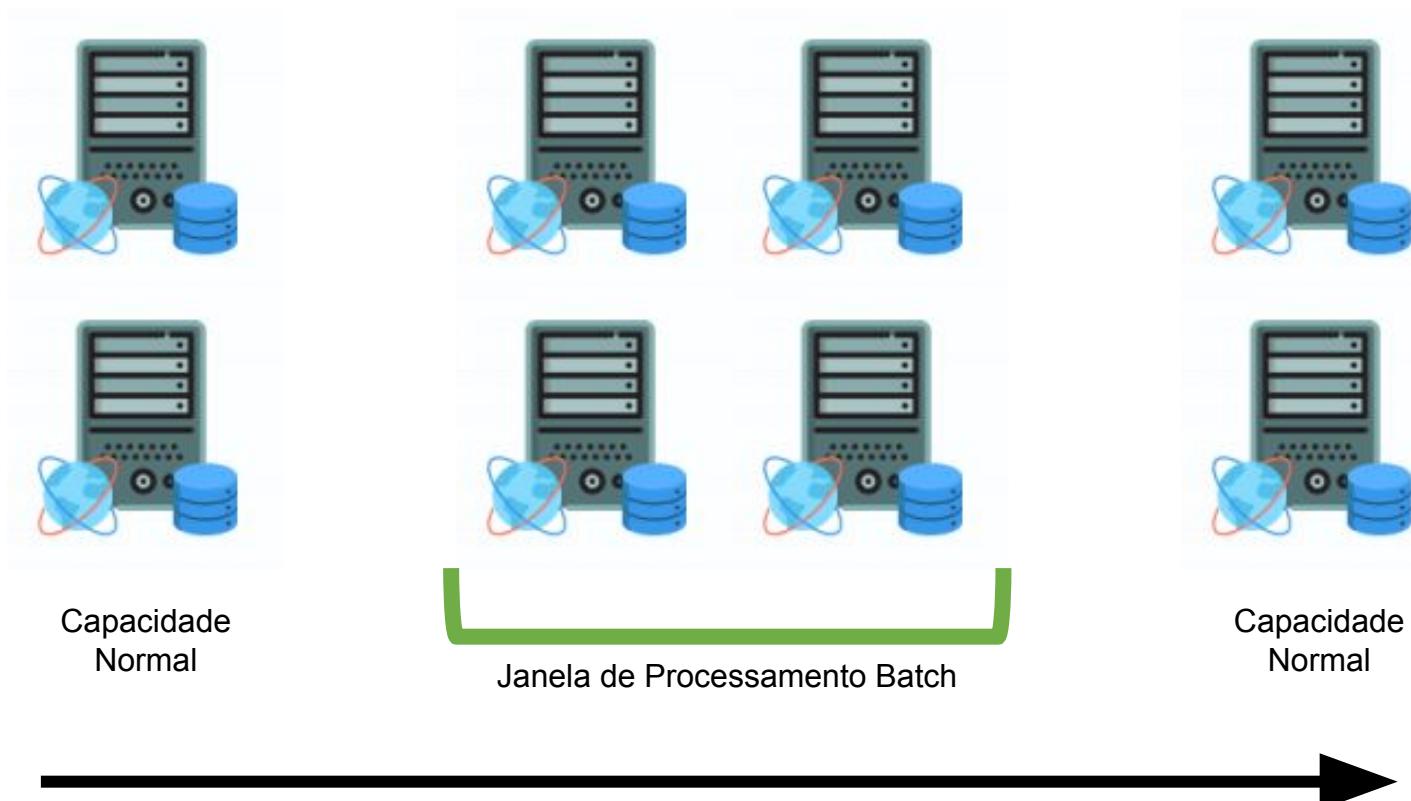
- **Análise:**

- Batch □ análises complexas.
- Streaming □ respostas simples, agregações ou cálculos, como média e desvio padrão.

Workloads x Escalabilidade

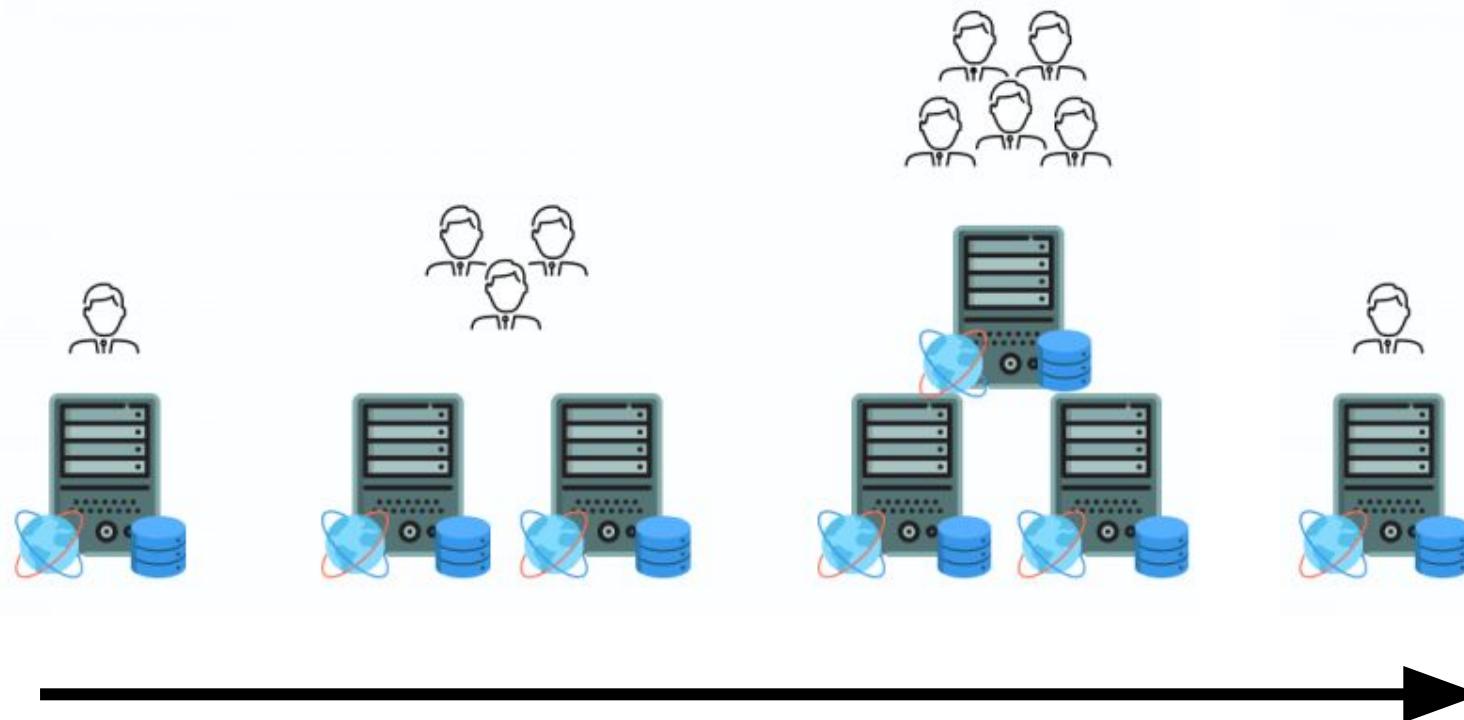
IGTI

- **BATCH:** escalabilidade programada.



Workloads x Escalabilidade

- **STREAMING e OLTP:** escalabilidade sobre demanda.



Próxima Aula



- Capítulo 2 - Armazenamento de Dados.

Serviços de Armazenamento, Banco de Dados e Analytics

Capítulo 2. Armazenamento de Dados

PROF. GUSTAVO AGUILAR

Armazenamento de Dados

CAPÍTULO 2. AULA 2.1. STORAGE ACCOUNT

PROF. GUSTAVO AGUILAR

Nesta Aula

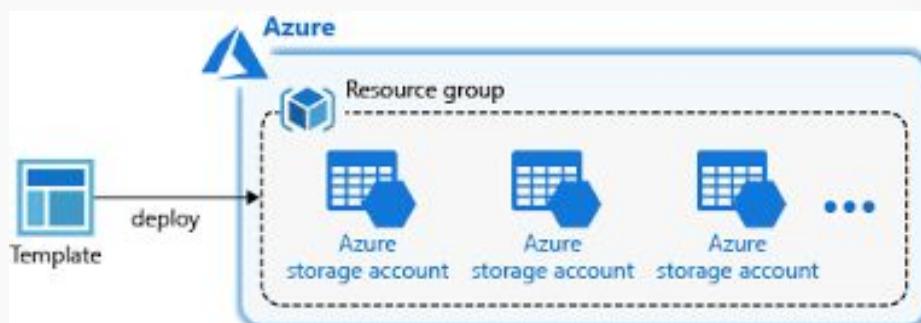
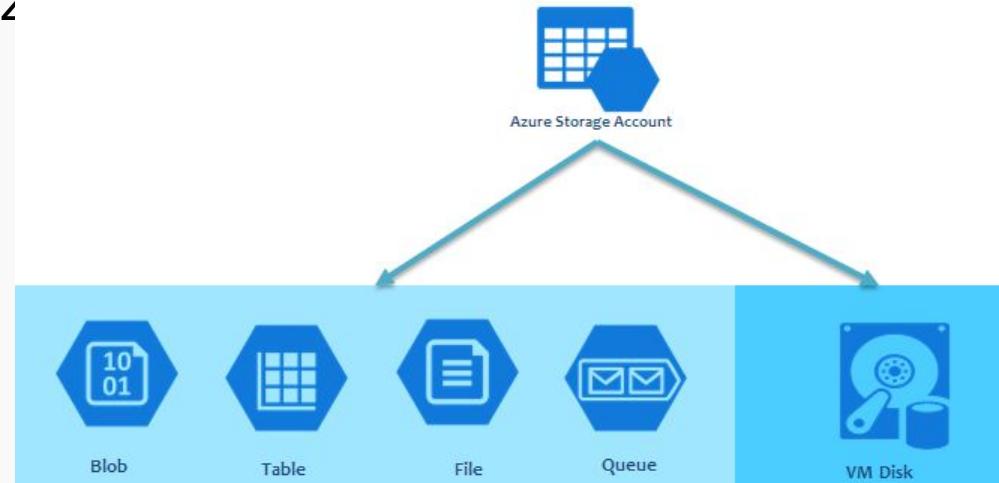


- Storage Account
- Configurações
- Ferramentas para Criação

Storage Account

IGTI

- **Conta de Armazenamento:** contêiner que agrupa um conjunto de serviços de armazenamento da Azure.
- Recurso do Azure, de uma assinatura (subscription), que está incluída em um grupo de recursos (Resource Manager).



Configurações

- Uma storage account define uma política que se aplica a todos os serviços de armazenamento na conta;
- Exemplo:
 - Datacenter Brazil South;
 - Acessíveis apenas por https;
 - Cobrados na assinatura do departamento de TI.
- **Subscription;**
- **Resource group;**
- **Storage account name;**

Configurações

- **Desempenho:** determina os serviços de dados e o tipo de disco;
 - **Standard:** qualquer serviço de dados (Blob, Arquivo, Fila, Tabela) e usa unidades de disco magnético;
 - **Premium:** tipo específico de blob chamado Page Blob e usa disco SSD.
- **Camada de acesso:** velocidade que você poderá acessar os blobs
 - **Hot** fornece acesso mais rápido que **Cool**, mas a um custo maior.
- **Redes virtuais:** método de conectividade (público / privado).
- **Proteção dos Dados**
 - Modelo de “lixeira” com expurgo automático □ blob e fileshare;
 - Opção de versionamento.
- **Transferência segura necessária (TLS):** ativado requer HTTPS, enquanto desativado permite HTTP.

Configurações

- **Redundância:** cópia dos dados contra falhas de hardware ou desastres
 - **Armazenamento com redundância LOCAL (LRS):** é o default
 - Protege contra falhas de hardware e de drives;
 - Não protege de um evento que indisponibiliza todo o datacenter.
 - **Armazenamento com redundância GEOGRÁFICA (GRS):**
 - Replicação em diferentes datacenters em todo o mundo;
 - Possibilidade de failover para a região secundária.
 - **Armazenamento com redundância de ZONA (ZRS):**
 - Protege contra falhas no nível de datacenter.
 - **Armazenamento com redundância de ZONA e REGIÃO (GZRS):**
 - GRS + ZRS.

Ferramentas para Criação



- Portal do Azure;
- Azure CLI
 - Command-line interface.
- Azure PowerShell;
- Management client libraries
 - Incorporar a criação dentro de um app.

The screenshot shows the 'Create a storage account' page in the Microsoft Azure portal. At the top, there's a navigation bar with the Microsoft Azure logo, a search bar, and links for Home, Storage accounts, Create a storage account, Basics, Advanced, Networking, Data protection, Tags, and Review + create. The main content area has a heading 'Create a storage account' with a '... more' link. Below it, a section titled 'Project details' asks to select a subscription and resource group. A note says: 'Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.' The 'Subscription' dropdown is set to 'Pay-As-You-Go', and the 'Resource group' dropdown is set to 'NetworkWatcherRG' with a 'Create new' link. The next section, 'Instance details', includes fields for 'Storage account name' (with a red asterisk), 'Region' (set to '(US) East US'), 'Performance' (radio button selected for 'Standard: Recommended for most scenarios (general-purpose v2 account)'), 'Redundancy' (dropdown set to 'Geo-redundant storage (GRS)'), and a checked checkbox for 'Make read access to data available in the event of regional unavailability'. There are also 'Previous' and 'Next' buttons at the bottom.

Próxima Aula



- Criando uma Storage Account.

Armazenamento de Dados

CAPÍTULO 2. AULA 2.2. CRIANDO UMA STORAGE ACCOUNT

PROF. GUSTAVO AGUILAR

Nesta Aula

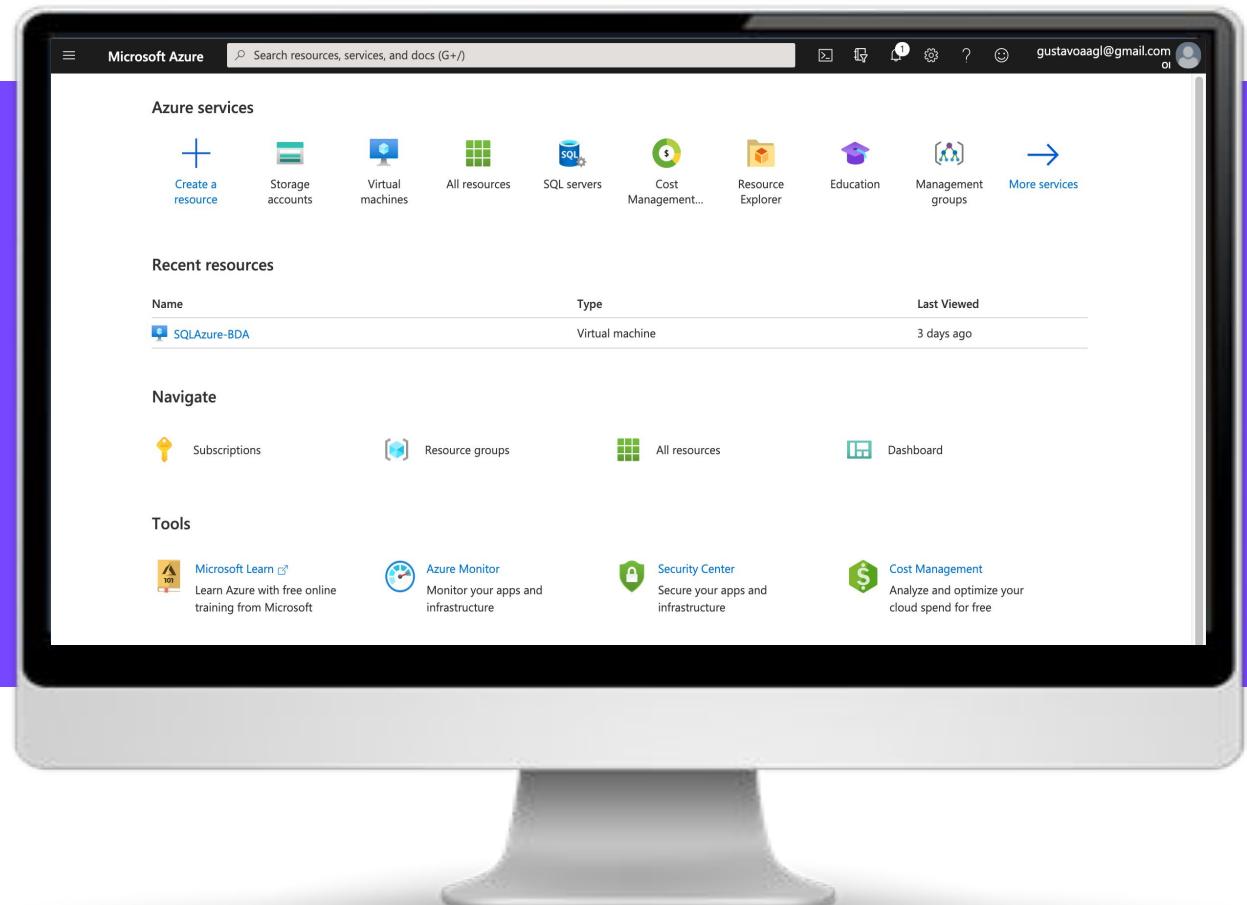


- Criando uma Storage Account

Criando uma Storage Account

IGTI

Demo



Próxima Aula



- Azure Files.

Armazenamento de Dados

CAPÍTULO 2. AULA 2.3. AZURE FILES

PROF. GUSTAVO AGUILAR

Nesta Aula



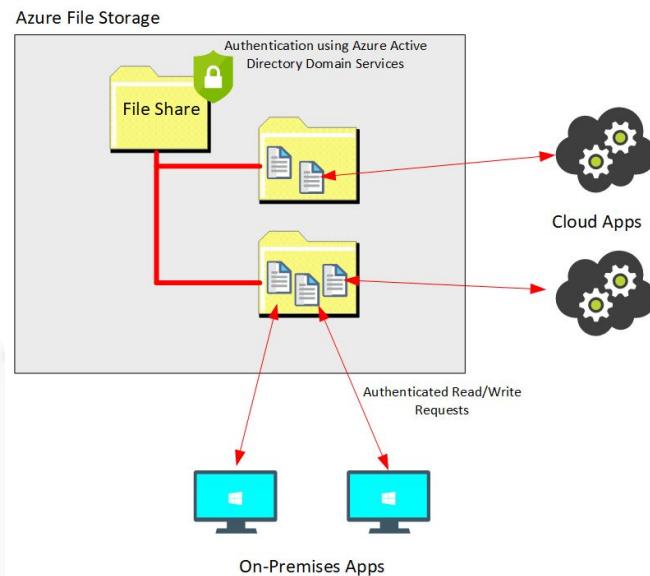
- Azure Files.
- Principais Características.

Azure Files



IGTI

- Serviço para criação de files shares (compartilhamento de arquivos) totalmente gerenciado;
- Podem ser acessados por meio dos protocolos Server Message Block (SMB) e Network File System (NFS);

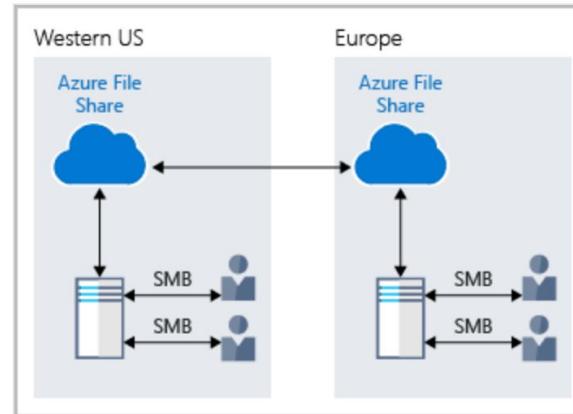


Azure Files

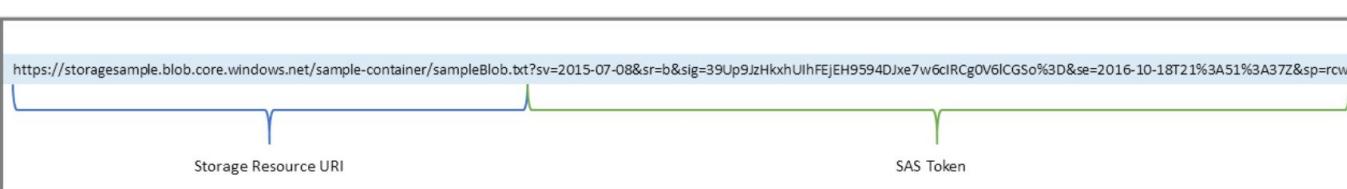


IGTI

- Podem ser usados para compartilhar dados entre duas localizações geográficas;



- Podem ser montados simultaneamente em implantações de Windows, Linux e macOS, na nuvem ou on-premise;
- Podem ser acessados também via URL;



Azure Files



IGTI

- 100 TB de dados por conta;
- 1 TiB (tebibytes) por arquivo (+- 1 TB);
- Permite gerenciamento individualizado da cota de armazenamento para cada file share;
- Até 2.000 conexões* simultâneas por arquivo compartilhado;
- Duas camadas de desempenho:
 - **Standard:** disco comum (até 300 MB/s).
 - **Premium:** disco SSD.

Próxima Aula



- Criando e Usando Azure Files.

Armazenamento de Dados

CAPÍTULO 2. AULA 2.4. CRIANDO E USANDO AZURE FILES

PROF. GUSTAVO AGUILAR

Nesta Aula



- ❑ Criando e Usando Azure Files.

Criando e Usando Azure Files



 Demo

The screenshot shows the Microsoft Azure portal interface. At the top, there's a search bar and a user profile. Below the header, the "Azure services" section features a "Create a resource" button and links for Storage accounts, Virtual machines, All resources, SQL servers, Cost Management, Resource Explorer, Education, Management groups, and More services. The "Recent resources" section lists a single item: "SQLAzure-BDA" (Virtual machine) last viewed 3 days ago. The "Navigate" section includes links for Subscriptions, Resource groups, All resources, and Dashboard. The "Tools" section includes links for Microsoft Learn, Azure Monitor, Security Center, and Cost Management.

Próxima Aula



- Ingestão de Dados.

Armazenamento de Dados

CAPÍTULO 2. AULA 2.5. INGESTÃO DE DADOS

PROF. GUSTAVO AGUILAR

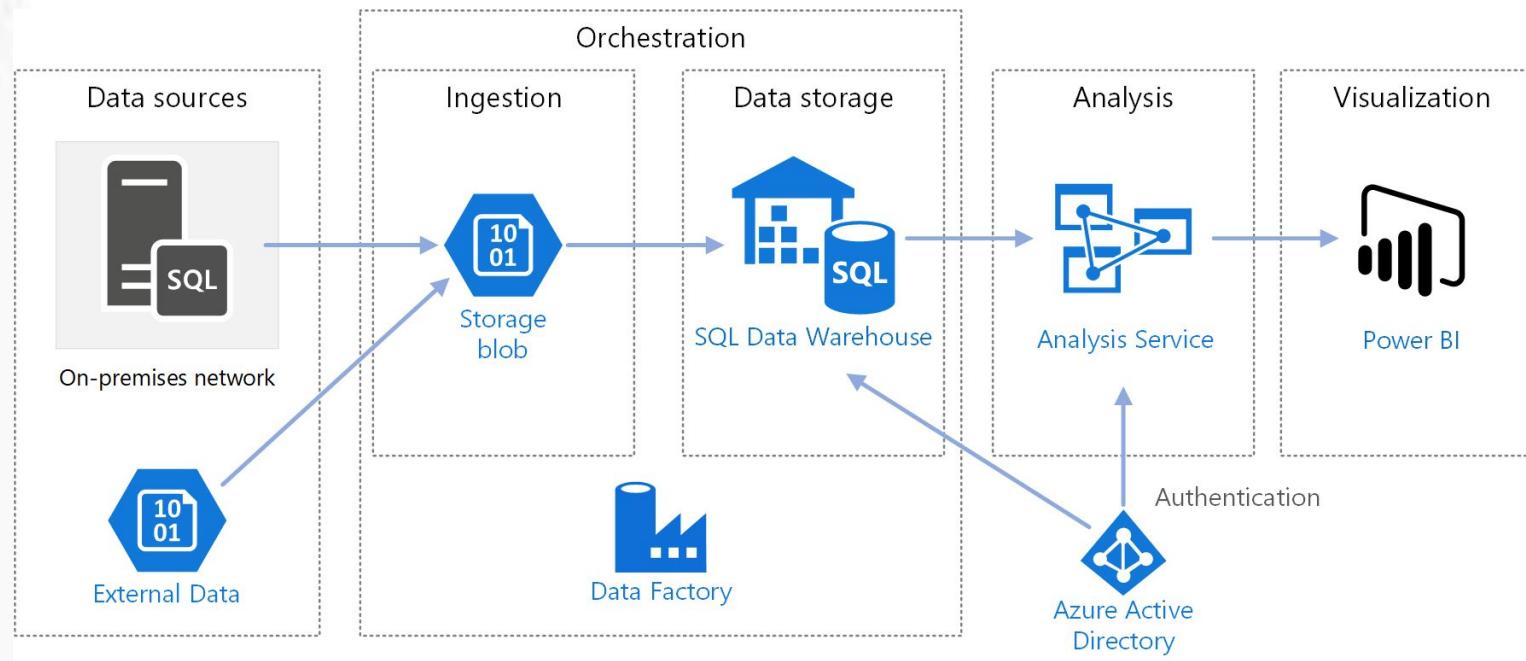
Nesta Aula



- Ingestão de Dados
- Azure Storage Explorer

Ingestão de Dados

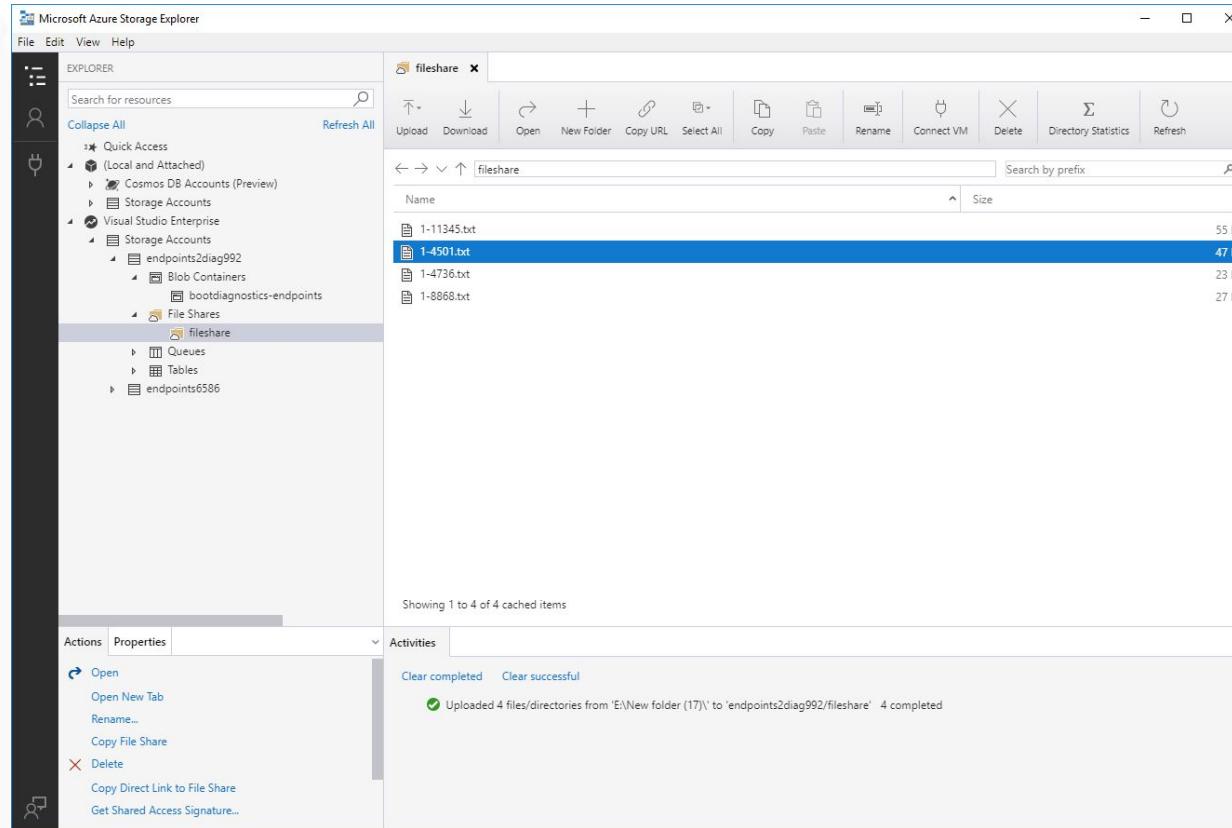
- Data Factory



Ingestão de Dados



- Azure Storage Explorer: www.storageexplorer.com



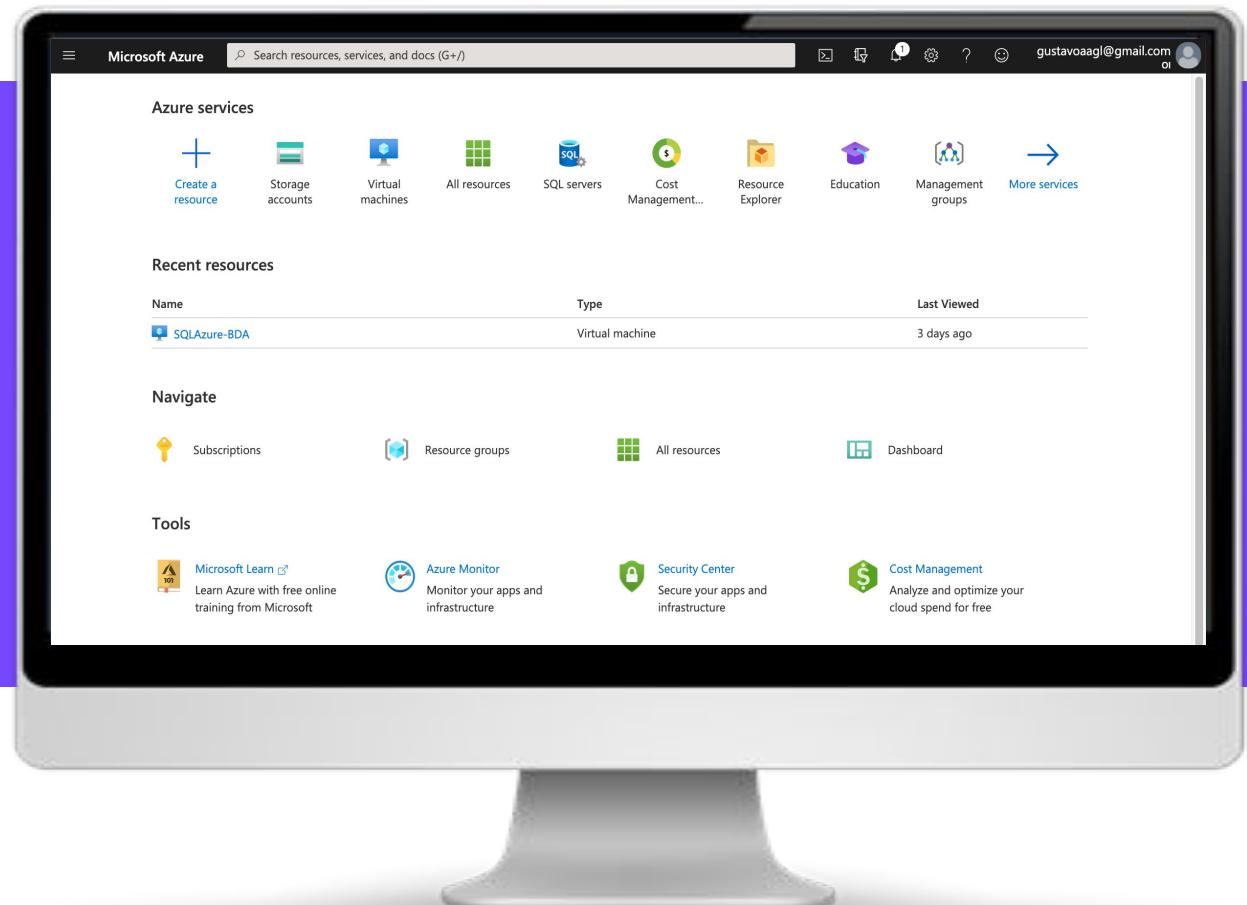
Ingestão de Dados

- **AzCopy;**
- **PowerShell;**
- **Visual Studio;**
- Arquivos acima de 2 GB □ PowerShell ou Visual Studio;
- AzCopy □ tamanho máximo de arquivo de 1 TB
 - Automaticamente dividido em vários arquivos se o arquivo exceder 200 GB.

Storage Explorer

IGTI

Demo



Próxima Aula



- Azure Table Storage.

Armazenamento de Dados

CAPÍTULO 2. AULA 2.6. AZURE TABLE STORAGE

PROF. GUSTAVO AGUILAR

Nesta Aula



- Azure Table Storage
- Vantagens
- Desvantagens

Azure Table Storage

- Armazenamento para o modelo de chave-valor;
 - Dados de um item (linha) são armazenados como um conjunto de colunas (campos) e o item é identificado por uma chave exclusiva □ 1 MB por linha;
 - Quantidade de colunas em cada linha pode ser diferente □ até 252;
 - Permite armazenamento de dados semiestruturados □ numérico / string / binário.

Key	Value (fields)			
AA	Data for AA	
BB	Data for BB	
CC	Data for CC	
...				
ZZ	Data for ZZ	

Key (Customer ID)	Value (Customer Data)						
C1	AAAAA	BBB	101 Block Street	YY	999	888	
C2	MM	NN	21 A Street	5 B Avenue			
C3	DDD	EEE	FFF	111	222	66	C Road

Azure Table Storage

Query:
Find the data for Device 2 recorded between
13:04:00 AM and 13:04:30 AM

Partition Key	Row Key	Value
Device 1	11:01:55 AM	Device data
	11:02:13 AM	Device data
	11:08:27 AM	Device data

Partition Key	Row Key	Value
Device 2	13:03:21 AM	Device data
	13:03:23 AM	Device data
	13:04:24 AM	Device data
	13:04:28 AM	Device data
	13:04:29 AM	Device data

Vantagens

- É mais simples de escalar:
 - Gasta o mesmo tempo para inserir dados em uma tabela vazia ou em uma tabela com bilhões de entradas;
 - Pode escalar até 500 TB de dados;
 - Gerencia automaticamente as partições em uma tabela e aloca o armazenamento conforme necessário.
- Alta disponibilidade:
 - Os dados de cada tabela são replicados 3 vezes em uma região do Azure;
 - Permite redundância geográfica ☐ + 3 réplicas e + custo.
- Especificando as chaves de partição e linha como critérios de consulta, a recuperação de dados é a mais rápida*.

Desvantagens

- A consistência deve ser levada em consideração, pois as atualizações transacionais em várias entidades não são garantidas;
- Não há integridade referencial ☐ quaisquer relações entre as linhas precisam ser mantidas externamente à tabela (no código);
- Consultas com base em campos não-chave podem resultar em varreduras completas da tabela.

Próxima Aula



- Criando e Usando Azure Table Storage.

Armazenamento de Dados

CAPÍTULO 2. AULA 2.7. CRIANDO E USANDO AZURE TABLE STORAGE

PROF. GUSTAVO AGUILAR

Nesta Aula

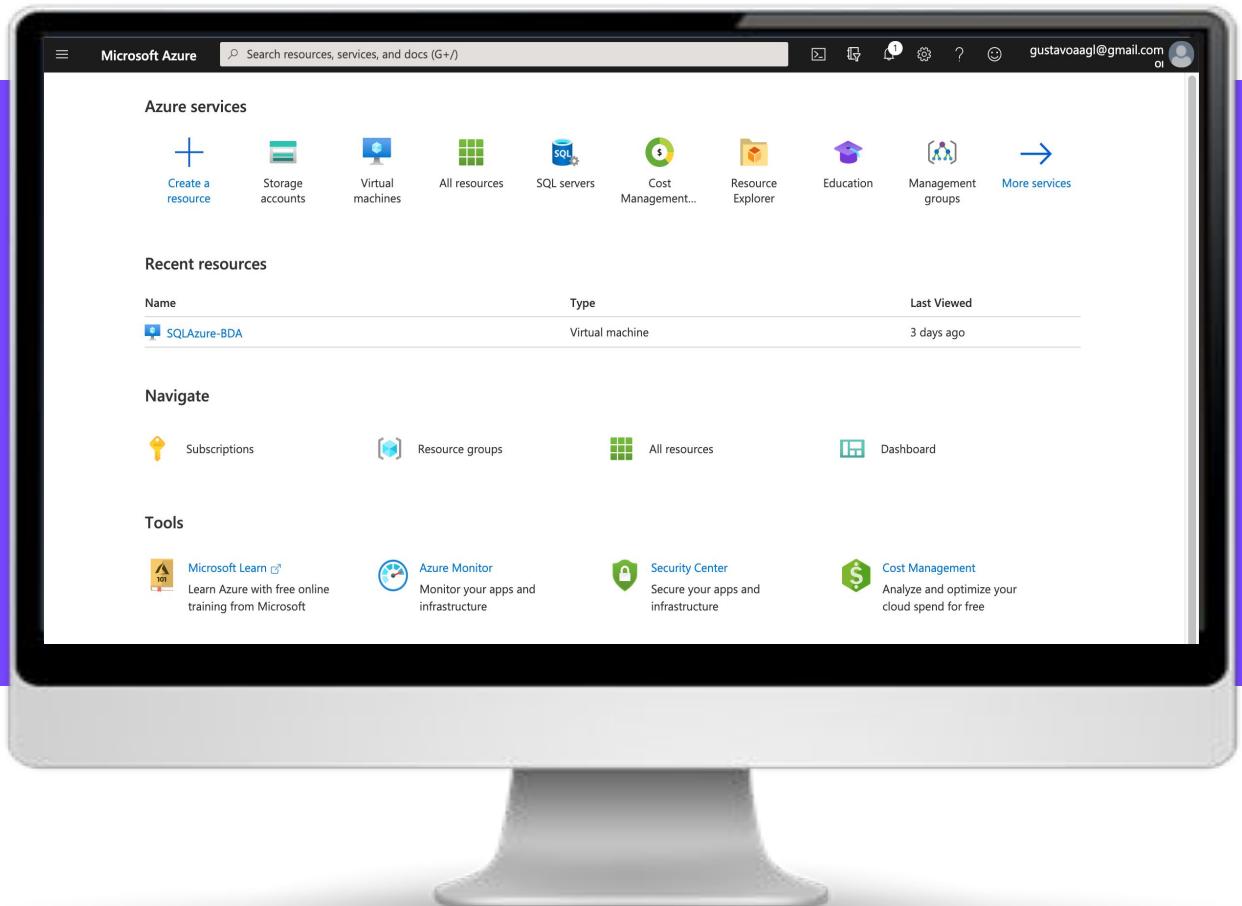


- ❑ Criando e Usando Azure Table Storage

Azure Table Storage



 Demo



The Microsoft Azure portal homepage is displayed on a monitor. The top navigation bar includes a search bar, user profile, and various icons. Below the header, the "Azure services" section features a grid of icons for Create a resource, Storage accounts, Virtual machines, All resources, SQL servers, Cost Management..., Resource Explorer, Education, Management groups, and More services. The "Recent resources" section shows a single entry: "SQLAzure-BDA" (Virtual machine) last viewed 3 days ago. The "Navigate" section provides links to Subscriptions, Resource groups, All resources, and Dashboard. The "Tools" section includes links to Microsoft Learn, Azure Monitor, Security Center, and Cost Management.

Próxima Aula



- Azure Blob Storage.

Armazenamento de Dados

CAPÍTULO 2. AULA 2.8. AZURE BLOB STORAGE

PROF. GUSTAVO AGUILAR

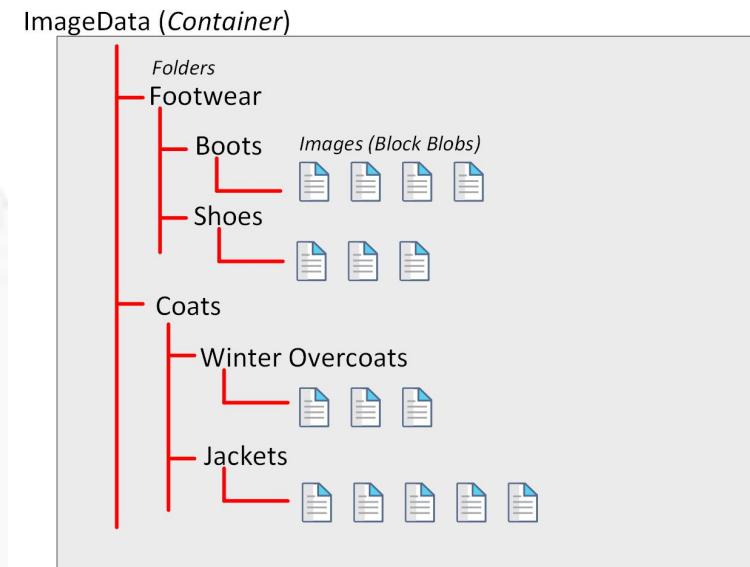
Nesta Aula



- ❑ Azure Blob Storage
- ❑ Tipos de Azure Blob Storage
- ❑ Cenários de Uso

Azure Blob Storage

- Serviço que permite armazenar grandes quantidades de dados não estruturados (**blobs**), na nuvem;
- Conta de armazenamento □ Container □ Blob;
- Pode-se organizar blobs em uma hierarquia de pastas,



Tipos de Azure Blob Storage

- **Blobs de Página (*Page Blobs*)**

- Organizado como uma coleção de páginas de 512 bytes de tamanho fixo;
- Um blob de página pode conter até 8 TB de dados;
- Otimizado para suportar operações aleatórias de leitura e gravação;
- Pode buscar e armazenar dados para uma única página, se necessário;
- O Azure usa blobs de página para implementar o armazenamento em disco virtual das máquinas virtuais.

Tipos de Azure Blob Storage

▪ **Blobs de Blocos (*Block Blobs*)**

- Conjunto de blocos □ cada bloco pode variar em tamanho, até 100 MB;
- Um blob de bloco pode conter até 50.000 blocos □ mais de 4,7 TB;
- O bloco é a menor quantidade de dados que pode ser lida ou gravada como uma unidade individual;
- Mais usados para armazenar objetos binários grandes que mudam com pouca frequência.

Tipos de Azure Blob Storage

- **Append Blobs (“Blob de Acréscimo”)**

- É um blob de bloco otimizado para suportar operações de append (acréscimo);
- Só permite adicionar blocos ao final □ atualizar ou excluir blocos existentes não é suportado;
- Cada bloco pode variar em tamanho, até 4 MB.
- O tamanho máximo de um append blob é pouco mais de 195 GB.

Cenários de Uso

- Envio de imagens ou documentos diretamente para um navegador;
- Armazenamento de arquivos para acesso distribuído;
- Streaming de vídeo e áudio;
- Armazenamento de dados para backup e restauração, recuperação de desastres e arquivamento;
 - Permite controle de versão.
- Armazenamento de dados para análise por um serviço local ou hospedado pelo Azure;
- Armazenamento para um Data Lake.

Próxima Aula



- Criando e Usando Azure Blob Storage.

Armazenamento de Dados

CAPÍTULO 2. AULA 2.9. CRIANDO E USANDO AZURE BLOB STORAGE

PROF. GUSTAVO AGUILAR

Nesta Aula

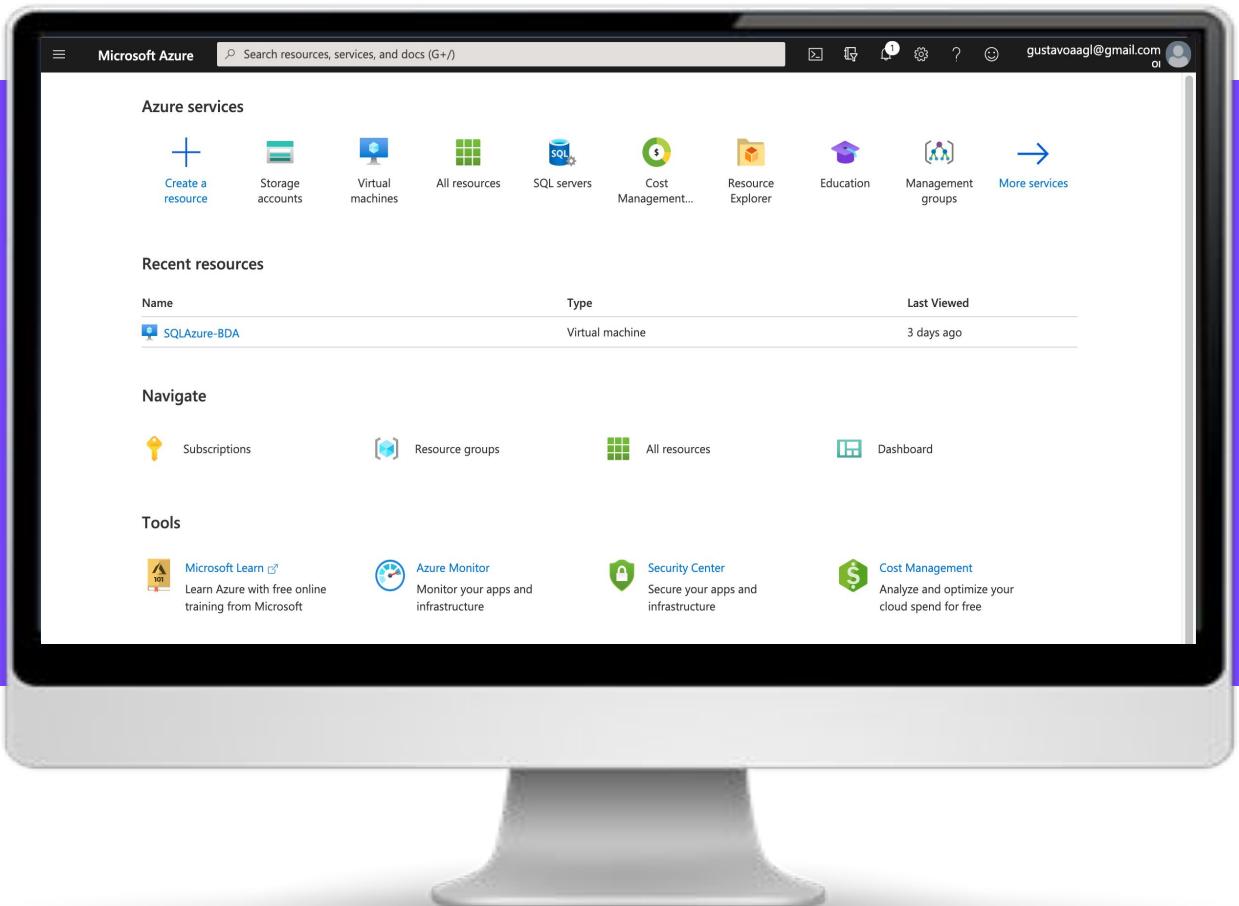


- ❑ Criando e Usando Azure Blob Storage

Azure Blob Storage



 Demo



The screenshot shows the Microsoft Azure portal interface. At the top, there's a search bar and a user profile. Below the header, the "Azure services" section features a "Create a resource" button and links for Storage accounts, Virtual machines, All resources, SQL servers, Cost Management, Resource Explorer, Education, Management groups, and More services. The "Recent resources" section lists a single item: "SQLAzure-BDA" (Virtual machine) last viewed 3 days ago. The "Navigate" section includes links for Subscriptions, Resource groups, All resources, and Dashboard. The "Tools" section includes links for Microsoft Learn, Azure Monitor, Security Center, and Cost Management.

Próxima Aula



- Data Lake Storage Gen2.

Armazenamento de Dados

CAPÍTULO 2. AULA 2.10. DATA LAKE STORAGE GEN2

PROF. GUSTAVO AGUILAR

Nesta Aula



- Data Lake Storage Gen2
- Características

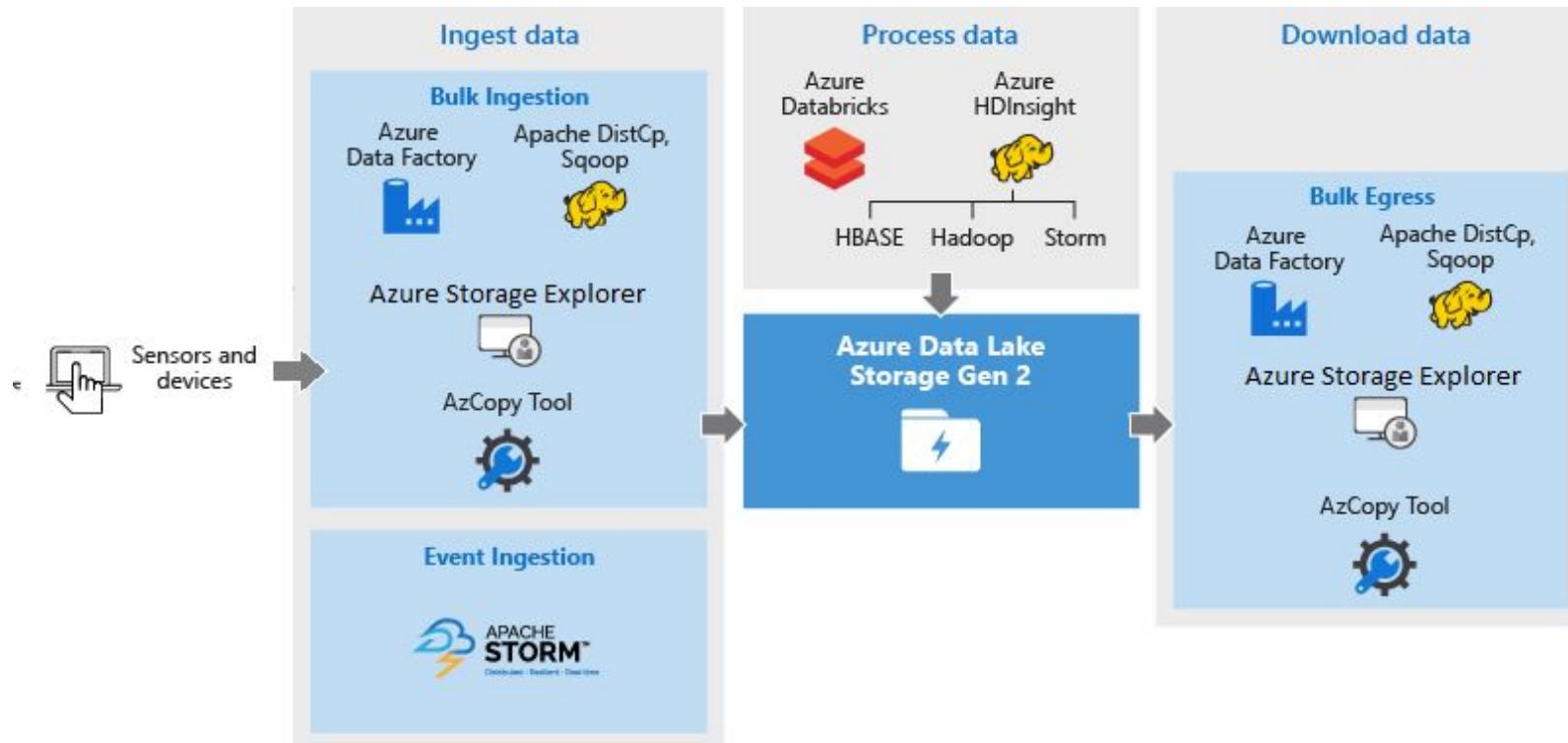
Data Lake Storage Gen2

- Repositório de dados compatível com HDFS
 - HDFS (Hadoop Distributed File System) ☐ Sistema de Arquivos Distribuído;
 - Pode armazenar qualquer tipo de dados.
- Gen2 combina os serviços de armazenamento do Gen1 com os benefícios do Azure Blob Storage;
- Desempenho ajustado para o processamento de soluções de análise de big data.



Data Lake Storage Gen2

IGTI



Data Lake Storage Gen2



Microsoft Azure Search resources, services, and docs (G+ /)

Home > Storage accounts >

Create storage account

Basics Networking Data protection Advanced Tags Review + create

Security

Secure transfer required Disabled Enabled

Allow shared key access Disabled Enabled

Minimum TLS version Version 1.2

Infrastructure encryption Disabled Enabled

Sign up is currently required to enable infrastructure encryption on a per-subscription basis. [Sign up for infrastructure encryption](#)

Blob storage

Allow Blob public access Disabled Enabled

Blob access tier (default) Cool Hot

NFS v3 Disabled Enabled

Sign up is currently required to utilize the NFS v3 feature on a per-subscription basis. [Sign up for NFS v3](#)

Data Lake Storage Gen2

Hierarchical namespace Disabled Enabled

Instance details

The default deployment model is Resource Manager, which supports the latest Azure features. You may choose to deploy using the classic deployment model instead. [Choose classic deployment model](#)

Storage account name * accountstoragebtc

Location * StorageV2 (general purpose v2)

Performance Storage (general purpose v1)

Account kind BlobStorage

StorageV2 (general purpose v2)

accountstoragebtc

StorageV2 (general purpose v2)

Storage (general purpose v1)

BlobStorage

StorageV2 (general purpose v2)

Próxima Aula



- Criando e Usando um Data Lake Storage Gen2.

Armazenamento de Dados

CAPÍTULO 2. AULA 2.11. CRIANDO E USANDO UM DATA LAKE STORAGE GEN2

PROF. GUSTAVO AGUILAR

Nesta Aula



- Criando um Data Lake Storage Gen2
- Usando um Data Lake Storage Gen2

Data Lake Storage Gen2



Demo

The screenshot shows the Microsoft Azure portal interface. At the top, there's a search bar labeled "Search resources, services, and docs (G+I)". On the right side of the header, the user's email "gustavoagl@gmail.com" and a profile icon are visible. Below the header, the "Azure services" section is displayed, featuring a "Create a resource" button and links to "Storage accounts", "Virtual machines", "All resources", "SQL servers", "Cost Management...", "Resource Explorer", "Education", "Management groups", and "More services". The "Recent resources" section lists a single item: "SQLAzure-BDA" (Type: Virtual machine, Last Viewed: 3 days ago). The "Navigate" section includes links to "Subscriptions", "Resource groups", "All resources", and "Dashboard". The "Tools" section features links to "Microsoft Learn", "Azure Monitor", "Security Center", and "Cost Management".

Próxima Aula



- Capítulo 3 - Armazenamento de Dados Relacionais no Azure.

Serviços de Armazenamento, Banco de Dados e Analytics

Capítulo 3. Armazenamento de Dados Relacionais no Azure

PROF. GUSTAVO AGUILAR

Armazenamento de Dados Relacionais no Azure

CAPÍTULO 3. AULA 3.1. BANCOS DE DADOS RELACIONAIS EM IAAS

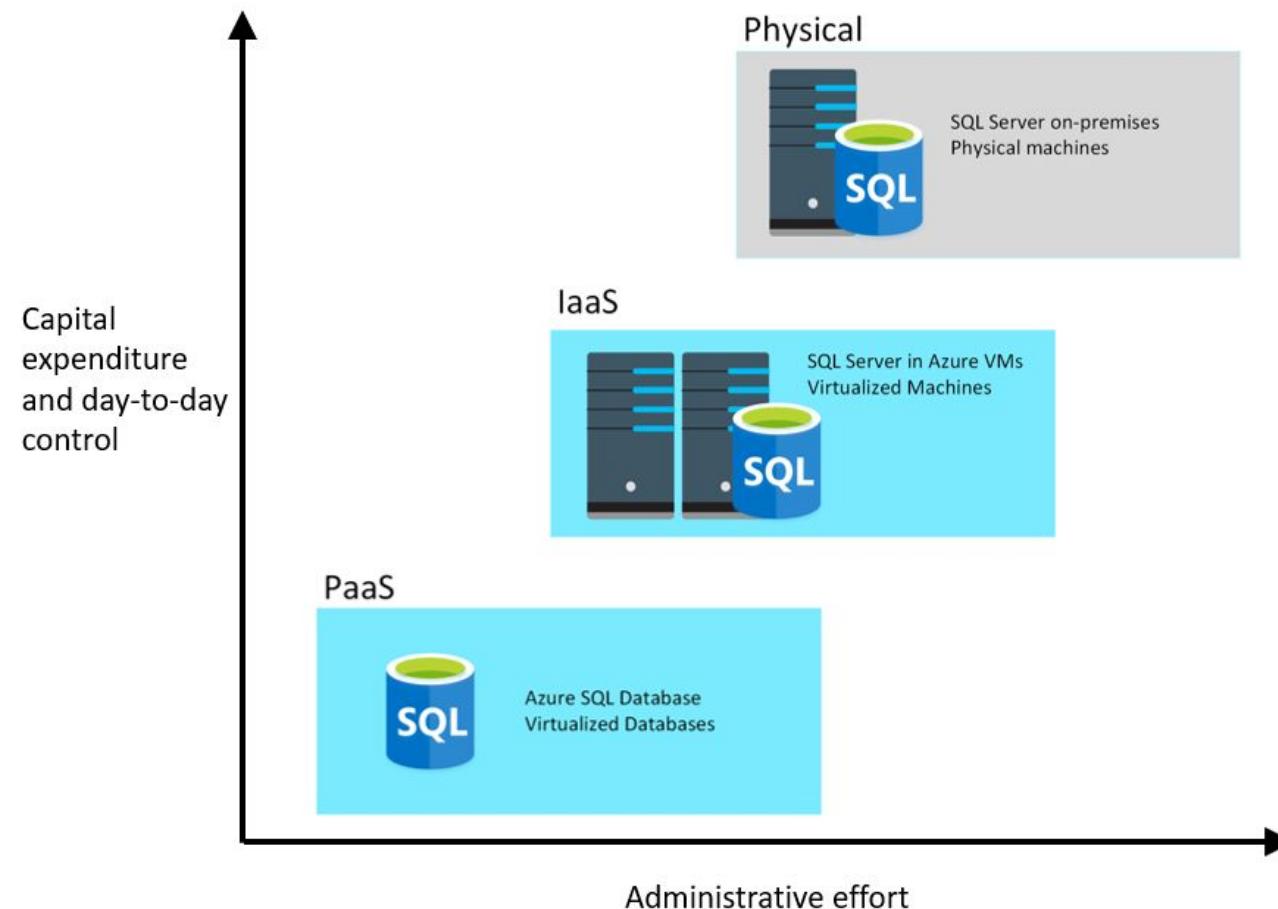
PROF. GUSTAVO AGUILAR

Nesta Aula



- Bancos de Dados Relacionais em IaaS
- SQL Server do Azure Marketplace

Banco de Dados em IaaS



SQL Server do Azure Marketplace



Microsoft Azure Search resources, services, and docs (G+/)

Home > New >

SQL Server 2019 on Windows Server 2019

Microsoft



SQL Server 2019 on Windows Server 2019

Microsoft

Select a plan SQL Server 2019 Web on Windows S... ▾

Create

Start with a pre-set configuration

Save for later

Deploy with Resource Manager (change to Classic)

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

IGTI

SQLAzure-BDA

Virtual machine

Search (Cmd+/)

Connect Start Restart Stop Move Delete Refresh Share to mobile

Overview Activity log Access control (IAM) Tags Diagnose and solve problems

Properties Monitoring Capabilities Recommendations Tutorials

Tags (change) Click here to add tags

Key Metrics See all metrics

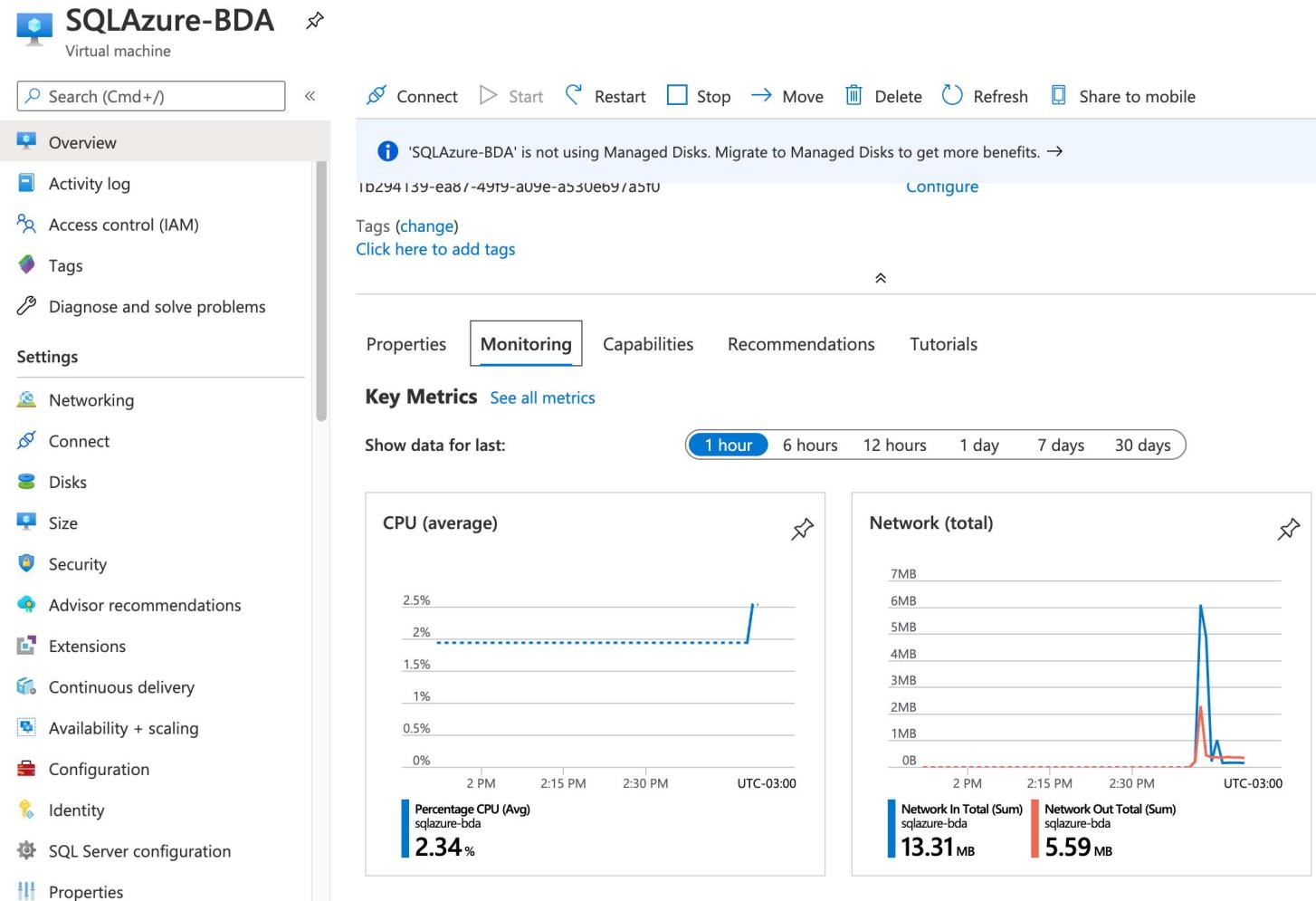
Show data for last: 1 hour 6 hours 12 hours 1 day 7 days 30 days

CPU (average)

Percentage CPU (Avg) sqlazure-bda 2.34%

Network (total)

Network In Total (Sum) sqlazure-bda 13.31 MB Network Out Total (Sum) sqlazure-bda 5.59 MB



Próxima Aula



- Demo Bancos de Dados Relacionais em IaaS.

Armazenamento de Dados Relacionais no Azure

CAPÍTULO 3. AULA 3.2. DEMO BANCOS DE DADOS RELACIONAIS EM IAAS

PROF. GUSTAVO AGUILAR

Nesta Aula

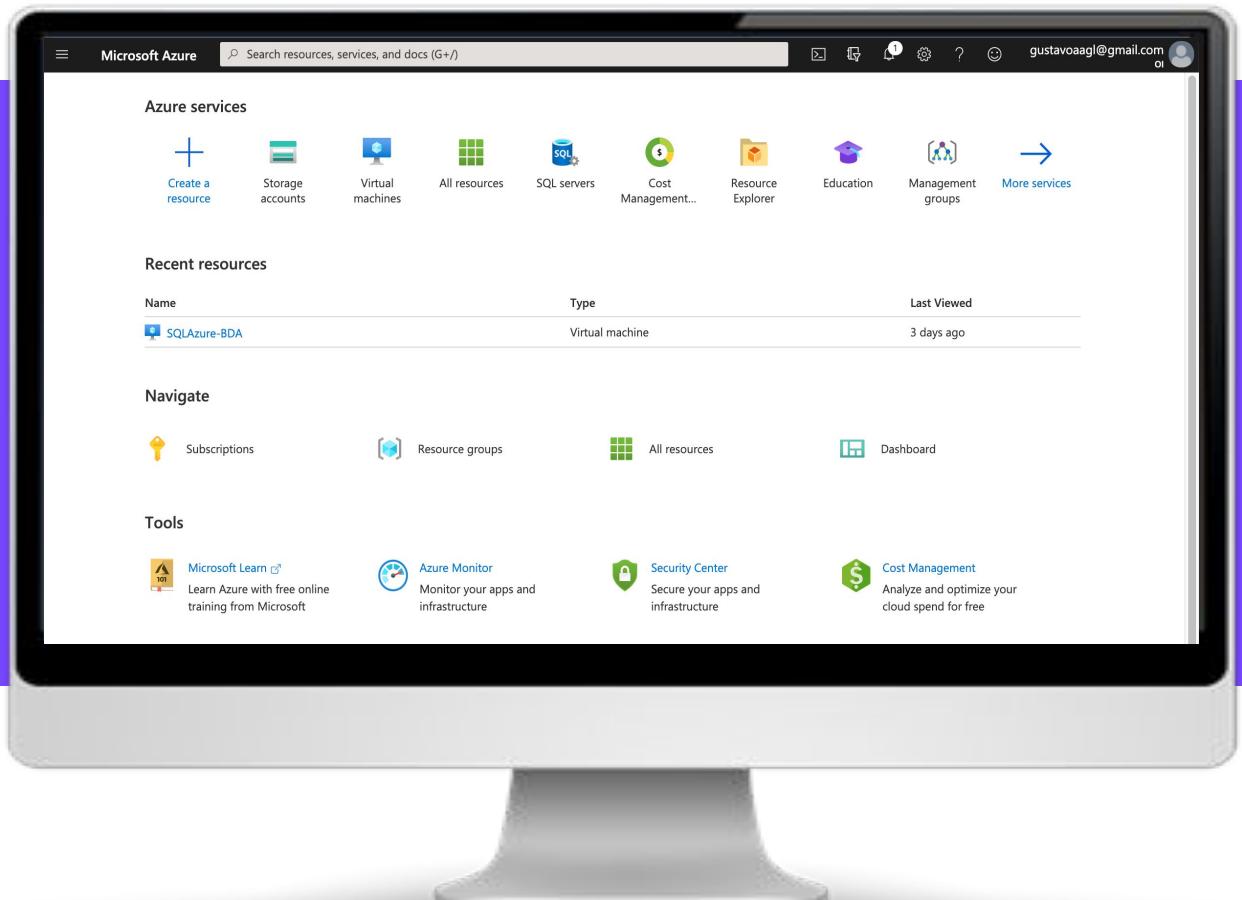


- Criar uma infraestrutura para banco de dados
- Instalar banco de dados em IaaS
- Usar banco de dados em IaaS

Bancos de Dados em IaaS



 Demo



The screenshot shows the Microsoft Azure portal's home dashboard. At the top, there's a search bar labeled "Search resources, services, and docs (G+/-)". Below the search bar is a navigation bar with icons for "Create a resource", "Storage accounts", "Virtual machines", "All resources", "SQL servers", "Cost Management...", "Resource Explorer", "Education", "Management groups", and "More services". The main area is titled "Recent resources" and lists a single item: "SQLAzure-BDA" (Virtual machine) last viewed 3 days ago. Below this is a "Navigate" section with links to "Subscriptions", "Resource groups", "All resources", and "Dashboard". The bottom section is titled "Tools" and includes links to "Microsoft Learn", "Azure Monitor", "Security Center", and "Cost Management".

Próxima Aula



- Azure SQL Database Managed Instance.

Armazenamento de Dados Relacionais no Azure

CAPÍTULO 3. AULA 3.3. AZURE SQL DATABASE MANAGED INSTANCE

PROF. GUSTAVO AGUILAR

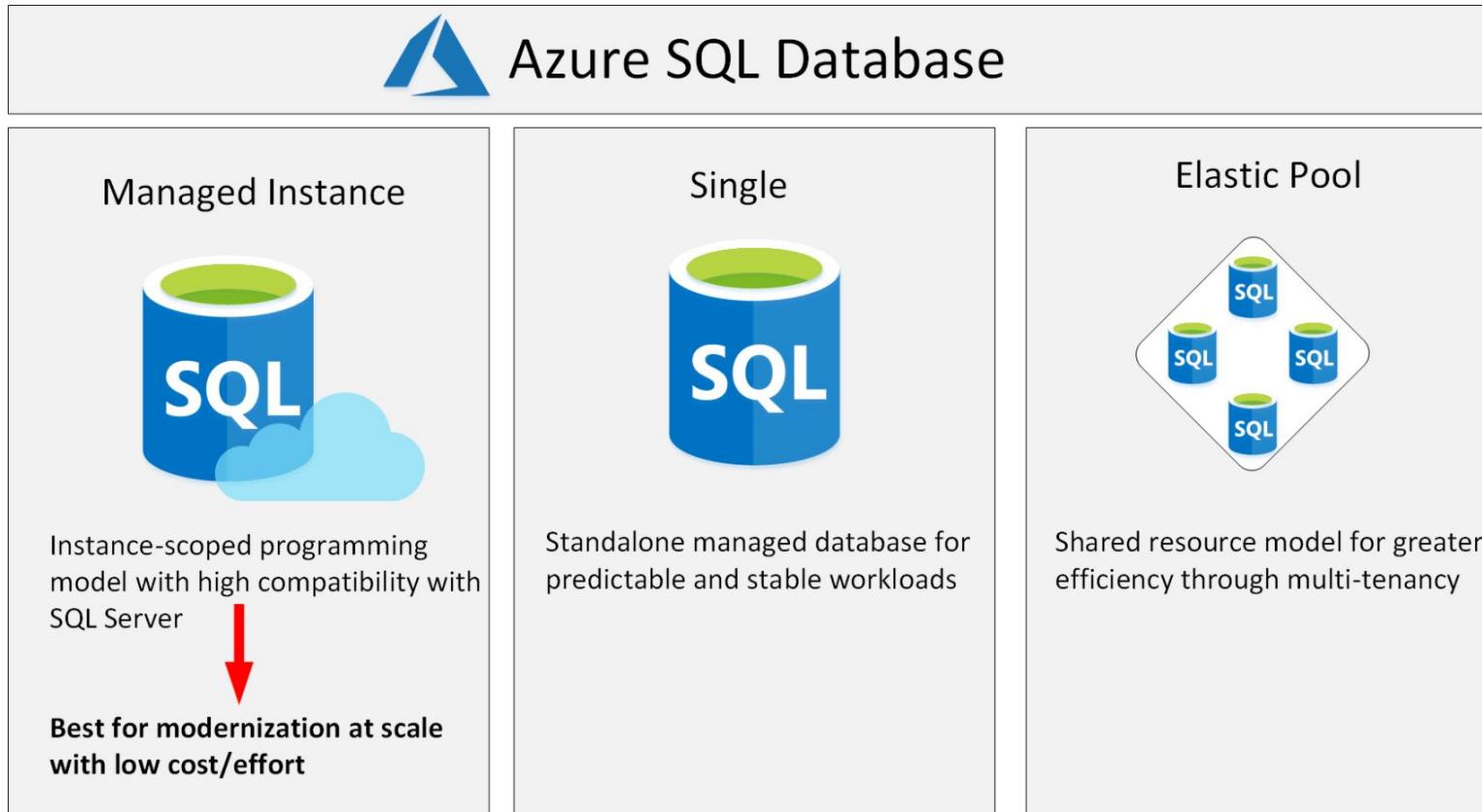
Nesta Aula



- SQL Managed Instance
- Ofertas de Instância Gerenciada
- Recursos da Instância Gerenciada

SQL Server Managed Instance

iG TI



Ofertas de Instância 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 atachado 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.

Próxima Aula



- Demo Azure SQL Database Managed Instance.

Armazenamento de Dados Relacionais no Azure

CAPÍTULO 3. AULA 3.4. DEMO AZURE SQL DATABASE MANAGED INSTANCE

PROF. GUSTAVO AGUILAR

Nesta Aula

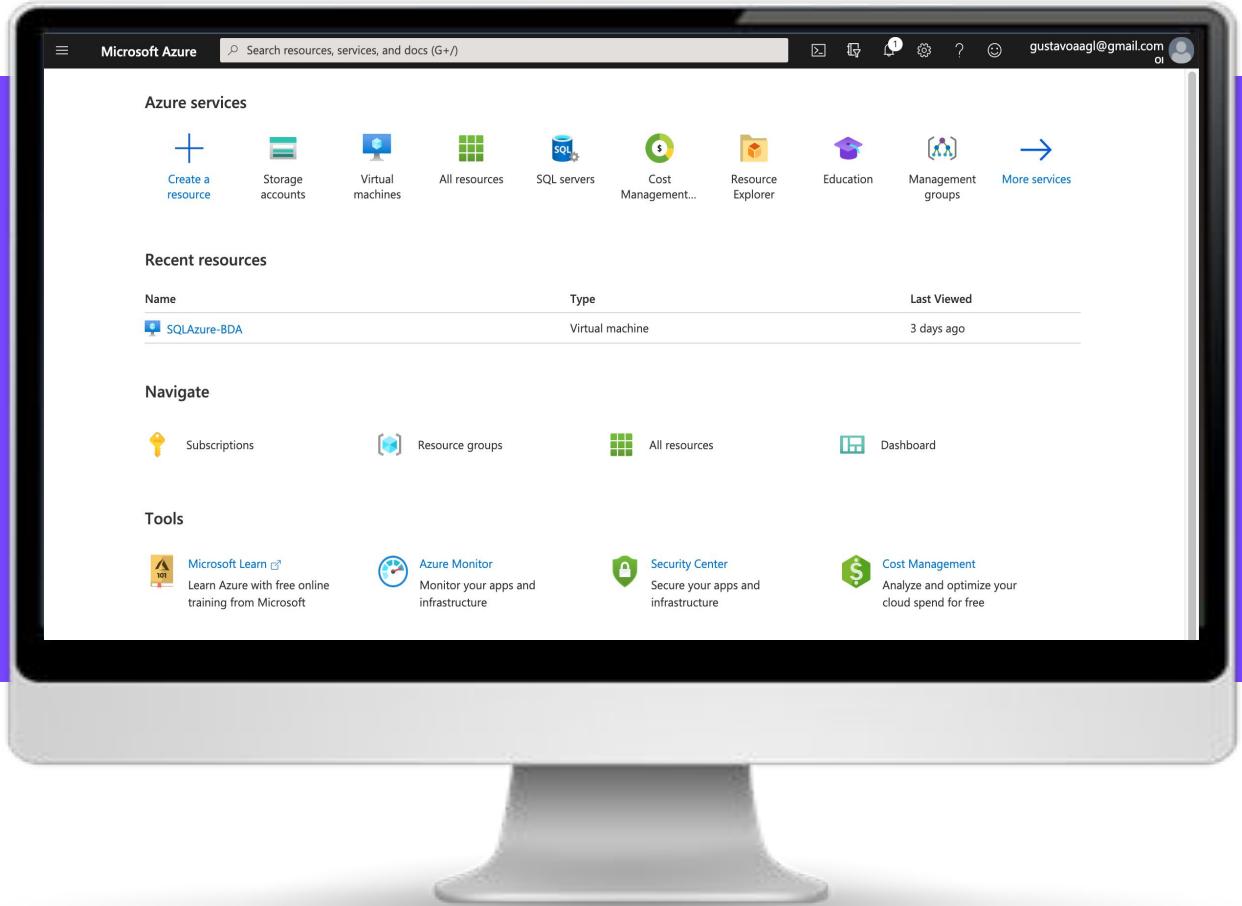


- ❑ Aprovisionando uma SQL Managed Instance

Aprovisionando uma SQL Managed Instance



 Demo



The screenshot shows the Microsoft Azure portal interface. At the top, there's a navigation bar with a search bar, user information, and a gear icon. Below the bar, the "Azure services" section features a grid of icons for "Create a resource", "Storage accounts", "Virtual machines", "All resources", "SQL servers", "Cost Management...", "Resource Explorer", "Education", "Management groups", and "More services". Under "Recent resources", there's a table with one item: "SQLAzure-BDA" (Virtual machine, last viewed 3 days ago). The "Navigate" section includes links for "Subscriptions", "Resource groups", "All resources", and "Dashboard". The "Tools" section contains links for "Microsoft Learn", "Azure Monitor", "Security Center", and "Cost Management".

Próxima Aula



- Azure SQL Database.

Armazenamento de Dados Relacionais no Azure

CAPÍTULO 3. AULA 3.5. AZURE SQL DATABASE

PROF. GUSTAVO AGUILAR

Nesta Aula



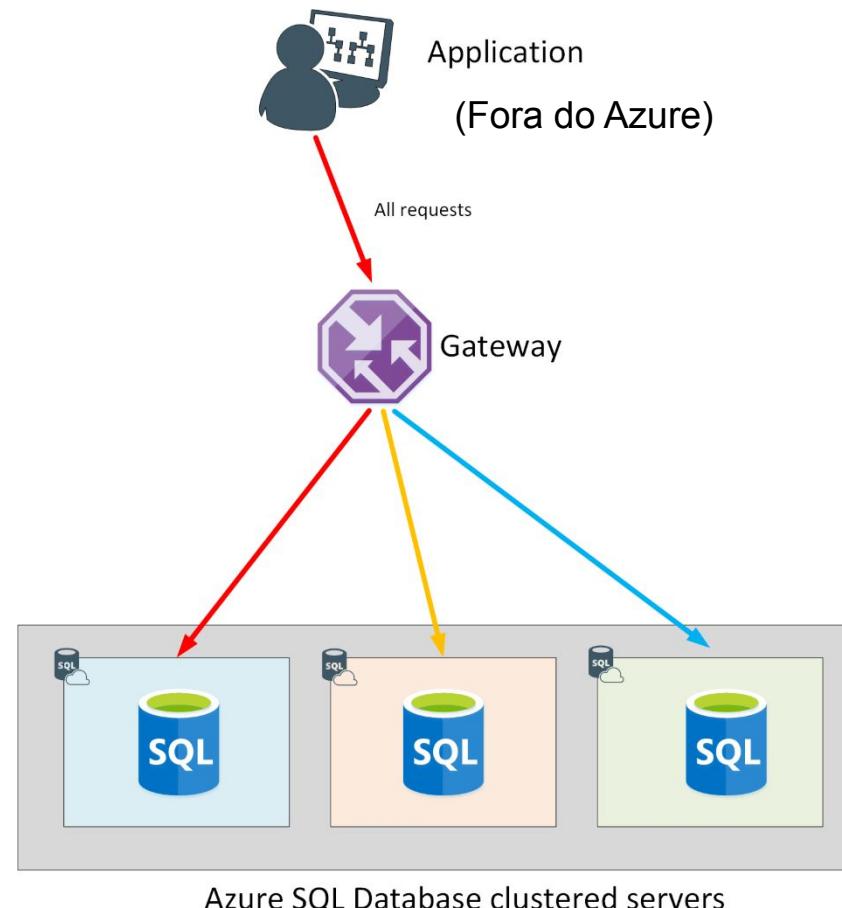
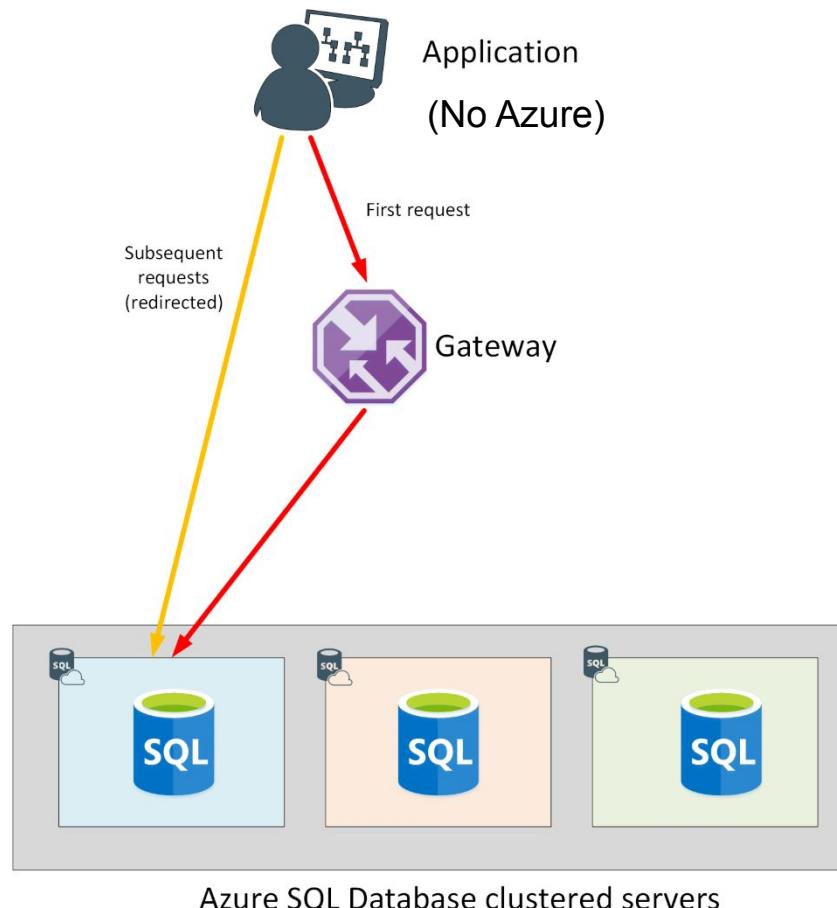
- Azure SQL Database
- Modelos de Implantação
- Modelos de Contratação

Azure SQL Database

- Oferta de banco de dados como serviço (PaaS) da Microsoft;
- Infra totalmente gerenciada pela Microsoft :
 - Gerenciamento do servidor e da instância SQL;
 - Backup e monitoração do banco de dados;
 - Patches do SO e do SQL;
 - Upgrades de versão do SQL.
- Usa uma topologia em cluster para fornecer alta disponibilidade;
 - Cada servidor e banco de dados são replicados de forma transparente;
 - Garantia que um servidor / banco de dados estará sempre acessível, mesmo no caso de um banco de dados ou servidor falhar.

Azure SQL Database

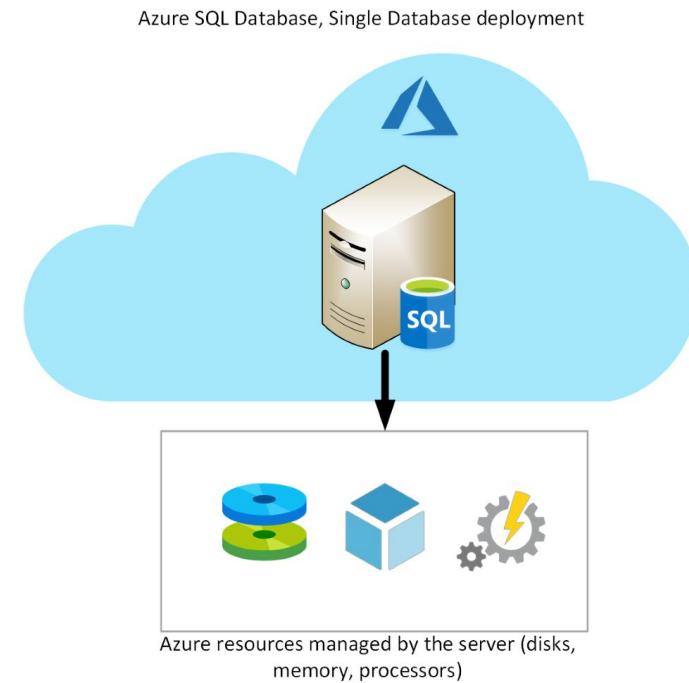
iGTTi



Modelos de Implantação

▪ Single Database:

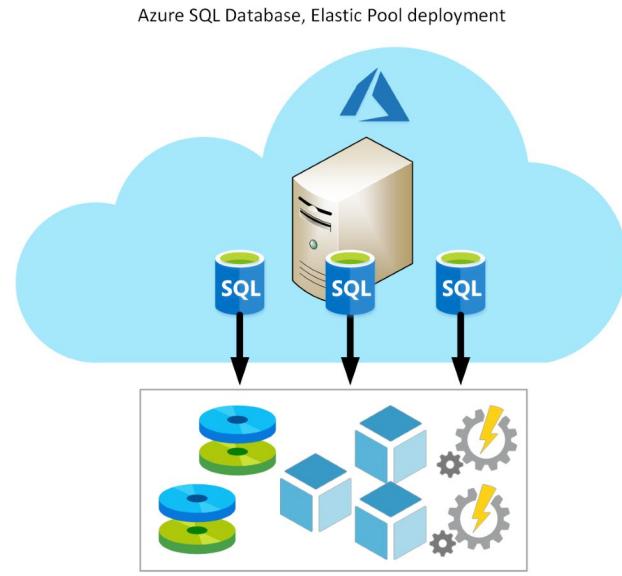
- Banco de dados isolado;
- Totalmente gerenciado pelo Azure;
- Recursos pré-alocados por banco;
- Pode escalar verticalmente (CPU/RAM/Disco);



Azure SQL Database

▪ Elastic Pool

- Coleção de single databases;
- Com um conjunto compartilhado de recursos (CPU / RAM / Disco);
- Útil para cenários com bancos de dados com requisitos de recursos que variam ao longo do tempo;
 - O Elastic Pool permite que os recursos disponíveis sejam usados sob demanda e liberados assim que o processamento for concluído;
 - Redução de custo comparando com Single Database, onde o recurso é alocado por banco de dados, sem compartilhamento entre eles.



Modelos de Contratação

- **Baseado em DTU (Database Transaction Units):** combinação de recursos de computação, memória e I/O.
 - **Basic:** 5 DTUs e tamanho máximo de 2 GB.
 - **Standard:** até 3.000 DTUs e máximo de 1 TB.
 - **Premium:**
 - ✓ Até 4.000 DTUs;
 - ✓ Máximo de 4 TB;
 - ✓ Permite leitura nas réplicas secundárias;
 - ✓ Permite redundância de zona.

Modelos de Contratação

- **Baseado em vCore:** permite selecionar a quantidade de CPU, memória RAM, velocidade do storage (IOPS) e tipo de hardware.

Available hardware configurations

Based on your workload requirements, select from available hardware configurations listed below.

Configuration	Description	Max vCores	Max memory	Max storage
Gen5	Balanced memory and compute	80	408 GB	4 TB
FSv2-series	Compute optimized	72	136 GB	4 TB
DC-series	Enables confidential computing	8	48 GB	4 TB
M-Series	Memory optimized	--	--	--

Modelos de Contratação

- Baseado em vCore:

- General Purpose

- ✓ 500 a 20.000 IPOS, com latência entre 2 e 10 ms;
 - ✓ Pode ser pré-alocado (**Provisioned**) ou com auto scaling (**Serverless**);
 - ✓ Permite também escalar verticalmente;
 - ✓ Permite redundância de zona.

- Hyperscale

- ✓ 500 a 204.800 IPOS, com latência entre 1 e 10 ms;
 - ✓ Para bancos de dados OLTP muito grandes: auto scaling até 100 TB;
 - ✓ Permite também escalar verticalmente;
 - ✓ Permite até 4 réplicas secundárias para escalar horizontalmente o workload de leitura e funcionar como DR.

Modelos de Contratação

- Baseado em vCore:

- **Business Critical / Premium:**

- ✓ Para aplicações OLTP com alto volume de transações;
 - ✓ Cenários que necessitam de baixa latência de I/O;
 - ✓ 5.000 a 204.800 IOPS e latência entre 1 e 2 ms;
 - ✓ Permite também escalar verticalmente;
 - ✓ Permite leitura nas réplicas secundárias;
 - ✓ Permite redundância de zona.

Próxima Aula



- Demo Azure SQL Database.

Armazenamento de Dados Relacionais no Azure

CAPÍTULO 3. AULA 3.6. DEMO AZURE SQL DATABASE

PROF. GUSTAVO AGUILAR

Nesta Aula

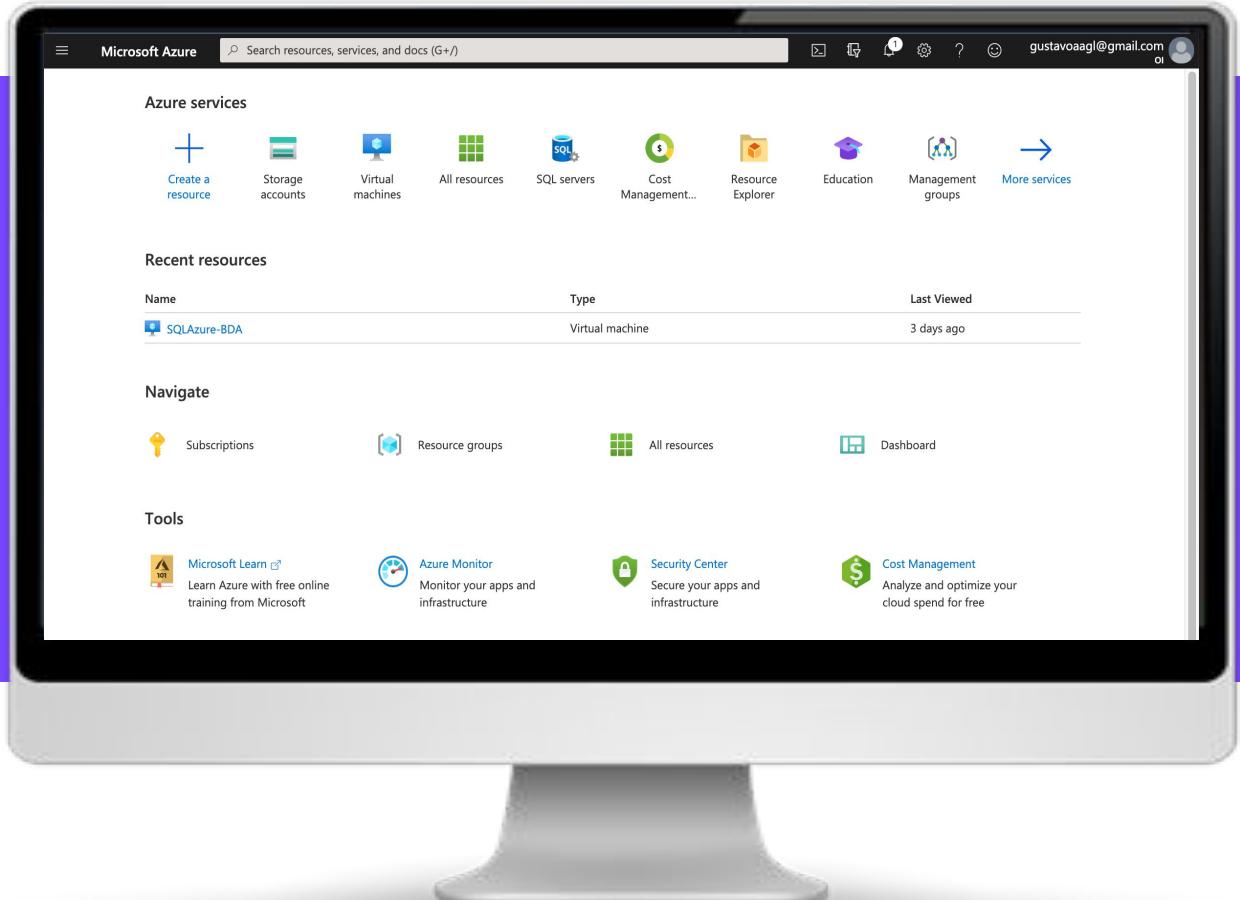


- ❑ Aprovisionando um Azure SQL Database

Aprovisionando um Azure SQL Database



 Demo



The screenshot shows the Microsoft Azure portal homepage. At the top, there's a search bar and a user profile. Below the header, the "Azure services" section features a "Create a resource" button and links for Storage accounts, Virtual machines, All resources, SQL servers, Cost Management, Resource Explorer, Education, Management groups, and More services. The "Recent resources" section lists a single item: "SQLAzure-BDA" (Virtual machine) last viewed 3 days ago. The "Navigate" section includes links for Subscriptions, Resource groups, All resources, and Dashboard. The "Tools" section includes links for Microsoft Learn, Azure Monitor, Security Center, and Cost Management.

Próxima Aula



- Bancos de Dados Open Source no Azure.

Armazenamento de Dados Relacionais no Azure

CAPÍTULO 3. AULA 3.7. BANCOS DE DADOS OPEN SOURCE NO AZURE

PROF. GUSTAVO AGUILAR

Nesta Aula

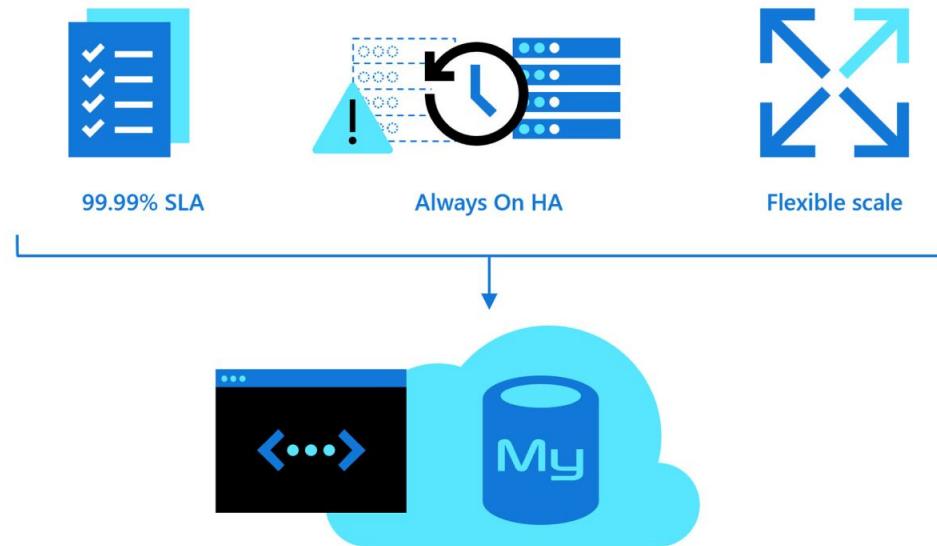


- Azure Database for MySQL
- Azure Database for MariaDB
- Azure Database for PostgreSQL

Bancos de Dados Open Source no Azure

IGTI

- 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.

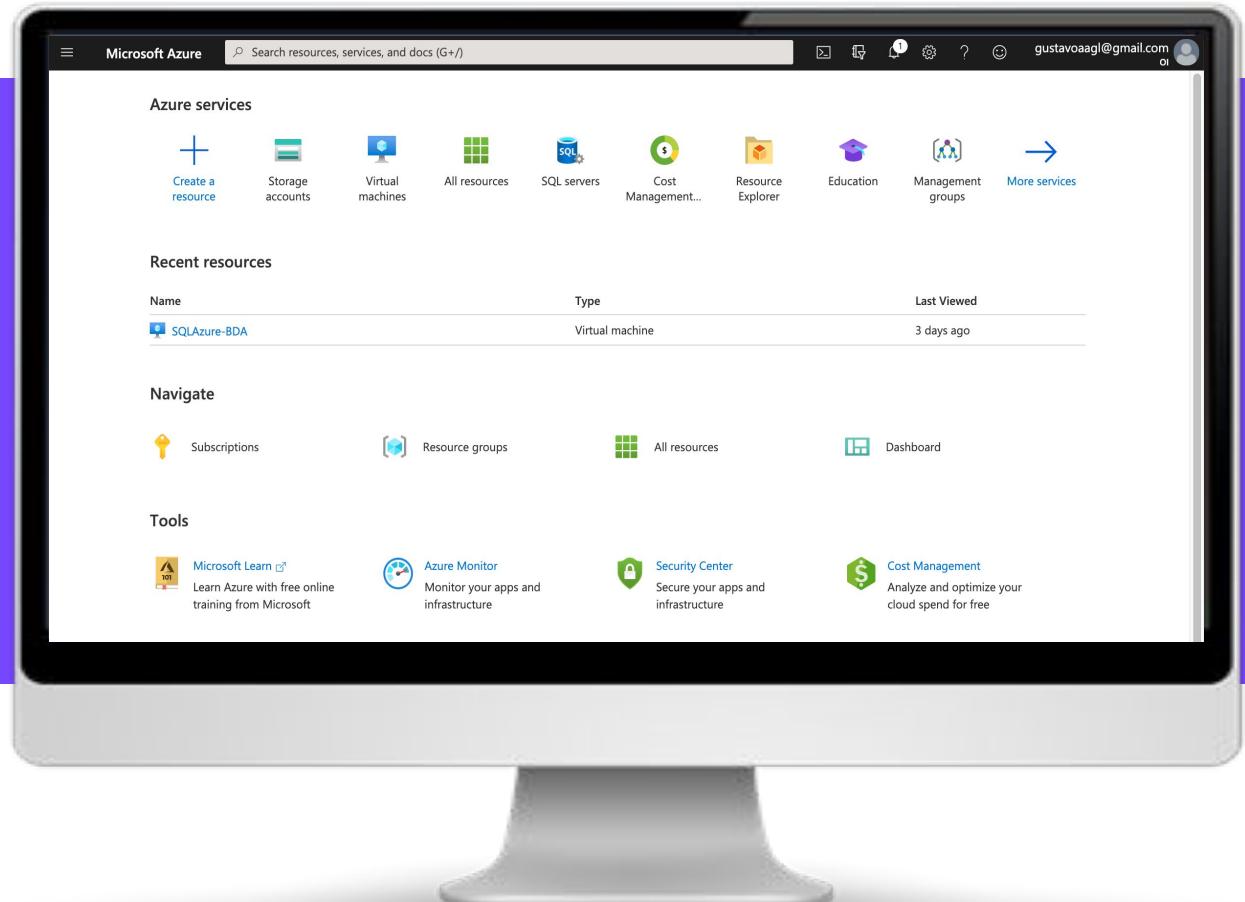


Criando Azure Database for MySQL e



for PostgreSQL

 **Demo**



Próxima Aula



- Capítulo 4 - Armazenamento de Dados Não Relacionais no Azure.

Serviços de Armazenamento, Banco de Dados e Analytics

Capítulo 4. Armazenamento de Dados Não Relacionais no Azure

PROF. GUSTAVO AGUILAR

Armazenamento de Dados Não Relacionais no Azure

CAPÍTULO 4. AULA 4.1. BANCOS DE DADOS NÃO RELACIONAIS NO AZURE

PROF. GUSTAVO AGUILAR

Nesta Aula

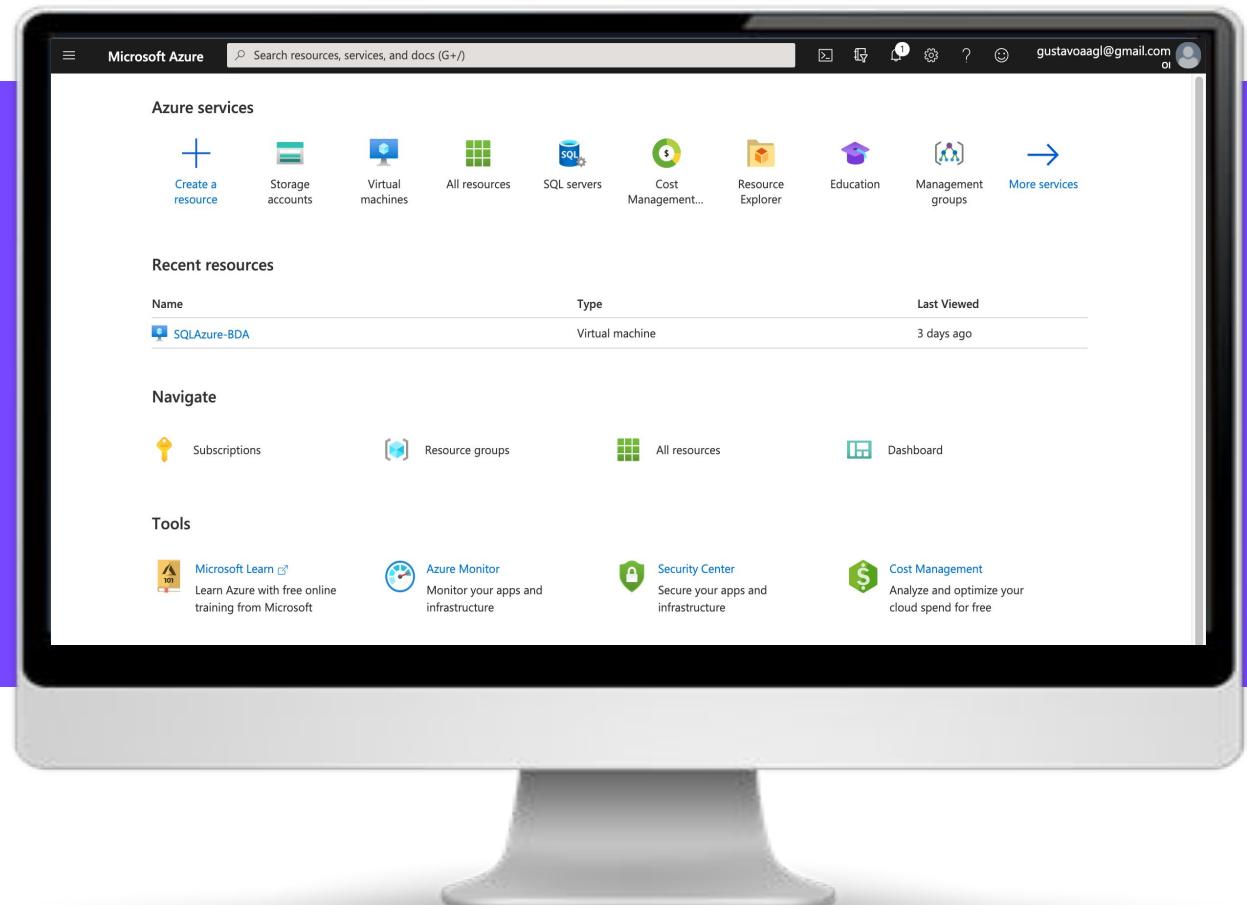


- Bancos de Dados Não Relacionais com IaaS
- Bancos de Dados Não Relacionais com PaaS

Banco de Dados Não Relacional com IaaS e PaaS



 **Demo**



Próxima Aula



- Azure Cosmos DB.

Armazenamento de Dados Não Relacionais no Azure

CAPÍTULO 4. AULA 4.2. AZURE COSMOS DB

PROF. GUSTAVO AGUILAR

Nesta Aula

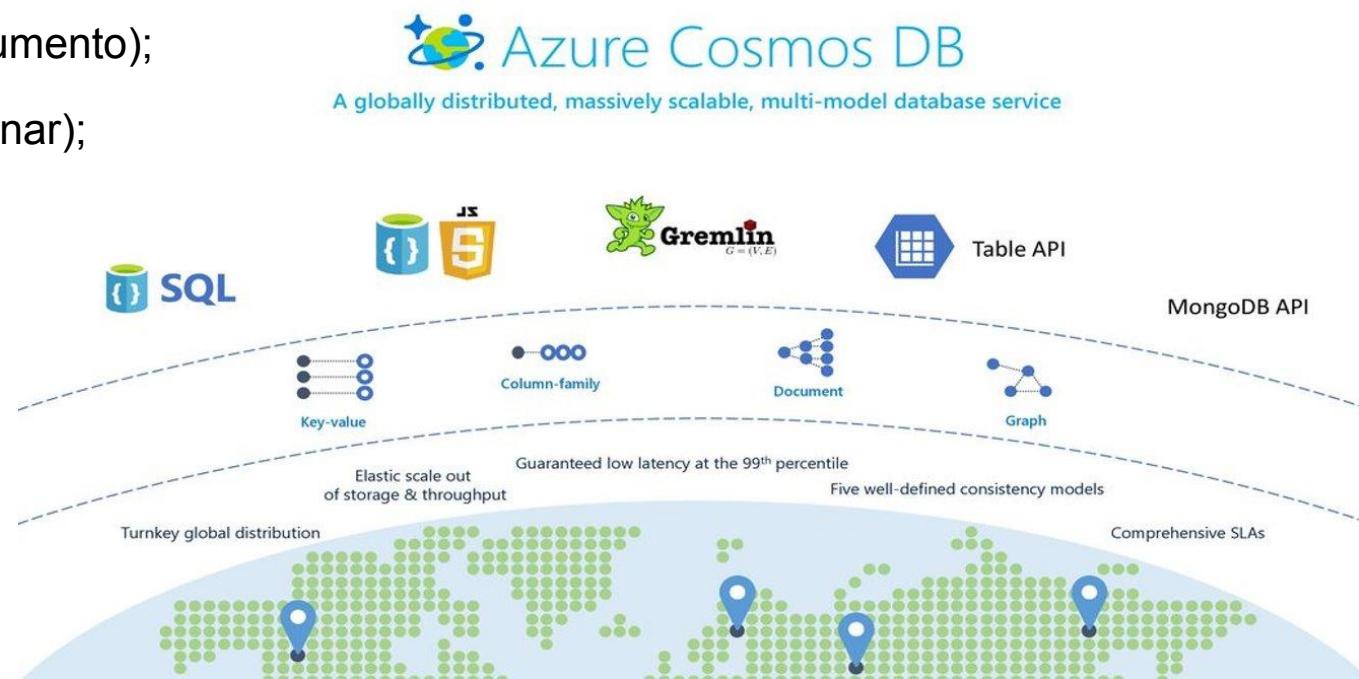


- Azure Cosmos DB
- Pré-requisitos para Criação

Azure Cosmos DB



- Serviço de banco de dados multimodelo e distribuídos globalmente;
- Database as a Service (PaaS);
- Suporte a 5 APIs (modelos de DB Engine):
 - SQL (relacional);
 - Mongo DB (documento);
 - Cassandra (colunar);
 - Gremlin (grafo);
 - Table.



Azure Cosmos DB

Home > jpwcosmosdb | Replicate data globally

jpwcosmosdb | Replicate data globally

Azure Cosmos DB account

Search (Ctrl+ /) Save Discard Manual Failover Automatic Failover

Click on a location to add or remove regions from your Azure Cosmos DB account.
* Each region is billable based on the throughput and storage for the account. Learn more

Tags Diagnose and solve problems Quick start Notifications Data Explorer Settings Features Replicate data globally Default consistency Firewall and virtual networks Private Endpoint Connections CORS Keys Add Azure Cognitive Search Add Azure Function Advanced security (preview)

Configure regions Multi-region writes [Enable](#)

Configure the regions for reads, writes and availability zone (supported in selected regions and can only be configured when a new region is added). [+ Add region](#)

Write Region	Availability Zone
West US	

Read Regions	Availability Zone	Action
The account has no read regions.		



Azure Cosmos DB

IGTI

Home > jpwcosmosdb | Default consistency

jpwcosmosdb | Default consistency

Azure Cosmos DB account

Save Discard

STRONG BOUNDED STALENESS SESSION CONSISTENT PREFIX **EVENTUAL**

i Eventual consistency is the weakest form of consistency wherein a client may get the values which are older than the ones it had seen before, over time.

Understand Eventual consistency

In the absence of any further writes, the replicas within the group will eventually converge. Eventual consistency is ideal where the application does not require any ordering guarantees. Examples include count of Retweets, Likes or non-threaded comments.

Azure Cosmos DB



- **Consistência dos Dados**
- **Eventual:**
 - Alterações não aparecem imediatamente nas réplicas;
 - Não há perda de dados.
 - Várias alterações em lote □ podem aparecer fora de ordem;
 - Menor consistência, mas menor latência.
- **Strong:**
 - Alterações visíveis para os demais usuários apenas depois delas serem confirmadas em todas as réplicas;
 - Não disponível para distribuição entre múltiplas regiões globais;
 - Maior consistência, mas maior latência.

Azure Cosmos DB



- **Consistência dos Dados**

- **Consistent Prefix:**

- Alterações aparecerão em ordem, embora possa haver um atraso antes de se tornarem visíveis;
 - Neste período, os aplicativos / usuários poderão estar vendo dados antigos.

- **Session:**

- Todas alterações ficam visíveis e em ordem para o “owner” delas;
 - Conhecida como *leia suas próprias gravações*.

- **Bounded Staleness:**

- Há um atraso entre a gravação e a leitura dos dados atualizados;
 - Tempo ou no. de versões anteriores que os dados estarão inconsistentes.

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.



Pré-requisitos para Criação

▪ 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;
- Pode aumentar ou diminuir RU/s a qualquer momento.

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

Próxima Aula



- Demo Azure Cosmos DB.

Armazenamento de Dados Não Relacionais no Azure

CAPÍTULO 4. AULA 4.3. DEMO AZURE COSMOS DB

PROF. GUSTAVO AGUILAR

Nesta Aula

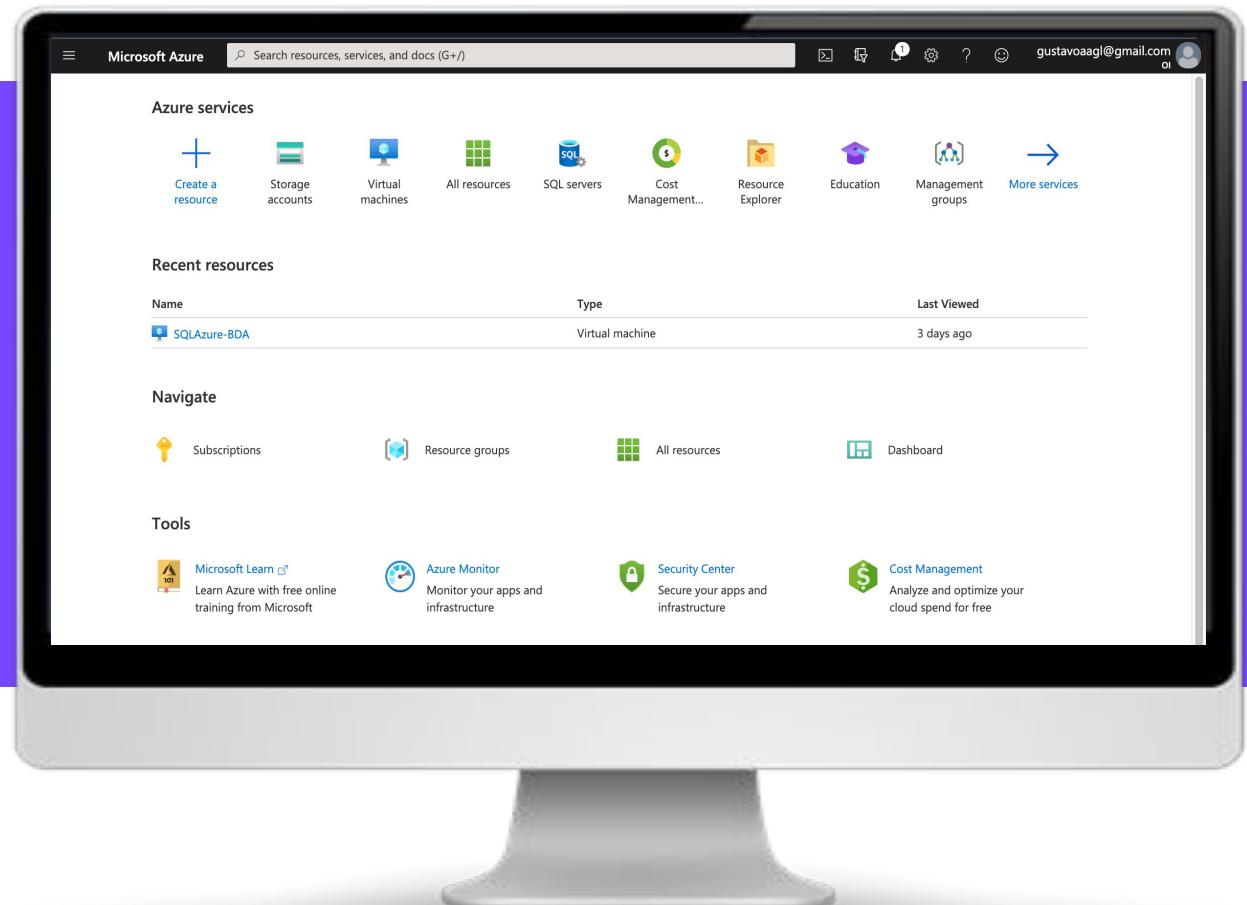


- ❑ Aprovisionando um Azure Cosmos DB com API SQL
- ❑ Usando um Azure Cosmos DB

Aprovisionando e Usando um Azure Cosmos DB



 Demo



Próxima Aula



- Capítulo 5 – Soluções para Big Data & Analytics.

Serviços de Armazenamento, Banco de Dados e Analytics

Capítulo 5. Soluções para Big Data & Analytics

PROF. GUSTAVO AGUILAR

Serviços de Armazenamento, Banco de Dados e Analytics

CAPÍTULO 5. AULA 5.1. INTRODUÇÃO À BIG DATA

PROF. GUSTAVO AGUILAR

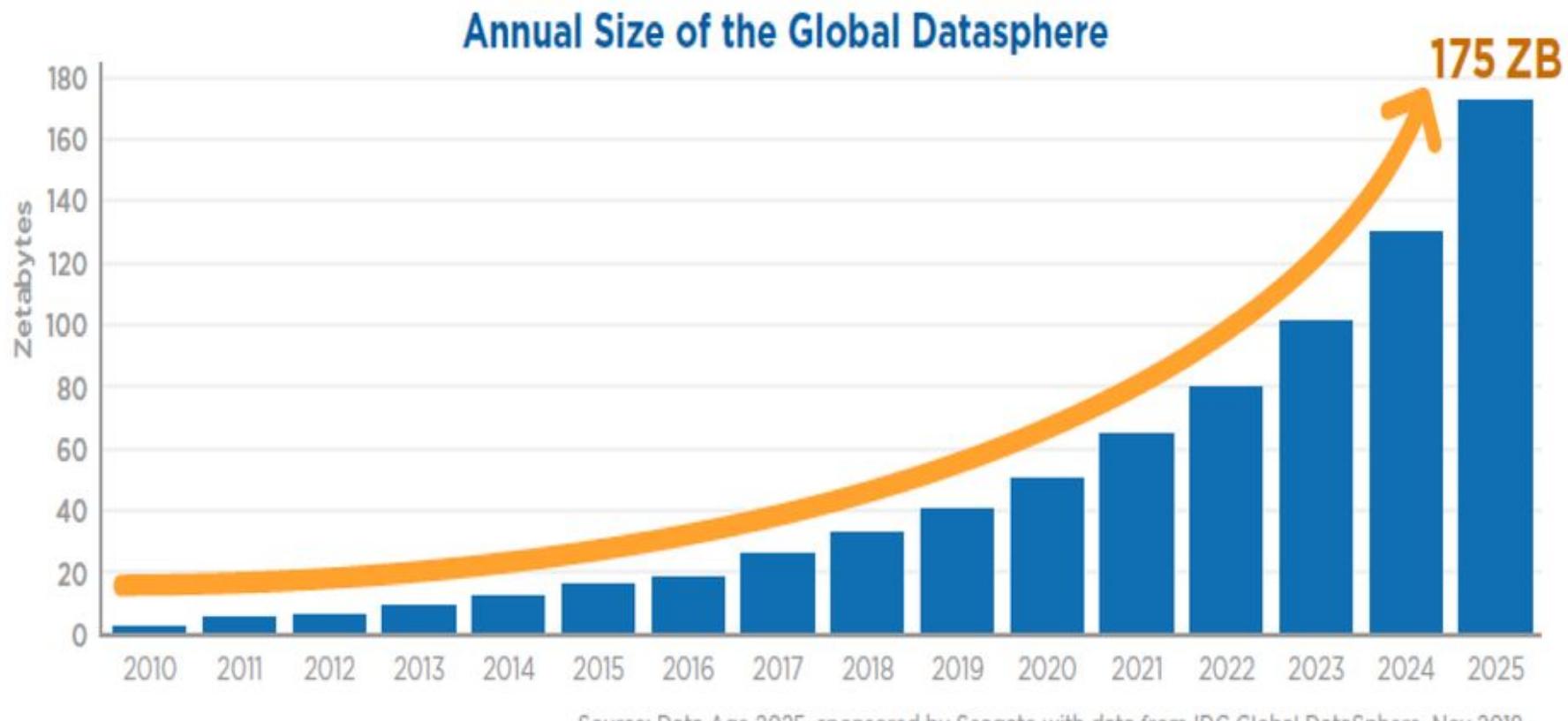
Nesta Aula



- O Que É Big Data?
- Fundamentos de Big Data

O Que É Big Data

IGTI



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)

O Que É Big Data

- A intensa utilização de redes sociais on-line, 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 a 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)

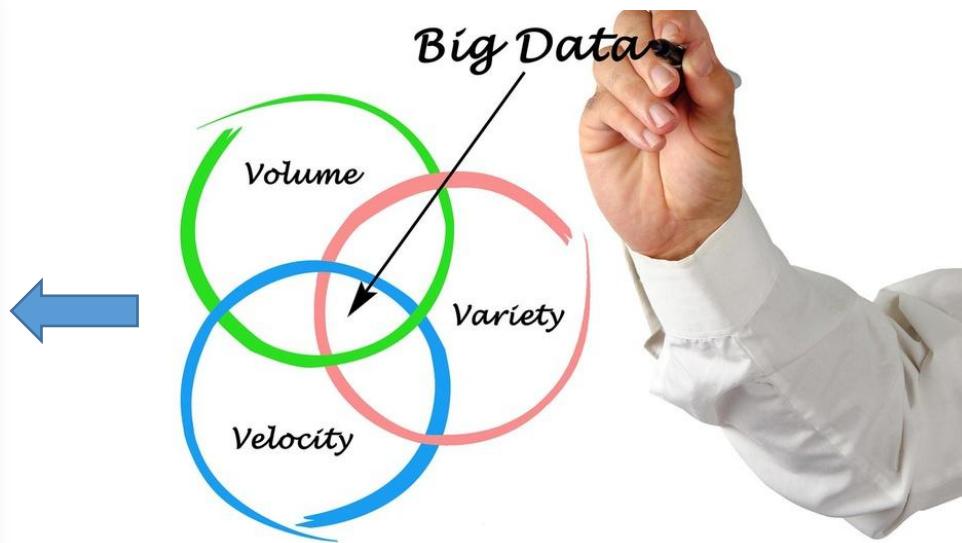
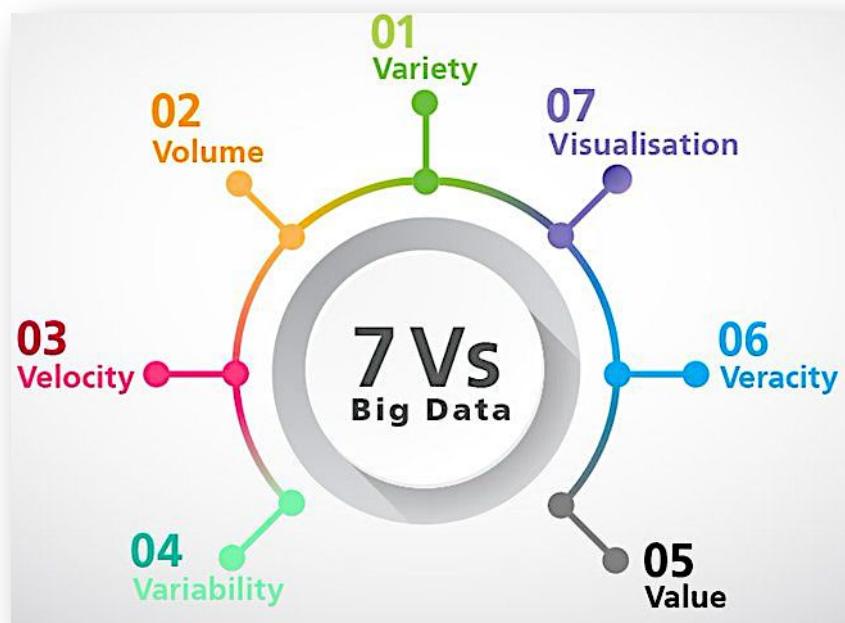
O Que É Big Data

- 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.



Fundamentos de Big Data

IGTI



Fundamentos de Big Data



- **VOLUME:** a quantidade de dados gerados, que costumava ser medida em Gigabytes agora é medida em Zettabytes (ZB), caminhando para Yottabytes (YB).
- **VELOCIDADE:** velocidade em que os dados são gerados e se tornam acessíveis.
- **VARIEDADE:** de formatos, de tipos (estruturado, não estruturado e semiestruturado), e da natureza (numérica, data, caractere, etc.).
- **VARIABILIDADE:** mesma combinação de dados cujo significado muda constantemente.

Fundamentos de Big Data

- **VERACIDADE:** garantir que os dados sejam precisos, verdadeiros, fidedignos.
 - **VISUALIZAÇÃO:** formas mais aderentes para visualizar grandes quantidades de dados complexos.
 - **VALOR:** transformar dados em valor, vantagem competitiva ou otimização operacional.



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.



Próxima Aula



- ❑ Introdução ao HDInsight

Serviços de Armazenamento, Banco de Dados e Analytics

CAPÍTULO 5. AULA 5.2. INTRODUÇÃO AO HDINSIGHT

PROF. GUSTAVO AGUILAR

Nesta Aula



- ❑ Overview do HDInsight
- ❑ Principais Componentes
- ❑ Ecossistema do Azure 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 Hadoop



Apache Spark



Apache Kafka



Apache HBase



Apache Hive LLAP



Apache Storm



Machine Learning

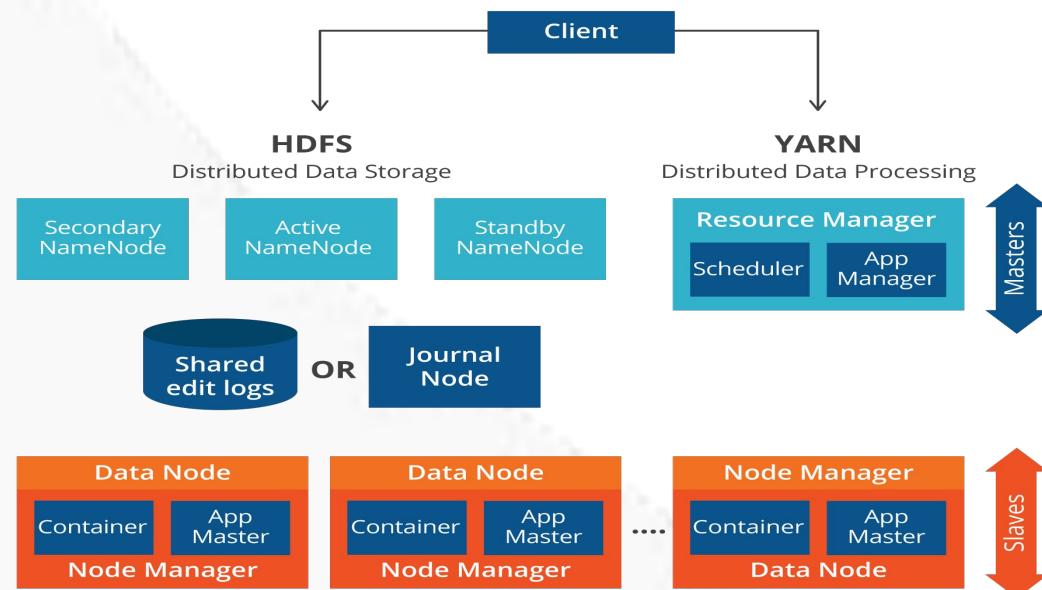
Principais Componentes

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 Hadoop



Principais Componentes

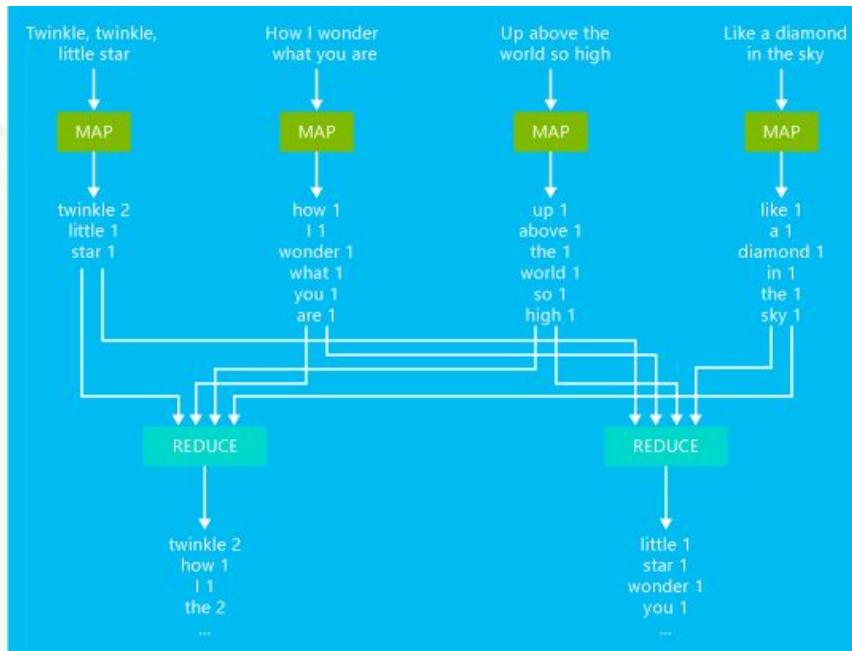
IGTI

APACHE HADOOP

- Utiliza um modelo de programação MapReduce simples para processar e analisar, paralelamente, dados em lote.



Apache Hadoop



Principais Componentes

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.

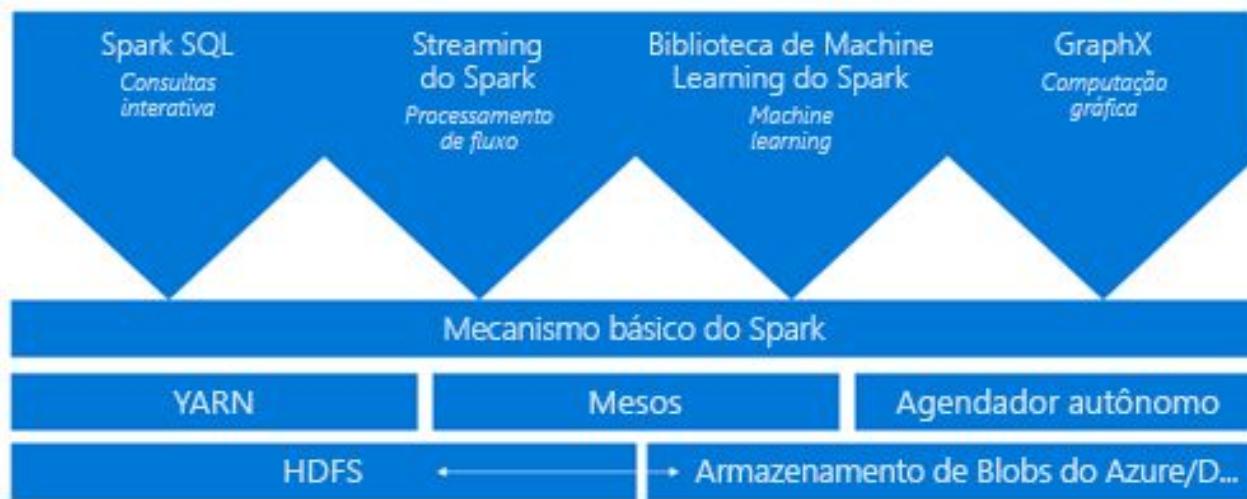


Apache Spark



Principais Componentes

APACHE SPARK



Principais Componentes

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.

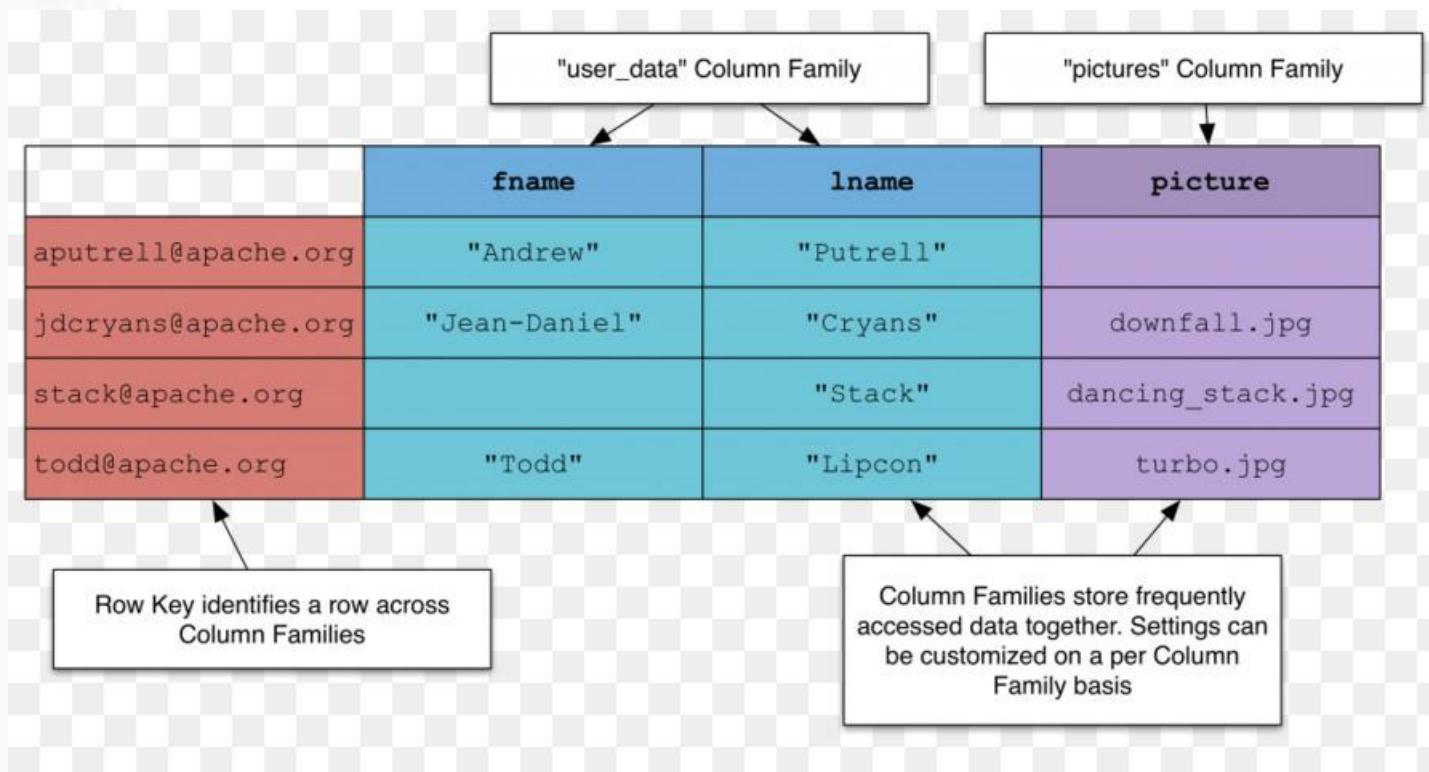


Apache HBase

Principais Componentes

IGTI

APACHE HBASE



Principais Componentes

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 Kafka



Principais Componentes

APACHE STORM

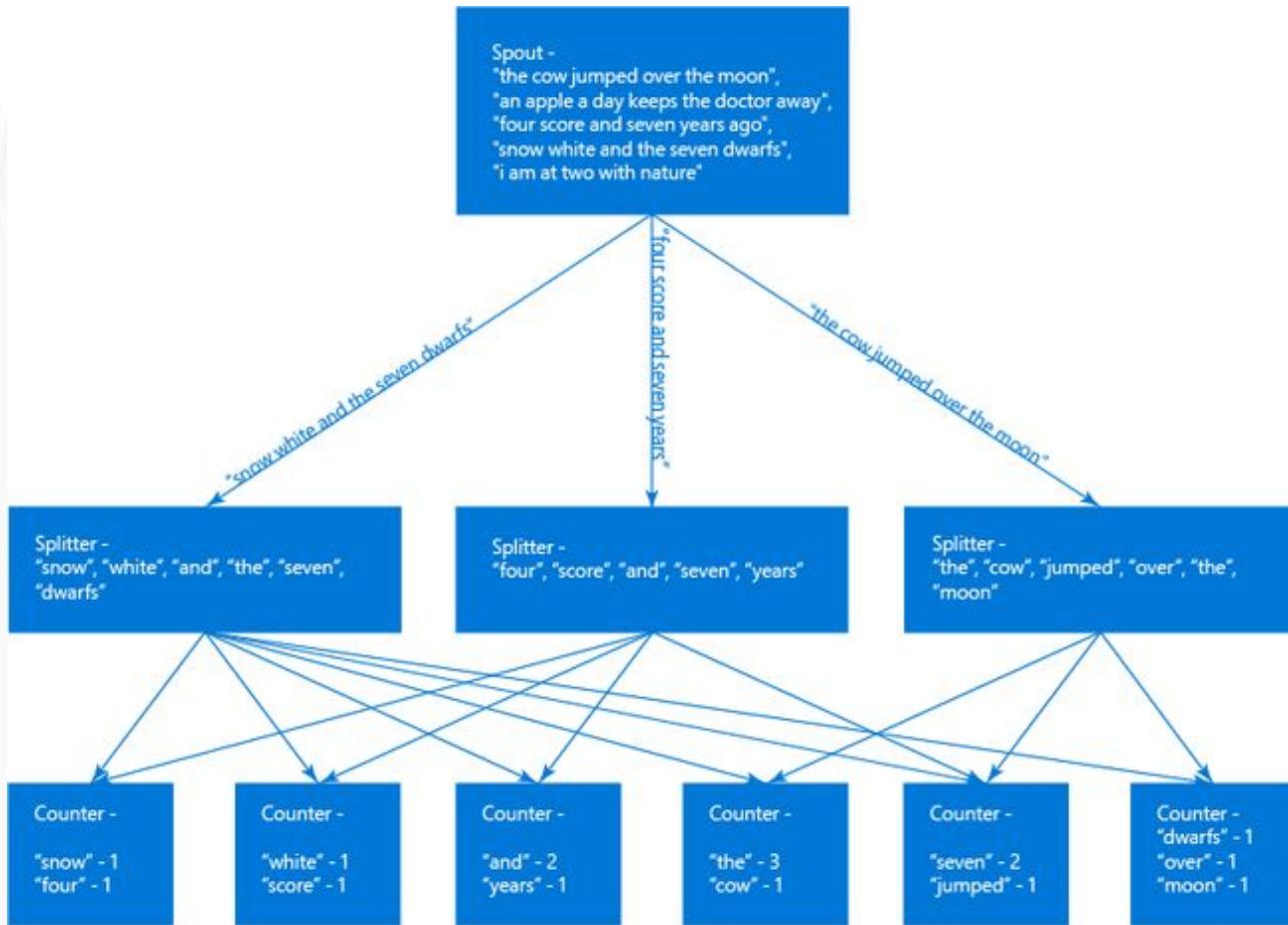
- Sistema de computação distribuído e em tempo real para processar rapidamente grandes fluxos de dados;
- Processa topologias ao invés de trabalhos MapReduce.



Apache Storm

Principais Componentes

IGTI



Apache Storm

Principais Componentes

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 Hive LLAP

Principais Componentes

APACHE ML SERVICES

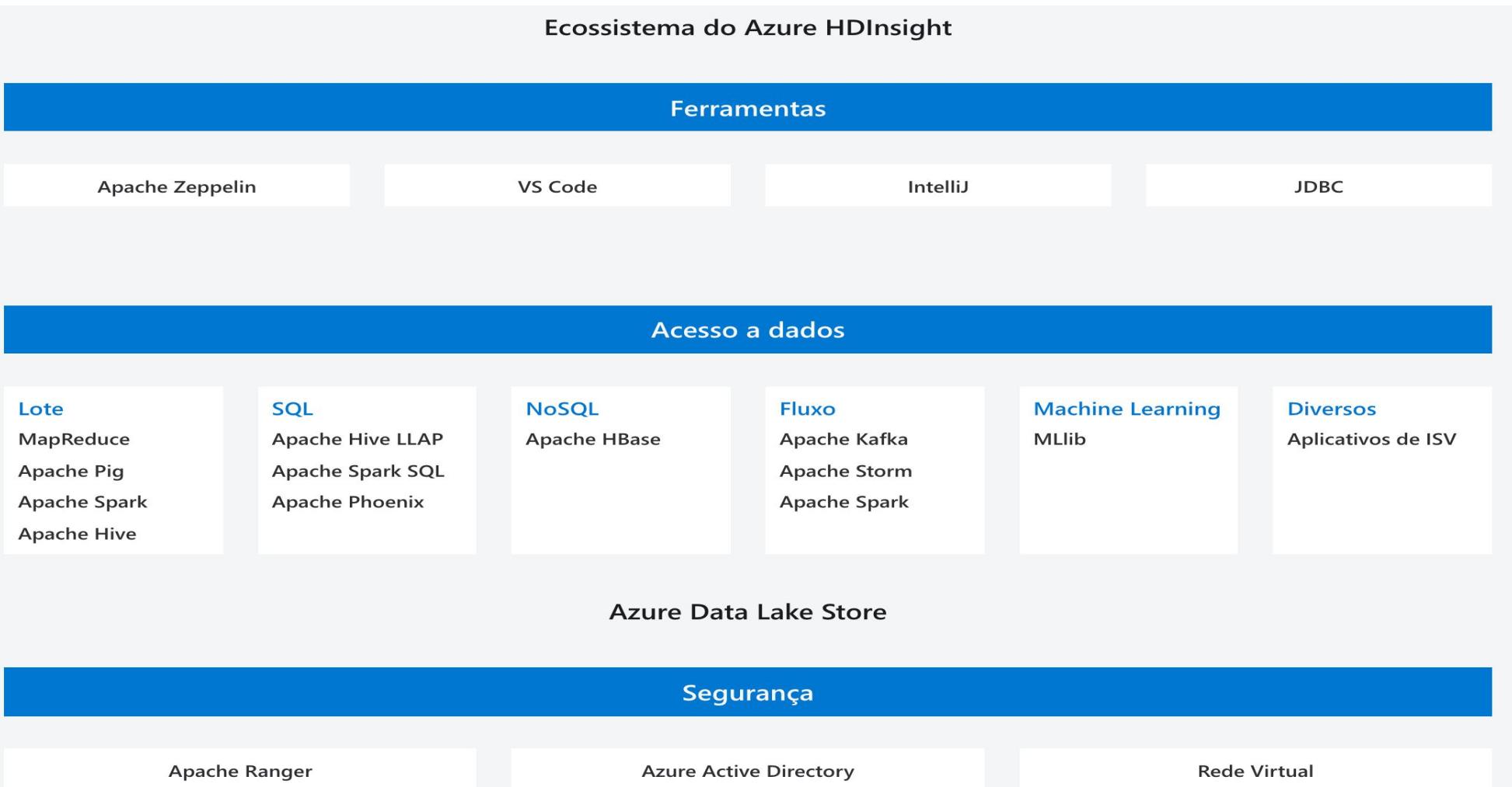
- Cluster para soluções de aprendizagem de máquina;
- Fornece aos cientistas de dados, estatísticos e programadores de R / Python o acesso sob demanda a métodos escalonáveis e distribuídos de análise no HDInsigh;
- Possui um conjunto de modelos e algoritmos de machine learning que podem ser adaptados.



Machine Learning

Ecossistema do Azure HDInsight

IGTI



Próxima Aula



- ❑ Aprovisionando um Ambiente do HDInsight.

Serviços de Armazenamento, Banco de Dados e Analytics

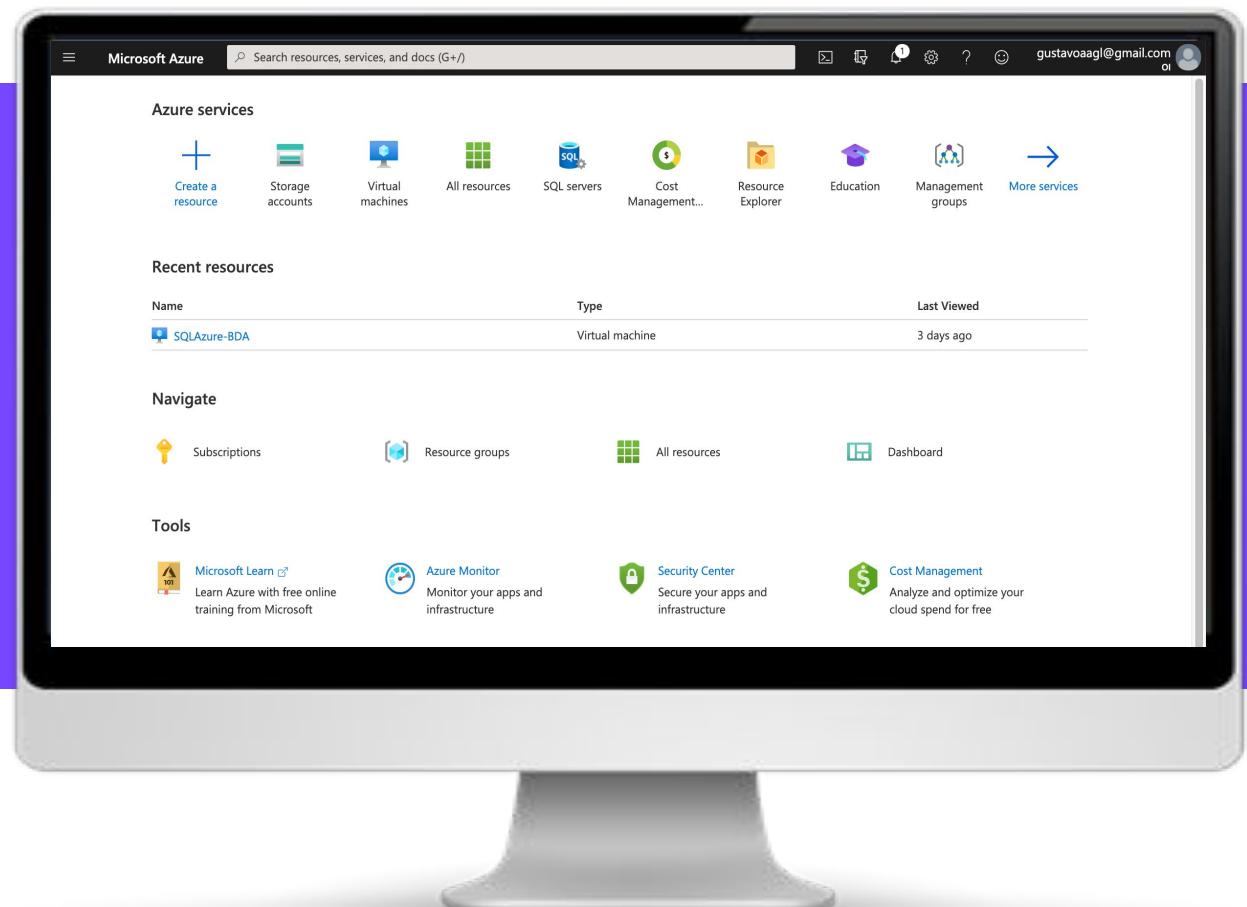
CAPÍTULO 5. AULA 5.3. APROVISIONANDO UM AMBIENTE DO HDINSIGHT

PROF. GUSTAVO AGUILAR

Aprovisionando um Ambiente de HDInsight



 Demo



Próxima Aula



- ❑ Introdução ao Azure DataBricks.

Serviços de Armazenamento, Banco de Dados e Analytics

CAPÍTULO 5. AULA 5.4. INTRODUÇÃO AO AZURE DATABRICKS

PROF. GUSTAVO AGUILAR

Nesta Aula



- ❑ Overview do Azure Databricks
- ❑ Componentes do Azure Databricks

Overview do Azure Databricks

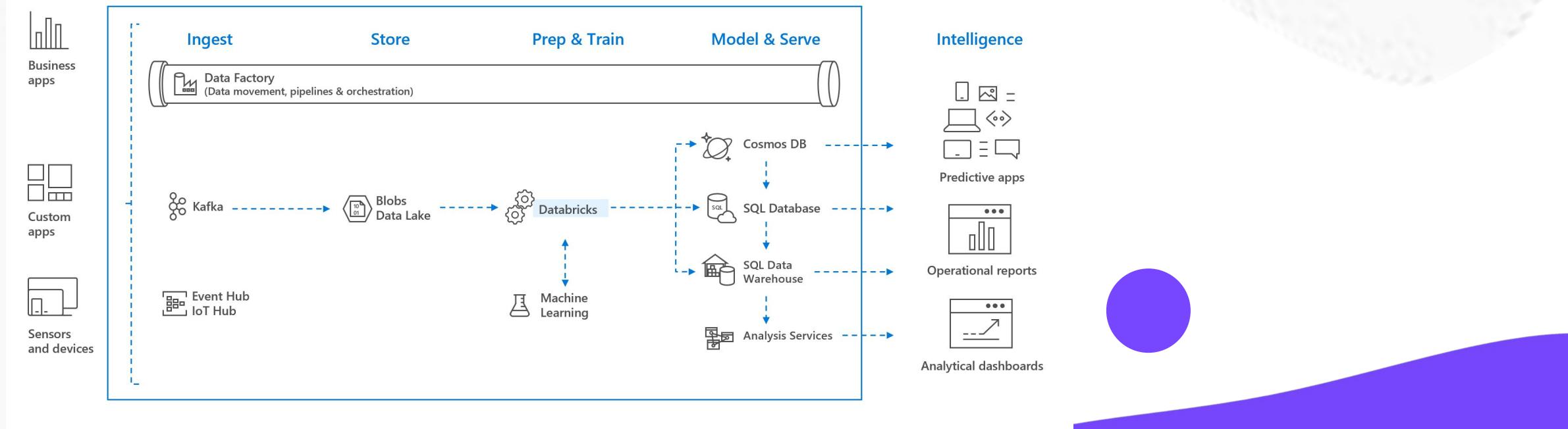


- Plataforma de análise baseada no Apache Spark e otimizada para a plataforma de serviços de nuvem do Microsoft Azure;
- Configuração rápida e automatizada do cluster Spark;
- Fluxos de trabalho simplificados;
- Workspace interativo que permite a colaboração entre os cientistas de dados, os engenheiros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



Overview do Azure Databricks

IGTI

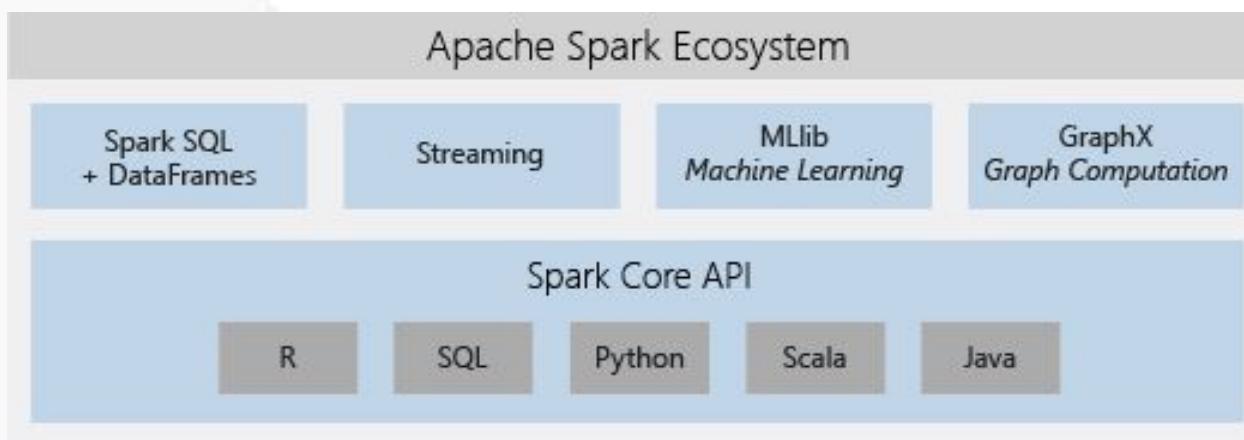


Componentes do Azure Databricks



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).

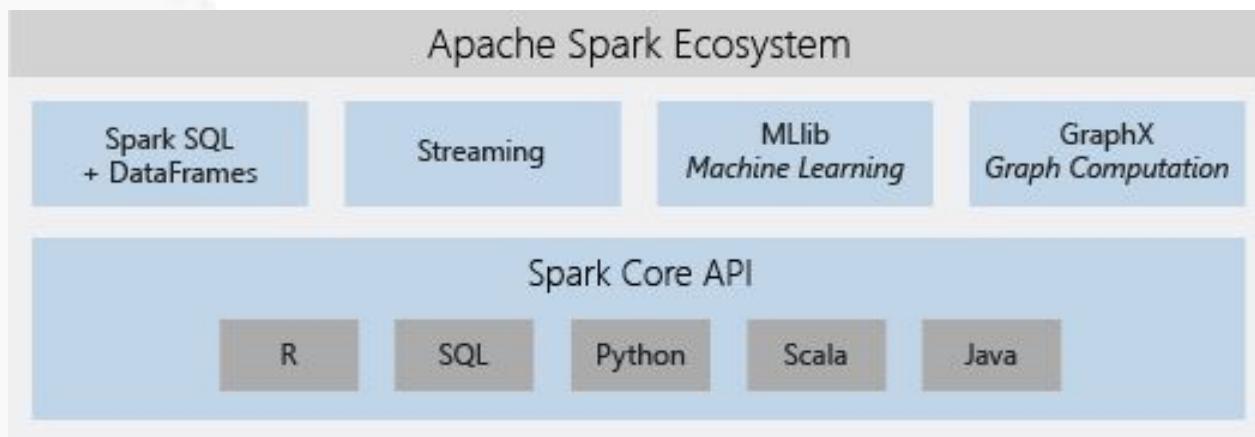


Componentes do Azure Databricks



STREAMING

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



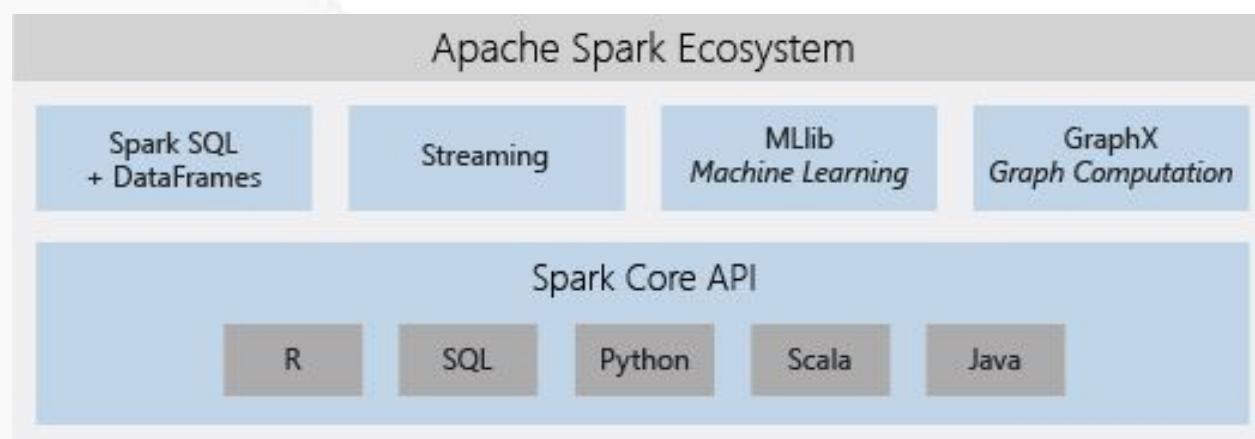
Componentes do Azure Databricks



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.



Próxima Aula



- ❑ Demonstração do Azure DataBricks.

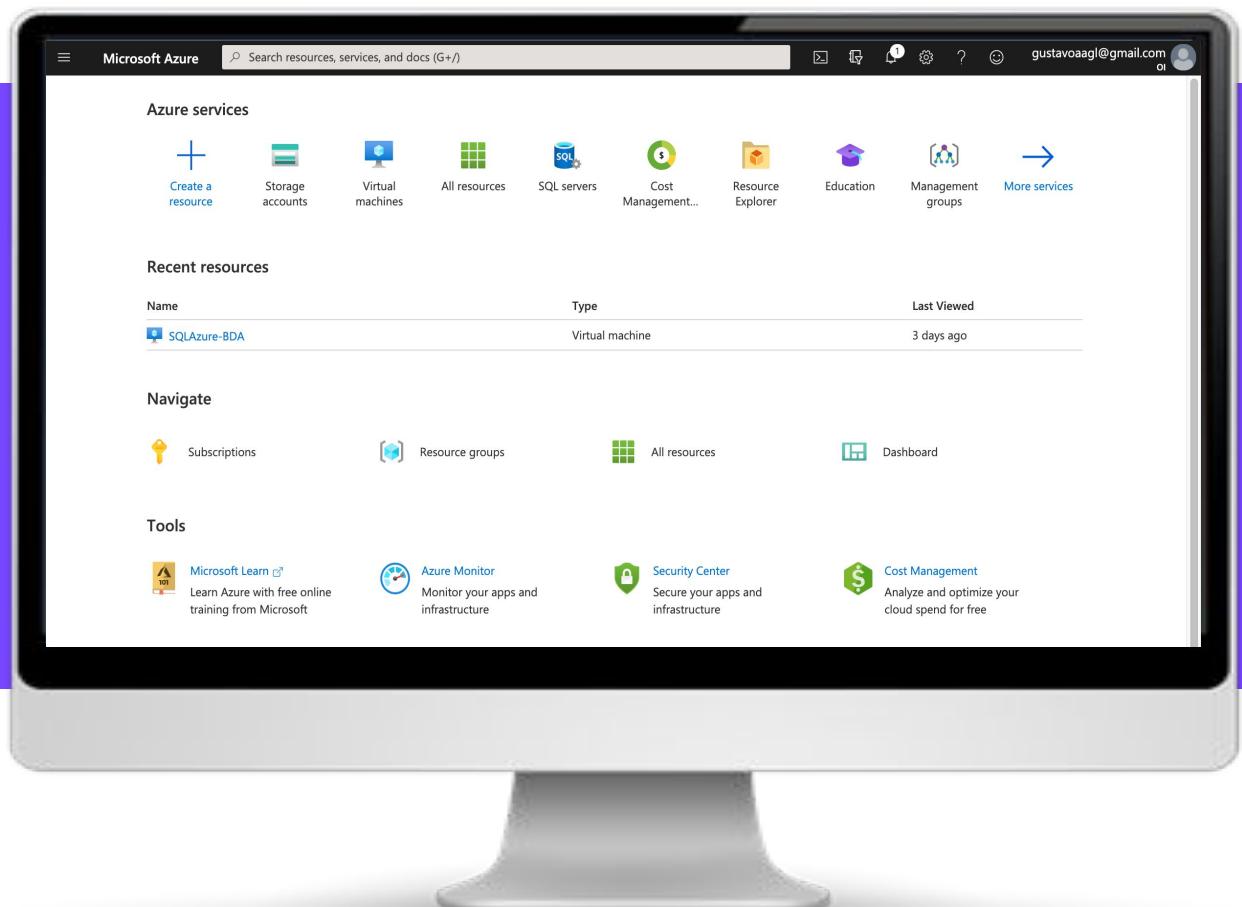
Serviços de Armazenamento, Banco de Dados e Analytics

CAPÍTULO 5. AULA 5.5. DEMONSTRAÇÃO DO AZURE DATABRICKS

PROF. GUSTAVO AGUILAR

Azure Databricks

 Demo



Próxima Aula



- ❑ Introdução ao Azure Synapse Analytics.

Serviços de Armazenamento, Banco de Dados e Analytics

CAPÍTULO 5. AULA 5.6. INTRODUÇÃO AO AZURE SYNAPSE ANALYTICS

PROF. GUSTAVO AGUILAR

Nesta Aula



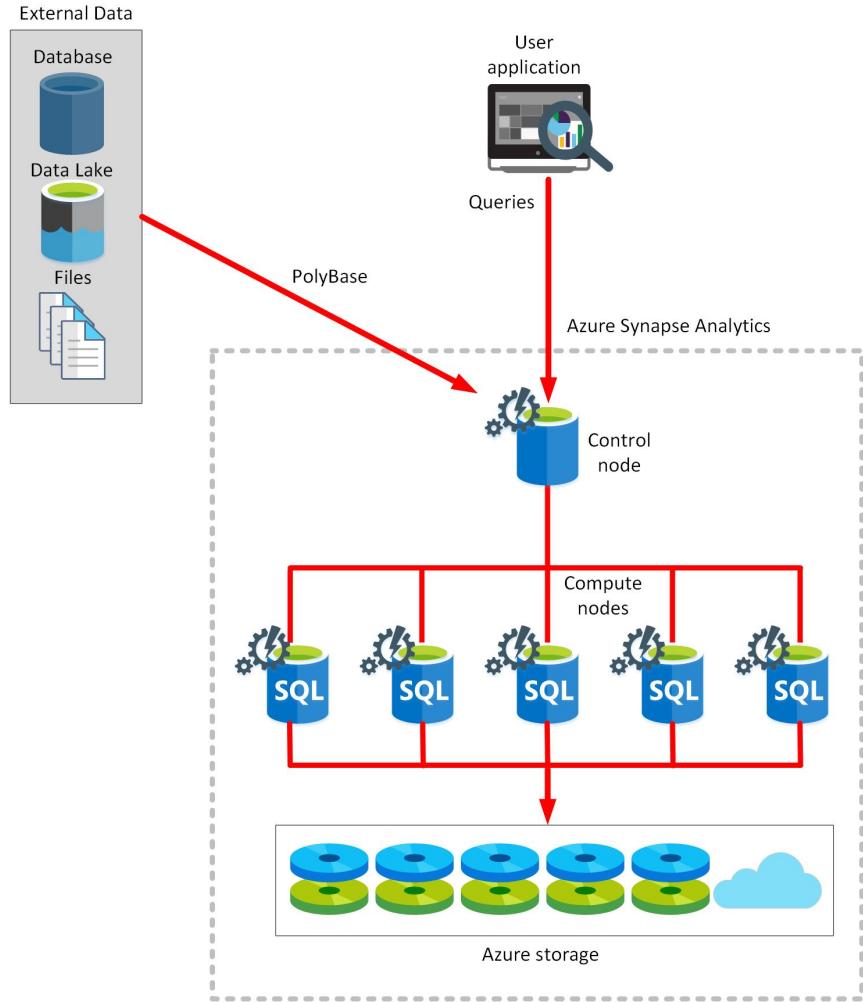
- Azure Synapse Analytics
- SQL Pools
- Workload Groups

Azure Synapse Analytics

- Anteriormente conhecido como SQL DW □ MPP;
- Fornece um ambiente unificado combinando data warehouse e os recursos de análise de Big Data do Spark;
- Azure Synapse tem quatro componentes:
 - Synapse SQL: análise completa baseada em T-SQL;
 - Pool de SQL: pagamento por DWU provisionado;
 - SQL sob demanda: pagamento por TB processado.
 - Apache Spark;
 - Data Integration: Integração de dados híbridos;
 - Studio: experiência de usuário unificada.

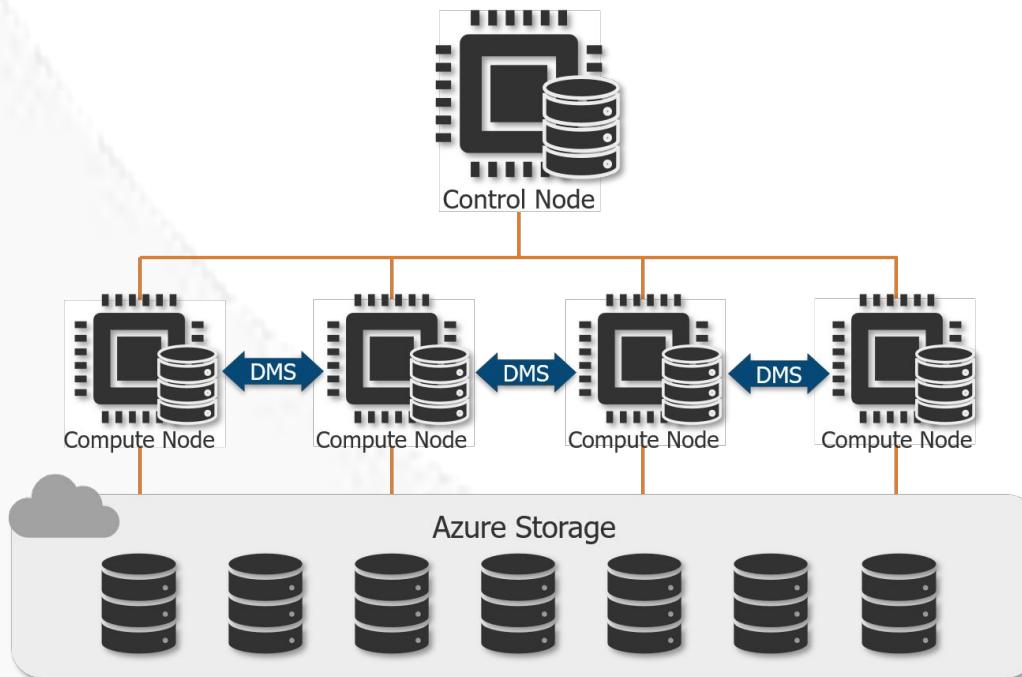
Azure Synapse Analytics

iGTTi



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 consumer;
 - 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'  
.....
```

Próxima Aula



- ❑ Demonstração do Azure Synapse Analytics.



Serviços de Armazenamento, Banco de Dados e Analytics

CAPÍTULO 5. AULA 5.7. DEMONSTRAÇÃO DO AZURE SYNAPSE ANALYTICS

PROF. GUSTAVO AGUILAR

Nesta Aula

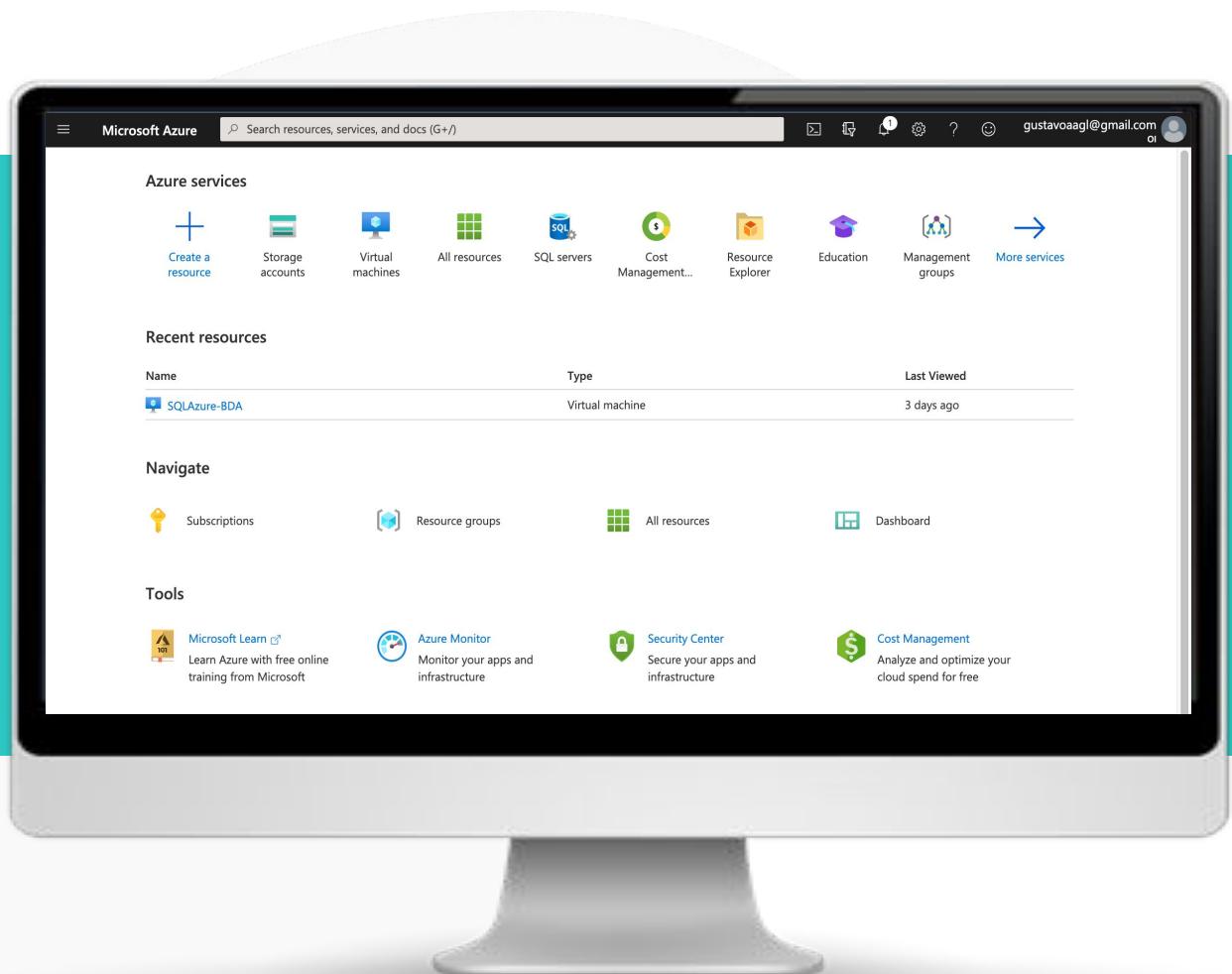


- Demonstração do Azure Synapse Analytics.

Azure Synapse Analytics



 Demo



Próxima Aula



- Capítulo 6 - Soluções para Pipeline de Dados.

Serviços de Armazenamento, Banco de Dados e Analytics

Capítulo 6. Soluções para Pipeline de Dados

PROF. GUSTAVO AGUILAR

Serviços de Armazenamento, Banco de Dados e Analytics

CAPÍTULO 6. AULA 6.1. INTRODUÇÃO AO AZURE DATA FACTORY

PROF. GUSTAVO AGUILAR

Nesta Aula



- Overview do Azure Data Factory
- Componentes do Azure Data Factory

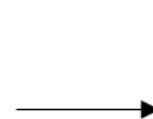
Overview do Azure Data Factory



- 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.



FTP server



Data Factory

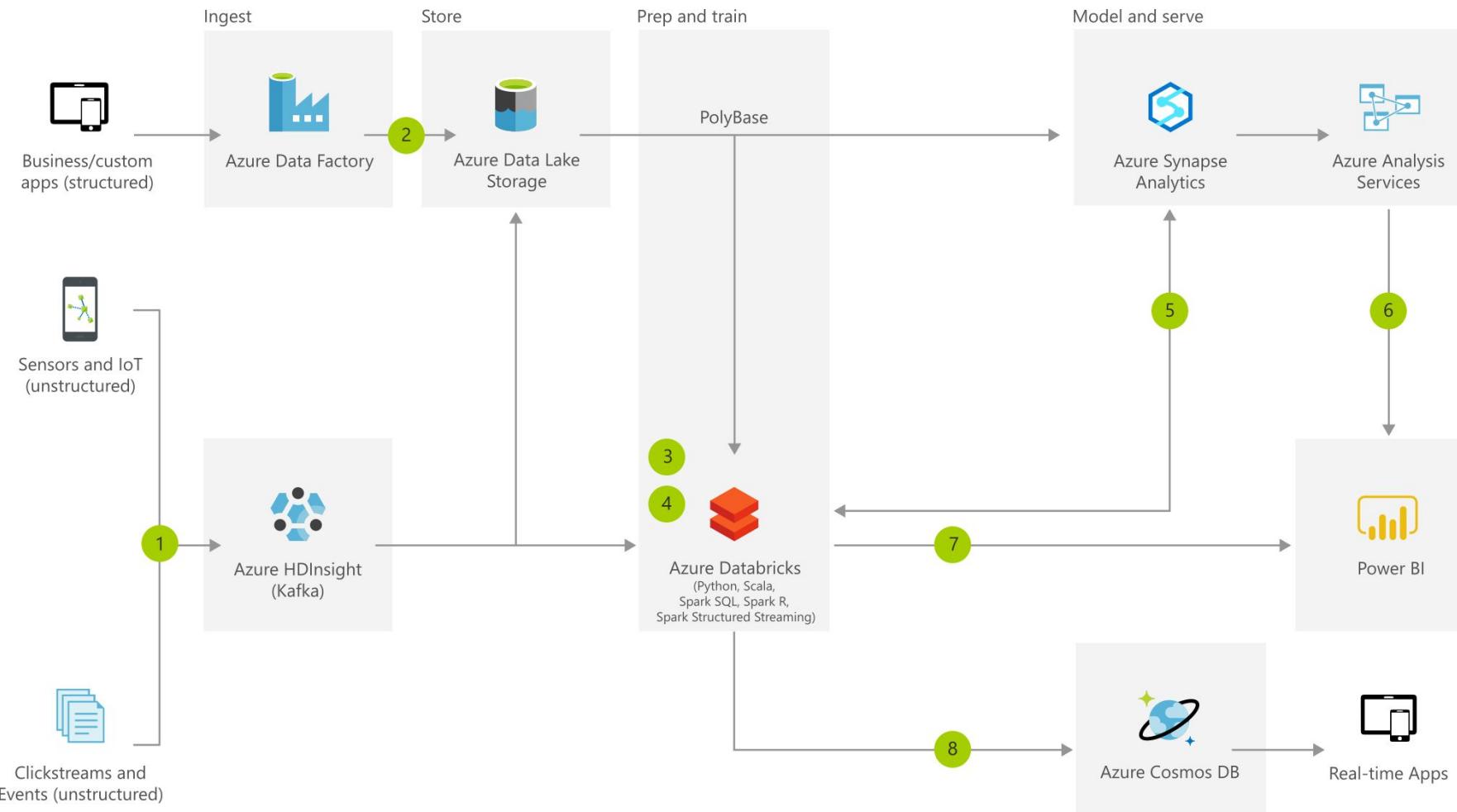


Blob Storage



Overview do Azure Data Factory

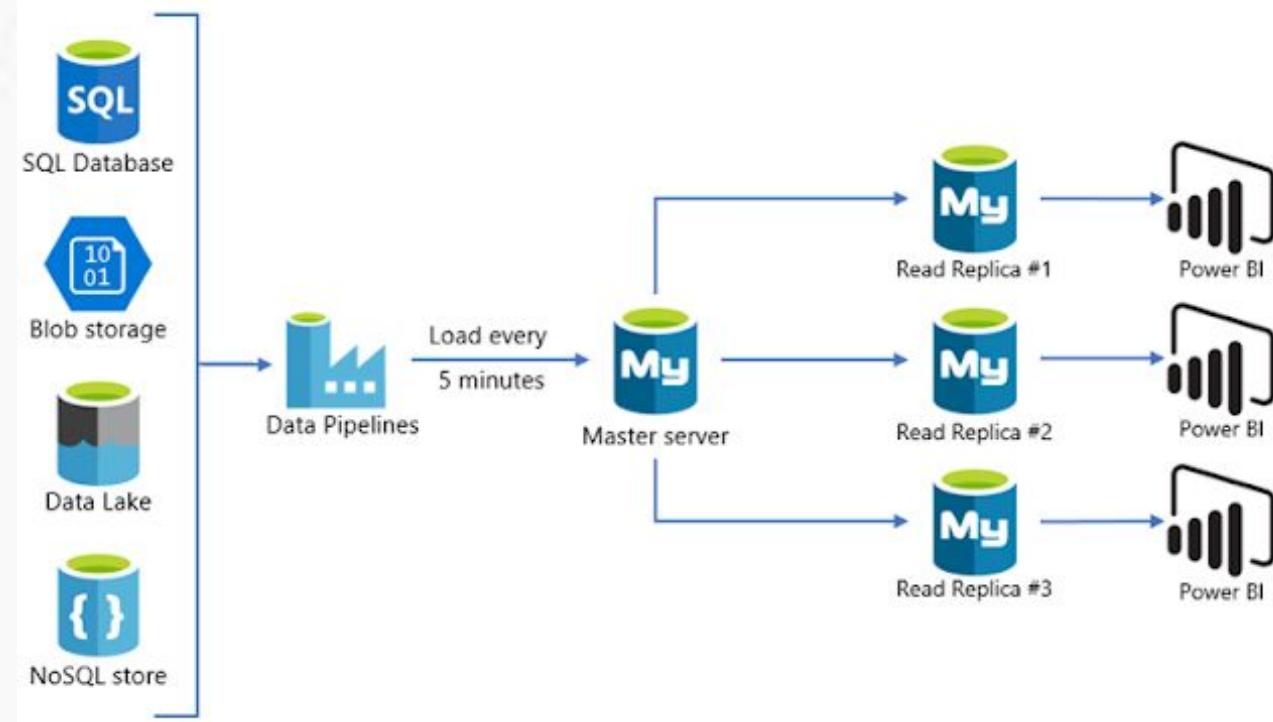
IGTI



Overview do Azure Data Factory

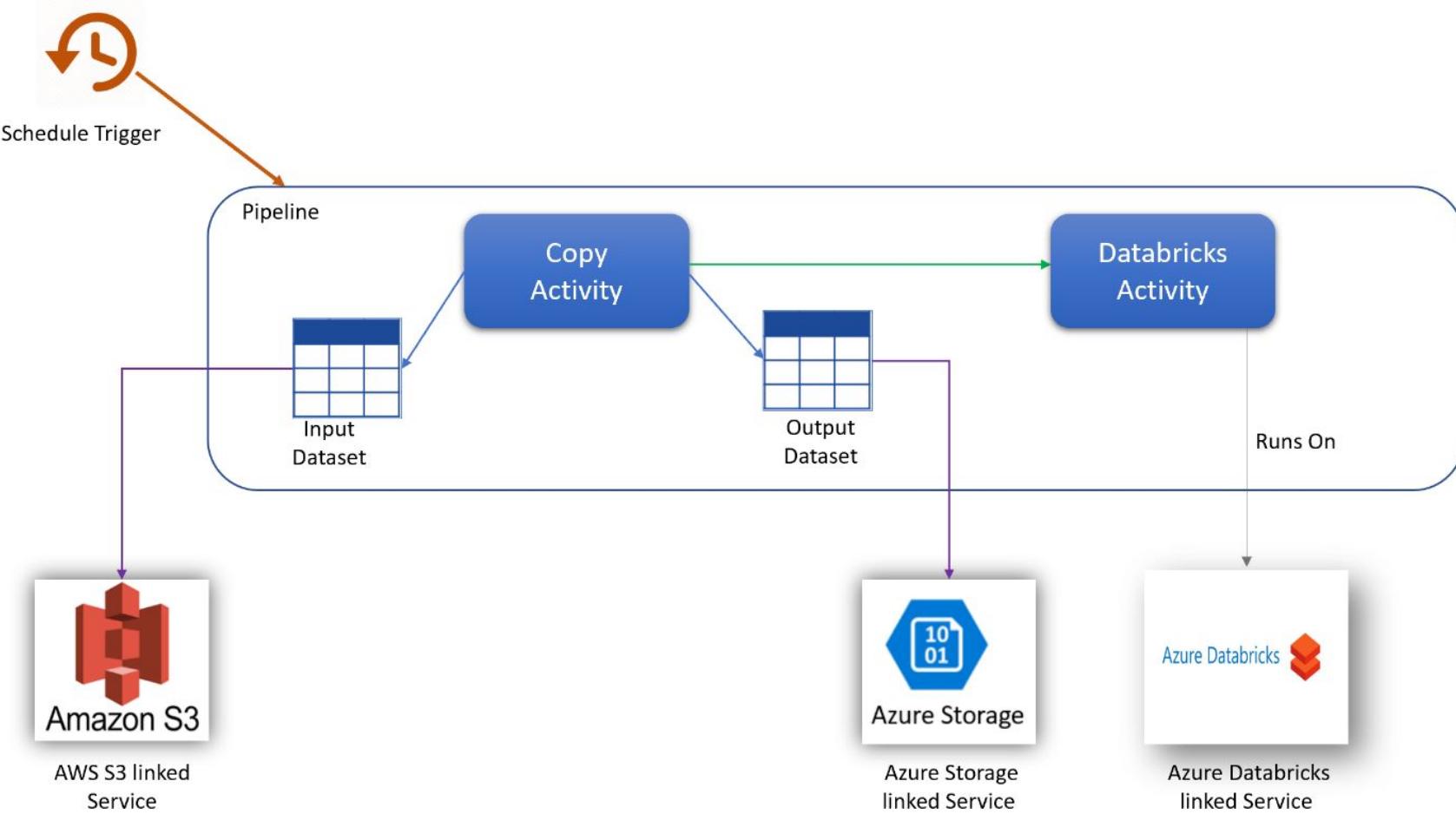
IGTI

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



Overview do Azure Data Factory

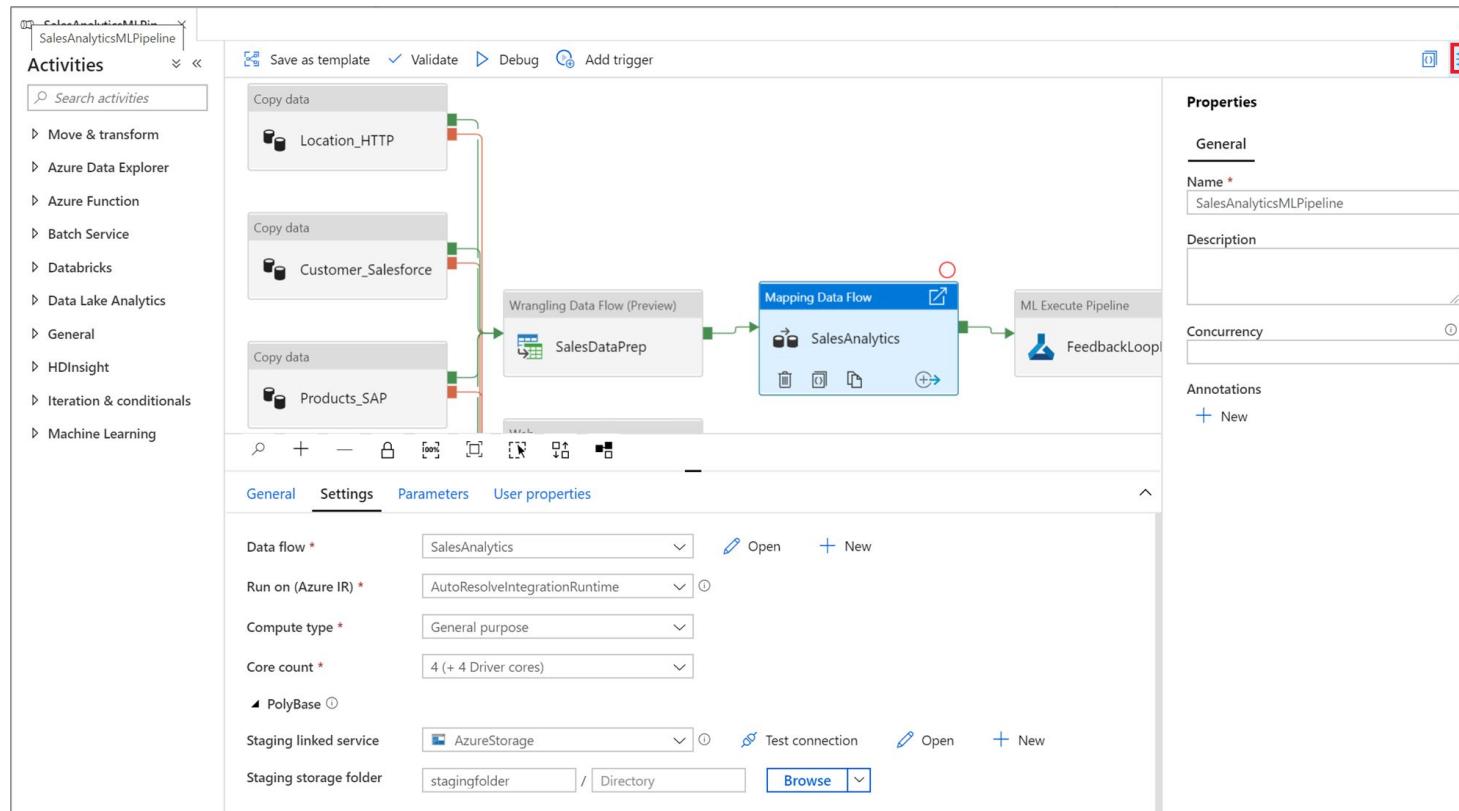
iGTTi



Overview do Azure Data Factory

IGTI

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



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.

Componentes do Azure Data Factory



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.

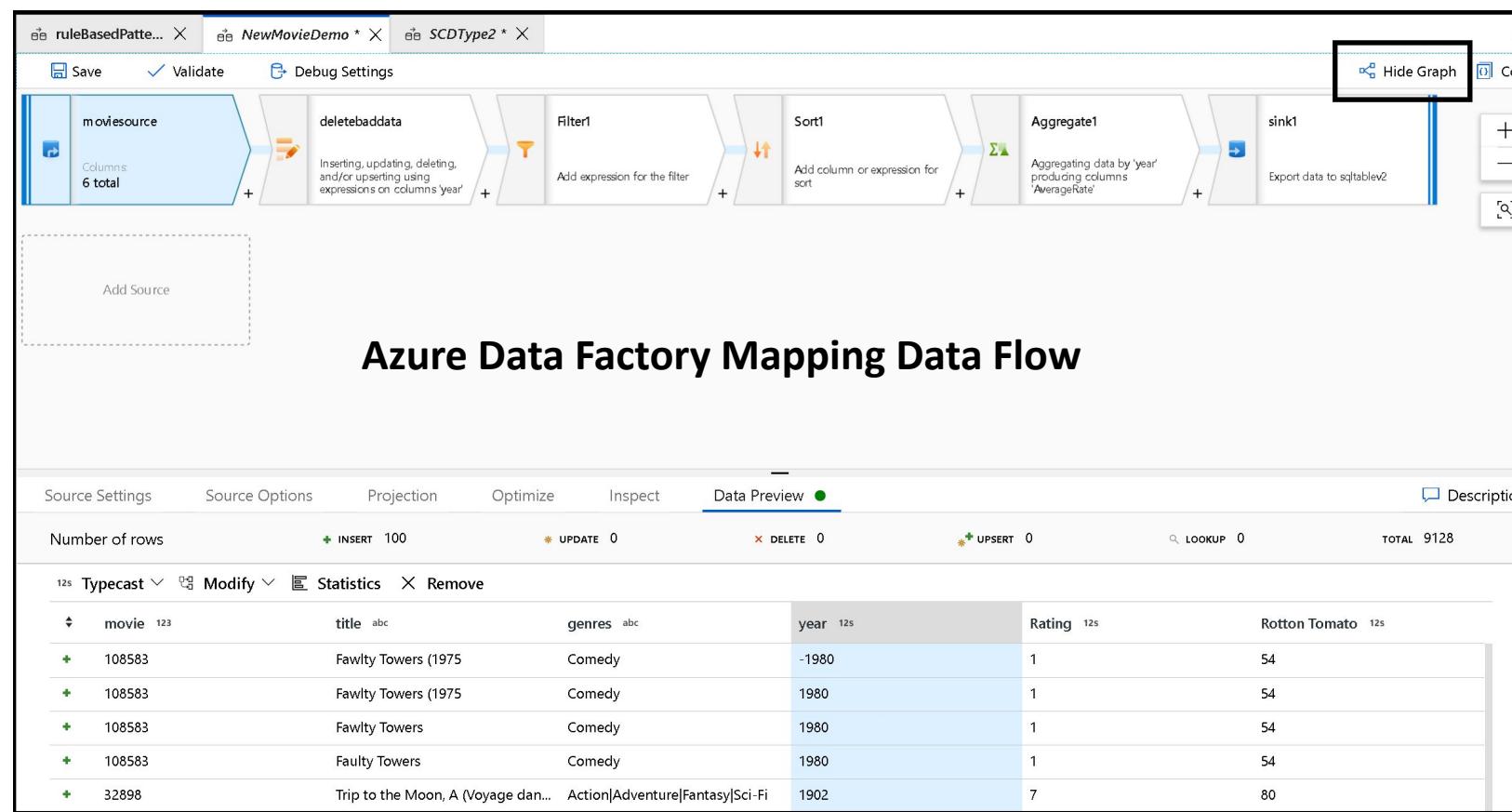
Componentes do Azure Data Factory



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.

Componentes do Azure Data Factory



Componentes do Azure Data Factory

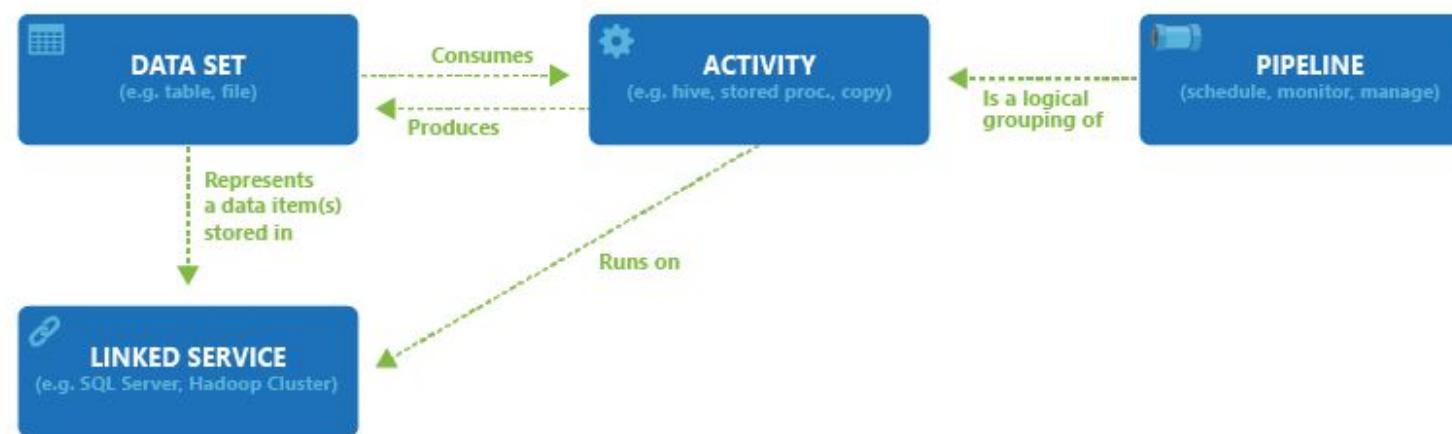


- **CONJUNTO DE DADOS (DATASET):** representam as estruturas de dados nos repositórios de dados, que simplesmente apontam para ou fazem referência aos dados que deseja-se usar em atividades, seja como entrada ou saída.
- **SERVIÇO VINCULADO (LINKED SERVICE):** define as informações de conexão necessárias 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.



Próxima Aula



- ❑ Criando um Pipeline de Dados com o Azure Data Factory.

Serviços de Armazenamento, Banco de Dados e Analytics

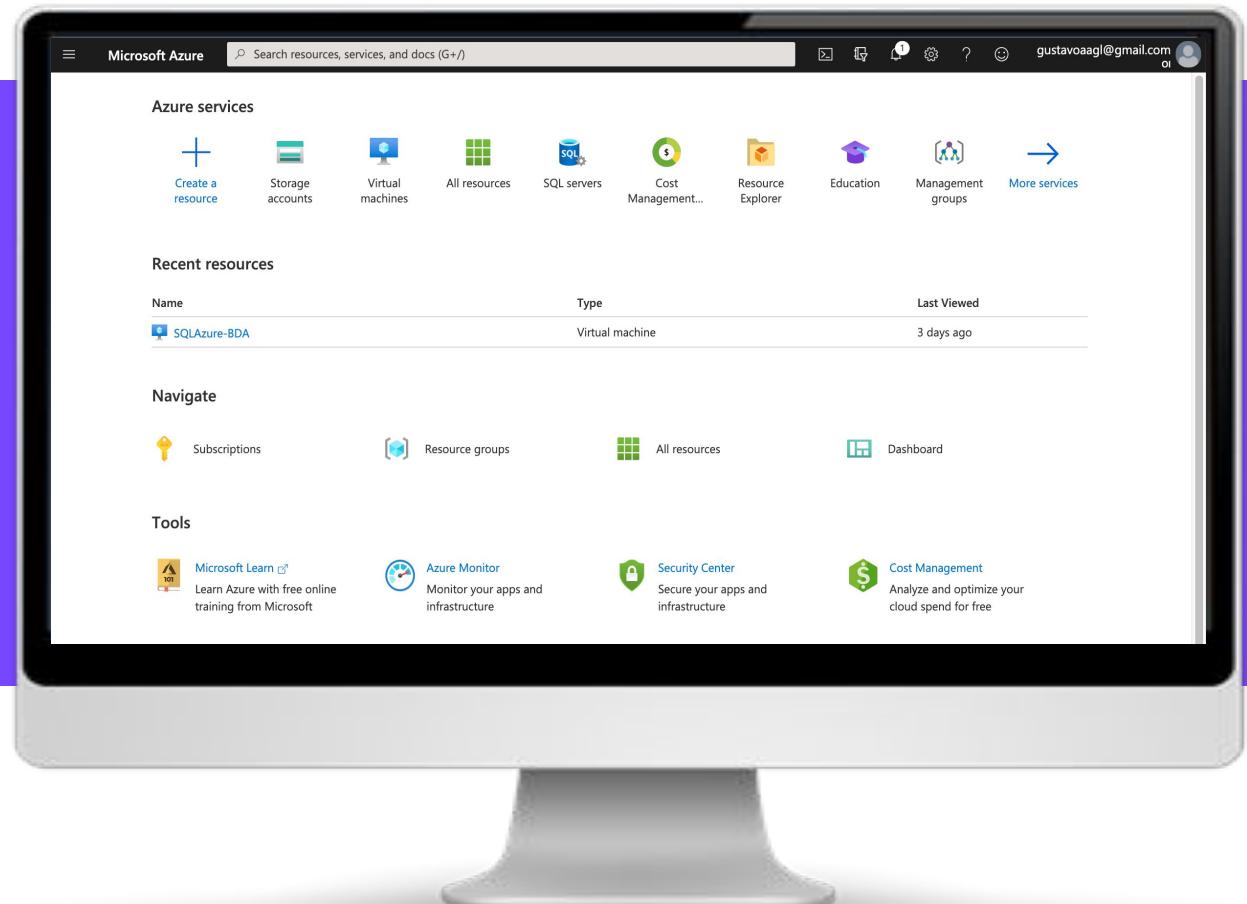
CAPÍTULO 6. AULA 6.2. CRIANDO UM PIPELINE DE DADOS COM O AZURE DATA FACTORY

PROF. GUSTAVO AGUILAR

Criando um Pipeline de Dados com o Azure Data Factory



 Demo



Próxima Aula



- ❑ Capítulo 7 - Soluções de Machine Learning.

Serviços de Armazenamento, Banco de Dados e Analytics

Capítulo 7. Soluções de Machine Learning

PROF. GUSTAVO AGUILAR

Serviços de Armazenamento, Banco de Dados e Analytics

CAPÍTULO 7. AULA 7.1. OVERVIEW DO AZURE MACHINE LEARNING

PROF. GUSTAVO AGUILAR

Nesta Aula



- Introdução ao Aprendizado de Máquina.
- Azure Machine Learning.

Introdução ao Aprendizado de Máquina



- 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.

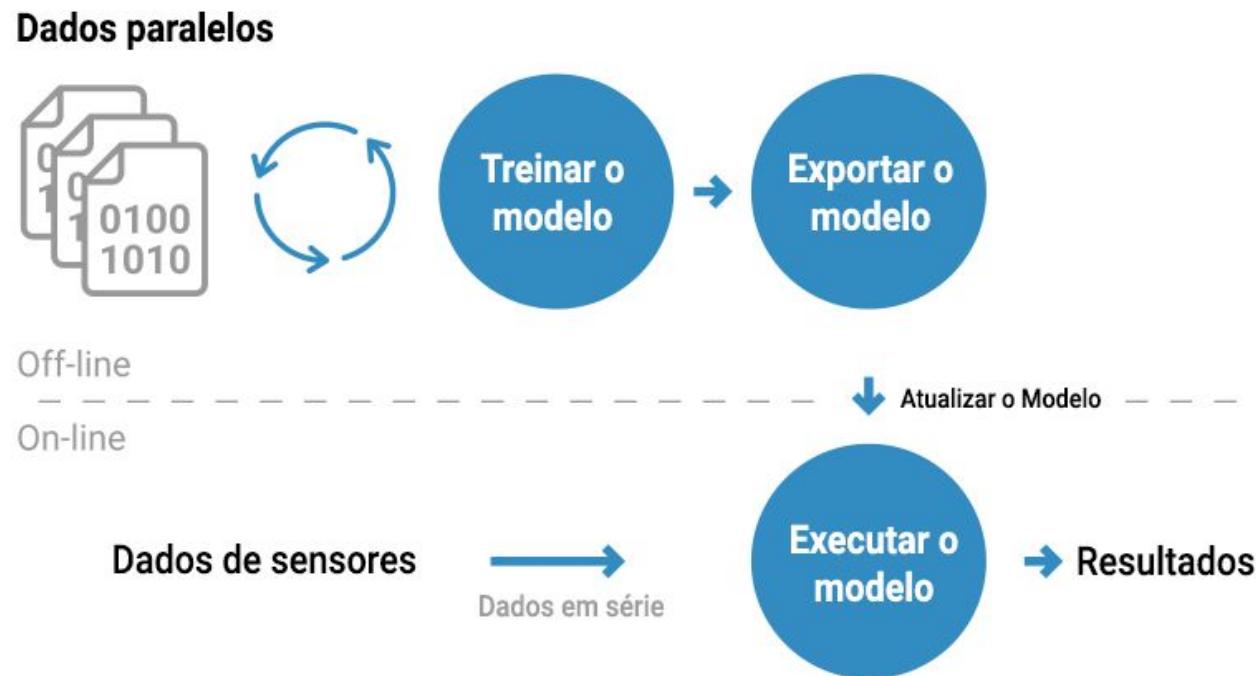
Introdução ao Aprendizado de Máquina

IGTI



Introdução ao Aprendizado de Máquina

IGTI



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.



Azure Machine Learning

IGTI

Microsoft Azure Search resources, services, and docs (G/)

Home > New >

Marketplace

My Saved List
Recently created
Service Providers

Categories

- Get Started
- AI + Machine Learning
- Analytics
- Blockchain
- Compute
- Containers
- Databases
- Developer Tools
- DevOps
- Identity
- Integration
- Internet of Things
- IT & Management Tools

Showing All Results

machine learning

Pricing : All Operating System : All Publisher : All

 Machine Learning Microsoft Enterprise-grade machine learning to build and deploy models faster	 Machine Learning Server Operationalization Microsoft Operationalize analytics with Machine Learning Server	 Weka Machine Learning on Windows 2019 Cloud Infrastructure Services Weka is a collection of machine learning algorithms for data mining tasks	 mahout Mahout machine learning algorithms powered by MIRI Miri Infotech Inc. Apache Mahout is a project of the Apache Software Foundation
 Python AI & Machine Learning Suit (Techlatest.net) TechLatest Save valuable time installing AI/ML for you or your entire team	 Machine Learning Studio (classic) Workspace Microsoft A workspace contains your Machine Learning experiments and predictive web services.	 KoçSistem Azure Machine Learning KoçSistem Bilgi ve İletişim Build, train and deploy machine learning models with Azure Machine Learning Services!	 Machine Learning Studio (classic) Web Service Microsoft Web Service for your machine learning model

Azure Machine Learning

IGTI

Microsoft Azure Search resources, services, and docs (G+) gustavoagl@gmail.com

Home > mlbtcclcmmod3 Machine Learning

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Events

Assets Compute

Settings Properties Locks Export template

Monitoring Alerts Metrics Diagnostic settings Logs

Support + troubleshooting Usage + quotas

Workspace

Download config.json Delete Upgrade

Workspace edition : Basic Resource group : rgbtcclcmmod32020 Location : East US 2 Subscription : MSDN Platforms Subscription ID : 1b294139-ea87-49f9-a09e-a530e697a5f0

Storage : mlbtcclcmmod30891520179 Registry : ... Key Vault : mlbtcclcmmod35752935466 Application Insights : mlbtcclcmmod30663281360

Azure Machine Learning studio

An immersive experience for managing the end-to-end machine learning lifecycle.

Launch now Learn more

Getting Started

View Documentation Learn how to use Azure Machine Learning.

View more samples at GitHub Get inspired by a large collection of machine learning examples.

View Forum Join the discussion of Azure Machine Learning.

Learn about Enterprise Edition (preview) Upgrade this workspace to Enterprise edition (preview) to use UI-based tools for all skill

Azure Machine Learning Studio

iG TI

Microsoft Azure Machine Learning

mlbtcclmod3 > Home

Welcome to the studio!

Create new

Start now

Notebooks

Code with Python SDK and run sample experiments.

Learn more

Automated ML (preview)

Enterprise

Automatically train and tune a model using a target metric.

Learn more

Designer (preview)

Enterprise

Drag-and-drop interface from prepping data to deploying models.

Learn more

Tutorials

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?

Deploy models with Azure Machine Learning

[View all tutorials →](#)

Links

[Blog](#)
Follow us and find updates

[Documentation](#)
Find step-by-step tutorials, concepts, how-to guides, and more

Azure Machine Learning Studio

IGTI

The screenshot shows the Microsoft Azure Machine Learning Studio interface. The left sidebar displays a file tree under 'mlbtcclmod3 > Notebooks'. The 'Sample Notebooks' tab is selected. The main area shows a notebook titled 'ML README.md' with the following content:

```
1  # Notebooks for Microsoft Azure Machine Learning Hardware Accelerated Models SDK
2
3  Easily create and train a model using various deep neural networks (DNNs) as a featurizer for deployment to Azure or a Data Box.
4
5  * ResNet 50
6  * ResNet 152
7  * DenseNet-121
8  * VGG-16
9  * SSD-VGG
10
11 To learn more about the azureml-accel-model classes, see the section [Model Classes](#model-classes) below or the [Azure ML Accelerated Models](https://aka.ms/accelerateAI) documentation.
12
13 ### Step 1: Create an Azure ML workspace
14 Follow [these instructions](https://docs.microsoft.com/en-us/azure/machine-learning/service/setup-create-workspace) to install
15 the Azure ML Accelerated Models SDK.
16
17 ### Step 2: Check your FPGA quota
18 Use the Azure CLI to check whether you have quota.
19
20 ````shell
21 az vm list-usage --location "eastus" -o table
22 `````
23
24 The other locations are ``southeastasia``, ``westeurope``, and ``westus2``.
25
26 Under the "Name" column, look for "Standard PBS Family vCPUs" and ensure you have at least 6 vCPUs under "CurrentValue."
27
28 If you do not have quota, then submit a request form [here](https://aka.ms/accelerateAI).
29
30 ### Step 3: Install the Azure ML Accelerated Models SDK
31 Once you have set up your environment, install the Azure ML Accel Models SDK. This package requires tensorflow >= 1.6,<2.0 to be installed.
32
33 If you already have tensorflow >= 1.6,<2.0 installed in your development environment, you can install the SDK package using:
34
35 `````
36 pip install azureml-accel-models
37 ``````
```

Azure Machine Learning



AZURE MACHINE LEARNING DESIGNER

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

The screenshot shows the Microsoft Azure Machine Learning Designer (preview) interface. The top navigation bar displays the project name "mlbtcclmod3" and the page title "Designer (preview)". On the left, a vertical sidebar contains icons for creating new pipelines, managing modules, and viewing runs. The main area is titled "New pipeline" and features five sample pipeline cards:

- Easy-to-use prebuilt modules
- Sample 1: Regression - Automobile Price Prediction
- Sample 2: Regression - Automobile Price Prediction
- Sample 3: Binary Classification with Feature Selection - Inc...
- Sample 4: Binary Classification with custom Python script - ...

Below these samples, under the "Pipelines" section, there are tabs for "Pipeline drafts" and "Pipeline runs". The "Pipeline drafts" tab is selected, and the message "No pipeline drafts found" is displayed, along with the instruction "Create a new pipeline or start from a sample".

Azure Machine Learning Studio

IGTI

Microsoft Azure Machine Learning

mlbtcclmod3 > Designer (preview) > Authoring

Search by name, tags and description

Autosave on

Submit Publish ...

Not started

Search in canvas

Modules

- Datasets
- Modules**
- Models

Default compute target: testeml
Select compute target

Pipeline parameters: No parameters selected

Default output settings: Select default datastore

Draft details:
Draft name: Sample 1: Regression - Automobile Price Pr...
Draft description (optional): This sample shows how to build a regression model to predict the automobile's price.

Created on: August 16, 2020 12:35 PM
Created by: Gustavo Aguilar de Araújo Gonzaga Lopes
Last edit time:

Navigator

Sample 1: Regression - Automobile Price Prediction (Basic)

The screenshot shows the Microsoft Azure Machine Learning Studio Designer (preview) interface. On the left, there is a sidebar with various icons for datasets, modules, models, and other components. The 'Modules' section is currently selected, showing categories like Data Input and Output, Data Transformation, Feature Selection, Statistical Functions, Machine Learning Algorithms, Model Training, Model Scoring & Evaluation, Python Language, R Language, Text Analytics, Computer Vision, Recommendation, Anomaly Detection, and Web Service. The 'Python Language' category is highlighted. The main workspace displays a pipeline titled 'Sample 1: Regression - Automobile Price Prediction (Basic)'. The pipeline starts with a 'Automobile price data (Raw)' dataset, which is processed by a 'Select Columns in Dataset' module (excluding normalized losses with many values). This is followed by a 'Clean Missing Data' module (removing missing value rows). The pipeline then branches into two parallel paths. One path leads to a 'Linear Regression' module, and the other leads to a 'Split Data' module (splitting the dataset into training set (0.7) and test). The outputs from these two modules converge at a 'Train Model' module. Finally, the output of 'Train Model' is used by a 'Score Model' module. The right side of the screen contains a 'Settings' panel with options for Default compute target (set to testeml), Pipeline parameters (none selected), Default output settings (select default datastore), and Draft details (name: Sample 1: Regression - Automobile Price Pr..., description: This sample shows how to build a regression model to predict the automobile's price). The pipeline was last created on August 16, 2020, at 12:35 PM by Gustavo Aguilar de Araújo Gonzaga Lopes.

Azure Machine Learning Designer

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.

Azure Machine Learning

IGTI

MACHINE LEARNING AUTOMATIZADO

The screenshot shows the Microsoft Azure Machine Learning interface for Automated ML (preview). The left sidebar contains navigation icons for Home, Data, Compute, Experiments, Datasets, Metrics, and Model registration. The main content area displays the 'Automated ML (preview)' page, which includes a message about letting it train models without code, a 'New Automated ML run' button, and a note that no recent runs are displayed. Below this is a 'Documentation' section with links to 'Concept: What is Automated ML?', 'Tutorial: Create your first classification model with Automated ML.', and 'Blog: Build more accurate forecasts with new capabilities in Automated ML.'

Microsoft Azure Machine Learning

mlbtccclmod3 > Automated ML (preview)

Automated ML (preview)

Let Automated ML train and find the best model based on your data without writing a single line of code. [Learn more about Automated ML](#)

+ New Automated ML run

No recent Automated ML runs to display.

Click "New Automated ML run" to create your first run

[Learn more on creating Automated ML runs](#)

Documentation

[View all documentation](#)

Concept: What is Automated ML?

Tutorial: [Create your first classification model with Automated ML.](#)

Blog: [Build more accurate forecasts with new capabilities in Automated ML](#)

Azure Machine Learning

IGTI

MACHINE LEARNING AUTOMATIZADO

Microsoft Azure Machine Learning

mlbtcclmod3 > Automated ML (preview) > Start run

Success: dbtestml dataset created successfully

Create a new Automated ML run

Select dataset

Select dataset from the list below, or create a new dataset. Automated ML currently only supports tabular data for authoring runs.

+ Create dataset | Show supported datasets only | Search to filter items...

Dataset name	Dataset type	Created on	Modified
dbtestml	Tabular	Aug 16, 2020 12:47 PM	Aug 16, 2020 12:47 PM

Back Next Cancel

Dataset name: dbtestml

Dataset type: Tabular

Created on: Aug 16, 2020 12:47 PM

Modified: Aug 16, 2020 12:47 PM

Azure Machine Learning

IGTI

MACHINE LEARNING AUTOMATIZADO

The screenshot shows the Microsoft Azure Machine Learning studio interface. On the left, there's a sidebar with various icons and a main panel titled "Create a new Automated ML run". A success message "Success: dbtestml dataset created successfully" is displayed. In the center, a modal window titled "Create dataset from Open Datasets" is open. The modal has two tabs: "Select Open Dataset" (selected) and "Dataset details". It includes a search bar "Type to filter..." and a list of datasets:

- San Francisco Safety Data**: Fire department calls for service and 311 cases in San Francisco. [Learn more](#)
- Sample: Diabetes**: The Diabetes dataset has 442 samples with 10 features, making it ideal for getting started with machine learning. [Learn more](#)
- US National Employment Hours and Earnings**: The Current Employment Statistics (CES) program produces detailed industry estimates of nonfarm...
- NOAA Global Forecast System (GFS)**: 15-day US hourly weather forecast data (example: temperature, precipitation, wind) produced by the Glob...
- US Labor Force Statistics**: Labor Force Statistics labor force, labor force participation rates, and the civilian noninstitutional population ... [Learn more](#)
- US Consumer Price Index**: The Consumer Price Index (CPI) is a measure of the average change over time in the prices paid by urban consumers for...

At the bottom of the modal are "Back", "Next", and "Cancel" buttons.

Azure Machine Learning

IGTI

MACHINE LEARNING AUTOMATIZADO

Microsoft Azure Machine Learning

mlbtcclmod3 > Automated ML (preview) > Start run

Success: dbtestml dataset created successfully

Create a new Automated ML run

Select task type

Select the machine learning task type for the experiment. Additional settings are available to fine tune the experiment if needed.

Classification
To predict one of several categories in the target column. yes/no, blue, red, green.

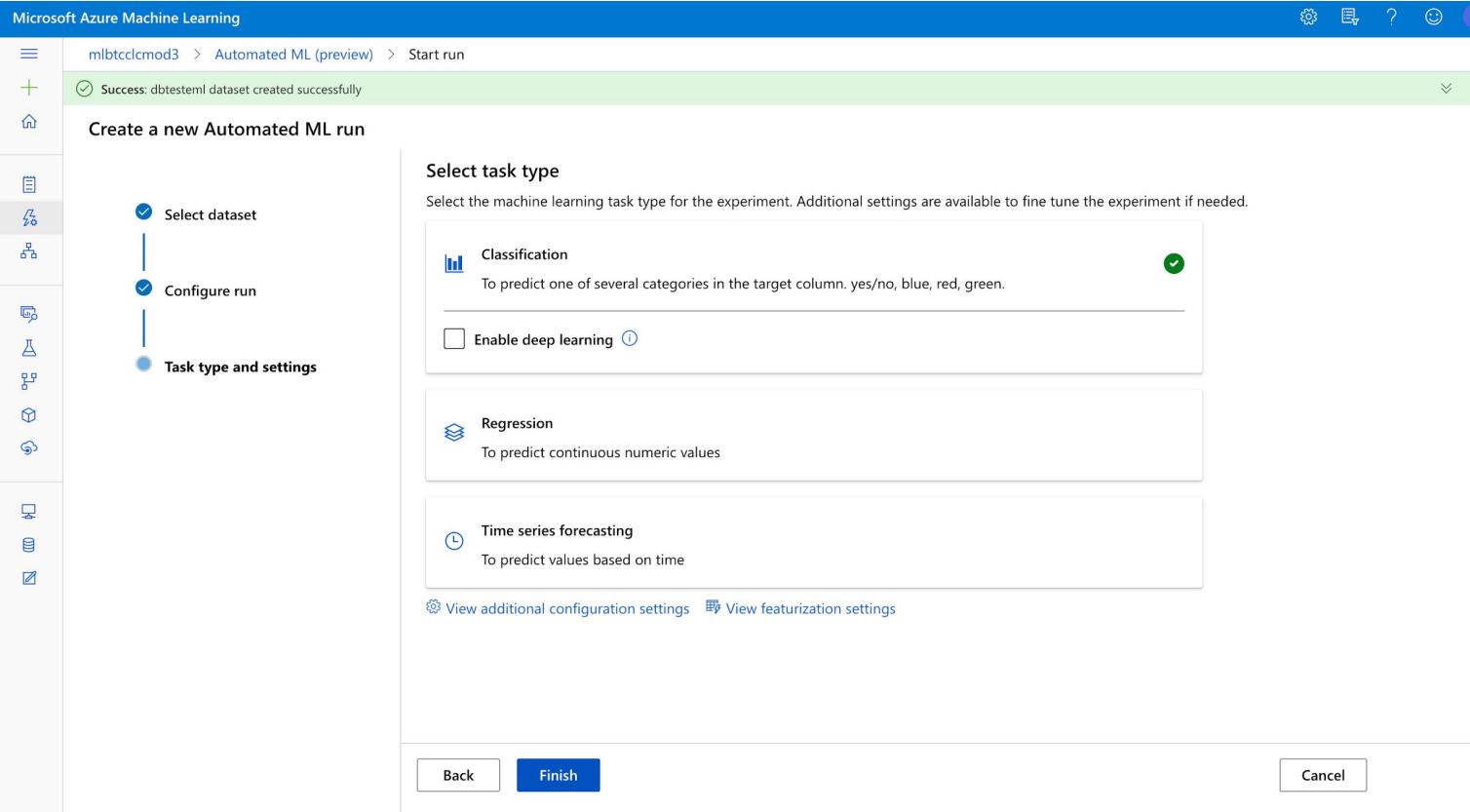
Enable deep learning ⓘ

Regression
To predict continuous numeric values

Time series forecasting
To predict values based on time

[View additional configuration settings](#) [View featurization settings](#)

Back [Finish](#) Cancel



Azure Machine Learning

IGTI

MACHINE LEARNING AUTOMATIZADO

Microsoft Azure Machine Learning

mlbtcclmod3 > Automated ML (preview) > teste > Run 1

Run 1 Completed

Refresh Cancel

Details Data guardrails Models Outputs + Logs Child runs Snapshot

Properties

Status Completed
Created Aug 16, 2020 1:00 PM
Duration 32m 4.193s
Compute target testeml
Run ID AutoML_5eb89d2d-92a0-4abb-8762-5ba73f0efccd
Run number 1
Script name --
Created by Gustavo Aguilar de Araújo Gonzaga Lopes
Input datasets Input name: input_data, ID: 37685d84-ae53-4fe0-b385-5d9521260348
Output datasets None

Best model summary

Algorithm name [VotingEnsemble](#)
Accuracy 0.06121 [View all other metrics](#)
Sampling 100.000 % ⓘ
Registered models No registration yet
Deploy status No deployment yet

Run summary

Task type Classification [View all run settings](#)
Primary metric Accuracy
Run status Completed
Experiment name teste



Serviços de Armazenamento, Banco de Dados e Analytics

PROF. GUSTAVO AGUILAR

FIM.