

# Fundamentos de Dados do Microsoft Azure

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

**DP-900T00-A**

## Índice

|   |     |
|---|-----|
| Módulo 0: Introdução .....  | 3   |
| Módulo 1: Explore os principais conceitos de dados .....              | 5   |
| Módulo 2: Explore dados relacionais no Azure.....                     | 88  |
| Módulo 3: Explore dados não relacionais no Azure.....                 | 161 |
| Módulo 4: Explore análises modernas de data Wharehouse no Azure ..... | 273 |

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## DP-900T00-A - Fundamentos de Dados do Microsoft Azure

### Módulo 0: Introdução

#### Sobre este curso

Bem-vindo a este curso de Fundamentos de dados do Azure!

Este curso foi desenvolvido para quem deseja aprender os fundamentos dos conceitos de banco de dados em um ambiente de nuvem, obter habilidades básicas em serviços de dados em nuvem e construir seu conhecimento básico de serviços de dados em nuvem no Microsoft Azure. O curso fornece uma abordagem prática, na qual você terá a chance de ver os dados em ação e experimentar os serviços de dados do Azure por conta própria.

Os materiais neste livro são projetados para serem usados junto com os módulos online em Microsoft Learn. Ao longo do curso, você encontrará referências aos módulos de Aprendizagem específicos que deve usar para obter experiência prática.

#### Objetivos de aprendizado

Depois de concluir este curso, você será capaz de:

- Descrever os principais conceitos de dados no Azure.
- Explicar os conceitos de dados relacionais no Azure.
- Explicar os conceitos de dados não relacionais no Azure.
- Identificar os componentes de um data warehouse moderno no Azure.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## Agenda do Curso

Este curso inclui os seguintes módulos:

### **Módulo 1: Explore os principais conceitos de dados**

Neste módulo, você explorará as ofertas de dados relacionais, provisionando e implantando bancos de dados relacionais, consultando dados relacionais por meio de soluções de dados em nuvem com o Microsoft Azure.

### **Módulo 2: Explore dados relacionais no Azure**

Neste módulo, você explorará ofertas de dados relacionais, provisionando e implantando bancos de dados relacionais e consultando dados relacionais por meio de soluções de dados em nuvem com o Azure.

### **Módulo 3: Explore dados não relacionais no Azure**

Neste módulo, você explorará ofertas de dados não relacionais, provisionando e implantando bancos de dados não relacionais, bem como seus armazenamentos de com o Microsoft Azure.

### **Módulo 4: Explore análises modernas de data warehouse no Azure**

Neste módulo, você explorará as opções de processamento disponíveis para criar soluções de análise de dados no Azure. Você explorará o Azure Synapse Analytics, o Azure Databricks e o Azure HDInsight para aprender o que é Power BI - incluindo seus blocos de construção e como eles funcionam juntos.

### **Prepare-se para os laboratórios**

Os materiais neste livro são projetados para serem usados junto com os módulos online no Microsoft Learn. Ao longo do curso, você encontrará referências a módulos específicos de Aprendizagem contendo laboratórios que devem ser utilizados para se obter experiência prática.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

## Módulo 1: Explorar os principais conceitos de dados

### Introdução

Nas últimas décadas, a quantidade de dados que sistemas, aplicativos e dispositivos aumentou significativamente. Os dados estão em toda parte, disponíveis em diferentes estruturas e formatos. Compreender os dados e explorá-los revela fatos interessantes, além de ajudar a obter insights significativos.

Nesta lição, você aprenderá como organizar e processar dados. Você aprenderá sobre bancos de dados relacionais e não relacionais, como os dados são tratados por meio do processamento transacional e por meio do processamento de dados em lote e streaming.

Imagine que você é um analista de dados de uma grande organização de consumidores. Você deseja entender os padrões de compra do cliente em supermercados e possui vários conjuntos de dados de diferentes fontes, como informações de caixa (ponto de venda), dados meteorológicos e dados de feriados. Para estes fins, as tecnologias do Azure podem ser utilizadas para entender e analisar os conjuntos de dados.

### Objetivos de aprendizado

Nesta lição você irá:

- Identificar como os dados são definidos e armazenados
- Identificar características de dados relacionais e não relacionais
- Descrever e diferenciar cargas de trabalho de dados
- Descrever e diferenciar dados em lote e streaming

THIAGO FERNANDES DA SILVA  
CPF 00296691208

## Identifique a necessidade de soluções de dados

Os dados agora são mais fáceis de coletar e mais baratos de hospedar, tornando-os acessíveis a quase todas as empresas. As soluções de dados incluem tecnologias e plataformas de software que podem ajudar a facilitar a coleta, análise e armazenamento de informações valiosas.

Todas as empresas gostariam de aumentar suas receitas e ter maiores lucros. No mercado competitivo, os dados são um ativo valioso e, quando analisados adequadamente, podem se transformar em uma riqueza de informações úteis e informar decisões de negócios críticas.

## O que são dados?

Dados são uma coleção de fatos, como números, descrições e observações, usados na tomada de decisões. Você pode classificar os dados como estruturados, semiestruturados ou não estruturados.

Dados estruturados são normalmente dados tabulares, representados por linhas e colunas em um banco de dados. Os bancos de dados que contêm tabelas neste formato são chamados de bancos de dados relacionais (o termo matemático refere-se a um conjunto organizado de dados mantidos como uma tabela). Cada linha de uma tabela possui o mesmo conjunto de colunas.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

A imagem abaixo ilustra um exemplo, mostrando duas tabelas em um banco de dados de comércio eletrônico. A primeira tabela contém os detalhes dos clientes de uma organização e a segunda contém informações sobre os produtos que a organização vende.

| CustomerID | Title | FirstName   | MiddleName | LastName  | Suffix | CompanyName                  | Phone               |
|------------|-------|-------------|------------|-----------|--------|------------------------------|---------------------|
| 1          | Mr.   | Orlando     | N.         | Gee       | NULL   | A Bike Store                 | 245-555-0173        |
| 2          | Mr.   | Keith       | NULL       | Harris    | NULL   | Progressive Sports           | 170-555-0127        |
| 3          | Ms.   | Donna       | F.         | Cameras   | NULL   | Advanced Bike Components     | 279-555-0130        |
| 4          | Ms.   | Janet       | M.         | Gates     | NULL   | Modular Cycle Systems        | 710-555-0173        |
| 5          | Mr.   | Lucy        | NULL       | Hemington | NULL   | Metropolitan Spots Supply    | 828-555-0186        |
| 6          | Ms.   | Rosanne     | J.         | Carroll   | NULL   | Aerobic Exercise Company     | 244-555-0112        |
| 7          | Mr.   | Dominic     | P.         | Gash      | NULL   | Associated Bikes             | 192-555-0173        |
| 10         | Ms.   | Kathleen    | M.         | Gerza     | NULL   | Rural Cycle Emporium         | 150-555-0127        |
| 11         | Ms.   | Katherine   | NULL       | Harding   | NULL   | Sharp Bikes                  | 926-555-0159        |
| 12         | Mr.   | Johnny      | A.         | Caprio    | Jr.    | Bikes and Motorbikes         | 112-555-0191        |
| 16         | Mr.   | Christopher | R.         | Beck      | Jr.    | Bulk Discount Store          | 1 (11) 500 555-0132 |
| 18         | Mr.   | David       | J.         | Liu       | NULL   | Catalog Store                | 440-555-0132        |
| 19         | Mr.   | John        | A.         | Beaver    | NULL   | Center Cycle Shop            | 521-555-0195        |
| 20         | Ms.   | Jean        | P.         | Handley   | NULL   | Central Discount Store       | 582-555-0113        |
| 21         | N...  | Jinghao     | NULL       | Liu       | NULL   | Chic Department Stores       | 928-555-0116        |
| 22         | Ms.   | Linda       | E.         | Burnett   | NULL   | Travel Systems               | 121-555-0121        |
| 23         | Mr.   | Kenn        | NULL       | Hanif     | NULL   | Bike World                   | 216-555-0122        |
| 24         | Mr.   | Kevin       | NULL       | Liu       | NULL   | Eastside Department Store    | 926-555-0164        |
| 25         | Mr.   | Donald      | L.         | Banton    | NULL   | Coalition Bike Company       | 357-555-0161        |
| 28         | Ms.   | Jackie      | E.         | Blackwell | NULL   | Computer Bicycle Store       | 972-555-0163        |
| 29         | Mr.   | Bryan       | NULL       | Hamilton  | NULL   | Cross-Country Riding Supp... | 344-555-0144        |
| 30         | Mr.   | Todd        | R.         | Logan     | NULL   | Cycle Merchants              | 783-555-0110        |
| 34         | Ms.   | Barbara     | J.         | Geman     | NULL   | Cycles Wholesaler & Mfg.     | 1 (11) 500 555-0181 |
| 37         | Mr.   | Jim         | NULL       | Grat      | NULL   | Two Bike Shops               | 724-555-0161        |

| ProductID | Name                        | ProductNumber | Color | StandardCost | ListPrice | Size | Weight  |
|-----------|-----------------------------|---------------|-------|--------------|-----------|------|---------|
| 680       | HL Road Frame - Black, 58   | FR-R92B-58    | Black | 1059.31      | 1431.50   | 58   | 1016.04 |
| 706       | HL Road Frame - Red, 58     | FR-R92R-58    | Red   | 1059.31      | 1431.50   | 58   | 1016.04 |
| 707       | Sport-100 Helmet, Red       | HL-U509-R     | Red   | 13.0863      | 34.99     | NULL | NULL    |
| 708       | Sport-100 Helmet, Black     | HL-U509       | Black | 13.0863      | 34.99     | NULL | NULL    |
| 709       | Mountain Bike Socks, M      | SO-B909-M     | White | 3.3963       | 9.50      | M    | NULL    |
| 710       | Mountain Bike Socks, L      | SO-B909-L     | White | 3.3963       | 9.50      | L    | NULL    |
| 711       | Sport-100 Helmet, Blue      | HL-U509-B     | Blue  | 13.0863      | 34.99     | NULL | NULL    |
| 712       | AWC Logo Cap                | CA-1098       | Multi | 6.9223       | 8.99      | NULL | NULL    |
| 713       | Long-Sleeve Logo Jersey, S  | LJ-0192-S     | Multi | 38.4923      | 49.99     | S    | NULL    |
| 714       | Long-Sleeve Logo Jersey, M  | LJ-0192-M     | Multi | 38.4923      | 49.99     | M    | NULL    |
| 715       | Long-Sleeve Logo Jersey, L  | LJ-0192-L     | Multi | 38.4923      | 49.99     | L    | NULL    |
| 716       | Long-Sleeve Logo Jersey, XL | LJ-0192-X     | Multi | 38.4923      | 49.99     | XL   | NULL    |
| 717       | HL Road Frame - Red, 62     | FR-R92R-62    | Red   | 868.6342     | 1431.50   | 62   | 1043.26 |
| 718       | HL Road Frame - Red, 44     | FR-R92R-44    | Red   | 868.6342     | 1431.50   | 44   | 961.61  |
| 719       | HL Road Frame - Red, 48     | FR-R92R-48    | Red   | 868.6342     | 1431.50   | 48   | 979.75  |
| 720       | HL Road Frame - Red, 52     | FR-R92R-52    | Red   | 868.6342     | 1431.50   | 52   | 957.90  |
| 721       | HL Road Frame - Red, 56     | FR-R92R-56    | Red   | 868.6342     | 1431.50   | 56   | 1016.04 |
| 722       | LL Road Frame - Black, 58   | FR-R38B-58    | Black | 204.6251     | 337.22    | 58   | 1115.83 |
| 723       | LL Road Frame - Black, 60   | FR-R38B-60    | Black | 204.6251     | 337.22    | 60   | 1124.90 |
| 724       | LL Road Frame - Black, 62   | FR-R38B-62    | Black | 204.6251     | 337.22    | 62   | 1133.98 |
| 725       | LL Road Frame - Red, 44     | FR-R38R-44    | Red   | 187.1571     | 337.22    | 44   | 1052.33 |
| 726       | LL Road Frame - Red, 48     | FR-R38R-48    | Red   | 187.1571     | 337.22    | 48   | 1070.47 |
| 727       | LL Road Frame - Red, 52     | FR-R38R-52    | Red   | 187.1571     | 337.22    | 52   | 1088.62 |

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

Dados semi-estruturados são informações que não residem em um banco de dados relacional, mas ainda possuem alguma estrutura. Os exemplos incluem documentos mantidos em formato Java Script Object Notation (JSON).

O exemplo abaixo mostra um par de documentos que representam as informações do cliente. Em ambos os casos, cada documento do cliente inclui documentos filho contendo o nome e endereço – porém, os campos nos documentos variam entre os clientes.

```
## Document 1 ##
{
  "customerID": "103248",
  "name":
  {
    "first": "AAA",
    "last": "BBB"
  },
  "address":
  {
    "street": "Main Street",
    "number": "101",
    "city": "Acity",
    "state": "NY"
  },
  "ccOnFile": "yes",
  "firstOrder": "02/28/2003"
}

## Document 2 ##
{
  "customerID": "103249",
  "name":
  {
    "title": "Mr",
    "forename": "AAA",
    "lastname": "BBB"
  },
  "address":
  {
    "street": "Another Street",
    "number": "202",
    "city": "Bcity",
    "county": "Gloucestershire",
    "country-region": "UK"
  },
  "ccOnFile": "yes"
}
```

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Existem outros tipos de dados semiestruturados também. Exemplos incluem valor-chave, lojas e gráficos bancos de dados. Um armazenamento de valor-chave é semelhante a uma tabela relacional. A diferença é que cada linha pode ter qualquer número de colunas.

Você pode usar um banco de dados gráfico para armazenar e consultar informações sobre relacionamentos complexos. Um gráfico contém nós (informações sobre objetos) e arestas (informações sobre as relações entre os objetos). A imagem abaixo mostra um exemplo de como você pode estruturar os dados em um banco de dados gráfico.



Nem todos os dados são estruturados, ou mesmo semiestruturados. Por exemplo, arquivos de áudio e vídeo e arquivos de dados binários podem não ter uma estrutura específica.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

### Como os dados são definidos e armazenados na computação em nuvem?

Dependendo do tipo de dados, como estruturados, semiestruturados ou não estruturados, os dados serão armazenados de forma diferente. Os dados estruturados são normalmente armazenados em um banco de dados relacional, como SQL Server ou Banco de Dados SQL do Azure.

O Banco de Dados SQL do Azure é um serviço executado na nuvem. Você pode utilizá-lo para criar e acessar tabelas relacionais. O serviço é gerenciado e executado pelo Azure, você apenas especifica que deseja que um servidor de banco de dados seja criado. O ato de configurar o servidor de banco de dados é chamado de provisionamento.

Você também pode provisionar outros serviços no Azure. Por exemplo, se você deseja armazenar dados não estruturados, como arquivos de vídeo ou áudio, você pode usar o armazenamento de Blob do Azure. Se você deseja armazenar dados semiestruturados, como documentos, pode usar um serviço como o Azure Cosmos DB.

Depois que seu serviço é provisionado, o serviço precisa ser configurado para que os usuários possam ter acesso aos dados. Normalmente, você pode definir vários níveis de acesso.

Somente leitura acesso significa que os usuários podem ler os dados, mas não podem modificar os dados existentes ou criar novos dados.

Ler escrever o acesso dá aos usuários a capacidade de visualizar e modificar os dados existentes.

Proprietário privilégio dá acesso total aos dados, incluindo gerenciamento de segurança, como adicionar novos usuários e remover o acesso aos usuários existentes.

Você também pode definir quais usuários devem ter permissão para acessar os dados em primeiro lugar. Se os dados forem confidenciais (ou secretos), você pode restringir o acesso a alguns usuários selecionados.

TITULO Fundamentos do Azure

CPF: 00296691208

## Descrever soluções de processamento de dados

As soluções de processamento de dados geralmente se enquadram em uma de duas categorias amplas: sistemas analíticos e sistemas de processamento de transações.

### O que é um sistema transacional?

Um sistema transacional é geralmente o que a maioria das pessoas considera a função primária da computação de negócios. Um sistema transacional registra transações.

Uma transação pode ser financeira, como a movimentação de dinheiro entre contas em um sistema bancário, ou pode ser parte de um sistema de varejo, rastreando pagamentos de mercadorias e serviços de clientes. Pense em uma transação como uma unidade de trabalho pequena e discreta.

Os sistemas transacionais costumam ser de alto volume, às vezes lidando com muitos milhões de transações em um único dia. Os dados, sendo processados, devem estar acessíveis muito rapidamente. O trabalho realizado por sistemas transacionais é frequentemente referido como Processamento Transacional Online (OLTP).

Para suportar o processamento rápido, os dados em um sistema transacional costumam ser divididos em pequenos pedaços. Por exemplo, se você estiver usando um sistema relacional, cada tabela envolvida em uma transação contém apenas as colunas necessárias para executar a tarefa transacional.

No exemplo da transferência bancária, uma tabela contendo informações sobre os fundos da conta pode conter apenas o número da conta e o saldo atual. Outras tabelas não envolvidas na operação de transferência conteriam informações como o nome e endereço do cliente e o histórico da conta. A divisão de tabelas em grupos separados de colunas é chamada de normalizada. A próxima unidade discute esse processo com mais detalhes.

A normalização pode permitir que um sistema transacional armazene em cache muitas das informações necessárias, para realizar transações na memória e acelerar o rendimento.

Embora a normalização permita um rendimento rápido para transações, ela pode tornar a consulta mais complexa. As consultas que envolvem tabelas normalizadas frequentemente precisarão reunir os dados mantidos em várias tabelas novamente. Isso pode dificultar os usuários de negócios que precisam examinar os dados.

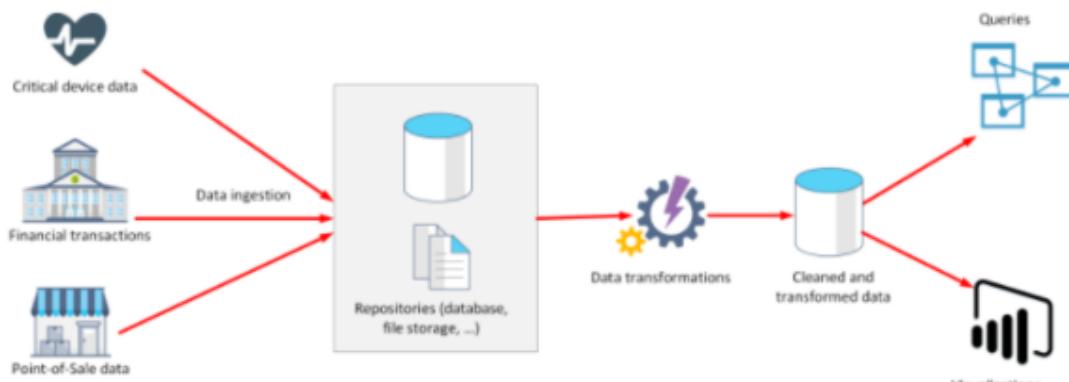
CNPJ 00296691208

## O que é um sistema analítico?

Em contraste com os sistemas projetados para suportar OLTP, um sistema analítico é projetado para oferecer suporte aos usuários de negócios que precisam consultar dados e obter uma ampla visualização das informações mantidas em um banco de dados.

Os sistemas analíticos estão preocupados em capturar dados brutos e usá-los para gerar insights. Uma organização pode usar esses insights para tomar decisões de negócios. Por exemplo, insights detalhados para uma empresa de manufatura podem indicar tendências, permitindo-lhes determinar em quais linhas de produtos se concentrarem, para lucratividade.

A maioria dos sistemas de processamento de dados analíticos precisa realizar tarefas semelhantes: ingestão de dados, transformação de dados, consulta de dados e visualização de dados. A imagem abaixo ilustra os componentes de um sistema típico de processamento de dados.



**Ingestão de Dados:** A ingestão de dados é o processo de captura dos dados brutos. Esses dados podem ser obtidos de dispositivos de controle que medem informações ambientais, como temperatura e pressão, dispositivos de ponto de venda que registram os itens comprados por um cliente em um supermercado, dados financeiros que registram o movimento de dinheiro entre contas bancárias e dados meteorológicos de estações.

Alguns desses dados podem vir de um sistema OLTP separado. Para processar e analisar esses dados, você deve primeiro armazenar os dados em um repositório de algum tipo. O repositório pode ser um armazenamento de arquivos, um banco de dados de documentos ou até mesmo um banco de dados relacional.

**Transformação / processamento de dados:** Os dados brutos podem não estar em um formato adequado para consulta. Os dados podem conter anomalias que devem ser filtradas ou podem exigir uma transformação de alguma forma. Por exemplo, datas ou endereços podem precisar ser convertidos em um formato padrão. Depois que os

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

dados são inseridos em um repositório de dados, você pode desejar fazer algumas operações de limpeza e remover quaisquer dados questionáveis ou inválidos, ou realizar algumas agregações, como cálculo de lucro, margem e outras Métricas de Desempenho Chave (KPIs). KPIs são como as empresas são avaliadas quanto ao crescimento e desempenho.

- Consulta de dados: Depois que os dados são inseridos e transformados, você pode consultar os dados para analisá-los. Você pode estar procurando tendências ou tentando determinar a causa dos problemas em seus sistemas. Muitos sistemas de gerenciamento de banco de dados fornecem ferramentas para permitir que você execute consultas ad hoc em seus dados e gere relatórios regulares.
- Visualização de dados: Os dados representados em tabelas, como linhas e colunas, ou como documentos, nem sempre são intuitivos. A visualização dos dados geralmente pode ser útil como uma ferramenta para examinar os dados. Você pode gerar gráficos como gráficos de barras, gráficos de linha, traçar resultados em mapas geográficos, gráficos de pizza ou ilustrar como os dados mudam ao longo do tempo. A Microsoft oferece ferramentas de visualização como o Power BI para fornecer uma representação gráfica rica de seus dados.

### Identifique os tipos de dados e armazenamento de dados

Você pode categorizar os dados de muitas maneiras diferentes, dependendo não apenas de como estão estruturados, mas também de como os dados são usados. Nesta unidade, você aprenderá sobre as características de diferentes tipos de dados.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

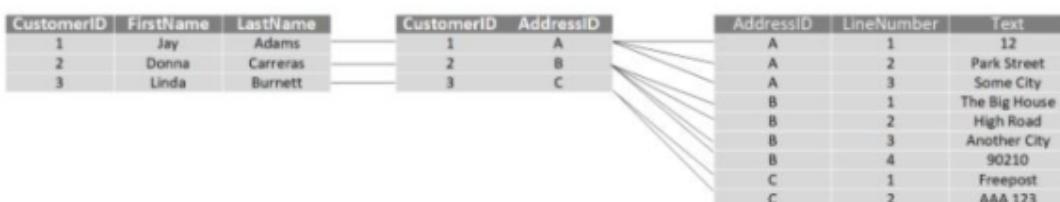
## Descreva as características dos dados relacionais e não relacionais

Os bancos de dados relacionais fornecem provavelmente o modelo mais bem compreendido para armazenar dados. A estrutura simples de tabelas e colunas torna-as fáceis de usar inicialmente, mas a estrutura rígida pode causar alguns problemas.

Por exemplo, em um banco de dados contendo informações de clientes, como você lida com clientes que têm mais de um endereço? Você adiciona colunas para conter os detalhes de cada endereço? Em caso afirmativo, quantas dessas colunas você deve adicionar? Se você permitir três endereços, o que acontecerá se um cliente tiver apenas um endereço? O que você armazena no poupar colunas? O que acontecerá se você de repente tiver um cliente com quatro endereços? Da mesma forma, quais informações você armazena em um endereço (nome da rua, número da casa, cidade, código postal)? O que acontece se uma casa tiver um nome em vez de um número, ou se estiver localizada em algum lugar que não use códigos postais?

Você pode resolver esses problemas usando um processo chamado normalização. Normalmente, o resultado final do processo de normalização é que seus dados são divididos em um grande número de tabelas estreitas e bem definidas (um “limitar tabela” é uma tabela com poucas colunas), com referências de uma tabela para outra, conforme mostra a imagem abaixo.

No entanto, consultar os dados geralmente requer a remontagem de informações de várias tabelas, juntando os dados novamente em tempo de execução (ilustrado pelas linhas no diagrama). Esses tipos de consultas podem ser caros.



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Bancos de dados não relacionais permitem que você armazene dados em um formato que corresponda mais de perto à estrutura original.

Por exemplo, em um banco de dados de documentos, você pode armazenar os detalhes de cada cliente em um único documento, conforme mostrado pelo exemplo na unidade anterior. Recuperar os detalhes de um cliente, incluindo o endereço, é uma questão de ler um único documento.

No entanto, existem algumas desvantagens em usar um banco de dados de documentos. Se dois clientes coabitam e têm o mesmo endereço, em um banco de dados relacional, você só precisa armazenar as informações de endereço uma vez. No diagrama abaixo, Jay e Frances Adams compartilham o mesmo endereço.



Em um banco de dados de documentos, o endereço seria duplicado nos documentos de Jay e Francis Adams. Essa duplicação não só aumenta o armazenamento necessário, mas também pode tornar a manutenção mais complexa (se o endereço mudar, você deve modificá-lo em dois documentos).

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

```
## Document for Jay Adams ##
{
  "customerID": "1",
  "name":
  {
    "firstname": "Jay",
    "lastname": "Adams"
  },
  "address":
  {
    "number": "12",
    "street": "Park Street",
    "city": "Some City",
  }
}

## Document for Frances Adams ##
{
  "customerID": "4",
  "name":
  {
    "firstname": "Francis",
    "lastname": "Adams"
  },
  "address":
  {
    "number": "12",
    "street": "Park Street",
    "city": "Some City",
  }
}
```

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 000296691208

## Descreva cargas de trabalho transacionais

Bancos de dados relacionais e não relacionais são adequados para diferentes cargas de trabalho. Um uso principal dos bancos de dados relacionais é lidar com o processamento de transações.

Uma transação é uma sequência de operações que são atômicas. Isso significa que todas as operações na sequência devem ser concluídas com êxito ou, se algo der errado, todas as operações executadas até o momento na sequência devem ser desfeitas. As transferências bancárias são um bom exemplo; você deduz fundos de uma conta e credita os fundos equivalentes em outra conta. Se o sistema falhar após a dedução dos fundos, eles devem ser restabelecidos na conta original (eles não devem ser perdidos). Você pode então tentar realizar a transferência novamente. Da mesma forma, você não deve ser capaz de creditar uma conta duas vezes com os mesmos fundos.

Cada transação do banco de dados tem um ponto inicial definido (seguido por etapas para modificar os dados no banco de dados). No final, o banco de dados confirma as alterações para torná-las permanentes, ou reverte as alterações para o ponto inicial, quando a transação pode ser tentada novamente.

Um banco de dados transacional deve aderir às propriedades ACID (Atomicity, Consistency, Isolation, Durability) para garantir que o banco de dados permaneça consistente durante o processamento de transações.

Atomicidade garante que cada transação seja tratada como uma única unidade, que é bem-sucedida ou falha totalmente. Se alguma das instruções que constituem uma transação não for concluída, a transação inteira falhará e o banco de dados permanecerá inalterado. Um sistema atômico deve garantir a atomicidade em todas as situações, incluindo falhas de energia, erros e travamentos.

Consistênciagarante que uma transação só pode levar os dados no banco de dados de um estado válido para outro. Um banco de dados consistente nunca deve perder ou criar dados de uma maneira que não pode ser contabilizada. No exemplo de transferência bancária, descrito anteriormente, se você adicionar fundos a uma conta, deve haver uma dedução correspondente de fundos em algum lugar, ou um registro que descreva de onde os fundos vieram (caso tenham sido recebidos externamente). Você não pode criar (ou perder) dinheiro de repente.

O isolamento garante que a execução simultânea de transações deixe o banco de dados no mesmo estado que teria sido obtido se as transações fossem executadas sequencialmente. Um processo simultâneo não pode visualizar os dados em um estado inconsistente (por exemplo, os fundos foram deduzidos de uma conta, mas ainda não foram creditados em outra).

CPF 00296691208

Durabilidade garante que uma vez que uma transação foi confirmada, ela permanecerá confirmada mesmo se houver uma falha do sistema, como queda de energia ou travamento.

Os sistemas de banco de dados que processam cargas de trabalho transacionais são inherentemente complexos. Eles precisam gerenciar usuários simultâneos, que possivelmente tentam acessar e modificar os mesmos dados ao mesmo tempo, processando as transações de forma isolada, mantendo o banco de dados consistente e recuperável.

Muitos sistemas implementam consistência relacional e isolamento, aplicando bloqueios aos dados quando eles são atualizados. O bloqueio impede que outro processo leia os dados até que o bloqueio seja liberado. O bloqueio só é liberado quando a transação é confirmada ou revertida. O bloqueio extensivo pode levar a um desempenho ruim, enquanto os aplicativos aguardam que os bloqueios sejam liberados.

Bancos de dados distribuídos são amplamente usados em muitas organizações. Um banco de dados distribuído é um banco de dados no qual os dados são armazenados em diferentes locais físicos. Ele pode ser mantido em vários computadores localizados no mesmo local físico (por exemplo, um datacenter) ou pode ser disperso em uma rede de computadores interconectados.

Quando comparado a sistemas de banco de dados não distribuídos, qualquer atualização de dados em um banco de dados distribuído levará tempo para ser aplicada em vários locais. Se você precisar de consistência transacional neste cenário, os bloqueios podem ser retidos por muito tempo, especialmente se houver uma falha de rede entre os bancos de dados em um momento crítico.

Para combater este problema, muitos sistemas de gerenciamento de banco de dados distribuídos relaxam os requisitos de isolamento estrito de transações e implementam "consistência eventual". Nessa forma de consistência, à medida que um aplicativo grava dados, cada alteração é registrada por um servidor e, em seguida, propagada para os outros servidores no sistema de banco de dados distribuído de forma assíncrona. Embora essa estratégia ajude a minimizar a latência, pode levar a inconsistências temporárias nos dados. A consistência eventual é ideal quando o aplicativo não exige nenhuma garantia de pedido. Os exemplos incluem contagens de compartilhamentos, curtidas ou comentários não encadeados em um sistema de mídia social.

CPF: 00296691208

## Descreva cargas de trabalho analíticas

Cargas de trabalho analíticas são normalmente sistemas somente leitura, que armazenam grandes volumes de dados históricos ou métricas de negócios, como desempenho de vendas e níveis de estoque. São utilizadas para análise de dados e tomada de decisão. As análises são geradas agregando os fatos apresentados pelos dados brutos em resumos, tendências e outros tipos de "Informação de negócios."

A análise pode ser baseada em um instantâneo dos dados em um determinado momento ou em uma série de instantâneos. Pessoas que estão em posições mais altas na hierarquia da empresa geralmente não exigem todos os detalhes de cada transação. Eles querem a imagem maior.

Um exemplo de informação analítica é um relatório de vendas mensais. Como chefe do departamento de vendas, você pode não precisar ver todas as transações diárias que ocorreram (informações transacionais), mas definitivamente gostaria de um relatório de vendas mensal para identificar tendências e tomar decisões (informações analíticas).

As informações transacionais, no entanto, são parte integrante das informações analíticas. Se você não tiver bons registros de vendas diárias, não poderá compilar um relatório útil para identificar tendências. É por isso que o manuseio eficiente de informações transacionais é muito importante.

## Descreva a diferença entre dados em lote e streaming

O processamento de dados é simplesmente a conversão de dados brutos em informações significativas, por meio de um processo. Dependendo de como os dados são inseridos em seu sistema, você pode processar cada item de dados conforme eles chegam ou armazenar em buffer os dados brutos e processá-los em grupos. O processamento de dados conforme eles chegam é chamado de transmissão. O armazenamento em buffer e o processamento dos dados em grupos são chamados processamento em lote.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Entenda o processamento em lote

No processamento em lote, os elementos de dados recém-chegados são coletados em um grupo. Todo o grupo é então processado em um momento futuro como um lote. O exato momento em que cada grupo é processado pode ser determinado de várias maneiras. Por exemplo, você pode processar dados com base em um intervalo de tempo programado (por exemplo, a cada hora) ou pode ser acionado quando uma certa quantidade de dados chega, ou como resultado de algum outro evento.

Um exemplo de processamento em lote é a maneira como as empresas de cartão de crédito lidam com o faturamento. O cliente não recebe uma fatura para cada compra separada com cartão de crédito, mas uma fatura mensal para todas as compras daquele mês.

As vantagens do processamento em lote incluem:

- Grandes volumes de dados podem ser processados em um momento conveniente.
- Ele pode ser programado para ser executado em um momento em que os computadores ou sistemas possam estar ociosos, como durante a noite ou fora do horário de pico.

As desvantagens do processamento em lote incluem:

- O atraso de tempo entre a ingestão dos dados e a obtenção dos resultados.

Todos os dados de entrada de uma tarefa em lote devem estar prontos antes que um lote possa ser processado. Isso significa que os dados devem ser verificados cuidadosamente. Problemas com dados, erros e travamentos do programa que ocorrem durante as tarefas em lote interrompem todo o processo. Os dados de entrada devem ser verificados cuidadosamente antes que o trabalho possa ser executado novamente. Mesmo pequenos erros de dados, como erros tipográficos em datas, podem impedir a execução de um trabalho em lote.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Entenda streaming e dados em tempo real

No processamento de fluxo, cada novo pedaço de dados é processado quando chega. Por exemplo, a ingestão de dados é inherentemente um processo de streaming.

O streaming lida com dados em tempo real. Ao contrário do processamento em lote, não há espera até o próximo intervalo de processamento em lote e os dados são processados como peças individuais em vez de serem processados um lote de cada vez. O processamento de dados de streaming é benéfico na maioria dos cenários em que dados novos e dinâmicos são gerados continuamente.

Exemplos de dados de streaming incluem:

- Uma instituição financeira rastreia as mudanças no mercado de ações em tempo real, calcula o valor em risco e reequilibra automaticamente as carteiras com base nos movimentos dos preços das ações.
- Uma empresa de jogos online coleta dados em tempo real sobre as interações jogador-jogo e os alimenta em sua plataforma de jogos. Em seguida, analisa os dados em tempo real, oferece incentivos e experiências dinâmicas para envolver seus jogadores.
- Um site de imobiliárias que rastreia um subconjunto de dados de dispositivos móveis de consumidores e faz recomendações de propriedades em tempo real para visitar com base em sua geolocalização.

O processamento de fluxo é ideal para operações de tempo crítico que requerem uma resposta instantânea em tempo real. Por exemplo, um sistema que monitora a fumaça e o calor de um prédio precisa disparar alarmes e destrancar portas, permitindo que os residentes escapem imediatamente em caso de incêndio.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

### Entenda as diferenças entre dados em lote e streaming

Além da maneira como o processamento em lote e o processamento de streaming lidam com os dados, existem outras diferenças:

- Escopo de Dados: Os dados em lote podem processar todos os dados no conjunto de dados. O processamento de fluxo normalmente só tem acesso aos dados mais recentes recebidos, ou dentro de uma janela de tempo de rolagem (os últimos 30 segundos, por exemplo).
- Tamanho dos Dados: Os dados em lote são adequados para lidar com grandes conjuntos de dados com eficiência. O processamento de fluxo é destinado a registros individuais ou microlotes, consistindo em poucos registros.
- Atuação: A latência para processamento em lote é normalmente de algumas horas. O processamento de fluxo normalmente ocorre imediatamente, com latência da ordem de segundos ou milissegundos (latência é o tempo que leva para os dados serem recebidos e processados).
- Análise: Normalmente, você usa o processamento em lote para realizar análises complexas. O processamento de fluxo é usado para funções de resposta simples, agregações ou cálculos, como médias contínuas.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Teste de conhecimento

### Questão 1

**Como os dados em uma tabela relacional são organizados?**

- a) Linhas e colunas
- b) Cabeçalho e rodapé
- c) Páginas e parágrafos

**Explicação:** está correto. Dados estruturados são tipicamente dados tabulares representados por linhas e colunas em uma tabela de banco de dados.

### Questão 2

**Qual das alternativas a seguir é um exemplo de dados não estruturados?**

- a) Uma tabela Employee com as colunas Employee ID, Employee Name e Employee Designation
- b) Arquivos de áudio e vídeo
- c) Uma tabela no banco de dados SQL Server

**Explicação:** está correto. Arquivos de áudio e vídeo são dados não estruturados.

### Questão 3

**Qual das alternativas a seguir é um exemplo de conjunto de dados de streaming?**

- a) Dados de sensores e dispositivos
- b) Dados de vendas do último mês
- c) Lista de funcionários que trabalham para uma empresa

**Explicação:** está correto. Feeds de sensores e dispositivos são exemplos de conjuntos de dados de streaming conforme são publicados continuamente.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Resumo

O Microsoft Azure fornece uma variedade de tecnologias para armazenar dados relacionais e não relacionais. Cada tecnologia tem seus próprios pontos fortes e é adequada a cenários específicos.

Nesta lição, você aprendeu como:

- Identificar como os dados são definidos e armazenados
- Identificar características de dados relacionais e não relacionais
- Descrever e diferenciar cargas de trabalho de dados
- Descrever e diferenciar dados em lote e streaming

Saber mais

- Introdução ao Banco de Dados SQL do Azure
- Introdução ao armazenamento de Blob do Azure
- Introdução ao Azure Cosmos DB
- Descrição dos fundamentos de normalização do banco de dados

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Introdução

Na última década, a quantidade de dados que os sistemas e dispositivos geram aumentou significativamente. Por causa desse aumento, novas tecnologias, funções e abordagens para trabalhar com dados estão afetando os profissionais de dados. Profissionais de dados normalmente desempenham funções diferentes ao gerenciar, usar e controlar dados. Neste módulo, você aprenderá sobre as várias funções que as organizações costumam aplicar aos profissionais de dados e as tarefas e responsabilidades associadas a essas funções.

## Objetivos de aprendizado

Nesta lição você irá:

- Explorar funções de trabalho de dados
- Explorar tarefas e ferramentas comuns para funções de trabalho de dados

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Explore cargos de trabalho no mundo dos dados

Há uma grande variedade de funções envolvidas no gerenciamento, controle e uso de dados. Algumas funções são orientadas para os negócios, algumas envolvem mais engenharia, algumas enfocam a pesquisa e algumas são funções híbridas que combinam diferentes aspectos do gerenciamento de dados. Nesta unidade, você explorará as funções de trabalho mais comuns no mundo dos dados. Sua organização pode definir funções de maneira diferente ou dar-lhes nomes diferentes, mas as funções descritas nesta unidade englobam a divisão mais comum de trabalho e responsabilidades.

## Quais são as funções no mundo dos dados?

Existem três funções principais que lidam com dados na maioria das organizações. Os administradores de banco de dados gerenciam bancos de dados, atribuindo permissões aos usuários, armazenando cópias de backup de dados e restaurando dados em caso de falhas.

Engenheiros de dados são vitais para trabalhar com dados, aplicando rotinas de limpeza de dados, identificando regras de negócios e transformando dados em informações úteis. Os analistas de dados exploram e analisam dados para criar visualizações e gráficos para permitir que as organizações tomem decisões informadas.

## Função de administrador de banco de dados do Azure



Um administrador de banco de dados do Azure é responsável pelo design, implementação, manutenção e aspectos operacionais de soluções de banco de dados locais e baseadas em nuvem criadas nos serviços de dados do Azure e SQL Server. Eles são responsáveis pela disponibilidade geral, desempenho consistente e otimizações das soluções de banco de dados. Eles trabalham com as partes interessadas para implementar políticas, ferramentas e processos para planos de backup e recuperação para recuperação após um desastre natural ou erro de fabricação humana.

O administrador do banco de dados também é responsável por gerenciar a segurança dos dados no banco de dados, concedendo privilégios sobre os dados, concedendo ou negando acesso aos usuários conforme apropriado.

## Função de engenheiro de dados



Um engenheiro de dados colabora com as partes interessadas para projetar e implementar ativos relacionados a dados que incluem pipelines de ingestão de dados, atividades de limpeza e transformação e armazenamentos de dados para cargas de trabalho analíticas. Eles usam uma ampla variedade de tecnologias de plataforma de dados, incluindo bancos de dados relacionais e não relacionais, armazenamentos de arquivos e fluxos de dados.

Eles também são responsáveis por garantir que a privacidade dos dados seja mantida dentro da nuvem e abrangendo desde os armazenamentos de dados locais aos na nuvem. Eles também possuem o gerenciamento e monitoramento de armazenamentos de dados e pipelines de dados para garantir que as cargas de dados sejam executadas conforme o esperado.

## Função de analista de dados



Um analista de dados permite que as empresas maximizem o valor de seus ativos de dados. Eles são responsáveis por projetar e construir modelos escaláveis, limpar e transformar dados e habilitar recursos analíticos avançados por meio de relatórios e visualizações.

Um analista de dados processa dados brutos em insights relevantes com base nos requisitos de negócios identificados para fornecer insights relevantes.

Revise tarefas e ferramentas para administração de banco de dados

Os administradores de banco de dados têm a tarefa de gerenciar e organizar bancos de dados. A principal função de um administrador de banco de dados é garantir que os dados estejam disponíveis, protegidos contra perda, corrupção ou roubo e sejam facilmente acessíveis conforme necessário.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Tarefas e responsabilidades do administrador de banco de dados

Algumas das funções e responsabilidades mais comuns de um administrador de banco de dados incluem:

- Instalando e atualizando o servidor de banco de dados e ferramentas de aplicativo.
- Alocação de armazenamento do sistema e planejamento de requisitos de armazenamento para o sistema de banco de dados.
- Modificar a estrutura do banco de dados, conforme necessário, a partir de informações fornecidas pelos desenvolvedores de aplicativos.
- Inscrevendo usuários e mantendo a segurança do sistema.
- Garantir a conformidade com o contrato de licença do fornecedor de banco de dados.
- Controlar e monitorar o acesso do usuário ao banco de dados.
- Monitorar e aperfeiçoar o desempenho do banco de dados.
- Planejamento para backup e recuperação de informações do banco de dados.
- Manter dados arquivados.
- Backup e restauração de bancos de dados.
- Entrando em contato com o fornecedor do banco de dados para obter suporte técnico.
- Gerar vários relatórios consultando a base de dados conforme a necessidade.
- Gerenciando e monitorando a replicação de dados.
- Atuando como elo de ligação com os usuários.

TITULO: FUNDAMENTOS DA SILVA OLIVEIRA

CPF: 00296691208

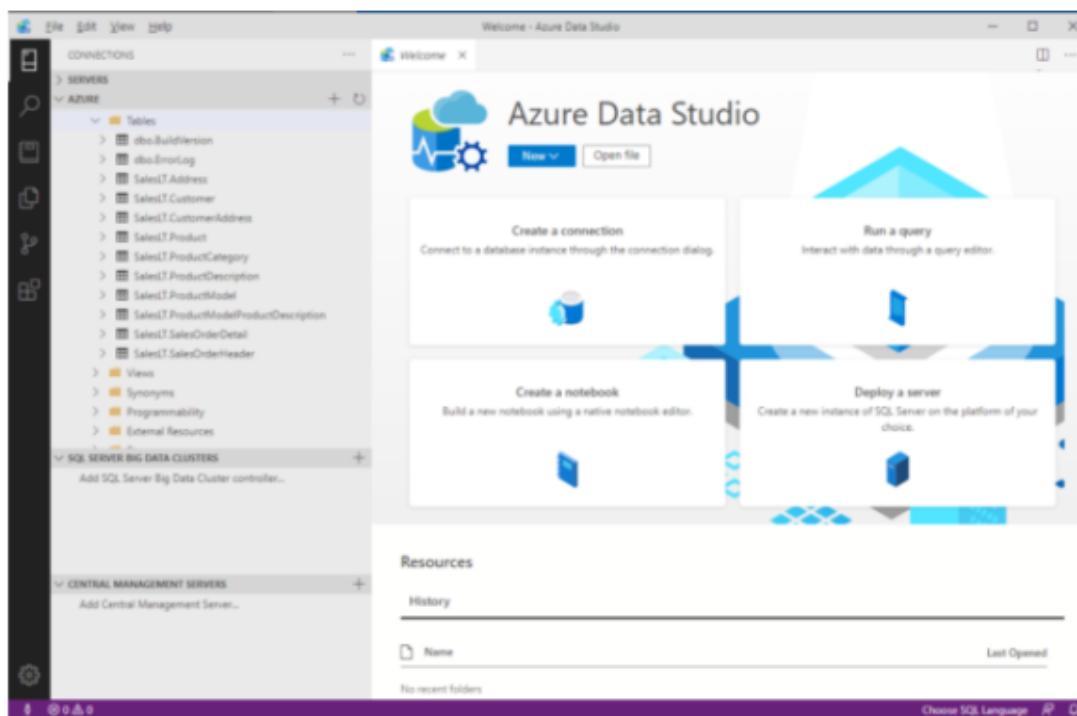
## Ferramentas comuns de administrador de banco de dados

A maioria dos sistemas de gerenciamento de banco de dados fornece seu próprio conjunto de ferramentas para auxiliar na administração do banco de dados. Por exemplo, os administradores de banco de dados do SQL Server usam SQL Server Management Studio para a maioria de suas atividades de manutenção de banco de dados no dia a dia.

Outros sistemas têm suas próprias interfaces específicas de banco de dados, como Pg Admin para sistemas PostgreSQL, ou MySQL Work Bench para MySQL. Existem também várias ferramentas de administração de banco de dados de plataforma cruzada disponíveis. Um exemplo é o Azure Data Studio.

## O que é Azure Data Studio?

O Azure Data Studio fornece uma interface gráfica do usuário para gerenciar muitos sistemas de banco de dados diferentes. Atualmente, ele fornece conexões com bancos de dados SQL Server locais, Banco de Dados SQL do Azure, PostgreSQL, Azure SQL Data Warehouse e Clusters de Big Data do SQL Server, entre outros. É uma ferramenta extensível e você pode baixar e instalar extensões de desenvolvedores terceirizados que se conectam a outros sistemas ou fornecem assistentes que ajudam a automatizar muitas tarefas administrativas.

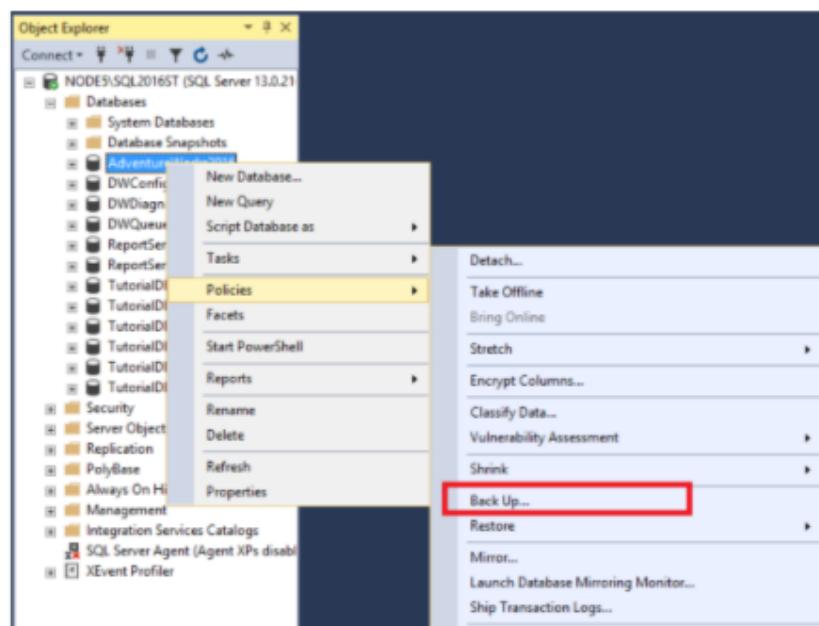


THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## O que é SQL Server Management Studio?

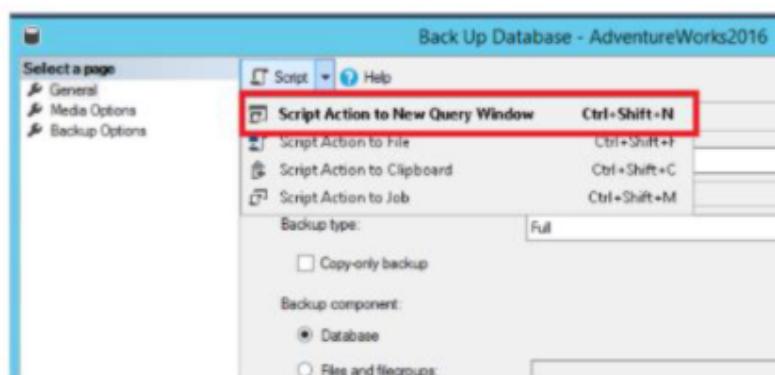
O SQL Server Management Studio fornece uma interface gráfica, permitindo que você consulte dados, execute tarefas gerais de administração de banco de dados e gere scripts para automatizar a manutenção do banco de dados e operações de suporte. O exemplo a seguir mostra o SQL Server Management Studio sendo usado para fazer backup de um banco de dados.



THIAGO FERNANDES DA SILVA OLIVEIRA

CNPJ 000296691208

Um recurso útil do SQL Server Management Studio é a capacidade de gerar scripts Transact-SQL para quase todas as funcionalidades fornecidas pelo SSMS. Isso dá ao DBA a capacidade de agendar e automatizar muitas tarefas comuns.

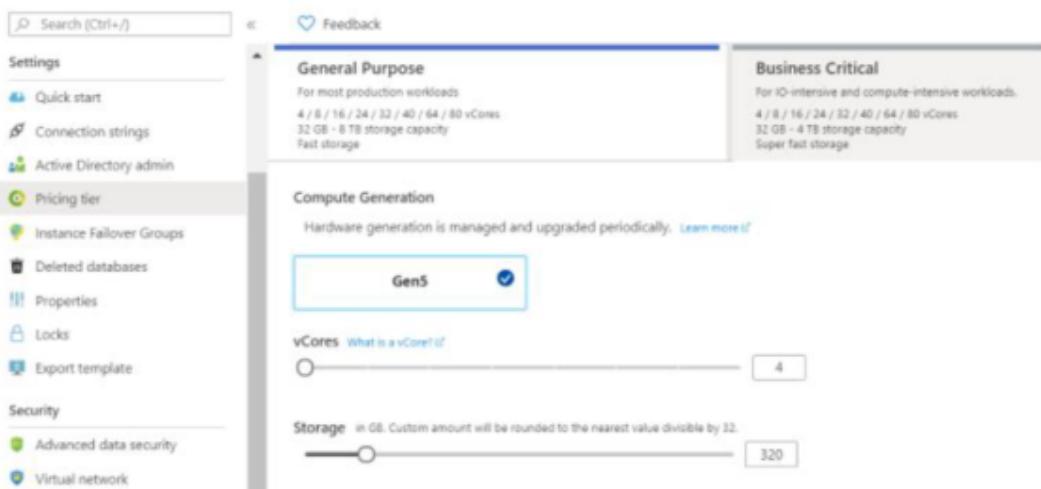


**NOTA:** Transact-SQL é um conjunto de extensões de programação da Microsoft que adiciona vários recursos à Structured Query Language (SQL), incluindo controle de transações, tratamento de exceções e erros, processamento de linhas e variáveis declaradas.

## Use o portal do Azure para gerenciar o Banco de Dados SQL do Azure

O banco de dados SQL do Azure fornece serviços de banco de dados no Azure. É semelhante ao SQL Server, exceto pelo fato de ser executado na nuvem. Você pode gerenciar o banco de dados SQL do Azure usando Azure portal.

Tarefas de configuração típicas, como aumentar o tamanho do banco de dados, criar um novo banco de dados e excluir um banco de dados existente, são feitas usando o portal do Azure.



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Você pode usar o portal do Azure para gerenciar e ajustar dinamicamente recursos, como o tamanho do armazenamento de dados e o número de núcleos disponíveis para o processamento do banco de dados. Essas tarefas exigiriam o suporte de um administrador do sistema se você estivesse executando o banco de dados local.

## Revise tarefas e ferramentas para engenharia de dados

Os engenheiros de dados têm a tarefa de gerenciar e organizar dados, ao mesmo tempo em que monitoram tendências ou inconsistências que afetarão as metas de negócios. É uma posição altamente técnica, que exige experiência e habilidades em áreas como programação, matemática e ciência da computação. Mas os engenheiros de dados também precisam de habilidades suaves para comunicar tendências de dados para outras pessoas na organização e para ajudar a empresa a usar os dados que coleta.

## Tarefas e responsabilidades do engenheiro de dados

Algumas das funções e responsabilidades mais comuns de um engenheiro de dados incluem:

- Desenvolvimento, construção, teste e manutenção de bancos de dados e estruturas de dados.
- Alinhando a arquitetura de dados com os requisitos de negócios.
- Aquisição de dados.
- Desenvolvimento de processos para criar e recuperar informações de conjuntos de dados.
- Usando linguagens de programação e ferramentas para examinar os dados.
- Identificar maneiras de melhorar a confiabilidade, eficiência e qualidade dos dados.
- Conduzindo pesquisas para questões de indústria e negócios.
- Implantar programas analíticos sofisticados, aprendizado de máquina e métodos estatísticos.
- Preparação de dados para modelagem preditiva e prescritiva.
- Usando dados para descobrir tarefas que podem ser automatizadas.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## Ferramentas comuns de engenharia de dados

Para dominar a engenharia de dados, você precisará estar familiarizado com uma variedade de ferramentas que permitem criar bancos de dados bem projetados, otimizados para os processos de negócios que serão executados. Você deve ter um entendimento completo da arquitetura do sistema de gerenciamento de banco de dados, a plataforma na qual o sistema é executado e os requisitos de negócios para os dados armazenados no banco de dados.

Se estiver usando um sistema de gerenciamento de banco de dados relacional, você precisa ser fluente em SQL. Você deve ser capaz de usar SQL para criar bancos de dados, tabelas, índices, visualizações e outros objetos exigidos pelo banco de dados. Muitos sistemas de gerenciamento de banco de dados fornecem ferramentas que permitem criar e executar scripts SQL. Por exemplo, SQL Server Management Studio (descrito na unidade anterior), permite criar e consultar tabelas visualmente, mas você também pode criar seus próprios scripts SQL manualmente.

Em alguns casos, você pode precisar interagir com um banco de dados a partir da linha de comando. Muitos sistemas de gerenciamento de banco de dados fornecem uma interface de linha de comando que suporta essas operações. Por exemplo, você pode usar osqlcmd utilitário para se conectar ao Microsoft SQL Server e ao Banco de Dados SQL do Azure e executar consultas e comandos ad-hoc.

Como profissional do SQL Server, sua principal ferramenta de manipulação de dados pode ser o Transact-SQL. Como engenheiro de dados, você pode usar tecnologias adicionais, como Azure Databricks, e Azure HD Insight para gerar e testar modelos preditivos. Se você está trabalhando no campo não relacional, pode usar Azure Cosmos DB como seu armazenamento de dados principal. Para manipular e consultar os dados, você pode usar linguagens como Hive QL, R ou Python.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Revise tarefas e ferramentas para visualização de dados e relatórios

Os analistas de dados são responsáveis por entender o que os dados realmente significam. Um analista de dados experiente explorará os dados e os usará para determinar tendências, problemas e obter outros insights que podem ser benéficos para a empresa.

Uma grande parte da função de analista de dados está relacionada à comunicação e visualização. A visualização de dados é a chave para apresentar grandes quantidades de informações de maneira que sejam universalmente comprehensíveis ou fáceis de interpretar e detectar padrões, tendências e correlações. Essas representações incluem tabelas, gráficos, infográficos e outros diagramas pictóricos.

Os analistas de visualização de dados usam ferramentas de visualização e software para comunicar informações dessas maneiras, para clientes ou para sua própria empresa. Um bom analista de dados requer experiência e habilidades em ferramentas de relatórios, como Microsoft Power BI e SQL Server Reporting Services.

## Tarefas e responsabilidades do analista de dados

As funções principais de um analista de dados geralmente incluem o seguinte:

- Tornar dados grandes ou complexos mais acessíveis, comprehensíveis e utilizáveis.
- Criar de tabelas e gráficos, histogramas, mapas geográficos e outros modelos visuais que ajudam a explicar o significado de grandes volumes de dados e isolar áreas de interesse.
- Transformar, melhorar e integrar dados de várias fontes, dependendo dos requisitos de negócios.
- Combinar os conjuntos de resultados de dados em várias fontes. Por exemplo, combinar dados de vendas e dados meteorológicos fornece uma visão útil de como o clima influenciou as vendas de certos produtos, como sorvetes.
- Encontrar padrões ocultos usando dados.
- Fornecer informações de uma maneira útil e atraente para os usuários, criando painéis e relatórios gráficos ricos.

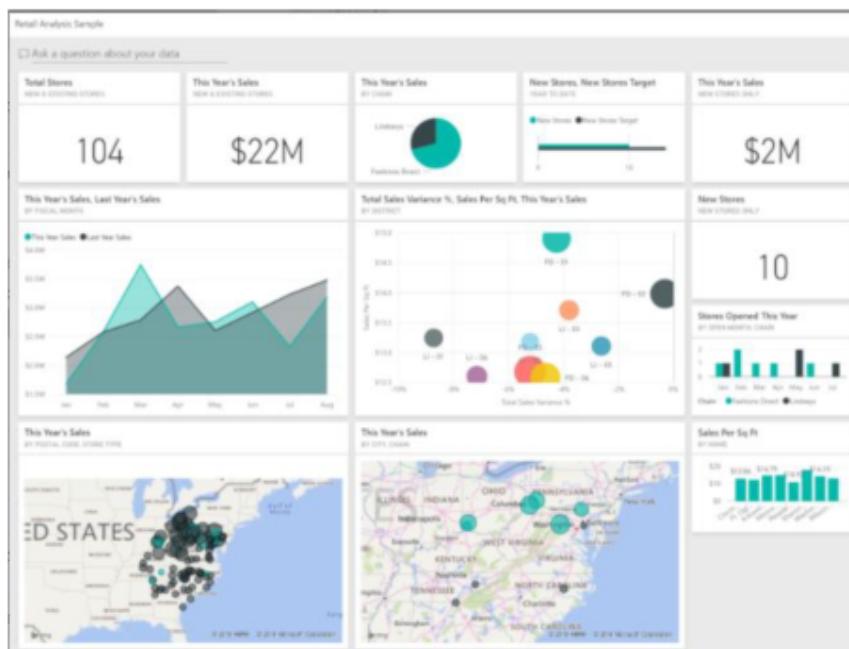
## Ferramentas comuns de visualização de dados

Tradicionalmente, muitos analistas de dados usavam aplicativos do Microsoft Office, como o Microsoft Excel, para criar relatórios visuais sofisticados. Muitos analistas agora usam o Microsoft Power BI, uma plataforma de visualização poderosa, para criar painéis e relatórios gráficos ricos sobre dados que podem variar dinamicamente.

O Power BI é uma coleção de serviços de software, aplicativos e conectores que funcionam juntos para transformar suas fontes de dados não relacionadas em percepções coerentes, visualmente imersivas e interativas. Seus dados podem ser mantidos em algum lugar local, como uma planilha do Excel, ou em uma coleção de bancos de dados locais e baseados em nuvem, ou algum outro conjunto de fontes de dados.

O Power BI permite que você se conecte facilmente às suas fontes de dados, descubra o que é importante nesses dados e compartilhe suas descobertas com outras pessoas na organização.

A imagem abaixo mostra um exemplo de um painel criado usando o Power BI. Neste exemplo, o analista está usando o Power BI para examinar os dados de vendas no varejo para itens vendidos em várias lojas e distritos. As métricas compararam o desempenho deste ano com o do ano passado para vendas, unidades, margem bruta e variação, bem como análise de novas lojas.



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

## Teste de conhecimento

### Questão 1

**Qual das seguintes tarefas é uma função de administrador de banco de dados?**

- a) Backup e restauração de bancos de dados
- b) Criação de painéis e relatórios
- c) Identificação de problemas de qualidade de dados

**Explicação:** está correto. Os administradores de banco de dados farão backup do banco de dados e restaurarão o banco de dados quando os dados forem perdidos ou corrompidos.

### Questão 2

**Qual das seguintes ferramentas é uma ferramenta de visualização e relatório?**

- a) SQL Server Management Studio
- b) Power BI
- c) SQL
- d)

**Explicação:** está correto. O Power BI é uma ferramenta padrão para criar painéis e relatórios gráficos ricos.

### Questão 3

**Qual das funções a seguir não é uma função de trabalho de dados?**

- a) Administrador de sistemas
- b) Analista de informações
- c) Administrador de banco de dados

**Explicação:** está correto. Os administradores de sistemas lidam com componentes de infraestrutura, como redes, máquinas virtuais e outros dispositivos físicos em um data center.

## Resumo

Gerenciar e trabalhar com dados é uma habilidade especializada. A maioria das organizações define funções de trabalho para as várias tarefas responsáveis pelo gerenciamento de dados.

Nesta lição você aprendeu:

- Algumas das funções comuns de trabalho para lidar com dados
- As tarefas normalmente realizadas por esses cargos e os tipos de ferramentas que eles usam

## Saber mais

- Visão geral do Azure Databricks
- Visão geral do Azure HDInsight
- Introdução ao Azure Cosmos DB
- Visão geral do Power BI
- Documentação técnica do SQL Server
- Introdução ao Azure Data Factory

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Introdução

Nos primeiros anos dos bancos de dados, cada aplicativo armazenava dados em sua própria estrutura única. Quando os desenvolvedores queriam construir aplicativos para usar esses dados, eles precisavam saber muito sobre uma estrutura específica para encontrar os dados de que precisavam.

Essas estruturas de dados eram ineficientes, difíceis de manter e difíceis de otimizar para fornecer bom desempenho de aplicativos. O modelo de banco de dados relacional foi projetado para resolver o problema de múltiplas estruturas de dados arbitrárias.

O modelo relacional forneceu uma maneira padrão de representar e consultar dados, que podem ser usados por qualquer aplicativo. Desde o início, os desenvolvedores reconheceram que a principal força do modelo de banco de dados relacional estava no uso de tabelas, que eram uma maneira intuitiva, eficiente e flexível de armazenar e acessar informações estruturadas.

O modelo relacional simples, mas poderoso, é usado por organizações de todos os tipos e tamanhos para uma ampla variedade de necessidades de gerenciamento de informações. Bancos de dados relacionais são usados para rastrear estoques, processar transações de comércio eletrônico, gerenciar grandes quantidades de informações de clientes de missão crítica e muito mais. Um banco de dados relacional é útil para armazenar qualquer informação que contenha elementos de dados relacionados que devem ser organizados de maneira consistente e baseada em regras.

Nesta lição, você aprenderá sobre as principais características dos dados relacionais e explorará as estruturas de dados relacionais.

## Objetivos de aprendizado

Nesta lição você irá:

- Explorar as características dos dados relacionais
- Definir tabelas, índices e visualizações
- Explorar as ofertas de carga de trabalho de dados relacionais no Azure

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Explore as características dos dados relacionais

Um dos principais benefícios dos bancos de dados de computador é que eles facilitam o armazenamento de informações, tornando-os rápidos e fáceis de encontrar. Por exemplo, um sistema de comércio eletrônico pode usar um banco de dados para registrar informações sobre os produtos que uma organização vende, detalhes dos clientes e os pedidos que eles fizeram. Um banco de dados relacional fornece um modelo para armazenar os dados e um recurso de consulta que permite recuperar os dados rapidamente.

Nesta unidade, você aprenderá mais sobre as características dos dados relacionais e como armazenar essas informações e consultá-las em um banco de dados relacional.

## Compreenda as características dos dados relacionais

Em um banco de dados relacional, você modela coleções de entidades do mundo real como tabelas. Uma entidade é descrita como algo sobre o qual a informação precisa ser conhecida ou mantida.

No exemplo de comércio eletrônico, você pode criar tabelas para clientes, produtos e pedidos. Uma tabela contém linhas e cada linha representa uma única instância de uma entidade. No cenário de comércio eletrônico, cada linha na tabela de clientes contém os dados de um único cliente, cada linha na tabela de produtos define um único produto e cada linha na tabela de pedidos representa um pedido feito por um cliente.

As linhas em uma tabela têm uma ou mais colunas que definem as propriedades da entidade, como o nome do cliente ou ID do produto. Todas as linhas da mesma tabela têm as mesmas colunas. Algumas colunas são usadas para manter relacionamentos entre tabelas. É daí que vem o nome do modelo relacional. Na imagem a seguir, a tabela Pedidos contém uma ID de cliente e uma ID de produto. O ID do cliente está relacionado à tabela Clientes para identificar o cliente que fez o pedido e o ID do produto está relacionado à tabela Produtos para indicar qual produto foi comprado.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Customers

| Customer ID | Customer Name | Customer Address |
|-------------|---------------|------------------|
| C1          | Fred          | ...              |
| C2          | Bert          | ...              |
| C3          | Jane          | ...              |

Products

| Product ID | Product Name | Description |
|------------|--------------|-------------|
| P1         | Shirt        | ...         |
| P2         | Tie          | ...         |
| P3         | Collar       | ...         |

Orders

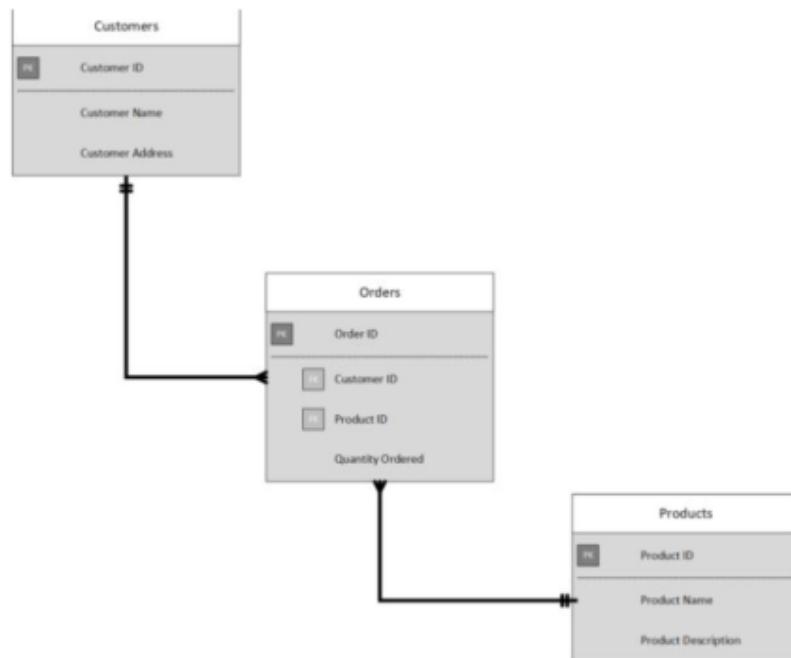
| Order ID | Customer ID | Product ID | Quantity |
|----------|-------------|------------|----------|
| 1000     | C1          | P1         | 1        |
| 1001     | C2          | P1         | 3        |
| 1002     | C1          | P3         | 1        |
| 1003     | C1          | P3         | 2        |
| 1004     | C2          | P2         | 4        |
| 1005     | C1          | P2         | 2        |
| 1006     | C3          | P3         | 1        |

Você projeta um banco de dados relacional criando um modelo de dados. O modelo a seguir mostra a estrutura das entidades do exemplo anterior. Neste diagrama, as colunas marcadas PK (Primary Key) são as chaves primárias para a mesa. A chave primária indica a coluna (ou combinação de colunas) que identifica exclusivamente cada linha. Cada tabela deve ter uma chave primária.

O diagrama também mostra as relações entre as tabelas. As linhas que conectam as tabelas indicam o tipo de relacionamento. Neste caso, a relação dos clientes com os pedidos é 1-para-muitos (um cliente pode fazer vários pedidos, mas cada pedido é para um único cliente). Da mesma forma, a relação entre pedidos e produtos é muitos-para-1 (vários pedidos podem ser para o mesmo produto).

As colunas marcadas FK (Foreign Key) são as colunas de chave estrangeira. Eles fazem referência ou se vinculam à chave primária de outra tabela e são usadas para manter os relacionamentos entre as tabelas. Uma chave estrangeira também ajuda a identificar e prevenir anomalias, como pedidos de clientes que não existem na tabela Clientes. No modelo abaixo, as colunas ID do cliente e ID do produto na tabela Pedidos vinculam-se ao cliente que fez o pedido e ao produto que foi pedido:

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208



As principais características de um banco de dados relacional são:

- Todos os dados são tabulares. As entidades são modeladas como tabelas, cada instância de uma entidade é uma linha na tabela e cada propriedade é definida como uma coluna.
- Todas as linhas da mesma tabela têm o mesmo conjunto de colunas.
- Uma tabela pode conter qualquer número de linhas.
- Uma chave primária identifica exclusivamente cada linha em uma tabela. Duas linhas não podem compartilhar a mesma chave primária.
- Uma chave estrangeira faz referência a linhas em outra tabela relacionada. Para cada valor na coluna de chave estrangeira, deve haver uma linha com o mesmo valor na coluna de chave primária correspondente na outra tabela.

**NOTA:** Criar um modelo de banco de dados relacional para uma grande organização não é uma tarefa trivial. Podem ser necessárias várias iterações para definir tabelas que correspondam às características descritas acima. Às vezes, você precisa dividir uma entidade em mais de uma tabela. Este processo é chamado de normalização.

A maioria dos bancos de dados relacionais oferece suporte a Structured Query Language (SQL). Você usa SQL para criar tabelas, inserir, atualizar e excluir linhas em tabelas e para consultar dados. Você utiliza o **CRIAR A TABELA** para criar uma mesa, o **INSERIR** declaração para armazenar dados em uma tabela, o **ATUALIZAR** declaração para modificar dados em uma tabela, e o **EXCLUIR** para remover linhas de uma tabela. O **SELECIONE** declaração recupera dados de uma tabela. O exemplo de consulta abaixo encontra os detalhes de cada cliente do banco de dados de exemplo mostrado acima.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

```
SELECT CustomerID, CustomerName, CustomerAddress  
FROM Customers
```

Em vez de recuperar todas as linhas, você pode filtrar os dados usando um ONDE cláusula. A próxima consulta busca o ID do pedido e o ID do produto para todos os pedidos feitos pelo cliente 1.

```
SELECT OrderID, ProductID  
FROM Orders  
WHERE CustomerID = 'C1'
```

Você pode combinar os dados de várias tabelas em uma consulta usando uma operação de junção. Uma operação de junção abrange os relacionamentos entre as tabelas, permitindo que você recupere os dados de mais de uma tabela por vez. A consulta a seguir recupera o nome de cada cliente, junto com o nome do produto e a quantidade de cada pedido feito. Observe que cada coluna é qualificada com a tabela a que pertence:

```
SELECT Customers.CustomerName, Orders.QuantityOrdered,  
Products.ProductName  
FROM Customers JOIN Orders  
ON Customers.CustomerID = Orders.CustomerID  
JOIN Products  
ON Orders.ProductID = Products.ProductID
```

Você pode encontrar detalhes completos sobre SQL no site da Microsoft, em Linguagem de consulta estruturada (SQL) página.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Explorar casos de uso de banco de dados relacional

Você pode usar um banco de dados relacional a qualquer momento, podendo modelar facilmente seus dados como uma coleção de tabelas com um conjunto fixo de colunas. Em teoria, você poderia modelar quase qualquer conjunto de dados dessa maneira, mas alguns cenários se prestam ao modelo relacional melhor do que outros.

Por exemplo, se você tem uma coleção de música, vídeo ou outros arquivos de mídia, pode ser difícil tentar forçar esses dados no modelo relacional. Talvez seja melhor usar armazenamento não estruturado, como o disponível no armazenamento de Blob do Azure. Da mesma forma, os sites de redes sociais usam bancos de dados para armazenar dados sobre milhões de usuários, junto com fotografias e outras informações sobre esses usuários e outras pessoas. Esse tipo de dados se presta mais a uma estrutura de banco de dados de gráficos, do que a uma coleção de tabelas relacionais.

Bancos de dados relacionais são comumente usados em sistemas de comércio eletrônico, mas um dos principais casos de uso para usar bancos de dados relacionais é o processamento de transações online (OLTP). Os aplicativos OLTP são focados em tarefas orientadas a transações que processam um grande número de transações por minuto. Os bancos de dados relacionais são adequados para aplicativos OLTP porque naturalmente suportam operações de inserção, atualização e exclusão. Muitas vezes, um banco de dados relacional pode ser ajustado para tornar essas operações rápidas. Além disso, a natureza do SQL torna mais fácil para os usuários realizarem consultas ad-hoc sobre os dados.

Exemplos de aplicativos OLTP que usam bancos de dados relacionais são soluções bancárias, aplicativos de varejo online, sistemas de reserva de voos e muitos aplicativos de compra.

## Explore estruturas de dados relacionais

Um banco de dados relacional compreende um conjunto de tabelas. Uma tabela pode ter zero (se a tabela estiver vazia) ou mais linhas. Cada tabela possui um conjunto fixo de colunas. Você pode definir relacionamentos entre tabelas usando chaves primárias e estrangeiras, e pode acessar os dados em tabelas usando SQL.

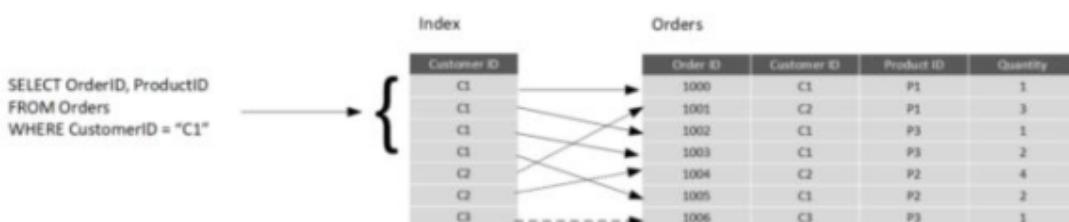
Além das tabelas, um banco de dados relacional típico contém outras estruturas que ajudam a otimizar a organização dos dados e a melhorar a velocidade de acesso. Nesta unidade, você verá duas dessas estruturas com mais detalhes: índices e visualizações.

## O que é um índice?

Um índice ajuda a pesquisar dados em uma tabela. Pense em um índice sobre uma tabela como um índice no final de um livro. Um índice de livro contém um conjunto classificado de referências, com as páginas em que cada referência ocorre. Quando você deseja encontrar uma referência a um item no livro, você a consulta no índice. Você pode usar os números das páginas no índice para ir diretamente para as páginas corretas do livro. Sem um índice, talvez você precise ler o livro inteiro para encontrar as referências que procura.

Ao criar um índice em um banco de dados, você especifica uma coluna da tabela e o índice irá conter uma cópia desses dados em uma ordem de classificação, com ponteiros para as linhas correspondentes na tabela.

Quando o usuário executa uma consulta que especifica esta coluna no **ONDE**, o sistema de gerenciamento de banco de dados pode usar este índice para buscar os dados mais rapidamente do que se tivesse que varrer toda a tabela linha por linha. No exemplo abaixo, a consulta recupera todos os pedidos do cliente C1. A tabela Pedidos possui um índice na coluna ID do cliente. O sistema de gerenciamento de banco de dados pode consultar o índice para localizar rapidamente todas as linhas correspondentes na tabela Pedidos.



Você pode criar muitos índices em uma tabela. Portanto, se você também quiser localizar todos os pedidos de um produto específico, será útil criar outro índice na coluna ID do produto, na tabela Pedidos. No entanto, os índices não são gratuitos. Um índice pode consumir espaço de armazenamento adicional e cada vez que você insere, atualiza ou exclui dados em uma tabela, os índices dessa tabela devem ser mantidos.

Esse trabalho adicional pode desacelerar as operações de inserção, atualização e exclusão e incorrer em custos de processamento adicionais. Portanto, ao decidir quais índices criar, você deve encontrar um equilíbrio entre ter índices que aceleram suas consultas e o custo de realizar outras operações.

Em uma tabela que é somente leitura, ou que contém dados que são modificados com pouca frequência, mais índices melhorarão o desempenho da consulta.

Alguns sistemas de gerenciamento de banco de dados relacional também oferecem suporte índices agrupados. Um índice clusterizado reorganiza fisicamente uma tabela pela chave do índice. Esse arranjo pode melhorar ainda mais o desempenho das consultas, porque o sistema de gerenciamento de banco de dados relacional não precisa seguir as referências do índice para encontrar os dados correspondentes na tabela subjacente.

A imagem abaixo mostra a tabela Pedidos com um índice clusterizado na coluna ID do cliente.

Orders

SELECT OrderID, ProductID  
FROM Orders  
WHERE CustomerID = "C1"

{

| Customer ID | Order ID | Product ID | Quantity |
|-------------|----------|------------|----------|
| C1          | 1000     | P1         | 1        |
| C1          | 1002     | P1         | 3        |
| C1          | 1003     | P3         | 1        |
| C1          | 1005     | P3         | 2        |
| C2          | 1001     | P2         | 4        |
| C2          | 1004     | P2         | 2        |
| C3          | 1006     | P3         | 1        |

Clustered Index

Em sistemas de gerenciamento de banco de dados que os suportam, uma tabela pode ter apenas um único índice clusterizado.

### O que é uma visão (view)?

Uma visão é uma tabela virtual baseada no conjunto de resultados de uma consulta. No caso mais simples, você pode pensar em uma visualização como uma janela em linhas especificadas em uma tabela subjacente. Por exemplo, você pode criar uma visualização na tabela Pedidos que liste os pedidos de um produto específico (neste caso, produto P1) como isso:

```
CREATE VIEW P1Orders AS  
SELECT CustomerID, OrderID, Quantity  
FROM Orders  
WHERE ProductID = "P1"
```

Você pode consultar a exibição e filtrar os dados da mesma maneira que uma tabela. A consulta a seguir localiza os pedidos do cliente C1 usando a visualização. Esta consulta retornará apenas pedidos do produto P1 feitos pelo cliente:

```
SELECT CustomerID, OrderID, Quantity  
FROM P1Orders  
WHERE CustomerID = "C1"
```

Uma visão também pode unir tabelas. Se você precisava encontrar regularmente os detalhes dos clientes e os produtos que eles pediram, pode criar uma visualização com base na consulta de junção mostrada na unidade anterior:

```
CREATE VIEW CustomersProducts AS
SQL
SELECT          Customers.CustomerName,          Orders.QuantityOrdered,
Products.ProductName
FROM Customers JOIN Orders
ON Customers.CustomerID = Orders.CustomerID
JOIN Products
ON Orders.ProductID = Products.ProductID
```

A consulta a seguir encontra o nome do cliente e os nomes dos produtos de todos os pedidos feitos pelo cliente C2, usando esta visualização:

```
SELECT CustomerName, ProductName
FROM CustomersProducts
WHERE CustomerID = "C2"
```

### Escolha a plataforma certa para uma carga de trabalho relacional

A popularidade da computação em nuvem cresceu, prometendo flexibilidade para empresas, oportunidades para economizar tempo e dinheiro e melhorar a agilidade e escalabilidade. Por outro lado, o software local, instalado nos próprios servidores de uma empresa e protegido por firewall, ainda tem seu apelo. Os aplicativos locais são confiáveis, seguros e permitem que as empresas mantenham um controle próximo.

Os sistemas de gerenciamento de banco de dados relacional são um exemplo de onde a nuvem permitiu que as organizações tirassem proveito da escalabilidade aprimorada. No entanto, essa escalabilidade deve ser balanceada com a necessidade de controle próximo sobre os dados. Os dados são indiscutivelmente um dos ativos mais valiosos que uma organização possui, e algumas empresas não desejam ou não podem passar a responsabilidade pela proteção desses dados para terceiros.

Nesta unidade, você verá algumas das vantagens e desvantagens de executar um sistema de gerenciamento de banco de dados na nuvem.

TÍTULO DA UNIDADE

CPF: 00296691208

### Compare a hospedagem local com a nuvem

Quer a empresa coloque sua carga de trabalho relacional na nuvem ou decida mantê-la no local, a segurança dos dados sempre será primordial. Mas para as empresas em setores altamente regulamentados, a decisão já pode ter sido feita quanto a hospedar seus aplicativos no local. Saber que seus dados estão localizados nos servidores internos e na infraestrutura de TI também pode proporcionar mais tranquilidade.

Hospedar um banco de dados relacional no local requer que uma empresa não apenas adquira o software do banco de dados, mas também mantenha o hardware necessário para executar o banco de dados. A organização é responsável por manter o hardware e software, aplicar patches, fazer backup dos bancos de dados, restaurá-los quando necessário e, geralmente, realizar todo o gerenciamento diário necessário para manter a plataforma operacional.

A escalabilidade também é uma preocupação. Se você precisar dimensionar seu sistema, precisará atualizar ou adicionar mais servidores. Em seguida, você precisa expandir seu banco de dados para esses servidores. Esta pode ser uma tarefa formidável que requer que você coloque um banco de dados offline enquanto a operação é executada. Na nuvem, muitas dessas operações podem ser gerenciadas para você pela equipe do data center, em muitos casos com nenhum (ou mínimo) tempo de inatividade. Você é livre para focar especificamente nos dados e deixar a condução de outras preocupações para terceiros. Afinal de contas, é para isso que você paga suas taxas da Azure.

Uma abordagem baseada em nuvem usa tecnologia virtual para hospedar os aplicativos de uma empresa fora do local. Não há despesas de capital, é possível fazer backup dos dados regularmente e as empresas só precisam pagar pelos recursos que usam. Para as organizações que planejam uma expansão agressiva em uma base global, a nuvem tem um apelo ainda maior porque permite que você se conecte com clientes, parceiros e outras empresas em qualquer lugar com o mínimo de esforço.

Além disso, a computação em nuvem oferece provisionamento quase instantâneo, porque tudo já está configurado. Assim, qualquer novo software integrado ao seu ambiente está pronto para uso imediatamente após a assinatura da empresa. Com o provisionamento instantâneo, qualquer tempo gasto na instalação e configuração é eliminado e os usuários podem acessar o aplicativo imediatamente.

CPF 00296691208

## Compreenda IaaS e PaaS

Você geralmente tem duas opções ao mover suas operações e bancos de dados para a nuvem. Você pode selecionar uma abordagem IaaS ou PaaS.

IaaS é um acrônimo para Infraestrutura como um serviço. O Azure permite que você crie uma infraestrutura virtual na nuvem que espelha a maneira como um data center local pode funcionar. Você pode criar um conjunto de máquinas virtuais, conectá-las usando uma rede virtual e adicionar uma variedade de dispositivos virtuais. Em muitos aspectos, essa abordagem é semelhante à maneira como você executa seus sistemas dentro de uma organização, exceto que você não precisa se preocupar em comprar ou manter o hardware.

No entanto, você ainda é responsável por muitas das operações diárias, como instalação e configuração do software, aplicação de patches, backups e restauração de dados quando necessário. Você pode pensar em IaaS como uma casa de recuperação para operações totalmente gerenciadas na nuvem. Você não precisa se preocupar com o hardware, mas executar e gerenciar o software ainda é sua responsabilidade.

Você pode executar qualquer software para o qual tenha as licenças apropriadas usando essa abordagem. Você não está restrito a nenhum sistema de gerenciamento de banco de dados específico.

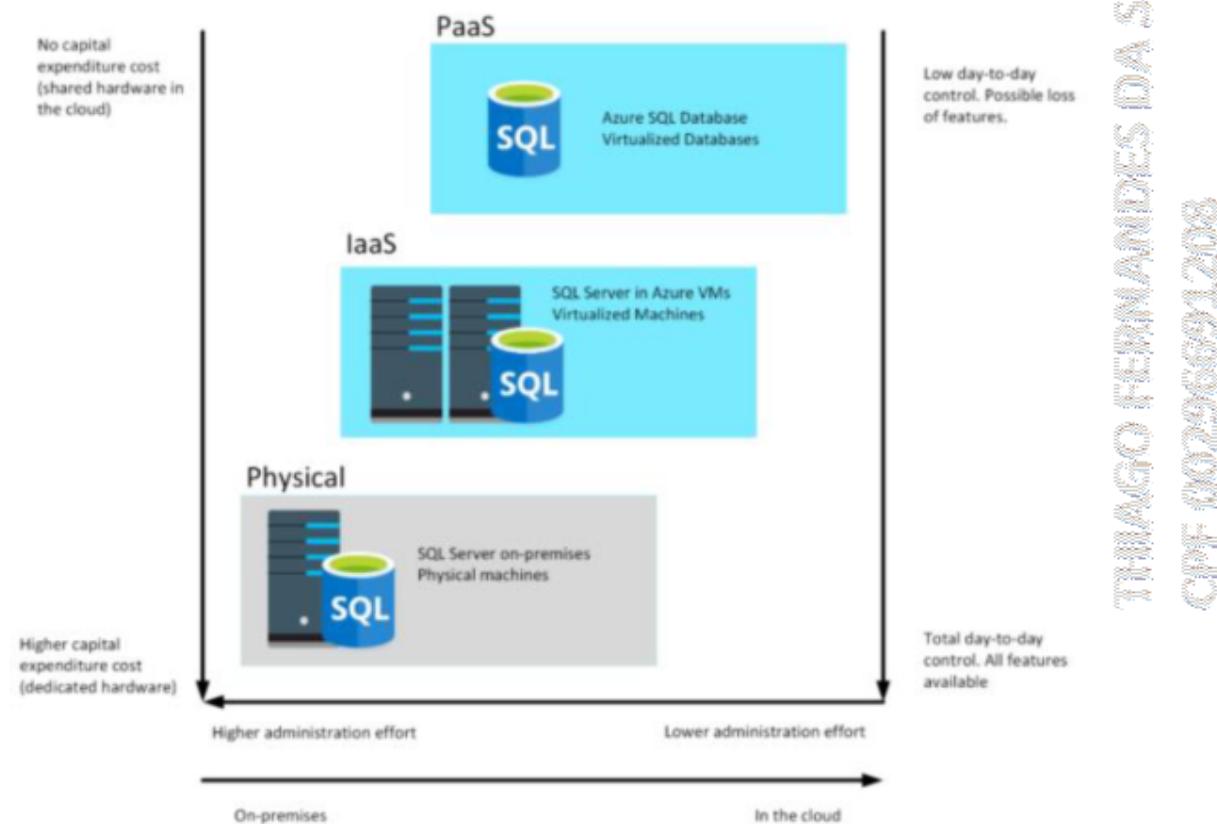
A abordagem IaaS é melhor para migrações e aplicativos que requerem acesso no nível do sistema operacional. Máquinas virtuais SQL são *lift-and-shift*. Ou seja, você pode copiar sua solução local diretamente para uma máquina virtual na nuvem. O sistema deve funcionar mais ou menos exatamente como antes em seu novo local, exceto por algumas pequenas mudanças de configuração (mudanças nos endereços de rede, por exemplo) para levar em conta a mudança no ambiente.

PaaS apoia Plataforma como serviço. Em vez de criar uma infraestrutura virtual e instalar e gerenciar o software de banco de dados você mesmo, uma solução PaaS faz isso por você. Você especifica os recursos de que precisa (com base no tamanho que você acha que seus bancos de dados serão, no número de usuários e no desempenho de que você precisa) e o Azure cria automaticamente as máquinas virtuais, redes e outros dispositivos necessários para você. Normalmente, você pode aumentar ou diminuir (aumentar ou diminuir o tamanho e o número de recursos) rapidamente, pois o volume de dados e a quantidade de trabalho que está sendo feito variam. O Azure lida com esse dimensionamento para você e você não precisa adicionar ou remover manualmente as máquinas virtuais ou realizar qualquer outra forma de configuração.

O Azure oferece várias soluções de PaaS para bancos de dados relacionais, incluindo Banco de Dados SQL do Azure, Banco de Dados do Azure para PostgreSQL, Banco de Dados do Azure para MySQL e Banco de Dados do Azure para MariaDB. Esses serviços executam versões gerenciadas dos sistemas de gerenciamento de banco de dados em seu nome. Você apenas se conecta a eles, cria seus bancos de dados e carrega seus dados.

No entanto, você pode descobrir que existem algumas restrições funcionais e nem todos os recursos do sistema de gerenciamento de banco de dados selecionado podem estar disponíveis. Essas restrições geralmente são devidas a problemas de segurança. Por exemplo, eles podem expor o sistema operacional e o hardware subjacentes aos seus aplicativos. Nesses casos, você pode precisar retrabalhar seus aplicativos para remover quaisquer dependências desses recursos.

A imagem abaixo ilustra os benefícios e desvantagens ao executar um sistema de gerenciamento de banco de dados (neste caso, SQL Server) no local, usando máquinas virtuais no Azure (IaaS) ou usando o Banco de Dados SQL do Azure (PaaS). As mesmas considerações generalizadas são verdadeiras para outros sistemas de gerenciamento de banco de dados.



## Questão 1

**Qual das seguintes afirmações é uma característica de um banco de dados relacional?**

- a) Todos os dados devem ser armazenados como cadeias de caracteres
- b) **Uma linha em uma tabela representa uma única entidade**
- c) Linhas diferentes na mesma tabela podem conter colunas diferentes

**Explicação:** está correto. Cada linha de uma tabela contém os dados de uma única entidade dessa tabela

## Questão 2

**O que é um índice?**

- a) **Uma estrutura que permite localizar linhas em uma tabela rapidamente, usando um valor indexado**
- b) Uma tabela virtual baseada no conjunto de resultados de uma consulta
- c) Uma estrutura que compreende linhas e colunas que você usa para armazenar dados

**Explicação:** está correto. Você cria índices para ajudar a acelerar as consultas.

## Questão 3

**Qual das seguintes afirmações é um benefício de usar um serviço PaaS, em vez de um sistema local, para executar seus sistemas de gerenciamento de banco de dados?**

- a) Aumento dos custos de gerenciamento do dia a dia
- b) **Maior escalabilidade**
- c) Maior funcionalidade

**Explicação:** está correto. As soluções de PaaS permitem que você amplie e expanda sem ter que adquirir seu próprio hardware.

## Resumo

Bancos de dados relacionais são amplamente usados para construir aplicativos do mundo real. Compreender as características dos dados relacionais é importante. Um banco de dados relacional é baseado em tabelas. Você pode executar muitos sistemas de gerenciamento de banco de dados no local e na nuvem.

Nesta lição você aprendeu:

- As características dos dados relacionais
- O que são tabelas, índices e visualizações
- As várias ofertas de carga de trabalho de dados relacionais disponíveis no Azure.

## Saber mais

- Descrição dos fundamentos de normalização do banco de dados
- Linguagem de consulta estruturada (SQL)
- Visão geral técnica do banco de dados SQL
- Documentação técnica do SQL Server
- Ofertas de PaaS vs IaaS do banco de dados SQL

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Introdução

Os dados vêm em todas as formas e tamanhos e podem ser usados para muitas finalidades. Muitas organizações usam bancos de dados relacionais para armazenar esses dados. No entanto, o modelo relacional pode não ser o esquema mais apropriado. A estrutura dos dados pode ser muito variada para ser facilmente modelada como um conjunto de tabelas relacionais. Por exemplo, os dados podem conter itens como vídeo, áudio, imagens, informações temporais, grandes volumes de texto livre ou outros tipos de dados que não são inherentemente relacionais. Além disso, os requisitos de processamento de dados podem não ser mais adequados ao tentar converter esses dados no formato relacional. Nessas situações, pode ser melhor usar repositórios não relacionais que podem armazenar dados em seu formato original, mas que permitem o armazenamento rápido e o acesso de recuperação a esses dados.

Suponha que você seja um engenheiro de dados trabalhando na Contoso, uma organização com uma grande operação de manufatura. A organização deve reunir e armazenar informações de uma variedade de fontes, como dados em tempo real de monitoramento do status das máquinas da linha de produção, dados de controle de qualidade do produto, registros históricos de produção, volumes de produtos em estoque e dados de estoque de matérias-primas. Essas informações são críticas para o funcionamento da organização. Você foi solicitado a determinar a melhor forma de armazenar essas informações, para que possam ser armazenadas rapidamente e consultadas facilmente.

## Objetivos de aprendizado

Nesta lição, você irá:

- Explorar as características dos dados não relacionais
- Definir tipos de dados não relacionais
- Descrever o NoSQL e os tipos de bancos de dados não relacionais

THIAGO FERNANDES  
CPF 00296691208

## Explorar características de dados não relacionais

Os bancos de dados relacionais são uma excelente ferramenta para armazenar e recuperar dados que possuem uma estrutura conhecida, contendo campos que você pode definir com antecedência. Em algumas situações, você pode não ter o conhecimento necessário da estrutura de seus dados, antes que eles cheguem ao seu banco de dados, para registrá-los como um conjunto organizado de linhas e colunas em um formato tabular. Este é um cenário comum em sistemas que consomem dados de uma ampla variedade de fontes, como pipelines de ingestão de dados. Nessas situações, um banco de dados não relacional pode ser extremamente útil.

Nesta unidade, você examinará com mais detalhes as características comuns dos bancos de dados não relacionais. Você aprenderá como eles permitem capturar dados rapidamente e modelar dados que podem variar em estrutura.

### Quais são as características dos dados não relacionais?

Você usa um banco de dados para modelar algum aspecto do mundo real. As entidades do mundo real costumam ter estruturas altamente variáveis. Por exemplo, em um banco de dados de comércio eletrônico que armazena informações sobre clientes, quantos números de telefone um cliente possui? Um cliente pode ter um telefone fixo e um número de celular, mas alguns clientes podem ter um número comercial, um número residencial adicional e talvez vários números de celular. Similarmente, os endereços dos clientes podem nem sempre seguir o mesmo formato; endereços de clientes em diferentes estados e regiões podem conter diferentes elementos, como CEPs ou códigos postais.

Em outro cenário, se você estiver ingerindo dados rapidamente, deseja capturar os dados e salvá-los de maneira igualmente ágil. Processar os dados e manipulá-los em um conjunto de linhas, em diferentes tabelas, em um banco de dados relacional, pode não ser apropriado neste ponto.

Você pode executar essas tarefas como parte em uma data posterior. No momento da ingestão, você simplesmente precisa armazenar os dados em seu estado e formato originais.

CURSO DE AZURE DATA FLOW

CPF: 000296691208

Um aspecto importante dos bancos de dados não relacionais é que eles permitem que você armazene dados de uma maneira muito flexível. Os bancos de dados não relacionais não impõem um esquema aos dados. Em vez disso, eles se concentram nos próprios dados, e não em como estruturá-los. Essa abordagem significa que você pode armazenar informações em um formato natural, que espelha a maneira como você as consumiria, consultaria e utilizaria.

Em um sistema não relacional, você armazena as informações de entidades em coleções ou contêineres, em vez de tabelas relacionais. Duas entidades na mesma coleção podem ter um conjunto diferente de campos em vez de um conjunto regular de colunas encontrado em uma tabela relacional. A falta de um esquema fixo significa que cada entidade deve ser autodescritiva. Frequentemente, isso é feito marcando cada campo com o nome dos dados que ele representa. Por exemplo, uma coleção não relacional de entidades de clientes pode ter a seguinte aparência:

```
## Customer 1
ID: 1
Name: Mark Hanson
Telephone: [ Home: 1-999-9999999, Business: 1-888-8888888, Cell: 1-777-7777777 ]
Address: [ Home: 121 Main Street, Some City, NY, 10110,
Business: 87 Big Building, Some City, NY, 10111 ]

## Customer 2
ID: 2
Title: Mr
Name: Jeff Hay
Telephone: [ Home: 0044-1999-333333, Mobile: 0044-17545-444444 ]
Address: [ UK: 86 High Street, Some Town, A County, GL8888, UK,
US: 777 7th Street, Another City, CA, 90111 ]
```

Neste exemplo, os campos são prefixados com um nome. Os campos também podem ter vários subcampos, também com nomes. No exemplo, vários subcampos são indicados ao colocá-los entre colchetes.

Adicionar um novo cliente é uma questão de inserir uma entidade com seus campos rotulados de forma significativa. Um aplicativo que consulta esses dados deve estar preparado para analisar as informações na entidade que ele recupera.

Os recursos de recuperação de dados de um banco de dados não relacional podem variar. Cada entidade deve ter um valor de chave exclusivo. As entidades em uma coleção geralmente são armazenadas na ordem dos valores-chave. No exemplo acima, a chave exclusiva é o campo ID. O tipo mais simples de banco de dados não relacional permite que um aplicativo especifique a chave exclusiva ou um intervalo de chaves como critérios de consulta.

No exemplo dos clientes, o banco de dados permitiria que um aplicativo consultasse os clientes apenas por ID. A filtragem de dados em outros campos exigiria a varredura de toda a coleção de entidades, analisando cada entidade por sua vez e, em seguida, aplicando quaisquer critérios de consulta a cada entidade para encontrar qualquer correspondência.

No exemplo abaixo, uma consulta que busca os detalhes de um cliente por ID pode identificar rapidamente qual entidade recuperar. Uma consulta que tenta localizar todos os clientes com um endereço no Reino Unido teria que iterar em todas as entidades e, para cada entidade, examinar cada campo por vez. Se o banco de dados contiver muitos milhões de entidades, essa consulta poderá levar um tempo considerável para ser executada.



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 000296691208

Sistemas não relacionais mais avançados suportam indexação, de maneira semelhante a um índice em um banco de dados relacional. As consultas podem então usar o índice para identificar e buscar dados com base em campos não-chave. Os sistemas não relacionais, como o Azure Cosmos DB (um sistema de gerenciamento de banco de dados não relacional disponível no Azure), oferecem suporte à indexação mesmo quando a estrutura dos dados indexados pode variar de registro para registro. Para mais informações, leia Indexação no Azure Cosmos DB - Visão geral.

Quando você projeta um banco de dados não relacional, é importante entender os recursos do sistema de gerenciamento de banco de dados e os tipos de consulta que ele deverá suportar.

**NOTA:** Bancos de dados não relacionais geralmente fornecem sua própria linguagem proprietária para gerenciar e consultar dados. Essa linguagem pode ser procedural ou pode ser semelhante ao SQL; depende de como o banco de dados é implementado pelo sistema de gerenciamento de banco de dados.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

### Identificar casos de uso de banco de dados não relacional

Os bancos de dados não relacionais são altamente adequados para os seguintes cenários:

- IoT e telemática. Esses sistemas geralmente ingerem grandes quantidades de dados em frequentes picos de atividade. Os bancos de dados não relacionais podem armazenar essas informações muito rapidamente. Os dados podem ser usados por serviços de análise, como Azure Machine Learning, Azure HDInsight e Microsoft Power BI. Além disso, você pode processar os dados em tempo real usando o Azure Functions que são acionados conforme os dados chegam ao banco de dados.
- Varejo e marketing. A Microsoft usa CosmosDB para suas próprias plataformas de comércio eletrônico que funcionam como parte da Windows Store e Xbox Live. Ele também é usado no setor de varejo para armazenar dados de catálogo e para sourcing de eventos em canais de processamento de pedidos.
- Jogos. A camada de banco de dados é um componente crucial dos aplicativos de jogos. Os jogos modernos executam processamento gráfico em clientes móveis / de console, mas contam com a nuvem para fornecer conteúdo personalizado e personalizado, como estatísticas no jogo, integração de mídia social e tabelas de classificação de alta pontuação. Os jogos geralmente exigem latências de um milissegundo para leituras e gravações para fornecer uma experiência envolvente no jogo. Um banco de dados de jogos precisa ser rápido e capaz de lidar com picos massivos nas taxas de solicitação durante o lançamento de novos jogos e atualizações de recursos.
- Aplicativos da web e móveis. Um banco de dados não relacional, como o Azure Cosmos DB, é comumente usado em aplicativos web e móveis e é adequado para modelar interações sociais, integração com serviços de terceiros e para criar experiências personalizadas ricas. Os Cosmos DB SDKs (kits de desenvolvimento de software) podem ser usados para criar aplicativos iOS e Android ricos, utilizando a popular estrutura Xamarin.

CPF: 00296691208

## Descreva os tipos de dados não relacionais

Os dados não relacionais geralmente se enquadram em duas categorias; semiestruturado e não estruturado. Nesta unidade, você aprenderá sobre o significado desses termos e verá alguns exemplos.

## O que são dados semiestruturados?

Dados semiestruturados são dados que contêm campos. Os campos não precisam ser iguais em todas as entidades. Você só define os campos necessários por entidade. As entidades do cliente mostradas na unidade anterior são exemplos de dados semiestruturados. Os dados devem ser formatados de forma que um aplicativo possa analisá-los e processá-los. Uma maneira comum de fazer isso é armazenar os dados de cada entidade como um documento JSON. O termo JSON significa JavaScript Object Notation. É o formato usado por aplicativos JavaScript para armazenar dados na memória, mas também pode ser usado para ler e gravar documentos de e para arquivos.

Um documento JSON é colocado entre chaves ({e}). Cada campo possui um nome (um rótulo), seguido por dois pontos e o valor do campo. Os campos podem conter valores simples ou subdocumentos (cada um começando e terminando com colchetes). Os campos também podem ter vários valores, mantidos como matrizes e entre colchetes ([e]). Literais em um campo são colocados entre aspas e os campos são separados por vírgulas.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

O exemplo abaixo mostra os clientes da unidade anterior, formatados como documentos JSON:

```
{  
    "ID": "1",  
    "Name": "Mark Hanson",  
    "Telephone": [  
        { "Home": "1-999-9999999" },  
        { "Business": "1-888-8888888" },  
        { "Cell": "1-777-7777777" }  
    ],  
    "Address": [  
        { "Home": [  
            { "StreetAddress": "121 Main Street" },  
            { "City": "Some City" },  
            { "State": "NY" },  
            { "Zip": "10110" }  
        ]},  
        { "Business": [  
            { "StreetAddress": "87 Big Building" },  
            { "City": "Some City" },  
            { "State": "NY" },  
            { "Zip": "10111" }  
        ]}  
    ]  
}  
  
{  
    "ID": "2",  
    "Title": "Mr",  
    "Name": "Jeff Hay",  
    "Telephone": [  
        { "Home": "0044-1999-333333" },  
        { "Mobile": "0044-17545-444444" }  
    ],  
    "Address": [  
        { "UK": [  
            { "StreetAddress": "86 High Street" },  
            { "Town": "Some Town" },  
            { "County": "A County" },  
            { "Postcode": "GL8888" },  
            { "Region": "UK" }  
        ]},  
        { "US": [  
            { "StreetAddress": "777 7th Street" },  
            { "City": "Another City" },  
            { "State": "CA" },  
            { "Zip": "90111" }  
        ]}  
    ]  
}
```

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Você é livre para definir os campos que desejar. O importante é que os dados sigam a gramática JSON. Quando um aplicativo lê um documento, ele pode usar um analisador JSON para dividir o documento em seus campos de componentes e extrair as partes individuais dos dados.

Outros formatos que você pode ver incluem Avro, ORC, e Parquet:

- Avro é um formato baseado em linha. Foi criado pela Apache. Cada registro contém um cabeçalho que descreve a estrutura dos dados no registro. Este cabeçalho é armazenado como JSON. Os dados são armazenados como informações binárias. Um aplicativo usa as informações no cabeçalho para analisar os dados binários e extrair os campos que ele contém. Avro é um formato muito bom para compactar dados e minimizar os requisitos de armazenamento e largura de banda da rede.
- ORC (*Optimized Row Columnar format*) organiza os dados em colunas em vez de linhas. Ele foi desenvolvido pela HortonWorks para otimizar as operações de leitura e gravação no Apache Hive. O Hive é um sistema de data warehouse que suporta summarização rápida de dados e consulta em conjuntos de dados muito grandes. O Hive oferece suporte a consultas semelhantes a SQL em dados não estruturados. Um arquivo ORC contém listras de dados. Cada faixa contém os dados de uma coluna ou conjunto de colunas. Uma faixa contém um índice nas linhas da faixa, os dados de cada linha e um rodapé que contém informações estatísticas (contagem, soma, máximo, mínimo e assim por diante) para cada coluna.
- Parquet outro formato de dados colunar. Foi criado por Cloudera e Twitter. Um arquivo Parquet contém grupos de linhas. Os dados de cada coluna são armazenados juntos no mesmo grupo de linhas. Cada grupo de linhas contém um ou mais blocos de dados. Um arquivo Parquet inclui metadados que descrevem o conjunto de linhas encontradas em cada bloco. Um aplicativo pode usar esses metadados para localizar rapidamente a parte correta para um determinado conjunto de linhas e recuperar os dados nas colunas especificadas para essas linhas. A Parquet é especializada em armazenar e processar tipos de dados aninhados com eficiência. Suporta compressão e esquemas de codificação muito eficientes.

CPF 00296691208

## O que são dados não estruturados?

Dados não estruturados são dados que não contêm campos naturalmente. Os exemplos incluem vídeo, áudio e outros fluxos de mídia. Cada item é uma bolha amorfa de dados binários. Você não pode pesquisar por elementos específicos nestes dados.

Você pode optar por armazenar dados como esse em um armazenamento projetado especificamente para essa finalidade. No Azure, você provavelmente armazenaria dados de vídeo e áudio, como blobs de bloco, em uma conta de Armazenamento. Um blob de bloco oferece suporte apenas para operações básicas de leitura e gravação e não tem capacidade de pesquisa interna.

Você também pode considerar os arquivos como uma forma de dados não estruturados, embora em alguns casos um arquivo possa incluir metadados que indicam o tipo de arquivo (fotografia, documento do Word, planilha do Excel e assim por diante), proprietário e outros elementos que podem ser armazenados como campos. No entanto, o conteúdo principal do arquivo não é estruturado.

## Descrever tipos de bancos de dados não relacionais e NoSQL

Dados não relacionais é um termo abrangente que significa qualquer coisa não estruturada como um conjunto de tabelas. Existem muitos tipos diferentes de dados não estruturados e as informações são usadas para uma ampla variedade de propósitos. Consequentemente, existem vários tipos diferentes de sistemas de gerenciamento de banco de dados não relacionais, cada um orientado para um conjunto específico de cenários.

Nesta unidade, você aprenderá sobre alguns dos tipos mais comuns de bancos de dados não relacionais.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## O que é NoSQL?

Você pode ver o termo NoSQL ao ler sobre bancos de dados não relacionais. NoSQL é um termo um tanto vago que significa simplesmente não relacional. Há algum debate sobre se a intenção é sugerir Não é SQL, ou Não apenas SQL. Alguns bancos de dados não relacionais oferecem suporte a uma versão do SQL adaptada para documentos em vez de tabelas (exemplos incluem Azure Cosmos DB).

Os bancos de dados NoSQL (não relacionais) geralmente se enquadram em quatro categorias: armazenamentos de valores-chave, bancos de dados de documentos, bancos de dados de famílias de colunas e bancos de dados de gráficos. As seções a seguir discutem esses tipos de bancos de dados NoSQL.

## O que é um armazenamento de valor-chave?

Um armazenamento de valor-chave é o tipo mais simples (e geralmente o mais rápido) de banco de dados NoSQL para inserir e consultar dados. Cada item de dados em um armazenamento de valor-chave possui dois elementos, uma chave e um valor. A chave identifica exclusivamente o item, e o valor contém os dados do item. O valor é *opaco* para o sistema de gerenciamento de banco de dados. Os itens são armazenados em ordem de chave.

**NOTA:** O termo *opaco* significa que o sistema de gerenciamento de banco de dados apenas vê o valor como um bloco não estruturado. Apenas o aplicativo entende como os dados no valor são estruturados e quais campos eles contêm. O oposto de opaco é transparente. Se os dados forem transparentes, o sistema de gerenciamento de banco de dados entende como os campos nos dados são organizados. Uma tabela relacional é um exemplo de estrutura transparente.

| Key    | Value                        |
|--------|------------------------------|
| AAAAAA | 1101001111010100110101111... |
| AABAB  | 1001100001011001101011110... |
| DFA766 | 0000000000101010110101010... |
| FABCC4 | 1110110110101010100101101... |

Opaque to data store

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

Uma consulta especifica as chaves para identificar os itens a serem recuperados. Você não pode pesquisar valores. Um aplicativo que recupera dados de um armazenamento de valor-chave é responsável por analisar o conteúdo dos valores retornados.

As operações de gravação são restritas a inserções e exclusões. Se você precisar atualizar um item, deverá recuperá-lo, modificá-lo na memória (no aplicativo) e, em seguida, gravá-lo de volta no banco de dados, sobrescrevendo o original (excluir e inserir).

O foco de um armazenamento de valor-chave é a capacidade de ler e gravar dados muito rapidamente. Os recursos de pesquisa são secundários. Um armazenamento de valor-chave é uma excelente escolha para ingestão de dados, quando um grande volume de dados chega como um fluxo contínuo e deve ser armazenado imediatamente.

O armazenamento de tabelas do Azure é um exemplo de armazenamento de valores-chave. Cosmos DB também implementa um armazenamento de valor-chave usando o API de tabela.

### O que é um banco de dados de documentos?

Um banco de dados de documentos representa a extremidade oposta do espectro NoSQL de um armazenamento de valor-chave. Em um banco de dados de documentos, cada documento possui um ID exclusivo, mas os campos nos documentos são transparentes para o sistema de gerenciamento de banco de dados. Os bancos de dados de documentos normalmente armazenam dados no formato JSON, conforme descrito na unidade anterior, ou podem ser codificados usando outros formatos como XML, YAML, JSON, BSON. Os documentos podem até ser armazenados como texto simples. Os campos em documentos são expostos ao sistema de gerenciamento de armazenamento, permitindo que um aplicativo consulte e filtre dados usando os valores desses campos.

Normalmente, um documento contém todos os dados de uma entidade. Quais itens constituem uma entidade são específicos do aplicativo. Por exemplo, uma entidade pode conter os detalhes de um cliente, um pedido ou uma combinação de ambos. Um único documento pode conter informações que seriam espalhadas por várias tabelas relacionais em um RDBMS (sistema de gerenciamento de banco de dados relacional).

Um armazenamento de documentos não exige que todos os documentos tenham a mesma estrutura. Essa abordagem de forma livre oferece uma grande flexibilidade. Os aplicativos podem armazenar dados diferentes em documentos conforme os requisitos de negócios mudam.

CPF 00296691208

| Key  | Document  |
|------|---|
| 1001 | <pre>{   "CustomerID": 99,   "OrderItems": [     { "ProductID": 2010,       "Quantity": 2,       "Cost": 520     },     { "ProductID": 4365,       "Quantity": 1,       "Cost": 18     }   ],   "OrderDate": "04/01/2017" }</pre> |
| 1002 | <pre>{   "CustomerID": 220,   "OrderItems": [     { "ProductID": 1285,       "Quantity": 1,       "Cost": 120     }   ],   "OrderDate": "05/08/2017" }</pre>  |

Um aplicativo pode recuperar documentos usando a chave de documento. A chave é um identificador exclusivo do documento. Alguns bancos de dados de documentos criam a chave do documento automaticamente. Outros permitem que você especifique um atributo do documento para usar como chave. O aplicativo também pode consultar documentos com base no valor de um ou mais campos. Alguns bancos de dados de documentos oferecem suporte à indexação para facilitar a pesquisa rápida de documentos com base em um ou mais campos indexados.

Alguns sistemas de gerenciamento de banco de dados de documentos oferecem suporte a atualizações no local, permitindo que um aplicativo modifique os valores de campos específicos em um documento sem reescrever todo o documento. Outros sistemas de gerenciamento de banco de dados de documentos (como Cosmos DB) só podem ler e gravar documentos inteiros. Nestes casos, uma atualização substitui todo o documento por uma nova versão. Essa abordagem ajuda a reduzir a fragmentação no banco de dados, o que pode, por sua vez, melhorar o desempenho.

A maioria dos bancos de dados de documentos ingerem grandes volumes de dados mais rapidamente do que um banco de dados relacional, mas não são tão ideais quanto um armazenamento de valor-chave para esse tipo de processamento. O foco de um banco de dados de documentos são seus recursos de consulta.

O Azure Cosmos DB implementa uma abordagem de banco de dados de documentos em sua API Core (SQL).

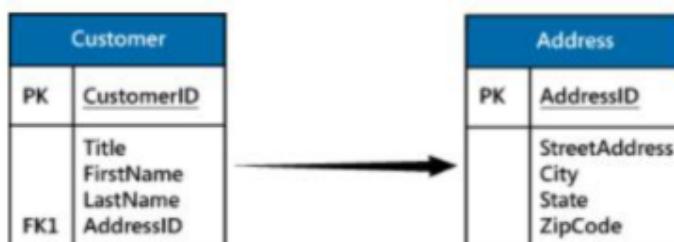
THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## O que é um banco de dados de família de colunas?

Um banco de dados de família de colunas organiza os dados em linhas e colunas. Exemplos dessa estrutura incluem arquivos ORC e Parquet, descritos na unidade anterior.

Em sua forma mais simples, um banco de dados de família de colunas pode parecer muito semelhante a um banco de dados relacional, pelo menos conceitualmente. O verdadeiro poder de um banco de dados de família de colunas está em sua abordagem desnormalizada para estruturar dados esparsos.

Por exemplo, se você precisar armazenar informações sobre clientes e seus endereços em um banco de dados relacional (ignorando a necessidade de manter dados históricos conforme descrito na seção anterior), você pode criar um esquema semelhante ao mostrado abaixo. Este diagrama também mostra alguns dados de amostra. Neste exemplo, o cliente 1 e o cliente 3 compartilham o mesmo endereço, e o esquema garante que essas informações de endereço não sejam duplicadas. Esta é uma maneira padrão de implementar um relacionamento um-para-muitos.



Customer Table

| CustomerID | Title | FirstName | LastName | AddressID |
|------------|-------|-----------|----------|-----------|
| 1          | Mr    | Mark      | Hanson   | 500       |
| 2          | Ms    | Lisa      | Andrews  | 501       |
| 3          | Mr    | Walter    | Harp     | 500       |

Address Table

| AddressID | StreetAddress   | City     | State | ZipCode |
|-----------|-----------------|----------|-------|---------|
| 500       | 999 500th Ave   | Bellevue | WA    | 12345   |
| 501       | 888 W. Front St | Boise    | ID    | 54321   |

THIAGO FERNANDES DA SILVA  
CPF 00296691208

O modelo relacional oferece suporte a uma abordagem muito generalizada para implementar esse tipo de relacionamento, mas para encontrar o endereço de qualquer cliente, um aplicativo precisa executar uma consulta que une duas tabelas. Se esta for a consulta mais comum realizada pelo aplicativo, a sobrecarga associada à execução dessa operação de junção pode rapidamente se tornar significativa se houver um grande número de solicitações e as próprias tabelas forem grandes.

O objetivo de um banco de dados de família de colunas é lidar de forma eficiente com situações tal como esta. Você pode pensar em um banco de dados de família de colunas como contendo dados tabulares compreendendo linhas e colunas, mas pode dividir as colunas em grupos conhecidos, como famílias de colunas.

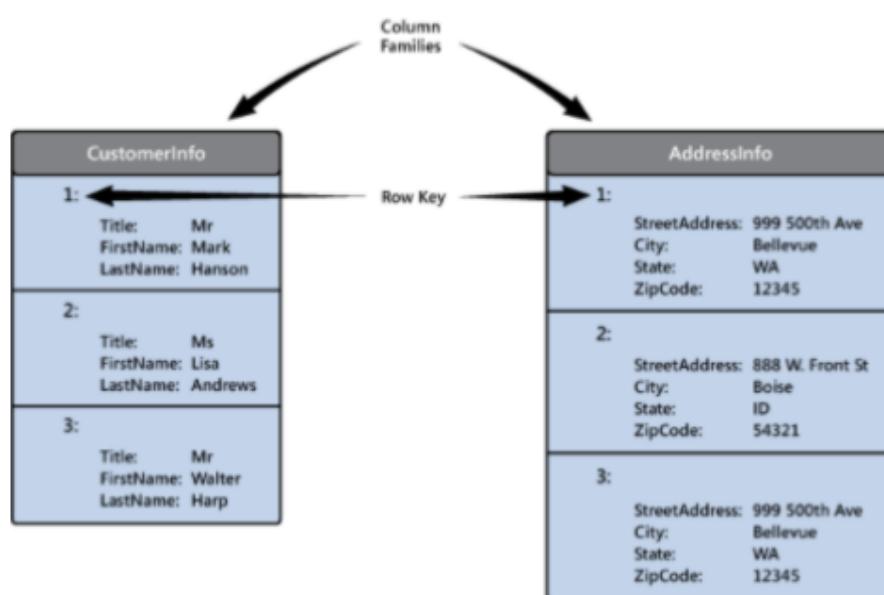
Cada família de colunas contém um conjunto de colunas que estão logicamente relacionadas entre si. A imagem abaixo mostra uma maneira de estruturar as mesmas informações da imagem anterior, usando um banco de dados de família de colunas para agrupar os dados em duas famílias de colunas contendo o nome do cliente e informações de endereço. Outras maneiras de organizar as colunas são possíveis, mas você deve implementar suas famílias de colunas para otimizar as consultas mais comuns que seu aplicativo executa. Nesse caso, as consultas que recuperam os endereços dos clientes podem buscar os dados com menos leituras do que seria necessário no banco de dados relacional correspondente; essas consultas podem buscar os dados diretamente da família de colunas AddressInfo.

| Row Key | Column Families   |                       |   |
|---------|---|-----------------------|---|
|         | CustomerID  | CustomerInfo          | AddressInfo   |
| 1       | CustomerInfo:Title<br>CustomerInfo:FirstName<br>CustomerInfo:LastName | Mr<br>Mark<br>Hanson  | AddressInfo:StreetAddress 999 500th Ave<br>AddressInfo:City Bellevue<br>AddressInfo:State WA<br>AddressInfo:ZipCode 12345 |
| 2       | CustomerInfo:Title<br>CustomerInfo:FirstName<br>CustomerInfo:LastName | Ms<br>Lisa<br>Andrews | AddressInfo:StreetAddress 888 W. Front St<br>AddressInfo:City Boise<br>AddressInfo:State ID<br>AddressInfo:ZipCode 54321  |
| 3       | CustomerInfo:Title<br>CustomerInfo:FirstName<br>CustomerInfo:LastName | Mr<br>Walter<br>Harp  | AddressInfo:StreetAddress 999 500th Ave<br>AddressInfo:City Bellevue<br>AddressInfo:State WA<br>AddressInfo:ZipCode 12345 |

A ilustração acima é mais conceitual do que física, e tem como objetivo mostrar a estrutura lógica dos dados - e não como eles podem ser organizados fisicamente. Cada linha em um banco de dados de família de colunas contém uma chave e você pode buscar os dados de uma linha usando esta chave.

Na maioria dos bancos de dados de famílias de colunas, as famílias de colunas são armazenadas separadamente. No exemplo anterior, a família de colunas CustomerInfo pode ser mantida em uma área de armazenamento físico e a família de colunas AddressInfo em outra, em uma forma simples de particionamento vertical.

Você realmente deve pensar na estrutura em termos de famílias de colunas, em vez de linhas. Os dados de uma única entidade que abrange vários grupos de colunas terão a mesma chave de linha em cada grupo de colunas. Como alternativa ao layout conceitual mostrado anteriormente, você pode visualizar os dados mostrados como o seguinte par de estruturas físicas.



O sistema de gerenciamento de banco de dados de família de colunas mais usado é o Apache Cassandra. O Azure Cosmos DB oferece suporte à abordagem de família de colunas por meio da API Cassandra.

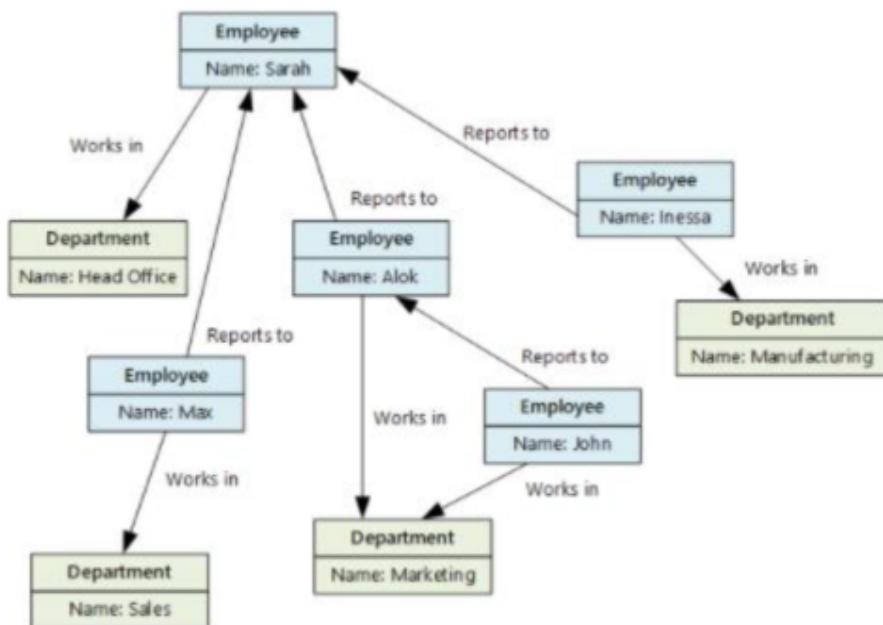
### O que é um banco de dados gráfico?

Os bancos de dados gráficos permitem que você armazene entidades, mas o foco principal está nos relacionamentos que essas entidades têm entre si. Um banco de dados gráfico armazena dois tipos de informações: nós (que você pode considerar como instâncias de entidades) e arestas (que especificam os relacionamentos entre os nós).

Nós e arestas podem ter propriedades que fornecem informações sobre esse nó ou aresta (como colunas em uma tabela). Além disso, as arestas podem ter uma direção que indica a natureza do relacionamento.

O objetivo de um banco de dados de gráficos é permitir que um aplicativo execute com eficiência consultas que atravessam a rede de nós e arestas e analise os relacionamentos entre as entidades. A imagem abaixo mostra o banco de dados de

pessoal de uma organização estruturado como um gráfico. As entidades são os funcionários e os departamentos da organização, e as bordas indicam as linhas hierárquicas e o departamento em que os funcionários trabalham. Neste gráfico, as setas nas bordas mostram a direção dos relacionamentos.



Uma estrutura como essa facilita a realização de consultas como "Encontre todos os funcionários que trabalham direta ou indiretamente para Sarah" ou "Quem trabalha no mesmo departamento que John?".

Para gráficos grandes com muitas entidades e relacionamentos, você pode realizar análises muito complexas muito rapidamente, e muitos bancos de dados de gráficos fornecem uma linguagem de consulta que você pode usar para percorrer uma rede de relacionamentos com eficiência. Frequentemente, você pode armazenar as mesmas informações em um banco de dados relacional, mas o SQL necessário para consultar essas informações pode exigir muitas operações de junção recursivas caras e subconsultas aninhadas.

O Azure Cosmos DB oferece suporte a bancos de dados gráficos usando o API Gremlin. A API Gremlin é uma linguagem padrão para criar e consultar gráficos.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Teste de conhecimento

### Questão 1

**Qual dos seguintes serviços você deve usar para implementar um banco de dados não relacional?**

- a) Azure Cosmos DB
- b) Banco de dados SQL do Azure
- c) A API Gremlin

**Explicação:** está correto. Cosmos DB suporta vários modelos comuns de banco de dados não relacional, incluindo armazenamentos de valores-chave, bancos de dados de gráficos, bancos de dados de documentos e armazenamentos de famílias de colunas.

### Questão 2

**Qual das opções a seguir é uma característica dos bancos de dados não relacionais?**

- a) Bancos de dados não relacionais contêm tabelas com registros de coluna fixa simples
- b) Bancos de dados não relacionais exigem que você use técnicas de normalização de dados para reduzir a duplicação de dados
- c) Bancos de dados não relacionais são livres de esquemas ou têm esquemas relaxados

**Explicação:** está correto. Cada entidade em um banco de dados não relacional possui apenas os campos de que precisa, e esses campos podem variar entre diferentes entidades.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

### Questão 3

Você está construindo um sistema que monitora a temperatura em um conjunto de blocos de escritórios e define o ar-condicionado de cada cômodo em cada bloco para manter uma temperatura ambiente agradável. Seu sistema tem que gerenciar o ar condicionado em vários milhares de edifícios espalhados pelo país ou região, e cada edifício contém pelo menos 100 quartos com ar condicionado. Que tipo de armazenamento de dados NoSQL é mais apropriado para capturar os dados de temperatura para permitir que sejam processados rapidamente?

- a) Um armazenamento de valor-chave
- b) Um banco de dados de família de colunas
- c) Grave as temperaturas em um blob no armazenamento de Blob do Azure

Explicação: está correto. Um armazenamento de valor-chave pode ingerir grandes volumes de dados rapidamente. Um termômetro em cada sala pode enviar os dados para o banco de dados.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Resumo

O Microsoft Azure fornece uma variedade de tecnologias para armazenar dados não relacionais. Cada tecnologia tem seus próprios pontos fortes e é adequada a cenários específicos.

## Você explorou:

- As características dos dados não relacionais
- Diferentes tipos de dados não relacionais
- NoSQL e os tipos de bancos de dados não relacionais

## Saber mais

- Escolha o armazenamento de dados certo
- Bem-vindo ao Azure Cosmos DB
- Indexação no Azure Cosmos DB - Visão geral
- Introdução ao Azure Cosmos DB: API de tabela
- Introdução ao Azure Cosmos DB: API Gremlin
- Introdução ao armazenamento de Blob do Azure

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Introdução

As empresas de sucesso tomam decisões informadas para encontrar novas oportunidades, identificar pontos fracos, aumentar a eficiência e melhorar a satisfação do cliente. A análise de dados é o processo de examinar dados brutos para descobrir tendências e descobrir informações usadas para fazer e responder perguntas relacionadas ao desempenho organizacional.

Por exemplo, resorts e cassinos podem combinar dados de visitas anteriores de clientes para determinar o melhor momento para executar atividades e jogos específicos. Um analista de dados pode pegar dados como gastos do cliente e procurar correlações com outros fatores, como clima, eventos regionais ou até mesmo a presença (ou ausência) de incentivos, como comida e bebida.

Outro exemplo é o setor de saúde. Há uma abundância de dados no setor de saúde, incluindo registros de pacientes e informações de seguros. Como há tantos dados, pode ser difícil gerenciá-los. A análise de dados permite uma análise completa dos dados e pode levar a um diagnóstico ou plano de tratamento mais rápido.

Nesta lição, você explorará os principais elementos envolvidos na análise de dados. Você verá como coletar dados, processar dados para gerar informações e visualizar resultados para detectar tendências.

## Objetivos de aprendizado

Nesta lição você irá:

- Aprender mais sobre ingestão e processamento de dados
- Explorar a visualização de dados
- Explorar a análise de dados

THIAGO FERNANDES DA SILVA  
CPF 00296691208

## Descreva a ingestão e processamento de dados

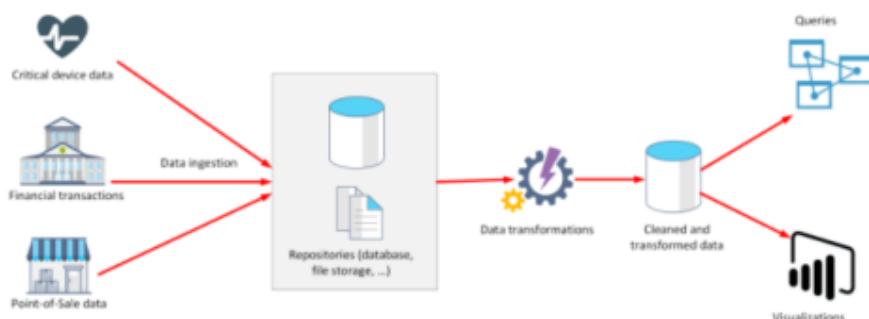
A análise de dados se preocupa em pegar os dados que sua organização produz e usá-los para estabelecer uma imagem de como está o desempenho de sua organização, ou que você pode fazer para manter o desempenho dos negócios. A análise de dados ajuda a identificar pontos fortes e fracos em sua organização e permite que você tome decisões de negócios adequadas.

Os dados que uma empresa usa podem vir de várias fontes. Pode haver uma massa de dados históricos para vasculhar e novos dados continuando a chegar o tempo todo. Esses dados podem ser o resultado de compras de clientes, transações bancárias, movimentos de preços de ações, dados meteorológicos em tempo real, dispositivos de monitoramento ou até mesmo câmeras. Em uma solução de análise de dados, você combina esses dados e constrói um data warehouse, que pode ser usado para fazer (e responder) perguntas sobre suas operações de negócios.

Construir um warehouse de dados requer que você possa capturar os dados de que precisa e disputá-los em um formato apropriado. Você pode então usar ferramentas de análise e visualizações para examinar as informações e tendências de identidade e suas causas.

**NOTA:** Wrangling é o processo pelo qual você transforma e mapeia dados brutos em um formato mais útil para análise. Pode envolver escrever código para capturar, filtrar, limpar, combinar e agregar dados de muitas fontes.

Nesta unidade, você aprenderá sobre dois estágios importantes na análise de dados: ingestão de dados e processamento de dados. O diagrama abaixo mostra como esses estágios se encaixam.



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## O que é ingestão de dados?

A ingestão de dados é o processo de obtenção e importação de dados para uso imediato ou armazenamento em um banco de dados. Os dados podem chegar como um fluxo contínuo ou em lotes, dependendo da fonte. O objetivo do processo de ingestão é capturar esses dados e armazená-los. Esses dados brutos podem ser mantidos em um repositório como um sistema de gerenciamento de banco de dados, um conjunto de arquivos ou algum outro tipo de armazenamento rápido e facilmente acessível.

O processo de ingestão também pode realizar a filtragem. Por exemplo, a ingestão pode rejeitar dados suspeitos, corrompidos ou duplicados. Dados suspeitos podem ser dados vindos de uma fonte inesperada. Dados corrompidos ou duplicados podem gerados devido a um erro do dispositivo, falha de transmissão ou violação.

Também pode ser possível realizar algumas transformações neste estágio, convertendo os dados em um formato padrão para processamento posterior. Por exemplo, você pode querer reformatar todos os dados de data e hora para usar as mesmas representações de data e hora e converter todos os dados de medição para usar as mesmas unidades. No entanto, essas transformações devem ser rápidas de realizar. Não tente executar cálculos complexos ou agregações nos dados neste estágio.

## O que é processamento de dados?

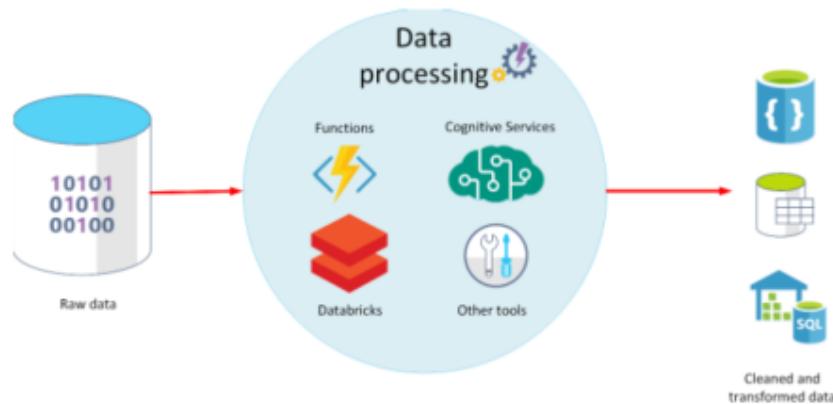
O estágio de processamento de dados ocorre após os dados serem ingeridos e coletados. O processamento de dados pega os dados em sua forma bruta, limpa-os e os converte em um formato mais significativo (tabelas, gráficos, documentos e assim por diante). O resultado é um banco de dados de dados que você pode usar para realizar consultas e gerar visualizações, dando-lhes a forma e o contexto necessários para serem interpretados por computadores e usados por funcionários em toda a organização.

**NOTA:** Limpeza de dados é um termo generalizado que abrange uma gama de ações, como remover anomalias e aplicar filtros e transformações que seriam muito demorados para serem executados durante o estágio de ingestão.

TUTORIO DE DATAS

CPF: 00296691208

O objetivo do processamento de dados é converter os dados brutos em um ou mais modelos de negócios. Um modelo de negócios descreve os dados em termos de entidades de negócios significativas e pode agregar itens e resumir informações. O estágio de processamento de dados também pode gerar modelos preditivos ou outros modelos analíticos a partir dos dados. O processamento de dados pode ser complexo e pode envolver scripts automatizados e ferramentas como o Azure Databricks, Azure Functions e Azure Cognitive Services para examinar e reformatar os dados e gerar modelos. Um analista de dados pode usar o aprendizado de máquina para ajudar a determinar tendências futuras com base nesses modelos.



### O que é ELT e ETL?

O mecanismo de processamento de dados pode assumir duas abordagens para recuperar os dados ingeridos, processar esses dados para transformá-los e gerar modelos e, em seguida, salvar os dados e modelos transformados. Essas abordagens são conhecidas como ETL e ELT.

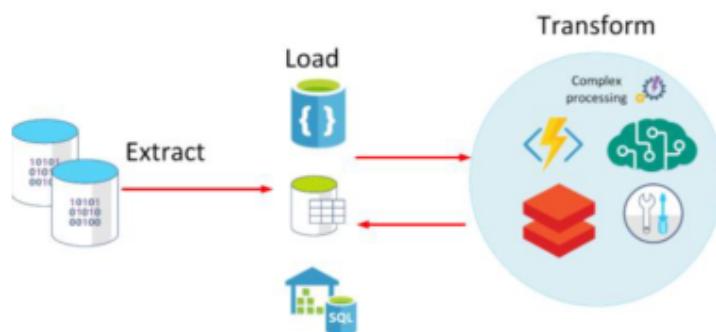
ETL significa Extrair, transformar e carregar. Os dados brutos são recuperados e transformados antes de serem salvos. As etapas de extração, transformação e carregamento podem ser realizadas como um pipeline contínuo de operações. É adequado para sistemas que requerem apenas modelos simples, com pouca dependência entre os itens. Por exemplo, esse tipo de processo é frequentemente usado para tarefas básicas de limpeza de dados, desduplicação de dados e reformatação do conteúdo de campos individuais.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208



Uma abordagem alternativa é ELT. ELT é uma abreviatura de Extrair, carregar e transformar. O processo difere do ETL porque os dados são armazenados antes de serem transformados. O mecanismo de processamento de dados pode ter uma abordagem iterativa, recuperando e processando os dados do armazenamento, antes de gravar os dados e modelos transformados de volta no armazenamento. ELT é mais adequado para construir modelos complexos que dependem de vários itens no banco de dados, geralmente usando processamento em lote periódico.



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

ELT é uma abordagem escalonável adequada para a nuvem porque pode usar o amplo poder de processamento disponível. A abordagem mais orientada para fluxo do ETL coloca mais ênfase no rendimento.

No entanto, o ETL pode filtrar os dados antes de serem armazenados. Dessa forma, o ETL pode ajudar com a privacidade e conformidade dos dados, removendo dados confidenciais antes que eles cheguem aos seus modelos de dados analíticos.

O Azure oferece várias opções que você pode usar para implementar as abordagens ELT e ETL. Por exemplo, se você estiver armazenando dados no Banco de Dados SQL do Azure, poderá usar o SQL Server Integration Services. O Integration Services pode extrair e transformar dados de uma ampla variedade de fontes, como arquivos de dados XML, arquivos simples e fontes de dados relacionais e, em seguida, carregar os dados em um ou mais destinos.

Outra abordagem mais generalizada é usar o Azure Data Factory. O Azure Data Factory é um serviço de integração de dados baseado em nuvem que permite criar fluxos de trabalho orientados a dados para orquestrar a movimentação de dados e transformá-los em escala. Usando o Azure Data Factory, você pode criar e agendar fluxos de trabalho baseados em dados (chamados de pipelines) que podem ingerir dados de armazenamentos de dados distintos.

Você pode construir processos ETL complexos que transformam dados visualmente com fluxos de dados ou usando serviços de computação, como Azure HDInsight Hadoop, Azure Databricks, e Banco de dados SQL do Azure.

### Explore a visualização de dados

Um modelo de negócios pode conter uma enorme quantidade de informações. O objetivo de produzir um modelo como este é ajudá-lo a raciocinar sobre as informações que ele contém, fazer perguntas e, com sorte, obter respostas que possam ajudá-lo a impulsionar seus negócios.

Esta unidade discute algumas das técnicas que você pode usar para analisar e compreender as informações em seus modelos.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## O que é relatório?

Relatório é o processo de organização de dados em resumos informativos para monitorar o desempenho de diferentes áreas de uma organização. Os relatórios ajudam as empresas a monitorar seus negócios online e saber quando os dados estão fora dos intervalos esperados. Bons relatórios devem levantar questões sobre o negócio de seus usuários finais. Os relatórios mostram o que aconteceu, enquanto a análise se concentra em explicar por que aconteceu e o que você pode fazer a respeito.

## O que é business intelligence?

O termo Business Intelligence (BI) refere-se a tecnologias, aplicativos e práticas para a coleta, integração, análise e apresentação de informações de negócios. O objetivo da inteligência de negócios é apoiar a melhor tomada de decisão.

Os sistemas de inteligência de negócios fornecem visões históricas, atuais e preditivas das operações de negócios, na maioria das vezes usando dados que foram reunidos em um data warehouse e, ocasionalmente, trabalhando a partir de dados operacionais ao vivo. Os elementos do software suportam relatórios, análise interativa de tabela dinâmica "slice-and-dice", visualização e mineração de dados estatísticos.

Os aplicativos lidam com vendas, produção, finanças e muitas outras fontes de dados de negócios para fins que incluem gerenciamento de desempenho de negócios. Frequentemente, são coletadas informações sobre outras empresas do mesmo setor para fins de comparação. Este processo é conhecido como avaliação comparativa.

## O que é visualização de dados?

A visualização de dados é a representação gráfica de informações e dados. Ao usar elementos visuais como tabelas, gráficos e mapas, as ferramentas de visualização de dados fornecem uma maneira acessível de identificar e compreender tendências, valores discrepantes e padrões nos dados. Se você estiver usando o Azure, a ferramenta de visualização de dados mais popular é o Power BI.

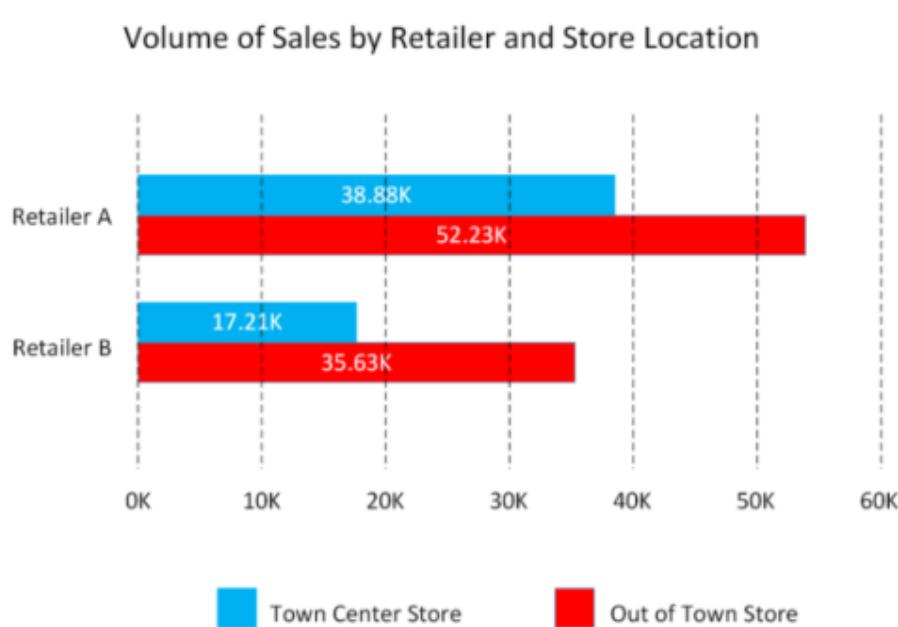
Usando o Power BI, você pode se conectar a várias fontes diferentes de dados e combiná-los em um modelo de dados. Este modelo de dados permite construir visuais e coleções de visuais, que você pode compartilhar como relatórios com outras pessoas dentro de sua organização.

CPF 000296691208

### Explore as opções de visualização para representar os dados

A visualização de dados ajuda você a se concentrar no significado dos dados, em vez de olhar para os próprios dados. Uma boa visualização de dados permite que você identifique rapidamente tendências, anomalias e problemas potenciais. As formas mais comuns de visualização são:

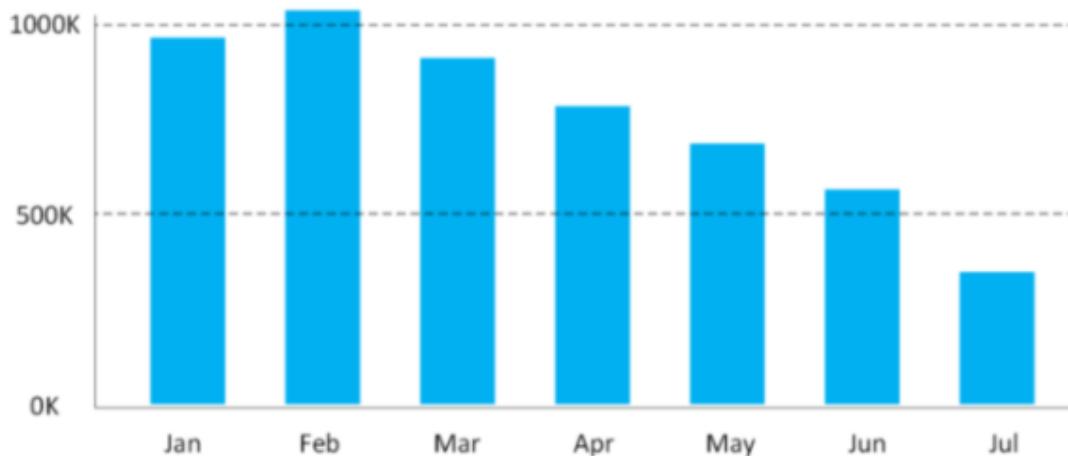
Gráficos de barras e colunas: Os gráficos de barras e colunas permitem que você veja como um conjunto de variáveis muda em diferentes categorias. Por exemplo, o primeiro gráfico abaixo mostra como as vendas de um par de varejistas fictícios variam entre os sites da loja.



Este gráfico mostra como as vendas variam por mês.

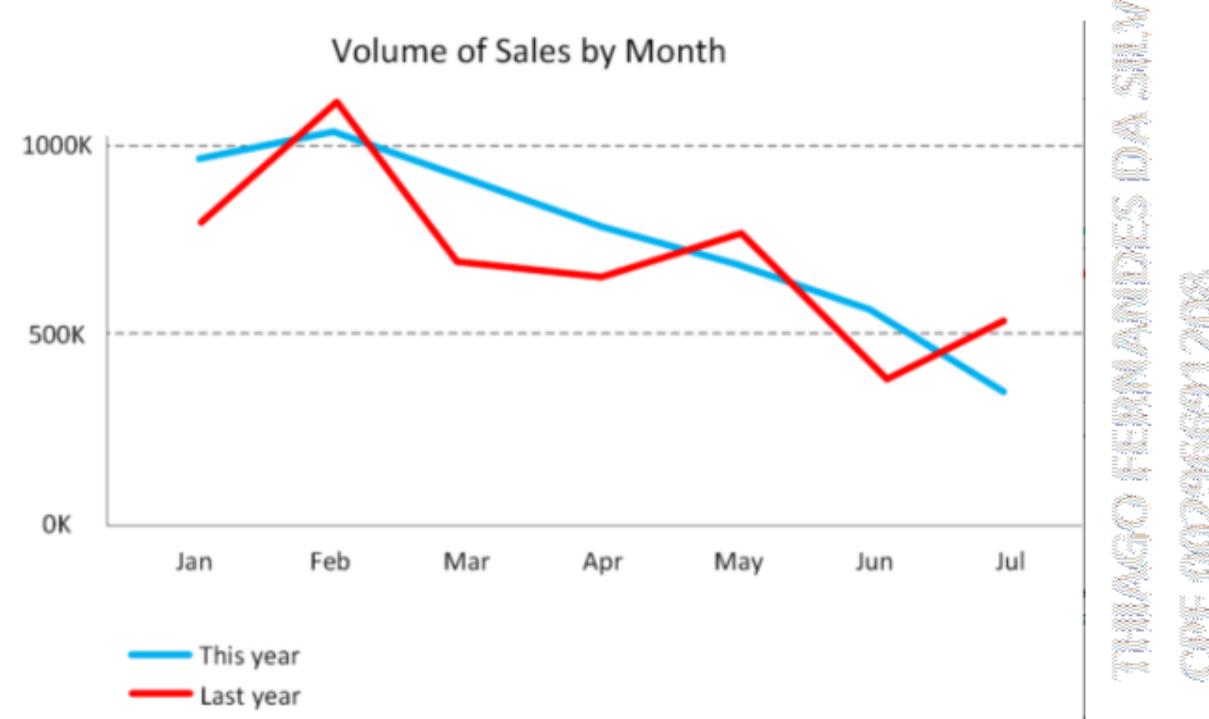
THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Volume of Sales by Month



Gráficos de linha: Os gráficos de linha enfatizam a forma geral de uma série inteira de valores, geralmente ao longo do tempo.

Volume of Sales by Month



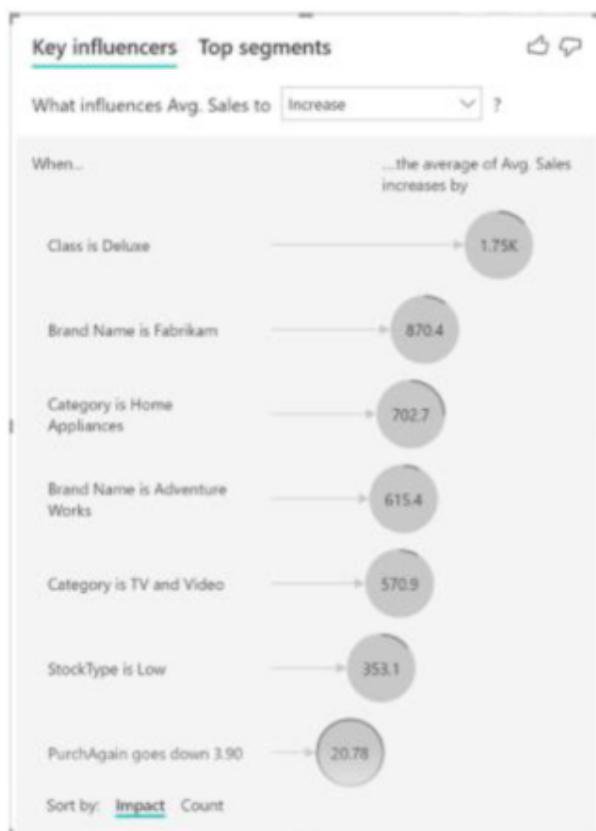
THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

Matriz: Uma matriz visual é uma estrutura tabular que resume os dados. Frequentemente, os designers de relatório incluem matrizes em relatórios e painéis para permitir que os usuários selecionem um ou mais elementos (linhas, colunas, células) na matriz para realçar outros elementos visuais em uma página de relatório.

| Quarter Year | Q1       |             | Q2       |             |
|--------------|----------|-------------|----------|-------------|
|              | Revenue  | YTD Revenue | Revenue  | YTD Revenue |
| 2015         | \$45,186 | \$45,186    | \$70,609 | \$115,795   |
| 2016         | \$52,154 | \$52,154    | \$73,542 | \$125,696   |
| 2017         | \$51,388 | \$51,388    | \$68,149 | \$118,537   |
| 2018         | \$48,281 | \$48,281    | \$66,853 | \$115,134   |
| 2019         | \$53,145 | \$53,145    | \$49,135 | \$102,280   |

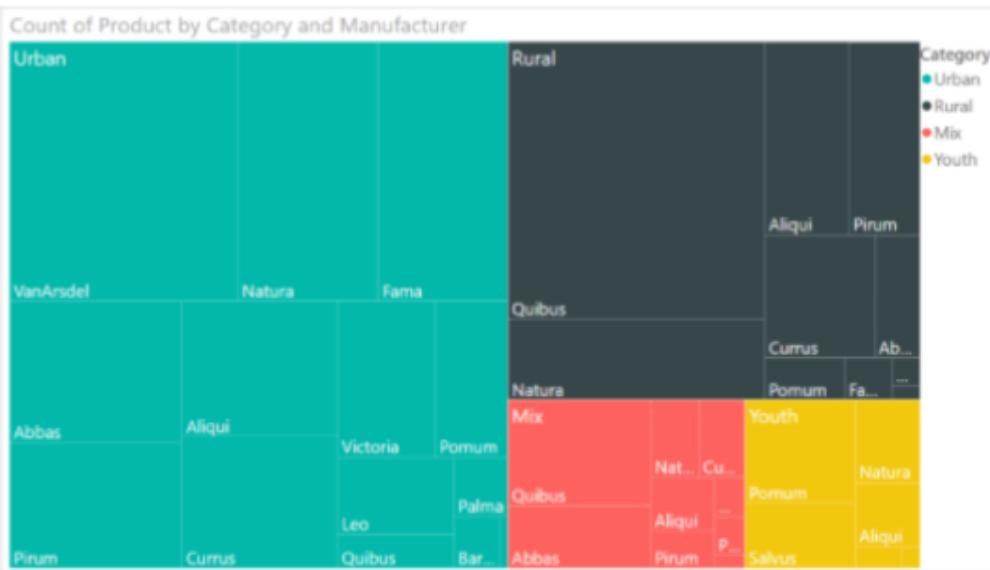
Principais influenciadores: Um gráfico de influenciador principal exibe os principais contribuintes para um resultado ou valor selecionado. Os influenciadores principais são uma ótima escolha para ajudá-lo a compreender os fatores que influenciam uma métrica principal. Por exemplo, o que influencia os clientes a fazerem um segundo pedido ou por que as vendas foram tão altas em junho passado.



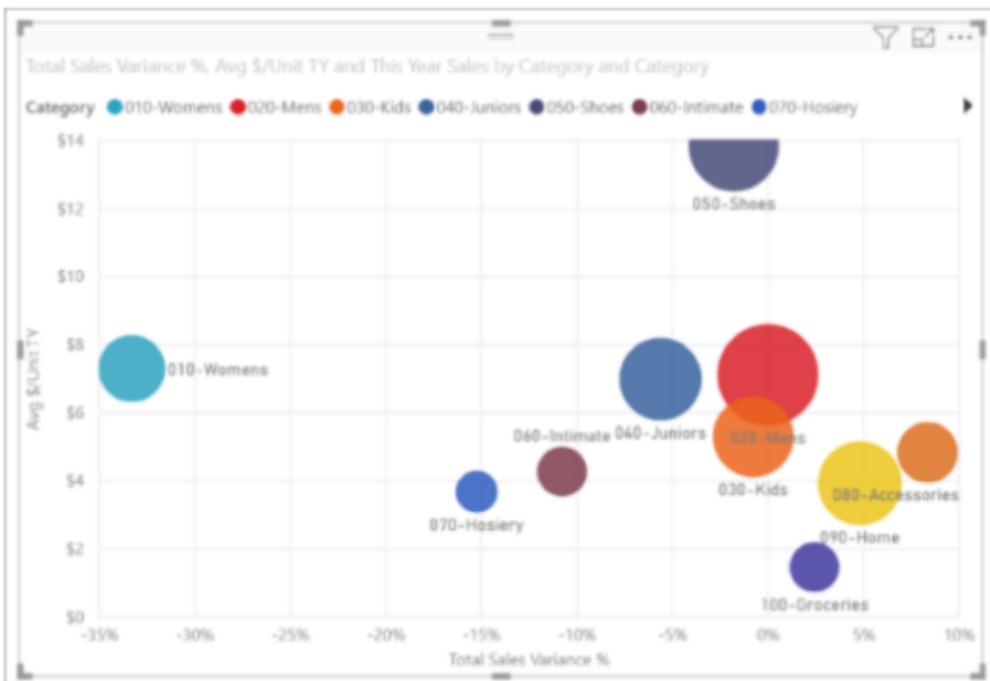
THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

Treemap: Treemaps são gráficos de retângulos coloridos, com o tamanho representando o valor relativo de cada item. Eles podem ser hierárquicos, com retângulos aninhados nos retângulos principais.

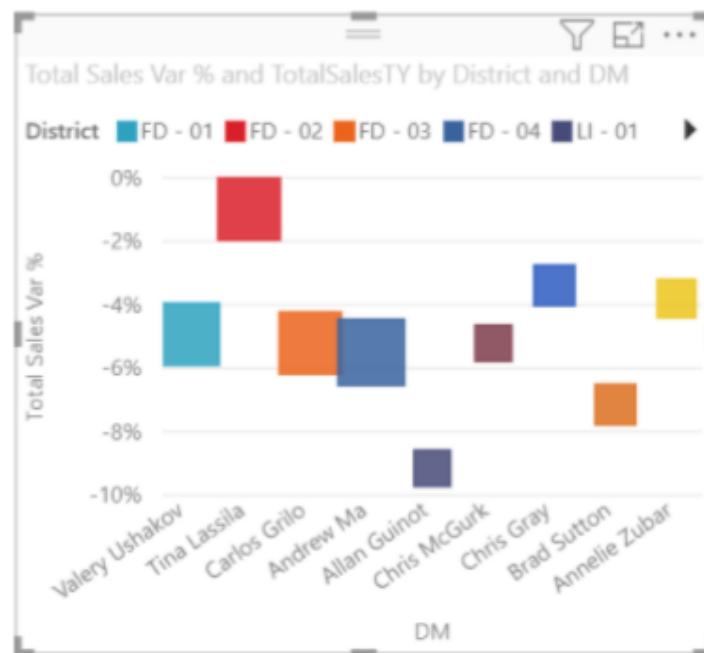


**Dispersão:** Um gráfico de dispersão mostra a relação entre dois valores numéricos. Um gráfico de bolhas é um gráfico de dispersão que substitui os pontos de dados por bolhas, com o tamanho da bolha representando uma terceira dimensão de dados adicionais.



Um gráfico de plot plot é semelhante a um gráfico de bolhas e um gráfico de dispersão, mas pode plotar dados categóricos ao longo do eixo X.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208



Mapa preenchido. Se você tiver dados geográficos, poderá usar um mapa preenchido para exibir como um valor difere em proporção em uma geografia ou região. Você pode ver diferenças relativas com sombreamento que varia de claro (menos frequente / mais baixo) a escuro (mais frequente / mais).



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Explore a análise de dados

A análise de dados se preocupa em examinar, transformar e organizar os dados para que você possa estudá-los e extrair informações úteis. A análise de dados é uma disciplina que cobre toda a gama de tarefas de gerenciamento de dados. Essas tarefas não incluem apenas análise, mas também coleta de dados, organização, armazenamento e todas as ferramentas e técnicas utilizadas.

O termo análise de dados é um resumo que cobre uma gama de atividades, cada uma com seu próprio foco e objetivos. Você pode categorizar essas atividades como descritivo, diagnóstico, preditivo, prescritivo, e cognitivo analítico.

Nesta unidade, você aprenderá sobre essas categorias de análise de dados.

## Análise descritiva

A análise descritiva ajuda a responder a perguntas sobre o que aconteceu, com base em dados históricos. As técnicas de análise descritiva resumem grandes conjuntos de dados para descrever os resultados às partes interessadas.

Ao desenvolver KPIs (indicadores-chave de desempenho), essas estratégias podem ajudar a rastrear o sucesso ou o fracasso de objetivos-chave. Métricas como retorno sobre o investimento (ROI) são usadas em muitos setores. Métricas especializadas são desenvolvidas para rastrear o desempenho em setores específicos.

Exemplos de análises descritivas incluem a geração de relatórios para fornecer uma visão das vendas e dados financeiros de uma organização.

## Análise de diagnóstico

A análise de diagnóstico ajuda a responder a perguntas sobre por que as coisas aconteceram. Técnicas de análise de diagnóstico complementam análises descritivas mais básicas. Eles pegam as descobertas da análise descritiva e se aprofundam para encontrar a causa. Os indicadores de desempenho são investigados posteriormente para descobrir por que eles melhoraram ou pioraram. Isso geralmente ocorre em três etapas:

- Identificar anomalias nos dados. Essas podem ser mudanças inesperadas em uma métrica ou em um mercado específico.
- Coletar dados relacionados a essas anomalias.
- Usar técnicas estatísticas para descobrir relacionamentos e tendências que explicam essas anomalias.

## Análise preditiva

A análise preditiva ajuda a responder a perguntas sobre o que acontecerá no futuro. As técnicas de análise preditiva usam dados históricos para identificar tendências e determinar se elas podem ocorrer novamente. Ferramentas analíticas preditivas fornecem informações valiosas sobre o que pode acontecer no futuro. As técnicas incluem uma variedade de técnicas estatísticas e de aprendizado de máquina, como redes neurais, árvores de decisão e regressão.

CPF: 00296691208

## Análise prescritiva

A análise prescritiva ajuda a responder a perguntas sobre quais ações devem ser tomadas para atingir uma meta ou alvo. Usando insights de análises preditivas, decisões baseadas em dados podem ser tomadas. Essa técnica permite que as empresas tomem decisões informadas diante da incerteza. As técnicas de análise prescritiva contam com estratégias de aprendizado de máquina para encontrar padrões em grandes conjuntos de dados. Ao analisar decisões e eventos passados, a probabilidade de resultados diferentes pode ser estimada.

## Analítica cognitiva

A análise cognitiva tenta tirar inferências de dados e padrões existentes, derivar conclusões com base em bases de conhecimento existentes e, em seguida, adicionar essas descobertas de volta à base de conhecimento para futuras inferências - um ciclo de feedback de autoaprendizagem. A análise cognitiva ajuda você a aprender o que pode acontecer se as circunstâncias mudarem e como você pode lidar com essas situações.

As inferências não são consultas estruturadas com base em um banco de dados de regras. São hipóteses não estruturadas, reunidas de várias fontes e expressas com vários graus de confiança. A análise cognitiva eficaz depende de algoritmos de aprendizado de máquina. Ele usa vários conceitos de PNL (Processamento de Linguagem Natural) para dar sentido a fontes de dados anteriormente inexploradas, como registros de conversas de call center e análises de produtos.

Aproveitando os benefícios da computação paralela / distribuída massiva e os custos decrescentes de armazenamento de dados e poder de computação, não há limite para o desenvolvimento cognitivo que esses sistemas podem alcançar.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Teste de conhecimento

### Questão 1

O que é ingestão de dados?

- a) O processo de transformar dados brutos em modelos contendo informações significativas
- b) Analisando dados para anomalias,
- c) Captura de streaming de dados brutos de várias fontes e armazenamento

Explicação: está correto. O objetivo da ingestão de dados é receber dados brutos e salvá-los o mais rápido possível. Os dados podem então ser processados e analisados.

### Questão 2

Qual dos seguintes recursos visuais exibe os principais contribuintes para um resultado ou valor selecionado?

- a) Principais influenciadores
- b) Gráfico de colunas e barras
- c) Gráfico de matriz

Explicação: está correto. Um gráfico de influenciador principal exibe os principais contribuintes para um resultado ou valor selecionado. Os influenciadores principais são uma ótima escolha para ajudá-lo a compreender os fatores que influenciam uma métrica principal.

### Questão 3

Que tipo de análise ajuda a responder a perguntas sobre o que aconteceu no passado?

- a) Análise descritiva
- b) Análise prescritiva
- c) Análise preditiva

Explicação: está correto. A análise descritiva ajuda a responder a perguntas sobre o que aconteceu.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Resumo

As organizações possuem enormes quantidades de dados. O objetivo da análise de dados é descobrir ideias importantes que podem ajudá-lo a impulsionar seus negócios.

Você explorou:

- Ingestão e processamento de dados
- Visualização de dados
- Análise de dados

## Saber mais

- Crie relatórios e painéis no Power BI - documentação
- Azure Databricks
- Serviços Cognitivos Azure
- Extrair, transformar e carregar (ETL)

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Módulo 2: Explore dados relacionais no Azure

### Introdução

Um banco de dados é uma coleção de dados. Ele pode ser tão simples quanto uma planilha de desktop ou tão complexo quanto um sistema global contendo petabytes de informações altamente estruturadas. Os dados podem ser estruturados de muitas maneiras diferentes. Uma abordagem comum é armazenar dados em um formato tabular, com linhas e colunas. Você pode definir relacionamentos entre tabelas. Esses bancos de dados são chamados de bancos de dados relacional.

Os bancos de dados também podem ser semi-estruturados ou não estruturados, compreendendo uma massa de dados brutos e não processados. Esses bancos de dados são normalmente chamados de não relacional. Os bancos de dados são gerenciados usando um sistema de gerenciamento de banco de dados (DBMS). O DBMS lida com os aspectos físicos de um banco de dados, como onde e como está armazenado, quem pode acessá-lo e como garantir que esteja disponível quando necessário.

Muitas organizações dependem das informações armazenadas em seus bancos de dados para ajudar a tomar decisões críticas de negócios. No passado, essas organizações executavam seus DBMSs no local. No entanto, essa abordagem exige que a organização mantenha sua própria infraestrutura de hardware. Portanto, um número crescente de empresas está migrando seus bancos de dados para a nuvem, onde os custos de configuração e manutenção da infraestrutura são bastante reduzidos.

Suponha que você seja um administrador de banco de dados na Wide World Importers. Você é responsável pelo design e manutenção do banco de dados, além de fornecer informações para a liderança e criar listas de clientes para o departamento de marketing. Você tem um banco de dados SQL Server existente que depende muito de procedimentos armazenados e outros recursos avançados de banco de dados, como servidores vinculados. O banco de dados está localizado em sua rede interna. Você foi solicitado a torná-lo globalmente disponível para seus parceiros em todo o mundo.

**NOTA:** A procedimento armazenado é um bloco de código executado dentro de seu banco de dados. Os aplicativos geralmente usam procedimentos armazenados porque são otimizados para serem executados no ambiente de banco de dados e podem acessar dados muito rapidamente. Um servidor vinculado é uma conexão de um servidor de banco de dados para outro. O SQL Server pode usar servidores vinculados para executar consultas em um servidor que podem incluir dados recuperados de outros servidores - estes são conhecidos como consultas distribuídas.

Nesta lição, você explorará as opções disponíveis ao escolher uma plataforma de dados relacional para hospedar um banco de dados no Azure.

## Objetivos de aprendizado

Nesta lição, você irá:

- Identificar serviços de dados relacionais do Azure
- Explorar as considerações na escolha de um serviço de dados relacionais

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## Explore os serviços relacionais de dados do Azure

O Azure oferece uma variedade de opções para executar um sistema de gerenciamento de banco de dados na nuvem. Por exemplo, você pode migrar seus sistemas locais para uma coleção de máquinas virtuais do Azure. Essa abordagem ainda requer que você gerencie seu DBMS com cuidado. Como alternativa você pode aproveitar as vantagens dos vários serviços de dados relacionais do Azure disponíveis. Esses serviços de dados gerenciam o DBMS para você, deixando-o livre para se concentrar nos dados que eles contêm e nos aplicativos que os utilizam.

## Compreenda IaaS, PaaS e SaaS

Antes de se aprofundar no Azure Data Services, você precisa entender alguns termos comuns usados para descrever as diferentes maneiras de hospedar um banco de dados no Azure.

IaaS é um acrônimo para Infraestrutura como um serviço. O Azure permite que você crie uma infraestrutura virtual na nuvem que espelha a maneira como um data center local pode funcionar. Você pode criar um conjunto de máquinas virtuais, conectá-las usando uma rede virtual e adicionar uma variedade de dispositivos virtuais. Você se responsabiliza pela instalação e configuração do software, como o DBMS, nessas máquinas virtuais. Em muitos aspectos, essa abordagem é semelhante à maneira como você executa seus sistemas dentro de uma organização, exceto que você não precisa se preocupar em comprar ou manter o hardware.

**NOTA:** Uma Rede Virtual do Azure é uma representação de sua própria rede na nuvem. Uma rede virtual permite que você conecte máquinas virtuais e serviços do Azure, da mesma forma que você pode usar uma rede física local. O Azure garante que cada rede virtual seja isolada de outras redes virtuais criadas por outros usuários e da Internet, permitindo que você especifique quais máquinas (reais e virtuais) e serviços têm permissão para acessar recursos na rede virtual e quais portas podem usar.

PaaS apoia Plataforma como serviço. Em vez de criar uma infraestrutura virtual e instalar e gerenciar o software de banco de dados você mesmo, uma solução PaaS faz isso por você. Você especifica os recursos de que precisa (com base no tamanho que você acha que seus bancos de dados serão, no número de usuários e no desempenho de que você precisa) e o Azure cria automaticamente as máquinas virtuais, redes e outros dispositivos necessários para você. Normalmente, você pode aumentar ou diminuir (aumentar ou diminuir o tamanho e o número de recursos) rapidamente, pois o volume de dados e a quantidade de trabalho que está sendo feito variam. O Azure lida com esse dimensionamento para você e você não precisa adicionar ou remover manualmente as máquinas virtuais, ou realizar qualquer outra forma de configuração.

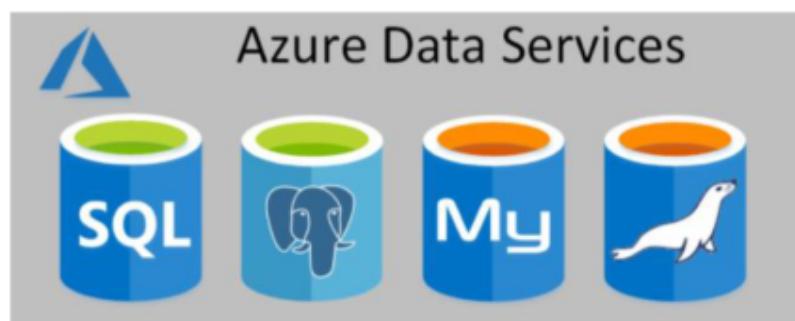
CPF 00296691208

SaaS é curto para Software-as-a-Service. As ofertas de SaaS são tipicamente pacotes de software específicos que são instalados e executados em hardware virtual na nuvem. Os pacotes SaaS são tipicamente aplicativos hospedados em vez de software mais generalizado, como um DBMS. Os pacotes SaaS comuns disponíveis no Azure incluem Microsoft 365 (antigo Office 365).

### O que são Serviços de Dados do Azure?

O Azure Data Services se enquadra na categoria PaaS. Esses serviços são uma série de DBMSs gerenciados pela Microsoft na nuvem. Cada serviço de dados cuida da configuração, do gerenciamento diário, das atualizações de software e da segurança dos bancos de dados que hospeda. Tudo o que você faz é criar seus bancos de dados sob o controle do serviço de dados.

O Azure Data Services está disponível para vários sistemas comuns de gerenciamento de banco de dados relacional. O serviço mais conhecido é o Banco de Dados SQL do Azure. Os outros disponíveis atualmente são Banco de Dados Azure para servidores MySQL, Banco de Dados Azure para servidores MariaDB e Banco de Dados Azure para servidores PostgreSQL. As demais unidades deste módulo descrevem os recursos fornecidos por esses serviços.



**NOTA:** A Microsoft também fornece serviços de dados para sistemas de gerenciamento de banco de dados não relacionais, como Cosmos DB.

Usar o Azure Data Services reduz a quantidade de tempo que você precisa investir para administrar um DBMS. No entanto, esses serviços também podem limitar o intervalo de tarefas de administração personalizadas que você pode executar, porque a execução manual de algumas tarefas pode comprometer a maneira como o serviço é executado. Por exemplo, alguns DBMSs permitem que você instale software personalizado em um banco de dados ou execute scripts como parte de uma operação de banco de dados.

CURSO DE MANHÃES DA SILVA

CPF: 00296691208

Este software pode não ser compatível com o serviço de dados, e permitir que um aplicativo execute um script de um banco de dados pode afetar a segurança do serviço. Você deve estar preparado para trabalhar com essas restrições em mente.

Além de reduzir a carga de trabalho administrativa, o Azure Data Services garante que seus bancos de dados estejam disponíveis por pelo menos 99,99% do tempo.

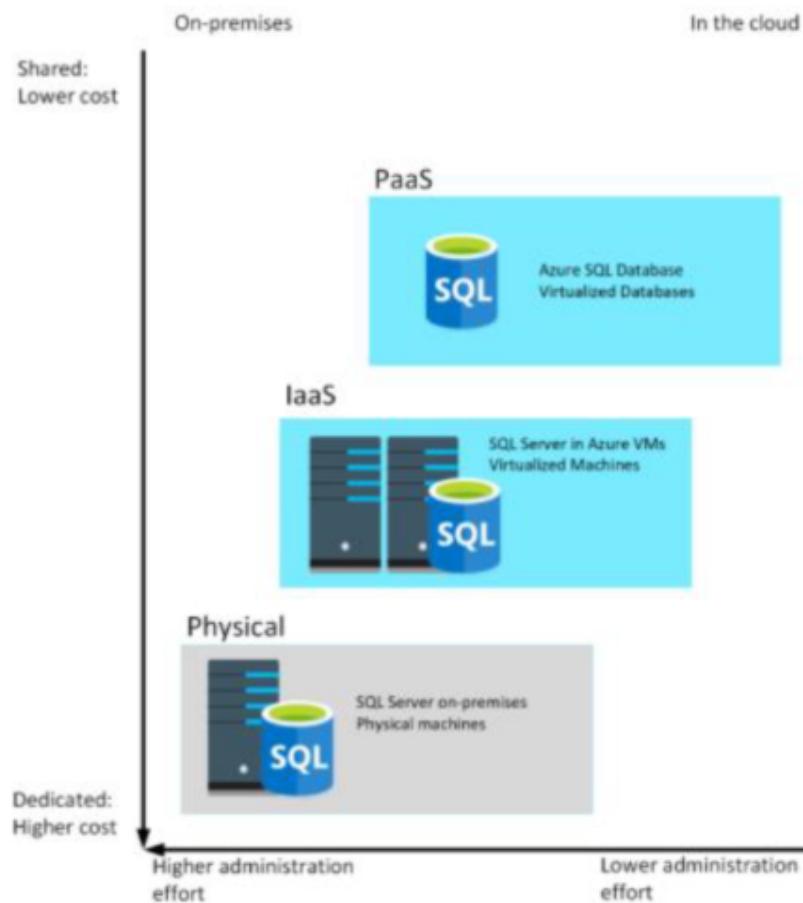
Existem custos associados à execução de um banco de dados no Azure Data Services. O preço base de cada serviço cobre a infraestrutura e o licenciamento subjacentes, junto com os encargos administrativos. Além disso, esses serviços são projetados para estarem sempre ligados. Isso significa que você não pode desligar um banco de dados e reiniciá-lo mais tarde.

Nem todos os recursos de um sistema de gerenciamento de banco de dados estão disponíveis no Azure Data Services. Isso ocorre porque o Azure Data Services assume a tarefa de gerenciar o sistema e mantê-lo em execução usando hardware situado em um datacenter do Azure. Expor algumas funções administrativas pode tornar a plataforma subjacente vulnerável ao uso indevido e até mesmo abrir algumas questões de segurança. Portanto, você não tem controle direto sobre a plataforma em que os serviços são executados.

Se você precisar de mais controle do que o permitido pelos Serviços de Dados do Azure, poderá instalar seu sistema de gerenciamento de banco de dados em uma máquina virtual executada no Azure. A próxima unidade examina essa abordagem com mais detalhes para o SQL Server, embora os mesmos problemas se apliquem a outros sistemas de gerenciamento de banco de dados com suporte pelo Azure Data Services.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

A imagem abaixo destaca as diferentes maneiras pelas quais você pode executar um DBMS, como o SQL Server, começando com um sistema local no canto esquerdo inferior até PaaS no canto superior direito. O diagrama ilustra os benefícios de mudar para a abordagem PaaS.



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## SQL Server em máquinas virtuais do Azure

O Microsoft SQL Server é um DBMS relacional popular. Ele tem uma longa história e possui recursos que fornecem gerenciamento de banco de dados para organizações de todos os tamanhos. No passado, as organizações executavam o SQL Server localmente. No entanto, muitas organizações agora procuram mudar as operações on-line para aproveitar as vantagens dos serviços disponíveis na nuvem.

O SQL Server oferece várias maneiras de executar um banco de dados no Azure. Nesta unidade, você verá como mover o SQL Server para uma Máquina Virtual do Azure.

### O que é SQL Server em Máquinas Virtuais do Azure?

O SQL Server em máquinas virtuais permite que você use versões completas do SQL Server na nuvem, sem ter que gerenciar nenhum hardware local. Este é um exemplo da abordagem IaaS.

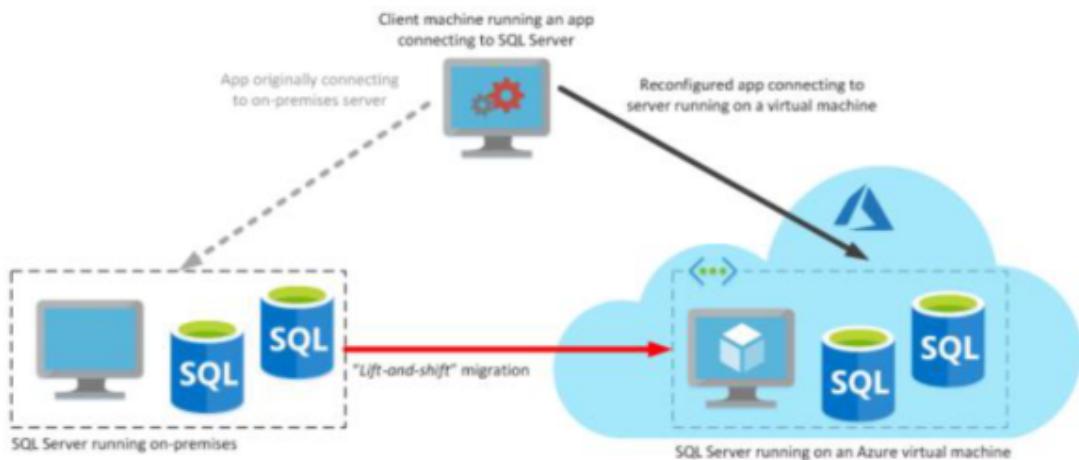
O SQL Server em execução em uma máquina virtual do Azure replica efetivamente o banco de dados em execução em hardware local real. Migrar do sistema em execução local para uma máquina virtual do Azure não é diferente de mover os bancos de dados de um servidor local para outro.

No cenário de exemplo descrito na introdução, o banco de dados executa procedimentos armazenados e scripts como parte da carga de trabalho do banco de dados. Se esses procedimentos armazenados e scripts dependerem de recursos que são restritos ao seguir uma abordagem PaaS, executar o SQL Server em suas próprias máquinas virtuais pode ser uma boa opção. No entanto, você continua responsável por manter o software SQL Server e executar as várias tarefas administrativas para manter o banco de dados em execução no dia a dia.

Essa abordagem é adequada para migrações e aplicativos que requerem acesso aos recursos do sistema operacional que podem não ser suportados no nível de PaaS. Máquinas virtuais SQL são lift-and-shift prontos para aplicativos existentes, que requerem migração rápida para a nuvem com mudanças mínimas.

TITULO DA UNIDADE

CPF: 00296691208

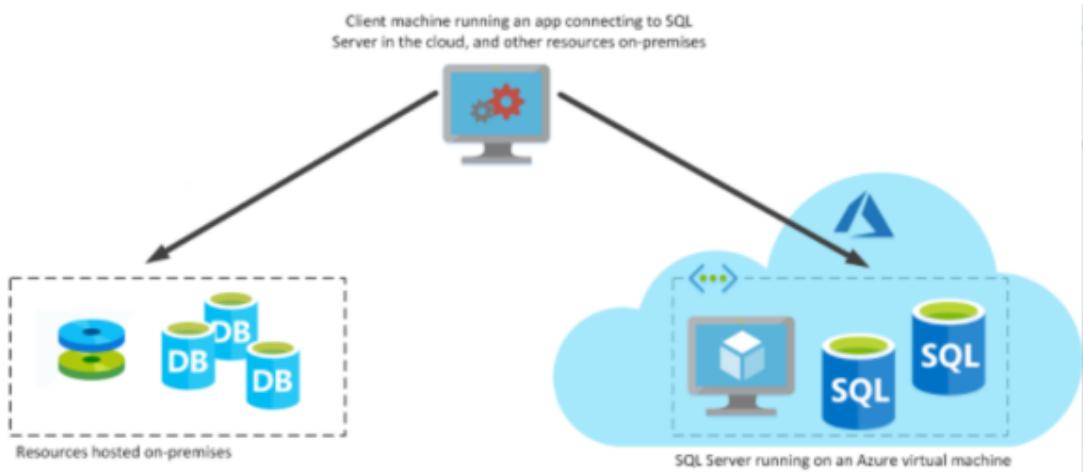


**NOTA:** O termo lift-and-shift refere-se à maneira como você pode mover um banco de dados diretamente de um servidor local para uma máquina virtual do Azure sem exigir que você faça alterações nele. Os aplicativos que se conectavam anteriormente ao banco de dados local podem ser reconfigurados rapidamente para se conectar ao banco de dados em execução na máquina virtual, mas, caso contrário, devem permanecer inalterados.

## Casos de uso

Essa abordagem é otimizada para migrar aplicativos existentes para o Azure, ou estender os aplicativos locais existentes para a nuvem em implantações híbridas.

**NOTA:** A implantação híbrida é um sistema em que parte da operação é executada no local e parte na nuvem. Seu banco de dados pode fazer parte de um sistema maior executado no local, embora os elementos do banco de dados possam estar hospedados na nuvem.



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Você pode usar o SQL Server em uma máquina virtual para desenvolver e testar aplicativos tradicionais do SQL Server. Com uma máquina virtual, você tem todos os

direitos administrativos sobre o DBMS e o sistema operacional. É a escolha perfeita quando uma organização já possui recursos de TI disponíveis para manter as máquinas virtuais.

Esses recursos permitem que você:

- Crie cenários de desenvolvimento e teste rápidos quando não quiser comprar hardware SQL Server de produção local.
- Torne-se pronto para mudanças e mudanças para os aplicativos existentes que requerem migração rápida para a nuvem com mudanças mínimas ou sem mudanças.
- Amplie a plataforma na qual o SQL Server está sendo executado, alocando mais memória, potência da CPU e espaço em disco para a máquina virtual. Você pode redimensionar rapidamente uma máquina virtual do Azure sem a necessidade de reinstalar o software que está sendo executado nela.

### Benefícios de negócios

A execução do SQL Server em máquinas virtuais permite que você atenda às necessidades comerciais exclusivas e diversas por meio de uma combinação de implantações locais e hospedadas na nuvem, usando o mesmo conjunto de produtos de servidor, ferramentas de desenvolvimento e experiência nesses ambientes.

Nem sempre é fácil para as empresas mudarem seu DBMS para um serviço totalmente gerenciado. Pode haver requisitos específicos que devem ser atendidos para migrar para um serviço gerenciado que exija alterações no banco de dados e nos aplicativos que o utilizam. Por esse motivo, o uso de máquinas virtuais pode oferecer uma solução, mas usá-las não elimina a necessidade de administrar seu DBMS com o mesmo cuidado que faria no local.

### Banco de dados SQL do Azure

Se você não deseja incorrer na sobrecarga de gerenciamento associada à execução do SQL Server em uma máquina virtual, pode usar o Banco de Dados SQL do Azure.

#### O que é Banco de Dados SQL do Azure?

O Banco de Dados SQL do Azure é uma oferta de PaaS da Microsoft. Você cria um servidor de banco de dados gerenciado na nuvem e, em seguida, implanta seus bancos de dados neste servidor.

**NOTA:** Um servidor de banco de dados SQL é uma construção lógica que atua como um ponto administrativo central para vários bancos de dados únicos ou em pool, logins, regras de firewall, regras de auditoria, políticas de detecção de ameaças e grupos de failover.

O Banco de Dados SQL do Azure está disponível com várias opções: Banco de dados único, Piscina elástica, e Instância gerenciada. As seções a seguir descrevem

Instância Única e Conjunto Elástico. A instância gerenciada é o assunto da próxima unidade.

### Banco de dados único

Essa opção permite que você configure e execute rapidamente um único banco de dados SQL Server. Você cria e executa um servidor de banco de dados na nuvem e acessa seu banco de dados por meio desse servidor. A Microsoft gerencia o servidor, então tudo que você precisa fazer é configurar o banco de dados, criar suas tabelas e preenchê-las com seus dados. Você pode dimensionar o banco de dados se precisar de espaço de armazenamento adicional, memória ou poder de processamento.

Por padrão, os recursos são pré-alocados e você é cobrado por hora pelos recursos solicitados. Você também pode especificar uma configuração serverless. Nessa configuração, a Microsoft cria seu próprio servidor, que pode ser compartilhado por vários bancos de dados pertencentes a outros assinantes do Azure. A Microsoft garante a privacidade de seu banco de dados. Seu banco de dados é dimensionado automaticamente e os recursos são alocados ou desalocados conforme necessário. Para mais informações, leia “O que é um único banco de dados no Banco de Dados SQL do Azure”.

Azure SQL Database, Single Database deployment



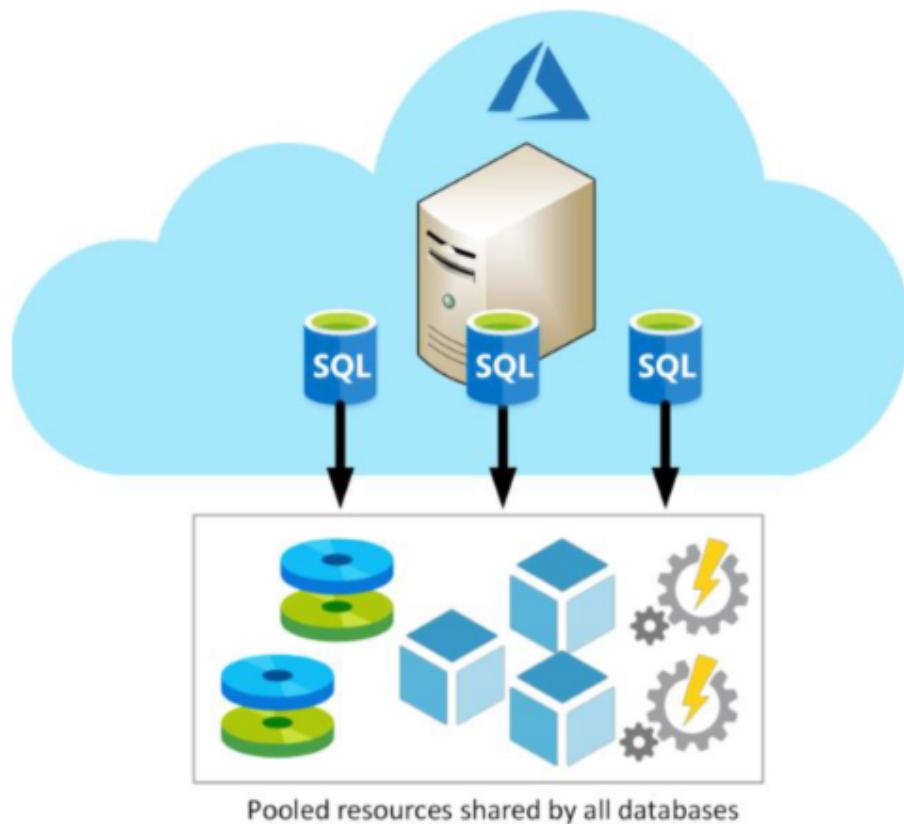
THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Piscina elástica

Esta opção é semelhante a Banco de dados único, exceto que, por padrão, vários bancos de dados podem compartilhar os mesmos recursos, como memória, espaço de armazenamento de dados e capacidade de processamento. Os recursos são chamados de piscina (ou pool, em inglês). Você cria o pool e apenas seus bancos de dados podem usá-lo. Este modelo é útil se você tiver bancos de dados com requisitos de recursos que variam ao longo do tempo e pode ajudá-lo a reduzir custos.

Por exemplo, o banco de dados da folha de pagamento pode exigir bastante potência da CPU no final de cada mês, conforme você lida com o processamento da folha de pagamento, mas em outros momentos o banco de dados pode se tornar muito menos ativo. Você pode ter outro banco de dados usado para executar relatórios. Este banco de dados pode ficar ativo por vários dias no meio do mês, conforme os relatórios de gerenciamento são gerados, mas com uma carga mais leve em outros momentos. O Elastic Pool permite que você use os recursos disponíveis no pool e libere os recursos assim que o processamento for concluído.

Azure SQL Database, Elastic Pool deployment



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Casos de uso

O Banco de Dados SQL do Azure oferece a melhor opção de baixo custo com administração mínima. Não é totalmente compatível com instalações locais do SQL

Server. Muitas vezes, é usado em novos projetos de nuvem em que o design do aplicativo pode acomodar quaisquer alterações necessárias em seus aplicativos.

Observação: você pode usar o Assistente de Migração de Dados para detectar problemas de compatibilidade com seus bancos de dados que podem afetar a funcionalidade do banco de dados no Banco de Dados SQL do Azure. Para mais informações, veja "Visão geral do assistente de migração de dados".

O Banco de Dados SQL do Azure costuma ser usado para:

- Aplicativos em nuvem modernos que precisam usar os recursos estáveis mais recentes do SQL Server.
- Aplicativos que requerem alta disponibilidade.
- Sistemas com carga variável, que precisam do servidor de banco de dados para aumentar ou diminuir rapidamente.

## Benefícios de negócios

O Banco de Dados SQL do Azure atualiza e corrige automaticamente o software SQL Server para garantir que você esteja sempre executando a versão mais recente e segura do serviço.

Os recursos de escalabilidade do Banco de Dados SQL do Azure garantem que você possa aumentar os recursos disponíveis para armazenar e processar dados sem ter que realizar uma atualização manual cara.

O serviço oferece garantias de alta disponibilidade, para assegurar que seus bancos de dados estejam disponíveis pelo menos 99,99% do tempo. O Banco de Dados SQL do Azure oferece suporte à restauração pontual, permitindo que você recupere um banco de dados para o estado em que estava em qualquer ponto no passado. Os bancos de dados podem ser replicados para diferentes regiões para fornecer garantia adicional e recuperação de desastres.

A proteção avançada contra ameaças oferece recursos avançados de segurança, como avaliações de vulnerabilidade, para ajudar a detectar e corrigir possíveis problemas de segurança com seus bancos de dados. A proteção contra ameaças também detecta atividades anômalas que indicam tentativas incomuns e potencialmente prejudiciais de acessar ou explorar seu banco de dados. Ele monitora continuamente seu banco de dados em busca de atividades suspeitas e fornece alertas de segurança imediatos sobre vulnerabilidades em potencial, ataques de injeção de SQL e padrões irregulares de acesso ao banco de dados. Os alertas de detecção de ameaças fornecem detalhes da atividade suspeita e recomendam ações sobre como investigar e atenuar a ameaça.

A auditoria rastreia eventos de banco de dados e os grava em um log de auditoria em sua conta de armazenamento do Azure. A auditoria pode ajudá-lo a manter a conformidade regulamentar, compreender a atividade do banco de dados e obter uma

visão das discrepâncias e anomalias que podem indicar preocupações comerciais ou suspeitas de violações de segurança.

O banco de dados SQL ajuda a proteger seus dados fornecendo criptografia. Para dados em movimento, ele usa a segurança da camada de transporte. Para dados em repouso, ele usa criptografia de dados transparente. Para dados em uso, ele sempre usa criptografados.

No cenário Wide World Importers, servidores vinculados são usados para realizar consultas distribuídas. No entanto, nem o Banco de Dados Único nem o Elastic Pool oferecem suporte a servidores vinculados. Se você deseja usar o Banco de Dados Único ou Elastic Pool, pode precisar modificar as consultas que usam servidores vinculados e retrabalhar as operações que dependem desses recursos.

### **Instância gerenciada de banco de dados SQL do Azure**

Uma empresa pode querer eliminar o máximo possível de sobrecarga de gerenciamento na administração de bancos de dados e servidores, mas as limitações das opções de Banco de Dados Único e Elastic Pool podem significar que essas opções não são adequadas. Nessas situações, A instância gerenciada do Banco de Dados SQL do Azure pode ser uma boa opção a ser considerada.

### **O que é instância gerenciada do Banco de Dados SQL do Azure?**

As opções Single Database e Elastic Pool restringem alguns dos recursos administrativos disponíveis para o SQL Server. A instância gerenciada executa com eficácia uma instância totalmente controlável do SQL Server na nuvem. Você pode instalar vários bancos de dados na mesma instância. Você tem controle total sobre esta instância, da mesma forma que teria para um servidor local. O serviço de instância gerenciada automatiza backups, patches de software, monitoramento de banco de dados e outras tarefas gerais, mas você tem controle total sobre a segurança e a alocação de recursos para seus bancos de dados. Você pode encontrar informações detalhadas em "O que é instância gerenciada do Banco de Dados SQL do Azure?".

As instâncias gerenciadas dependem de outros serviços do Azure, como Armazenamento do Azure para backups, Hubs de Eventos do Azure para telemetria, Azure Active Directory para autenticação, Azure Key Vault para Transparent Data Encryption (TDE) e alguns serviços da plataforma do Azure que fornecem recursos de segurança e suporte. As instâncias gerenciadas fazem conexões com esses serviços.

Todas as comunicações são criptografadas e assinadas usando certificados. Para verificar a confiabilidade das partes comunicantes, as instâncias gerenciadas verificam constantemente esses certificados por meio de listas de revogação de certificados. Se os certificados forem revogados, a instância gerenciada fecha as conexões para proteger os dados.

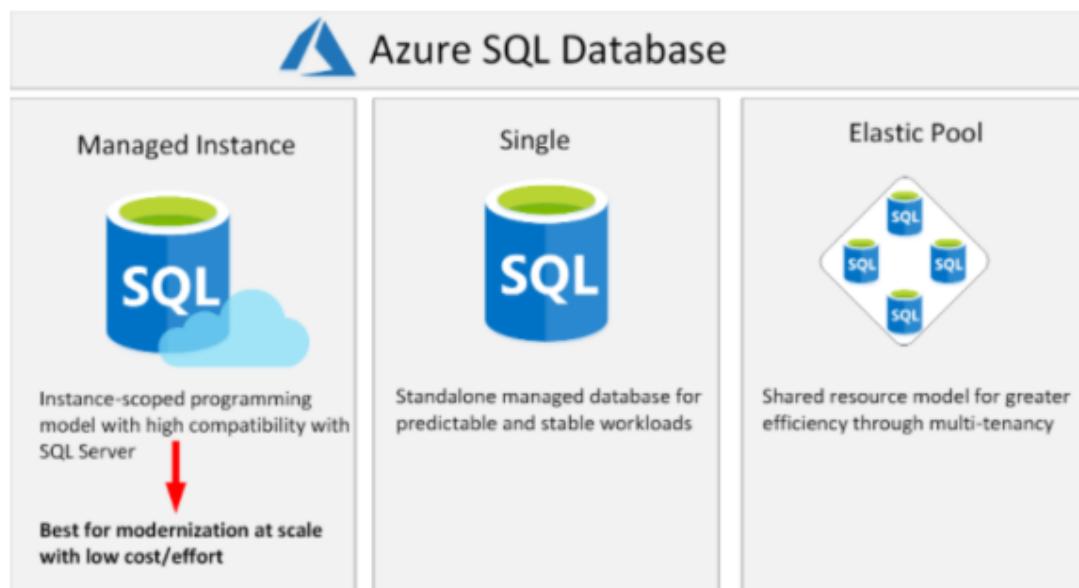
CURSO DE AZURE DATA FUNDAMENTOS

CPF: 00296691208

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

A imagem a seguir resume as diferenças entre a instância gerenciada do Banco de Dados SQL, Banco de Dados Único e Elastic Pool



## Casos de uso

Considere a instância gerenciada do Banco de Dados SQL do Azure se desejar elevar e mudar uma instância local do SQL Server e todos os seus bancos de dados para a nuvem, sem incorrer na sobrecarga de gerenciamento de executar o SQL Server em uma máquina virtual.

A instância gerenciada do Banco de Dados SQL fornece recursos não disponíveis com as opções Banco de Dados Único ou Elastic Pool. Se o seu sistema usa recursos como servidores vinculados, Service Broker (um sistema de processamento de mensagens que pode ser usado para distribuir trabalho entre servidores) ou Database Mail (que permite que seu banco de dados envie mensagens de e-mail aos usuários), então você deve usar a instância gerenciada.

Para verificar a compatibilidade com um sistema local existente, você pode instalar o Assistente de migração de dados (DMA). Essa ferramenta analisa seus bancos de dados no SQL Server e relata quaisquer problemas que possam bloquear a migração para uma instância gerenciada.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Benefícios de negócios

A instância gerenciada do Banco de Dados SQL fornece todos os benefícios de gerenciamento e segurança disponíveis ao usar o Banco de Dados Único e Elastic Pool. A implantação de instância gerenciada permite que um administrador de sistema gaste menos tempo em tarefas administrativas porque o serviço de banco de dados SQL as executa para você ou simplifica muito essas tarefas. As tarefas automatizadas incluem instalação e correção de software de sistema de gerenciamento de banco de dados e sistema operacional, redimensionamento e configuração de instância dinâmica, backups, replicação de banco de dados (incluindo bancos de dados do sistema), configuração de alta disponibilidade e configuração de fluxos de dados de monitoramento de integridade e desempenho.

A instância gerenciada tem quase 100% de compatibilidade com o SQL Server Enterprise Edition, rodando localmente.

A opção de implantação de instância gerenciada do Banco de Dados SQL oferece suporte a logons e logons tradicionais do mecanismo de Banco de Dados do SQL Server, integrados ao Azure Active Directory (AD). Os logins tradicionais do mecanismo de banco de dados do SQL Server incluem um nome de usuário e uma senha. Você deve inserir suas credenciais sempre que se conectar ao servidor. Os logons do Azure AD usam as credenciais associadas à entrada do seu computador atual e você não precisa fornecê-las sempre que se conectar.

No cenário Wide World Importers, a instância gerenciada do SQL Database pode ser uma escolha mais adequada do que Single Database ou Elastic Pool. A instância gerenciada do Banco de Dados SQL oferece suporte a servidores vinculados, embora alguns dos outros recursos avançados exigidos pelo banco de dados possam não estar disponíveis. Se você deseja uma correspondência completa, executar o SQL Server em uma máquina virtual pode ser sua única opção, mas você precisa equilibrar os benefícios da funcionalidade completa com a sobrecarga administrativa e de manutenção necessária.

## PostgreSQL, MariaDB e MySQL

Assim como o Banco de Dados SQL do Azure, os Serviços de Dados do Azure estão disponíveis para outras soluções populares de banco de dados baseadas em SQL. Atualmente, os serviços de dados estão disponíveis para PostgreSQL, MySQL e MariaDB. O principal motivo para esses serviços é permitir que as organizações que executam PostgreSQL, MySQL ou MariaDB mudem para o Azure rapidamente, sem fazer alterações em seus aplicativos.

CNPJ 1002966912008

### O que são MySQL, MariaDB e PostgreSQL

PostgreSQL, MariaDB e MySQL são sistemas de gerenciamento de banco de dados relacional, adaptados para diferentes especializações.

O MySQL começou como um sistema de gerenciamento de banco de dados de código aberto simples de usar. Está disponível em várias edições: Comunidade (Community), Padrão (Standard) e Empresa (Enterprise). A edição Community está disponível gratuitamente e tem sido historicamente popular como um sistema de gerenciamento de banco de dados para aplicativos da web, rodando em Linux. Versões também estão disponíveis para Windows. A edição Standard oferece melhor desempenho e usa uma tecnologia diferente para armazenar dados. A edição Enterprise fornece um conjunto abrangente de ferramentas e recursos, incluindo segurança, disponibilidade e escalabilidade aprimoradas. As edições Standard e Enterprise são as versões mais frequentemente usadas por organizações comerciais, embora essas versões do software não sejam gratuitas.

MariaDB é um sistema de gerenciamento de banco de dados mais recente, criado pelos desenvolvedores originais do MySQL. O mecanismo de banco de dados foi reescrito e otimizado para melhorar o desempenho. O MariaDB oferece compatibilidade com o banco de dados Oracle (outro sistema popular de gerenciamento de banco de dados comercial). Um recurso notável do MariaDB é seu suporte integrado para dados temporais. Uma tabela pode conter várias versões de dados, permitindo que um aplicativo consulte os dados conforme eles apareciam em algum ponto no passado.

PostgreSQL é um banco de dados de objetos relacionais híbrido. Você pode armazenar dados em tabelas relacionais, mas um banco de dados PostgreSQL também permite armazenar tipos de dados personalizados, com suas próprias propriedades não relacionais. O sistema de gerenciamento de banco de dados é extensível; você pode adicionar módulos de código ao banco de dados, que podem ser executados por consultas. Outro recurso importante é a capacidade de armazenar e manipular dados geométricos, como linhas, círculos e polígonos.

PostgreSQL tem sua própria linguagem de consulta chamada pgsql. Essa linguagem é uma variante da linguagem de consulta relacional padrão, SQL, com recursos que permitem escrever procedimentos armazenados que são executados dentro do banco de dados.

CURSO DE AZURE DATA FLOW

## O que é Banco de Dados do Azure para MySQL?

O Banco de Dados do Azure para MySQL é uma implementação de PaaS do MySQL na nuvem Azure, com base no MySQL Community Edition.

O serviço de Banco de Dados do Azure para MySQL inclui alta disponibilidade sem custo adicional e escalabilidade conforme necessário. Você só paga pelo que usa. Backups automáticos são fornecidos, com restauração pontual.

O servidor fornece segurança de conexão para impor regras de firewall e, opcionalmente, requer conexões SSL. Muitos parâmetros do servidor permitem definir as configurações do servidor, como modos de bloqueio, número máximo de conexões e tempos limite.

O Banco de Dados do Azure para MySQL fornece um sistema de banco de dados global que pode ser dimensionado para grandes bancos de dados sem a necessidade de gerenciar hardware, componentes de rede, servidores virtuais, patches de software e outros componentes subjacentes.

Determinadas operações não estão disponíveis com o Banco de Dados do Azure para MySQL. Essas funções estão principalmente relacionadas à segurança e administração. O Azure gerencia esses aspectos do próprio servidor de banco de dados.

## Benefícios do Banco de Dados Azure para MySQL

Você obtém os seguintes recursos com o Banco de Dados do Azure para MySQL:

- Recursos de alta disponibilidade integrados.
- Desempenho previsível.
- Escalonamento fácil que responde rapidamente à demanda.
- Dados seguros, tanto em repouso quanto em movimento.
- Backups automáticos e restauração pontual nos últimos 35 dias.
- Segurança de nível empresarial e conformidade com a legislação.
- O sistema usa preços pré-pagos para que você pague apenas pelo que usar.
- O Banco de Dados do Azure para servidores MySQL fornece funcionalidade de monitoramento para adicionar alertas e exibir métricas e logs.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

## O que é Banco de Dados do Azure para MariaDB?

O Banco de Dados Azure para MariaDB é uma implementação do sistema de gerenciamento de banco de dados MariaDB adaptado para ser executado no Azure. É baseado no MariaDB Community Edition.

O banco de dados é totalmente gerenciado e controlado pelo Azure. Depois de provisionar o serviço e transferir seus dados, o sistema quase não requer administração adicional.

## Benefícios do Banco de Dados Azure para MariaDB

O Banco de Dados Azure para MariaDB oferece:

- Alta disponibilidade integrada sem custo adicional.
- Desempenho previsível, usando preços pré-pagos inclusivos.
- Escalando conforme necessário em segundos.
- Proteção segura de dados confidenciais em repouso e em movimento.
- Backups automáticos e restauração pontual por até 35 dias.
- Segurança e conformidade de nível empresarial.

## O que é Banco de Dados do Azure para PostgreSQL?

Se preferir PostgreSQL, você pode escolher Banco de dados do Azure para PostgreSQL para executar uma implementação PaaS de PostgreSQL na nuvem do Azure. Este serviço oferece a mesma disponibilidade, desempenho, escala, segurança e benefícios administrativos que o serviço MySQL.

Alguns recursos de bancos de dados PostgreSQL locais não estão disponíveis no Banco de Dados do Azure para PostgreSQL. Esses recursos estão relacionados principalmente às extensões que os usuários podem adicionar a um banco de dados para realizar tarefas especializadas, como escrever procedimentos armazenados em várias linguagens de programação (além do pgsql, que está disponível) e interagir diretamente com o sistema operacional. Um conjunto básico das extensões usadas com mais frequência é compatível, e a lista de extensões disponíveis está em revisão contínua.

O Banco de Dados do Azure para PostgreSQL tem duas opções de implantação: Servidor único e Hyperscale.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Banco de dados Azure para servidor único PostgreSQL

A opção de implantação de servidor único para PostgreSQL oferece benefícios semelhantes aos do Banco de Dados do Azure para MySQL. Você escolhe entre três níveis de preços: Básico, Uso Geral e Otimizado para Memória. Cada camada oferece suporte a diferentes números de CPUs, memória e tamanhos de armazenamento - você seleciona um com base na carga que espera suportar.

## Banco de dados Azure para PostgreSQL Hyperscale (Citus)

Hyperscale (Citus) é uma opção de implantação que dimensiona as consultas em vários nós de servidor para suportar grandes cargas de banco de dados. Seu banco de dados é dividido em nós. Os dados são divididos em blocos com base no valor de uma chave de partição ou chave de fragmentação. Considere usar esta opção de implantação para as maiores implantações de banco de dados PostgreSQL na nuvem do Azure.

## Benefícios do Banco de Dados Azure para PostgreSQL

O Banco de Dados do Azure para PostgreSQL é um serviço altamente disponível. Ele contém detecção de falha embutida e mecanismos de failover.

Os usuários do PostgreSQL estarão familiarizados com o pgAdmin ferramenta, que você pode usar para gerenciar e monitorar um banco de dados PostgreSQL. Você pode continuar a usar essa ferramenta para se conectar ao Banco de Dados do Azure para PostgreSQL. No entanto, algumas funcionalidades focadas no servidor, como a execução de backup e restauração do servidor, não estão disponíveis porque o servidor é gerenciado e mantido pela Microsoft.

O Banco de Dados do Azure para servidores PostgreSQL registra informações sobre as consultas executadas em bancos de dados no servidor e as salva em um banco de dados denominado `azure_sys`. Você consulta `oquery_store.qs_viewview` para ver essas informações e use-as para monitorar as consultas que os usuários estão executando. Essas informações podem ser inestimáveis se você precisar ajustar as consultas realizadas por seus aplicativos.

TITULO DA AULA

CPF: 00296691208

### Migrar dados para o Azure

Se você tiver bancos de dados MySQL, MariaDB ou PostgreSQL existentes em execução no local e desejar mover os dados para um banco de dados executando os serviços de dados correspondentes no Azure, poderá usar o Serviço de migração de banco de dados do Azure (DMS).

O Serviço de Migração de Banco de Dados permite que você restaure um backup de seus bancos de dados locais diretamente para bancos de dados em execução no Azure Data Services. Você também pode configurar a replicação de um banco de dados local, para que todas as alterações feitas nos dados desse banco de dados sejam copiadas para o banco de dados em execução no Azure Data Services. Essa estratégia permite reconfigurar usuários e aplicativos para se conectar ao banco de dados na nuvem enquanto o sistema local ainda está ativo; você não precisa desligar o sistema local enquanto transfere usuários para a nuvem.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Teste de conhecimento

### Questão 1

**Qual implantação requer menos alterações ao migrar uma solução existente do SQL Server local?**

- a) Instância gerenciada de banco de dados SQL do Azure
- b) **SQL Server em execução em uma máquina virtual**
- c) Banco de dados SQL do Azure Banco de dados único

Explicação: está correto. O SQL Server em execução em uma máquina virtual oferece suporte a tudo o que uma solução local possui.

### Questão 2

**Qual das afirmações a seguir é verdadeira sobre o SQL Server em execução em uma máquina virtual?**

- a) Você deve instalar e manter o software para o sistema de gerenciamento de banco de dados sozinho, mas os backups são automatizados
- b) A instalação e manutenção do software são automatizadas, mas você deve fazer seus próprios backups
- c) **Você é responsável por toda a instalação e manutenção do software e pela execução de backups**

Explicação: está correto. Com o SQL Server em execução em uma máquina virtual, você é responsável por aplicar patches e fazer backups.

### Questão 3

**Qual das afirmações a seguir é verdadeira sobre o Banco de Dados SQL do Azure?**

- a) A ampliação não tem efeito até que você reinicie o banco de dados
- b) A ampliação não tem efeito até que você reinicie o banco de dados
- c) **O aumento ou redução terá efeito sem reiniciar o banco de dados SQL**

Explicação: está correto, você pode aumentar ou diminuir a escala sem interromper o uso do banco de dados.

## Questão 4

Ao usar uma instância gerenciada do Banco de Dados SQL do Azure, qual é a maneira mais simples de implementar backups?

- a) Configuração manual do servidor SQL
- b) Crie uma tarefa agendada para fazer backup
- c) Os backups são gerenciados automaticamente

Explicação: está correto. Uma instância gerenciada vem com o benefício de backups automáticos e a capacidade de restaurar em um determinado momento.

## Questão 5

Qual é a melhor maneira de transferir os dados de um banco de dados PostgreSQL em execução local para um banco de dados que executa o serviço de Banco de Dados do Azure para PostgreSQL?

- a) Exporte os dados do banco de dados local e importe-os manualmente para o banco de dados em execução no Azure
- b) Carregue um arquivo de backup de banco de dados PostgreSQL para o banco de dados em execução no Azure
- c) Use os Serviços de Migração de Banco de Dados do Azure

Explicação: está correto. O Serviço de Migração de Banco de Dados oferece a maneira mais segura de enviar seu banco de dados PostgreSQL local para o Azure.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Resumo

Nesta lição, você aprendeu sobre as opções de implantação de PaaS e IaaS para executar bancos de dados na nuvem. Você viu como o Azure Data Services fornece uma variedade de serviços PaaS para executar bancos de dados relacionais no Azure. Você aprendeu como as opções de PaaS fornecem suporte para gerenciamento e administração automatizados, em comparação com uma abordagem de IaaS.

## Recursos adicionais

- Escolha a opção de implantação certa
- O que é um único banco de dados no Banco de Dados SQL do Azure
- O que é instância gerenciada do Banco de Dados SQL do Azure?
- Assistente de migração de dados (DMA)
- Serviço de migração de banco de dados do Azure (DMS)
- Escolha o armazenamento de dados certo

## Introdução

O Azure oferece suporte a vários serviços de banco de dados, permitindo que você execute sistemas populares de gerenciamento de banco de dados, como SQL Server, PostgreSQL e MySQL, na nuvem.

Os serviços de banco de dados do Azure são totalmente gerenciados, liberando um tempo valioso que você gastaria gerenciando seu banco de dados. O desempenho de nível empresarial com alta disponibilidade integrada significa que você pode escalar rapidamente e alcançar a distribuição global sem se preocupar com paralisações dispendiosas. Os desenvolvedores podem tirar proveito das inovações líderes do setor, como segurança integrada com monitoramento automático e detecção de ameaças, ajuste automático para melhor desempenho. Além de todos esses recursos, você tem disponibilidade garantida.

Suponha que você seja um engenheiro de dados da Contoso e seja responsável por criar e gerenciar bancos de dados. Você foi solicitado a configurar três novos armazenamentos de dados relacionais: banco de dados SQL do Azure, PostgreSQL e MySQL.

Nesta lição, você explorará as opções disponíveis para criar e configurar os serviços de dados relacionais do Azure.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

## Objetivos de aprendizado

Nesta lição, você irá:

- Provisionar serviços de dados relacionais
- Configurar serviços de dados relacionais
- Explorar problemas básicos de conectividade
- Explorar a segurança de dados
- Descrever o provisionamento de serviços de dados relacionais

No cenário de exemplo, a Contoso decidiu que a organização exigirá vários armazenamentos relacionais diferentes. Como engenheiro de dados, você foi solicitado a configurar armazenamentos de dados usando o Banco de Dados SQL do Azure, PostgreSQL e MySQL.

Neste módulo, você aprenderá como provisionar esses serviços.

### O que é provisionamento?

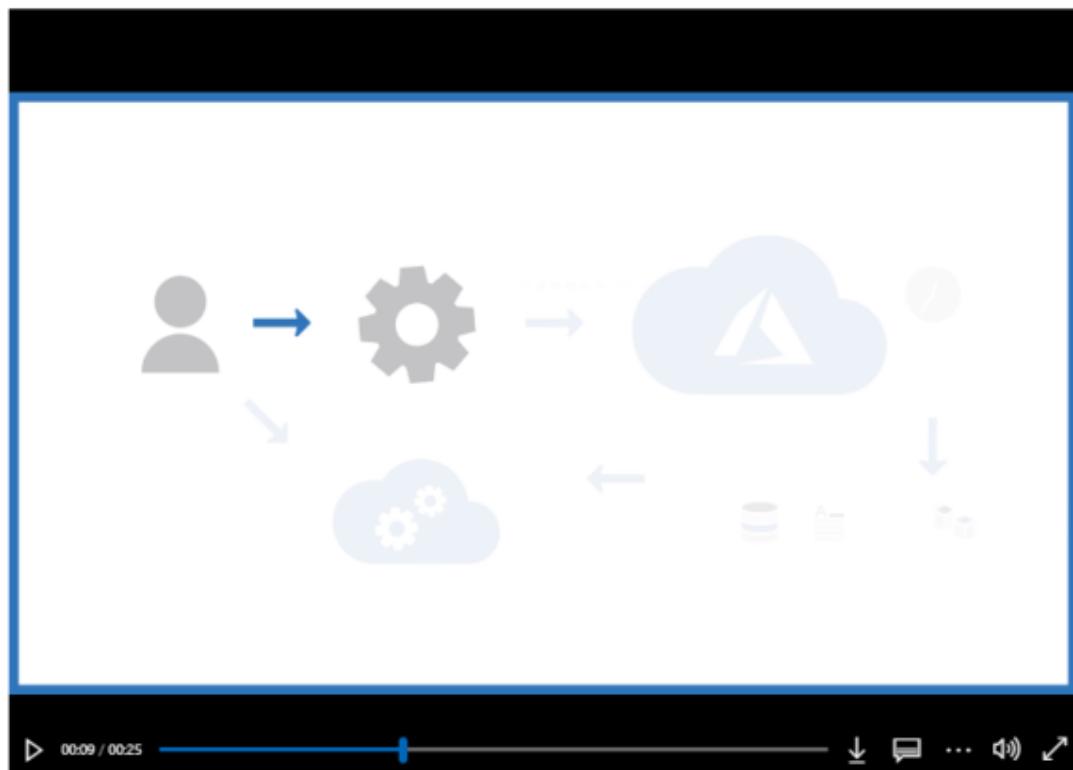
Provisionamento é o ato de executar uma série de tarefas que um provedor de serviços, como o Banco de Dados SQL do Azure, executa para criar e configurar um serviço. Nos bastidores, o provedor de serviços configurará os vários recursos (discos, memória, CPUs, redes e assim por diante) necessários para executar o serviço. Esses recursos serão atribuídos a você e eles permanecerão alocados a você (e cobrados de você), até que você exclua o serviço.

O modo como o provedor de serviços provisiona os recursos é obscuro e você não precisa se preocupar com o funcionamento desse processo. Tudo o que você faz é especificar parâmetros que determinam o tamanho dos recursos necessários (quanto espaço em disco, memória, capacidade de computação e largura de banda da rede). Esses parâmetros são determinados estimando o tamanho da carga de trabalho que você pretende executar usando o serviço. Em muitos casos, você pode modificar esses parâmetros após a criação do serviço, talvez aumentando a quantidade de espaço de armazenamento ou memória se a carga de trabalho for maior do que o previsto inicialmente. O ato de aumentar (ou diminuir) os recursos usados por um serviço é denominado escala.

TITULO DO MÓDULO  
DATA DA CRIAÇÃO  
CPF 00296691208

O vídeo a seguir resume o processo que o Azure executa quando você provisiona um serviço.

[vídeo]



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

O Azure fornece várias ferramentas que você pode usar para provisionar serviços:

- O portal do Azure. Essa é a maneira mais conveniente de fornecer um serviço para a maioria dos usuários. O portal do Azure exibe uma série de páginas específicas do serviço que solicitam as configurações necessárias, validando essas configurações antes de realmente provisionar o serviço.
- A interface de linha de comando (CLI) do Azure. A CLI fornece um conjunto de comandos que você pode executar no prompt de comando do sistema operacional ou no Cloud Shell no portal do Azure. Você pode usar esses comandos para criar e gerenciar recursos do Azure. A CLI é adequada se você precisar automatizar a criação de serviço; você pode armazenar comandos CLI em scripts e pode executar esses scripts programaticamente. A CLI pode ser executada em computadores Windows, macOS e Linux. Para obter informações detalhadas sobre a CLI do Azure, leia “O que é CLI do Azure”.

- Azure PowerShell. Muitos administradores estão familiarizados com o uso de comandos do PowerShell para criar scripts e automatizar tarefas administrativas. O Azure fornece uma série de commandlets (comandos específicos do Azure) que você pode usar no PowerShell para criar e gerenciar recursos do Azure. Você pode encontrar mais informações sobre o Azure PowerShell online, em "Documentação do Azure PowerShell". Como a CLI, o PowerShell está disponível para Windows, macOS e Linux.
- Modelos do Azure Resource Manager. Um modelo do Azure Resource Manager descreve o serviço (ou serviços) que você deseja implantar em um arquivo de texto, em um formato conhecido como JSON (JavaScript Object Notation). O exemplo abaixo mostra um modelo que você pode usar para provisionar uma instância do Banco de Dados SQL do Azure.

```
"resources": [
{
  "name": "sql-server-dev",
  "type": "Microsoft.Sql/servers",
  "apiVersion": "2014-04-01-preview",
  "location": "[parameters('location')]",
  "tags": {
    "displayName": "SqlServer"
  },
  "properties": {}
}
]
```

Você envia o modelo para o Azure usando o comando az group create na CLI do Azure, ou New-AzResourceGroupDeployment no Azure PowerShell. Para obter mais informações sobre como criar e usar modelos do Azure Resource Manager para provisionar recursos do Azure, consulte "O que são modelos do Azure Resource Manager?".

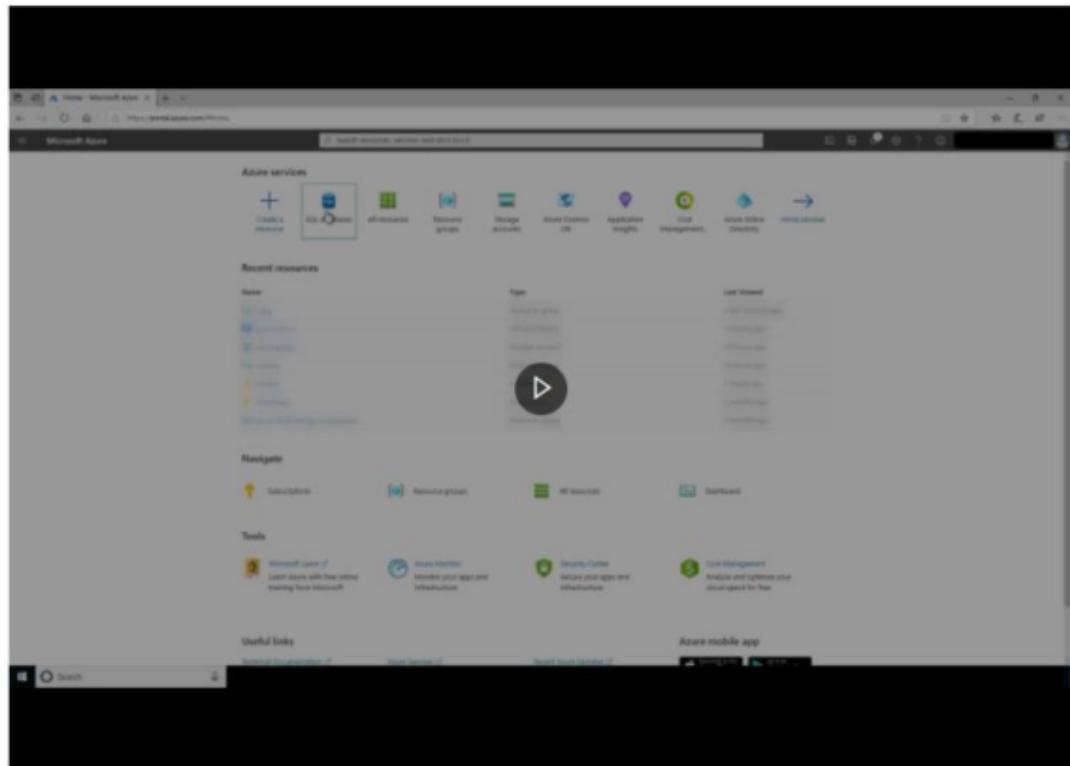
THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

## Demonstração: Provisionando Banco de Dados SQL do Azure

Uma das implantações mais populares nos serviços de dados relacionais do Azure é o Banco de Dados SQL do Azure. Este vídeo demonstra como provisionar uma instância do Banco de Dados SQL do Azure para criar um banco de dados e um servidor.

[vídeo]



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

## Descreva o provisionamento PostgreSQL e MySQL

Os serviços de dados relacionais do Azure permitem que você trabalhe com outros provedores de banco de dados relacionais líderes, como PostgreSQL e MySQL. Esses serviços são chamados de Banco de Dados Azure para PostgreSQL e Banco de Dados Azure para MySQL.

Nesta unidade, você verá como provisionar esses armazenamentos de dados no Azure.

- Como provisionar Banco de Dados do Azure para PostgreSQL e Banco de Dados do Azure para MySQL
- Como acontece com o Banco de Dados SQL do Azure, você pode provisionar um banco de dados PostgreSQL ou MySQL interativamente usando o portal do Azure.

Você pode encontrar esses dois serviços no Azure Marketplace:

The screenshot shows the Azure Marketplace search results for "Azure Database". The search bar at the top has "Azure Database" typed into it. Below the search bar, there are filters for "Pricing: All", "Operating System: All", and "Publisher: All". On the left, there is a sidebar with categories like "Get Started", "AI + Machine Learning", "Analytics", "Blockchain", "Compute", "Containers", "Databases", "Developer Tools", "DevOps", "Identity", "Integration", "Internet of Things", and "IT & Management Tools". The main area displays several service cards. Two services are highlighted with red boxes: "Azure Database for MySQL" (Microsoft) and "Azure Database for PostgreSQL" (Microsoft). Both are described as managed database services for SQL developers.

Os processos de provisionamento do Banco de Dados do Azure para PostgreSQL e do Banco de Dados do Azure para MySQL são muito semelhantes.

**NOTA:** PostgreSQL também oferece a opção de hiperescala, que oferece suporte a cargas de trabalho de desempenho ultra-alto.

The screenshot shows the "Select Azure Database for PostgreSQL deployment option" page. At the top, there is a breadcrumb navigation: Home > Marketplace > Azure Database for PostgreSQL > Select Azure Database for PostgreSQL deployment option. Below the breadcrumb, the title "Select Azure Database for PostgreSQL deployment option" is displayed, followed by the Microsoft logo. The page is divided into two main sections: "How do you plan to use the service?". The first section, "Single server", features a blue icon of a tooth-like shape, a title "Single server", a description "Best for broad range of traditional transactional workloads.", and a note "Enterprise ready, fully managed community PostgreSQL server with up to 64 vCores, optional geospatial support, full-text search and more.". It includes a "Create" button and a "Learn more" link. The second section, "Hyperscale (Citus) server group", features a blue icon with a green circle, a title "Hyperscale (Citus) server group", a description "Best for ultra-high performance and data needs beyond 100GB.", and a note "Ideal for multi-tenant applications and real-time analytical workloads that need sub-second response. Supports both transactional/operational workloads and hybrid transactional analytics workloads.". It also includes a "Create" button and a "Learn more" link.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

A opção de implantação em hiperescala oferece suporte a:

- Escala horizontal em várias máquinas. Esta opção permite que o serviço adicione e remova computadores à medida que as cargas de trabalho aumentam ou diminuem.
- Paralelização de consulta entre esses servidores. O serviço pode dividir as consultas com uso intensivo de recursos em partes que podem ser executadas em paralelo nos diferentes servidores. Os resultados de cada servidor são agregados novamente para produzir um resultado final. Esse mecanismo pode fornecer respostas mais rápidas em consultas em grandes conjuntos de dados.
- Excelente suporte para aplicativos multi-tenant, análise operacional em tempo real e cargas de trabalho transacionais de alto rendimento.

As informações abaixo resumem os campos e configurações necessários ao provisionar um serviço de banco de dados PostgreSQL ou MySQL:

The screenshot shows the 'Create MySQL server' wizard in the Azure portal. The 'Basics' step is selected. Key configuration details include:

- Project details:** Subscription is set to 'mysubscription' and Resource group is 'myresourcegroup'.
- Server details:** Server name is 'mydemoserver'. Data source is 'None' (Backup). Location is '(US) Central US'. Version is '5.7'. Compute + storage is set to 'General Purpose' (4 vCores, 100 GB storage).
- Administrator account:** Admin username is 'myadmin' and Password is 'XXXXXXXXXX' (twice).
- Buttons at the bottom:** 'Review + create' and 'Next : Tags >'.

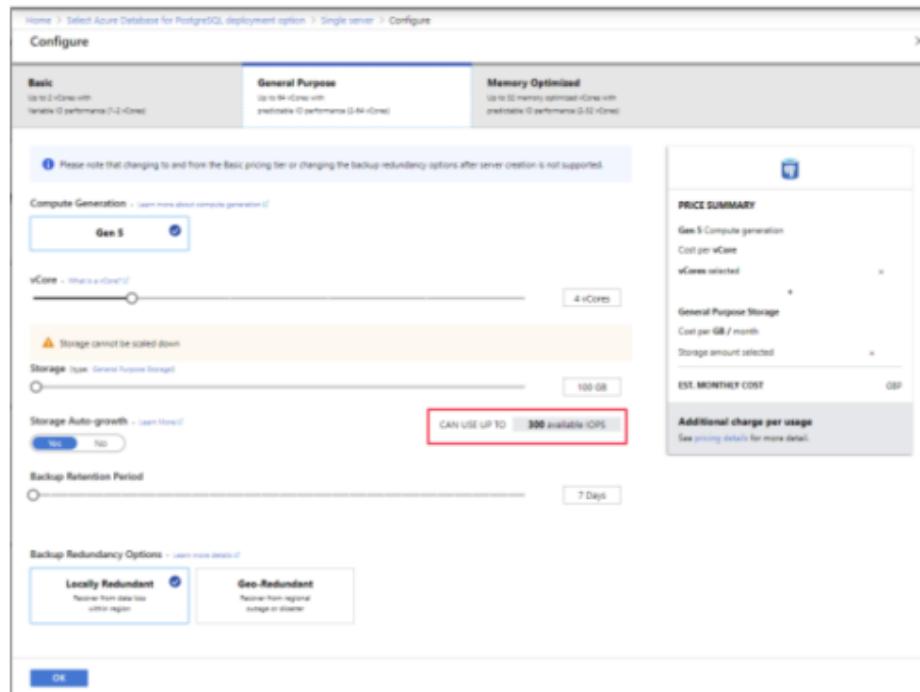
THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

A guia Básicos (Basics), solicita os seguintes detalhes:

- Inscrição. Selecione sua assinatura do Azure.
- Grupo de Recursos. Escolha um grupo de recursos existente ou selecione **Crie um novo** para construir um novo.
- Nome do servidor. Cada banco de dados MySQL ou PostgreSQL deve ter um nome exclusivo que ainda não foi usado por outra pessoa. O nome deve ter entre 3 e 31 caracteres e pode conter apenas letras minúsculas, dígitos e o caractere “\_”.
- Fonte de dados. Selecione **Nenhum** para criar um novo servidor do zero. Você pode selecionar **Cópia de segurança** se você estiver criando um servidor a partir de um backup geográfico de um Banco de Dados do Azure existente para servidor MySQL.
- Localização. Selecione a região mais próxima de você ou a região mais próxima de seus usuários.
- Versão. A versão do MySQL ou PostgreSQL a ser implantada.
- Computação + armazenamento. As configurações de computação, armazenamento e backup para seu novo servidor. O **Configurar link do servidor** permite que você selecione os recursos necessários para suportar suas cargas de trabalho de banco de dados. Esses recursos incluem a quantidade de capacidade de computação, memória, backups e opções de redundância (para alta disponibilidade).

**NOTA:** O termo calcular refere-se à quantidade de energia do processador disponível, mas em termos de tamanho e número de CPUs alocadas para o serviço.



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Você pode selecionar entre três níveis de preços, cada um projetado para suportar diferentes cargas de trabalho:

- Básico. Essa camada é adequada para cargas de trabalho que requerem computação leve e desempenho de I/O. Os exemplos incluem servidores usados para desenvolvimento ou teste ou aplicativos de pequena escala usados com pouca frequência.
- Propósito geral. Use esta faixa de preço para cargas de trabalho de negócios que requerem computação e memória balanceadas com rendimento de I/O escalonável. Os exemplos incluem servidores para hospedar aplicativos da web e móveis e outros aplicativos corporativos.
- Memória Otimizada Essa camada oferece suporte a cargas de trabalho de banco de dados de alto desempenho que exigem desempenho na memória para processamento de transações mais rápido e maior simultaneidade. Os exemplos incluem servidores para processamento de dados em tempo real e aplicativos transacionais ou analíticos de alto desempenho.
- Você pode ajustar os recursos disponíveis para a camada selecionada e aumentar esses recursos posteriormente, se necessário.

**NOTA:** O Configurar página exibe o desempenho que as configurações de finalidade geral e memória otimizada fornecem em termos de IOPS. IOPS é um acrônimo para “Operações de entrada / saída por segundos”, e é uma medida da capacidade de leitura e gravação disponível usando os recursos configurados.

- Nome de usuário do administrador. Uma conta de login para usar ao se conectar ao servidor. O nome de login do administrador não pode ser azure\_superuser, admin, administrador, raiz, convidado, ou público.
- Senha. Forneça uma nova senha para a conta de administrador do servidor. Deve conter de 8 a 128 caracteres. Sua senha deve conter caracteres de três das seguintes categorias: letras maiúsculas em inglês, letras minúsculas em inglês, números (0-9) e caracteres não alfanuméricos (!, \$, #, % E assim por diante).

Depois de especificar as configurações apropriadas, selecione Rever + criar para provisionar o servidor.

TUTORIAIS DE AZURE

CPF 00296691208

## Descreva a configuração de serviços de dados relacionais

Depois de provisionar um recurso, você geralmente precisará configurá-lo para atender às necessidades de seus aplicativos e ambiente. Por exemplo, você pode precisar configurar o acesso à rede ou abrir uma porta de firewall para permitir que seus aplicativos se conectem ao recurso.

Nesta unidade, você aprenderá como habilitar o acesso à rede para seus recursos e como evitar a exposição acidental de seus recursos a terceiros. Você verá como usar autenticação e controle de acesso para proteger os dados gerenciados por seus recursos.

### Configurar conectividade e firewalls

A conectividade padrão para os serviços de dados relacionais do Azure é desabilitar o acesso ao mundo.

### Configure a conectividade para redes virtuais e computadores locais

Para habilitar a conectividade, use Firewalls e redes páginas virtuais de serviço. Para habilitar a conectividade, escolha Redes selecionadas. Três outras seções aparecerão, rotuladas Virtual network, Firewall, e Exceções.

**NOTA:** Uma Rede Virtual do Azure é uma representação de sua própria rede na nuvem. Uma rede virtual permite que você conecte máquinas virtuais e serviços do Azure, da mesma forma que você pode usar uma rede física local. O Azure garante que cada rede virtual seja isolada de outras redes virtuais criadas por outros usuários e da Internet. O Azure permite que você especifique quais máquinas (reais e virtuais) e serviços têm permissão para acessar recursos na rede virtual e quais portas podem usar.

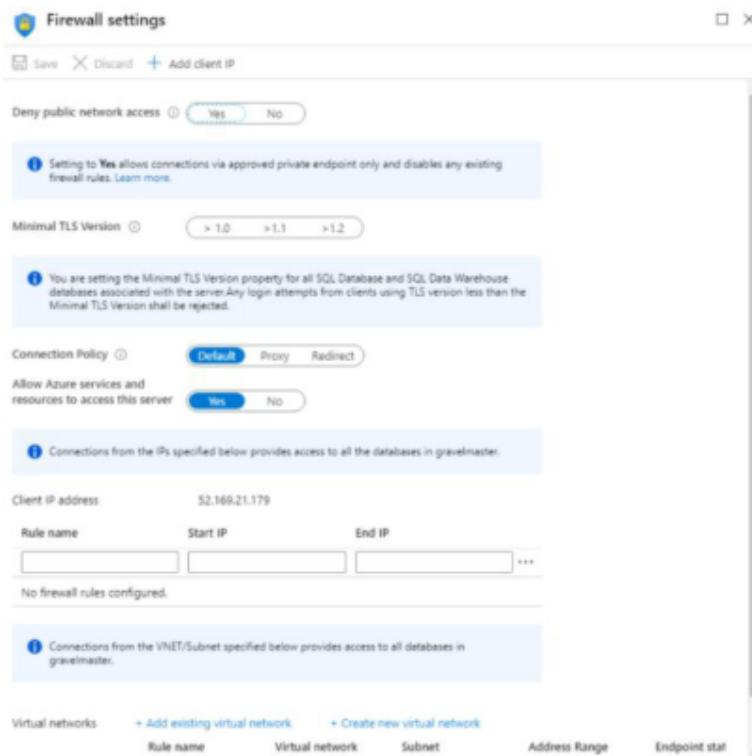
Na seção Redes virtuais, você pode especificar quais redes virtuais têm permissão para rotear o tráfego para o serviço. Ao criar itens como aplicativos da web e máquinas virtuais, você pode adicioná-los a uma rede virtual. Se esses aplicativos e máquinas virtuais exigirem acesso ao seu recurso, adicione a rede virtual que contém esses itens à lista de redes permitidas.

Se você precisar se conectar ao serviço de um computador local, na seção Firewall, adicione o endereço IP do computador. Essa configuração cria uma regra de firewall que permite que o tráfego desse endereço chegue ao serviço.

A configuração Exceções permite que você habilite o acesso a qualquer outro de seus serviços criados em sua assinatura do Azure.

A imagem abaixo mostra o Firewalls e página de redes virtuais para um banco de dados SQL do Azure. MySQL e PostgreSQL têm uma página semelhante.

CPF: 00296691208



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF: 00296691208

**NOTA:** O Banco de Dados SQL do Azure se comunica pela porta 1433. Se você estiver tentando se conectar a partir de uma rede corporativa, o tráfego de saída pela porta 1433 pode não ser permitido pelo firewall da sua rede. Nesse caso, você não pode se conectar ao servidor de Banco de Dados SQL do Azure, a menos que seu departamento de TI abra a porta 1433.

**IMPORTANTE:** Uma regra de firewall de 0.0.0.0 permite que todos os serviços do Azure passem pela regra de firewall no nível do servidor e tentem se conectar a um banco de dados único ou em pool por meio do servidor.

## Configure a conectividade de terminais privados.

Ponto final privado do Azure é uma interface de rede que conecta você de forma privada e segura a um serviço fornecido pelo Azure Private Link. O Private Endpoint usa um endereço IP privado de sua rede virtual, trazendo efetivamente o serviço para sua rede virtual. O serviço pode ser um serviço do Azure, como o Serviço de Aplicativo do Azure ou seu próprio Serviço de Link Privado. Para obter informações detalhadas, leia "O que é ponto final privado do Azure?".

A página Conexões de endpoint privadas para um serviço permite que você especifique quais terminais privados, se houver, têm acesso permitido ao seu serviço. Você pode usar as configurações nesta página, junto com o Firewalls e redes virtuais página, para bloquear completamente os usuários e aplicativos de acessar pontos de extremidade públicos para se conectar à sua conta do Banco de Dados SQL do Azure.

## Configurar autenticação

Com a autenticação do Azure Active Directory (AD), você pode gerenciar centralmente as identidades dos usuários do banco de dados e outros serviços da Microsoft em um local central. O gerenciamento de ID central fornece um local único para gerenciar usuários de banco de dados e simplifica o gerenciamento de permissões.

Você pode usar essas identidades e configurar o acesso aos seus serviços de dados relacionais.

Para obter informações detalhadas sobre como usar o Azure AD com o banco de dados SQL do Azure, visite a página “O que é autenticação do Azure Active Directory para banco de dados SQL” no site da Microsoft. Você também pode autenticar usuários que se conectam a Banco de dados Azure para PostgreSQL e Banco de dados do Azure para MySQL com AD.

### Configurar o controle de acesso

O Azure AD permite que você especifique quem ou o quê pode acessar seus recursos. O controle de acesso define o que um usuário ou aplicativo pode fazer com seus recursos depois de autenticados.

O gerenciamento de acesso para recursos de nuvem é uma função crítica para qualquer organização que usa a nuvem. O controle de acesso baseado em função do Azure (Azure RBAC) ajuda a gerenciar quem tem acesso aos recursos do Azure e o que eles podem fazer com esses recursos. Por exemplo, usando RBAC, você pode:

- Permitir que um usuário gerencie máquinas virtuais em uma assinatura e outro usuário gerencie redes virtuais.
- Permitir que um grupo de administradores de banco de dados gerencie bancos de dados SQL em uma assinatura.
- Permitir que um usuário gerencie todos os recursos em um grupo de recursos, como máquinas virtuais, sites e sub-redes.
- Permitir que um aplicativo acesse todos os recursos em um grupo de recursos.

Você controla o acesso aos recursos usando o Azure RBAC para criar atribuições de função. Uma atribuição de função consiste em três elementos: uma entidade de segurança, uma definição de função e um escopo.

A security principal é um objeto que representa um usuário, grupo, entidade de serviço ou identidade gerenciada que está solicitando acesso aos recursos do Azure.

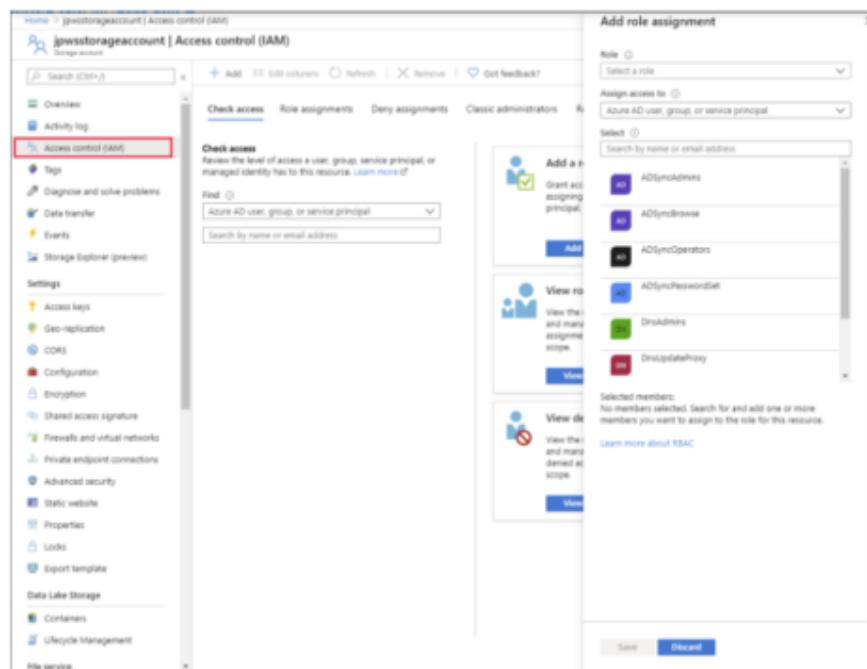
A definição de papel, frequentemente abreviado para Função, é uma coleção de permissões. Uma definição de função lista as operações que podem ser executadas, como leitura, gravação e exclusão. As funções podem receber nomes de alto nível, como proprietário, ou nomes específicos, como leitor de máquina virtual. O Azure inclui várias funções internas que você pode usar, incluindo:

- Proprietário - Tem acesso total a todos os recursos, incluindo o direito de delegar acesso a outros.
- Contribuinte - Pode criar e gerenciar todos os tipos de recursos do Azure, mas não pode conceder acesso a outros.
- Leitor- Pode visualizar os recursos existentes do Azure.
- Administrador de acesso do usuário - Permite gerenciar o acesso do usuário aos recursos do Azure.

Você também pode criar suas próprias funções personalizadas. Para obter informações detalhadas, consulte Criar ou atualizar funções personalizadas do Azure usando o portal do Azure no site da Microsoft.

O escopo lista o conjunto de recursos aos quais o acesso se aplica. Ao atribuir uma função, você pode limitar ainda mais as ações permitidas definindo um escopo. Isso é útil se, por exemplo, você quiser tornar alguém um Contribuidor do Site, mas apenas para um grupo de recursos.

Você adiciona atribuições de função a um recurso no portal do Azure usando o Access control (IAM). O guia Atribuições de funções permite que você associe uma função a uma entidade de segurança, definindo o nível de acesso da função ao recurso. Para mais informações, leia “Adicionar ou remover atribuições de função do Azure usando o portal do Azure”.



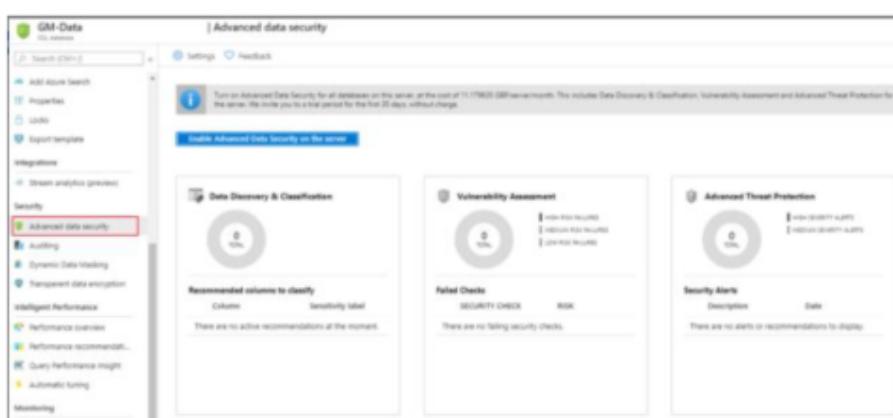
THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Configurar segurança avançada de dados

Além de autenticação e autorização, muitos serviços fornecem proteção adicional por meio de segurança avançada de dados.

A segurança de dados avançada implementa proteção e avaliação de ameaças. A proteção contra ameaças adiciona inteligência de segurança ao seu serviço. Essa inteligência monitora o serviço e detecta padrões incomuns de atividade que podem ser prejudiciais ou comprometer os dados gerenciados pelo serviço. A avaliação identifica vulnerabilidades de segurança em potencial e recomenda ações para mitigá-las.

A imagem abaixo mostra o Segurança de dados avançado página para banco de dados SQL. As páginas correspondentes para MySQL e PostgreSQL são semelhantes.



## Descrever a configuração do Banco de Dados SQL do Azure, Banco de Dados do Azure para PostgreSQL e Banco de Dados do Azure para MySQL

Esta unidade explora as opções de configuração específicas disponíveis para cada tipo de armazenamento de dados nos serviços de dados relacionais do Azure.

### Configurar Banco de Dados SQL do Azure

O princípio abrangente para a segurança de rede da oferta do Banco de Dados SQL do Azure é permitir apenas a conexão e a comunicação necessárias para permitir a operação do serviço. Todas as outras portas, protocolos e conexões são bloqueados por padrão. As redes locais virtuais (VLANs) e as listas de controle de acesso (ACLs) são usadas para restringir as comunicações de rede por redes de origem e destino, protocolos e números de porta.

**NOTA:** Uma ACL contém uma lista de recursos e os objetos (usuários, computadores e aplicativos) que têm permissão para acessar esses recursos. Quando um objeto tenta usar um recurso que está protegido por uma ACL, se não estiver na lista, ele não terá acesso.

Os itens que implementam ACLs baseados em rede incluem roteadores e平衡adores de carga. O fluxo de tráfego por meio desses itens é controlado pela definição de regras de firewall.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

As etapas a seguir descrevem como uma conexão é estabelecida com um banco de dados SQL do Azure:

- Os clientes se conectam a um Porta de entrada que tem um endereço IP público e escuta na porta 1433.
- Dependendo da política de conexão efetiva, o gateway redireciona o tráfego para o cluster de banco de dados ou atua como um proxy para o cluster de banco de dados.

**NOTA:** O Banco de Dados SQL do Azure usa uma topologia em cluster para fornecer alta disponibilidade. Cada servidor e banco de dados são replicados de forma transparente para garantir que um servidor esteja sempre acessível, mesmo no caso de um banco de dados ou falha do servidor.

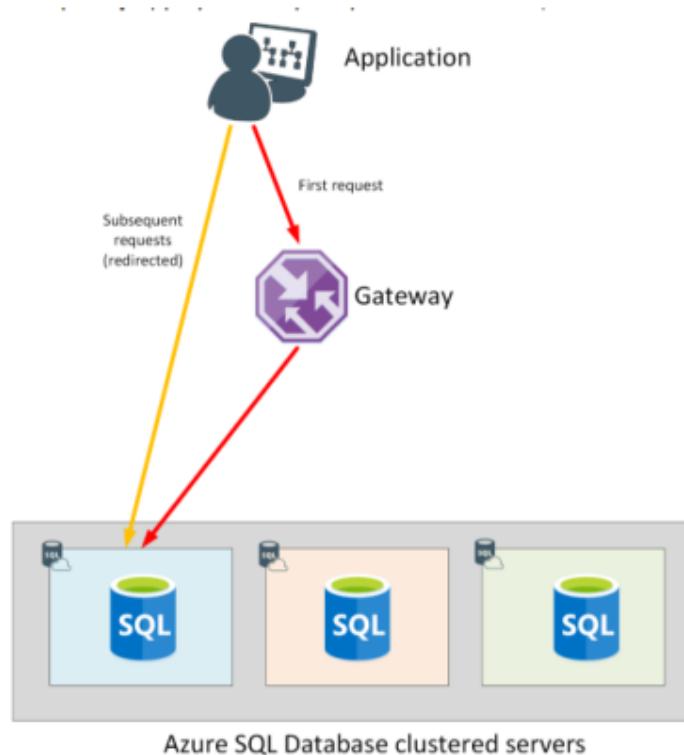
Dentro do cluster de banco de dados, o tráfego é encaminhado para o banco de dados SQL do Azure apropriado.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## Conectividade de dentro do Azure

Se você estiver se conectando a partir de outro serviço do Azure, como um aplicativo da web em execução no Serviço de Aplicativo do Azure, suas conexões têm uma política de conexão de Redirecionar por padrão. Uma política de redirecionamento significa que, depois que seu aplicativo estabelece uma conexão com o banco de dados SQL do Azure por meio do gateway, todas as solicitações a seguir do seu aplicativo irão diretamente para o banco de dados, em vez de pelo gateway. Se a conectividade com o banco de dados falhar posteriormente, seu aplicativo terá que se reconectar por meio do gateway, quando pode ser direcionado para uma cópia diferente do banco de dados em execução em outro servidor no cluster.

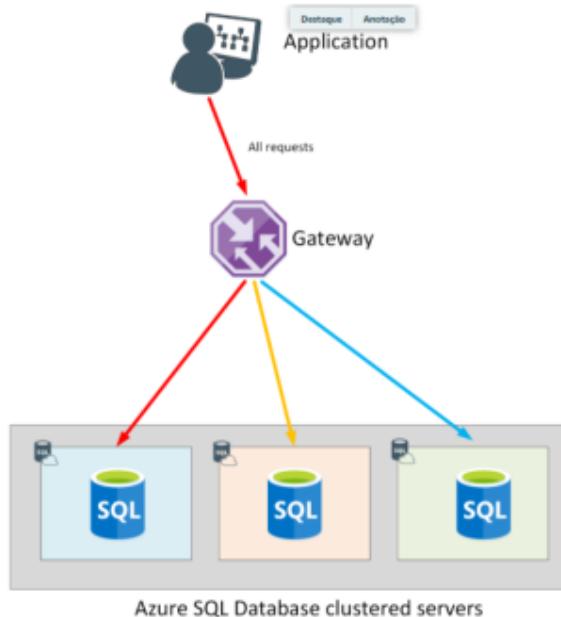


## Conectividade de fora do Azure

Se você estiver se conectando de fora do Azure, como um aplicativo local, suas conexões têm uma política de conexão de Proxy por padrão. Uma política de Proxy significa que a conexão é estabelecida por meio do gateway e todas as solicitações subsequentes fluem pelo gateway. Cada solicitação pode (potencialmente) ser atendida por um banco de dados diferente no cluster.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 000296691208



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF: 00296691208

## Configurar DoSGuard

Ataques de negação de serviço (DoS) são reduzidos por um serviço de gateway de banco de dados SQL chamado DoSGuard. DoSGuard rastreia ativamente logins com falha de endereços IP. Se houver vários logins com falha de um endereço IP específico dentro de um período de tempo, o endereço IP será impedido de acessar quaisquer recursos no serviço por um curto período.

Além disso, o gateway do Banco de Dados SQL do Azure executa as seguintes tarefas:

- Ele valida todas as conexões com os servidores de banco de dados, para garantir que sejam de clientes genuínos.
- Ele criptografa todas as comunicações entre um cliente e os servidores de banco de dados.
- Ele inspeciona cada pacote de rede enviado por uma conexão de cliente. O gateway valida as informações de conexão no pacote e as encaminha ao servidor físico apropriado com base no nome do banco de dados especificado na string de conexão.

## Configurar Banco de Dados do Azure para PostgreSQL

Quando você cria seu Banco de Dados do Azure para o servidor PostgreSQL, um banco de dados padrão chamado `postgres` é criado. Para se conectar ao servidor de banco de dados, você precisa do nome completo do servidor e das credenciais de login do administrador. Você pode encontrar facilmente o nome do servidor e informações de login no servidor na página Visão geral do portal. Esta página contém o nome do servidor e o nome de login do administrador do servidor.

The screenshot shows the Azure portal interface for a PostgreSQL server named "mydemoserver". The left sidebar has a red box around the "Overview" tab. The main content area shows the server's resource group ("myresourcegroup"), status ("Available"), location ("Southeast Asia"), subscription name ("Free Trial"), and subscription ID ("000000-0000-0000-0000-000000000000"). On the right, there's a section for "Server name" which is "mydemoserver.postgres.database.azure.com" and "Server admin login name" which is "myadmin@mydemoserver". Below that, it shows "PostgreSQL version" (9.6), "Performance configuration" (General Purpose, 2 vCore(s), 5 GB), and "SSL enforcement status" (ENABLED). A red box highlights the "Server name" and "Server admin login name" fields.

**NOTA:** As conexões com o Banco de Dados do Azure para o servidor PostgreSQL se comunicam pela porta 5432. Quando você tenta se conectar a partir de uma rede corporativa, o tráfego de saída pela porta 5432 pode não ser permitido pelo firewall da sua rede. Nesse caso, você não pode se conectar ao seu servidor, a menos que seu departamento de TI abra a porta 5432.

## Configure os parâmetros e extensões do servidor

Um servidor de banco de dados PostgreSQL possui vários parâmetros de configuração que você pode definir. Esses parâmetros oferecem suporte ao ajuste fino do banco de dados e à depuração de código no banco de dados. Você pode modificar esses parâmetros usando a página Parâmetros do servidor no portal do Azure.

The screenshot shows the "Server parameters" section in the Azure portal. The left sidebar has a red box around the "Server parameters" tab under the "Settings" category. The main table lists various PostgreSQL parameters with their current values and descriptions. Some parameters shown include: `array_nulls` (ON), `backslash_quote` (`SQL_BACKSLASH`), `bytea_output` (`HEX`), `check_function_bodies` (OFF), `client_encoding` (`SQL_ASCII`), `client_min_messages` (`NOTICE`), `connection_throttling` (OFF), `constraint_exclusion` (`PARALLEL`), `cpu_index_tuple_cost` (0.005), `cpu_operator_cost` (0.0025), `cpu_tuple_cost` (0.01), `cursor_tuple_fraction` (0.1), `datestyle` (`iso, mdy`)), `deadlock_timeout` (1000), `debug_print_parse` (ON), `debug_print_plan` (ON), `debug_print_rewritten` (ON), `default_statistics_target` (100), `default_text_search_config` (`pg_catalog_english`), `default_transaction_deferrable` (ON), `default_transaction_isolation` (`READ_COMMITTED`), `default_transaction_read_only` (ON), and `default_withoids` (ON). A red box highlights the "Parameter name" column.

Se estiver familiarizado com o PostgreSQL, você descobrirá que nem todos os parâmetros têm suporte no Azure. A página Parâmetros do servidor no site da Microsoft descreve os parâmetros PostgreSQL que estão disponíveis.

O PostgreSQL também oferece a capacidade de estender a funcionalidade de seu banco de dados usando extensões. As extensões agrupam vários objetos SQL relacionados em um único pacote que pode ser carregado ou removido do banco de dados com um único comando. Depois de carregadas no banco de dados, as extensões funcionam como recursos integrados. Você instala uma extensão em seu banco de dados antes de usá-la. Para instalar uma extensão específica, execute o comando `CRIAR EXTENSÃO` da ferramenta `psql` para carregar os objetos empacotados em seu banco de dados. Nem todas as extensões PostgreSQL têm suporte no Azure. Para uma lista completa, leia “Extensões PostgreSQL no Banco de Dados Azure para PostgreSQL - Servidor Único”.

### Configurar réplicas de leitura

Você pode replicar dados de um Banco de Dados do Azure para o servidor PostgreSQL para um servidor somente leitura. O Banco de Dados do Azure para PostgreSQL oferece suporte à replicação do servidor mestre para até cinco réplicas. As réplicas são atualizadas de forma assíncrona com a tecnologia de replicação nativa do mecanismo PostgreSQL.

As réplicas de leitura ajudam a melhorar o desempenho e a escala de cargas de trabalho com leitura intensiva. As cargas de trabalho de leitura podem ser isoladas nas réplicas, enquanto as cargas de trabalho de gravação podem ser direcionadas ao mestre.

Um cenário comum é fazer com que as cargas de trabalho analíticas e de BI usem réplicas de leitura como fonte de dados para relatórios.

Como as réplicas são somente leitura, elas não reduzem diretamente a carga das operações de gravação no mestre. Este recurso não é direcionado a cargas de trabalho com intensa gravação.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

As réplicas são novos servidores que você gerencia de forma semelhante ao Banco de Dados do Azure regular para servidores PostgreSQL. Para cada réplica de leitura, você é cobrado pela computação provisionada em vCores e armazenamento em GB / mês.

Use o Replicação página de um servidor PostgreSQL no portal do Azure para adicionar réplicas de leitura ao seu banco de dados:

| Master    | Name      | Pricing tier           | Location | Status    |
|-----------|-----------|------------------------|----------|-----------|
| gutenberg | gutenberg | Basic, 2 vCores, 50 GB | East US  | Available |

| Replica          | Name             | Pricing tier           | Location | Status    |
|------------------|------------------|------------------------|----------|-----------|
| gutenbergreplica | gutenbergreplica | Basic, 2 vCores, 50 GB | East US  | Available |

## Configurar Banco de Dados do Azure para MySQL

Para se conectar ao banco de dados MySQL que você provisionou, você precisará inserir as informações de conexão. Essas informações incluem o nome do servidor totalmente qualificado e as credenciais de entrada. Você pode encontrar essas informações na página Visão geral para o seu servidor:

**NOTA:** As conexões com o Banco de Dados do Azure para servidor MySQL se comunicam pela porta 3306. Quando você tenta se conectar a partir de uma rede corporativa, o tráfego de saída pela porta 3306 pode não ser permitido pelo firewall da sua rede. Nesse caso, você não pode se conectar ao seu servidor a menos que seu departamento de TI abra a porta 3306.

**IMPORTANTE:** Por padrão, a segurança da conexão SSL é necessária e aplicada em seu Banco de Dados do Azure para servidor MySQL.

## Configure os parâmetros do servidor

Como o PostgreSQL, um servidor de banco de dados MySQL possui vários parâmetros de configuração que você pode definir. Você pode modificar esses parâmetros usando a página Parâmetros do servidor no portal do Azure.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

The screenshot shows the Azure portal interface for managing a MySQL database. On the left, there's a sidebar with various navigation options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings, Connection security, Connection strings, Server parameters (which is highlighted with a red box), Replication, Active Directory admin, Pricing tier, Properties, Locks, Export template, Security, Advanced Threat Protection ..., Private endpoint connections, Intelligent Performance, Query Performance Insight, Performance recommendations, Monitoring, Alerts, Metrics, and Diagnostic settings. The main area displays a table of server parameters with columns: Parameter name, Value, Parameter Type, and Description. Some of the visible parameters include audit\_log\_enabled (OFF), audit\_log\_events (CONNECTION), audit\_log\_exclude\_users (azure\_superuser), audit\_log\_include\_users (empty), binlog\_expire\_logs\_seconds (43200), binlog\_group\_commit\_sync\_delay (0), binlog\_group\_commit\_sync\_no\_delay\_count (0), binlog\_row\_image (MINIMAL), character\_set\_server (LATIN1), collation\_server (LATIN1\_SWEDISH\_CI), default\_week\_format (0), div\_precision\_increment (4), eq\_range\_index\_dive\_limit (200), event\_scheduler (OFF), flush\_time (0), ft\_query\_expansion\_limit (20), group\_concat\_max\_len (1024), init\_connect (empty), innodb\_adaptive\_flushing (ON), innodb\_adaptive\_flushing\_lwm (10), innodb\_adaptive\_hash\_index (ON), innodb\_adaptive\_hash\_index\_parts (8), and innodb\_adaptive\_max\_sleep\_delay (150000). There are also buttons for Save, Discard, and Reset all to default.

Você pode encontrar mais informações sobre os parâmetros disponíveis para MySQL no Azure em “Como configurar os parâmetros do servidor no Banco de Dados do Azure para MySQL usando o portal do Azure” na página da Microsoft.

## Configurar réplicas de leitura

Este recurso é semelhante ao disponível para PostgreSQL. Você pode criar até cinco réplicas de leitura para um banco de dados MySQL. Esse recurso permite que você replique geograficamente os dados entre as regiões e distribua a sobrecarga associada a cargas de trabalho com leitura intensiva. A replicação é assíncrona do servidor mestre, portanto, pode haver algum atraso entre os registros sendo gravados no mestre e tornando-se disponíveis em todas as réplicas.

- A replicação de leitura não se destina a suportar cargas de trabalho pesadas de gravação.
- Use o Replicação página de um servidor MySQL no portal do Azure para adicionar réplicas de leitura ao seu banco de dados.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Laboratório: Provisionar serviço de banco de dados relacional do Azure

Como parte de sua função na Contoso como engenheiro de dados, você foi solicitado para criar e configurar servidores SQL Server, PostgreSQL e MySQL para Azure. Você pode escolher usar o Banco de Dados SQL do Azure, PostgreSQL ou MySQL.

Vá ao Exercício: “Provisionar serviços de dados não relacionais do Azure” no Microsoft Learn e siga as instruções do módulo para criar os seguintes armazenamentos de dados:

- Um Cosmos DB para conter informações sobre o volume de itens em estoque. Você precisa armazenar informações atuais e históricas sobre os níveis de volume, para que possa acompanhar como os níveis variam ao longo do tempo. Os dados são registrados diariamente.
- Uma loja Data Lake para armazenar dados de produção e qualidade.
- Um contêiner blob para armazenar imagens dos produtos que a empresa fabrica.
- Armazenamento de arquivos para compartilhamento de relatórios.

## Resumo

Nesta lição, você aprendeu como provisionar e implantar bancos de dados relacionais usando diferentes tipos de armazenamentos de dados. Você viu como pode implantar os serviços de dados do Azure por meio do portal do Azure, da CLI do Azure e do Azure PowerShell. Você também aprendeu como configurar a conectividade com esses bancos de dados para permitir o acesso local ou em uma rede virtual do Azure. Você também viu como proteger seu banco de dados usando ferramentas como o firewall e configurando a autenticação.

## Recursos adicionais

- Crie um Banco de Dados do Azure para PostgreSQL
- Crie um Banco de Dados do Azure para MySQL
- Crie um único banco de dados do Azure
- Documentação do Banco de Dados SQL do Azure
- Parâmetros do servidor PostgreSQL
- Extensões PostgreSQL no Banco de Dados Azure para PostgreSQL - Servidor Único
- Como configurar os parâmetros do servidor no Banco de Dados do Azure para MySQL usando o portal do Azure

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Introdução

O Azure permite que você crie bancos de dados relacionais usando várias tecnologias, incluindo Banco de Dados SQL do Azure, Banco de Dados do Azure para PostgreSQL, Banco de Dados do Azure para MySQL e Banco de Dados do Azure para MariaDB.

Imagine que você trabalhe como desenvolvedor para uma grande rede de supermercados chamada Contoso. A empresa criou um armazenamento de dados que será usado para armazenar o estoque de produtos. A equipe de desenvolvimento usou um banco de dados SQL do Azure para armazenar seus dados. Eles precisam saber como consultar e manipular esses dados usando SQL.

Nesta lição, você aprenderá como usar esses serviços de banco de dados para armazenar e recuperar dados. Você examinará como usar algumas das ferramentas comuns disponíveis para esses sistemas de gerenciamento de banco de dados para se conectar aos serviços de banco de dados em execução no Azure.

**NOTA:** Esta lição se concentra no uso do Banco de Dados SQL do Azure, Banco de Dados do Azure para PostgreSQL e Banco de Dados do Azure para MySQL. Se você estiver usando o Banco de Dados do Azure para MariaDB, o dialeto do SQL é muito semelhante ao usado pelo MySQL.

## Objetivos de aprendizado

Nesta lição, você irá:

- Descrever técnicas de consulta para dados usando a linguagem SQL
- Consultar dados relacionais
- Ser introduzido ao SQL

SQL significa Structured Query Language. SQL é usado para se comunicar com um banco de dados relacional. É a linguagem padrão para sistemas de gerenciamento de banco de dados relacional. As instruções SQL são usadas para executar tarefas como atualizar dados em um banco de dados ou recuperar dados de um banco de dados. Alguns sistemas de gerenciamento de banco de dados relacional comuns que usam SQL incluem Microsoft SQL Server, MySQL, PostgreSQL, MariaDB e Oracle.

TUTORIAIS DE APRENDIZADO DA SÉRIE

CPF: 00296691208

**NOTA:** SQL foi originalmente padronizado pelo American National Standards Institute (ANSI) em 1986, e pela International Organization for Standardization (ISO) em 1987. Desde então, o padrão foi estendido várias vezes, à medida que fornecedores de banco de dados relacional adicionaram novos recursos a seus sistemas. Além disso, a maioria dos fornecedores de banco de dados inclui suas próprias extensões proprietárias que não fazem parte do padrão, o que resultou em uma variedade de dialetos do SQL.

Nesta unidade, você aprenderá sobre SQL. Você verá como ele é usado para consultar e manter dados em um banco de dados e os diferentes dialetos disponíveis.

## Entenda os dialetos SQL

Você pode usar instruções SQL como SELECIONAR, INSERIR, ATUALIZAR, EXCLUIR, CRIAR, e SOLTAR para realizar quase tudo o que é necessário fazer com um banco de dados. Embora essas instruções SQL façam parte do padrão SQL, muitos sistemas de gerenciamento de banco de dados também possuem suas próprias extensões proprietárias adicionais para lidar com as especificidades desse sistema de gerenciamento de banco de dados. Essas extensões fornecem funcionalidades não cobertas pelo padrão SQL e incluem áreas como gerenciamento de segurança e programação. O Microsoft SQL Server, por exemplo, usa Transact-SQL. Essa implementação inclui extensões proprietárias para escrever procedimentos armazenados e gatilhos (código de aplicativo que pode ser armazenado no banco de dados) e gerenciar contas de usuário. PostgreSQL e MySQL também têm suas próprias versões desses recursos.

Alguns dialetos populares do SQL incluem:

- Transact-SQL (SQL). Esta versão do SQL é usada pelo Microsoft SQL Server e pelo Banco de Dados SQL do Azure.
- pgSQL. Este é o dialeto, com extensões implementadas no PostgreSQL.
- PL / SQL. Este é o dialeto usado pelo Oracle. PL / SQL significa Linguagem de Procedimentos / SQL.

Os usuários que planejam trabalhar especificamente com um único sistema de banco de dados devem aprender as complexidades de sua plataforma e dialeto SQL preferidos.

## Entenda os tipos de instrução SQL

As instruções SQL são agrupadas em dois grupos lógicos principais e são:

- Linguagem de manipulação de dados (DML)
- Linguagem de definição de dados (DDL)

## Use instruções DML

Você usa instruções DML para manipular as linhas em uma tabela relacional. Essas instruções permitem recuperar (consultar) dados, inserir novas linhas ou editar linhas existentes. Você também pode excluir linhas se não precisar mais delas.

As quatro principais declarações DML são:

| Declaração | Descrição                               |
|------------|---|
| SELECIONE  | Selecionar / ler linhas de uma tabela   |
| INSERIR    | Insira novas linhas em uma tabela       |
| ATUALIZAR  | Editar / atualizar linhas existentes    |
| EXCLUIR    | Excluir linhas existentes em uma tabela |

A forma básica de uma declaração INSERIR irá inserir uma linha de cada vez. Por padrão, as instruções SELECIONE, ATUALIZAR, e EXCLUIR são aplicadas a todas as linhas de uma tabela. Você geralmente aplica um ONDE com essas instruções para especificar critérios; apenas as linhas que correspondem a esses critérios serão selecionadas, atualizadas ou excluídas.

**ATENÇÃO:** SQL não fornece prompts “você tem certeza?”, portanto, tome cuidado ao usar DELETE ou UPDATE sem uma cláusula ONDE, pois você pode perder ou modificar muitos dados.

O código a seguir é um exemplo de uma instrução SQL que seleciona todas as linhas que correspondem a um único filtro de uma tabela. A cláusula PARTIR DE especifica a tabela a ser usada:

```
SELECT *
FROM MyTable
WHERE MyColumn2 = 'contoso'
```

Se uma consulta retornar muitas linhas, elas não aparecerão necessariamente em uma sequência específica. Se você quiser classificar os dados, pode adicionar um ORDENAR POR cláusula. Os dados serão classificados pela coluna especificada:

```
SELECT *
FROM MyTable
ORDER BY MyColumn1
```

Você também pode executar instruções SELECT que recuperam dados de várias tabelas usando o JUNTE-SE. As junções indicam como as linhas de uma tabela são conectadas às linhas da outra para determinar quais dados retornar. Uma condição de junção define a maneira como duas tabelas estão relacionadas em uma consulta por:

Especificando a coluna de cada tabela a ser usada para a junção. Uma condição de junção típica especifica uma chave estrangeira de uma tabela e sua chave primária associada na outra tabela.

Especificando um operador lógico (por exemplo, = ou <>) a ser usado na comparação de valores das colunas.

A consulta a seguir mostra um exemplo que une duas tabelas, chamadas Inventário e Pedido do cliente. Ele recupera todas as linhas onde o valor na coluna ID na tabela Inventário corresponde ao valor noInventoryId coluna no Inventário tabela corresponde ao valor no InventoryID coluna Pedido de Cliente da tabela.

```
SELECT *
FROM Inventory
JOIN CustomerOrder
WHERE Inventory.ID = CustomerOrder.InventoryID
```

SQL fornece funções agregadas. Uma função de agregação calcula um único resultado em um conjunto de linhas ou em uma tabela inteira. O exemplo abaixo encontra o valor mínimo noMyColumn1 coluna em todas as linhas na tabela Minha Tabela:

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

```
SELECT MIN(MyColumn1)  
FROM MyTable
```

Várias outras funções de agregação estão disponíveis, incluindo MAX (que retorna o maior valor em uma coluna), AVG (que retorna o valor médio, mas apenas se a coluna contiver dados numéricos) e SUM (que retorna a soma de todos os valores na coluna, mas novamente, apenas se a coluna for numérica).

O próximo exemplo mostra como atualizar uma linha existente usando SQL. Ele modifica o valor da segunda coluna, mas apenas para linhas que têm o valor 3 em MyColumn3. Todas as outras linhas permanecem inalteradas:

```
UPDATE MyTable  
SET MyColumn2 = 'contoso'  
WHERE MyColumn1 = 3
```

**ATENÇÃO:** Se você omitir o ONDE, uma declaração ATUALIZAR irá modificar cada linha na tabela.

Use a declaração EXCLUIR para remover linhas. Você especifica a tabela a ser excluída e um ONDE que identifica as linhas a serem excluídas:

```
DELETE FROM MyTable  
WHERE MyColumn2 = 'contoso'
```

**ATENÇÃO:** Se você omitir o ONDE, a EXCLUIR declaração irá remover cada linha da mesa.

o INSERIR assume uma forma ligeiramente diferente. Você especifica uma tabela e colunas em um PARA DENTRO e uma lista de valores a serem armazenados nessas colunas. O SQL padrão oferece suporte apenas à inserção de uma linha por vez, conforme mostrado no exemplo a seguir. Alguns dialetos permitem que você especifique vários VALORES para adicionar várias linhas de uma vez:

```
INSERT INTO MyTable(MyColumn1, MyColumn2, MyColumn3)  
VALUES (99, 'contoso', 'hello')
```

## Use instruções DDL

Você usa instruções DDL para criar, modificar e remover tabelas e outros objetos em um banco de dados (tabela, procedimentos armazenados, visualizações e assim por diante).

As instruções DDL mais comuns são:

| Declaração | Descrição   |
|------------|---|
| CRIAR      | Crie um novo objeto no banco de dados, como uma tabela ou visualização.                             |
| ALTERAR    | Modifique a estrutura de um objeto. Por exemplo, alterar uma tabela para adicionar uma nova coluna. |
| SOLTAR     | Remova um objeto do banco de dados.   |
| RENOMEAR   | Renomeie um objeto existente.   |

**ATENÇÃO:** A declaração SOLTA é muito poderosa. Quando você remove uma tabela, todas as linhas dessa tabela são perdidas. A menos que você tenha um backup, não será possível recuperar esses dados.

O exemplo a seguir cria uma nova tabela de banco de dados. Os itens entre parênteses especificam os detalhes de cada coluna, incluindo o nome, o tipo de dados, se a coluna deve sempre conter um valor (NÃO NULO) e se os dados na coluna são usados para identificar exclusivamente uma linha (PRIMARY KEY). Cada tabela deve ter uma chave primária, embora o SQL não imponha essa regra.

**NOTA:** Colunas marcadas como NÃO NULO são referenciados como colunas obrigatórias. Se você omitir o NÃO NULO, você pode criar linhas que não contenham um valor na coluna. Diz-se que uma coluna vazia em uma linha tem um valor NULO.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

```
CREATE TABLE MyTable  
(  
    MyColumn1 INT NOT NULL PRIMARY KEY,  
    MyColumn2 VARCHAR(50) NOT NULL,  
    MyColumn3 VARCHAR(10) NULL  
);
```

Os tipos de dados disponíveis para colunas em uma tabela variam entre os sistemas de gerenciamento de banco de dados. No entanto, a maioria dos sistemas de gerenciamento de banco de dados suporta tipos numéricos, como INT, e tipos de string, como VARCHAR (VARCHAR significa dados de caracteres de comprimento variável). Para obter mais informações, consulte a documentação do sistema de gerenciamento de banco de dados selecionado.

## Consultar dados relacionais no Banco de Dados SQL do Azure

Você executa comandos SQL a partir de ferramentas e utilitários que se conectam ao banco de dados apropriado. As ferramentas disponíveis dependem do sistema de gerenciamento de banco de dados que você está usando. Nesta unidade, você aprenderá sobre as ferramentas que pode usar para se conectar ao Banco de Dados SQL do Azure.

## Recuperar informações de conexão para o Banco de Dados SQL do Azure

Você pode usar qualquer uma dessas ferramentas para consultar dados mantidos no Banco de Dados SQL do Azure:

- O editor de consulta no portal do Azure
- O sqlcmd utilitário da linha de comando ou do Azure Cloud Shell
- SQL Server Management Studio
- Azure Data Studio
- Ferramentas de dados do SQL Server

Para usar essas ferramentas, primeiro você precisa estabelecer uma conexão com o banco de dados. Você exigirá os detalhes do servidor ao qual se conectar, uma conta do Banco de Dados SQL do Azure (um nome de usuário e uma senha) que tenha acesso a este servidor e o nome do banco de dados a ser usado neste servidor.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Você pode encontrar o nome do servidor de um banco de dados usando o portal do Azure: vá para a página do seu banco de dados e no Visão geral página, observe o nome do servidor totalmente qualificado no Nome do servidor campo.

The screenshot shows the Azure portal interface for a SQL database named 'coursedatabase101'. On the left, there's a sidebar with various navigation options like Overview, Activity log, Tags, and Power Platform. The main content area displays the 'Overview' page for the database. At the top right, there are several actions: Copy, Restore, Export, Set server firewall, Delete, Connect with..., and Feedback. Below these actions, the 'Server name' field is highlighted with a red box, containing the value 'courseserver102.database.windows.net'. Other details shown include Status (Online), Location (East US), Subscription (change), Pricing tier (Basic), and Compute utilization. There are also sections for Tags and Show data for last 1 hour, 24 hours, or 7 days.

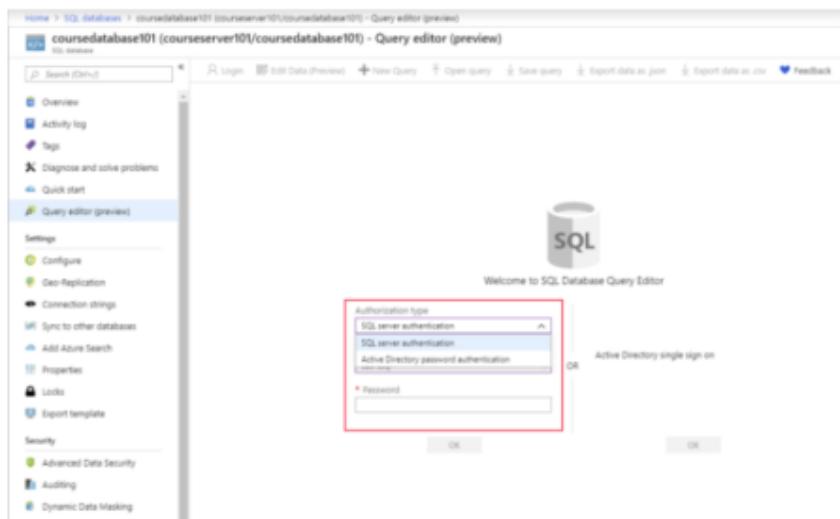
Algumas ferramentas e aplicativos requerem uma string de conexão que identifica o servidor, banco de dados, nome da conta e senha. Você pode encontrar essas informações no Visão geral página para um banco de dados no portal do Azure: selecione Mostrar strings de conexão de banco de dados.

**NOTA:** A cadeia de conexão do banco de dados mostrada no portal do Azure não inclui a senha da conta. Você deve entrar em contato com o administrador do banco de dados para obter essas informações.

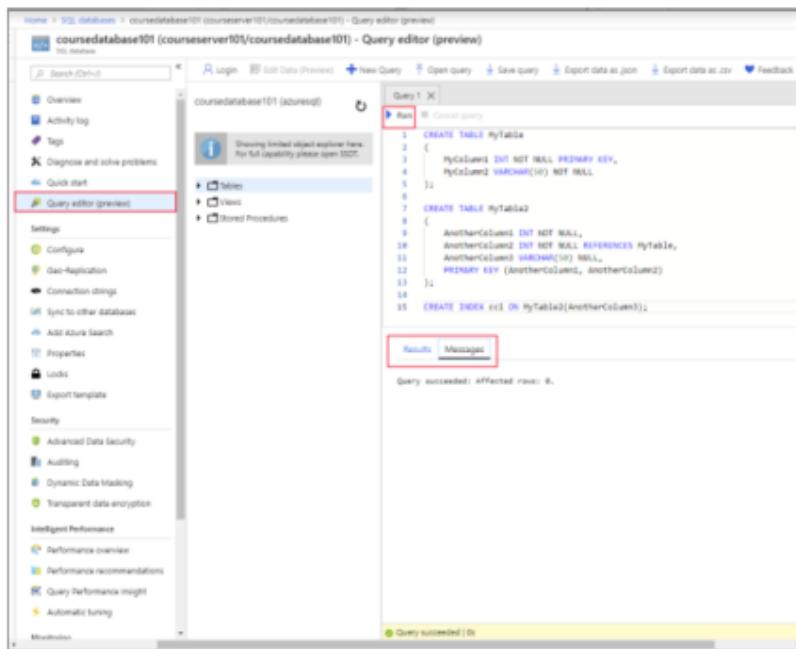
## Use o portal do Azure para consultar um banco de dados

Para acessar o editor de consulta no portal do Azure, vá para a página do seu banco de dados e selecione Editor de consultas. Você será solicitado a fornecer credenciais. Você pode definir o Tipo de autorização para Autenticação SQL Servere insira o nome de usuário e a senha que você configurou ao criar o banco de dados. Ou você pode selecionar Autenticação de senha do Active Directory fornecer as credenciais de um usuário autorizado no Azure Active Directory. Se o logon único do Active Directory estiver habilitado, você pode se conectar usando sua identidade do Azure.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208



Você insere sua consulta SQL no painel de consulta e clique em Correpara executá-lo. Quaisquer linhas retornadas aparecem no painel Resultados. O painel Mensagens exibe informações como o número de linhas retornadas ou quaisquer erros que ocorreram:



Você também pode inserir as instruções INSERT, UPDATE, DELETE, CREATE e DROP no painel de consulta.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Use SQLCMD para consultar um banco de dados

O utilitário sqlcmdO é executado na linha de comando e também está disponível no Cloud Shell. Você especifica parâmetros que identificam o servidor, banco de dados e suas credenciais. O código abaixo mostra um exemplo. Substituir<servidor> com o nome do servidor de banco de dados que você criou, <banco de dados> com o nome do seu banco de dados, e <nome do usuário> e <senha> com suas credenciais.

**NOTA:** Para usar o utilitário sqlcmd na linha de comando, você deve instalar os utilitários de linha de comando da Microsoft em seu computador. Você pode encontrar instruções de download e mais detalhes sobre a execução do utilitário sqlcmd no utilitário sqlcmd página da web.

```
sqlcmd -S <server>.database.windows.net -d <database> -U <username> -P  
<password>
```

Se o comando de login for bem-sucedido, você verá um 1>pronto. Você pode inserir comandos SQL e, em seguida, digitar IR em uma linha por si só para executá-los.

## Use o Azure Data Studio

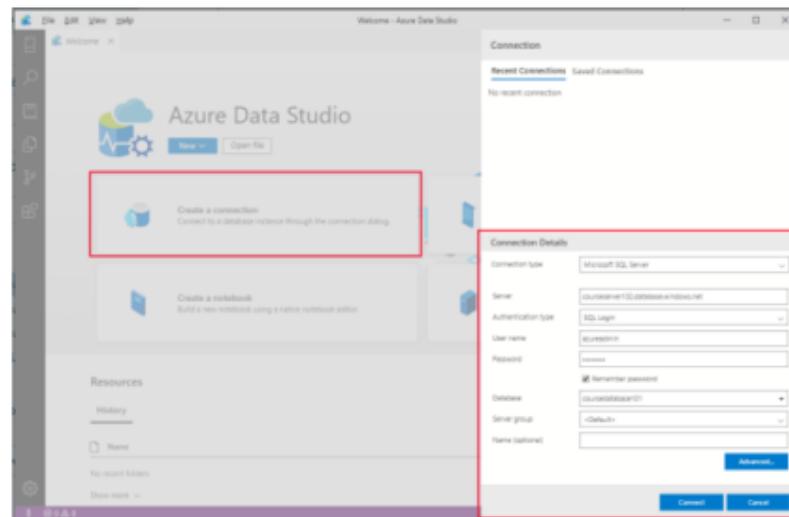
Azure Data Studio é um utilitário gráfico para criar e executar consultas SQL a partir de sua área de trabalho. Para obter instruções de download e instalação, visite o “Baixe e instale o Azure Data Studio” página da Microsoft.

A primeira vez que você executa o Azure Data Studio, a página Bem-vindo deve abrir. Se você não vê a página, selecione Socorro selecione Bem-vindo. Selecione Crie uma conexão para abrir o Conexão painel:

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

- Preencha os seguintes campos usando o nome do servidor, o nome de usuário e a senha do Azure SQL Server:

| Configuração           | Descrição  |
|------------------------|--|
| Nome do servidor       | O nome completo do servidor. Você pode encontrar o nome do servidor no portal do Azure, conforme descrito anteriormente.   |
| Autenticação           | Login do SQL ou autenticação do Windows. A menos que você esteja usando o Azure Active Directory, selecione Login do SQL.  |
| Nome do usuário        | O nome de usuário da conta do administrador do servidor. Especifique o nome de usuário da conta usada para criar o servidor.   |
| Senha                  | A senha que você especificou ao provisionar o servidor.  |
| Nome do banco de dados | O nome do banco de dados ao qual você deseja se conectar.  |
| Grupo de Servidores    | Se você tiver muitos servidores, poderá criar grupos para ajudar a categorizá-los. Esses grupos são para conveniência no Azure Data Studio e não afetam o banco de dados ou o servidor no Azure. |



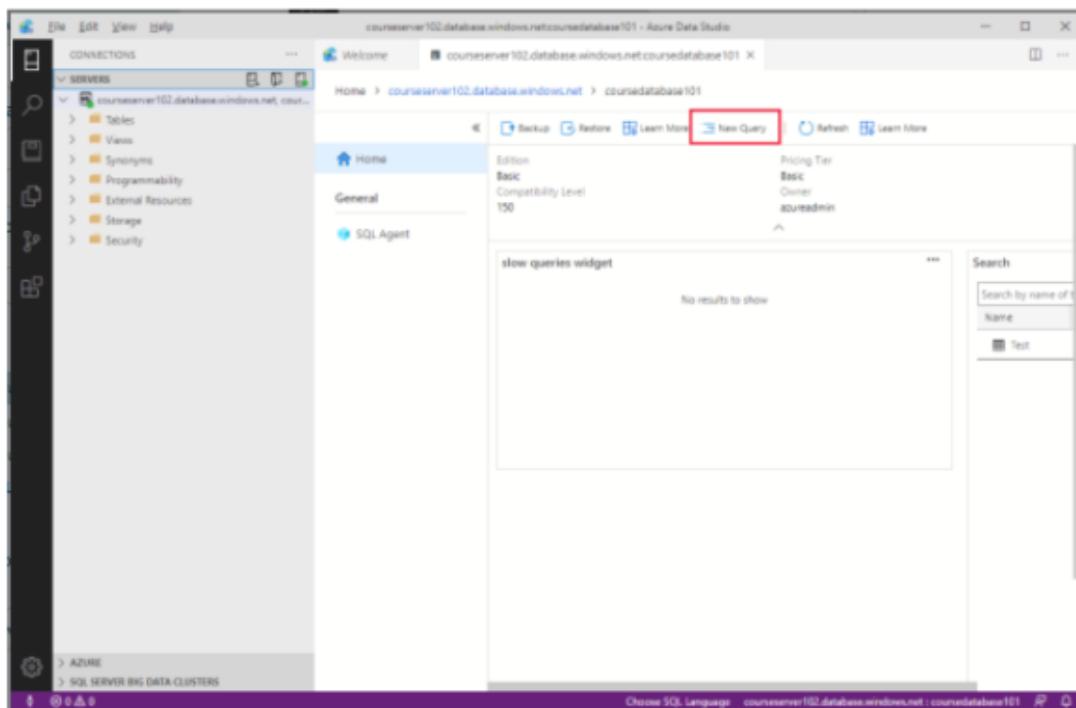
THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 000296691208

- Selezione Conectar.

Se o seu servidor não tiver uma regra de firewall permitindo a conexão do Azure Data Studio, o formulário Criar nova regra de firewall será aberto. Preencha o formulário para criar uma nova regra de firewall. Para detalhes, veja crie uma regra de firewall no nível do servidor usando o portal do Azure.

- Após conectar com sucesso, seu servidor estará disponível na barra lateral SERVIDORES em Conexões. Agora você pode usar o comando Nova Consulta para criar e executar scripts de comandos SQL.



O exemplo abaixo usa comandos Transact-SQL para criar um novo banco de dados (os comandos CREATE DATABASE e ALTER DATABASE fazem parte do dialeto Transact-SQL e não fazem parte do SQL padrão). O script então cria uma nova tabela chamada Clientes e insere quatro linhas nessa tabela. Novamente, a versão da instrução INSERT, com quatro cláusulas VALUES, faz parte do dialeto Transact-SQL. Os caracteres -- iniciam um comentário em Transact-SQL. Os caracteres [ e ] circundam os identificadores, como o nome de uma tabela, banco de dados, coluna ou tipo de dados. O caractere N na frente de uma string indica que a string usa o conjunto de caracteres Unicode.

**NOTA:** Você não pode criar novos bancos de dados SQL a partir de uma conexão no Azure Data Studio se estiver executando um único banco de dados SQL ou pools elásticos. Você só pode criar novos bancos de dados dessa maneira se estiver usando a instância gerenciada do Banco de Dados SQL.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

```
IF NOT EXISTS (
    SELECT name
    FROM sys.databases
    WHERE name = N'TutorialDB'
)
CREATE DATABASE [TutorialDB];
GO

ALTER DATABASE [TutorialDB] SET QUERY_STORE=ON;
GO

-- Switch to the TutorialDB database
USE [TutorialDB]
GO

-- Create a new table called 'Customers' in schema 'dbo'
-- Drop the table if it already exists
IF OBJECT_ID('dbo.Customers', 'U') IS NOT NULL
DROP TABLE dbo.Customers;
GO

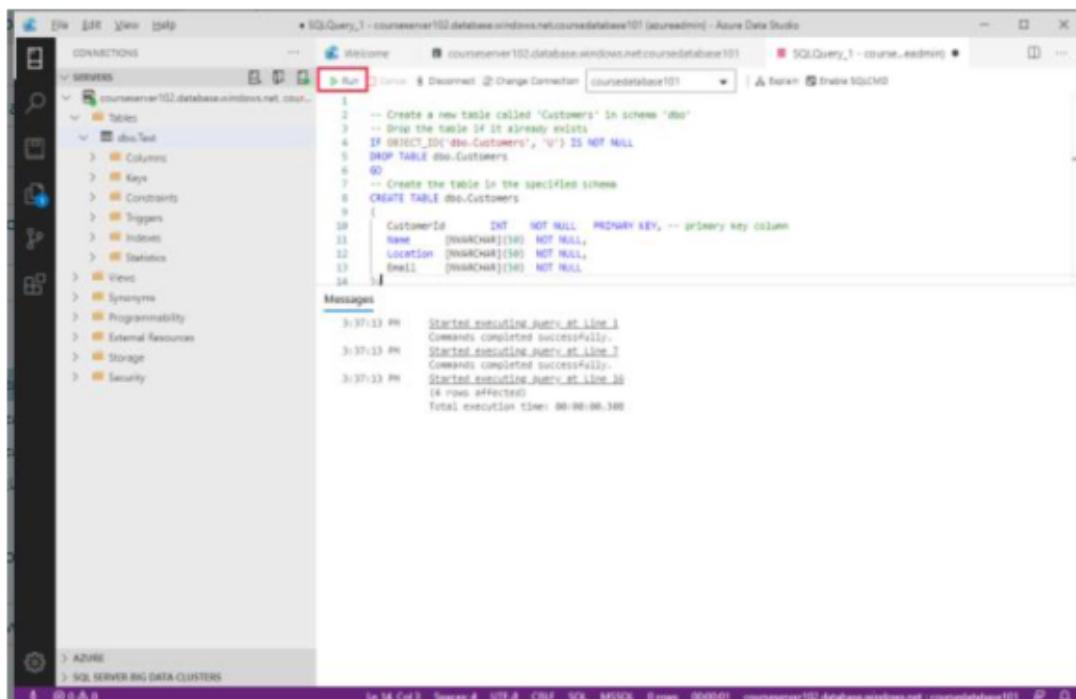
-- Create the table in the specified schema
CREATE TABLE dbo.Customers
(
    CustomerId      INT      NOT NULL  PRIMARY KEY, -- primary key column
    Name            NVARCHAR(40) NOT NULL,
    Location        NVARCHAR(40) NOT NULL,
    Email           NVARCHAR(40) NOT NULL
);
GO

-- Insert rows into table 'Customers'
INSERT INTO dbo.Customers
    ([CustomerId],[Name],[Location],[Email])
VALUES
    ( 1, N'Orlando', N'Australia', N""),
    ( 2, N'Keith', N'India', N'keith0@adventure-works.com'),
    ( 3, N'Donna', N'Germany', N'donna0@adventure-works.com'),
    ( 4, N'Janet', N'United States', N'janet1@adventure-works.com');
GO
```

Para executar o script, selecione Corre na barra de ferramentas. As notificações aparecem no painel MENSAGENS mostrando o andamento da consulta.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208



The screenshot shows the Azure Data Studio interface. On the left, the 'Connections' sidebar is open, showing a connection to 'courseserver102.database.windows.net'. Under the 'Tables' section, there is a 'dbsTest' database listed. In the main pane, a SQL query is being run to create a 'Customers' table:

```
-- Create a new table called 'Customers' in schema 'dbo'  
-- Drop the table if it already exists  
IF OBJECT_ID('dbo.Customers', 'U') IS NOT NULL  
DROP TABLE dbo.Customers  
GO  
-- Create the table in the specified schema  
CREATE TABLE dbo.Customers  
(  
    CustomerId INT NOT NULL PRIMARY KEY, -- primary key column  
    Name NVARCHAR(100) NOT NULL,  
    Location NVARCHAR(100) NOT NULL,  
    Email NVARCHAR(100) NOT NULL  
)  
GO
```

The 'Messages' pane at the bottom shows the execution log:

```
3:37:13 PM Started executing query at Line 1  
Commands completed successfully.  
3:37:13 PM Started executing query at Line 2  
Commands completed successfully.  
3:37:13 PM Started executing query at Line 3  
Commands completed successfully.  
3:37:13 PM Started executing query at Line 30  
14 rows affected  
Total execution time: 00:00:00.386
```

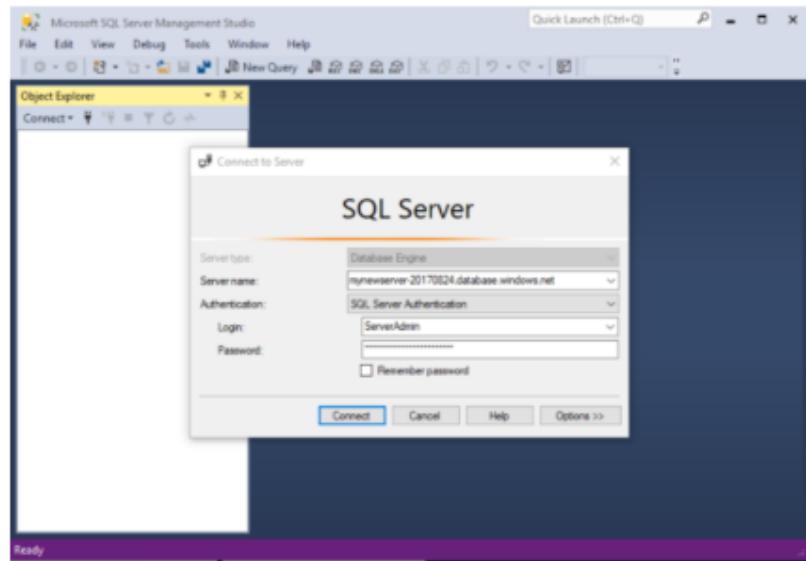
## Use o SQL Server Management Studio

O SQL Server Management Studio é outra ferramenta que você pode baixar e executar em sua área de trabalho. Veja “Baxe o SQL Server Management Studio (SSMS)” para detalhes.

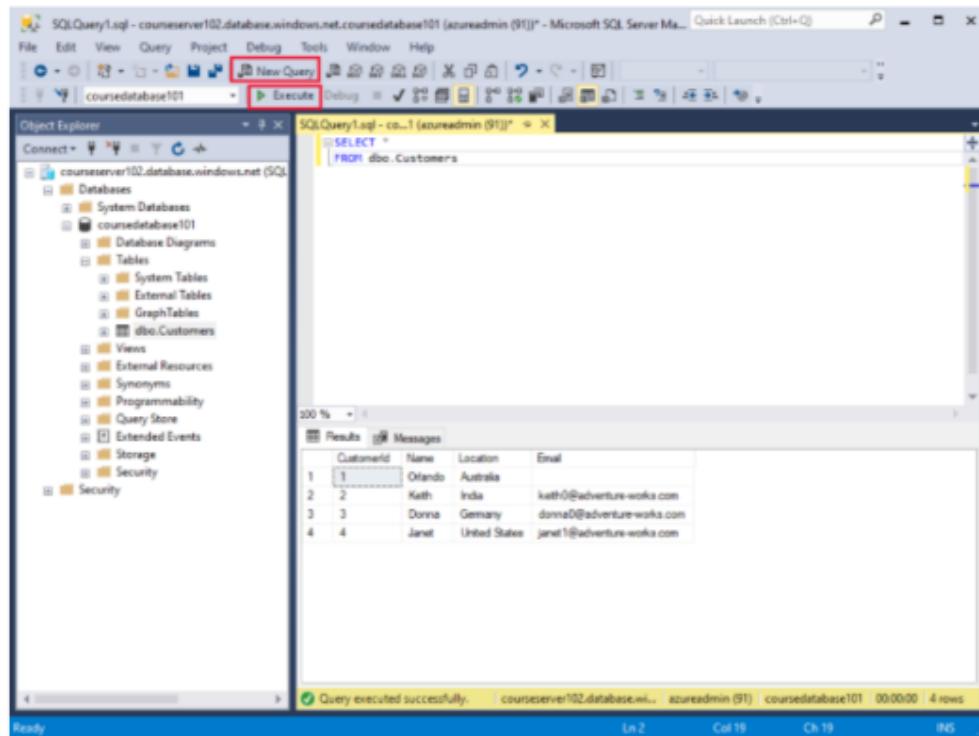
Para se conectar a um servidor e banco de dados, execute as seguintes etapas:

1. Abra o SQL Server Management Studio.
2. Quando a caixa de diálogo Conecte-se ao servidor for exibida, insira as seguintes informações:

| Configuração     | Valor   |
|------------------|---|
| Tipo de servidor | Motor de banco de dados   |
| Nome do servidor | O nome do servidor totalmente qualificado, da Visão geral página no portal do Azure |
| Autenticação     | Autenticação SQL Server   |
| Login            | O ID do usuário da conta admin do servidor usada para criar o servidor.             |
| Senha            | senha da conta de administrador do servidor   |



3. Selecione Conectar. A janela Object Explorer é aberta.
4. Para ver os objetos do banco de dados, expanda Bancos de dados e, em seguida, expanda o nó do banco de dados.
5. Na barra de ferramentas, selecione Nova Consulta para abrir uma janela de consulta.
6. Insira suas instruções SQL e selecione Executar para executar consultas e recuperar dados das tabelas do banco de dados.



THIAGO FERNANDES DA SILVA OLIVEIRA

CNPJ 000296691208

## Use as ferramentas de dados do SQL Server no Visual Studio

Visual Studio é uma ferramenta de desenvolvimento popular para construir aplicativos. Está disponível em várias edições. Você pode baixar a edição gratuita da comunidade no Visual Studio Downloads página da Microsoft.

As ferramentas de dados do SQL Server estão disponíveis no menu Ferramentas no Visual Studio. Para se conectar a uma instância existente do Banco de Dados SQL do Azure:

1. No Visual Studio, em Ferramentas, selecione servidor SQL e selecione Nova Consulta.
2. No Conectar caixa de diálogo, insira as seguintes informações e selecione Conectar:

| Cofragação             | Valor   |
|------------------------|---|
| Nome do servidor       | O nome do servidor totalmente qualificado, em <b>Visão geral</b> no portal do Azure |
| Autenticação           | Autenticação SQL Server   |
| Login                  | O ID do usuário da conta do administrador do servidor usado para criar o servidor   |
| Senha                  | Senha da conta de administrador do servidor   |
| Nome do banco de dados | Seu nome de banco de dados  |



No Inquerir janela, insira sua consulta SQL e selecione o Executar na barra de ferramentas. Os resultados aparecem no painel Resultados.

A screenshot of the Microsoft Data Studio interface. At the top, there's a menu bar with File, Edit, View, Git, Project, Build, Debug, SQL, Test, Analyze, and Help. Below the menu is a toolbar with various icons. A red box highlights the 'Execute' button (a green triangle icon) in the toolbar. The main area shows a query editor window titled 'SQLQuery1.sql' with the following code:

```
1 | SELECT *
2 | FROM dbo.Customers
```

Below the query editor is a results pane titled 'Results' which displays a table with four rows of data from the 'Customers' table:

| CustomerID | Name  | Location     | Email                      |
|------------|-------|--------------|----------------------------|
| 1          | Orbit | Australia    |                            |
| 2          | Keith | India        | keith0@adventure-works.com |
| 3          | Donna | Germany      | donna0@adventure-works.com |
| 4          | Janet | United Stat. | janet1@adventure-works.com |

## Consultar dados relacionais no Banco de Dados SQL do Azure para PostgreSQL

O PostgreSQL fornece muitas ferramentas que você pode usar para se conectar a um banco de dados PostgreSQL e executar consultas. Essas ferramentas incluem pgAdmin, interface gráfica do usuário, e o psql utilitário de linha de comando. Há um grande número de utilitários de terceiros que você também pode usar.

Nesta unidade, você verá como se conectar a um banco de dados PostgreSQL em execução no Banco de Dados do Azure para PostgreSQL a partir da linha de comando usando psql do Azure Data Studio.

### Recuperar informações de conexão para Banco de Dados do Azure para PostgreSQL

Para se conectar a um banco de dados PostgreSQL, você precisa do nome do servidor e das credenciais de uma conta que tenha direitos de acesso para se conectar ao servidor. Você pode encontrar o nome do servidor e o nome da conta de administrador padrão em Visão geralpágina do Banco de Dados do Azure para instância PostgreSQL no portal do Azure. Contate seu administrador para obter a senha.

The screenshot shows the Azure portal interface for a PostgreSQL server named 'postgresqlserver101'. The left sidebar has a search bar and navigation links: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings, Connection security, and Connection strings. The main panel displays resource group information (Resource group: change), status (Available), location (East US), subscription (change), and connection string details. A red box highlights the 'Server name' field containing 'postgresqlserver101.postgres.database.azure.com' and the 'Admin username' field containing 'azuresadmin@postgresqlserver101'. Other visible details include PostgreSQL version 10, General Purpose 4 vCore(s), 100 GB storage, and SSL enforcement status set to ENABLED.

Assim como no Banco de Dados SQL do Azure, você deve abrir o firewall PostgreSQL para permitir que os aplicativos cliente se conectem ao serviço. Para obter informações detalhadas, consulte “Regras de firewall no Banco de Dados do Azure para PostgreSQL - Servidor Único”.

## Use o psql para consultar um banco de dados

O psqlutilitário está disponível no Azure Cloud Shell. Você também pode executá-lo a partir de um prompt de comando em seu computador desktop, mas deve baixar e instalar o cliente psql. Você pode encontrar o cliente psql no [postgresql.org](http://postgresql.org) local na rede Internet.

Para se conectar ao Banco de Dados do Azure para PostgreSQL usando psql, execute as seguintes operações:

1. Execute o seguinte comando. Certifique-se de substituir o nome do servidor e o nome do administrador pelos valores do portal do Azure.

```
psql --host=<server-name>.postgres.database.azure.com --username=<admin-user>@<server-name> --dbname=postgres
```

Digite sua senha quando solicitado.

**NOTA:** postgres é o banco de dados de gerenciamento padrão criado com o Banco de Dados do Azure para PostgreSQL. Você pode criar bancos de dados adicionais usando o comando CRIAR BASE DE DADOS do psql.

2. Se sua conexão for bem-sucedida, você verá o prompt postgres =>.
3. Você pode criar um novo banco de dados com o seguinte comando SQL:

```
CREATE DATABASE "Adventureworks";
```

**NOTA:** Você pode inserir comandos em várias linhas. O caractere ponto-e-vírgula [; ] atua como o terminador do comando.

4. Dentro do psql, você pode executar o comando \c Adventureworks para se conectar ao banco de dados.

5. Você pode criar tabelas e inserir dados usando os comandos CRIAR e INSERIR, conforme mostrado nos exemplos a seguir:

```
CREATE TABLE PEOPLE(NAME TEXT NOT NULL, AGE INT NOT NULL);

INSERT INTO PEOPLE(NAME, AGE) VALUES ('Bob', 35);

INSERT INTO PEOPLE(NAME, AGE) VALUES ('Sarah', 28);

CREATE TABLE LOCATIONS(CITY TEXT NOT NULL, STATE TEXT NOT NULL);

INSERT INTO LOCATIONS(CITY, STATE) VALUES ('New York', 'NY');

INSERT INTO LOCATIONS(CITY, STATE) VALUES ('Flint', 'MI');
```

6. Você pode recuperar os dados que acabou de adicionar usando os seguintes comandos SQL:

```
SELECT * FROM PEOPLE;

SELECT * FROM LOCATIONS;
```

7. Outros comandos psql incluem:

- \l para listar bancos de dados.
- \dt para listar as tabelas no banco de dados atual.

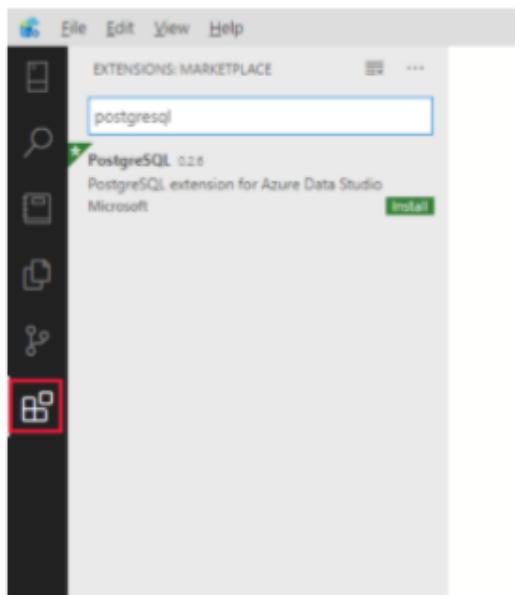
8. Você pode usar o \q comando para sair do psql.

### Conecte-se ao banco de dados PostgreSQL usando o Azure Data Studio

1. Para se conectar ao Banco de Dados do Azure para PostgreSQL a partir do Azure Data Studio, você deve primeiro instalar a extensão PostgreSQL.
2. Em Extensões, procure por postgresql.
3. Conecte-se ao banco de dados PostgreSQL usando o Azure Data Studio
4. Para se conectar ao Banco de Dados do Azure para PostgreSQL a partir do Azure Data Studio, você deve primeiro instalar a extensão PostgreSQL.
5. Em Extensões, procure por postgresql.

THIAGO FERNANDES DA SILVA OLIVEIRA

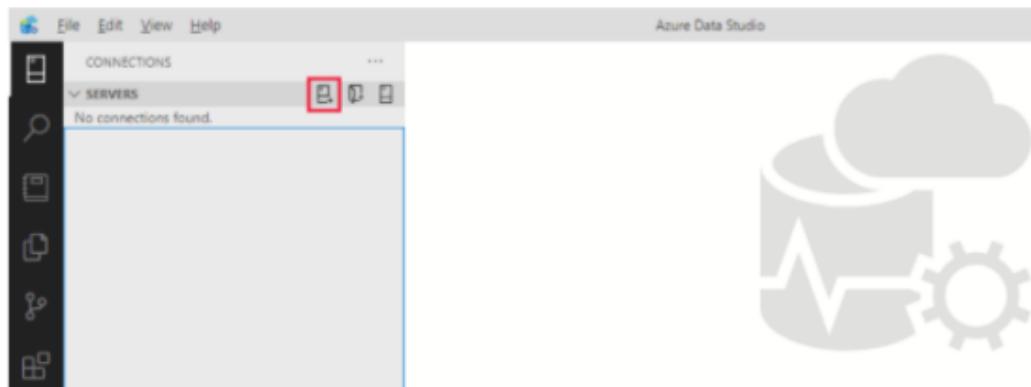
CPF 00296691208



6. Selecione os PostgreSQL extensão e selecione Instalar.

Você pode então usar a extensão para se conectar ao PostgreSQL:

1. No Azure Data Studio, vá para o SERVIDORES barra lateral e selecione Nova Conexão.



2. No Conexão caixa de diálogo, no Tipo de conexão caixa de lista suspensa, selecione PostgreSQL.
3. Preencha os campos restantes usando o nome do servidor, nome de usuário e senha do seu servidor PostgreSQL.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Connection

Recent Connections Saved Connections

No recent connection

Connection Details

Connection type: PostgreSQL

Server name: postgres/server101.postgres.database.azure.com

Authentication type: Password

User name: azureadmin@postgres/server101

Password: \*\*\*\*\*

Remember password

Database name: postgres

Server group: <Default>

Name (optional):

Advanced...

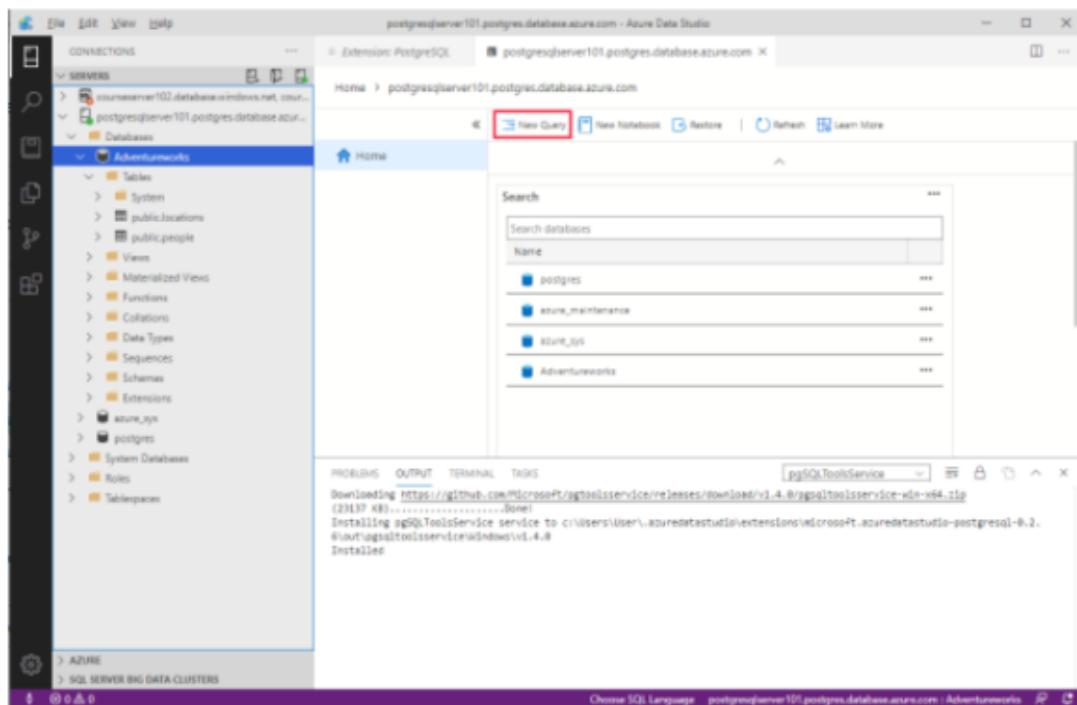
Connect Cancel



| Configuração           | Descrição  |
|------------------------|--|
| Nome do servidor       | O nome do servidor totalmente qualificado do portal do Azure.  |
| Nome do usuário        | O nome de usuário com o qual você deseja entrar. Deve estar no formato mostrado no portal do Azure, <username> @ <hostname>. |
| Senha                  | Senha da conta com a qual você está fazendo login.   |
| Nome do banco de dados | Preencha isto se desejar que a conexão especifique um banco de dados.  |
| Grupo de Servidores    | Esta opção permite atribuir essa conexão a um grupo de servidores específico criado por você.                                |
| Nome (opcional)        | Esta opção permite especificar um nome amigável para o seu servidor.   |

Selecione Conectar para estabelecer a conexão. Depois de se conectar com sucesso, seu servidor abre no SERVIDORES. Você pode expandir o Bancos de dados para se

conectar a bancos de dados no servidor e visualizar seu conteúdo. Use o Nova Consulta na barra de ferramentas para criar e executar consultas.



O exemplo a seguir adiciona uma nova tabela ao banco de dados e insere quatro linhas.

```
-- Create a new table called 'customers'  
CREATE TABLE customers(  
    customer_id SERIAL PRIMARY KEY,  
    name VARCHAR (50) NOT NULL,  
    location VARCHAR (50) NOT NULL,  
    email VARCHAR (50) NOT NULL  
);  
  
-- Insert rows into table 'customers'  
INSERT INTO customers  
    (customer_id, name, location, email)  
VALUES  
    ( 1, 'Orlando', 'Australia', "),  
    ( 2, 'Keith', 'India', 'keith0@adventure-works.com'),  
    ( 3, 'Donna', 'Germany', 'donna0@adventure-works.com'),  
    ( 4, 'Janet', 'United States','janet1@adventure-works.com');
```

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Na barra de ferramentas, selecione Correpara executar a consulta. Tal como acontece com o Azure SQL, as notificações aparecem no painel MENSAGENS para mostrar o progresso da consulta.

Para consultar os dados, insira uma instrução SELECIONAR e clique em EXECUTAR:

```
-- Select rows from table 'customers'
```

```
SELECT * FROM customers;
```

Os resultados da consulta devem aparecer no painel de resultados.

## Consultar dados relacionais no Banco de Dados SQL do Azure para MySQL

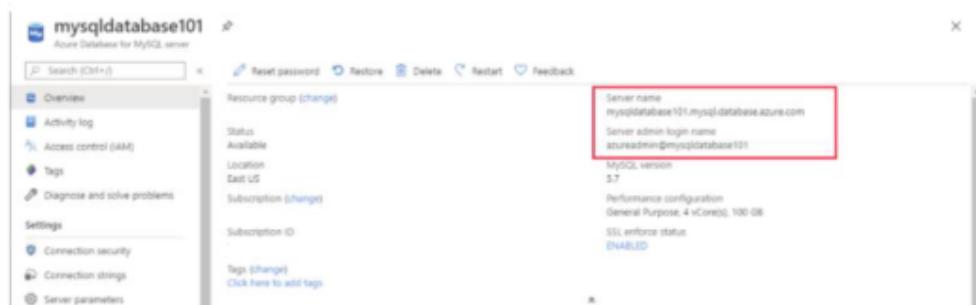
Assim como no PostgreSQL, existem muitas ferramentas disponíveis para conectar ao MySQL que permitem criar e executar scripts de comandos SQL. Você pode usar omysql utilitário de linha de comando, que também está disponível no Azure Cloud Shell, ou você pode usar ferramentas gráficas da área de trabalho, como o MySQL Workbench.

Nesta unidade, você verá como se conectar ao Banco de Dados do Azure para MySQL usando o MySQL Workbench.

**NOTA:** Atualmente, não há extensões disponíveis para conexão ao MySQL no Azure Data Studio.

## Recuperar informações de conexão para Banco de Dados do Azure para MySQL

Como o banco de dados SQL e o PostgreSQL, você precisa do nome do servidor e das credenciais de uma conta que tenha direitos de acesso para se conectar ao servidor. Você pode encontrar o nome do servidor e o nome da conta de administrador padrão na Visão geral página para a instância do Banco de Dados do Azure para MySQL, no portal do Azure. Contate seu administrador para obter a senha.



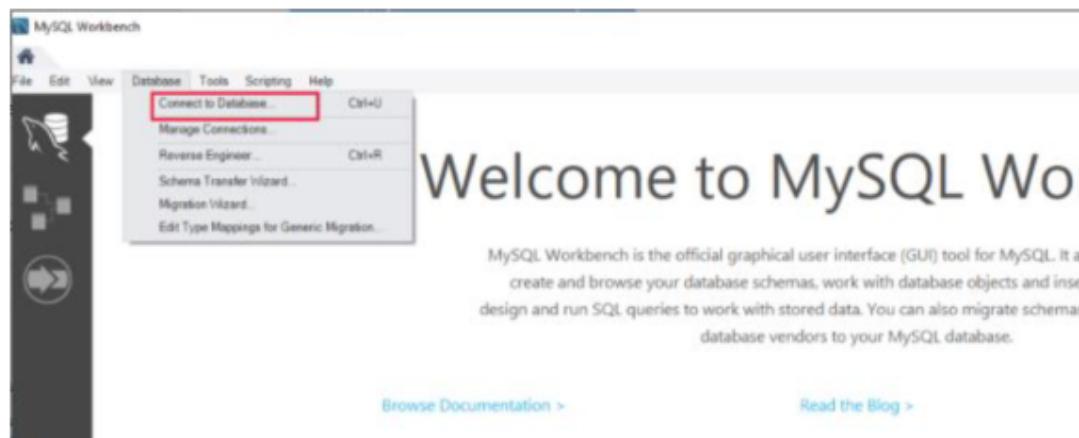
Você também deve abrir o firewall MySQL para permitir que aplicativos cliente se conectem ao serviço. Para obter informações detalhadas, consulte “Banco de dados do Azure para regras de firewall do servidor MySQL”.

## Use o MySQL Workbench para consultar um banco de dados

Você pode baixar e instalar o MySQL Workbench do Downloads da comunidade MySQL página.

Para se conectar ao Azure MySQL Server usando o MySQL Workbench, execute as seguintes etapas:

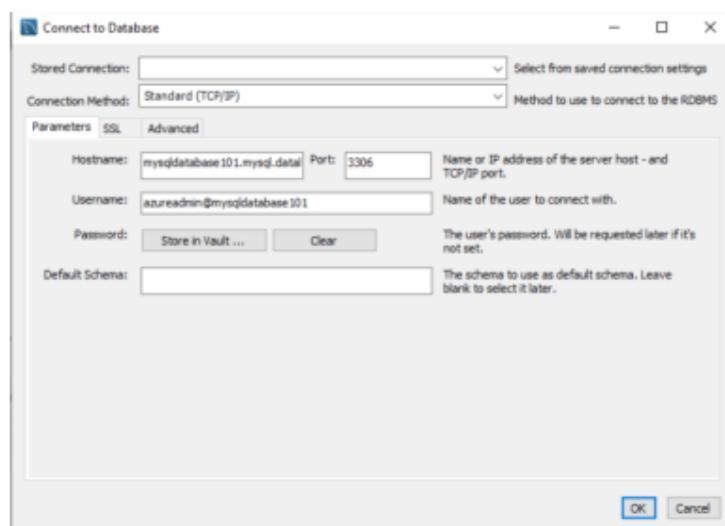
1. Inicie o MySQL Workbench em seu computador.
2. Na Página inicial, selecione Conectar ao banco de dados.



THIAGO FERNANDES DA SILVA OLIVEIRA

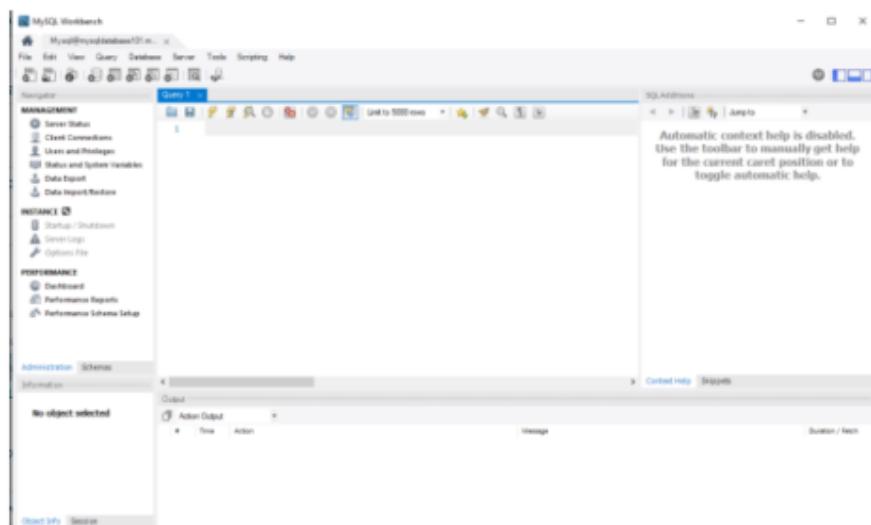
CNPJ 000296691208

1. No Conectar ao banco de dados caixa de diálogo, insira as seguintes informações na aba Parâmetros:



| Configuração       | Descrição   |
|--------------------|---|
| Conexão armazenada | Deixe em branco   |
| Método de Conexão  | Padrão (TCP / IP)   |
| Nome de anfitrião  | Especifique o nome do servidor totalmente qualificado do portal do Azure  |
| Porta              | 3306  |
| Nome do usuário    | Insira o nome de usuário de login do administrador do servidor no portal do Azure, no formato <nome de usuário><databaseusername> |
| Senha              | Selecione <b>Armazenar no Vault</b> digite a senha do administrador especificada quando o servidor foi criado                     |

Selecione OK para criar a conexão. Se a conexão for bem-sucedida, o editor de consultas será aberto.



Você pode usar este editor para criar e executar scripts de comandos SQL. O exemplo a seguir cria um banco de dados denominado quickstartdb, em seguida, adiciona uma tabela chamada inventário. Ele insere algumas linhas e, em seguida, lê as linhas. Ele altera os dados com uma instrução de atualização e lê as linhas novamente. Por fim, ele exclui uma linha e depois as lê novamente.

```
-- Create a database  
-- DROP DATABASE IF EXISTS quickstartdb;  
CREATE DATABASE quickstartdb;  
USE quickstartdb;  
  
-- Create a table and insert rows  
DROP TABLE IF EXISTS inventory;  
CREATE TABLE inventory (id serial PRIMARY KEY, name VARCHAR(50), quantity  
INTEGER);  
INSERT INTO inventory (name, quantity) VALUES ('banana', 150);  
INSERT INTO inventory (name, quantity) VALUES ('orange', 154);  
INSERT INTO inventory (name, quantity) VALUES ('apple', 100);  
  
-- Read  
SELECT * FROM inventory;  
  
-- Update  
UPDATE inventory SET quantity = 200 WHERE id = 1;  
SELECT * FROM inventory;  
  
-- Delete  
DELETE FROM inventory WHERE id = 2;  
SELECT * FROM inventory;
```

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Para executar o código SQL de amostra, selecione o ícone de raio na barra de ferramentas

The screenshot shows the MySQL Workbench interface. On the left, the Navigator pane displays categories like MANAGEMENT, INSTANCE, PERFORMANCE, and ADMINISTRATION. The central area has a toolbar with various icons. A query editor window is open, containing the following SQL code:

```
-- Update
17 • Execute the selected portion of the script or everything, if there is no selection.
18 • SELECT * FROM inventory;
19
20 -- Delete
21 • DELETE FROM inventory WHERE id = 1;
22 • SELECT * FROM inventory;
```

Below the query editor is a Result Grid showing the following data:

| ID | name   | quantity |
|----|--------|----------|
| 1  | banana | 200      |
| 2  | apple  | 300      |
| 3  | orange | 300      |

At the bottom of the interface, a Results pane displays the execution history:

| Action | Time     | Action  | Message  | Duration / Batch      |
|--------|----------|---|--|-----------------------|
| 7      | 10:44:19 | INSERT INTO `inventory`(`name`, `quantity`) VALUES ('apple', 100) | 1 rows affected  | 0.094 sec             |
| 8      | 10:44:19 | SELECT * FROM `inventory` LIMIT 0, 5000                           | 3 rows returned  | 0.098 sec / 0.098 sec |
| 9      | 10:44:19 | UPDATE `inventory` SET `quantity` = 200 WHERE `id` = 1            | 1 rows affected Rows matched: 1 Changed: 1 Warnings: 0 | 0.094 sec             |
| 10     | 10:44:19 | SELECT * FROM `inventory` LIMIT 0, 5000                           | 3 rows returned  | 0.090 sec / 0.090 sec |
| 11     | 10:44:19 | DELETE FROM `inventory` WHERE `id` = 2                            | 1 rows affected  | 0.078 sec             |
| 12     | 10:44:19 | SELECT * FROM `inventory` LIMIT 0, 5000                           | 3 rows returned  | 0.094 sec / 0.094 sec |

Os resultados da consulta aparecem na Grade de resultados, no meio da página. A lista Resultado na parte inferior da página mostra o status de cada comando conforme ele é executado.

## Laboratório: Use SQL para consultar o Banco de Dados SQL do Azure

A Contoso provisionou o banco de dados SQL e importou todos os dados de inventário para o armazenamento de dados. Como desenvolvedor líder, você foi solicitado a executar algumas consultas sobre os dados.

Vou ao Use SQL para consultar o Banco de Dados SQL do Azure no Microsoft Learn e siga as instruções para consultar o banco de dados para descobrir quantos produtos estão no banco de dados e o número de itens em estoque para um determinado produto.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## Resumo

Nesta lição, você aprendeu como usar o SQL para armazenar e recuperar dados no Banco de Dados SQL do Azure, Banco de Dados do Azure para PostgreSQL e Banco de Dados do Azure para MySQL. Você viu como se conectar a esses sistemas de gerenciamento de banco de dados usando algumas das ferramentas comuns disponíveis atualmente.

## Saber mais

- Utilitário sqlcmd
- Baixe e instale o Azure Data Studio
- Baixe o SQL Server Management Studio (SSMS)
- Tutorial: projetar um banco de dados relacional em um único banco de dados no Azure SQL usando SSMS
- Downloads da comunidade MySQL
- Banco de dados do Azure para MySQL: use o MySQL Workbench para conectar e consultar dados
- Início rápido: use o editor de consulta do portal do Azure para consultar um banco de dados
- Consultas DML com SQL
- Associações (SQL Server)

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Módulo 3: Explore dados não relacionais no Azure

### Introdução

Os dados vêm em todas as formas e tamanhos e podem ser usados para uma grande variedade de finalidades. Muitas organizações usam bancos de dados relacionais para armazenar esses dados. No entanto, o modelo relacional pode não ser o esquema mais apropriado. A estrutura dos dados pode ser muito variada para ser facilmente modelada como um conjunto de tabelas relacionais. Por exemplo, os dados podem conter itens como vídeo, áudio, imagens, informações temporais, grandes volumes de texto livre, informações criptografadas ou outros tipos de dados que não são inherentemente relacionais. Além disso, os requisitos de processamento de dados podem não ser mais adequados ao tentar converter esses dados no formato relacional. Nessas situações, pode ser melhor usar repositórios não relacionais que podem armazenar dados em seu formato original, mas que permitem o armazenamento rápido e o acesso de recuperação a esses dados.

Suponha que você seja um engenheiro de dados trabalhando na Contoso, uma organização com uma grande operação de manufatura. A organização deve reunir e armazenar informações de uma variedade de fontes, como dados em tempo real de monitoramento do status das máquinas da linha de produção, dados de controle de qualidade do produto, registros históricos de produção, volumes de produtos em estoque e dados de estoque de matérias-primas. Essas informações são críticas para o funcionamento da organização. Você foi solicitado a determinar a melhor forma de armazenar essas informações, para que possam ser armazenadas rapidamente e consultadas facilmente.

### Objetivos de aprendizado

Nesta lição, você irá:

- Explorar casos de uso e benefícios de gerenciamento de usar o armazenamento de tabelas do Azure
- Explorar casos de uso e benefícios de gerenciamento do armazenamento de Blob do Azure
- Explorar casos de uso e benefícios de gerenciamento de usar o armazenamento de arquivos do Azure
- Explorar casos de uso e benefícios de gerenciamento do uso do Azure Cosmos DB

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Explorando o armazenamento da Tabela Azure

O Armazenamento de Tabela do Azure implementa o modelo de valor-chave NoSQL. Nesse modelo, os dados de um item são armazenados como um conjunto de campos e o item é identificado por uma chave exclusiva.

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

### O que é Armazenamento de Tabela do Azure?

O Azure Table Storage é um armazenamento de valor-chave escalonável mantido na nuvem. Você cria uma tabela usando uma conta de armazenamento do Azure.

Em uma tabela de Armazenamento de Tabela do Azure, os itens são chamados de filas, e os campos são conhecidos como colunas. No entanto, não deixe que essa terminologia o confunda pensando que uma tabela do Armazenamento de Tabela do Azure é como uma tabela em um banco de dados relacional. Uma tabela do Azure permite que você armazene dados semiestruturados. Todas as linhas de uma tabela devem ter uma chave, mas, fora isso, as colunas em cada linha podem variar.

Ao contrário dos bancos de dados relacionais tradicionais, as tabelas do Azure Table Storage não têm conceito de relacionamentos, procedimentos armazenados, índices secundários ou chaves estrangeiras. Os dados geralmente serão desnormalizados, com cada linha contendo todos os dados de uma entidade lógica.

Por exemplo, uma tabela contendo informações do cliente pode armazenar o nome, sobrenome, um ou mais números de telefone e um ou mais endereços para cada cliente. O número de campos em cada linha pode ser diferente, dependendo do número de números de telefone e endereços de cada cliente, e os detalhes registrados para cada endereço.

Em um banco de dados relacional, essas informações seriam divididas em várias linhas em várias tabelas. Neste exemplo, o uso do Armazenamento de Tabela do Azure fornece acesso muito mais rápido aos detalhes de um cliente porque os dados estão disponíveis em uma única linha, sem exigir que você execute associações entre relacionamentos.

| Key<br>(Customer ID) | Value<br>(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 |  |

Para ajudar a garantir o acesso rápido, o Armazenamento de Tabela do Azure divide uma tabela em partições. O particionamento é um mecanismo para agrupar linhas relacionadas, com base em uma propriedade comum ou chave de partição. As linhas que compartilham a mesma chave de partição serão armazenadas juntas. O particionamento não apenas ajuda a organizar os dados, mas também pode melhorar a escalabilidade e o desempenho:

- As partições são independentes umas das outras e podem aumentar ou diminuir à medida que as linhas são adicionadas ou removidas de uma partição. Uma tabela pode conter qualquer número de partições.
- Ao pesquisar dados, você pode incluir a chave de partição nos critérios de pesquisa. Isso ajuda a reduzir o volume de dados a serem examinados e melhora o desempenho, reduzindo a quantidade de E / S (leituras e gravações) necessária para localizar os dados.

A chave em uma tabela do Azure Table Storage compreende dois elementos; a chave de partição que identifica a partição que contém a linha (conforme descrito acima) e uma chave de linha que é exclusiva para cada linha na mesma partição. Os itens na mesma partição são armazenados em ordem de chave de linha. Se um aplicativo adicionar uma nova linha a uma tabela, o Azure garante que a linha seja colocada na posição correta na tabela. No exemplo abaixo, tirado de um cenário de IoT, a chave de linha é um valor de data e hora.

| 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 | ... |
|               | ...         | ...         | ... |
|               | ...         | ...         | ... |
| ...           |             | ...         |     |
| 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 | ... |
|               | ...         | ...         | ... |

Este esquema permite que um aplicativo execute rapidamente Consultas de ponto que identificam uma única linha, e Consultas de alcance que buscam um bloco contíguo de linhas em uma partição.

Em uma consulta de ponto, quando um aplicativo recupera uma única linha, a chave de partição permite que o Azure aprimore rapidamente a partição correta e a chave de linha permite que o Azure identifique a linha nessa partição. Você pode ter centenas de milhões de linhas, mas se tiver definido as chaves de partição e linha com cuidado ao projetar seu aplicativo, a recuperação de dados pode ser muito rápida. A chave de partição e a chave de linha definem efetivamente um índice clusterizado sobre os dados.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Query:  
Find the data for Device 2 recorded at 13:03:23 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 ... ... |
| ...           |             | ...                 |
| 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 ... ... |
|               | ...         |                     |

Em uma consulta de intervalo, o aplicativo procura um conjunto de linhas em uma partição, especificando o ponto inicial e final do conjunto como chaves de linha. Este tipo de consulta também é muito rápido, desde que você tenha desenhado suas chaves de linha de acordo com os requisitos das consultas realizadas por sua aplicação.

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 ... ... |
| ...           |             | ...                 |
| 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 ... ... |
| ...           |             | ...                 |

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

As colunas em uma tabela podem conter dados numéricos, de string ou binários de até 64 KB. Uma tabela pode ter até 252 colunas, além das chaves de partição e linha. O tamanho máximo da linha é 1 MB. Para mais informações, leia "Compreendendo o modelo de dados do serviço Tabela".

### Casos de uso e benefícios de gerenciamento do armazenamento de tabelas do Azure

As tabelas do Azure Table Storage não têm esquemas. É fácil adaptar seus dados conforme as necessidades de seu aplicativo evoluem. Você pode usar tabelas para conter conjuntos de dados flexíveis, como dados do usuário para aplicativos da web, catálogos de endereços, informações do dispositivo ou outros tipos de metadados que seu serviço requer. O importante é escolher as chaves de partição e linha com cuidado.

As principais vantagens de usar tabelas do Azure Table Storage sobre outras formas de armazenamento de dados incluem:

- É mais simples de escalar. Leva o mesmo tempo para inserir dados em uma tabela vazia ou em uma tabela com bilhões de entradas. Uma conta de armazenamento do Azure pode conter até 500 TB de dados
- Uma tabela pode conter dados semiestruturados
- Não há necessidade de mapear e manter os relacionamentos complexos normalmente exigidos por um banco de dados relacional normalizado.
- A inserção da linha é rápida
- A recuperação de dados é rápida, se você especificar as chaves de partição e linha como critérios de consulta

No entanto, há desvantagens em armazenar dados dessa maneira, incluindo:

- 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
- É difícil filtrar e classificar dados não essenciais. Consultas que pesquisam com base em campos não-chave podem resultar em varreduras completas da tabela

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

O Azure Table Storage é um excelente mecanismo para:

- Armazenamento de TBs de dados estruturados capazes de atender a aplicativos de escala da web. Os exemplos incluem catálogos de produtos para aplicativos de comércio eletrônico e informações do cliente, onde os dados podem ser rapidamente identificados e ordenados por uma chave composta. No caso de um catálogo de produtos, a chave de partição pode ser a categoria do produto (como calçados), e a chave de linha identifica o produto específico nessa categoria (como botas de escalada).
- Armazenamento de conjuntos de dados que não requerem junções complexas, chaves estrangeiras ou procedimentos armazenados e que podem ser desnormalizados para acesso rápido. Em um sistema IoT, você pode usar o Azure Table Storage para capturar os dados do sensor do dispositivo. Cada dispositivo pode ter sua própria partição e os dados podem ser ordenados pela data e hora em que cada medição foi capturada.
- Captura de registro de eventos e dados de monitoramento de desempenho. O log de eventos e as informações de desempenho geralmente contêm dados estruturados de acordo com o tipo de evento ou medida de desempenho que está sendo registrado. Os dados podem ser particionados por evento ou tipo de medição de desempenho e ordenados pela data e hora em que foram registrados. Como alternativa, você pode partitionar os dados por data, se precisar analisar uma série ordenada de eventos e medidas de desempenho cronologicamente. Se você deseja analisar os dados por tipo e data / hora, considere armazenar os dados duas vezes, particionados por tipo e novamente por data. Gravar dados é rápido e os dados ficam estáticos depois de gravados.
- O Azure Table Storage destina-se a oferecer suporte a grandes volumes de dados, de até várias centenas de TB. Conforme você adiciona linhas a uma tabela, o Armazenamento de Tabela do Azure gerencia automaticamente as partições em uma tabela e aloca o armazenamento conforme necessário. Você não precisa realizar nenhuma etapa adicional.
- O Armazenamento de Tabela do Azure oferece garantias de alta disponibilidade em uma única região. Os dados de cada tabela são replicados três vezes em uma região do Azure. Para maior disponibilidade, mas com custo adicional, você pode criar tabelas em armazenamento com redundância geográfica. Nesse caso, os dados de cada tabela são replicados mais três vezes em outra região a várias centenas de quilômetros de distância. Se uma réplica na região local ficar indisponível, o Azure mudará de forma transparente para uma réplica funcional enquanto a réplica com falha é recuperada. Se uma região inteira for atingida por uma interrupção, suas mesas estarão seguras em uma região remota e você poderá alternar rapidamente seu aplicativo para se conectar a essa região remota.
- O Armazenamento de Tabela do Azure ajuda a proteger seus dados. Você pode configurar a segurança e o controle de acesso baseado em funções para

garantir que apenas as pessoas ou aplicativos que precisam ver seus dados possam realmente recuperá-los.

## Crie e visualize uma tabela usando o portal do Azure

A maneira mais simples de criar uma tabela no Armazenamento de Tabela do Azure é usar o portal do Azure. Siga esses passos:

1. Entre no portal do Azure usando sua conta do Azure.
2. Na página inicial do portal do Azure, selecione + Criar um recurso.

The screenshot shows the Microsoft Azure portal's main interface. At the top, there is a navigation bar with back, forward, refresh, and search icons, followed by the URL 'portal.azure.com/#home'. Below the URL is the Microsoft Azure logo and a search bar containing the placeholder 'Search resources, services, and docs (G+/-)'. The main content area is titled 'Azure services' and features several icons: a red-bordered 'Create a resource' button with a plus sign icon; 'Storage accounts' with a bar chart icon; 'Azure Cosmos DB' with a globe icon; 'Application Insights' with a lightbulb icon; and 'Resource groups' with a hexagon icon.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

3. No Novo página, selecione Conta de armazenamento - blob, arquivo, tabela, fila

The screenshot shows the Microsoft Azure portal's 'New' blade at the URL <https://portal.azure.com/#create/hub>. The 'Popular' section is displayed, featuring various Azure services with their icons and names. The 'Storage account' option is highlighted with a red box. Other visible items include Ubuntu Server 18.04 LTS, Web App, SQL Database, Function App, Azure Cosmos DB, Kubernetes Service, DevOps Starter, and Storage account.

4. No Criar conta de armazenamento página, insira os seguintes detalhes e selecione Rever + criar.

| Campo                          | Valor   |
|--------------------------------|---|
| Inscrição                      | Selecionar sua assinatura do Azure  |
| Grupo de recursos              | Selecionar Crie um novo e especifique o nome de um novo grupo de recursos do Azure. Use um nome de sua escolha, como grupo de mistérios |
| Nome da conta de armazenamento | Insira um nome de sua escolha para a conta de armazenamento. O nome deve ser único embora   |
| Localização                    | Selecionar seu local mais próximo   |
| Atuação                        | Padrão  |
| Tipo de conta                  | StorageV2 (uso geral v2)  |
| Replicação                     | rmazenamento georredundante de acesso de leitura (RA-GRS)   |
| Camada de acesso               | Quente  |

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

The screenshot shows the 'Create storage account' wizard in the Microsoft Azure portal. The 'Basics' tab is selected. The page describes Azure Storage as a highly available, secure, durable, scalable, and redundant service. It includes options for Blob, Data Lake Storage Gen2, File, Queue, and Table storage. A note states that the cost of your storage account depends on usage and options chosen.

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

**Subscription:** [dropdown menu]  
**Resource group:** (New) mystoragegroup  
Create new

**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:** myuniqueaccountname  
**Location:** (US) Central US  
**Performance:** Standard (selected)  
**Account kind:** StorageV2 (general purpose v2)  
**Replication:** Read-access geo-redundant storage (RA-GRS)  
**Access tier (default):** Cool (selected)

**Review + create** (button highlighted with a red box)

5. Na página de validação, clique em Crie e aguarde enquanto a nova conta de armazenamento é configurada.
6. Quando o Sua implantação aparecer, selecione Vá para o recurso.

The screenshot shows the 'Overview' page for a Microsoft Storage Account named 'Microsoft.StorageAccount-20200513113353'. The deployment status is shown as complete with a green checkmark. Deployment details include:  
Deployment name: Microsoft.StorageAccount-20200513113353  
Subscription: Freebie  
Resource group: mystoragegroup  
Start time: 5/13/2020, 11:44:57 AM  
Correlation ID: d639f043-261c-4951-b39a-9e1e5bcdca1

**Next steps:** Go to resource (button highlighted with a red box)

7. No Visão geral página para a nova conta de armazenamento, selecione Mesas.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

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

Home > Microsoft.StorageAccount-20200513113353 | Overview > myuniqueaccountname

myuniqueaccountname | Storage account

Search (Ctrl+)

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Data transfer

Events

Storage Explorer (preview)

Resource group (change) : mystoragegroup

Status : Primary: Available, Secondary: Available

Location : Central US, East US 2

Subscription (change) :

Subscription ID :

Tags (change) : Click here to add tags

Performance/Access tier : Standard/Hot

Replication : Read-access geo-redundant storage (RA-GRS)

Account kind : StorageV2 (general purpose v2)

Containers

File shares

Tables

Queues

8. No Mesas página, selecione + Mesa.

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

Home > Microsoft.StorageAccount-20200513113353 | Overview > myuniqueaccountname | Tables

myuniqueaccountname | Tables

Search (Ctrl+)

+ Table Refresh Delete

Check out premium Table experience with Azure Cosmos DB →

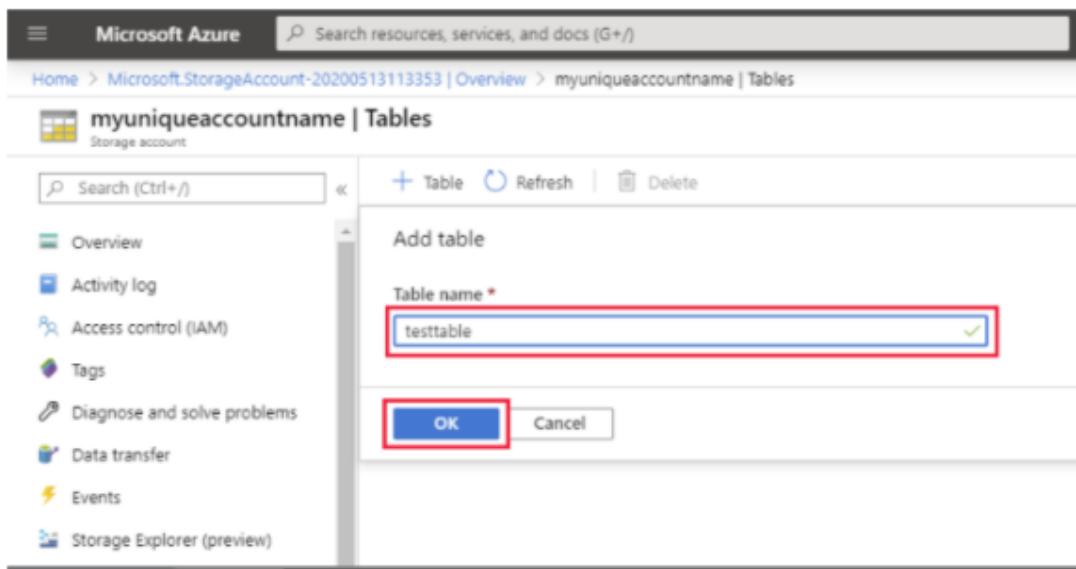
Search tables by prefix:

Table Url

You don't have any tables yet.

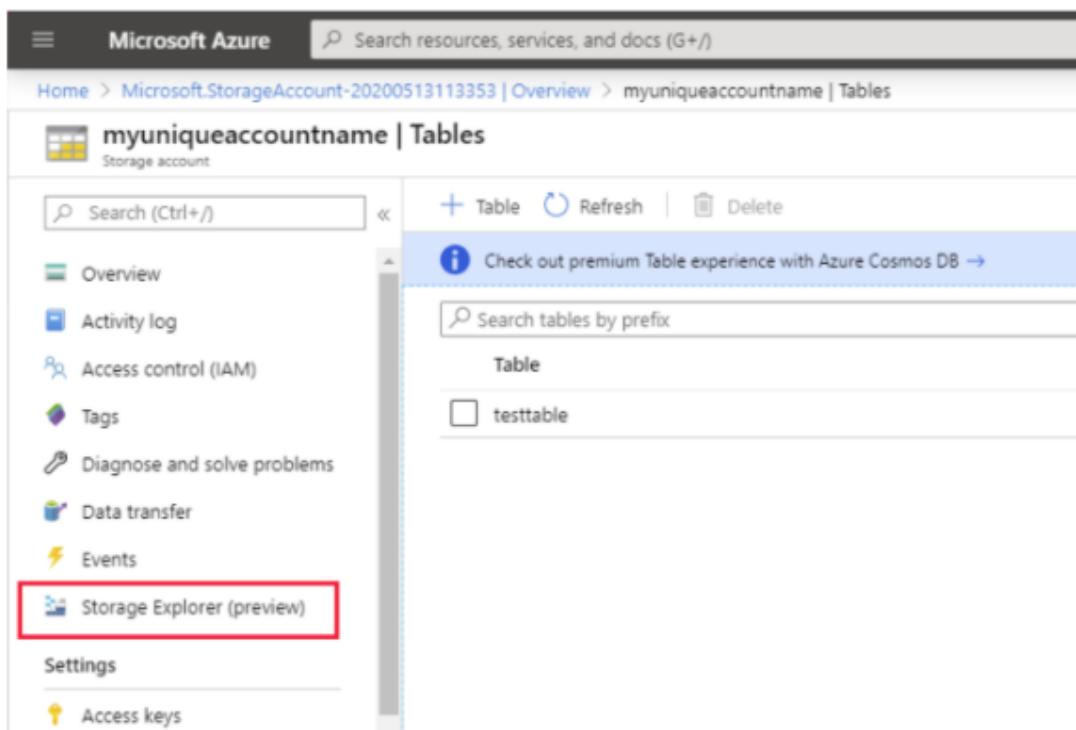
9. No Adicionar mesa caixa de diálogo, digite mesa de teste para o nome da tabela e, em seguida, selecione OK.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208



A screenshot of the Microsoft Azure Storage Account Tables creation dialog. The dialog shows a 'Table name' field containing 'testtable' with a red border around it. Below the field is an 'OK' button with a red border. On the left, there's a sidebar with various storage account management options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Data transfer, Events, and Storage Explorer (preview). The 'Storage Explorer (preview)' option is highlighted with a red box.

10. Quando a nova tabela for criada, selecione Explorador de Armazenamento.



A screenshot of the Microsoft Azure Storage Account Tables list. A message at the top says 'Check out premium Table experience with Azure Cosmos DB →'. Below is a search bar and a table with one row labeled 'testtable'. On the left, there's a sidebar with various storage account management options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Data transfer, Events, and Storage Explorer (preview). The 'Storage Explorer (preview)' option is highlighted with a red box.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

11. No Explorador de Armazenamento página, expanda Mesase selecione mesa de teste. Selecione Adicionar para inserir uma nova entidade na tabela.

**NOTA:** Dentro Explorador de Armazenamento, as linhas também são chamadas entidades.

The screenshot shows the Microsoft Azure Storage Explorer (preview) interface. On the left, there's a sidebar with various options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Data transfer, Events, and Storage Explorer (preview). The main area has a search bar at the top. Below it, there's a tree view with categories: BLOB CONTAINERS, FILE SHARES, QUEUES, and TABLES. The TABLES category is expanded, and 'testtable' is selected. A red box highlights the '+ Add' button in the toolbar above the table list. The table list itself is empty, showing 'No data available in table'. There are also buttons for Query, Edit, Select All, Column Options, and More.

12. No Adicionar Entidade caixa de diálogo, insira seus próprios valores para o PartitionKey e RowKey e, em seguida, selecione Adicionar propriedade. Adicione uma propriedade String chamada Nome e defina o valor para o seu nome. Seleccione Adicionar novamente, e adicione uma propriedade Double (isso é numérico) chamado IDADE e defina o valor para sua idade. Seleccione Inserir para salvar a entidade.

The screenshot shows the 'Add Entity' dialog box. It has a table with columns 'Property Name', 'Type', and 'Value'. The rows are:

| Property Name | Type   | Value      |
|---------------|--------|------------|
| PartitionKey  | String | PartitionA |
| RowKey        | String | Row1       |
| Name          | String | John       |
| Age           | Double | 21         |

At the bottom left of the dialog, there is a red box around the 'Add Property' button. At the bottom right, there are 'Insert' and 'Cancel' buttons, with 'Insert' also having a red box around it.

13. Verifique se a nova entidade foi criada. A entidade deve conter os valores que você especificou, junto com um carimbo de data / hora que contém a data e a hora em que a entidade foi criada.

THIAGO FERNANDES DA SILVA

CPF: 000296691208

The screenshot shows the Azure Storage Explorer interface. On the left, there's a sidebar with options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Data transfer, and Events. The main area has a search bar and a tree view under 'myuniqueaccountname' showing 'Storage account', 'BLOB CONTAINERS', 'FILE SHARES', 'QUEUES', 'TABLES', and 'testable'. To the right, there's a table with columns: PARTITIONKEY, ROWKEY, TIMESTAMP, NAME, and AGE. A single row is highlighted with a red border, showing 'PartitionK' as the PARTITIONKEY, 'Row1' as the ROWKEY, '2020-05-12T11:14:14.937968Z' as the TIMESTAMP, 'John' as the NAME, and '21' as the AGE.

Se o tempo permitir, experimente criar entidades adicionais. Nem todas as entidades devem ter as mesmas propriedades. Você pode usar o Editar para modificar os valores na entidade e adicionar ou remover propriedades. A função Query permite que você encontre entidades que possuem propriedades com um conjunto específico de valores.

## Explorar o armazenamento de Blob do Azure

Muitos aplicativos precisam armazenar grandes objetos de dados binários, como imagens e fluxos de vídeo. As máquinas virtuais do Microsoft Azure usam o armazenamento de blobs para manter imagens de disco da máquina virtual. Esses objetos podem ter várias centenas de GB de tamanho.

**NOTA:** O termo bolha é um acrônimo para Objeto Binário Grande.

## O que é armazenamento de Blob do Azure?

O armazenamento de Blob do Azure é um serviço que permite armazenar grandes quantidades de dados não estruturados ou bolhas, na nuvem. Como o armazenamento de Tabela do Azure, você cria blobs usando uma conta de armazenamento do Azure.

Atualmente, o Azure oferece suporte a três tipos diferentes de blob:

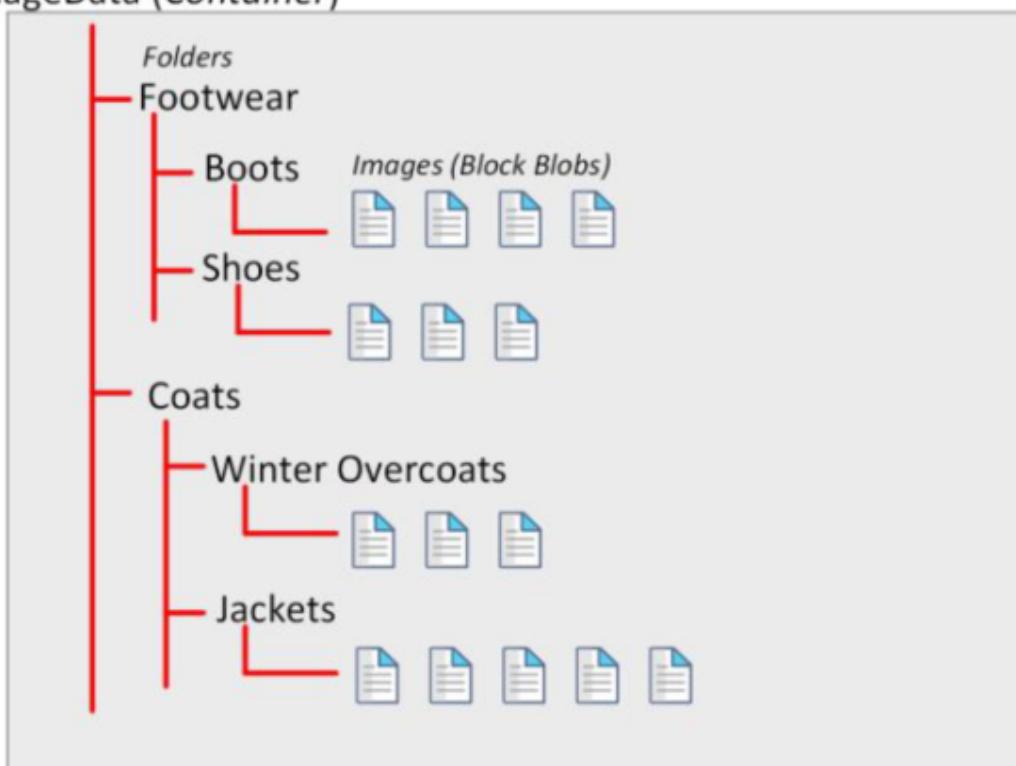
- **Blobs de blocos.** Um blob de bloco é tratado como um conjunto de blocos. Cada bloco pode variar em tamanho, até 100 MB. Um blob de bloco pode conter até 50.000 blocos, dando um tamanho máximo de mais de 4,7 TB. O bloco é a menor quantidade de dados que pode ser lida ou gravada como uma unidade individual. Blobs de bloco são mais bem usados para armazenar objetos binários grandes e discretos que mudam com pouca frequência.
- **Blobs de página.** Um blob de página é organizado como uma coleção de páginas de 512 bytes de tamanho fixo. Um blob de página é otimizado para suportar operações aleatórias de leitura e gravação; você pode buscar e armazenar dados para uma única página, se necessário. Um blob de página pode conter até 8 TB de dados. O Azure usa blobs de página para implementar o armazenamento em disco virtual para máquinas virtuais.

CIFP 100296691208

- Anexar blobs. Um blob de acréscimo é um blob de bloco otimizado para suportar operações de acréscimo. Você só pode adicionar blocos ao final de um blob de acréscimo; atualizar ou excluir blocos existentes não é suportado. Cada bloco pode variar em tamanho, até 4 MB. O tamanho máximo de um blob de acréscimo é pouco mais de 195 GB.

Dentro de uma conta de armazenamento do Azure, você cria blobs dentro contêiners. Um contêiner fornece uma maneira conveniente de agrupar blobs relacionados, e você pode organizar blobs em uma hierarquia de pastas, semelhante a arquivos em um sistema de arquivos no disco. Você controla quem pode ler e gravar blobs dentro de um contêiner no nível do contêiner.

### ImageData (Container)



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 000296691208

O armazenamento de blob fornece três níveis de acesso, que ajudam a equilibrar a latência de acesso e o custo de armazenamento:

- Hot (padrão). Você usa essa camada para blobs que são acessados com frequência. Os dados de blob são armazenados em mídia de alto desempenho.
- Cool. Essa camada tem desempenho inferior e incorre em custos de armazenamento reduzidos em comparação com a camada Hot. Use a camada Cool para dados que são acessados com pouca frequência. É comum que blobs recém-criados sejam acessados com frequência inicialmente, mas nem tanto com o passar do tempo. Nessas situações, você pode criar o blob na camada Hot, mas migrá-lo para a camada Cool posteriormente. Você pode migrar um blob da camada Cool de volta para a camada Hot.

- Arquivo (Archive). Essa camada oferece o menor custo de armazenamento, mas com latência aumentada. A camada Archive se destina a dados históricos que não devem ser perdidos, mas raramente são necessários. Blobs na camada Archive são efetivamente armazenados em um estado offline. A latência de leitura típica para as camadas Hot e Cool é de alguns milissegundos, mas para a camada Archive, pode levar horas para que os dados fiquem disponíveis. Para recuperar um blob da camada Archive, você deve alterar a camada de acesso para Hot ou Cool. A bolha então será reidratado. Você pode ler o blob somente quando o processo de reidratação for concluído.

Você pode criar políticas de gerenciamento de ciclo de vida para blobs em uma conta de armazenamento. Uma política de gerenciamento de ciclo de vida pode mover automaticamente um blob de Hot para Cool e, em seguida, para a camada Archive, à medida que envelhece e é usada com menos frequência (a política é baseada no número de dias desde a modificação). Uma política de gerenciamento de ciclo de vida também pode organizar a exclusão de blobs desatualizados.

### Casos de uso e benefícios de gerenciamento do armazenamento de Blob do Azure

Os usos comuns do Armazenamento de Blob do Azure incluem:

- Envio de imagens ou documentos diretamente para um navegador, na forma de um site estático. VisitaHospedagem de site estático no armazenamento do Azure para obter informações detalhadas.
- Armazenamento de arquivos para acesso distribuído
- Streaming de vídeo e áudio
- Armazenamento de dados para backup e restauração, recuperação de desastres e arquivamento
- Armazenamento de dados para análise por um serviço local ou hospedado pelo Azure

**NOTA:** O armazenamento de Blob do Azure também é usado como base para o armazenamento do Azure Data Lake. Você pode usar o armazenamento do Azure Data Lake para realizar análise de big data. Para mais informações visite "Introdução ao Azure Data Lake Storage Gen2".

Para garantir a disponibilidade, o armazenamento de Blob do Azure fornece redundância. Os blobs são sempre replicados três vezes na região em que você criou sua conta, mas você também pode selecionar a redundância geográfica, que replica seus dados em uma segunda região (com custo adicional).

Outros recursos disponíveis com o armazenamento de Blob do Azure incluem:

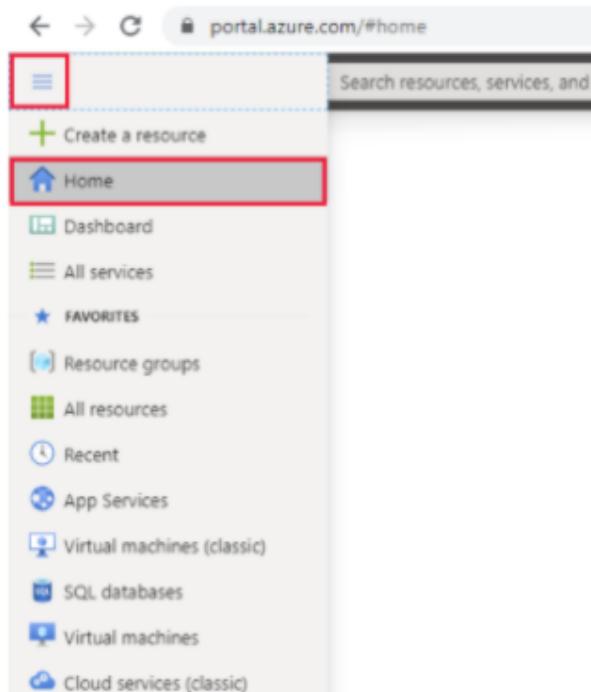
CPF: 00296691208

- Controle de versão. Você pode manter e restaurar versões anteriores de um blob.
- Exclusão suave. Este recurso permite que você recupere um blob que foi removido ou sobreescrito, por acidente ou de outra forma.
- Instantâneos. Um instantâneo é uma versão somente leitura de um blob em um determinado momento.
- Alterar feed. O feed de alterações para um blob fornece um registro ordenado, somente leitura, das atualizações feitas em um blob. Você pode usar o feed de mudança para monitorar essas mudanças e realizar operações como:
- Atualize um índice secundário, sincronize com um cache, mecanismo de pesquisa ou qualquer outro cenário de gerenciamento de conteúdo.
- Extraia insights e métricas de analítica de negócios, com base nas mudanças que ocorrem em seus objetos, seja em modo streaming ou em lote.
- Armazene, audite e analise alterações em seus objetos, em qualquer período de tempo, para segurança, conformidade ou inteligência para gerenciamento de dados corporativos.
- Crie soluções para fazer backup, espelhar ou replicar o estado do objeto em sua conta para gerenciamento de desastres ou conformidade.
- Crie pipelines de aplicativos conectados que reagem a eventos de mudança ou planeje execuções com base em objetos criados ou alterados.
- Crie e visualize um blob de bloco usando o portal do Azure

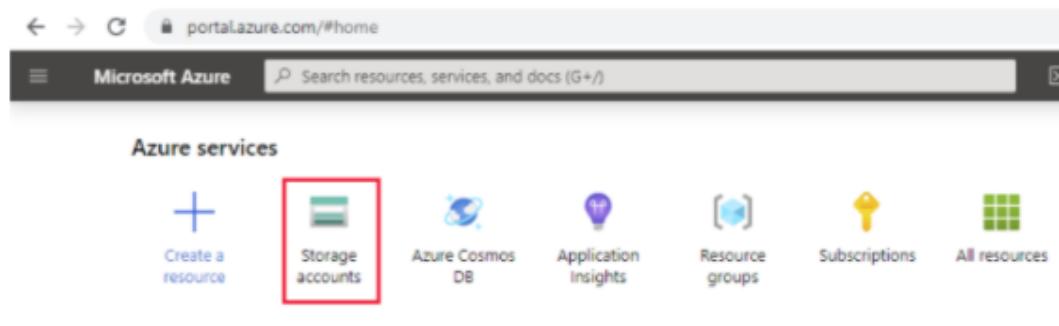
THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Você pode criar blobs de bloco usando o portal do Azure. Lembre-se de que os blobs são armazenados em contêineres e você cria um contêiner usando uma conta de armazenamento. As etapas a seguir pressupõem que você criou a conta de armazenamento descrita na unidade anterior.

1. No portal do Azure, no menu de navegação à esquerda, selecione Casa.



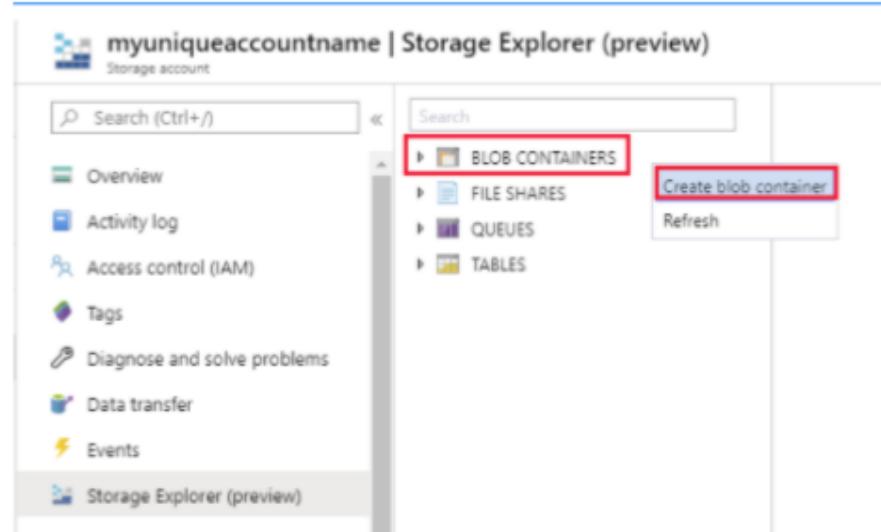
2. Na página inicial, selecione Contas de armazenamento.



3. No Contas de armazenamento página, selecione a conta de armazenamento que você criou na unidade anterior.
4. No Visão geral página para sua conta de armazenamento, selecione Explorador de Armazenamento.
5. No Explorador de Armazenamento página, clique com o botão direito CONTÊINERS DE BLOB e selecione Criar contêiner de blob.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 000296691208

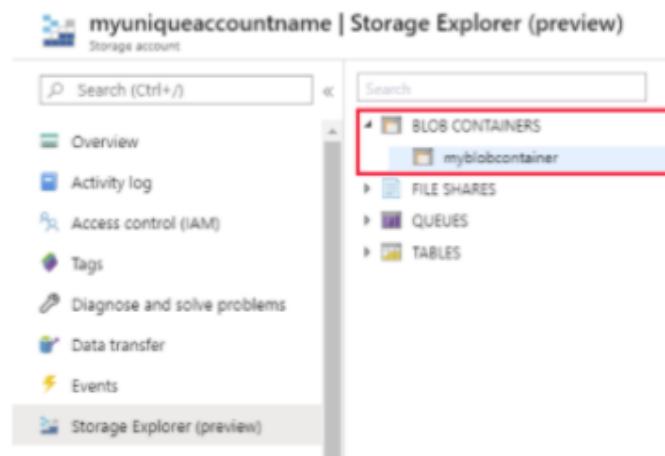


6. No Novo Container caixa de diálogo, dê um nome ao seu contêiner, aceite o nível de acesso público padrão e selecione Criar.

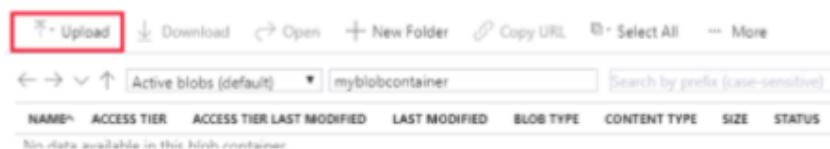


7. No Explorador de Armazenamento janela, expandir CONTÊINERS DE BLOB e selecione seu novo contêiner de blob.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208



8. Na janela de blobs, selecione Envio.



9. No Blob de upload caixa de diálogo, use o botão de arquivos para escolher um arquivo de sua escolha em seu computador e selecione Envio



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

10. Quando o upload for concluído, feche o Blob de uploadcaixa de diálogo.  
Verifique se o blob de bloco aparece em seu contêiner.

The screenshot shows the Storage Explorer interface. On the left, there's a sidebar with options: BLOB CONTAINERS, FILE SHARES, QUEUES, and TABLES. Under BLOB CONTAINERS, 'myblobcontainer' is selected. The main pane shows a list of blobs with columns: NAME, ACCESS TIER, ACCESS TIER LAST MODIFIED, LAST MODIFIED, BLOB TYPE, CONTENT TYPE, and SIZE. A single row is highlighted for 'Quiz.pdf', which has a size of 525. The status bar at the bottom indicates '13/05/2020, 12:58:07'.

Se você tiver tempo, pode experimentar fazer upload de outros arquivos como blobs de blocos. Você também pode baixar blobs de volta para o seu computador usando o botão Baixar.

## Explorar o armazenamento de arquivos do Azure

Muitos sistemas locais que compreendem uma rede de computadores internos fazem uso de compartilhamentos de arquivos. Um compartilhamento de arquivo permite armazenar um arquivo em um computador e conceder acesso a esse arquivo a usuários e aplicativos executados em outros computadores. Essa estratégia pode funcionar bem para computadores na mesma rede local, mas não se adapta bem à medida que o número de usuários aumenta ou se os usuários estão localizados em sites diferentes.

## O que é Armazenamento de Arquivos do Azure?

O Armazenamento de Arquivos do Azure permite que você crie compartilhamentos de arquivos na nuvem e acesse esses compartilhamentos de arquivos de qualquer lugar com uma conexão à Internet. O Armazenamento de Arquivos do Azure expõe compartilhamentos de arquivos usando o protocolo Server Message Block 3.0 (SMB). Este é o mesmo protocolo de compartilhamento de arquivos usado por muitos aplicativos locais existentes. Esses aplicativos devem continuar a funcionar inalterados se você migrar seus compartilhamentos de arquivos para a nuvem. Os aplicativos podem ser executados no local ou na nuvem. Você pode controlar o acesso aos compartilhamentos no Armazenamento de Arquivos do Azure usando os serviços de autenticação e autorização disponíveis por meio dos Serviços de Domínio Active Directory do Azure.

CNPJ 0007296691208

The screenshot shows the Azure Storage Account Overview page for 'jpwsstorageaccount'. The left sidebar contains navigation links like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Data transfer, Events, and Storage Explorer (preview). The main area displays account details: Resource group (changed) to 'learning', Status Primary: Available, Secondary: Available, Location UK West, UK South, Subscription (changed) to '...', Subscription ID '...', and Tags (changed) with a link to add tags. It also shows Performance/Access tier as Standard/Hot, Replication as Read-access geo-redundant storage (RA-GRS), and Account kind as StorageV2 (general purpose v2). Below this, there are four cards: 'Containers' (Massively scalable data lake storage), 'File shares' (Serverless SMB file share, highlighted with a red box), 'Tables' (Tabular data storage), and 'Queues' (Effectively scale apps according to traffic). At the bottom, it lists Tools and SDKs: Storage Explorer (preview), PowerShell, Azure CLI, .NET, Java, Python, and Node.js.

Você cria o armazenamento de Arquivos do Azure em uma conta de armazenamento. O Armazenamento de Arquivos do Azure permite que você compartilhe até 100 TB de dados em uma única conta de armazenamento. Esses dados podem ser distribuídos em qualquer número de compartilhamentos de arquivos na conta. O tamanho máximo de um único arquivo é 1 TiB, mas você pode definir cotas para limitar o tamanho de cada compartilhamento abaixo desta figura. Atualmente, o Armazenamento de Arquivos do Azure dá suporte a até 2.000 conexões simultâneas por arquivo compartilhado.

Depois de criar uma conta de armazenamento, você pode carregar arquivos no Armazenamento de Arquivos do Azure usando o portal do Azure ou ferramentas como o AzCopy. Você também pode usar o serviço de Sincronização de Arquivos do Azure para sincronizar cópias em cache local de arquivos compartilhados com os dados no Armazenamento de Arquivos do Azure.

O Armazenamento de Arquivos do Azure oferece duas camadas de desempenho. A camada Padrão usa hardware baseado em disco rígido em um datacenter, e o Prêmiocamada usa discos de estado sólido. A camada Prêmio oferece maior rendimento, mas é cobrada a uma taxa mais alta.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

### Casos de uso e benefícios de gerenciamento do armazenamento de arquivos do Azure

O Armazenamento de Arquivos do Azure foi projetado para oferecer suporte a muitos cenários, incluindo o seguinte:

- Migrar os aplicativos existentes para a nuvem. Muitos aplicativos existentes acessam dados usando APIs baseadas em arquivo e são projetados para compartilhar dados usando compartilhamentos de arquivos SMB. O Armazenamento de Arquivos do Azure permite que você migre seu arquivo local ou aplicativos baseados em compartilhamento de arquivos para o Azure sem precisar provisionar ou gerenciar máquinas virtuais de servidor de arquivos altamente disponíveis.
- Compartilhe dados do servidor localmente e na nuvem. Os clientes agora podem armazenar dados do servidor, como arquivos de log, dados de eventos e backups na nuvem para aproveitar a disponibilidade, durabilidade, escalabilidade e redundância geográfica incorporadas à plataforma de armazenamento Azure. Com a criptografia no SMB 3.0, você pode montar com segurança os compartilhamentos do Armazenamento de Arquivos do Azure de qualquer lugar. Os aplicativos em execução na nuvem podem compartilhar dados com os aplicativos locais usando as mesmas garantias de consistência implementadas por servidores SMB locais.
- Integre aplicativos modernos ao Armazenamento de Arquivos do Azure. Aproveitando a API REST moderna que o Armazenamento de Arquivos do Azure implementa além do SMB 3.0, você pode integrar aplicativos herdados com aplicativos de nuvem modernos ou desenvolver novos arquivos ou aplicativos baseados em compartilhamento de arquivos.
- Simplifique a hospedagem de dados de carga de trabalho de alta disponibilidade (HA). O Armazenamento de Arquivos do Azure oferece disponibilidade contínua para simplificar o esforço de hospedar dados de carga de trabalho HA na nuvem. Os identificadores persistentes habilitados no SMB 3.0 aumentam a disponibilidade do compartilhamento de arquivos, o que torna possível hospedar aplicativos como SQL Server e IIS no Azure com dados armazenados no armazenamento de arquivos compartilhados.

CPF: 00296691208

**NOTA:** Não use o Armazenamento de Arquivos do Azure para arquivos que podem ser gravados por vários processos simultâneos simultaneamente. Vários gravadores requerem sincronização cuidadosa, caso contrário, as alterações feitas por um processo podem ser substituídas por outro. A solução alternativa é bloquear o arquivo à medida que é escrito e, em seguida, liberar o bloqueio quando a operação de gravação for concluída. No entanto, essa abordagem pode impactar de várias formas a simultaneidade e limitar o desempenho.

O Armazenamento de Arquivos do Azure é um serviço totalmente gerenciado. Seus dados compartilhados são replicados localmente em uma região, mas também podem ser replicados geograficamente para uma segunda região.

O Azure visa fornecer até 300 MB / segundo de taxa de transferência para um único compartilhamento de arquivo Standard, mas você pode aumentar a capacidade de taxa de transferência criando um compartilhamento de arquivo Premium, por um custo adicional.

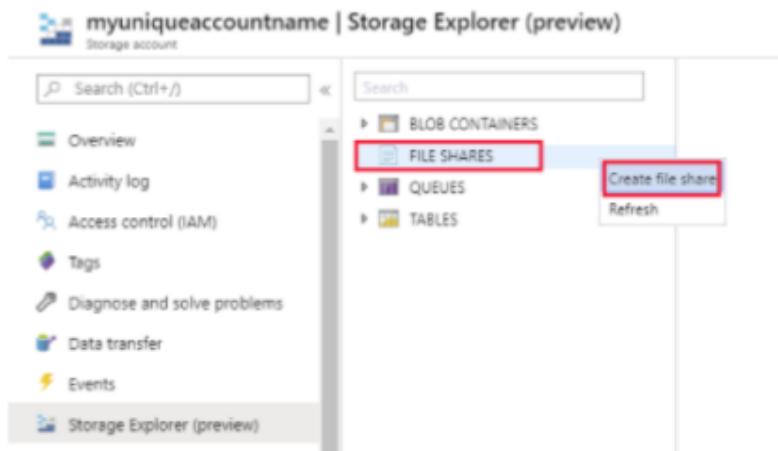
Todos os dados são criptografados em repouso e você pode habilitar a criptografia para dados em trânsito entre o Armazenamento de Arquivos do Azure e seus aplicativos.

Para obter informações adicionais sobre como gerenciar e planejar o uso do Armazenamento de Arquivos do Azure, leia Planejando uma implantação de Arquivos do Azure.

Crie um compartilhamento de arquivo de armazenamento do Azure usando o portal do Azure

Você pode criar compartilhamentos de arquivos de armazenamento do Azure usando o portal do Azure. As etapas a seguir pressupõem que você criou a conta de armazenamento descrita na unidade 2.

1. No portal do Azure, no menu, selecione Casa.
2. Na página inicial, selecione Contas de armazenamento.
3. Na Contas de armazenamento página, selecione a conta de armazenamento que você criou na unidade 2.
4. Na Visão geral página para sua conta de armazenamento, selecione Explorador de Armazenamento.
5. No Explorador de Armazenamento, clique com o botão direito AÇÕES DE ARQUIVOS e selecione Criar compartilhamento de arquivo.



6. No Novo compartilhamento de arquivo caixa de diálogo, digite um nome para o compartilhamento de arquivos e Citar e selecione Crio.

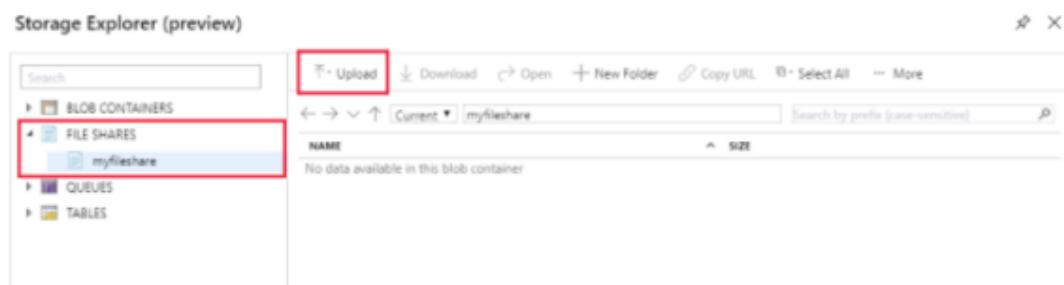


THIAGO FERNANDES DA SILVA OLIVEIRA

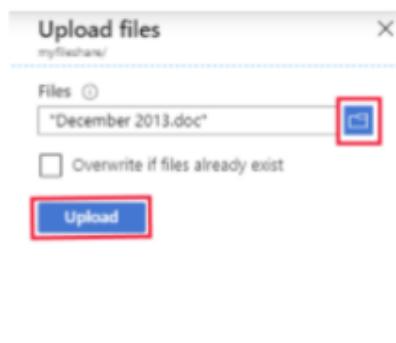
CPF 00296691208

- No Explorador de Armazenamento janela, expandir AÇÕES DE ARQUIVOS e selecione seu novo compartilhamento de arquivo e, em seguida, selecione Envio.

**DICA:** Se o seu novo compartilhamento de arquivos não aparecer, clique com o botão direito AÇÕES DE ARQUIVOSe selecione Atualizar.

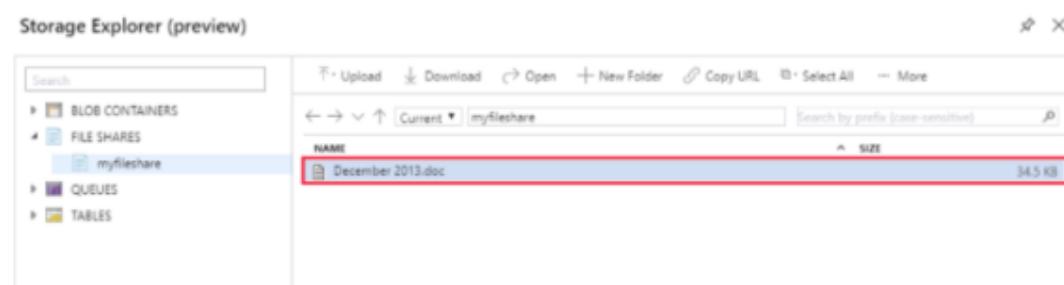


- No Fazer upload de arquivos caixa de diálogo, use o botão de arquivos para escolher um arquivo de sua escolha em seu computador e selecione Enviar



- Quando o upload for concluído, feche o Fazer upload de arquivos caixa de diálogo. Verifique se o arquivo aparece no compartilhamento de arquivos.

**DICA:** Se o arquivo não aparecer, clique com o botão direito em AÇÕES DE ARQUIVOS e selecione Atualizar.



TÍTULO: FUNDAMENTOS DA SILVA OLIVEIRA

CNPJ: 000296691208

### Explorar o Azure Cosmos DB

Tabelas, blobs e arquivos são todos tipos de armazenamento especializados, destinados a ajudar a resolver problemas específicos. Ler e gravar uma tabela é uma tarefa significativamente diferente de armazenar dados em um blob ou processar um arquivo. Às vezes, você precisa de uma solução mais generalizada, que permite armazenar e consultar dados com mais facilidade, sem ter que se preocupar com o mecanismo exato para realizar essas operações. É aqui que um sistema de gerenciamento de banco de dados se mostra útil.

Os bancos de dados relacionais armazenam dados em tabelas relacionais, mas às vezes a estrutura imposta por este modelo pode ser muito rígida e geralmente leva a um desempenho ruim, a menos que você gaste tempo implementando ajustes detalhados. Outros modelos, conhecidos coletivamente como NoSQL bancos de dados. Esses modelos armazenam dados em outras estruturas, como documentos, gráficos, armazenamentos de valores-chave e armazenamentos de famílias de colunas.

### O que é Azure Cosmos DB?

O Azure Cosmos DB é um sistema de gerenciamento de banco de dados NoSQL com vários modelos. Cosmos DB gerencia dados como um conjunto particionado de documentos. Um documento é uma coleção de campos, identificados por uma chave. Os campos em cada documento podem variar e um campo pode conter documentos filho. Muitos bancos de dados de documentos usam JSON (JavaScript Object Notation) para representar a estrutura do documento. Nesse formato, os campos em um documento são colocados entre colchetes, {e}, e cada campo é prefixado com seu nome. O exemplo abaixo mostra um par de documentos que representam as informações do cliente. Em ambos os casos, cada documento do cliente inclui documentos filho contendo o nome e endereço, mas os campos nesses documentos filhos variam entre os clientes.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

```
## Document 1 ##
{
  "customerID": "103248",
  "name":
  {
    "first": "AAA",
    "last": "BBB"
  },
  "address":
  {
    "street": "Main Street",
    "number": "101",
    "city": "Acity",
    "state": "NY"
  },
  "ccOnFile": "yes",
  "firstOrder": "02/28/2003"
}

## Document 2 ##
{
  "customerID": "103249",
  "name":
  {
    "title": "Mr",
    "forename": "AAA",
    "lastname": "BBB"
  },
  "address":
  {
    "street": "Another Street",
    "number": "202",
    "city": "Bcity",
    "county": "Gloucestershire",
    "country-region": "UK"
  },
  "ccOnFile": "yes"
}
```

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Um documento pode conter até 2 MB de dados, incluindo pequenos objetos binários. Se você precisar armazenar blobs maiores como parte de um documento, use o armazenamento de Blob do Azure e adicione uma referência ao blob no documento.

Cosmos DB fornece APIs que permitem acessar esses documentos usando um conjunto de interfaces conhecidas.

**NOTA:** API é um Interface de programação de aplicativo. Os sistemas de gerenciamento de banco de dados (e outras estruturas de software) fornecem um conjunto de APIs que os desenvolvedores podem usar para escrever programas que precisam acessar dados. As APIs geralmente serão diferentes para diferentes sistemas de gerenciamento de banco de dados.

As APIs que o Cosmos DB suporta atualmente incluem:

- API SQL. Esta interface fornece uma linguagem de consulta semelhante a SQL sobre documentos, permite identificar e recuperar documentos usando SELECIONE. O exemplo abaixo encontra o endereço do cliente 103248 nos documentos mostrados acima:

```
SELECT a.address  
FROM customers a  
WHERE a.customerID = "103248"
```

- API de tabela. Essa interface permite que você use a API do Azure Table Storage para armazenar e recuperar documentos. O objetivo desta interface é permitir que você alterne do Table Storage para o Cosmos DB sem exigir que você modifique seus aplicativos existentes.
- API MongoDB. O MongoDB é outro banco de dados de documentos bem conhecido, com sua própria interface programática. Muitas organizações executam o MongoDB localmente. Você pode usar a API MongoDB para Cosmos DB para permitir que um aplicativo MongoDB seja executado sem alterações em um banco de dados Cosmos DB. Você pode migrar os dados do banco de dados MongoDB para o Cosmos DB em execução na nuvem, mas continue executando seus aplicativos existentes para acessar esses dados.
- API Cassandra. Cassandra é um sistema de gerenciamento de banco de dados de família de colunas. Este é outro sistema de gerenciamento de banco de dados que muitas organizações executam no local. A API Cassandra para Cosmos DB fornece uma interface programática semelhante a Cassandra para Cosmos DB. As solicitações da API Cassandra são mapeadas para as solicitações de documentos do Cosmos DB. Assim como com a API do MongoDB, o objetivo principal da API do Cassandra é permitir que você migre rapidamente os bancos de dados e aplicativos do Cassandra para o Cosmos DB.

CIFP 00296691208

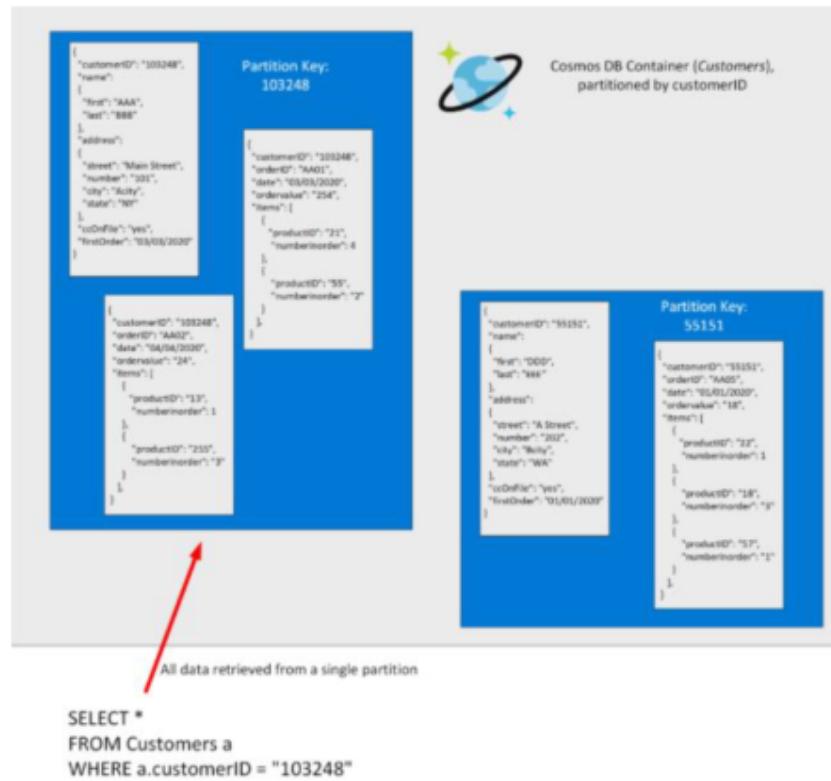
- API Gremlin. A API Gremlin implementa uma interface de banco de dados gráfico para o Cosmos DB. Um gráfico é uma coleção de objetos de dados e relacionamentos direcionados. Os dados ainda são mantidos como um conjunto de documentos no Cosmos DB, mas a API do Gremlin permite que você execute consultas de gráfico sobre os dados. Usando a API do Gremlin, você pode percorrer os objetos e relacionamentos no gráfico para descobrir todos os tipos de relacionamentos complexos, como "Qual é o nome do animal de estimação do proprietário de Sam?" no gráfico mostrado abaixo.



**NOTA:** O objetivo principal das APIs Table, MongoDB, Cassandra e Gremlin é oferecer suporte aos aplicativos existentes. Se você estiver construindo um novo aplicativo e banco de dados, deverá usar a API SQL.

Os documentos em um banco de dados Cosmos DB são organizados em contêineres. Os documentos em um contêiner são agrupados em partições. Uma partição contém um conjunto de documentos que compartilham uma chave de partição comum. Você designa um dos campos em seus documentos como a chave de partição. Você deve selecionar uma chave de partição que reúna todos os documentos relacionados juntos. Essa abordagem ajuda a reduzir a quantidade de I/O (leituras de disco) que as consultas podem precisar realizar ao recuperar um conjunto de documentos para uma determinada entidade. Por exemplo, em um banco de dados de documentos para um sistema de comércio eletrônico que registra os detalhes dos clientes e os pedidos que eles fizeram, você pode particionar os dados por ID do cliente e armazenar os detalhes do cliente e do pedido de cada cliente na mesma partição.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF: 00296691208



Há uma semelhança superficial entre um contêiner Cosmos DB e uma tabela no armazenamento de tabelas do Azure: em ambos os casos, os dados são particionados e os documentos (linhas em uma tabela) são identificados por uma ID exclusiva dentro de uma partição. No entanto, a semelhança termina aí. Ao contrário do armazenamento de Tabela do Azure, os documentos em uma partição Cosmos DB não são classificados por ID. Em vez disso, o Cosmos DB mantém um índice separado. Este índice contém não apenas os IDs dos documentos, mas também rastreia o valor de todos os outros campos em cada documento. Este índice é criado e mantido automaticamente. Este índice permite que você execute consultas que especificam critérios que fazem referência a quaisquer campos em um contêiner, sem incorrer na necessidade de verificar toda a partição para localizar esses dados. Para obter uma descrição detalhada de como a indexação do Cosmos DB funciona, leia “Indexação no Azure Cosmos DB - Visão geral”.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

### Casos de uso e benefícios de gerenciamento do uso do Azure Cosmos DB

Cosmos DB é um sistema de gerenciamento de banco de dados altamente escalonável. O Cosmos DB aloca espaço automaticamente em um contêiner para suas partições, e cada partição pode crescer até 10 GB de tamanho. Os índices são criados e mantidos automaticamente. Praticamente não há sobrecarga administrativa.

Para garantir a disponibilidade, todos os bancos de dados são replicados em uma única região. Essa replicação é transparente e o failover de uma réplica com falha é automático. Cosmos DB garante 99,99% de alta disponibilidade.

Além disso, você pode optar por replicar dados entre regiões, a um custo adicional. Esse recurso permite que você coloque cópias de dados em qualquer lugar do mundo e permita que os aplicativos se conectem à cópia dos dados que por acaso seja a mais próxima, reduzindo a latência da consulta. Todas as réplicas são sincronizadas, embora possa haver uma pequena janela enquanto as atualizações são transmitidas e aplicadas. O protocolo de replicação multimestre oferece suporte a cinco opções de consistência bem definidas - forte, envelhecimento limitado, sessão, prefixo consistente e eventual. Para mais informações, veja "Níveis de consistência no Azure Cosmos DB".

O Cosmos DB garante latências inferiores a 10 ms para leituras (indexadas) e gravações no percentil 99, em todo o mundo. Esse recurso permite a ingestão sustentada de dados e consultas rápidas para aplicativos altamente responsivos.

Cosmos DB é certificado por uma ampla gama de padrões de conformidade. Além disso, todos os dados no Cosmos DB são criptografados em repouso e em movimento. Cosmos DB fornece autorização de nível de linha e segue padrões de segurança rígidos.

Cosmos DB é um serviço básico no Azure. Cosmos DB tem sido usado por muitos dos produtos da Microsoft para aplicativos de missão crítica em escala global, incluindo Skype, Xbox, Office 365, Azure e muitos outros. Cosmos DB é altamente adequado para os seguintes cenários:

- IoT e telemática. Esses sistemas geralmente ingerem grandes quantidades de dados em frequentes picos de atividade. O Cosmos DB pode aceitar e armazenar essas informações muito rapidamente. Os dados podem ser usados por serviços de análise, como Azure Machine Learning, Azure HDInsight e Power BI. Além disso, você pode processar os dados em tempo real usando o Azure Functions que são acionados conforme os dados chegam ao banco de dados.
- Varejo e marketing. A Microsoft usa CosmosDB para suas próprias plataformas de e-commerce que funcionam como parte da Windows Store e Xbox Live. Ele também é usado no setor de varejo para armazenar dados de catálogo e para sourcing de eventos em canais de processamento de pedidos.

- Jogos. A camada de banco de dados é um componente crucial dos aplicativos de jogos. Os jogos modernos executam processamento gráfico em clientes móveis / de console, mas contam com a nuvem para fornecer conteúdo personalizado e personalizado, como estatísticas no jogo, integração de mídia social e tabelas de classificação de alta pontuação. Os jogos geralmente exigem latências de um milissegundo para leituras e gravações para fornecer uma experiência envolvente no jogo. Um banco de dados de jogos precisa ser rápido e capaz de lidar com picos massivos nas taxas de solicitação durante o lançamento de novos jogos e atualizações de recursos.
- Aplicativos da web e móveis. O Azure Cosmos DB é comumente usado em aplicativos web e móveis e é adequado para modelar interações sociais, integração com serviços de terceiros e para criar experiências personalizadas ricas. Os SDKs do Cosmos DB podem ser usados para construir aplicativos iOS e Android sofisticados usando a popular estrutura Xamarin.

Para obter informações adicionais sobre os usos do Cosmos DB, leia “Casos de uso comuns do Azure Cosmos DB”.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Teste de conhecimento

### Questão 1

**Quais são os elementos de uma chave de armazenamento Azure Table?**

- a) Nome da tabela e nome da coluna
- b) Chave de partição e chave de linha
- c) Número da linha

**Explicação:** está correto. A chave de partição identifica a partição na qual uma linha está localizada e as linhas em cada partição são armazenadas na ordem da chave de linha.

### Questão 2

**Quando você deve usar um blob de bloco e quando deve usar um blob de página?**

- a) Use um blob de bloco para dados não estruturados que requerem acesso aleatório para realizar leituras e gravações. Use um blob de página para objetos discretos que raramente mudam.
- b) Use um blob de bloco para dados ativos armazenados usando a camada de acesso de dados Hot e um blob de página para dados armazenados usando as camadas de acesso de dados Cool ou Archive.
- c) Use um bloco de página para blobs que requerem acesso aleatório de leitura e gravação. Use um blob de bloco para objetos discretos que mudam com pouca frequência.

**Explicação:** está correto. Use um bloco de página para blobs que requerem acesso aleatório de leitura e gravação. Use um blob de bloco para objetos discretos que mudam com pouca frequência.

### Questão 3

**Por que você pode usar o armazenamento de arquivos do Azure?**

- a) Para compartilhar arquivos armazenados no local com usuários localizados em outros sites.
- b) Para permitir que usuários em sites diferentes compartilhem arquivos.
- c) Para armazenar grandes arquivos de dados binários contendo imagens ou outros dados não estruturados.

**Explicação:** está correto. Você pode criar um compartilhamento de arquivos no armazenamento de arquivos do Azure, carregar arquivos para este compartilhamento de arquivos e conceder acesso ao compartilhamento de arquivos para usuários remotos.

### Questão 4

Você está construindo um sistema que monitora a temperatura em um conjunto de blocos de escritórios e define o ar-condicionado de cada cômodo em cada bloco para manter uma temperatura ambiente agradável. Seu sistema deve gerenciar o ar condicionado em vários milhares de edifícios espalhados por todo o país / região, e cada edifício normalmente contém pelo menos 100 quartos com ar condicionado. Que tipo de armazenamento de dados NoSQL é mais apropriado para capturar os dados de temperatura para permitir que sejam processados rapidamente?

- a) Envie os dados para um banco de dados do Azure Cosmos DB e use o Azure Functions para processar os dados.
- b) Armazene os dados em um arquivo armazenado em um compartilhamento criado usando o Armazenamento de Arquivos do Azure.
- c) Grave as temperaturas em um blob no armazenamento de Blob do Azure.

Explicação: está correto. O Cosmos DB pode ingerir grandes volumes de dados rapidamente. Um termômetro em cada sala pode enviar os dados para um banco de dados Cosmos DB. Você pode organizar uma Função do Azure para ser executada conforme cada item é armazenado. A função pode examinar a temperatura e iniciar um processo remoto para configurar o ar condicionado na sala.

Ka Solution

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## Resumo

O Microsoft Azure fornece uma variedade de tecnologias para armazenar dados não relacionais. Cada tecnologia tem seus próprios pontos fortes e é adequada a cenários específicos.

Nesta lição, você aprendeu sobre as tecnologias a seguir e como usá-las para atender aos requisitos de vários cenários:

- Armazenamento da Tabela Azure
- Armazenamento de Blob do Azure
- Armazenamento de arquivos do Azure
- Azure Cosmos DB

## Saber mais

- Compreendendo o modelo de dados do serviço Tabela
- Guia de design da tabela de armazenamento da Tabela Azure: tabelas escalonáveis e de desempenho
- Introdução ao armazenamento de Blob do Azure
- Introdução ao Azure Data Lake Storage Gen2
- Hospedagem de site estático no Armazenamento do Azure
- O que são arquivos do Azure?
- Planejando uma implantação de Arquivos do Azure
- Bem-vindo ao Azure Cosmos DB
- Indexação no Azure Cosmos DB - Visão geral
- Níveis de consistência no Azure Cosmos DB

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

**Explore o provisionamento e a implantação de serviços de dados não relacionais no Azure**

## Introdução

O Microsoft Azure oferece suporte a vários serviços de dados não relacionais, incluindo armazenamento de arquivos do Azure, armazenamento de Blob do Azure, Azure Data Lake Store e Azure Cosmos DB. Esses serviços oferecem suporte a diferentes tipos de dados não relacionais. Por exemplo, você pode usar o Cosmos DB para armazenar documentos e armazenamento de Blob como um repositório para grandes objetos binários, como dados de vídeo e áudio.

Antes de usar um serviço, você deve provisionar uma instância desse serviço. Você pode então configurar o serviço para permitir que você armazene e recupere dados, e para torná-los acessíveis aos usuários e aplicativos que os necessitem.

Suponha que você seja um engenheiro de dados trabalhando na Contoso, uma organização com uma grande operação de manufatura. A organização deve reunir e armazenar informações de uma variedade de fontes, como dados em tempo real de monitoramento do status das máquinas da linha de produção, dados de controle de qualidade do produto, registros históricos de produção, volumes de produtos em estoque e dados de estoque de matérias-primas. Essas informações são críticas para o funcionamento da organização. A Contoso decidiu armazenar essas informações em vários bancos de dados não relacionais, de acordo com os diferentes requisitos de processamento de dados para cada conjunto de dados. Você foi solicitado a provisionar uma variedade de serviços de dados do Azure para permitir que os aplicativos armazenem e processem as informações.

## Objetivos de aprendizado

Nesta lição, você irá:

- Aprender sobre fornecimento de serviços de dados não relacionais
- Configurar serviços de dados não relacionais
- Explorar problemas básicos de conectividade
- Explorar os componentes de segurança de dados

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Descreva o provisionamento de serviços de dados não relacionais

No cenário de exemplo, a Contoso decidiu que a organização exigirá vários armazenamentos não relacionais diferentes. Como engenheiro de dados, você é solicitado a configurar armazenamentos de dados usando o Azure Cosmos DB, armazenamento de Blob do Azure, armazenamento do Azure Data Lake e armazenamento de arquivos do Azure.

Nesta unidade, você aprenderá mais sobre o que o processo de provisionamento envolve e o que realmente acontece quando você provisiona um serviço.

### O que é provisionamento?

Provisionamento é o ato de executar uma série de tarefas que um provedor de serviços, como o Azure Cosmos DB, executa para criar e configurar um serviço. Nos bastidores, o provedor de serviços configurará os vários recursos (discos, memória, CPUs, redes e assim por diante) necessários para executar o serviço. Esses recursos serão atribuídos a você e eles permanecerão alocados a você (e cobrados de você), até que você exclua o serviço.

O modo como o provedor de serviços provisiona os recursos é obscuro e você não precisa se preocupar com o funcionamento desse processo. Tudo o que você faz é especificar parâmetros que determinam o tamanho dos recursos necessários (quanto espaço em disco, memória, capacidade de computação e largura de banda da rede). Esses parâmetros são determinados estimando o tamanho da carga de trabalho que você pretende executar usando o serviço. Em muitos casos, você pode modificar esses parâmetros após a criação do serviço, talvez aumentando a quantidade de espaço de armazenamento ou memória se a carga de trabalho for maior do que o previsto inicialmente. O ato de aumentar (ou diminuir) os recursos usados por um serviço é denominado escala.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

O vídeo a seguir resume o processo que o Azure executa quando você provisiona um serviço.

[vídeo]



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

O Azure fornece várias ferramentas que você pode usar para provisionar serviços:

- O portal do Azure. Essa é a maneira mais conveniente de fornecer um serviço para a maioria dos usuários. O portal do Azure exibe uma série de páginas específicas do serviço que solicitam as configurações necessárias e valida essas configurações antes de realmente provisionar o serviço.
- A interface de linha de comando (CLI) do Azure. A CLI fornece um conjunto de comandos que você pode executar no prompt de comando do sistema operacional ou no Cloud Shell no portal do Azure. Você pode usar esses comandos para criar e gerenciar recursos do Azure. A CLI é adequada se você precisar automatizar a criação de serviço; você pode armazenar comandos CLI em scripts e pode executar esses scripts programaticamente. A CLI pode ser executada em computadores Windows, macOS e Linux. Para obter informações detalhadas sobre a CLI do Azure, leia o que é CLI do Azure.
- Azure PowerShell. Muitos administradores estão familiarizados com o uso de comandos do PowerShell para criar scripts e automatizar tarefas administrativas. O Azure fornece uma série de commandlets (comandos específicos do Azure) que você pode usar no PowerShell para criar e gerenciar recursos do Azure. Você pode encontrar mais informações sobre o Azure PowerShell online, em Documentação do Azure PowerShell. Como a CLI, o PowerShell está disponível para Windows, macOS e Linux.
- Modelos do Azure Resource Manager. Um modelo do Azure Resource Manager descreve o serviço (ou serviços) que você deseja implantar em um

arquivo de texto, em um formato conhecido como JSON (JavaScript Object Notation). O exemplo abaixo mostra um modelo que você pode usar para provisionar uma conta de Armazenamento do Azure.

```
"resources": [
  {
    "type": "Microsoft.Storage/storageAccounts",
    "apiVersion": "2016-01-01",
    "name": "mystorageaccount",
    "location": "westus",
    "sku": {
      "name": "Standard_LRS"
    },
    "kind": "Storage",
    "properties": {}
  }
]
```

Você envia o modelo para o Azure usando o comando az group create na CLI do Azure, ou New-AzResourceGroupDeployment no Azure PowerShell. Para obter mais informações sobre como criar e usar modelos do Azure Resource Manager para provisionar recursos do Azure, consulte “O que são modelos do Azure Resource Manager?”.

## Provisionar Azure Cosmos DB

O Azure Cosmos DB é um banco de dados de documentos, adequado para uma variedade de aplicativos. No cenário de exemplo, a Contoso decidiu usar o Cosmos DB para pelo menos parte do armazenamento e processamento de dados.

No Cosmos DB, você organiza seus dados como uma coleção de documentos armazenados em contêineres. Os contêineres são mantidos em um banco de dados. Um banco de dados é executado no contexto de uma conta Cosmos DB. Você deve criar a conta antes de configurar qualquer banco de dados.

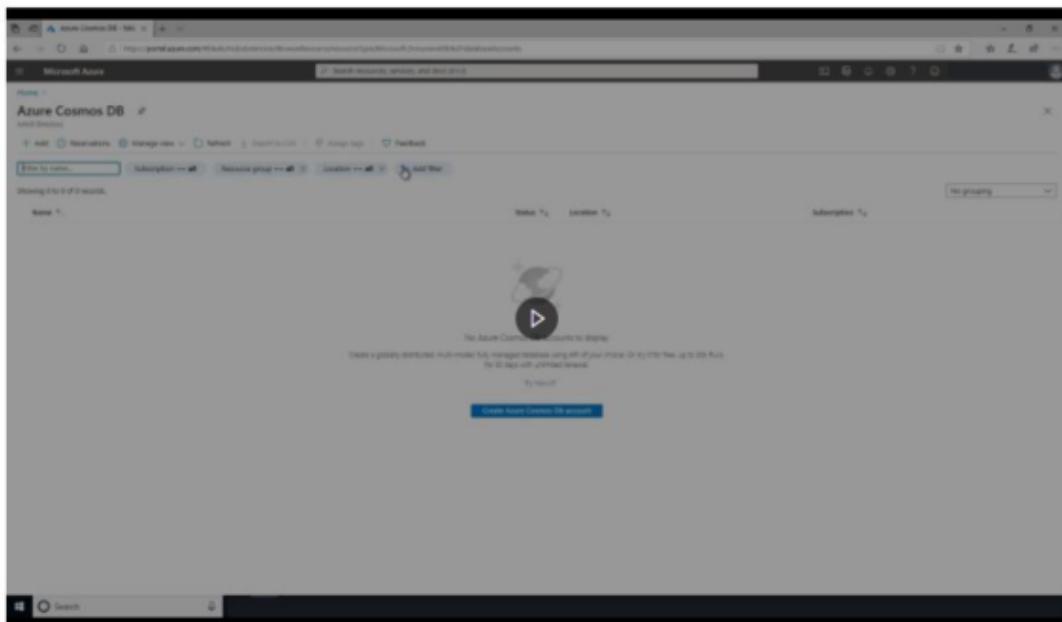
Esta unidade descreve como provisionar uma conta do Cosmos DB e, em seguida, criar um banco de dados e um contêiner nessa conta.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Como provisionar uma conta Cosmos DB

Você pode provisionar uma conta Cosmos DB de maneira interativa usando o portal do Azure ou pode executar essa tarefa programaticamente por meio da CLI do Azure, do Azure PowerShell ou de um modelo do Azure Resource Manager. O vídeo a seguir descreve como usar o portal do Azure.

[vídeo]



Se preferir usar a CLI do Azure ou o Azure PowerShell, você pode executar os comandos a seguir para criar uma conta do Cosmos DB. Os parâmetros para esses comandos correspondem a muitas das opções que você pode selecionar usando o portal do Azure. Os exemplos mostrados abaixo criam uma conta para a API Core (SQL), com redundância geográfica entre as regiões EastUS e WestUS, e suporte para gravações multirregionais. Para obter mais informações sobre esses comandos, consulte [criar az cosmosdb](#) para a CLI do Azure ou o [New-AzCosmosDBAccount](#) para PowerShell.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

```
## Azure CLI
```

```
az cosmosdb create \
--subscription <your-subscription> \
--resource-group <resource-group-name> \
--name <cosmosdb-account-name> \
--locations regionName=eastus failoverPriority=0 \
--locations regionName=westus failoverPriority=1 \
--enable-multiple-write-locations
```

```
## Azure PowerShell
```

```
New-AzCosmosDBAccount ` 
-ResourceGroupName "<resource-group-name>" ` 
-Name "<cosmosdb-account-name>" ` 
-Location @("West US", "East US") ` 
-EnableMultipleWriteLocations
```

**NOTA:** Para usar o Azure PowerShell para provisionar uma conta Cosmos DB, você deve primeiro instalar o módulo Az.CosmosDB PowerShell:

```
Install-Module -Name Az.CosmosDB
```

A outra opção de implantação é usar um modelo do Azure Resource Manager. O modelo para Cosmos DB pode ser bastante longo, devido ao número de parâmetros. Para facilitar a vida, a Microsoft publicou uma série de modelos de exemplo para lidar com diferentes configurações. Você pode baixar esses modelos no site da Microsoft, em Gerenciar recursos de API do Azure Cosmos DB Core (SQL) com modelos do Azure Resource Manager.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

### Como criar um banco de dados e um container

Uma conta do Azure Cosmos DB por si só não fornece nenhum recurso além de algumas peças de infraestrutura estática. Bancos de dados e contêineres são os principais consumidores de recursos. Os recursos são alocados em termos de espaço de armazenamento necessário para manter seus bancos de dados e contêineres e a capacidade de processamento necessária para armazenar e recuperar dados. O Azure Cosmos DB usa o conceito de unidades de solicitação por segundo (RU / s) para gerenciar o desempenho e o custo dos bancos de dados. Essa medida abstrai os recursos físicos subjacentes que precisam ser provisionados para oferecer suporte ao desempenho necessário.

Você pode pensar em uma unidade de solicitação como a quantidade de recursos de computação e E / S necessários para satisfazer uma solicitação de leitura simples feita ao banco de dados. A Microsoft fornece uma medida de aproximadamente um RU como os recursos necessários para ler um documento de 1 KB com 10 campos. Portanto, uma taxa de transferência de um RU por segundo (RU / s) dará suporte a um aplicativo que lê um único documento de 1 KB a cada segundo. Você pode especificar quantos RU / s de taxa de transferência você precisa ao criar um banco de dados ou ao criar contêineres individuais em um banco de dados. Se você especificar a taxa de transferência de um banco de dados, todos os contêineres nesse banco de dados compartilharão essa taxa de transferência. Se você especificar a taxa de transferência de um contêiner, o contêiner obterá toda a taxa de transferência para si.

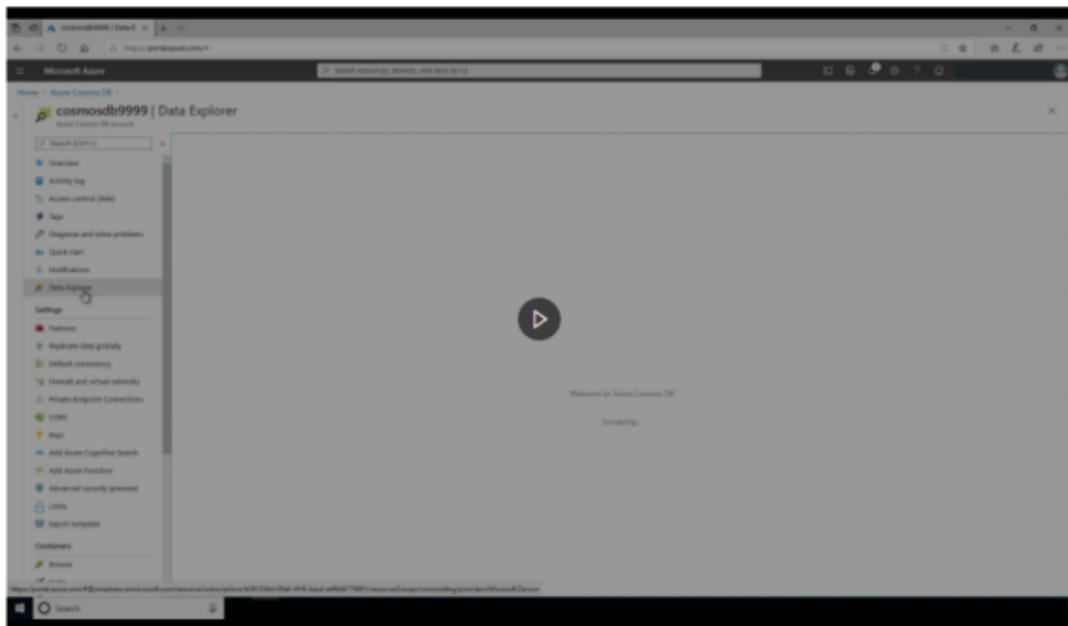
Se você subprovisionar (especificando poucos RU / s), o Cosmos DB começará a limitar o desempenho. Assim que a limitação começar, as solicitações serão solicitadas a tentar novamente mais tarde, quando houver esperança de que haja recursos disponíveis para satisfazê-la. Se um aplicativo fizer muitas tentativas de repetir uma solicitação limitada, a solicitação poderá ser cancelada. A taxa de transferência mínima que você pode alocar para um banco de dados ou contêiner é 400 RU / s. Você pode aumentar e diminuir o RU / s para um contêiner a qualquer momento. Alocar mais RU / s aumenta o custo. No entanto, depois de alocar a taxa de transferência para um banco de dados ou contêiner, você será cobrado pelos recursos provisionados, quer os use ou não.

**NOTA:** Se você aplicou o desconto de nível gratuito à sua conta do Cosmos DB, você receberá os primeiros 400 RU / s para um único banco de dados ou contêiner gratuitamente. 400 RU / s é a capacidade suficiente para a maioria dos bancos de dados pequenos a moderados.

CPF: 00296691208

O vídeo a seguir mostra como usar o portal do Azure para criar um banco de dados e um contêiner.

[vídeo]



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Se preferir usar a CLI do Azure ou Azure PowerShell, você pode executar os comandos a seguir para criar documentos e contêineres. O código abaixo mostra alguns exemplos:

```
## Azure CLI - create a database

az cosmosdb sql database create \
--account-name <cosmos-db-account-name> \
--name <database-name> \
--resource-group <resource-group-name> \
--subscription <your-subscription> \
--throughput <number-of-RU/s>

## Azure CLI - create a container

az cosmosdb sql container create \
--account-name <cosmos-db-account-name> \
--database-name <database-name> \
--name <container-name> \
--resource-group <resource-group-name> \
--partition-key-path <key-field-in-documents>
```

```
## Azure PowerShell - create a database

Set-AzCosmosDBSqlDatabase ` 
-ResourceGroupName "<resource-group-name>" ` 
-AccountName "<cosmos-db-account-name>" ` 
-Name "<database-name>" ` 
-Throughput <number-of-RU/s>

## Azure PowerShell - create a container

Set-AzCosmosDBSqlContainer ` 
-ResourceGroupName "<resource-group-name>" ` 
-AccountName "<cosmos-db-account-name>" ` 
-DatabaseName "<database-name>" ` 
-Name "<container-name>" ` 
-PartitionKeyKind Hash ` 
-PartitionKeyPath "<key-field-in-documents>"
```

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208  
CNPJ: 003296691208

## Provisionar outros serviços de dados não relacionais

Além do Cosmos DB, o Azure oferece suporte a outros serviços de dados não relacionais. Esses serviços são otimizados para casos mais específicos do que um armazenamento de banco de dados generalizado.

No cenário de exemplo, a Contoso deseja usar o armazenamento de Blob do Azure para armazenar arquivos de vídeo e áudio, o armazenamento do Azure Data Lake para dar suporte a grandes volumes de dados e o armazenamento de Arquivos do Azure para criar compartilhamentos de arquivos.

Esta unidade descreve como provisionar armazenamento Data Lake, armazenamento de Blob e Armazenamento de arquivo. Assim como no Cosmos DB, você pode provisionar esses serviços usando o portal do Azure, a CLI do Azure, o Azure PowerShell e os modelos do Azure Resource Manager. O armazenamento Data Lake, o armazenamento de Blob e o Armazenamento de arquivos exigem que você primeiro crie uma conta de armazenamento do Azure.

## Como criar uma conta de armazenamento

1. Use o portal do Azure
2. Use o Criar conta de armazenamento página para configurar uma nova conta de armazenamento usando o portal do Azure.

The screenshot shows the 'Create storage account' wizard in the Azure portal. The title bar says 'Microsoft Azure' and 'Destaque'. The URL is 'Home > New > Create storage account'. The page title is 'Create storage account'. A horizontal navigation bar at the top includes 'Basics', 'Networking', 'Advanced', 'Tags', and 'Review + create'. Below this, a descriptive text block states: 'Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below.' A link 'Learn more about Azure storage accounts' is provided. The 'Project details' section asks to select a subscription and resource group. The 'Subscription' dropdown is empty, and the 'Resource group' dropdown contains '(New) mystoragegroup' with a 'Create new' link. The 'Instance details' section includes fields for 'Storage account name' (set to 'myuniqueaccountname'), 'Location' (set to '(US) Central US'), 'Performance' (radio buttons for 'Standard' and 'Premium', with 'Standard' selected), 'Account kind' (dropdown set to 'StorageV2 (general purpose v2)'), 'Replication' (dropdown set to 'Read-access geo-redundant storage (RA-GRS)'), and 'Access tier (default)' (radio buttons for 'Cool' and 'Hot', with 'Hot' selected). At the bottom, there are 'Review + create', '< Previous', and 'Next : Networking >' buttons.

3. No Fundamentos guia, forneça os seguintes detalhes:

- Inscrição. Selecione sua assinatura do Azure.
- Grupo de Recursos. Selecione um grupo de recursos existente ou crie um novo, conforme apropriado.
- Nome da conta de armazenamento. Como acontece com uma conta Cosmos DB, cada conta de armazenamento deve ter um nome exclusivo que ainda não foi usado por outra pessoa.
- Localização. Selecione a região mais próxima de você se estiver no processo de desenvolvimento de um novo aplicativo, ou a região mais próxima de seus usuários se estiver implantando um aplicativo existente.
- Atuação. Essa configuração tem duas opções:

Padrão: contas de armazenamento são baseadas em discos rígidos. Eles são o menor custo das duas opções de armazenamento, mas têm latência mais alta. Esse tipo de conta de armazenamento é adequado para aplicativos que requerem armazenamento em massa que é acessado com pouca frequência, como arquivos.

Premium: o armazenamento usa unidades de estado sólido e tem latência muito menor e melhor desempenho de leitura / gravação do que o armazenamento padrão. As unidades de estado sólido são mais utilizadas para aplicativos de E / S intensivos, como bancos de dados. Você também pode usar o armazenamento premium para manter os discos da máquina virtual do Azure. Uma conta de armazenamento premium é mais cara do que uma conta padrão.

**NOTA:** O armazenamento Data Lake está disponível apenas com uma conta de armazenamento padrão, não premium.

- Tipo de conta. O armazenamento do Azure dá suporte a vários tipos diferentes de conta:
- V2 de uso geral. Você pode usar esse tipo de conta de armazenamento para blobs, arquivos, filas e tabelas e é recomendado para a maioria dos cenários que exigem o Armazenamento do Azure. Se você deseja provisionar o Azure Data Lake Storage, deve especificar esse tipo de conta.
- V1 de uso geral. Este é um tipo de conta herdada para blobs, arquivos, filas e tabelas. Use contas v2 de uso geral, quando possível.
- BlockBlobStorage. O tipo de conta de armazenamento está disponível apenas para contas premium. Você usa esse tipo de conta para blobs de bloco e blobs de anexo. É recomendado para cenários com altas taxas de transação ou que usam objetos menores ou exigem latência de armazenamento consistentemente baixa.
- Armazenamento de arquivo. Esse tipo também está disponível apenas para contas premium. Use-o para criar contas de armazenamento apenas de arquivos com características de desempenho premium. É recomendado para aplicativos de escala corporativa ou de alto desempenho. Use este tipo se estiver criando uma conta para oferecer suporte ao armazenamento de arquivos.

- BlobStorage. Este é outro tipo de conta herdada que só pode conter blobs. Use contas v2 de uso geral, quando possível. Você pode usar esse tipo de conta para armazenamento do Azure Data Lake, mas o tipo de conta de uso geral v2 é preferível.
- Replicação. Os dados em uma conta de Armazenamento do Azure são sempre replicados três vezes na região que você especifica como o local principal para a conta. O Armazenamento do Azure oferece duas opções de como seus dados são replicados na região primária:
- Armazenamento localmente redundante (LRS) copia seus dados de maneira síncrona três vezes em um único local físico na região. LRS é a opção de replicação menos dispendiosa, mas não é recomendada para aplicativos que requerem alta disponibilidade.
- Armazenamento georredundante (GRS) copia seus dados de maneira síncrona três vezes em um único local físico na região primária usando LRS. Em seguida, ele copia seus dados de forma assíncrona para um único local físico na região secundária. Essa forma de replicação protege você contra interrupções regionais.
- Armazenamento georredundante de acesso de leitura (RA-GRS) - a replicação é uma extensão do GRS que fornece acesso direto somente leitura aos dados no local secundário. Em contraste, a opção GRS não expõe os dados no local secundário e só é usada para se recuperar de uma falha no local primário. A replicação RA-GRS permite que você armazene uma cópia somente leitura dos dados perto de usuários localizados em um local geograficamente distante, ajudando a reduzir os tempos de latência de leitura.

**NOTA:** Para manter o desempenho, as contas de armazenamento premium suportam apenas a replicação LRS. Isso ocorre porque a replicação é realizada de forma síncrona para manter a integridade dos dados. A replicação de dados para uma região distante pode aumentar a latência a ponto de perder todas as vantagens de usar o armazenamento premium.

- Camada de acesso. Esta opção está disponível apenas para contas de armazenamento padrão. Você pode selecionar entre Hot (quente) e Cool (frio).

A camada de acesso Hot tem custos de armazenamento mais altos do que as camadas de refrigeração e arquivamento, mas os custos de acesso mais baixos. Exemplos de cenários de uso para o nível de acesso quente incluem:

Dados que estão em uso ativo ou devem ser acessados (lidos e gravados) com frequência.

Dados que são preparados para processamento e eventual migração para o nível de acesso legal.

O nível de acesso frio tem custos de armazenamento mais baixos e custos de acesso mais altos em comparação ao armazenamento quente. Esta camada é destinada a dados que permanecerão na camada fria por pelo menos 30 dias. Exemplos de cenários de uso para o nível de acesso legal incluem:

Backup de curto prazo e conjuntos de dados de recuperação de desastres.

O conteúdo de mídia mais antigo não é mais visto com frequência, mas deve estar disponível imediatamente quando acessado.

Grandes conjuntos de dados que precisam ser armazenados de maneira econômica enquanto mais dados são coletados para processamento futuro. Por exemplo, armazenamento de longo prazo de dados científicos ou dados de telemetria brutos de uma fábrica.

### Use a CLI do Azure

Se você estiver usando a CLI do Azure, execute o comando para criar uma nova conta de armazenamento para criar uma nova conta de armazenamento. O exemplo abaixo resume as opções disponíveis:

```
az storage account create \
--name <storage-account-name> \
--resource-group <resource-group> \
--location <your-location> \
--sku <sku> \
--kind <kind> \
--access-tier <tier>
```

O sku é a combinação da camada de desempenho e das opções de replicação. Pode ser Premium\_LRS, Premium\_ZRS, Standard\_GRS, Standard\_GZRS, Standard\_LRS, Standard\_RAGRS, Standard\_RAGZRS ou Standard\_ZRS.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 000296691208

**NOTA:** ZRS em alguns desses skus significa Armazenamento redundante de zona. O armazenamento com redundância de zona replica seus dados de armazenamento do Azure de forma síncrona em três zonas de disponibilidade do Azure na região primária. Cada zona de disponibilidade é um local físico separado com energia, refrigeração e rede independentes. Isso é útil para aplicativos que requerem alta disponibilidade.

O tipo O parâmetro deve ser BlobStorage, BlockBlobStorage, FileStorage, Storage ou StorageV2.

A camada de acesso parâmetro pode ser Cool ou Hot.

## Use o Azure PowerShell

Você usa o New-AzStorageAccount Cmdlet PowerShell para criar uma nova conta de armazenamento, da seguinte maneira:

```
New-AzStorageAccount  
  -Name "<storage-account-name>"  
  -ResourceGroupName "<resource-group-name>"  
  -Location "<your-location>"  
  -SkuName "<sku>"  
  -Kind "<kind>"  
  -AccessTier "<tier>"
```

Os valores para SkuName, Tipo, e AccessTier são iguais aos do comando da CLI do Azure.

## Como provisionar o armazenamento do Data Lake em uma conta de armazenamento

### Use o portal do Azure

**IMPORTANTE:** Se você estiver provisionando um armazenamento Data Lake, você deve especificar as definições de configuração apropriadas ao criar a conta de armazenamento. Não é possível configurar o armazenamento do Data Lake após a configuração da conta de armazenamento.

No portal do Azure, no Advanced em Criar conta de armazenamento página, no Data Lake Storage Gen2, selecione ativado para o Namespace hierárquico.

The screenshot shows the 'Create storage account' wizard in the Microsoft Azure portal. The 'Advanced' tab is selected. Under the 'Data Lake Storage Gen2' section, the 'Hierarchical namespace' option is set to 'Enabled'. A note indicates that this combination does not support large file shares. Other sections like 'Data protection' and 'Versioning' also have notes about unsupported combinations.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Depois que a conta de armazenamento for criada, você pode adicionar um ou mais contêineres do Data Lake Storage à conta. Cada contêiner oferece suporte a uma estrutura de diretório para armazenar arquivos do Data Lake.

The screenshot shows the Azure Storage Explorer interface. The left sidebar has 'Containers' selected under 'Data Lake Storage'. The main pane displays a table of containers, with one entry named 'mydatalakecontainer'.

| Name                | Last modified         | Public access level | Lease state |
|---------------------|-----------------------|---------------------|-------------|
| mydatalakecontainer | 5/22/2020, 3:07:50 PM | Private             | Available   |

## Use a CLI do Azure

Execute o comando `az storage account create` com o parâmetro `--enable-hierarchical-namespace` para criar uma nova conta de armazenamento compatível com Data Lake Storage:

```
az storage account create \
--name <storage-account-name> \
--resource-group <resource-group> \
--location <your-location> \
--sku <sku> \
--kind <kind> \
--access-tier <tier> \
--enable-hierarchical-namespace true
```

## Use o Azure PowerShell

Use o Cmdlet `New-AzStorageAccount` PowerShell com o parâmetro `EnableHierarchicalNamespace`, da seguinte forma:

```
New-AzStorageAccount ` 
-Name "<storage-account-name>" ` 
-ResourceGroupName "<resource-group-name>" ` 
-Location "<your-location>" ` 
-SkuName "<sku>" ` 
-Kind "<kind>" ` 
-AccessTier "<tier>" ` 
-EnableHierarchicalNamespace $True
```

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Como provisionar armazenamento de Blob em uma conta de armazenamento

### Use o portal do Azure

Blobs são armazenados em contêineres e você cria contêineres depois devocê criou uma conta de armazenamento. No portal do Azure, você pode adicionar um contêiner usando os recursos em Visão geral para sua conta de armazenamento.

The screenshot shows the Azure Storage Account Overview page for the 'contosodata' account. On the left, there's a sidebar with various navigation options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Data transfer, Events, Storage Explorer (preview), Settings, Access keys, Geo-replication, CORS, Configuration, Encryption, Shared access signature, Firewalls and virtual networks, Private endpoint connections, Advanced security, Static website, Properties, and Logos. The main content area displays account details such as Resource group (changed), Name (contosodata), Status (Primary: Available, Secondary: Available), Location (UK South, UK West), Subscription ID (b38133bd-05d0-41f6-bacd-a4ff08779001), and Tags (changed). Below these details, there are four cards: 'Containers' (Scalable, cost-effective storage for unstructured data), 'File shares' (Serverless SMB file shares), 'Tables' (Tabular data storage), and 'Queues' (Effectively scale apps according to traffic). The 'Containers' card is highlighted with a red border.

A página Contêiners permite criar e gerenciar contêineres. Cada contêiner deve ter um nome exclusivo na conta de armazenamento. Você também pode especificar o nível de acesso. Por padrão, os dados mantidos em um contêiner só podem ser acessados pelo proprietário do contêiner. Você pode definir o nível de acesso para Blob para permitir o acesso público de leitura a qualquer blobs criado no contêiner, ou Contêiner para permitir acesso de leitura a todo o conteúdo do contêiner, incluindo a capacidade de listar todos os blobs. Você também pode configurar o controle de acesso baseado em função para um blob se precisar de um nível de segurança mais granular.

TITULO: FUNDAMENTOS DA SILVA OLIVEIRA

CPF: 00296691208

The screenshot shows the Azure Storage Container creation dialog. At the top, there are buttons for '+ Container', 'Change access level', 'Refresh', and 'Delete'. Below that is a search bar labeled 'Search containers by prefix'. A message says 'You don't have any containers yet. Click '+ Container' to get started.' On the right, there's a form for creating a new container. It has a 'Name' field with 'images' entered, a 'Public access level' dropdown set to 'Blob (anonymous read access for blobs only)', and a warning message: 'Blobs within the container can be read by anonymous request, but container data is not available. Anonymous clients cannot enumerate the blobs within the container. Anonymous access bypasses Access Control List (ACL) settings.' At the bottom are 'Create' and 'Discard' buttons.

Depois de provisionar um contêiner, seus aplicativos podem carregar blobs para o contêiner.

## Use a CLI do Azure

O comando az storage container create estabelece um novo contêiner de blob em uma conta de armazenamento.

```
az storage container create \
--name <container-name> \
--account-name <storage-account-name> \
--public-access <access>
```

O parâmetro público de acesso pode ser bolha, contêiner, ou fora (apenas para acesso privado).

## Use o Azure PowerShell

Use o New-AzStorageContainer cmdlet para adicionar um contêiner a uma conta de armazenamento. Você deve primeiro recuperar um objeto de conta de

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

armazenamento com o Get-AzStorageAccount cmdlet. O código abaixo mostra um exemplo:

```
Get-AzStorageAccount `

-ResourceGroupName "<resource-group>" `

-Name "<storage-account-name>" | New-AzStorageContainer `

-Name "<container-name>" `

-Permission <permission>
```

O parâmetro Permissão aceita os valores Blob, Contêiner, ou Fora.

## Como provisionar o armazenamento de arquivos em uma conta de armazenamento

### Use o portal do Azure

Você provisiona o armazenamento de arquivos criando um ou mais compartilhamentos de arquivos na conta de armazenamento. No portal do Azure, selecione o compartilhamento de arquivos na Visão geral da conta.

The screenshot shows the Azure Storage Account Overview page for 'jpwsstorageaccount'. The 'File shares' section is highlighted with a red box. It contains the text: 'File shares Serverless SMB file shares Learn more'. Other sections visible include 'Containers' (Massively scalable data lake storage), 'Tables' (Tabular data storage), and 'Queues' (Effectively scale apps according to traffic). The left sidebar lists various storage account management options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Data transfer, Events, and Storage Explorer (preview).

THIAGO FERNANDES DA SILVA OLIVEIRA

CNPJ 000296691208

Usando o Compartilhamento de arquivos, crie um novo compartilhamento de arquivo. Dê um nome ao compartilhamento de arquivos e, opcionalmente, defina uma cota para limitar o tamanho dos arquivos no compartilhamento. O tamanho total de todos os arquivos em todos os compartilhamentos de arquivos em uma conta de armazenamento não pode exceder 5120 GB.



Depois de criar o compartilhamento de arquivos, os aplicativos podem ler e gravar arquivos compartilhados usando o compartilhamento de arquivos.

## Use a CLI do Azure

A CLI do Azure fornece o comando para criar um novo compartilhamento de arquivo em uma conta de armazenamento:

```
az storage share create \
  --name <share-name> \
  --account-name <storage-account-name>
```

## Use o Azure PowerShell

O New-AzStorageShare cmdlet cria um novo compartilhamento de arquivo em uma conta de armazenamento. Você deve recuperar os detalhes da conta de armazenamento primeiro.

```
Get-AzStorageAccount \
  -ResourceGroupName "<resource-group>" \
  -Name "<storage-account-name>" |New-AzStorageShare \
  -Name "<share-name>"
```

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF: 00296691208

## Descreva a configuração de serviços de dados não relacionais

Depois de provisionar um recurso, você geralmente precisará configurá-lo para atender às necessidades de seus aplicativos e ambiente. Por exemplo, você pode precisar configurar o acesso à rede ou abrir uma porta de firewall para permitir que seus aplicativos se conectem ao recurso.

Nesta unidade, você aprenderá como habilitar o acesso à rede para seus recursos e como evitar a exposição acidental de seus recursos a terceiros. Você verá como usar autenticação e controle de acesso para proteger os dados gerenciados por seus recursos.

### Configurar conectividade e firewalls

A conectividade padrão para o Azure Cosmos DB e o Armazenamento do Azure é permitir o acesso ao mundo em geral. Você pode se conectar a esses serviços de uma rede local, a Internet ou de uma rede virtual do Azure. Embora esse nível de acesso pareça arriscado, a maioria dos serviços do Azure atenua esse risco exigindo autenticação antes de conceder acesso. A autenticação é descrita posteriormente nesta unidade.

**NOTA:** Uma Rede Virtual do Azure é uma representação de sua própria rede na nuvem. Uma rede virtual permite que você conecte máquinas virtuais e serviços do Azure, da mesma forma que você pode usar uma rede física local. O Azure garante que cada rede virtual seja isolada de outras redes virtuais criadas por outros usuários e da Internet. O Azure permite que você especifique quais máquinas (reais e virtuais) e serviços têm permissão para acessar recursos na rede virtual e quais portas podem usar.

### Configure a conectividade para redes virtuais e computadores locais

Para restringir a conectividade, use o Firewalls e redes virtuais página de um serviço. Para limitar a conectividade, escolha Redes selecionadas. Três outras seções aparecerão, Virtual Network, Firewall, e Exceções.

Na seção Redes virtuais, você pode especificar quais redes virtuais têm permissão para rotear o tráfego para o serviço. Ao criar itens como aplicativos da web e máquinas virtuais, você pode adicioná-los a uma rede virtual. Se esses aplicativos e máquinas virtuais exigirem acesso ao seu recurso, adicione a rede virtual que contém esses itens à lista de redes permitidas.

Se você precisar se conectar ao serviço de um computador local, no Firewall seção, adicione o endereço IP do computador. Essa configuração cria uma regra de firewall que permite que o tráfego desse endereço chegue ao serviço.

CPF 000296691208

A Configuração de Exceções permite que você habilite o acesso a qualquer outro de seus serviços criados em sua assinatura do Azure.

Para informações detalhadas, leia “Configurar firewalls e redes virtuais do Armazenamento do Azure”.

A imagem abaixo mostra o Firewalls e redes virtuais página para uma conta de armazenamento do Azure. Outros serviços têm a mesma página ou página semelhante.

The screenshot shows the 'Firewalls and virtual networks' section of the Azure Storage account settings. It includes:

- Allow access from:** Selected networks (radio button selected).
- Virtual networks:** A table showing a single entry: myvnet (Subnet 1, Endpoint Status: learning, Resource Group: Freebie, Subscription: Freebie).
- Firewall:** A section for adding IP ranges, with 'Add your client IP address ('90.248.112.156')' checked.
- Address range:** An input field for 'IP address or CIDR'.
- Exceptions:** Three checkboxes:
  - Allow trusted Microsoft services to access this storage account (checked)
  - Allow read access to storage logging from any network (unchecked)
  - Allow read access to storage metrics from any network (checked)

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## Configure a conectividade de endpoints privados

Ponto final privado do Azure é uma interface de rede que conecta você de forma privada e segura a um serviço fornecido pelo Azure Private Link. Endpoint privado usa um endereço IP privado de sua VNet, trazendo efetivamente o serviço para sua VNet. O serviço pode ser um serviço do Azure, como Armazenamento do Azure, Azure Cosmos DB, SQL ou seu próprio Private Link Service. Para obter informações detalhadas, leia “O que é ponto final privado do Azure?”.

As Conexões de endpoint privadas de um serviço permitem que você especifique quais terminais privados, se houver, têm acesso permitido ao seu serviço. Você pode usar as configurações nesta página, junto com o Firewalls e redes virtuais, para bloquear completamente o acesso de usuários e aplicativos a endpoints públicos para se conectar à sua conta do Cosmos DB.

## Configurar autenticação

Muitos serviços incluem uma chave de acesso que você pode especificar ao tentar se conectar ao serviço. Se você fornecer uma chave incorreta, seu acesso será negado. A imagem abaixo mostra como encontrar a chave de acesso para uma conta de Armazenamento do Azure; você seleciona Chaves de acesso sob Configurações na página principal da conta. Muitos outros serviços permitem que você visualize a chave de acesso da mesma maneira no portal do Azure. Se sua chave estiver comprometida, você pode gerar uma nova chave de acesso.

The screenshot shows the 'Access keys' section of the Azure Storage Account settings. Key1 and Key2 are listed with their respective keys and connection strings. The 'Access keys' option in the left sidebar is highlighted with a red box.

**NOTA:** Os serviços do Azure fornecem, na verdade, duas chaves, rotuladas chave1 e chave2. Um aplicativo pode usar qualquer uma das teclas para se conectar ao serviço.

Qualquer usuário ou aplicativo que conhece a chave de acesso de um recurso pode se conectar a esse recurso. No entanto, as chaves de acesso fornecem um nível de autenticação bastante granular. Além disso, se você precisar gerar novamente uma chave de acesso (após divulgação acidental, por exemplo), pode ser necessário atualizar todos os aplicativos que se conectam usando essa chave.

Azure Active Directory (Azure AD) oferece segurança superior e facilidade de uso em relação à autorização de chave de acesso. A Microsoft recomenda usar a autorização do Azure AD quando possível para minimizar as vulnerabilidades de segurança potenciais inerentes ao uso de chaves de acesso.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Azure AD é um serviço separado do Azure. Você adiciona usuários e outras entidades de segurança (como um aplicativo) a um domínio de segurança gerenciado pelo Azure AD. O vídeo a seguir descreve como a autenticação funciona com o Azure.

[vídeo]



Para obter informações detalhadas sobre como usar o Azure AD, visite a página “O que é o Azure Active Directory?” no site da Microsoft.

### Configurar o controle de acesso

O Azure AD permite que você especifique quem ou o quê pode acessar seus recursos. O controle de acesso define o que um usuário ou aplicativo pode fazer com seus recursos depois de serem autenticados.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

O gerenciamento de acesso para recursos de nuvem é uma função crítica para qualquer organização que usa a nuvem. O controle de acesso baseado em função do Azure (Azure RBAC) ajuda a gerenciar quem tem acesso aos recursos do Azure e o que eles podem fazer com esses recursos. Por exemplo, usando RBAC, você pode:

- Permitir que um usuário gerencie máquinas virtuais em uma assinatura e outro usuário gerencie redes virtuais.
- Permitir que um grupo de administradores de banco de dados gerencie bancos de dados SQL em uma assinatura.
- Permitir que um usuário gerencie todos os recursos em um grupo de recursos, como máquinas virtuais, sites e sub-redes.
- Permitir que um aplicativo acesse todos os recursos em um grupo de recursos.
- Você controla o acesso aos recursos usando o Azure RBAC para criar atribuições de função. Uma atribuição de função consiste em três elementos: uma entidade de segurança, uma definição de função e um escopo.
- A security principal é um objeto que representa um usuário, grupo, serviço ou identidade gerenciada que está solicitando acesso aos recursos do Azure.

A definição de papel, frequentemente abreviado para Função, é uma coleção de permissões. Uma definição de função lista as operações que podem ser executadas, como leitura, gravação e exclusão. As funções podem receber nomes de alto nível, como proprietário, ou nomes específicos, como leitor de máquina virtual. O Azure inclui várias funções internas que você pode usar, incluindo:

- Proprietário - Tem acesso total a todos os recursos, incluindo o direito de delegar acesso a outros.
- Contribuinte - Pode criar e gerenciar todos os tipos de recursos do Azure, mas não pode conceder acesso a outros.
- Leitor - Pode visualizar os recursos existentes do Azure.
- Administrador de acesso do usuário - Permite gerenciar o acesso do usuário aos recursos do Azure.

Você também pode criar suas próprias funções personalizadas. Para obter informações detalhadas, consulte Criar ou atualizar funções personalizadas do Azure usando o portal do Azure no site da Microsoft.

A lista escopo o conjunto de recursos aos quais o acesso se aplica. Ao atribuir uma função, você pode limitar ainda mais as ações permitidas definindo um escopo. Isso é útil se, por exemplo, você quiser tornar alguém um Contribuidor do Site, mas apenas para um grupo de recursos.

Você adiciona atribuições de função a um recurso no portal do Azure usando o Access control (IAM)página. Atribuições de funções permite que você associe uma função a uma entidade de segurança, definindo o nível de acesso da função ao recurso. Para

mais informações, leia “Adicionar ou remover atribuições de função do Azure usando o portal do Azure”.

The screenshot shows the Azure Storage Account Access Control (IAM) interface. On the left, there's a sidebar with various options like Overview, Activity log, and Access control (IAM). The Access control (IAM) option is highlighted with a red box. The main area has tabs for Check access, Role assignments, and Deny assignments, with Role assignments selected. A sub-menu titled 'Check access' is open, showing a search bar and a 'Find' dropdown set to 'Azure AD user, group, or service principal'. Below it is a search bar with placeholder 'Search by name or email address'. To the right, a modal window titled 'Add role assignment' is open. It has a 'Role' dropdown set to 'Select a role' and a 'Assign access to' dropdown set to 'Azure AD user, group, or service principal'. A 'Select' button is followed by a search bar with placeholder 'Search by name or email address'. A list of roles is shown, each with a color-coded icon: ADSyncAdministrators (blue), ADSyncBrowne (purple), ADSyncOperators (black), ADSyncPasswordSet (blue), DnsAdmins (green), and DnsUpdateProxy (red). Below the list, a section titled 'Selected members:' says 'No members selected. Search for and add one or more members you want to assign to the role for this resource.' At the bottom of the modal are 'Save' and 'Discard' buttons.

## Configurar segurança avançada

Além de autenticação e autorização, muitos serviços fornecem proteção adicional por meio de segurança avançada.

Segurança avançada implementa proteção e avaliação de ameaças. A proteção contra ameaças adiciona inteligência de segurança ao seu serviço. Essa inteligência monitora o serviço e detecta padrões incomuns de atividade que podem ser prejudiciais ou comprometer os dados gerenciados pelo serviço. A avaliação identifica vulnerabilidades de segurança em potencial e recomenda ações para mitigá-las.

Você é cobrado uma taxa adicional por este recurso. A imagem abaixo mostra a página Segurança Avançada para armazenamento do Azure. A página correspondente para outros serviços não relacionais, como Cosmos DB, é semelhante.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

The screenshot shows the Azure Storage Advanced security settings page. On the left, there's a sidebar with various options like Overview, Activity log, Access control (IAM), Tags, etc., and a 'Settings' section with 'Advanced security' highlighted with a red box. The main content area is titled 'Advanced Threat Protection for Azure Storage' and includes a brief description, a trial offer, and a 'Enable Advanced Threat Protection' button. Below this, the 'Recommendations' section states 'No recommendations to display' and 'There are no security recommendations for this resource'. A blue button at the bottom right of this section says 'View all recommendations in Security Center'.

## Configurar o Azure Cosmos DB e o Armazenamento do Azure

Além das configurações gerais aplicáveis a muitos serviços, a maioria dos serviços também possui recursos específicos que você pode configurar. Por exemplo, no cenário de amostra, depois de provisionar uma conta do Cosmos DB, você pode precisar configurar a replicação ou as configurações de consistência do banco de dados.

Nesta unidade, você verá as definições de configuração específicas para contas do Azure Cosmos DB e do Armazenamento do Azure.

### Configurar Cosmos DB

#### Configurar replicação

O Azure Cosmos DB permite que você replique os bancos de dados e contêineres em sua conta em várias regiões. Ao provisionar inicialmente uma conta, você pode especificar que deseja copiar os dados para outra região. Você não tem controle sobre qual região é usada, pois a próxima região mais próxima é selecionada automaticamente. A página [Replicar dados globalmente](#) permite que você configure a replicação com mais detalhes. Você pode replicar para várias regiões e selecionar as regiões a serem usadas. Dessa forma, você pode escolher as regiões mais próximas de seus consumidores, para ajudar a minimizar a latência das solicitações feitas por esses consumidores.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Você também pode usar esta página para configurar o failover automático para ajudar a garantir a alta disponibilidade. Se os bancos de dados na região primária (a região em que você criou a conta) ficarem indisponíveis, uma das regiões replicadas assumirá o processamento e se tornará a nova região primária.

Por padrão, apenas a região na qual você criou a conta oferece suporte a operações de gravação; as réplicas são todas somente leitura. No entanto, você pode habilitar gravações multirregionais. No entanto, gravações multirregionais podem causar conflitos, se os aplicativos executados em regiões diferentes modificarem os mesmos dados. Nesse caso, a gravação mais recente substituirá as alterações feitas anteriormente quando os dados são replicados, embora você possa escrever seu próprio código para aplicar uma estratégia diferente.

A replicação é assíncrona, portanto, é provável que haja um atraso entre uma alteração feita em uma região e essa alteração se tornar visível em outras regiões.

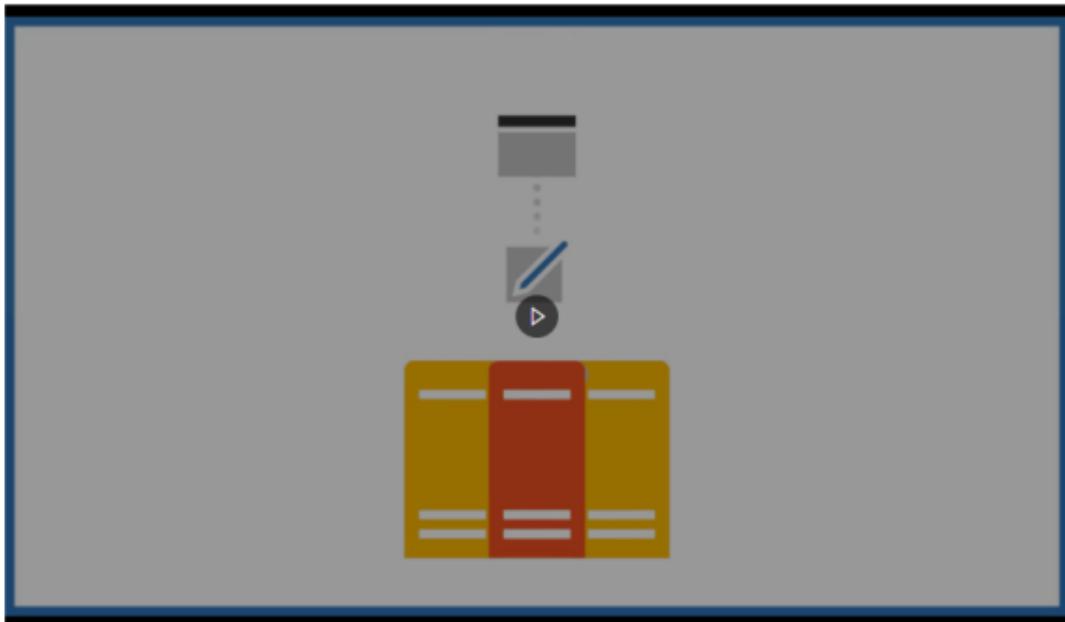
The screenshot shows the Azure portal interface for managing a Cosmos DB account named 'jpwscosmosdb'. The left sidebar has a 'Replicate data globally' option highlighted with a red box. The main content area displays a world map with various regions marked. A callout box on the map says: 'Click on a 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'. To the right, there's a 'Configure regions' section with a 'Multi-region writes' toggle set to 'Enable'. Below it is a table for 'Write Region' and 'Read Regions' with no data listed. A vertical watermark on the right side reads 'THIAGO FERNANDES DA SILVA OLIVEIRA' and 'CPF 00296691208'.

**NOTA:** Cada réplica aumenta o custo do serviço Cosmos DB. Por exemplo, se você replicar sua conta para duas regiões, seus custos serão três vezes maiores que os de uma conta não replicada.

## Configure a consistência

Em uma única região, o Cosmos DB usa um cluster de servidores. Essa abordagem ajuda a melhorar a escalabilidade e a disponibilidade. Uma cópia de todos os dados é mantida em cada servidor do cluster. O vídeo a seguir explica como isso funciona e os efeitos que pode ter na consistência.

[vídeo]



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

O Cosmos DB permite que você especifique como essas inconsistências devem ser tratadas. Ele oferece as seguintes opções:

- Eventual. Esta opção é a menos consistente. É baseado na situação que acabamos de descrever. As alterações não serão perdidas, elas aparecerão eventualmente, mas eles podem não aparecer imediatamente. Além disso, se um aplicativo fizer várias alterações, algumas delas podem ficar imediatamente visíveis, mas outras podem ser atrasadas; as alterações podem parecer fora de ordem.
- Prefixo consistente. Esta opção garante que as alterações aparecerão em ordem, embora possa haver um atraso antes de se tornarem visíveis. Neste período, os aplicativos podem ver dados antigos.
- Seção. Se um aplicativo fizer várias alterações, todas ficarão visíveis para esse aplicativo e em ordem. Outros aplicativos podem ver dados antigos, embora quaisquer alterações apareçam em ordem, como aconteceu com o Prefixo consistente opção. Esta forma de consistência às vezes é conhecida como leia suas próprias escritas.

- Staleness limitado. Há um atraso entre a gravação e a leitura dos dados atualizados. Você especifica esta desatualização como um período de tempo ou número de versões anteriores para as quais os dados serão inconsistentes.
- Forte: Neste caso, todas as gravações são visíveis apenas para os clientes depois que as alterações são confirmadas como gravadas com sucesso em todas as réplicas. Esta opção não estará disponível se você precisar distribuir seus dados por várias regiões globais.

A consistência eventual fornece a menor latência e a menor consistência. A consistência forte resulta na latência mais alta, mas também na maior consistência. Você deve selecionar um nível de consistência padrão que equilibre o desempenho e os requisitos de seus aplicativos.

Você pode alterar a consistência padrão para uma conta Cosmos DB usando o Consistência padrão página no portal do Azure. Os aplicativos podem substituir o nível de consistência padrão para operações de leitura individuais. No entanto, eles não podem aumentar a consistência acima da especificada nesta página; eles podem apenas diminuí-lo.

The screenshot shows the Azure portal interface for managing a Cosmos DB account named 'jpwscosmosdb'. In the left sidebar, under 'Settings', the 'Default consistency' option is highlighted. At the top of the main content area, there are tabs for 'STRONG', 'BOUNDED STALENESS', 'SESSION', 'CONSISTENT PREFIX', and 'EVENTUAL'. The 'EVENTUAL' tab is selected. A note below the tabs states: '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.' To the right of the note is a diagram featuring a world map and four staves of music. Each staff has a different color (green, red, purple, blue) and contains musical notes. The notes on each staff are slightly offset from each other, illustrating how multiple replicas converge over time.

THIAGO FERNANDES DA SILVA  
CPF 00296691208

## Configurar contas de armazenamento

### Configuração geral

A página Configuração para uma conta de armazenamento permite que você modifique algumas configurações gerais da conta. Você pode:

- Habilitar ou desabilitar as comunicações seguras com o serviço. Por padrão, todas as solicitações e respostas são criptografadas usando o protocolo HTTPS à medida que atravessam a Internet. Você pode desativar a criptografia se necessário, embora isso não seja recomendado.
- Alternar o nível de acesso padrão entre Frio e Quente.
- Alterar a maneira como a conta é replicada.
- Habilitar ou desabilitar a integração com o Azure AD para solicitações que acessam compartilhamentos de arquivos.

Outras opções, como tipo de conta e nível de desempenho, são exibidas nesta página apenas para informação; você não pode mudá-las.

The cost of your storage account depends on the usage and the options you choose below.  
Learn more

Account kind  
StorageV2 (general purpose v2)

Performance  
Standard Premium

Secure transfer required  
Disabled Enabled

Access tier (default)  
Hot Cool

Replication  
Read-access geo-redundant storage (RA-GRS)

Large file shares  
Disabled Enabled

The current combination of storage account kind, performance, replication and location does not support large file shares.  
Learn more

Identity-based access for file shares  
Azure Active Directory Domain Service (AAD DS)  
Disabled Enabled

Active Directory (AD)  
How to domain join this storage account

Data Lake Storage Gen2  
Hierarchical namespace  
Disabled Enabled

NFS v3  
Disabled Enabled

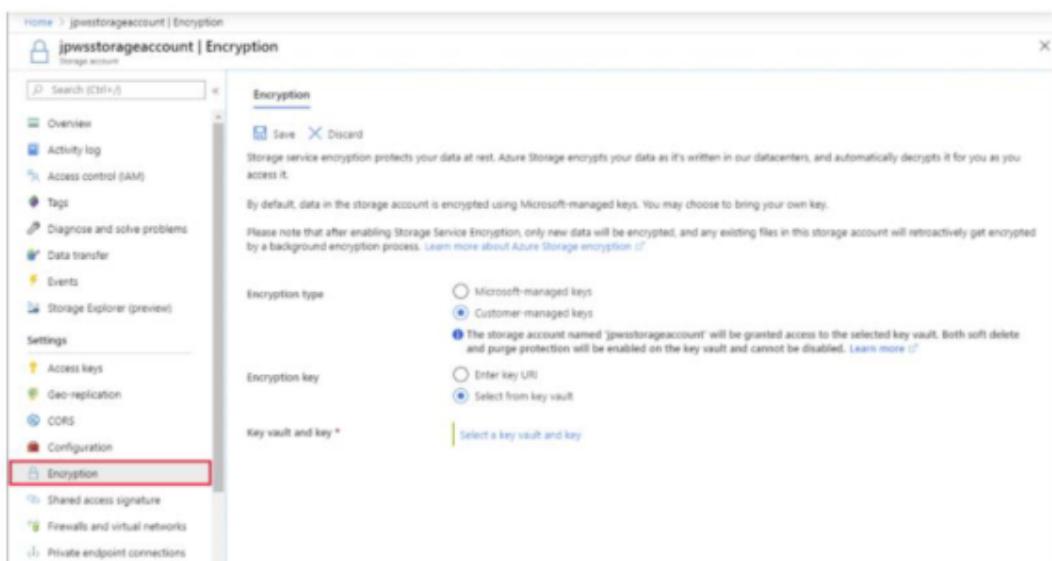
The current combination of storage account kind, performance, replication, and location does not support the NFS v3 feature.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Configurar criptografia

Todos os dados mantidos em uma conta de Armazenamento do Azure são criptografados automaticamente. Por padrão, a criptografia é executada usando chaves gerenciadas e de propriedade da Microsoft. Se preferir, você pode fornecer suas próprias chaves de criptografia.

Para usar suas próprias chaves, adicione-as ao Azure Key Vault. Você então fornece os detalhes do cofre e da chave, ou o URI da chave no cofre. Todos os novos dados serão criptografados conforme são gravados. Os dados existentes serão criptografados usando um processo executado em segundo plano; este processo pode demorar um pouco.



## Configurar assinaturas de acesso compartilhado

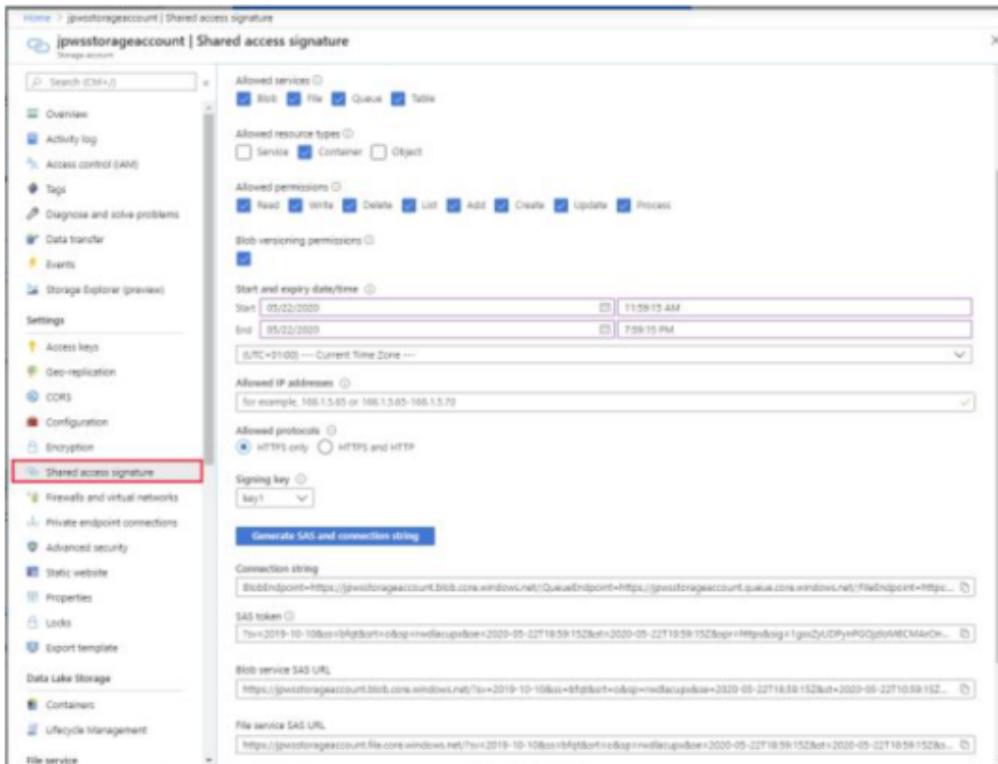
Você pode usar assinaturas de acesso compartilhado (SAS) para conceder direitos limitados a recursos em uma conta de armazenamento do Azure por um período de tempo especificado. Esse recurso permite que os aplicativos acessem recursos como blobs e arquivos, sem exigir que sejam autenticados primeiro. Você só deve usar o SAS para dados que pretende tornar públicos.

Um SAS é um token que um aplicativo pode usar para se conectar ao recurso. O aplicativo anexa o token à URL do recurso. O aplicativo pode então enviar solicitações para ler ou gravar dados usando este URL e token.

Você pode criar um token que conceda acesso temporário a todo o serviço, contêineres no serviço ou objetos individuais, como blobs e arquivos.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF: 00296691208

Use a Assinatura de acesso compartilhado no portal do Azure para gerar tokens SAS. Você especifica as permissões (você pode fornecer acesso somente leitura a um blob, por exemplo), o período para o qual o token SAS é válido e o intervalo de endereços IP de computadores com permissão para usar o token SAS. O token SAS é criptografado usando uma das chaves de acesso; você especifica qual chave usar (chave1 ou chave2).



## Laboratório: Provisionar serviços de dados não relacionais do Azure

No cenário de amostra, você decidiu criar os seguintes armazenamentos de dados:

- Um Cosmos DB para conter informações sobre o volume de itens em estoque. Você precisa armazenar informações atuais e históricas sobre os níveis de volume, para que possa acompanhar como os níveis variam ao longo do tempo. Os dados são registrados diariamente.
- Uma loja Data Lake para armazenar dados de produção e qualidade.
- Um contêiner blob para armazenar imagens dos produtos que a empresa fabrica.
- Armazenamento de arquivos para compartilhamento de relatórios.

TÍTULO: FERNANDES DA SILVA OLIVEIRA

CNPJ: 002796691208

Vá Exercício: Provisionar serviços de dados não relacionais do Azure no Microsoft Learn e siga as instruções do módulo para provisionar e configurar a conta Cosmos DB e testá-la criando um banco de dados, um contêiner e um documento de amostra. Você também provisionará uma conta de Armazenamento do Azure que pode fornecer armazenamento de blob, arquivo e Data Lake.

Você realizará este exercício usando o portal do Azure.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## Teste de conhecimento

### Questão 1

#### O que é provisionamento?

- a) O ato de executar uma série de tarefas que um provedor de serviços executa para criar e configurar um serviço.
- b) Fornecer a outros usuários acesso a um serviço existente.
- c) Ajustando um serviço para melhorar o desempenho.

Explicação: está correto. No Azure, você deve provisionar um serviço antes de usá-lo.

### Questão 2

#### O que é um principal de segurança?

- a) Uma coleção nomeada de permissões que podem ser concedidas a um serviço, como a capacidade de usar o serviço para ler, gravar e excluir dados. No Azure, os exemplos incluem Proprietário e Contribuinte.
- b) Um conjunto de recursos gerenciados por um serviço ao qual você pode conceder acesso.
- c) Um objeto que representa um usuário, grupo, serviço ou identidade gerenciada que está solicitando acesso aos recursos do Azure.

Explicação: está correto. A autenticação do Azure usa princípios de segurança para ajudar a determinar se uma solicitação de acesso a um serviço deve ser concedida.

### Questão 3

#### Qual das opções a seguir é uma vantagem de usar a replicação multirregional com o Cosmos DB?

- a) Os dados sempre serão consistentes em todas as regiões.
- b) A disponibilidade é aumentada.
- c) Maior segurança para seus dados.

Explicação: está correto. A replicação melhora a disponibilidade. Se uma região se tornar inacessível, os dados ainda estarão disponíveis em outras regiões.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Resumo

Provisionamento é o ato de criar uma instância de um serviço. O Azure se encarrega de alocar os recursos necessários para executar um serviço como parte do processo de provisionamento. Depois de provisionar um serviço, você pode configurá-lo para permitir que seus aplicativos e usuários acessem o serviço.

- Nesta lição, você aprendeu como:
- Fornecimento de serviços de dados não relacionais
- Configurar serviços de dados não relacionais
- Explorar problemas básicos de conectividade
- Explorar os componentes de segurança de dados

## Saber mais

- O que é CLI do Azure
- Documentação do Azure PowerShell
- O que são modelos do Azure Resource Manager?
- Notebooks Jupyter integrados no Azure Cosmos DB
- O que é ponto final privado do Azure?
- Gerenciar recursos de API do Azure Cosmos DB Core (SQL) com modelos do Azure Resource Manager
- Configurar firewalls e redes virtuais do Armazenamento do Azure
- Criar ou atualizar funções personalizadas do Azure usando o portal do Azure
- Adicionar ou remover atribuições de função do Azure usando o portal do Azure
- O que é o Azure Active Directory?

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Introdução

Os armazenamentos de dados não relacionais podem assumir várias formas. O Azure permite que você crie bancos de dados não relacionais usando o Azure Cosmos DB. O Cosmos DB oferece suporte a vários modelos NoSQL, incluindo armazenamentos de documentos, bancos de dados gráficos, armazenamentos de valores-chave e bancos de dados de famílias de colunas. Outros armazenamentos não relacionais disponíveis no Azure incluem o Armazenamento do Azure, que você pode usar para armazenar blobs e arquivos. Nesta lição, você aprenderá como usar esses vários serviços de armazenamento para armazenar e recuperar dados.

Suponha que você seja um engenheiro de dados trabalhando na Contoso, uma organização com uma grande operação de manufatura. A organização deve reunir e armazenar informações de uma variedade de fontes, como dados em tempo real de monitoramento do status das máquinas da linha de produção, dados de controle de qualidade do produto, registros históricos de produção, volumes de produtos em estoque e dados de estoque de matérias-primas. Essas informações são críticas para o funcionamento da organização. A Contoso criou lojas para armazenar essas informações. Você foi solicitado a carregar dados para esses armazenamentos e investigar como consultar esses dados usando os recursos fornecidos pelo Azure.

## Objetivos de aprendizado

Nesta lição, você irá:

- Carregar dados em um banco de dados Cosmos DB e aprenda como consultar esses dados.
- Carregar e baixar dados em uma conta de armazenamento do Azure.

THIAGO FERNANDES DA SILVA  
CPF 00296691208

## Gerenciar Azure Cosmos DB

O Azure Cosmos DB é um sistema de gerenciamento de banco de dados NoSQL. É compatível com alguns sistemas NoSQL existentes, incluindo MongoDB e Cassandra. No cenário da Contoso, você criou um banco de dados Cosmos DB para armazenar informações sobre a quantidade de itens em estoque. Agora você precisa entender como preencher esse banco de dados e como consultá-lo.

Nesta unidade, você revisará como o Cosmos DB armazena dados. Em seguida, você aprenderá como fazer upload de dados para um banco de dados Cosmos DB e configurar o Cosmos DB para suportar o carregamento em massa.

### O que é Azure Cosmos DB?

Cosmos DB gerencia dados como um conjunto de documentos. Um documento é uma coleção de campos, identificados por uma chave. Os campos em cada documento podem variar e um campo pode conter documentos filho. Cosmos DB usa JSON (JavaScript Object Notation) para representar a estrutura do documento. Nesse formato, os campos em um documento são colocados entre colchetes, {e}, e cada campo é prefixado com seu nome. O exemplo abaixo mostra um par de documentos que representam as informações do cliente. Em ambos os casos, cada documento do cliente inclui documentos filho contendo o nome e endereço, mas os campos nesses documentos filhos variam entre os clientes.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

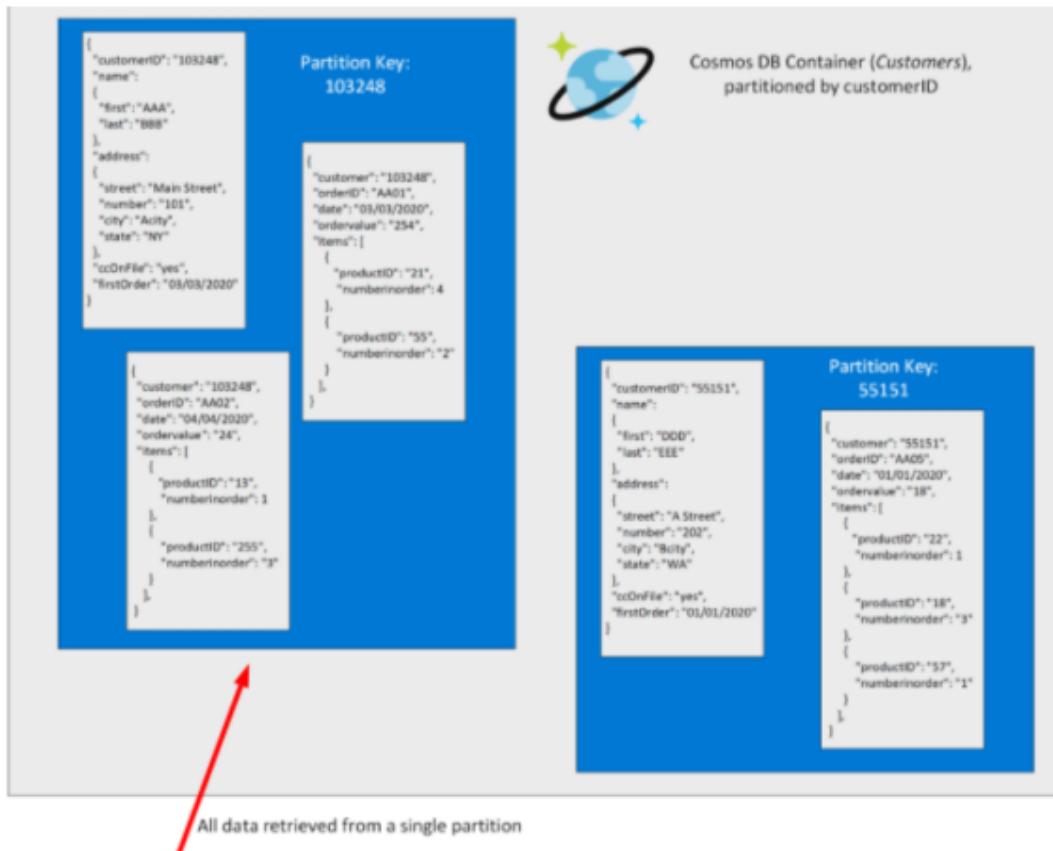
```
## Document 1 ##
{
  "customerID": "103248",
  "name":
  {
    "first": "AAA",
    "last": "BBB"
  },
  "address":
  {
    "street": "Main Street",
    "number": "101",
    "city": "Acity",
    "state": "NY"
  },
  "ccOnFile": "yes",
  "firstOrder": "02/28/2003"
}

## Document 2 ##
{
  "customerID": "103249",
  "name":
  {
    "title": "Mr",
    "forename": "AAA",
    "lastname": "BBB"
  },
  "address":
  {
    "street": "Another Street",
    "number": "202",
    "city": "Bcity",
    "county": "Gloucestershire",
    "country-region": "UK"
  },
  "ccOnFile": "yes"
}
```

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Os documentos em um banco de dados Cosmos DB são organizados em contêineres. Os documentos em um contêiner são agrupados em partições. Uma partição contém um conjunto de documentos que compartilham uma chave de partição comum. Você designa um dos campos em seus documentos como a chave de partição. Seleciona uma chave de partição que reúna todos os documentos relacionados. Essa abordagem ajuda a reduzir a quantidade de operações de leitura de disco que as consultas usam ao recuperar um conjunto de documentos para uma determinada entidade. Por exemplo, em um banco de dados de documentos para um sistema de comércio eletrônico que registra os detalhes dos clientes e os pedidos que eles fizeram, você pode particionar os dados por ID do cliente e armazenar os detalhes do cliente e do pedido de cada cliente na mesma partição. Para encontrar todas as informações e pedidos de um cliente, você simplesmente precisa consultar essa partição única:



```
SELECT *
FROM Customers a
WHERE a.customerID = "103248"
```

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Cosmos DB é um serviço básico no Azure. Cosmos DB é usado por muitos dos produtos da Microsoft para aplicativos de missão crítica executados em escala global, incluindo Skype, Xbox, Office 365 e Azure. Cosmos DB é altamente adequado para IoT, varejo e marketing, jogos e aplicativos da Web. Para obter informações adicionais sobre os usos do Cosmos DB, leia “Casos de uso comuns do Azure Cosmos DB”.

### O que são APIs do Cosmos DB?

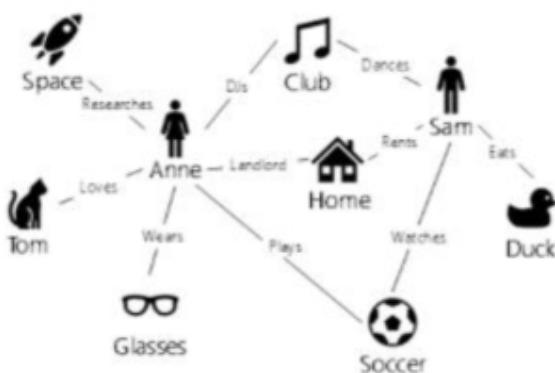
Você acessa os dados em um banco de dados Cosmos DB por meio de um conjunto de comandos e operações, conhecidos coletivamente como um FOGO, ou Interface de programação de aplicativo. Cosmos DB fornece sua própria API nativa, chamada de API SQL. Esta API fornece uma linguagem de consulta semelhante a SQL sobre documentos, que permite recuperar documentos usando SELECIONE. O exemplo abaixo encontra o endereço do cliente 103248 nos documentos mostrados acima:

```
SELECT c.address  
FROM customers c  
WHERE c.customerID = "103248"
```

O Cosmos DB também fornece outras APIs que permitem acessar esses documentos usando os conjuntos de comandos de outros sistemas de gerenciamento de banco de dados NoSQL. Essas APIs são:

- *API de tabela.* Essa interface permite que você use a API do Azure Table Storage para armazenar e recuperar documentos. O objetivo desta interface é permitir que você altere do Table Storage para o Cosmos DB sem exigir que você modifique seus aplicativos existentes.
- *API MongoDB.* O MongoDB é outro banco de dados de documentos bem conhecido, com sua própria interface programática. Muitas organizações usam no local. Você pode usar a API MongoDB para Cosmos DB para permitir que um aplicativo MongoDB seja executado sem alterações em um banco de dados Cosmos DB. Você pode migrar os dados do banco de dados MongoDB para o Cosmos DB em execução na nuvem, mas continue executando seus aplicativos existentes para acessar esses dados.
- *API Cassandra.* Cassandra é um sistema de gerenciamento de banco de dados de família de colunas. Este é outro sistema de gerenciamento de banco de dados que muitas organizações executam no local. A API Cassandra para Cosmos DB fornece uma interface programática semelhante a Cassandra para Cosmos DB. As solicitações da API Cassandra são mapeadas para as solicitações de documentos do Cosmos DB. Assim como com a API do MongoDB, o objetivo principal da API do Cassandra é permitir que você migre rapidamente os bancos de dados e aplicativos do Cassandra para o Cosmos DB.

- **API Gremlin.** A API Gremlin implementa uma interface de banco de dados gráfico para o Cosmos DB. Um gráfico é uma coleção de objetos de dados e relacionamentos direcionados. Os dados ainda são mantidos como um conjunto de documentos no Cosmos DB, mas a API do Gremlin permite que você execute consultas de gráfico sobre os dados. Usando a API do Gremlin, você pode percorrer os objetos e relacionamentos no gráfico para descobrir todos os tipos de relacionamentos complexos, como "Qual é o nome do animal de estimação do proprietário de Sam?" no gráfico mostrado abaixo.



O principal uso das APIs Table, MongoDB e Cassandra é dar suporte a aplicativos existentes escritos usando esses armazenamentos de dados. Se você estiver construindo um novo aplicativo e banco de dados, deve usar a API SQL ou a API Gremlin.

### Realizar operações de dados no Cosmos DB

O Cosmos DB oferece várias opções para fazer upload de dados para um banco de dados Cosmos DB e consultar esses dados. Você pode:

- Usar **Data Explorer** no portal do Azure para executar consultas ad-hoc. Você também pode usar essa ferramenta para carregar dados, mas só pode carregar um documento por vez. A funcionalidade de carregamento de dados tem como objetivo principal o upload de um pequeno número de documentos (até 2 MB no tamanho total) para fins de teste, em vez de importar grandes quantidades de dados.
- Use o Ferramenta de migração de dados do Cosmos DB para realizar um carregamento em massa ou transferência de dados de outra fonte de dados.
- Usar Azure Data Factory para importar dados de outra fonte.
- Escreva um aplicativo personalizado que importe dados usando o Cosmos DB BulkExecutor. Essa estratégia está além do escopo deste módulo.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

- Crie seu próprio aplicativo que usa as funções disponíveis através do Biblioteca cliente da Cosmos DB SQL API para armazenar dados. Essa abordagem também está além do escopo deste módulo.

### Carregue dados usando a ferramenta Cosmos DB Data Migration

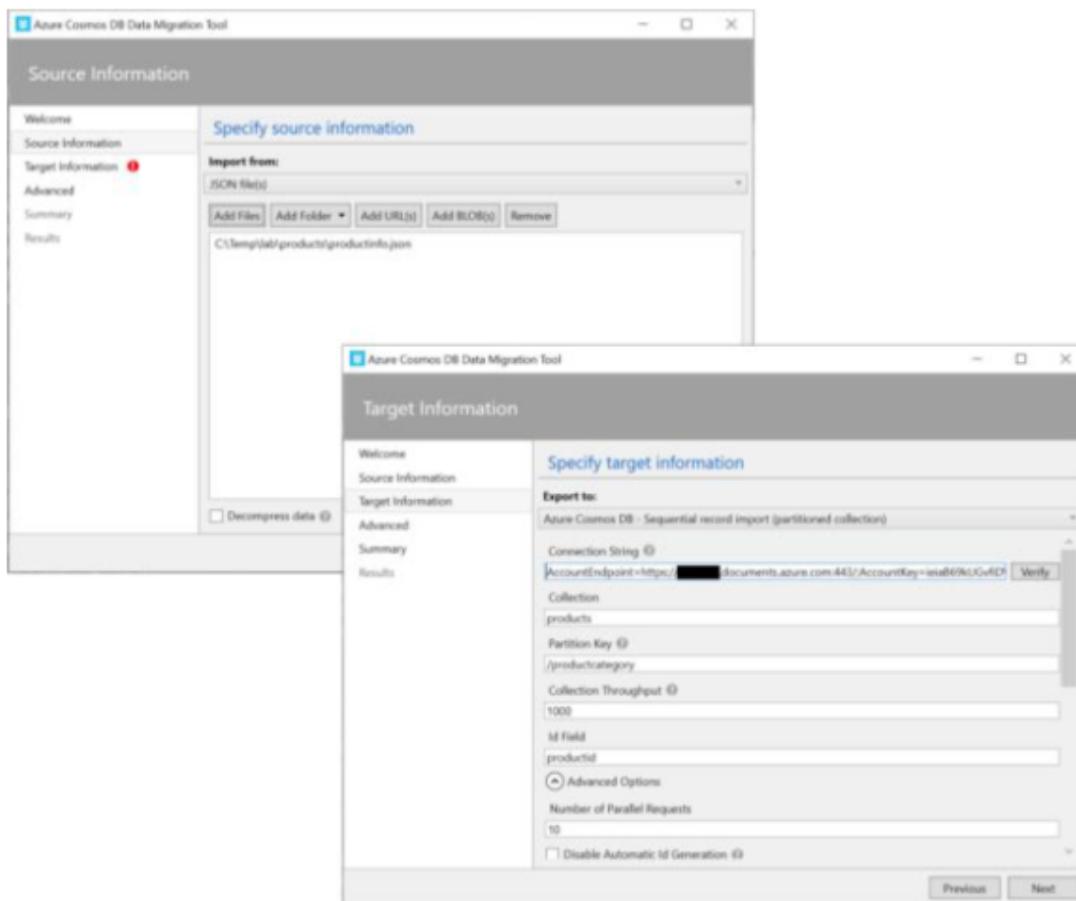
Você pode usar a ferramenta de migração de dados para importar dados para o Azure Cosmos DB de uma variedade de fontes, incluindo:

- Arquivos JSON
- MongoDB
- servidor SQL
- Arquivos CSV
- Armazenamento da Tabela Azure
- Amazon DynamoDB
- HBase
- Contêineres do Azure Cosmos

A ferramenta de migração de dados está disponível como um baixar do GitHub. A ferramenta orienta você no processo de migração de dados para um banco de dados Cosmos DB. Você será solicitado a fornecer a origem dos dados (um dos itens listados acima) e o destino (o banco de dados e o contêiner do Cosmos DB). A ferramenta pode preencher um container existente ou criar um novo se o container especificado ainda não existir.

**NOTA:** Você também pode usar a ferramenta de migração de dados para exportar dados de um contêiner Cosmos DB para um arquivo JSON, mantido localmente ou no armazenamento de Blob do Azure

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208



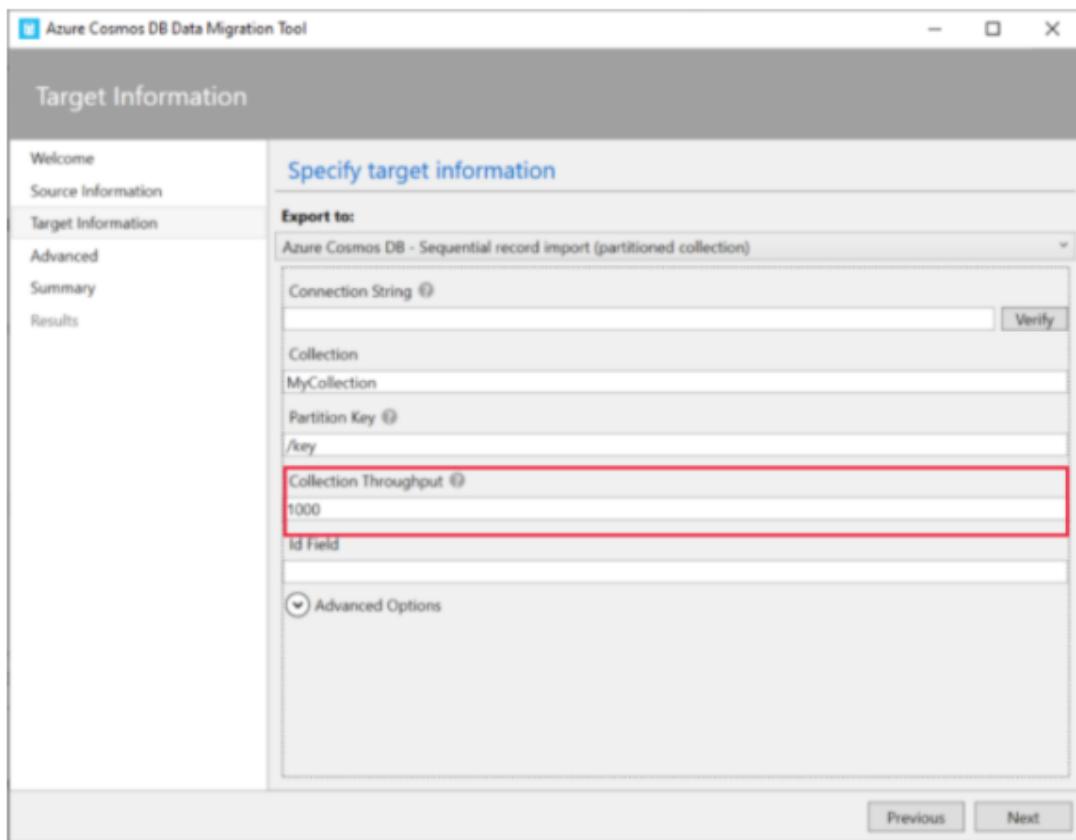
## Configure o Cosmos DB para suportar o carregamento em massa

Se você tiver uma grande quantidade de dados, a ferramenta de migração de dados pode usar vários threads simultâneos para agrupar seus dados em blocos e carregá-los em paralelo. Cada thread atua como uma conexão de cliente separada com o banco de dados. O carregamento em massa pode se tornar uma tarefa de gravação intensiva.

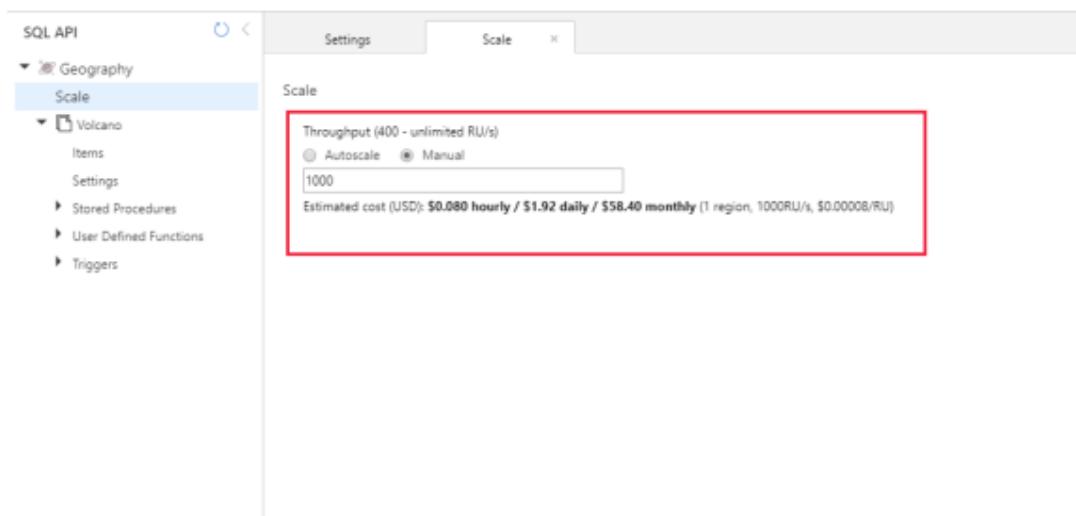
Ao fazer upload de dados para um contêiner, se você tiver capacidade de rendimento insuficiente configurada para suportar o volume de operações de gravação que ocorrem simultaneamente, algumas das solicitações de upload falharão. Cosmos DB relata um erro HTTP 429 (a taxa de solicitação é grande). Portanto, se você estiver planejando realizar uma grande importação de dados, deverá aumentar os recursos de taxa de transferência disponíveis para o contêiner Cosmos de destino. Se você estiver usando a ferramenta de migração de dados para criar o contêiner e também para preenchê-lo, a Informação do alvo permite que você especifique os recursos de taxa de transferência a serem alocados.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208



Se você já criou o contêiner, use o **Escala** configurações do banco de dados na página Data Explorer para seu banco de dados no portal do Azure para especificar a taxa de transferência máxima ou definir a taxa de transferência para **Escala automática**.



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Depois que os dados forem carregados, você poderá reduzir os recursos de taxa de transferência para diminuir os custos do banco de dados.

## Consultar Azure Cosmos DB

Embora o Azure Cosmos DB seja descrito como um sistema de gerenciamento de banco de dados NoSQL, a API SQL permite que você execute consultas Semelhante a SQL em bancos de dados Cosmos DB. Essas consultas usam uma sintaxe semelhante à do SQL, mas existem algumas diferenças. Isso ocorre porque os dados em um Cosmos DB são estruturados como documentos em vez de tabelas.

Nesta unidade, você aprenderá sobre o dialeto do SQL implementado pela API SQL. Você verá como usar o **Data Explorer** no portal do Azure para executar consultas.

### Use a API SQL para consultar documentos

A API Cosmos DB SQL oferece suporte a um dialeto de SQL para consultar documentos usando instruções SELECT que serão familiares se você tiver escrito instruções SELECT em um banco de dados relacional usando um mecanismo de banco de dados compatível com ANSI SQL. A API SQL retorna resultados na forma de documentos JSON. Todas as consultas são executadas no contexto de um único contêiner.

### Compreender uma consulta de API SQL

Uma consulta SELECT da API SQL inclui as seguintes cláusulas:

1. **Cláusula SELECT.** A cláusula começa com a palavra-chave SELECT seguida por uma lista separada por vírgulas de propriedades a serem retornadas. A palavra-chave "\*" significa todas as propriedades do documento.
2. **Cláusula FROM.** Esta cláusula começa com a palavra-chave FROM seguida por um identificador, representando a origem dos registros, e um alias que você pode usar para esse identificador em outras cláusulas (o alias é opcional). Em uma consulta de banco de dados relacional, a cláusula FROM conteria um nome de tabela. Na API SQL, todas as consultas são limitadas ao escopo de um contêiner, portanto, o identificador representa o nome do contêiner.
3. **Cláusula WHERE.** Esta cláusula é opcional. Ele começa com a palavra-chave WHERE seguida por uma ou mais condições lógicas que devem ser satisfeitas por um documento retornado pela consulta. Você usa a cláusula WHERE para filtrar os resultados de uma consulta.
4. **Cláusula ORDER BY.** Esta cláusula também é opcional. Ele começa com a frase ORDER BY seguida por uma ou mais propriedades usadas para ordenar o conjunto de resultados de saída.

**NOTA:** Uma consulta também pode conter uma cláusula JOIN. Em um sistema de gerenciamento de banco de dados relacional, como o Banco de Dados SQL do Azure, as cláusulas JOIN são usadas para conectar dados de diferentes tabelas. Na API

CURSO DE AZURE DATA FLOW

CPF: 000296691208

SQL, você usa cláusulas JOIN para conectar campos em um documento com campos em um subdocumento que faz parte do mesmo documento. Você não pode realizar junções em documentos diferentes.

Os exemplos abaixo mostram algumas consultas simples:

```
// Simple SELECT. The identifier "c" is an alias for the container being queried
SELECT c.*
FROM customers c

// Projection - limit the output to specified fields
SELECT c.Title, c.Name
FROM customers c

// Projection - Address is a subdocument that contains fields named "state" and "city",
// amongst others

SELECT c.Name, c.Address.State, c.Address.City
FROM customers c

// Filter that limits documents to customers living in California
SELECT c.Name, c.Address.City
FROM customers c
WHERE c.Address.State = "CA"

// Retrieve customers living in California in Name order
SELECT c.Name, c.Address.City
FROM customers c
WHERE c.Address.State = "CA"
ORDER BY c.Name
```

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## Compreenda os operadores suportados

A API SQL inclui muitas operações matemáticas e de string comuns, além de funções para trabalhar com matrizes e verificar tipos de dados. Os operadores com suporte em consultas de API SQL incluem:

| Tipo                | Operador               |
|---------------------|------------------------|
| Unário              | +,-,~, NÃO             |
| Aritmética          | +,-,*/,%               |
| Bit a bit           | , &, ^, <<, >>, >>>    |
| Lógico              | AND, OR                |
| Comparação          | =,! =, <,>, <=,> =, <> |
| String (concatenar) |                        |
| Ternário (se)       | ?                      |

A API SQL também suporta:

- O operador DISTINCT que você usa como parte da cláusula SELECT para eliminar duplicatas nos dados de resultado.
- O operador TOP que você pode usar para recuperar apenas as primeiras linhas retornadas por uma consulta que poderia gerar um grande conjunto de resultados.
- A operação BETWEEN que você usa como parte da cláusula WHERE para definir um intervalo inclusivo de valores. O campo de condição BETWEEN a AND b é equivalente à condição `campo >= a AND campo <= b`.
- O operador IS\_DEFINED que você pode usar para detectar se um campo especificado existe em um documento.

TITULO FUNDAMENTOS DA AZURE

CPF: 00296691208

A consulta abaixo mostra alguns exemplos de uso desses operadores.

/ Liste todas as cidades dos clientes (remova as duplicatas) para clientes que moram em estados com códigos entre AK (Alasca) e MD (Maryland)

// List all customer cities (remove duplicates) for customers living in states with codes between AK (Alaska) and MD (Maryland)

```
SELECT DISTINCT c.Address.City  
FROM c  
WHERE c.Address.State BETWEEN "AK" AND "MD"
```

```
// Find the 3 most common customer names  
SELECT TOP 3 *  
FROM c  
ORDER BY c.Name
```

```
// Display the details of every customer for which the data of birth is recorded  
SELECT * FROM p  
WHERE IS_DEFINED(p.DateOfBirth)
```

### Compreenda as funções de agregação

Você pode usar funções de agregação para resumir dados em consultas SELECT; você coloca funções agregadas na cláusula SELECT. A linguagem de consulta da API SQL oferece suporte às seguintes funções de agregação:

- **CONTAGEM (p).** Esta função retorna uma contagem do número de instâncias do campo p no conjunto de resultados. Para contar todos os itens no conjunto de resultados, defina p como um valor escalar, como 1.
- **Soma (p).** Esta função retorna a soma de todas as instâncias do campo p no conjunto de resultados. Os valores de p devem ser numéricos.
- **AVG (p).** Esta função retorna a média matemática de todas as instâncias do campo p no conjunto de resultados. Os valores de p devem ser numéricos.
- **MAX (p).** Esta função retorna o valor máximo do campo p no conjunto de resultados.
- **MIN (p).** Esta função retorna o valor mínimo do campo p no conjunto de resultados.

Embora a sintaxe das funções agregadas seja semelhante ao ANSI SQL, ao contrário do ANSI SQL, a linguagem de consulta da API SQL não oferece suporte à cláusula GROUP BY; você não pode gerar subtotais para valores diferentes do mesmo campo em uma única consulta. Você pode incluir mais de uma função agregada na cláusula SELECT de suas consultas.

No exemplo a seguir, a consulta retorna a média, o máximo e a soma do campo dos documentos em uma coleção, além de uma contagem de todos os documentos na coleção:

```
SELECT AVG(c.age) AS avg,  
       MAX(c.age) AS max,  
       SUM(c.age) AS sum,  
       COUNT(1) AS count  
FROM c
```

A API SQL também oferece suporte a um grande número de funções matemáticas, trigonométricas, de string, de matriz e espaciais. Para obter informações detalhadas sobre a sintaxe de consultas e as funções e operadores suportados pela API Cosmos DB SQL, visite a página [Introdução às consultas SQL no Azure Cosmos DB](#) no site da Microsoft.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Consultar documentos com a API SQL usando Data Explorer

Você pode usar o Data Explorer no portal do Azure para criar e executar consultas em um contêiner do Cosmos DB. A página Itens de um contêiner fornece o **Nova Consulta SQL**, comando na barra de ferramentas:

The screenshot shows the Azure Data Explorer interface. On the left, there's a sidebar with 'SQL API' selected. Under 'Volcano', 'Items' is selected. In the main area, a query 'SELECT \* FROM c' is run against the 'Volcano' container. The results are displayed in a table with columns 'id' and '/Type'. The data includes various volcano types like Shield volcano, Stratovolcano, and Maar.

| id             | /Type          |
|----------------|----------------|
| 4cb67ab0-b...  | Shield volcano |
| 246927ec-1...  | Stratovolcano  |
| a6297b2d-...   | Stratovolcano  |
| cd080a05-b...  | Maar           |
| 9e3c494e-8...  | Stratovolcano  |
| 81ed06ee-8...  | Stratovolcano  |
| 6802c282-2...  | Stratovolcano  |
| 1edb7d7-4...   | Stratovolcano  |
| 7971f48f-af... | Volcanic field |
| e26b4342-2...  | Stratovolcano  |

No painel de consulta que aparece, você pode inserir uma consulta SQL. Selecione **Executar consulta** para executá-la. Os resultados serão exibidos como uma lista de documentos JSON

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

The screenshot shows the Azure Data Explorer interface. On the left, there's a navigation pane with 'SQL API' selected, followed by 'Geography', 'Scale', and 'Volcano'. Under 'Volcano', 'Items' is selected. At the top, there's a toolbar with icons for back, forward, search, and a red-bordered 'Execute Query' button. To the right of the toolbar is a 'Save Query' button. Below the toolbar, the main area has tabs for 'Items' and 'Query 1'. The 'Query 1' tab contains the following T-SQL code:

```
1 SELECT *
2 FROM c
3 WHERE c.Country = "Japan"
```

Below the code, there are 'Results' and 'Query Stats' tabs. The 'Results' tab shows the first 100 results, which are JSON objects representing volcanoes. One result is expanded to show its location details:

```
{
  "Volcano Name": "Abu",
  "Country": "Japan",
  "Region": "Honshu-Japan",
  "Location": {
    "type": "Point",
    "coordinates": [
      131.6,
      34.5
    ]
  },
  "Elevation": 571,
  "Type": "Shield volcano",
  "Status": "Holocene",
  "Last Known Eruption": "Unknown",
  "id": "4cb67ab0-ba1a-0e8a-8dfc-a48472fd5766",
  "_rid": "ckxaAIyZRGYBAAAAAAAAAsw==",
  "_self": "dbs/ckxaA==/colls/ckxaAIyZRGY=/docs/ckxaAIyZRGYBAAAAAAAAAsw==/",
  "_etag": "\"1b005dc5-0000-0700-0000-5ece7f850000\"",
  "_attachments": "Attachments/",
  "_ts": 1590591365
}, {
  "Volcano Name": "Adatara",
  "Country": "Japan",
  "Region": "Honshu-Japan",
  "Location": {
    "type": "Point",
    "coordinates": [
      140.28,
      37.62
    ]
  },
  "Elevation": 1000,
  "Type": "Stratovolcano",
  "Status": "Extinct",
  "Last Known Eruption": "Unknown",
  "id": "4cb67ab0-ba1a-0e8a-8dfc-a48472fd5767",
  "_rid": "ckxaAIyZRGYBAAAAAAAAAsw==",
  "_self": "dbs/ckxaA==/colls/ckxaAIyZRGY=/docs/ckxaAIyZRGYBAAAAAAAAAsw==/",
  "_etag": "\"1b005dc5-0000-0700-0000-5ece7f850000\"",
  "_attachments": "Attachments/",
  "_ts": 1590591365
}
```

Você pode salvar o texto da consulta se precisar repeti-lo no futuro. A consulta é salva em um contêiner separado. Você pode recuperá-la mais tarde usando **Consulta aberta** na barra de ferramentas.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

The screenshot shows the Azure portal interface. On the left, there's a navigation tree with 'SQL API', 'Geography', 'Scale', and 'Volcano' selected. Under 'Volcano', 'Items' is selected. In the center, there's a search bar with 'Open Query' and 'Items' dropdown, and a 'SELECT \* FROM c' query editor with an 'Edit Filter' button. To the right is a table listing documents with columns 'id' and '/Type'. The table contains the following data:

| id             | /Type          |
|----------------|----------------|
| 4cb67ab0-b...  | Shield volcano |
| 246927ec-1...  | Stratovolcano  |
| a6297b2d-...   | Stratovolcano  |
| cd080a05-b...  | Maar           |
| 9e3c494e-8...  | Stratovolcano  |
| 81ed06ee-8...  | Stratovolcano  |
| 6802c282-2...  | Stratovolcano  |
| 1edb7fd7-4...  | Stratovolcano  |
| 7971f48f-af... | Volcanic field |

**NOTA:** A página itens também permite modificar e excluir documentos. Selecione um documento da lista para exibi-lo no painel principal. Você pode modificar qualquer um dos campos e selecionar **Atualizar** para salvar as alterações. Selecione **Excluir** para remover o documento da coleção. O **Novo item** permite adicionar manualmente um novo documento à coleção. Você pode usar o **Carregar item** para criar novos documentos a partir de um arquivo contendo dados JSON.

## Gerenciar armazenamento de Blob do Azure

O armazenamento de Blob do Azure é um repositório adequado para armazenar grandes objetos binários, como imagens, vídeo e arquivos de áudio. No cenário da Contoso, você criou um contêiner de blob para conter imagens dos produtos que a empresa fabrica.

Atualmente, o Azure oferece suporte a três tipos diferentes de blobs; Blobs de blocos, Blobs de página, e Anexar blobs. Normalmente, você usa blobs de página para implementar armazenamento em disco virtual para máquinas virtuais do Azure; eles são otimizados para suportar operações aleatórias de leitura e gravação. Blobs de anexo são adequados para armazenar dados que crescem em blocos, como logs ou outros dados de arquivo. Blobs de bloco são melhores para dados estáticos e são o tipo mais apropriado de armazenamento para manter os dados de imagem mantidos pela Contoso.

Nesta unidade, você aprenderá a criar e gerenciar blobs e os contêineres que os contêm.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

**NOTA:** Esta unidade se concentra no uso do portal do Azure, a CLI do Azure e o Azure PowerShell para gerenciar blobs e armazenamento de blobs. Você também pode usar o AzCopy utilitário para fazer upload e download de arquivos, incluindo blobs. A próxima unidade descreve como usar AzCopy.

## Crie um contêiner de armazenamento do Azure

Em uma conta de armazenamento do Azure, você armazena blobs em *contêiners*. Um contêiner fornece uma maneira conveniente de agrupar blobs relacionados, e você pode organizar blobs em uma hierarquia de pastas dentro de um contêiner, semelhante aos arquivos em um sistema de arquivos no disco.

Você cria um contêiner em uma conta de Armazenamento do Azure. Você pode fazer isso usando o portal do Azure ou usando a CLI do Azure ou o Azure PowerShell na linha de comando.

## Use o portal do Azure

No portal do Azure, vá para **Visão geral** da sua conta de Armazenamento do Azure e selecione **Contêiners**.

The screenshot shows the Azure Storage Account Overview page for the 'contosodata' account. The left sidebar lists various settings like Activity log, Access control (IAM), Tags, and Configuration. The main content area displays account details such as Resource group (learning), status (Primary: Available, Secondary: Available), Location (UK South, UK West), and Subscription (Free). Below these details are sections for Tags and Container storage types. The 'Containers' section is highlighted with a red box. It describes 'Containers' as scalable, cost-effective storage for unstructured data and provides a 'Learn more' link. Other storage options shown include File shares, Tables, and Queues.

Na página, selecione **+** **Container** e forneça um nome para o novo contêiner. Você também pode especificar o nível de acesso público. Para um contêiner que será usado para conter blobs, o nível de acesso mais apropriado é **Blob**. Essa configuração oferece suporte a acesso anônimo somente leitura para blobs. No entanto, os clientes não autenticados não podem listar os blobs no contêiner. Isso significa que eles só podem baixar um blob se souberem seu nome e localização dentro do contêiner.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

[+ Container](#) [Change access level](#) [Refresh](#) [Delete](#)

Search containers by prefix:

| Name   | Last modified |
|--|---------------|
| You don't have any containers yet. Click '+ Container' to get started. |               |

**Name \***  
images

**Public access level** ⓘ  
Blob (anonymous read access for blobs only)

⚠ Blobs within the container can be read by anonymous request, but container data is not available. Anonymous clients cannot enumerate the blobs within the container. Anonymous access bypasses Access Control List (ACL) settings.

The screenshot shows a Microsoft Power BI Data Studio interface. On the left, a sidebar titled "Geography" lists items like Scale, Items, Settings, Stored Procedures, User Defined Functions, and Triggers. The "Items" section is selected. In the main area, a table titled "Items" is displayed with columns "Id" and "Type". The table contains 22 rows of data. The first few rows are:

| Type             | Id             |
|------------------|----------------|
| Shield volcano   | 4cb6Tab0-b...  |
| Stratovolcano    | 248R27ec1...   |
| Stratovolcano    | a6297b2d...    |
| Maar             | cd00a05-l...   |
| Stratovolcano    | 9e3c494e-8...  |
| Stratovolcano    | 8ed60ee-a...   |
| Stratovolcano    | 6802282-2...   |
| Stratovolcano    | 7ed6ef07-4...  |
| Volcanic field   | 7977448f-a...  |
| Stratovolcano    | e28e6f42-2...  |
| Stratovolcano    | 4a9f9255d-f... |
| Stratovolcano    | 644c94b6-8...  |
| Stratovolcano    | 740b6446-f...  |
| Stratovolcano    | 670b15d5-8...  |
| Submarine vol... | e581b14e-4...  |
| Stratovolcano    | 00546666d-0... |
| Stratovolcanoes  | 767a71-a...    |

A red box highlights the row with Id "a6297b2d...". The "Type" column for this row is "Stratovolcano". The "Id" column for this row is "a6297b2d...". The "Type" column for the row with Id "4cb6Tab0-b..." is "Shield volcano". The "Id" column for the row with Type "Stratovolcano" is "248R27ec1...". The "Type" column for the row with Id "cd00a05-l..." is "Mhaar". The "Type" column for the row with Id "9e3c494e-8..." is "Stratovolcano". The "Type" column for the row with Id "8ed60ee-a..." is "Stratovolcano". The "Type" column for the row with Id "6802282-2..." is "Stratovolcano". The "Type" column for the row with Id "7ed6ef07-4..." is "Stratovolcano". The "Type" column for the row with Id "7977448f-a..." is "Volcanic field". The "Type" column for the row with Id "e28e6f42-2..." is "Stratovolcano". The "Type" column for the row with Id "4a9f9255d-f..." is "Stratovolcano". The "Type" column for the row with Id "644c94b6-8..." is "Stratovolcano". The "Type" column for the row with Id "740b6446-f..." is "Stratovolcano". The "Type" column for the row with Id "670b15d5-8..." is "Stratovolcano". The "Type" column for the row with Id "e581b14e-4..." is "Submarine volcano". The "Type" column for the row with Id "00546666d-0..." is "Stratovolcano". The "Type" column for the row with Id "767a71-a..." is "Stratovolcanoes".

THIAGO FERNANDES DA SILVA OLIVEIRA

卷之三

## Use a CLI do Azure

Se você preferir usar a CLI do Azure, o comando `az storage container create` cria um novo contêiner. Este comando tem uma série de parâmetros opcionais e você pode encontrar os detalhes completos no site da Microsoft. O exemplo abaixo cria um container chamado `images` para armazenar blobs. O contêiner é criado em uma conta de armazenamento chamada `contosodata`. O contêiner fornece acesso anônimo a blob.

```
az storage container create \
--name images \
--account-name contosodata \
--resource-group contoso-group \
--public-access blob
```

## Use o Azure PowerShell

Você pode usar o `New-AzStorageContainer` cmdlet PowerShell para criar um novo contêiner de armazenamento. Os detalhes estão disponíveis na [página](#) no site da Microsoft. Você deve primeiro obter uma referência à conta de armazenamento usando o `Get-AzStorageAccount` comando. O código abaixo mostra um exemplo:

```
Get-AzStorageAccount `

-ResourceGroupName "contoso-group" `

-Name "contosodata" | New-AzStorageContainer `

-Name "images" `

-Permission Blob
```

## Carregar um blob para o Armazenamento do Azure

Depois de criar um contêiner, você pode fazer upload de blobs. Dependendo de como deseja organizar seus blobs, você também pode criar pastas no contêiner.

## Use o portal do Azure

Se você estiver usando o portal do Azure, vá para a página da sua conta de armazenamento e selecione Contêiners sob Serviço de blob. Na página, selecione o contêiner que deseja usar.

**NOTA:** Se você criou a conta de armazenamento com suporte para namespaces hierárquicos (para Data Lake Storage), o Serviço de blob seção não aparece no portal do Azure. Em vez disso, selecione Contêiners sob Data Lake Storage.

| Name   | Last modified          | Public access level | Lease state |
|--------|------------------------|---------------------|-------------|
| images | 5/28/2020, 10:23:21 AM | Blob                | Available   |

Na página do container, na barra de ferramentas, selecione Envio. No Blob de upload, navegue até o contêiner de arquivo dos dados para carregar. A seção suspensa Advanced fornece opções para você modificar as opções padrão. Por exemplo, você pode especificar o nome de uma pasta no contêiner (a pasta será criada se não existir), o tipo de blob e a camada de acesso. O blob criado tem o nome do arquivo que você carregou.

TUTORIAL AZURE DATA FUNDAMENTALS 1.208

**NOTA:** Você pode selecionar vários arquivos. Cada um deles será carregado em blobs separados.

## Use a CLI do Azure

Use o upload do blob de armazenamento comando para fazer upload de um arquivo para um blob em um contêiner. Os detalhes que descrevem os parâmetros para este comando estão disponíveis no upload do blob de armazenamento, no site da Microsoft.

O exemplo a seguir carrega um arquivo local chamado racer\_green\_large.gif no dados pasta para um blob chamado racer\_green na pasta \*bicicletas de imagens contêiner na conta de armazenamento.

```
az storage blob upload \ --container-name images \ --account-name contosodata \ --
file "\ data \ racer_green_large.gif" \ --name "bikes \ racer_green"
```

Se você precisar fazer upload de vários arquivos, use o lote de upload de blob de armazenamento azcomando. Este comando pega o nome de uma pasta local em vez de um nome de arquivo e carrega os arquivos nessa pasta para blobs separados. O exemplo abaixo carrega todos arquivos gif em dados para a pasta bicicletas em imagens .

```
Get-AzStorageAccount ` 
-ResourceGroupName "contoso-group" ` 
-Name "contosodata" | Set-AzStorageBlobContent ` 
-Container "images" ` 
-File "\data\racer_green_large.gif" ` 
-Blob "bikes\racer_green"
```

O Azure PowerShell não inclui atualmente um comando de carregamento de blob em lote. Se precisar carregar vários arquivos, você pode escrever seu próprio script PowerShell (use o Get-ChildItem cmdlet) para iterar pelos arquivos e carregar cada um individualmente.

### Liste os blobs em um contêiner

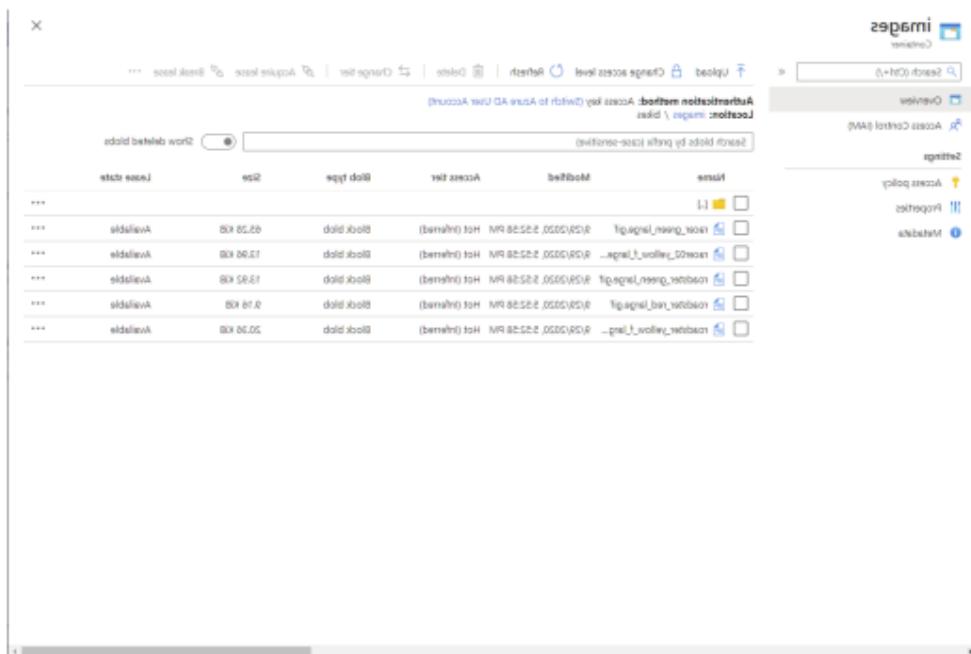
Se você recebeu os direitos de acesso apropriados, pode visualizar os blobs em um contêiner.

### Use o portal do Azure

Se você estiver usando o portal do Azure, vá para a página da sua conta de armazenamento e selecione Contêiners sob Serviço de blob.

Na página contêiners, selecione o contêiner que contém seus blobs. Se o contêiner tiver uma estrutura de pastas, vá para a pasta que contém os blobs que deseja ver. Os blobs nessa pasta devem ser exibidos.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208



## Use a CLI do Azure

Na CLI do Azure, você pode usar a lista de blob de armazenamento azcommand para visualizar os blobs em um contêiner. Este comando itera recursivamente por meio de qualquer pasta no contêiner. O exemplo abaixo lista os blobs enviados anteriormente para o contêiner de imagens:

```
az storage blob list \ --account-name contosodata \ --container-name "imagens"
```

## Use o Azure PowerShell

No Azure PowerShell, execute o [Get-AzStorageBlob](#) cmdlet, conforme ilustrado no exemplo a seguir:

```
az storage blob list \  
  --account-name contosodata \  
  --container-name "images"
```

## Baixe um blob de um contêiner

Você pode recuperar um blob do Armazenamento do Azure e salvá-lo em um arquivo local em seu computador.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF: 00296691208

## Use o portal do Azure

Se você estiver usando o portal do Azure, vá para a página da sua conta de armazenamento e selecione Contêineres sob Serviço de blob. No página, selecione o contêiner que contém seus blobs. Se o contêiner tiver uma estrutura de pastas, vá para a pasta que contém os blobs dos quais deseja fazer download. Selecione o blob para ver seus detalhes. Na página de detalhes, selecione Baixar.

The screenshot shows the Azure Storage Blob service interface. On the left, there's a list of blobs in the 'images' container, including 'racer\_green\_large.gif' which is selected. On the right, the details for this blob are shown in a modal window. The 'Download' button is highlighted with a red box. The blob properties include:

| Property                  | Value   |
|---------------------------|---|
| URL                       | <a href="https://contosostorage.blob.core.windows.net/images/bikes/racer_green_large.gif">https://contosostorage.blob.core.windows.net/images/bikes/racer_green_large.gif</a> |
| LAST MODIFIED             | 9/29/2020, 5:52:58 PM   |
| CREATION TIME             | 9/29/2020, 5:52:58 PM   |
| VERSION ID                | -   |
| TYPE                      | Block blob  |
| SIZE                      | 65.28 KB  |
| ACCESS TIER               | Hot (infmed)  |
| ACCESS TIER LAST MODIFIED | N/A   |
| SERVER ENCRYPTED          | true  |
| ETAG                      | 0x80064901EC70035   |
| CONTENT-TYPE              | image/gif   |
| CONTENT-MD5               | f8ykm3/yb5++zluOFslA==  |
| LEASE STATUS              | Unlocked  |
| LEASE STATE               | Available   |
| LEASE DURATION            | -   |
| COPY STATUS               | -   |
| COPY COMPLETION TIME      | -   |

## Use a CLI do Azure

A CLI do Azure fornece o download de blob de armazenamento az e lote de download de blob de armazenamento azcomandos. Esses comandos são análogos aos disponíveis para upload de blobs. O exemplo abaixo recupera o "racer\_green" bolha do bicicletas na pasta imagens .

```
az storage blob list \
    --account-name contosodata \
    --container-name "images"
```

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

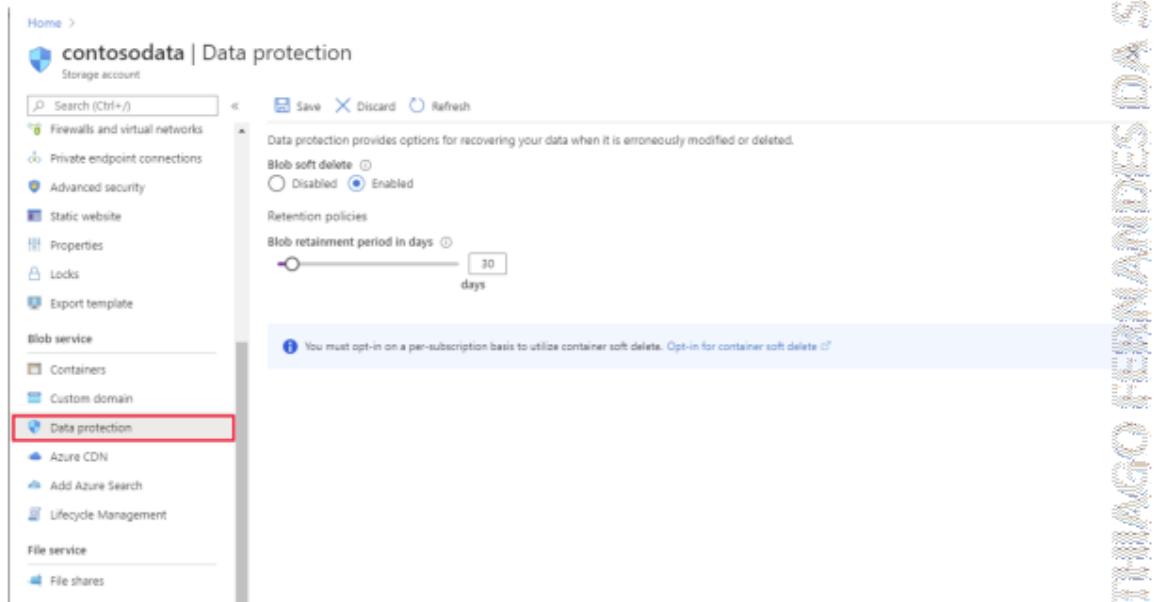
## Use o Azure PowerShell

No Azure PowerShell, use o Get-AzStorageBlobContent cmdlet.

```
Get-AzStorageAccount  
  -ResourceGroupName "contoso-group"  
  -Name "contosodata" | Get-AzStorageBlob  
    -Container "images"
```

## Excluir um blob de um contêiner

A exclusão de um blob pode recuperar os recursos usados no contêiner de armazenamento. No entanto, se você habilitou a opção de exclusão reversível para a conta de armazenamento, o blob é oculto em vez de removido e você pode restaurá-lo posteriormente. Você pode habilitar ou desabilitar a exclusão reversível no portal do Azure e especificar por quanto tempo o blob é retido. Selecione Proteção de dados sob Serviço de blob. Se o blob não for restaurado até o final do período de retenção, ele será removido do armazenamento.

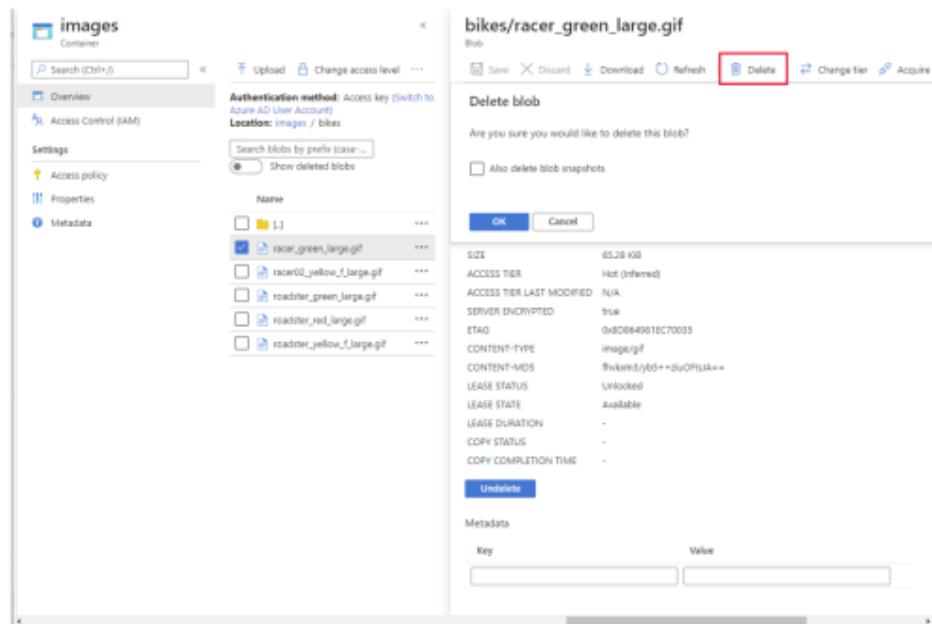


THIAGO FERNANDES DA SILVA  
CPF 00296691208

**ATENÇÃO:** Se você criou a conta de armazenamento com suporte para namespaces hierárquicos (para Data Lake Storage), a opção de exclusão reversível não está disponível. Todas as operações de exclusão de blob serão finais.

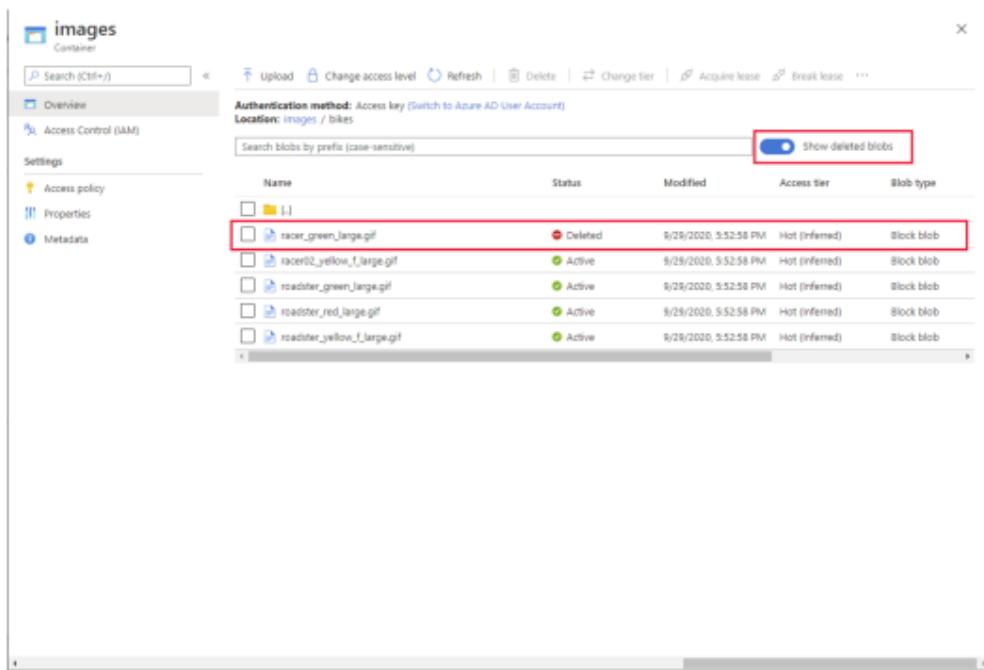
## Use o portal do Azure

Se você estiver usando o portal do Azure, vá para a página da sua conta de armazenamento e selecione Contêineres sob Serviço de blob. Na página, selecione o contêiner que contém seus blobs. Se o contêiner tiver uma estrutura de pastas, vá para a pasta que contém os blobs dos quais deseja fazer download. Selecione o blob para ver seus detalhes. Na página de detalhes, selecione Excluir. Você será solicitado a confirmar a operação.



Se você habilitou a exclusão reversível para a conta de armazenamento, a página de blobs que lista os blobs em um contêiner inclui a opção Mostrar blobs excluídos. Se você selecionar esta opção, você pode ver e restaurar um blob excluído.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208



| Name                        | Status  | Modified              | Access tier    | Blob type  |
|-----------------------------|---------|-----------------------|----------------|------------|
| racer_green_large.gif       | Deleted | 8/28/2020, 5:52:58 PM | Hot (inferred) | Block blob |
| racer02_yellow_t_large.gif  | Active  | 8/28/2020, 5:52:58 PM | Hot (inferred) | Block blob |
| roadster_green_large.gif    | Active  | 8/28/2020, 5:52:58 PM | Hot (inferred) | Block blob |
| roadster_red_large.gif      | Active  | 8/28/2020, 5:52:58 PM | Hot (inferred) | Block blob |
| roadster_yellow_t_large.gif | Active  | 8/28/2020, 5:52:58 PM | Hot (inferred) | Block blob |

## Use a CLI do Azure

Você pode excluir um único blob com o comando `az storage blob delete`, ou um conjunto de blobs com o comando `az storage blob delete-batch`. O comando abaixo remove a bolha racer green da pasta bicicletas em imagens do contêiner:

```
az storage blob download \
    --container-name images \
    --account-name contosodata \
    --file "racer_green_large.gif" \
    --name "bikes\racer_green"
```

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

CNPJ 00296691208

## Use o Azure PowerShell

Use o Remove-AzStorageBlobcmdlet para excluir um blob de armazenamento do Azure PowerShell. Por padrão, a exclusão é silenciosa. Você pode adicionar o -Confirmador para solicitar que o usuário confirme se realmente deseja excluir o blob:

```
Get-AzStorageAccount  
  -ResourceGroupName "contoso-group"  
  -Name "contosodata" | Get-AzStorageBlobContent  
    -Container "images"  
    -Blob "bikes\racer_green_large.gif"  
    -Destination "racer_green_large.gif"
```

## Excluir um contêiner de armazenamento do Azure

A remoção de um contêiner exclui automaticamente todos os blobs mantidos naquele contêiner. Se você não tiver cuidado, pode perder muitos dados.

## Use o portal do Azure

No portal do Azure, selecione Contêiners sob Serviço de blob, selecione o contêiner a ser excluído e, em seguida, selecione Excluir na barra de ferramentas.

## Use a CLI do Azure

Na CLI do Azure, use o deletar contêiner de armazenamento azcomando. O exemplo a seguir exclui o imagens container referenciado nos exemplos anteriores.

```
az storage container delete \  
  --account-name contosodata \  
  --name "images"
```

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

## Use o Azure PowerShell

O Remove-AzStorageContainer cmdlet exclui um contêiner de armazenamento. O ConfirmMesinalizador solicita que o usuário confirme a operação de exclusão. O código abaixo mostra um exemplo:

```
Get-AzStorageAccount  
    -ResourceGroupName "contoso-group"  
    -Name "contosodata" | Remove-AzStorageContainer  
        -Name "images"  
        -Confirm
```

## Gerenciar o armazenamento de arquivos do Azure

Você pode usar o armazenamento de arquivos do Azure para armazenar arquivos compartilhados. Os usuários podem se conectar a uma pasta compartilhada (também conhecida como compartilhamento de arquivo) e ler e gravar arquivos (se eles tiverem os privilégios apropriados) da mesma forma que usariam uma pasta em uma máquina local. No cenário da Contoso, o armazenamento de arquivos do Azure é usado para armazenar relatórios e documentação do produto que os usuários da empresa precisam ler.

Nesta unidade, você aprenderá como criar e gerenciar compartilhamentos de arquivos e carregar e baixar arquivos no armazenamento de arquivos do Azure.

**NOTA:** Os arquivos em um compartilhamento de arquivos tendem a ser tratados de maneira diferente dos blobs. Em muitos casos, os usuários simplesmente leem e gravam arquivos como se fossem objetos locais. Por esse motivo, embora a CLI do Azure e o Azure PowerShell forneçam acesso programático ao armazenamento de Arquivos do Azure, esta unidade se concentra nas ferramentas disponíveis no portal do Azure e no comando AzCopy .

## Crie um compartilhamento de arquivos

A Microsoft fornece duas ferramentas gráficas que você pode usar para criar e gerenciar compartilhamentos de arquivos no Armazenamento do Azure: o portal do Azure e o Explorador de Armazenamento do Azure.

CNPJ 000296691208

## Use o portal do Azure

Selecione Compartilhamento de arquivos no painel principal da Visão geral para uma conta de armazenamento do Azure, e também disponível no Serviço de arquivo da barra de comando:

The screenshot shows the Azure Storage Account overview page for 'contosodata'. On the left, there's a sidebar with sections like Blob service, File service (with 'File shares' highlighted by a red box), Table service, Queue service, Monitoring, and Alerts. The main content area displays details about the storage account, including its resource group, status (Primary: Available, Secondary: Available), location (UK South, UK West), subscription (Freelive), and a note about classic alerts being retired. Below this, there are three cards: 'Containers' (scalable, cost-effective storage for unstructured data), 'File shares' (selected, showing 'Serverless SMB file shares'), and 'Tables' (tabular data storage). A vertical watermark 'THIAGO FERNANDES DA SILVA OLIVEIRA' and a horizontal watermark 'CPF 00296691208' are visible on the right side of the page.

**No Compartilhamento de arquivos página, selecione + Compartilhamento de arquivos.** Dê um nome ao compartilhamento de arquivos e, opcionalmente, especifique uma cota. O Azure permite que você armazene até 5 PiB de arquivos em todos os compartilhamentos de arquivos em uma conta de armazenamento. Uma cota permite que você limite a quantidade de espaço que um compartilhamento de arquivo individual consome, para evitar que outros compartilhamentos de arquivo fiquem sem armazenamento. Se você tiver apenas um compartilhamento de arquivo, pode deixar a cota vazia.

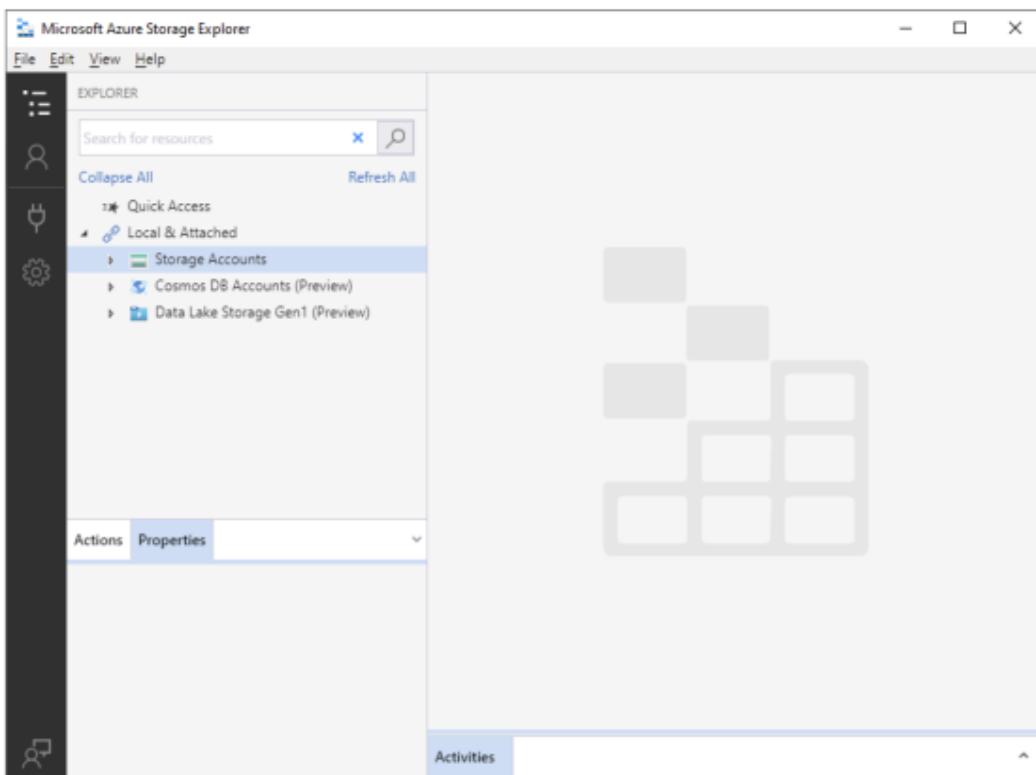
The screenshot shows the Azure Storage blade for the 'contosodata' storage account. Under the 'File service' section, 'File shares' is selected. A 'New file share' dialog is open, prompting for a 'Name' (set to 'reports') and a 'Quota' (set to 2500 GB). The main pane shows a table with columns 'Name' and 'Modified', currently displaying 'No results'.

Depois de criar um compartilhamento, você pode usar o portal do Azure para adicionar diretórios ao compartilhamento, carregar arquivos para o compartilhamento e excluí-lo. O comando **Conectar** gera um script PowerShell que você pode executar para anexar ao compartilhamento de seu computador local. Você pode então usar o compartilhamento como se fosse uma unidade de disco local.

The screenshot shows the 'reports' file share blade. In the top navigation bar, the 'Connect' button is highlighted with a red box. The 'Connect' dialog is open, specifically for the Windows tab. It displays a warning about SMB encryption and provides PowerShell commands to connect from a Windows terminal. The commands involve running Test-NetConnection and cmdkey commands to establish a connection via port 445.

## Use o Azure Storage Explorer

O Explorador de Armazenamento do Azure é um utilitário que permite gerenciar contas de Armazenamento do Azure de seu computador desktop. Você pode baixá-lo do Azure Storage Explorer no site da Microsoft. Você pode usar o Storage Explorer para criar contêineres de blob e compartilhamentos de arquivos, bem como fazer upload e download de arquivos.



Uma versão desse utilitário também está disponível no portal do Azure, em Visão geral para uma conta de armazenamento do Azure.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

The screenshot shows the Azure Storage Explorer (preview) interface. On the left, there's a sidebar with navigation links like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Data transfer, Events, and Storage Explorer (preview). Below that is a 'Settings' section with options such as Access keys, Geo-replication, CORS, Configuration, Encryption, Shared access signature, Firewalls and virtual networks, Private endpoint connections, Advanced security, and Static website. The main area displays a large, stylized icon of a grid of squares.

Para criar um novo compartilhamento de arquivos, clique com o botão direito Compartilhamentos de arquivos e selecione Criar compartilhamento de arquivo. No portal do Azure, o Storage Explorer exibe a mesma caixa de diálogo que você viu anteriormente. Na versão desktop, você simplesmente insere um nome para o novo compartilhamento de arquivos; você não tem a opção de definir uma cota neste ponto.

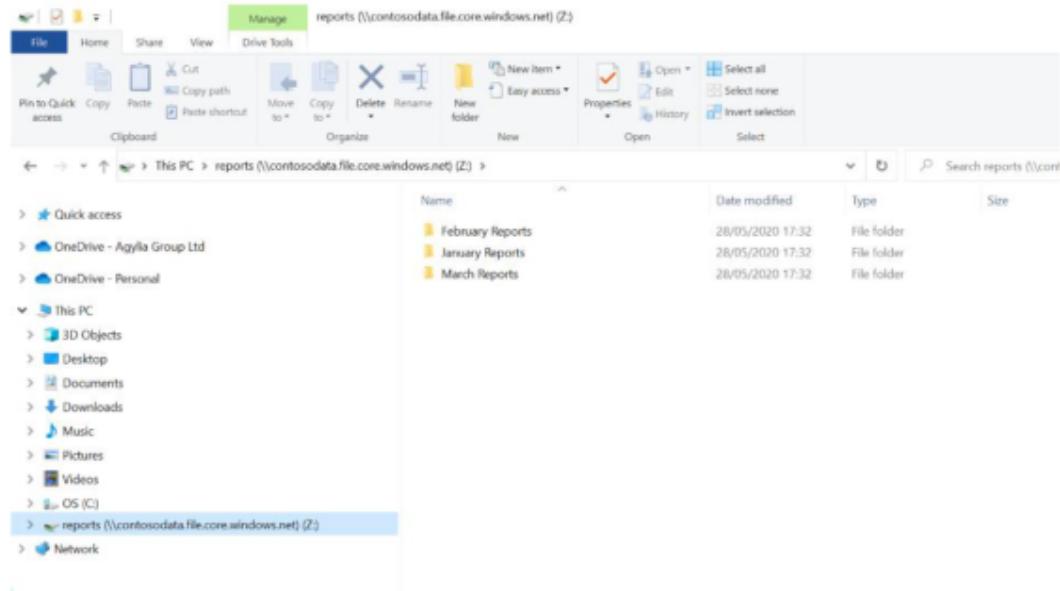
Assim como no portal do Azure, depois de criar um novo compartilhamento, você pode usar o Gerenciador de Armazenamento para criar pastas e carregar e baixar arquivos.

The screenshot shows the Microsoft Azure Storage Explorer desktop application. The left pane is the 'EXPLORER' view, showing a tree structure of storage accounts, blob containers, file shares, and tables. A folder named 'complaints' is selected. The right pane shows a list of files and folders within 'complaints', including 'delivery complaints' and 'product complaints'. At the top, there's a toolbar with buttons for Upload, Download, Open, and New Folder, where the 'New Folder' button is highlighted with a red box. Other buttons include Refresh All, Select All, Copy, Paste, Clone with New Name, Connect VM, Delete, and View.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Upload e download de arquivos

Você pode carregar e baixar arquivos individuais de e para o armazenamento de arquivos do Azure manualmente, usando o Storage Explorer, o portal do Azure ou conectando o compartilhamento de arquivos ao seu computador desktop e arrastando e soltando arquivos no File Explorer.



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

No entanto, se você precisar transferir um número significativo de arquivos dentro e fora do armazenamento de arquivos do Azure, deverá usar o utilitário AzCopy. AzCopy é um utilitário de linha de comando otimizado para transferir arquivos grandes (e blobs) entre o computador local e o armazenamento de arquivos do Azure. Ele pode detectar falhas de transferência e reiniciar uma transferência com falha no ponto em que ocorreu um erro - você não precisa repetir a operação inteira.

### Gerar um token SAS

Antes de usar o AzCopy, você gera um Token de assinatura de acesso compartilhado (SAS). Um token SAS fornece acesso controlado, limitado no tempo e anônimo a serviços e recursos em uma conta de armazenamento; os usuários não precisam fornecer credenciais adicionais. Os tokens SAS são úteis em situações em que você não sabe com antecedência quais usuários precisarão de acesso aos seus recursos.

**NOTA:** O comando AzCopy também oferece suporte à autenticação usando o Azure Active Directory, mas essa abordagem exige a adição de todos os seus usuários ao Azure Active Directory primeiro.

Você pode criar um token SAS para se conectar ao armazenamento de arquivos do Azure usando o portal do Azure. Na página da sua conta de armazenamento, em Configurações, selecione Assinatura de acesso compartilhado. Em Assinatura de acesso compartilhado, sob Serviços permitidos, selecione Arquivo. Sob Tipos de recursos permitidos, selecione Contêiner e Objeto. Sob Permissões, selecione os privilégios que deseja conceder aos usuários. Defina o horário de início e término do token SAS e especifique o intervalo de endereços IP dos computadores que seus usuários usarão. Selecione Gerar SAS e string de conexão para criar o token SAS. Copie o valor noToken SAS campo em algum lugar seguro.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

The screenshot shows the Azure Storage Account settings for 'contosodata'. The 'Shared access signature' section is selected. It includes fields for Allowed services (Blob, File checked), Allowed resource types (Container, Object checked), Allowed permissions (Read, Write, Delete, List, Add, Create, Update, Process), Blob versioning permissions (Enabled checked), Start and expiry date/time (Start: 05/29/2020, End: 05/29/2020), Allowed IP addresses (example: 168.1.5.65 or 168.1.5.65-168.1.5.70), Allowed protocols (HTTPS only checked), and a Signing key (key1). Below these, there are two text fields: 'Connection string' and 'SAS token'. The 'SAS token' field is highlighted with a red box and contains a long URL. The 'File service SAS URL' field below it also contains a similar URL.

## Fazer upload de arquivos

Para transferir um único arquivo para o armazenamento de arquivos usando AzCopy, use a forma do comando mostrado no exemplo a seguir. Execute este comando a partir da linha de comando. Neste exemplo, substitua <storage-account-name> pelo nome da conta de armazenamento, substitua <file-share> pelo nome de um compartilhamento de arquivos nesta conta e substitua <SAS-token> com o token que você criou usando o portal do Azure. Você deve incluir as aspas onde mostrado.

**NOTA:** Não se esqueça de incluir cópia de palavra-chave após o comando azcopy. O AzCopy oferece suporte a outras operações, como exclusão de arquivos e blobs, listagem de arquivos e blobs e criação de novos compartilhamentos de arquivos. Cada uma dessas operações possui sua própria palavra-chave.

```
azcopy copy "myfile.txt" "https://<storage-account-name>.file.core.windows.net/<file-share-name>/myfile.txt<SAS-token>"
```

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Você pode transferir todo o conteúdo de uma pasta local para o armazenamento de arquivos do Azure usando um comando semelhante. Você substitui o nome do arquivo ("myfile.txt") pelo nome da pasta. Se a pasta contiver subpastas que você deseja copiar, adicione o -Recurciva bandeira.

```
azcopy copy "myfolder" "https://<storage-account-name>.file.core.windows.net/<file-share-name>/myfolder<SAS-token>" --recursive
```

Conforme o processo é executado, o AzCopy exibe um relatório de progresso:

```
INFO: Scanning...
INFO: Any empty folders will be processed, because source and destination both support folders

Job b86eeb8b-1f24-614e-6302-de066908d4a2 has started
Log file is located at: C:\Users\User\azcopy\b86eeb8b-1f24-614e-6302-de066908d4a2.log

11.5 %, 126 Done, 0 Failed, 48 Pending, 0 Skipped, 174 Total, 2-sec Throughput (Mb/s): 8.2553
```

Quando a transferência for concluída, você verá um resumo do trabalho realizado.

```
Job b86eeb8b-1f24-614e-6302-de066908d4a2 summary
Elapsed Time (Minutes): 0.6002
Number of File Transfers: 161
Number of Folder Property Transfers: 13
Total Number of Transfers: 174
Number of Transfers Completed: 174
Number of Transfers Failed: 0
Number of Transfers Skipped: 0
TotalBytesTransferred: 43686370
Final Job Status: Completed
```

O comando AzCopy tem outras opções também. Para mais informações, consulte a página "Fazer upload de arquivos" no site da Microsoft.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF: 00296691208

## Download de arquivos

Você também pode usar o AzCopy para transferir arquivos e pastas do Armazenamento de Arquivos do Azure para seu computador local. O comando é semelhante ao de upload de arquivos, exceto que você muda a ordem dos argumentos; especifique os arquivos e pastas no compartilhamento de arquivos primeiro, e depois os arquivos e pastas locais. Por exemplo, para baixar os arquivos de uma pasta chamada *minha pasta* em um compartilhamento de arquivo chamado *minha parte* para uma pasta local chamada *pasta local*, use o seguinte comando:

```
azcopy copy "https://<storage-account-name>.file.core.windows.net/myshare/myfolder<SAS-token>" "localfolder" --recursive
```

Para obter detalhes completos sobre o download de arquivos usando o AzCopy, consulte Download de arquivos.

## Laboratório: faça upload, download e consulte dados em um armazenamento de dados não relacional

No cenário de amostra, suponha que você tenha criado os seguintes armazenamentos de dados:

- Um Cosmos DB para conter informações sobre os produtos que a Contoso fabrica.
- Um contêiner de blob no Armazenamento do Azure para conter as imagens dos produtos.
- Um compartilhamento de arquivo, na mesma conta de Armazenamento do Azure, para manter a documentação do produto.

Neste laboratório, você fará upload de dados para esses armazenamentos de dados. Você executará consultas nos dados do banco de dados Cosmos DB. Por fim, você baixará e visualizará as imagens e documentos mantidos no Armazenamento do Azure.

Vou ao Exercício: Upload, download e consulta de dados em um armazenamento de dados não relacional módulo no Microsoft Learn e siga as instruções do módulo.

Você realizará este exercício usando o portal do Azure e a linha de comando.

CURSO DE AZURE DATA FUNDAMENTALS

CPF: 00296691208

## Resumo

Nesta lição, você viu como usar as contas do Azure Cosmos DB e do Armazenamento do Azure para armazenar e recuperar dados não relacionais. Você aprendeu como:

- Carregar dados em um banco de dados Cosmos DB e consulte esses dados.
- Carregar e baixe dados em uma conta de armazenamento do Azure.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Saber mais

- Casos de uso comuns do Azure Cosmos DB
- Migrar o esquema de banco de dados normalizado do Banco de Dados SQL do Azure para o contêiner desnormalizado do Azure CosmosDB
- Tutorial: Use a ferramenta de migração de dados para migrar seus dados para o Azure Cosmos DB
- Copie e transforme dados no Azure Cosmos DB (API SQL) usando o Azure Data Factory
- Início rápido: crie um aplicativo de console usando o SDK do .NET V4 para gerenciar os recursos da conta da API SQL do Azure Cosmos DB
- Primeiros passos com consultas SQL
- Criar contêiner de armazenamento az
- New-AzStorageContainer
- o upload do blob de armazenamento az
- Set-AzStorageBlobContent
- lista de blob de armazenamento az
- Get-AzStorageBlob
- Download de blob de armazenamento az
- Lote de download de blob de armazenamento az
- Get-AzStorageBlobContent
- Deletar blob de armazenamento az
- Remove-AzStorageBlob
- Apagar lote de blob de armazenamento az
- Deletar contêiner de armazenamento az
- Remove-AzStorageContainer
- Comece com o AzCopy
- Azure Storage Explorer
- Transfira dados com AzCopy e armazenamento de arquivos

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Módulo 4: Explore análises modernas de data Wharehouse no Azure

### Introdução

A maioria das organizações possui vários armazenamentos de dados, geralmente com estruturas e formatos diferentes. Eles geralmente têm fluxos de dados de entrada ao vivo, como dados de sensores, cuja análise pode ser cara. Geralmente, há uma abundância de informações úteis disponíveis fora das organizações. Essas informações podem ser combinadas com dados locais para adicionar insights e enriquecer o entendimento. Ao combinar todos os dados locais com informações externas úteis, muitas vezes é possível obter insights sobre os dados que não eram possíveis anteriormente. O processo de combinação de todas as fontes de dados locais é conhecido como data warehousing. O processo de análise de streaming de dados e dados da Internet é conhecido como análise de Big Data. O Azure Synapse Analytics combina armazenamento de dados com análises de Big Data.

Suponha que você seja um engenheiro de dados trabalhando na Contoso, uma organização com uma grande operação de manufatura. A organização deve reunir e armazenar informações de uma variedade de fontes, como dados em tempo real de monitoramento do status das máquinas da linha de produção, dados de controle de qualidade do produto, registros históricos de produção, volumes de produtos em estoque e dados de estoque de matérias-primas. Essas informações são críticas para o funcionamento da organização. Você foi solicitado a determinar a melhor forma de armazenar essas informações, para que possam ser analisadas rapidamente e consultadas facilmente.

### Objetivos de aprendizado

Nesta lição, você irá:

- Explorar os conceitos de armazenamento de dados
- Explorar os serviços de dados do Azure para armazenamento de dados moderno
- Explorar a arquitetura moderna de data warehouse e a carga de trabalho
- Explorar os serviços de dados do Azure no portal do Azure

TITULO DO COFFEE BREAK 1208

## Descreva o armazenamento de dados moderno

Um data warehouse reúne dados de muitas fontes diferentes dentro de uma organização. Esses dados são então usados como fonte para análise, relatórios e processamento analítico online (OLAP). O foco de um data warehouse é fornecer respostas a consultas complexas, ao contrário de um banco de dados relacional tradicional, que se concentra no desempenho transacional.

Armazéns de dados precisam lidar big data. Big data é o termo usado para grandes quantidades de dados coletados em volumes crescentes, em velocidades mais altas e em uma variedade maior de formatos do que nunca. Pode ser histórico (ou seja, armazenado) ou em tempo real (ou seja, transmitido da fonte). Normalmente, as empresas dependem de seus big data para ajudar a tomar decisões críticas de negócios.

## O que é armazenamento de dados moderno?

Um data warehouse moderno pode conter uma mistura de dados relacionais e não relacionais, incluindo arquivos, fluxos de mídia social e dados do sensor da Internet das Coisas (IoT). O Azure fornece uma coleção de serviços que você pode usar para criar uma solução de data warehouse, incluindo Azure Data Factory, Azure Data Lake Storage, Azure Databricks, Azure Synapse Analytics e Azure Analysis Services. Você pode usar ferramentas como o Power BI para analisar e visualizar os dados, gerando relatórios, gráficos e painéis.

O vídeo a seguir descreve os componentes comumente usados para criar um data warehouse e como os dados podem fluir por eles. Este vídeo mostra uma abordagem particular.

THIAGO FERNANDES  
CPF 00296691208

[vídeo]



A próxima unidade descreve cada um desses serviços com um pouco mais de detalhes.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Combine processamento em lote e fluxo

Um negócio típico de grande escala requer uma combinação de dados atualizados e informações históricas. Os dados atualizados podem ser usados para ajudar a monitorar processos de fabricação críticos em tempo real, onde uma decisão instantânea é necessária. Outros exemplos incluem fluxos de dados do mercado de ações, onde os preços atuais são necessários para tomar decisões informadas de compra ou venda em frações de segundo.

Os dados históricos são igualmente importantes para dar a uma empresa uma visão mais estabilizada das tendências de desempenho. Uma organização de manufatura exigirá informações como os volumes de vendas por produtos ao longo de um mês, um trimestre ou um ano, para determinar se deve continuar a produzir vários itens ou se deve aumentar ou diminuir a produção de acordo com as flutuações sazonais. Esses dados históricos podem ser gerados por processos em lote em intervalos regulares, com base nos dados de vendas ao vivo que podem ser capturados continuamente.

Qualquer solução de data warehouse moderna deve ser capaz de fornecer acesso aos fluxos de dados raw, e informações cooked de negócios derivadas desses dados.

## Explore os serviços de dados do Azure para armazenamento de dados moderno

Como engenheiro de dados que trabalha em uma organização com uma grande operação de manufatura, você deseja entender mais sobre os componentes que formam um data warehouse moderno. Essas informações o ajudarão a determinar quais elementos atendem melhor aos requisitos de sua organização.

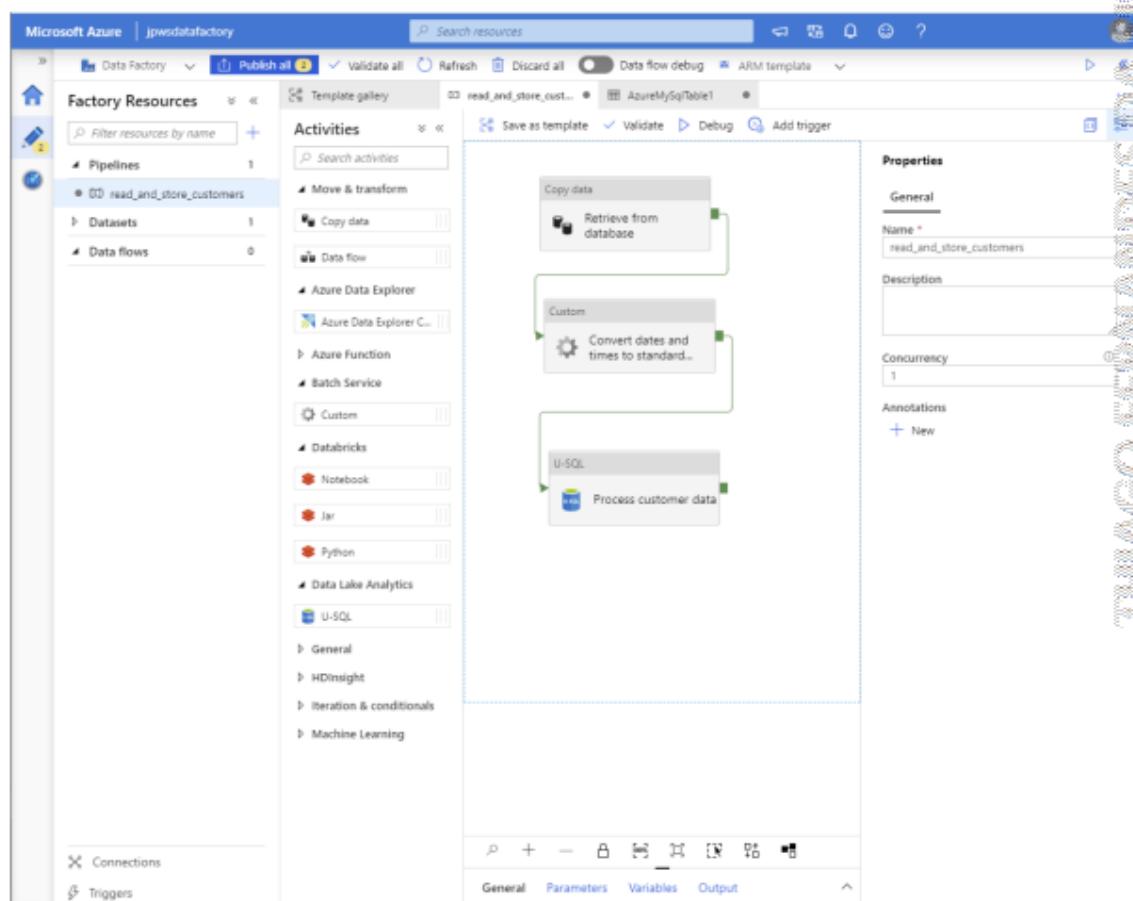
Nesta unidade, você aprenderá mais sobre os serviços de dados que o Azure fornece. Esses serviços permitem combinar dados de várias fontes, reformatá-los em modelos analíticos e salvar esses modelos para consultas, relatórios e visualizações subsequentes.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## O que é Azure Data Factory?

Azure Data Factory é descrito como um serviço de integração de dados. O objetivo do Azure Data Factory é recuperar dados de uma ou mais fontes de dados e convertê-los em um formato que você processa. As fontes de dados podem apresentar dados de maneiras diferentes e conterá que você precisa filtrar. O Azure Data Factory permite que você extraia os dados interessantes e descarte o restante. Os dados podem não estar em um formato adequado para processamento por outros serviços em sua solução de warehouse, portanto, você pode transformá-los. Por exemplo, seus dados podem conter datas e horas formatadas de maneiras diferentes em fontes de dados diferentes. Você pode usar o Azure Data Factory para transformar esses itens em uma única estrutura uniforme. O Azure Data Factory pode gravar os dados ingeridos em um armazenamento de dados para processamento subsequente.

Você define o trabalho executado pelo Azure Data Factory como um pipeline de operações. Um pipeline pode ser executado continuamente, à medida que os dados são recebidos de várias fontes de dados. Você pode criar pipelines usando a interface gráfica do usuário fornecida pela Microsoft ou escrevendo seu próprio código. A imagem abaixo mostra o editor de pipeline no Azure Data Factory.



CIFP 00296691208

### O que é Azure Data Lake Storage?

Um data lake é um repositório para grandes quantidades de dados brutos. Como os dados são brutos e não processados, é muito rápido carregar e atualizar, mas os dados não foram colocados em uma estrutura adequada para uma análise eficiente. Você pode pensar em um data lake como um ponto de preparação para seus dados ingeridos, antes de serem massageados e convertidos em um formato adequado para realizar análises.

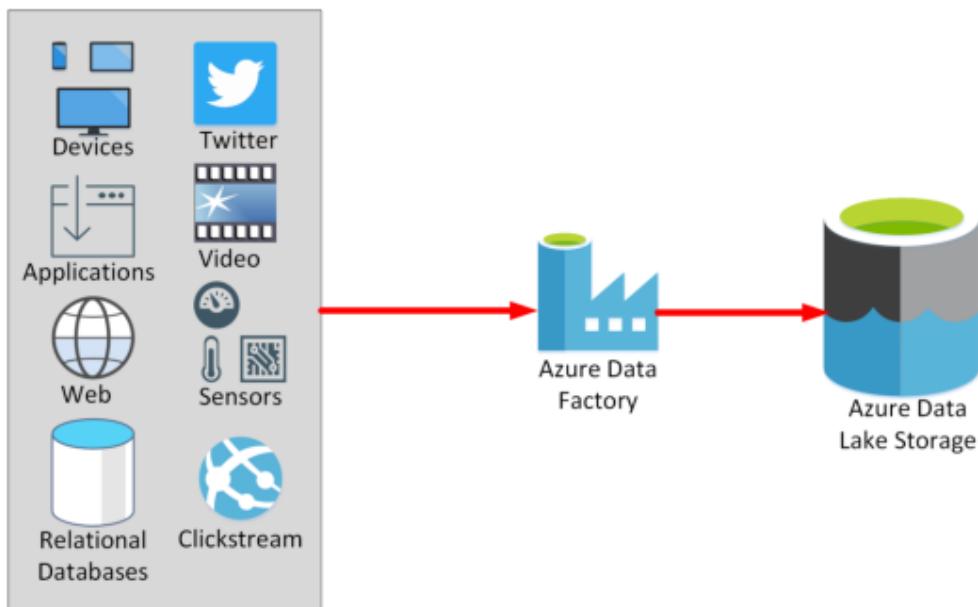
**NOTA:** Um data warehouse também armazena grandes quantidades de dados, mas os dados em um warehouse foram processados para convertê-los em um formato para análise eficiente. Um data lake contém dados não tratados, mas um data warehouse contém informação estruturada.

O Azure Data Lake Storage combina a estrutura hierárquica de diretórios e a semântica do sistema de arquivos de um sistema de arquivos tradicional com a segurança e escalabilidade fornecidas pelo Azure. O Azure Data Lake Storage é essencialmente uma extensão do armazenamento de Blob do Azure, organizado como um sistema de arquivos quase infinito. Possui as seguintes características:

- O Data Lake Storage organiza seus arquivos em diretórios e subdiretórios para melhor organização de arquivos. O armazenamento de blob pode apenas simular uma estrutura de diretório.
- O Data Lake Storage oferece suporte ao arquivo POSIX (Portable Operating System Interface) e às permissões de diretório para habilitar o controle de acesso baseado em função (RBAC) granular em seus dados.
- O Armazenamento do Azure Data Lake é compatível com o Hadoop Distributed File System (HDFS). Hadoop é um serviço de análise altamente flexível e programável, usado por muitas organizações para examinar grandes quantidades de dados. Todos os ambientes Apache Hadoop podem acessar dados no Azure Data Lake Storage Gen2.

Em uma solução de data warehouse do Azure Data Services, os dados são normalmente carregados no Azure Data Lake Storage antes de serem processados em uma estrutura que permite uma análise eficiente no Azure Synapse Analytics. Você pode usar um serviço como o Azure Data Factory (descrito acima) para ingerir e carregar os dados de uma variedade de fontes no Armazenamento do Azure Data Lake.

CPF 000296691208



## O que é Azure Databricks?

Azure Databricks é um ambiente Apache Spark em execução no Azure para fornecer processamento de big data, streaming e aprendizado de máquina. Apache Spark é um mecanismo de processamento de dados altamente eficiente que pode consumir e processar grandes quantidades de dados muito rapidamente. Há um número significativo de bibliotecas do Spark que você pode usar para realizar tarefas como processamento de SQL, agregações e para construir e treinar modelos de aprendizado de máquina usando seus dados.

O Azure Databricks fornece uma interface gráfica do usuário onde você pode definir e testar seu processamento passo a passo, antes de enviá-lo como um conjunto de tarefas em lote. Você pode criar scripts Databricks e consultar dados usando linguagens como R, Python e Scala. Você escreve seu código Spark usando notebooks. Um caderno contém células, cada um contendo um bloco de código separado. Quando você executa um notebook, o código em cada célula é passado para o Spark para execução. A imagem abaixo mostra uma célula em uma pasta de trabalho que executa uma consulta e gera um gráfico.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

Adventure-Works (Scala)

jpscluster

File Edit View: Code Permissions Run All Clear

Cmd 1

```
1 import com.microsoft.azure.sqldb.spark.config.Config
2 import com.microsoft.azure.sqldb.spark.connect._
3
4 val config = Config(Map(
5   "url"          -> "[REDACTED].database.windows.net",
6   "databaseName" -> "adventure-works",
7   "QueryCustom"  -> "SELECT P.Name, SUM(S.OrderQty) As TotalSales FROM SalesLT.SalesOrderDetail S JOIN SalesLT.Product P ON S.ProductID = P.ProductID WHERE P.ProductSubcategoryID = 1 GROUP BY P.Name",
8   "user"         -> "[REDACTED]",
9   "password"     -> "[REDACTED]",
10  "connectTimeout" -> "5", //seconds
11  "queryTimeout"  -> "5" //seconds
12 ))
13
14 val sales = sqlContext.read.sqlDB(config)
15 display(sales)
```

(1) Spark Jobs

sales: org.apache.spark.sql.DataFrame = [Name: string, TotalSales: integer]

Azure Databricks também oferece suporte ao processamento de fluxo estruturado. Neste modelo, o Databricks executa seus cálculos de forma incremental e atualiza continuamente o resultado conforme os dados de streaming chegam.

## O que é o Azure Synapse Analytics?

O Azure Synapse Analytics é um mecanismo de análise. Ele foi projetado para processar grandes quantidades de dados muito rapidamente.

Usando o Synapse Analytics, você pode ingerir dados de fontes externas, como arquivos simples, Azure Data Lake ou outros sistemas de gerenciamento de banco de dados e, em seguida, transformar e agregar esses dados em um formato adequado para processamento analítico. Você pode realizar consultas complexas sobre esses dados e gerar relatórios, gráficos e tabelas.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Ler e transformar dados de uma fonte externa pode consumir recursos consideráveis. O Azure Synapse Analytics permite que você armazene os dados que leu e processou localmente, dentro do serviço (isso será descrito posteriormente). Essa abordagem permite que você consulte repetidamente os mesmos dados sem a sobrecarga de buscá-los e convertê-los todas as vezes. Você também pode usar esses dados como entrada para processamento analítico adicional, usando o Azure Analysis Services.

O Azure Synapse Analytics aproveita uma arquitetura de processamento paralelo maciço (MPP). Essa arquitetura inclui um nó de controle e um pool de nós de computação.

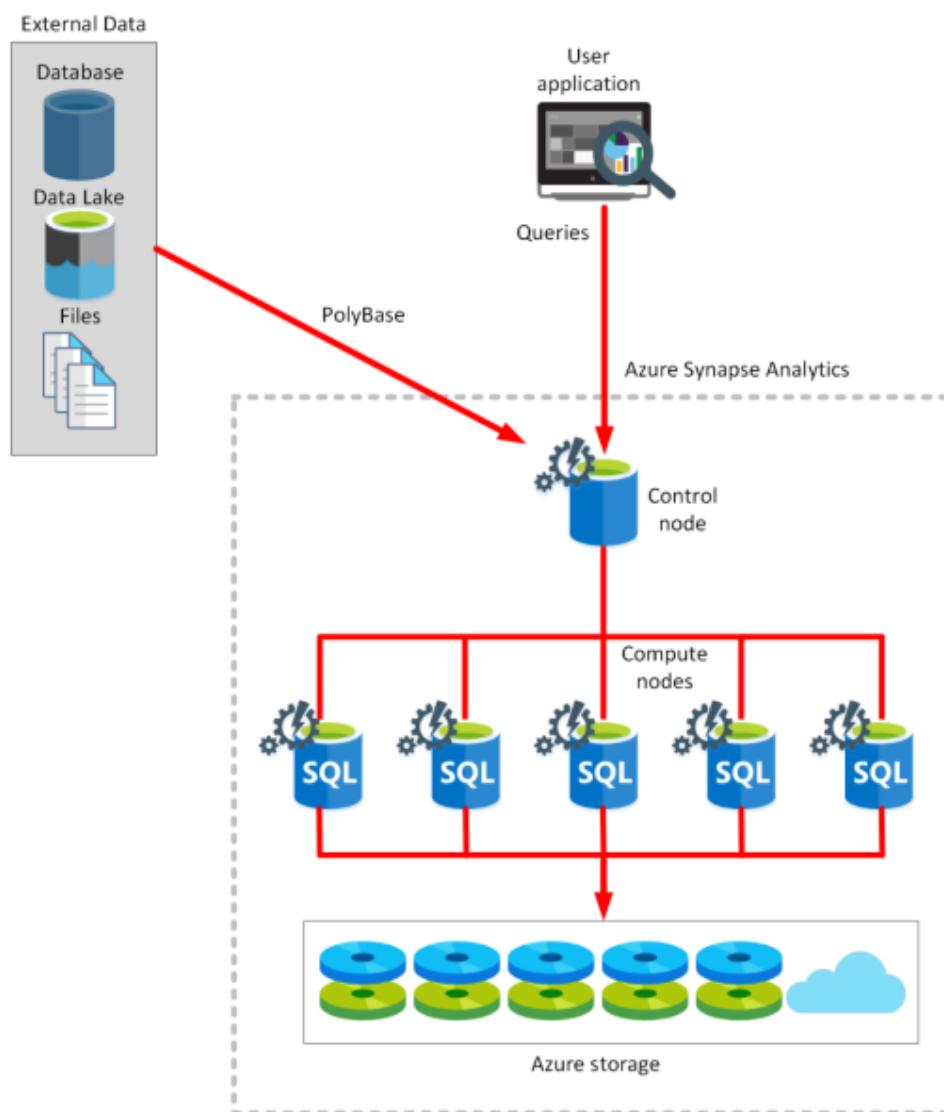
O controle nó é o cérebro da arquitetura. É o front end que interage com todos os aplicativos. O mecanismo MPP é executado no nó de controle para otimizar e coordenar consultas paralelas. Quando você envia uma solicitação de processamento, o nó de Controle a transforma em solicitações menores que são executadas em subconjuntos distintos de dados em paralelo.

o Calcular nós fornecem o poder computacional. Os dados a serem processados são distribuídos uniformemente pelos nós. Os usuários e aplicativos enviam solicitações de processamento para o nó de controle. O nó de controle envia as consultas aos nós de computação, que executam as consultas sobre a parte dos dados que cada um contém. Quando cada nó termina seu processamento, os resultados são enviados de volta ao nó de controle, onde são combinados em um resultado geral.

O Azure Synapse Analytics oferece suporte a dois modelos computacionais: pools SQL e pools Spark.

Em um pool SQL, cada nó de computação usa um Banco de Dados SQL do Azure e o Armazenamento do Azure para lidar com uma parte dos dados.

THIAGO FERNANDES  
CPF 00296691208



Você envia consultas na forma de instruções Transact-SQL e o Azure Synapse Analytics as executa. No entanto, ao contrário de um mecanismo de banco de dados SQL Server comum, o Azure Synapse Analytics pode receber dados de uma ampla variedade de fontes. Para fazer isso, o Azure Synapse Analytics usa uma tecnologia chamada PolyBase. O PolyBase permite que você recupere dados de fontes relacionais e não relacionais, como arquivos de texto delimitados, Azure Blob Storage e Azure Data Lake Storage. Você pode salvar os dados lidos como tabelas SQL no serviço Synapse Analytics.

Você especifica o número de nós ao criar um pool SQL. Você pode dimensionar o pool SQL manualmente para adicionar ou remover nós de computação conforme necessário.

**NOTA:** Você só pode dimensionar um pool SQL quando ele não está executando uma consulta Transact-SQL.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF: 00296691208

Em um pool Spark, os nós são substituídos por um cluster Spark. Você executa trabalhos do Spark compreendendo código escrito em Notebooks, da mesma forma que os Databricks do Azure. Você pode escrever o código para notebook em C #, Python, Scala ou Spark SQL (um dialeto de SQL diferente do Transact-SQL). Assim como acontece com um pool SQL, o cluster Spark divide o trabalho em uma série de tarefas paralelas que podem ser realizadas simultaneamente. Você pode salvar os dados gerados por seus blocos de anotações no Armazenamento do Azure ou no Data Lake Storage.

**NOTA:** Spark é otimizado para processamento na memória. Um trabalho do Spark pode carregar e armazenar dados em cache na memória e consultá-los repetidamente. A computação na memória é muito mais rápida do que os aplicativos baseados em disco, mas requer recursos de memória adicionais.

Você especifica o número de nós ao criar o cluster Spark. Os Spark pools podem ter escalonamento automático habilitado, para que os pools sejam escalonados adicionando ou removendo nós conforme necessário. O escalonamento automático pode ocorrer enquanto o processamento está ativo.

**NOTA:** Azure Synapse Analytics pode consumir muitos recursos. Se você não planeja realizar nenhum processamento por um tempo, pode pausar o serviço. Esta ação libera os recursos do pool para outros usuários e reduz seus custos.

### O que é o Azure Analysis Services?

O Azure Analysis Services permite que você crie modelos tabulares para oferecer suporte a consultas de processamento analítico online (OLAP). Você pode combinar dados de várias fontes, incluindo Banco de Dados SQL do Azure, Azure Synapse Analytics, armazenamento do Azure Data Lake, Azure Cosmos DB e muitos outros. Você usa essas fontes de dados para construir modelos que incorporam seu conhecimento de negócios. Um modelo é essencialmente um conjunto de consultas e expressões que recuperam dados de várias fontes de dados e geram resultados. Os resultados podem ser armazenados em cache na memória para uso posterior ou podem ser calculados dinamicamente, diretamente das fontes de dados subjacentes.

O Analysis Services inclui um designer gráfico para ajudá-lo a conectar fontes de dados e definir consultas que combinam, filtram e agregam dados. Você pode explorar esses dados de dentro do Analysis Services ou pode usar uma ferramenta como o Microsoft Power BI para visualizar os dados apresentados por esses modelos.

## Compare Analysis Services com Synapse Analytics

O Azure Analysis Services tem uma sobreposição funcional significativa com o Azure Synapse Analytics, mas é mais adequado para processamento em escala menor.

Use o Azure Synapse Analytics para:

- Volumes de dados muito altos (conjuntos de dados de vários terabytes a petabytes).
- Consultas e agregações muito complexas.
- Mineração de dados e exploração de dados.
- Operações ELT complexas. *ELT* apoia *Extrair, transformar e carregare* referindo-se à maneira como você pode recuperar dados brutos de várias fontes, converter esses dados em um formato padrão e armazená-los.
- Concorrência baixa a média (128 usuários ou menos).

Use o Azure Analysis Services para:

- Volumes de dados menores (alguns terabytes).
- Várias fontes que podem ser correlacionadas.
- Alta simultaneidade de leitura (milhares de usuários).
- Análise detalhada e pesquisa de dados, usando funções no Power BI.
- Desenvolvimento rápido de painel a partir de dados tabulares.

## Combine Analysis Services com Synapse Analytics

Muitos cenários podem se beneficiar do uso do Synapse Analytics e do Analysis Services juntos. Se você tiver grandes quantidades de dados ingeridos que requerem pré-processamento, pode usar o Synapse Analytics para ler esses dados e manipulá-los em um modelo que contém informações de negócios em vez de uma grande quantidade de dados brutos. A escalabilidade do Synapse Analytics oferece a capacidade de processar e reduzir muitos terabytes de dados em um conjunto de dados menor e sucinto que resume e agrupa muitos desses dados. Você pode então usar o Analysis Services para realizar interrogatórios detalhados dessas informações e visualizar os resultados dessas consultas com o Power BI.

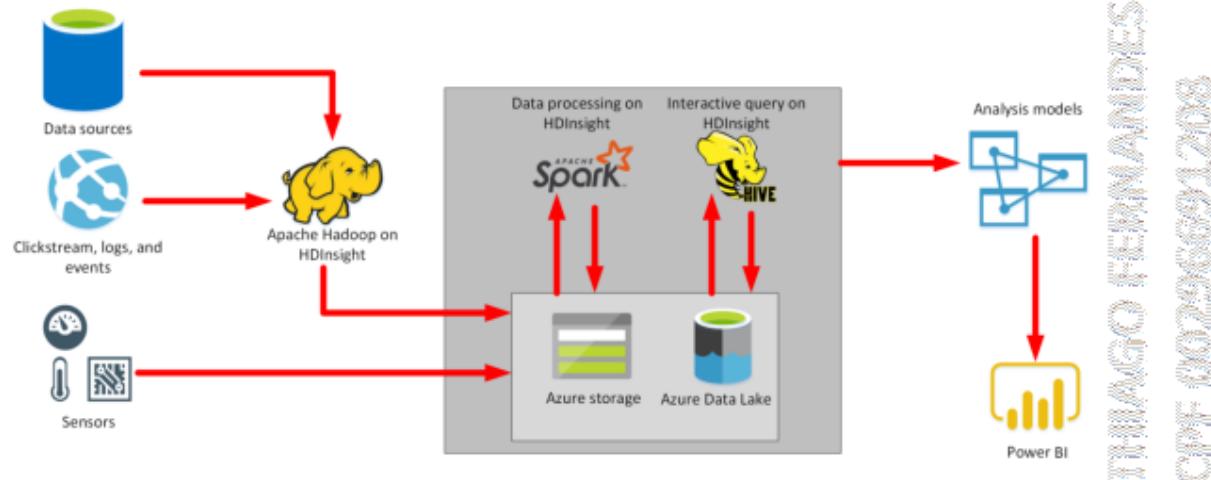
## O que é Azure HDInsight?

Azure HDInsight é um serviço de processamento de big data, que fornece a plataforma para tecnologias como Spark em um ambiente Azure. O HDInsight implementa um modelo em cluster que distribui o processamento em um conjunto de computadores. Este modelo é semelhante ao usado pelo Synapse Analytics, exceto que os nós estão executando o mecanismo de processamento Spark em vez do Banco de Dados SQL do Azure.

Você pode usar o Azure HDInsight em conjunto com, ou em vez do Azure Synapse Analytics. Bem como Spark, HDInsight oferece suporte a tecnologias de streaming, como Apache Kafka e o modelo de processamento Apache Hadoop. A imagem abaixo mostra onde você pode usar os componentes do HDInsight em uma solução de armazenamento de dados.

**NOTA:** Nesta imagem, Hadoop é uma estrutura de código aberto que divide grandes problemas de processamento de dados em pedaços menores e os distribui em um cluster de servidores, semelhante à maneira como o Synapse Analytics opera.

Colmeia é um recurso de consulta semelhante a SQL que você pode usar com um cluster HDInsight para examinar dados mantidos em uma variedade de formatos. Você pode usá-lo para criar, carregar e consultar tabelas externas, de maneira semelhante ao PolyBase para Azure Synapse Analytics.



## Teste de conhecimento

### Questão 1

#### Quando você deve usar o Azure Synapse Analytics?

- a) Para realizar consultas e agregações muito complexas
- b) Para criar painéis de dados tabulares
- c) Para permitir que um grande número de usuários consultem dados analíticos

Explicação: está correto. O Azure Synapse Analytics é adequado para realizar tarefas de computação intensiva como essas.

### Questão 2

#### Qual é o propósito da ingestão de dados?

- a) Para realizar transformações de dados complexas em dados recebidos de fontes externas
- b) Para capturar dados que fluem para um sistema de data warehouse o mais rápido possível
- c) Para visualizar os resultados da análise de dados

Explicação: está correto. A ingestão de dados pode receber dados de várias fontes, incluindo fluxos, e deve ser executada com rapidez suficiente para não perder nenhum dado recebido.

### Questão 3

#### Qual é a principal diferença entre um data lake e um armazém de dados?

- a) Um data lake contém informação estruturada, mas um data warehouse contém dados brutos de negócios
- b) Um data lake contém dados não tratados, mas um data warehouse contém informação estruturada
- c) Os dados armazenados em um data lake são dinâmicos, mas as informações armazenadas em um data warehouse são estáticas

Explicação: está correto. Uma solução de data warehouse converte os dados brutos de um data lake em informações de negócios significativas em um data warehouse.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Resumo

Esta lição descreveu como uma solução de data warehouse funciona e deu a você uma visão geral dos serviços que você pode usar para construir um data warehouse moderno no Azure.

Nesta lição, você viu como:

- Explorar os conceitos de armazenamento de dados
- Explorar os serviços de dados do Azure para armazenamento de dados moderno
- Explorar a arquitetura moderna de data warehouse e a carga de trabalho
- Explorar os serviços de dados do Azure no portal do Azure

## Saber mais

- Data Factory
- Armazenamento Azure Data Lake
- Azure Databricks
- Azure Synapse Analytics
- O que é o Azure Analysis Services?
- O que é Power BI?
- HDInsight
- Orientação para projetar tabelas distribuídas no pool Synapse SQL
- Estratégias de carregamento de dados para o pool Synapse SQL
- O que é PolyBase?

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Introdução

A ingestão de dados é o processo usado para carregar dados de uma ou mais fontes em um armazenamento de dados. Depois de ingeridos, os dados ficam disponíveis para uso. Os dados podem ser ingeridos usando processamento em lote ou streaming, dependendo da natureza da fonte de dados.

As organizações geralmente têm várias fontes de dados distintas. Para fornecer uma solução de nuvem completa, é importante ter uma abordagem flexível para ingestão de dados em um armazenamento de dados do Azure. O Azure oferece muitas maneiras de ingerir dados.

Nesta lição, você explorará algumas dessas ferramentas e técnicas que podem ser usadas para ingerir dados com o Azure.

## Objetivos de aprendizado

Nesta lição, você irá:

- Descrever a ingestão de dados no Azure
- Descrever os componentes do Azure Data Factory
- Aprender como usar o Azure Data Factory para carregar dados em um data warehouse

## Descreva as práticas comuns para carregamento de dados

A ingestão de dados é a primeira parte de qualquer solução de armazenamento de dados. É sem dúvida a parte mais importante. Se você perder algum dado neste ponto, qualquer informação resultante pode ser imprecisa, deixando de representar os fatos nos quais você pode basear suas decisões de negócios. Em um sistema de big data, a ingestão de dados deve ser rápida o suficiente para capturar as grandes quantidades de dados que podem estar vindo em sua direção e ter poder de computação suficiente para processar esses dados em tempo hábil.

O Azure fornece vários serviços que você pode usar para ingerir dados. Esses serviços podem operar com quase qualquer fonte. Nesta unidade, você examinará algumas das ferramentas mais populares usadas com o Azure: Azure Data Factory, PolyBase, SQL Server Integration Services e Azure Databricks.

CURSO DE INGRESSO NA AZURE DATA FUNDAMENTOS DA SILVA OLIVEIRA

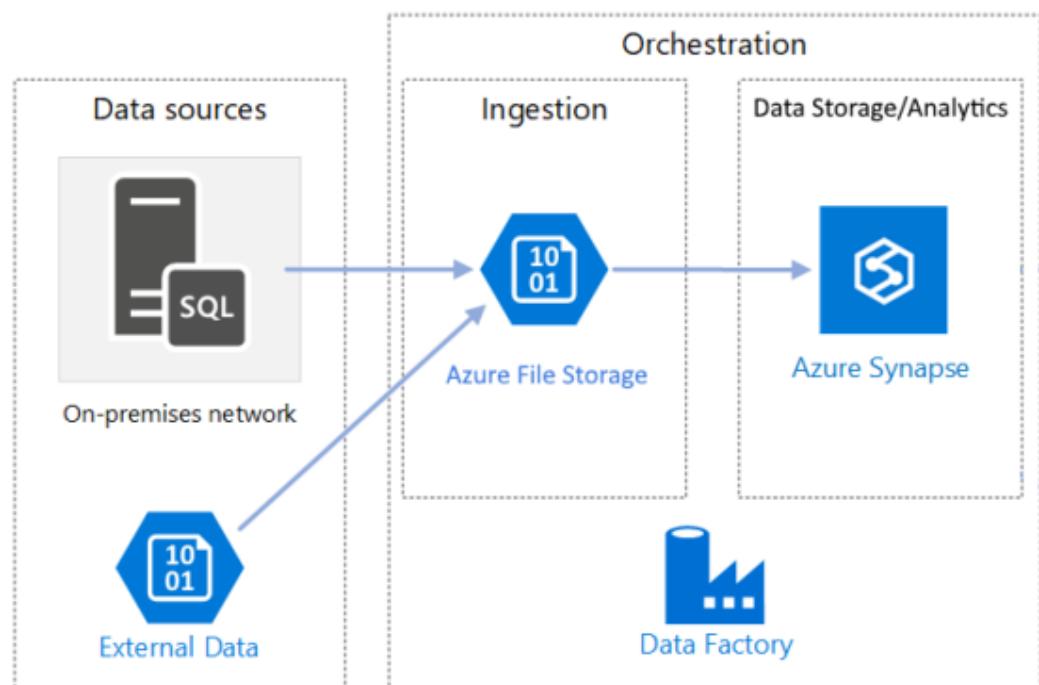
CPF: 00296691208

## Ingerir dados usando o Azure Data Factory

O Azure Data Factory é um serviço de ingestão e transformação de dados que permite carregar dados brutos de muitas fontes diferentes, tanto no local quanto na nuvem. À medida que ingere os dados, o Data Factory pode limpar, transformar e reestruturar os dados, antes de carregá-los em um repositório como um data warehouse. Assim que os dados estiverem no data warehouse, você pode analisá-los.

Data Factory contém uma série de sistemas interconectados que fornecem uma plataforma ponta a ponta completa para engenheiros de dados. Você pode carregar dados estáticos, mas também pode ingerir dados de streaming. Carregar dados de um fluxo oferece uma solução em tempo real para dados que chegam rapidamente ou que mudam rapidamente. Usando streaming, você pode usar o Azure Data Factory para atualizar continuamente as informações em um data warehouse com os dados mais recentes.

Data Factory fornece um *orquestração*. Orquestração é o processo de direcionar e controlar outros serviços, e conectá-los, para permitir que os dados fluam entre eles. Data Factory usa orquestração para combinar e automatizar sequências de tarefas que usam serviços diferentes para realizar operações complexas.



THIAGO FERNANDES DA SILVA  
CPF 00296691208

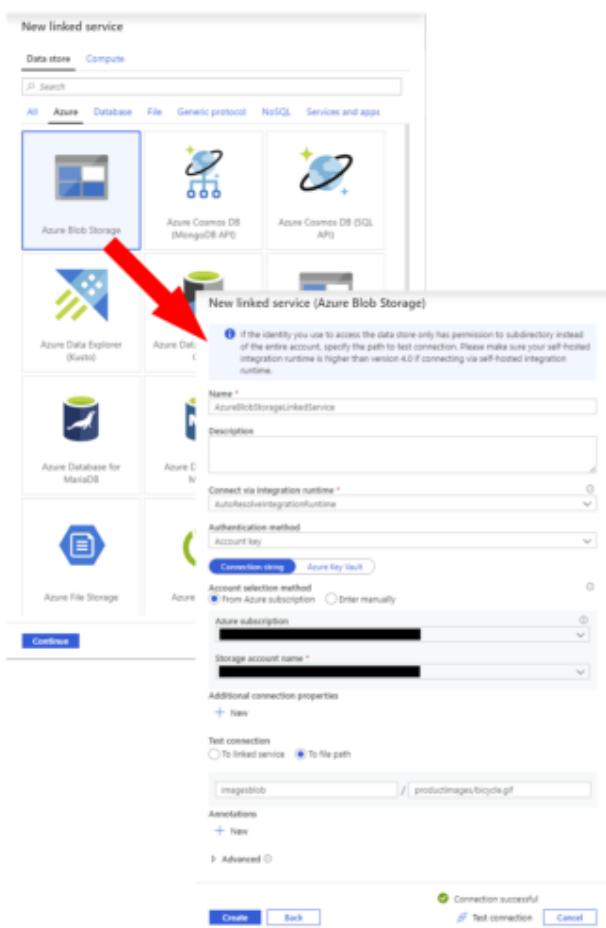
O Azure Data Factory usa vários recursos diferentes: serviços vinculados, conjuntos de dados e pipelines. As seções a seguir descrevem como o Data Factory usa esses recursos.

## Entenda os serviços vinculados

Data Factory move dados de uma fonte de dados para um destino. Um serviço vinculado fornece as informações necessárias para o Data Factory se conectar a uma origem ou destino. Por exemplo, você pode usar um serviço vinculado do Armazenamento de Blob do Azure para conectar uma conta de armazenamento ao Data Factory ou o serviço vinculado do Banco de Dados SQL do Azure para se conectar a um banco de dados SQL.

As informações que um serviço vinculado contém variam de acordo com o recurso. Por exemplo, para criar um serviço vinculado para o Armazenamento de Blob do Azure, você fornece informações como o nome da assinatura do Azure que possui a conta de armazenamento, o nome da conta de armazenamento e as informações necessárias para autenticação na conta de armazenamento. Para criar um serviço vinculado a um recurso diferente, como Banco de Dados SQL do Azure, você especifica o nome do servidor de banco de dados, o nome do banco de dados e as credenciais apropriadas.

A imagem abaixo mostra a interface gráfica do usuário fornecida pelo Azure Data Factory para criar serviços vinculados.

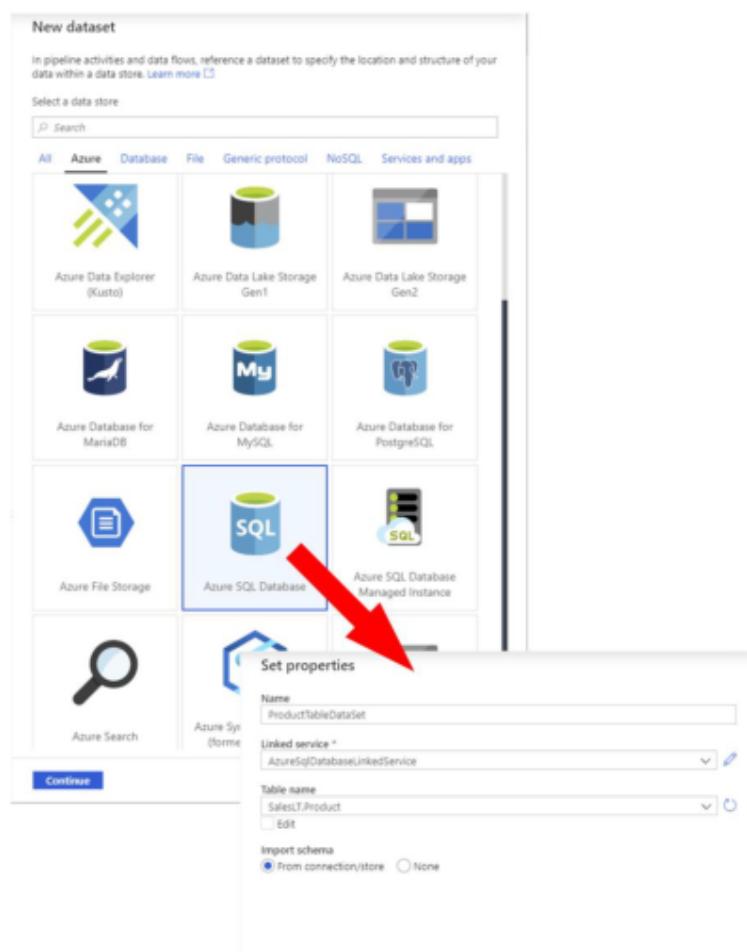


THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Entenda conjuntos de dados

Um conjunto de dados no Azure Data Factory representa os dados que você deseja ingerir (entrada) ou armazenar (saída). Se seus dados tiverem uma estrutura, um conjunto de dados especifica como os dados são estruturados. Nem todos os conjuntos de dados são estruturados. Blobs mantidos no armazenamento de Blob do Azure são um exemplo de dados não estruturados.

Um conjunto de dados se conecta a uma entrada ou saída usando um serviço vinculado. Por exemplo, se você estiver lendo e processando dados do armazenamento de Blob do Azure, você criará um conjunto de dados de entrada que usa um serviço vinculado ao Armazenamento de Blob para especificar os detalhes da conta de armazenamento. O conjunto de dados especificaria qual blob ingerir e o formato das informações no blob (dados binários, JSON, texto delimitado e assim por diante). Se estiver usando o Azure Data Factory para armazenar dados em uma tabela em um banco de dados SQL, você definirá um conjunto de dados de saída que usa um serviço vinculado ao Banco de Dados SQL para se conectar ao banco de dados e especifica qual tabela usar nesse banco de dados.

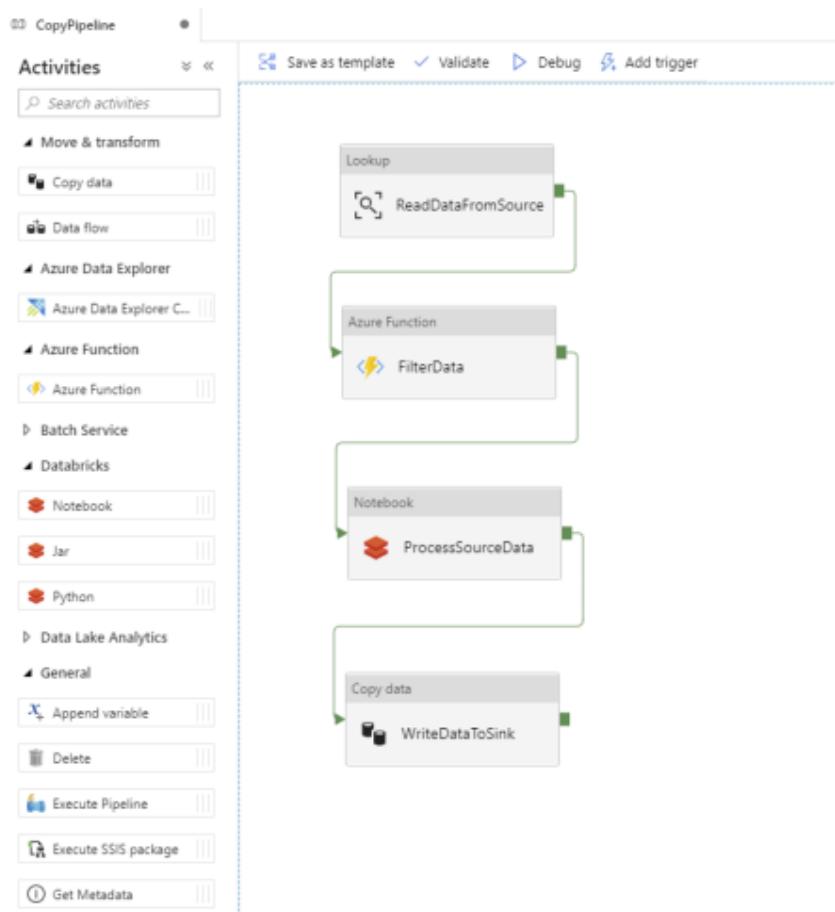


THIAGO FERNANDES DA SILVA

CPF: 00296691208

## Entenda pipelines

Um pipeline é um agrupamento lógico de atividades que, juntas, executam uma tarefa. As atividades em um pipeline definem ações a serem executadas em seus dados. Por exemplo, você pode usar uma atividade de cópia para transformar dados de um conjunto de dados de origem em um conjunto de dados de destino. Você pode incluir atividades que transformam os dados à medida que são transferidos ou pode combinar dados de várias fontes. Outras atividades permitem incorporar elementos de processamento de outros serviços. Por exemplo, você pode usar uma Função Azure modificar e filtrar dados, ou um Azure Databricks Notebook atividade para executar um notebook que execute um processamento mais avançado.

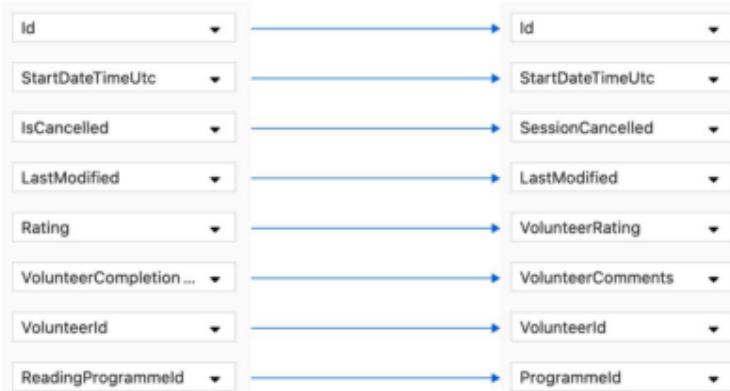


THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Os pipelines não precisam ser lineares. Você pode incluir atividades lógicas que executam repetidamente uma série de tarefas enquanto alguma condição é verdadeira usando uma atividade “Para cada”, ou seguir diferentes caminhos de processamento, dependendo do resultado do processamento anterior usando uma atividade “Se condicional”.

Às vezes, ao ingerir dados, os dados que você está trazendo podem ter nomes de coluna e tipos de dados diferentes daqueles exigidos pela saída. Nesses casos, você pode usar um mapeamento para transformar seus dados do formato de entrada para o formato de saída. A captura de tela abaixo mostra a tela de mapeamento para o *Copiar Dados* atividade. Ele ilustra como as colunas dos dados de entrada podem ser mapeadas para o formato de dados exigido pela saída.



Você pode executar um pipeline manualmente ou pode fazer com que seja executado posteriormente usando um gatilho. Um gatilho permite que você agende um pipeline para ocorrer de acordo com uma programação planejada (todos os sábados à noite, por exemplo) ou em intervalos repetidos (a cada poucos minutos ou horas), ou quando ocorre um evento, como a chegada de um arquivo no Azure Data Lake Storage ou a exclusão de um blob no Azure Blob Storage.

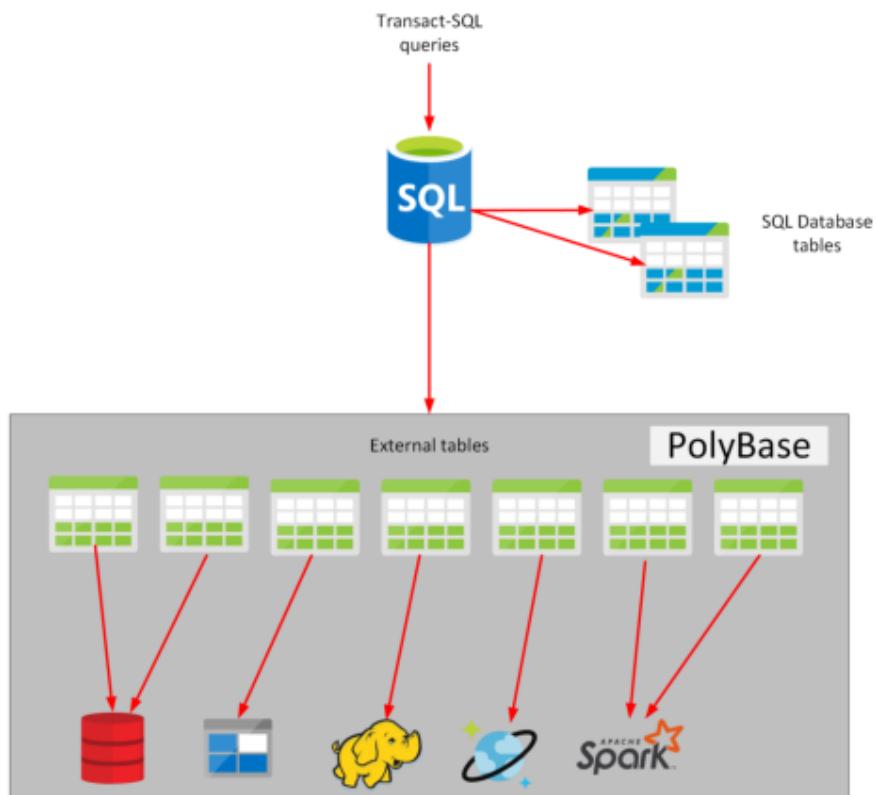
### Ingerir dados usando PolyBase

PolyBase é um recurso do SQL Server e do Azure Synapse Analytics que permite que você execute consultas Transact-SQL que leem dados de fontes de dados externas. O PolyBase faz com que essas fontes de dados externas pareçam tabelas em um banco de dados SQL. Usando o PolyBase, você pode ler dados gerenciados por Hadoop, Spark e Azure Blob Storage, bem como outros sistemas de gerenciamento de banco de dados, como Cosmos DB, Oracle, Teradata e MongoDB.

**NOTA:** Spark é um mecanismo de processamento paralelo que oferece suporte a análises em grande escala.

O PolyBase permite que você transfira dados de uma fonte de dados externa para uma tabela, bem como copie dados de uma fonte de dados externa no Azure Synapse Analytics ou SQL Server. Você também pode executar consultas que unem tabelas em um banco de dados SQL com dados externos, permitindo que você execute análises que abrangem vários armazenamentos de dados.

**NOTA:** O Banco de Dados SQL do Azure não oferece suporte ao PolyBase.



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF: 00296691208

O Azure Data Factory oferece suporte ao PolyBase para carregar dados. Por exemplo, o Data Factory pode invocar diretamente o PolyBase em seu nome se os seus dados estiverem em um armazenamento de dados compatível com o PolyBase.

## Ingerir dados usando o SQL Server Integration Services

O SSIS (SQL Server Integration Services) é uma plataforma para construir soluções de integração e transformação de dados de nível empresarial. Você pode usar o SSIS para resolver problemas de negócios complexos, copiando ou baixando arquivos, carregando data warehouses, limpando e minerando dados e gerenciando objetos e dados de banco de dados SQL. O SSIS faz parte do Microsoft SQL Server.

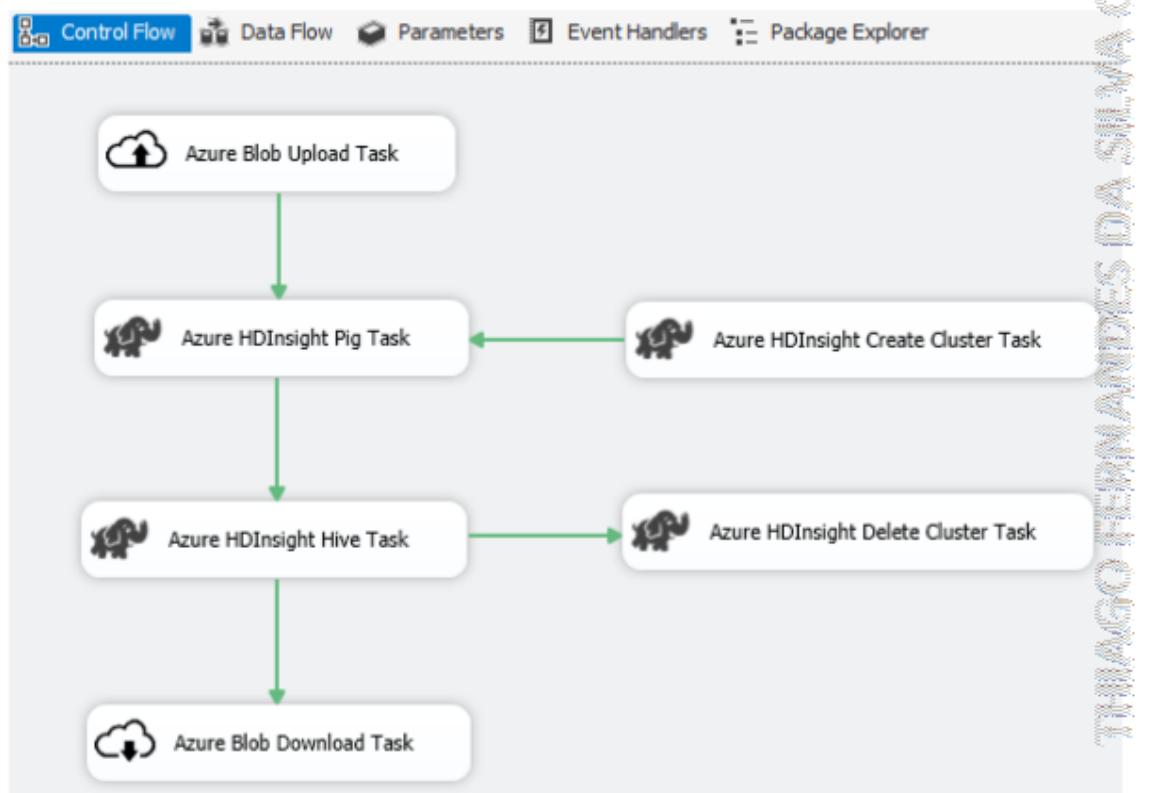
O SSIS pode extrair e transformar dados de uma ampla variedade de fontes, como arquivos de dados XML, arquivos simples e fontes de dados relacionais e, em seguida, carregar os dados em um ou mais destinos.

O SSIS inclui um rico conjunto de tarefas e transformações internas, ferramentas gráficas para a construção de pacotes e o banco de dados do Catálogo do Integration Services, onde você armazena, executa e gerencia pacotes. Um pacote é uma coleção organizada de conexões, elementos de fluxo de controle, elementos de fluxo de dados, manipuladores de eventos, variáveis, parâmetros e configurações, que você monta usando as ferramentas de design gráfico que o SQL Server Integration Services fornece ou constrói programaticamente. Em seguida, salve o pacote concluído no SQL Server, no Armazenamento de Pacotes do Integration Services ou no sistema de arquivos.

Você pode usar as ferramentas gráficas do SSIS para criar soluções sem escrever uma única linha de código. Você também pode programar o extenso modelo de objeto do Integration Services para criar pacotes programaticamente e codificar tarefas personalizadas e outros objetos de pacote.

SSIS é um utilitário local. No entanto, o Azure Data Factory permite que você execute seus pacotes SSIS existentes como parte de um pipeline na nuvem. Isso permite que você comece rapidamente, sem ter que reescrever sua lógica de transformação existente.

O SSIS Feature Pack para Azure é uma extensão que fornece componentes que se conectam aos serviços do Azure, transferem dados entre o Azure e fontes de dados locais e processam dados armazenados no Azure. Os componentes do pacote de recursos oferecem suporte à transferência de ou para o armazenamento do Azure, Azure Data Lake e Azure HDInsight. Usando esses componentes, você pode executar processamento em grande escala de dados ingeridos.



### Ingerir dados usando Azure Databricks

Azure Databricks é uma plataforma de análise otimizada para a plataforma de serviços em nuvem Microsoft Azure. O Databricks é baseado no Spark e integrado ao Azure para otimizar os fluxos de trabalho. Ele fornece um espaço de trabalho interativo que permite a colaboração entre cientistas de dados, engenheiros de dados e analistas de negócios.

Os Databricks podem processar dados mantidos em muitos tipos diferentes de armazenamento, incluindo armazenamento de Blob do Azure, Azure Data Lake Store, armazenamento Hadoop, arquivos simples, bancos de dados SQL e data warehouses e serviços do Azure, como Cosmos DB. Os Databricks também podem processar dados de streaming. Por exemplo, você pode capturar dados transmitidos de sensores e outros dispositivos.

Você escreve e executa o código Spark usando notebooks. Um notebook é como um programa que contém uma série de etapas (chamadas células). Um bloco de notas pode conter células que leem dados de uma ou mais fontes de dados, processam os dados e gravam os resultados em um armazenamento de dados. A escalabilidade do Azure Databricks o torna uma plataforma ideal para realizar a ingestão de dados complexos e tarefas analíticas.

O Azure Data Factory pode incorporar blocos de anotações do Azure Databricks em um pipeline. Um pipeline pode passar parâmetros para um notebook. Esses parâmetros podem especificar quais dados ler e analisar. O bloco de notas pode salvar os resultados, que o pipeline do Azure Data Factory pode usar em atividades subsequentes.

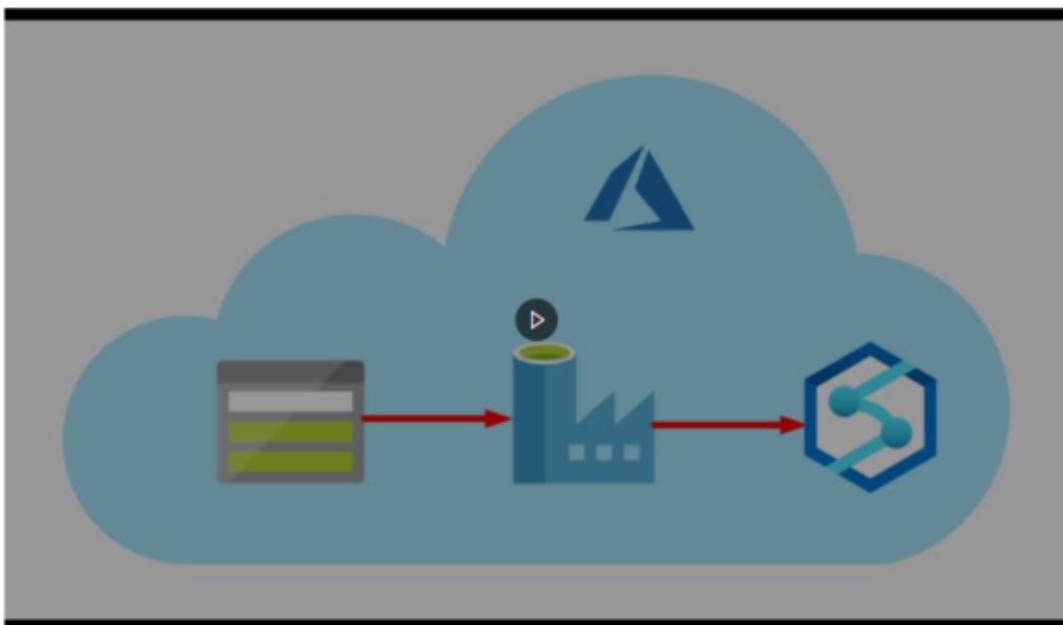
THIAGO FERNANDES  
CPF 00296691208

## Demonstração: carregar dados no Azure Synapse Analytics

Imagine que você faz parte de uma equipe que analisa dados de preços de casas. O conjunto de dados que você recebe contém informações sobre preços de casas para várias regiões. Sua equipe precisa relatar como os preços das casas em cada região variaram nos últimos meses. Para conseguir isso, você precisa ingerir os dados no Azure Synapse Analytics. Você decidiu usar o Azure Data Factory para executar esta tarefa.

Neste vídeo, você verá como usar o Azure Data Factory para ingerir e processar dados de preços de casas para análise. Você armazenará os dados no Azure Synapse Analytics para análise posterior.

[vídeo]



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## Teste de conhecimento

### Questão 1

**Qual componente de um Azure Data Factory pode ser acionado para executar tarefas de ingestão de dados?**

- a) Arquivo CSV
- b) Pipeline
- c) Serviço vinculado

Explicação: está correto. Os pipelines podem ser acionados para executar atividades de ingestão de dados.

### Questão 2

**Quando você pode usar o PolyBase?**

- a) Para consultar dados de fontes de dados externas do Azure Synapse Analytics
- b) Para ingerir dados de streaming usando Azure Databricks
- c) Para orquestrar atividades no Azure Data Factory

Explicação: está correto. Este é o objetivo do PolyBase

### Questão 3

**Qual desses serviços pode ser usado para ingerir dados no Azure Synapse Analytics?**

- a) Azure Data Factory
- b) Power BI
- c) Azure Active Directory

Explicação: está correto. O Azure Data Factory pode ser usado para ingerir dados no Azure Synapse Analytics de quase qualquer fonte.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Resumo

Nesta lição, você aprendeu sobre ferramentas para ingestão de dados em um banco de dados do Azure. Você viu como usar o Azure Data Factory para ler, processar e armazenar dados em um data warehouse.

Nesta lição, você aprendeu como:

- Descrever a ingestão de dados no Azure
- Descrever os componentes do Azure Data Factory
- Carregar dados no Azure Synapse Analytics

## Saber mais

- Pipelines e atividades no Azure Data Factory
- Início rápido: crie um data factory usando a IU do Azure Data Factory
- Azure SQL Data Warehouse agora é Azure Synapse Analytics
- BI empresarial automatizado com Azure Synapse Analytics e Azure Data Factory

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Introdução

Os dados residem em muitos locais em uma organização. Ao projetar sua solução de dados em nuvem, você desejará ingerir seus dados brutos em um armazenamento de dados para análise. Uma abordagem comum que você pode usar com o Azure Synapse Analytics é extrair os dados de onde estão armazenados atualmente, carregar esses dados em um armazenamento de dados analíticos, e depois transformar os dados, moldando-os para análise. Essa abordagem é conhecida como ELT, para extrair, carregar e transformar.

O Azure Synapse Analytics é particularmente adequado para essa abordagem. Usando o Apache Spark e pipelines automatizados, o Synapse Analytics pode executar tarefas de processamento paralelo em grandes conjuntos de dados e realizar análises big data.

**NOTA:** O termo big data refere-se a dados que são muito grandes ou complexos para sistemas de banco de dados tradicionais. Os sistemas que processam big data precisam realizar ingestão e processamento rápidos de dados; eles devem ter capacidade para armazenar os resultados e poder de computação suficiente para realizar análises sobre esses resultados.

Outra opção é analisar os dados operacionais em seu local original. Esta estratégia é conhecida como processamento analítico transacional híbrido (HTAP). Você pode executar esse estilo de análise sobre os dados mantidos em repositórios, como o Azure Cosmos DB usando o Azure Synapse Link.

## Objetivos de aprendizado

Nesta lição, você irá:

- Descrever as opções de processamento de dados para realizar análises no Azure
- Explorar o Azure Synapse Analytics

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Descreva o armazenamento e processamento de dados com o Azure

As organizações geram dados em todos os seus negócios. Para fins de análise, esses dados podem ser deixados em seu formato bruto, ingeridos, ou os dados podem ser processados e salvos em um armazenamento de dados ou warehouse de dados especialmente projetado. O Azure permite que as empresas implementem qualquer um desses cenários.

As opções mais comuns para processar dados no Azure incluem Azure Databricks, Azure Data Factory, Azure Synapse Analytics e Azure Data Lake. Nesta unidade, você explorará essas opções com mais detalhes.

### O que é o Azure Synapse Analytics?

O Azure Synapse Analytics é um serviço de análise generalizado. Você pode usá-lo para ler dados de muitas fontes, processar esses dados, gerar várias análises e modelos e salvar os resultados.

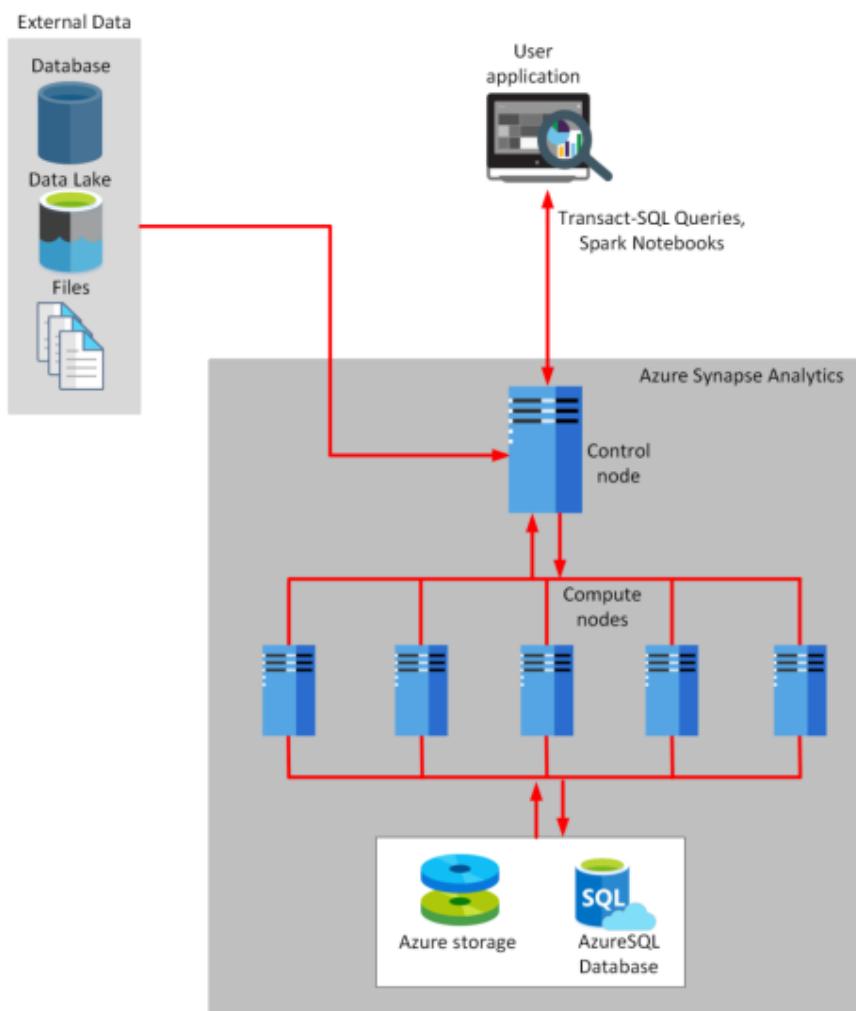
Você pode selecionar entre duas tecnologias para processar dados:

- Transact-SQL. Esse é o mesmo dialeto do SQL usado pelo Banco de Dados SQL do Azure, com algumas extensões para ler dados de fontes externas, como bancos de dados, arquivos e armazenamento do Azure Data Lake. Você pode usar essas extensões para carregar dados rapidamente, gerar agregações e outras análises, criar tabelas e exibições e armazenar informações usando essas tabelas e exibições. Você pode usar os resultados para relatórios e processamento posteriores.
- Spark. Esta é a mesma tecnologia de código aberto usada para alimentar os Azure Databricks. Você escreve seu código analítico usando notebooks em uma linguagem de programação como C #, Scala, Python ou SQL. As bibliotecas Spark fornecidas com o Azure Synapse Analytics permitem que você leia dados de fontes externas e também grave dados em uma variedade de formatos diferentes se você precisar salvar seus resultados para análise posterior.

O Azure Synapse Analytics usa uma arquitetura em cluster. Cada cluster tem um nó de controle que é usado como ponto de entrada para o sistema. Quando você executa instruções Transact-SQL ou inicia trabalhos do Spark em um bloco de notas, a solicitação é enviada ao nó de controle. O nó de controle executa um mecanismo de processamento paralelo que divide a operação em um conjunto de tarefas que podem ser executadas simultaneamente. Cada tarefa executa parte da carga de trabalho em um subconjunto dos dados de origem. Cada tarefa é enviada para calcular um nó para realmente fazer o processamento. O nó de controle reúne os resultados dos nós de computação e os combina em um resultado geral.

AULA 01 - ARMAZENAMENTO E PROCESSAMENTO DE DADOS NO AZURE

CPF: 000296691208



A próxima unidade descreve os componentes do Azure Synapse Analytics em mais detalhes.

Para mais informações, leia O que é o Azure Synapse Analytics?

## O que é Azure Databricks?

Azure Databricks é uma plataforma de análise otimizada para a plataforma de serviços em nuvem Microsoft Azure. Projetado com os fundadores do Apache Spark, o Databricks é integrado ao Azure para fornecer configuração com um clique, fluxos de trabalho simplificados e um espaço de trabalho interativo que permite a colaboração entre cientistas de dados, engenheiros de dados e analistas de negócios.

Os Databricks podem processar dados mantidos em muitos tipos diferentes de armazenamento, incluindo armazenamento de Blob do Azure, Azure Data Lake Store, armazenamento do Hadoop, arquivos simples, bancos de dados e data warehouses. Os Databricks também podem processar dados de streaming. O Databricks usa uma arquitetura extensível baseada em drivers.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

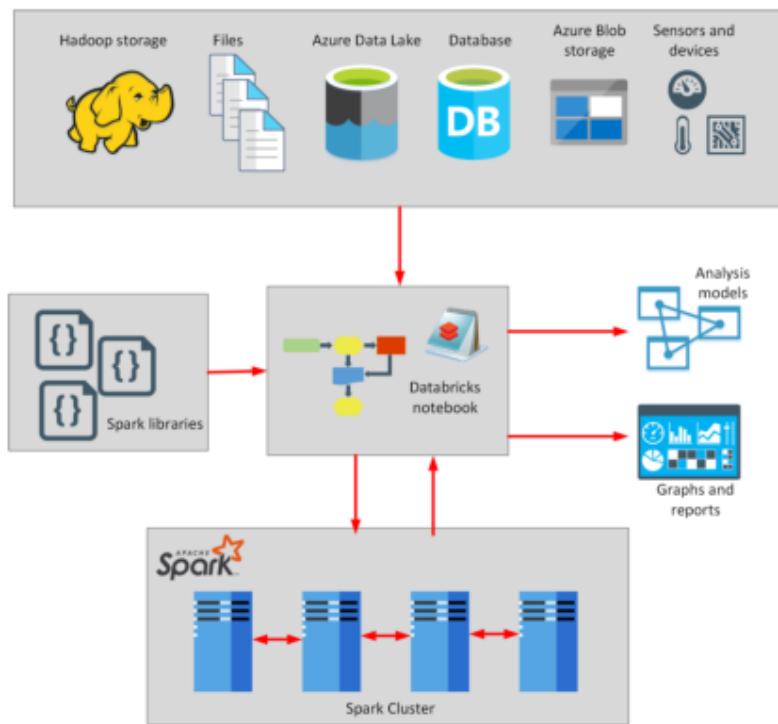
**NOTA:** Driver é um trecho de código que se conecta a uma fonte de dados específica e permite que você leia e grave essa fonte. Um driver é normalmente fornecido como parte de uma biblioteca que pode ser carregada no ambiente do Databricks. Os drivers estão disponíveis para muitos serviços do Azure, incluindo Banco de Dados SQL do Azure, Azure Cosmos DB, armazenamento de Blob do Azure e armazenamento do Azure Data Lake, bem como muitos serviços e bancos de dados produzidos por terceiros, como MySQL e PostgreSQL.

O mecanismo de processamento é fornecido pelo Apache Spark. Spark é um mecanismo de processamento paralelo que oferece suporte a análises em grande escala. Você escreve o código do aplicativo que consome dados de uma ou mais fontes e mescla, reformata, filtra e remodela esses dados e, em seguida, armazena os resultados. O Spark distribui o trabalho por um cluster de computadores. Cada computador pode processar seus dados em paralelo com os outros computadores. A estratégia ajuda a reduzir o tempo necessário para a execução do trabalho. O Spark foi projetado para lidar com grandes quantidades de dados.

Você pode escrever o código do aplicativo Spark usando várias linguagens, como Python, R, Scala, Java e SQL. O Spark possui várias bibliotecas para essas linguagens, fornecendo rotinas analíticas complexas que foram otimizadas para o ambiente em cluster. Essas bibliotecas incluem módulos para aprendizado de máquina, análise estatística, modelagem linear e não linear, análise preditiva e gráficos.

Você escreve aplicativos Databricks usando um Notebook. Um bloco de notas contém uma série de etapas (células), cada um contendo um bloco de código. Por exemplo, uma célula pode conter o código que se conecta a uma fonte de dados, a próxima célula lê os dados dessa fonte e os converte em um modelo na memória, a próxima célula plota um gráfico e uma célula final salva os dados do modelo na memória para um repositório.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208



Para mais informações, leia “O que é Azure Databricks?”.

## O que é Azure HDInsight?

Azure HDInsight é um serviço de análise gerenciado na nuvem. Ele é baseado no Apache Hadoop, uma coleção de ferramentas e utilitários de código aberto que permitem executar tarefas de processamento em grandes quantidades de dados. O HDInsight usa um modelo em cluster, semelhante ao do Synapse Analytics. HDInsight armazena dados usando o armazenamento Azure Data Lake. Você pode usar o HDInsight para analisar dados usando estruturas como Hadoop Map / Reduce, Apache Spark, Apache Hive, Apache Kafka, Apache Storm, R e mais.

Hadoop Map / Reduce usa uma estrutura simples para dividir uma tarefa em um grande conjunto de dados em uma série de tarefas menores em subconjuntos de dados que podem ser executados em paralelo e os resultados então combinados. Você escreve seu código Map / Reduce em uma linguagem como Java e, em seguida, envia esse código como um trabalho para o cluster do Hadoop. Hadoop Map / Reduce foi amplamente substituído pelo Spark, que oferece um conjunto mais avançado de operações e uma interface mais simples.

Como os trabalhos de Mapear / Reduzir, os trabalhos do Spark são paralelizados em uma série de tarefas de subtarefas que são executadas no cluster. Você pode escrever trabalhos do Spark como parte de um aplicativo ou usar blocos de anotações interativos. Esses blocos de anotações são iguais aos que você pode executar a partir do Azure Databricks. O Spark inclui bibliotecas que você pode usar para ler e gravar dados em uma ampla variedade de armazenamentos de dados (não apenas

THIAGO FERNANDES DA SILVA OLIVEIRA

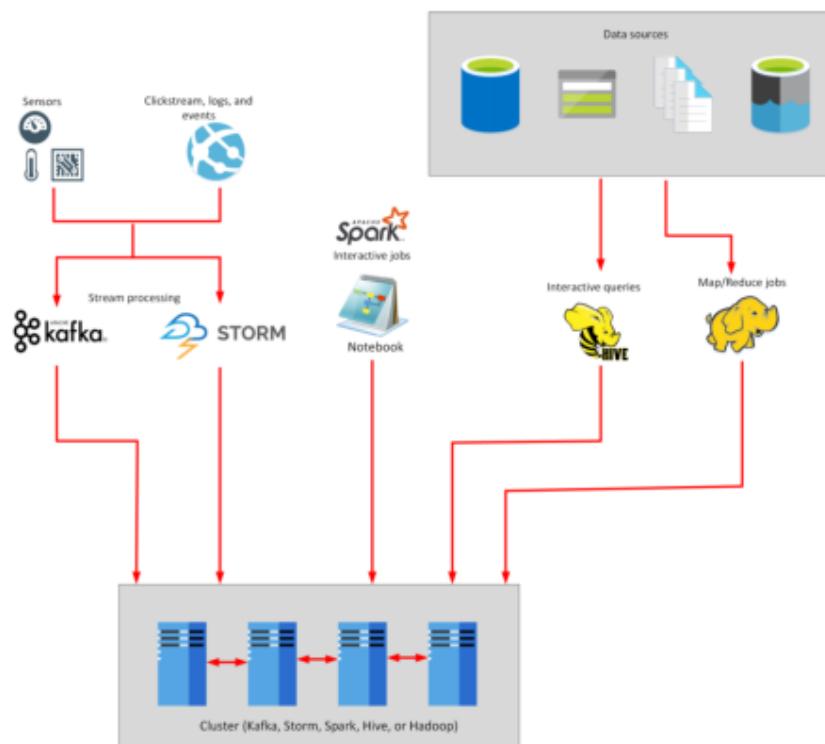
CPF: 00296691208

HDFS). Por exemplo, você pode se conectar a bancos de dados relacionais, como o Banco de Dados SQL do Azure, e outros serviços, como o Azure Cosmos DB.

O Apache Hive fornece recursos interativos semelhantes a SQL para consultar, agragar e resumir dados. Os dados podem vir de muitas fontes diferentes. As consultas são convertidas em tarefas e paralelizadas. Cada tarefa pode ser executada em um nó separado no cluster HDInsight e os resultados são combinados antes de serem retornados ao usuário.

Apache Kafka é um serviço de streaming em cluster que pode ingerir dados em tempo real. É uma solução altamente escalonável que oferece recursos de publicação e assinatura.

Apache Storm é uma plataforma escalável e tolerante a falhas para executar aplicativos de processamento de dados em tempo real. Storm pode processar grandes volumes de dados de streaming usando requisitos computacionais relativamente modestos. O Storm foi projetado para ser confiável, para que os eventos não sejam perdidos. As soluções Storm também podem fornecer processamento garantido de dados, com a capacidade de reproduzir dados que não foram processados com sucesso na primeira vez. Storm pode interoperar com uma variedade de fontes de eventos, incluindo Hubs de Eventos do Azure, Hub IoT do Azure, Apache Kafka e RabbitMQ (um serviço de enfileiramento de mensagens). Storm também pode gravar em armazenamentos de dados, como bancos de dados HDFS, Hive, HBase, Redis e SQL. Você escreve um aplicativo Storm usando as APIs fornecidas pelo Apache.



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## O que é Azure Data Factory?

Azure Data Factory é um serviço que pode ingerir grandes quantidades de dados brutos e não organizados de sistemas relacionais e não relacionais e converter esses dados em informações significativas. O Data Factory fornece um mecanismo de ingestão escalonável e programável que você pode usar para implementar projetos complexos de extração-transformação-carga (ETL), extração-carga-transformação (ELT) e integração de dados.

Por exemplo, imagine uma empresa de jogos que coleta petabytes de logs de jogos que são produzidos por jogos na nuvem. A empresa deseja analisar esses logs para obter insights sobre as preferências do cliente, dados demográficos e comportamento de uso. Ela também deseja identificar oportunidades de vendas adicionais e cruzadas, desenvolver novos recursos atraentes, impulsionar o crescimento dos negócios e fornecer uma experiência melhor aos clientes.

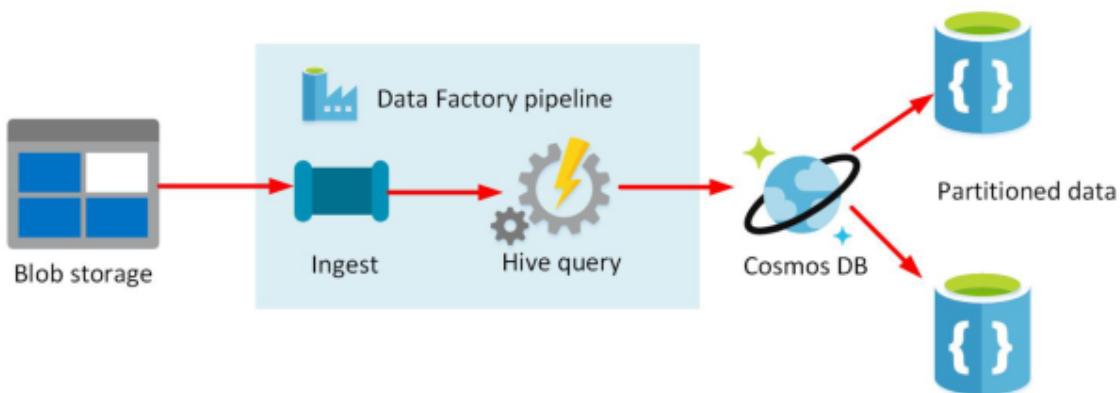
Para analisar esses logs, a empresa precisa usar dados de referência, como informações do cliente, informações do jogo e informações da campanha de marketing que estão em um armazenamento de dados local. A empresa deseja utilizar esses dados do armazenamento de dados local, combinando-os com dados de log adicionais que possui em um armazenamento de dados na nuvem.

Para extrair insights, a empresa deseja processar os dados associados usando um cluster Spark na nuvem (usando Azure HDInsight) e publicar os dados transformados em um data warehouse em nuvem, como o Azure Synapse Analytics. A empresa pode usar as informações no data warehouse para gerar e publicar relatórios. Eles desejam automatizar esse fluxo de trabalho e monitorar e gerenciá-lo diariamente. Eles também desejam executá-lo quando os arquivos chegam a um contêiner de armazenamento de blob.

Usando o Azure Data Factory, você pode criar e agendar fluxos de trabalho baseados em dados (chamados de pipelines) que podem ingerir dados de armazenamentos de dados distintos usados pela empresa de jogos. Você pode criar processos ETL complexos que transformam dados visualmente com fluxos de dados ou usando serviços de computação como Azure HDInsight, Azure Databricks e Banco de Dados SQL do Azure. Em seguida, você pode publicar os dados transformados no Azure Synapse Analytics para que os aplicativos de business intelligence consumam.

Um pipeline é um agrupamento lógico de atividades que executa uma unidade de trabalho. Juntas, as atividades em um pipeline executam uma tarefa. Por exemplo, um pipeline pode conter uma série de atividades que ingerem dados brutos do armazenamento de Blob do Azure e, em seguida, executar uma consulta Hive em um cluster HDInsight para particionar os dados e armazenar os resultados em um banco de dados Cosmos DB.

CPF: 00296691208



## O que é Azure Data Lake?

Azure Data Lake é uma coleção de serviços de análise e armazenamento que você pode combinar para implementar uma solução de big data. É composto por três elementos principais:

- Data Lake Store
- Data Lake Analytics
- HDInsight

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## O que é Data Lake Store?

O Data Lake Store fornece um sistema de arquivos que pode armazenar quantidades quase ilimitadas de dados. Ele usa uma organização hierárquica (como os sistemas de arquivos Windows e Linux), mas você pode conter grandes quantidades de dados brutos (blobs) e dados estruturados. Ele é otimizado para cargas de trabalho analíticas.

O Azure Data Lake Store é compatível com o Hadoop Distributed File System (HDFS). Você pode executar trabalhos do Hadoop usando o Azure HDInsight (veja abaixo) que pode ler e gravar dados no Data Lake Store com eficiência.

O Azure Data Lake Store fornece segurança granular sobre os dados, usando listas de controle de acesso. Uma Lista de Controle de Acesso especifica quais contas podem acessar quais arquivos e pastas na loja. Se estiver mais familiarizado com o Linux, você pode usar permissões de estilo POSIX para conceder acesso de leitura, gravação e pesquisa com base na propriedade do arquivo e associação de grupo de usuários.

Serviços como Azure Data Factory, Azure Databricks, Azure HDInsight, Azure Data Lake Analytics e Azure Stream Analytics podem ler e gravar o Data Lake Store diretamente.

## O que é Data Lake Analytics?

Azure Data Lake Analytics é um serviço de trabalho de análise sob demanda que você pode usar para processar big data. Ele fornece uma estrutura e um conjunto de ferramentas que você usa para analisar dados mantidos no Microsoft Azure Data Lake Store e em outros repositórios. Você escreve trabalhos que contêm consultas para transformar dados e extrair insights.

Você define um trabalho usando uma linguagem chamada U-SQL. Esta é uma linguagem híbrida que usa recursos de SQL e C # e fornece recursos declarativos e procedurais que você pode usar para processar dados.

O exemplo de bloco U-SQL abaixo lê dados de um arquivo chamado StockPrices.csv, que é mantido em uma pasta chamada Mercado de ações no Data Lake Storage. Este é um arquivo de texto que contém informações do mercado de ações (cotações e preços e possivelmente outros dados), mantido em formato separado por vírgulas. oEXTRAIR declaração lê o arquivo linha por linha e extrai os dados no Ticker, e Preçocampos (ele pula a primeira linha, onde um arquivo CSV normalmente contém informações de nome de campo em vez de dados). oSELEÇÃO declaração calcula o preço máximo para cada ticker. oRESULTADO declaração armazena os resultados em outro arquivo no Data Lake Storage.

DATA LAKES DA AZURE

CPF: 000296691208

**NOTA:** Em um arquivo CSV, cada linha consiste em um ou mais campos, e cada campo é separado por uma vírgula. A primeira linha do arquivo geralmente contém os nomes dos campos.

```
@priceData =  
    EXTRACT Ticker string,  
            Price int  
    FROM "/StockMarket/StockPrices.csv"  
    USING Extractors.Csv(skipFirstNRows: 1);  
  
@maxPrices =  
    SELECT Ticker, MAX(Price) AS MaxPrice  
    FROM @priceData  
    GROUP BY Ticker;  
  
OUTPUT @maxPrices  
    TO "/output/MaxPrices.csv"  
    USING Outputters.Csv(outputHeader: true);
```

É importante entender que o código U-SQL fornece apenas uma descrição do trabalho a ser executado. O Azure Data Lake Analytics determina a melhor forma de realmente realizar esse trabalho. O Data Lake Analytics pega a descrição U-SQL de um trabalho, analisa-a para se certificar de que está sintaticamente correta e, em seguida, compila-a em uma representação interna. O Data Lake Analytics divide essa representação interna em estágios de execução. Cada estágio executa uma tarefa, como extrair os dados de uma fonte especificada, dividir os dados em partições, processar os dados em cada partição, agregar os resultados em uma partição e, em seguida, combinar os resultados de todas as partições. O particionamento é usado para melhorar a paralelização e o processamento para diferentes partições é executado simultaneamente em diferentes nós de processamento.

Um trabalho U-SQL pode gerar resultados em um único arquivo CSV, partitionar os resultados em vários arquivos ou pode gravar em outros destinos. Por exemplo, o Data Lake Analytics permite que você crie saídas personalizadas se desejar salvar dados em um formato específico (como XML ou HTML). Você também pode gravar dados no Data Lake Catalog. O catálogo fornece uma interface semelhante a SQL para o Data Lake Storage, permitindo que você crie tabelas e exibições e execute instruções INSERT, UPDATE e DELETE nessas tabelas e exibições.

THIAGO FERNANDES DA SILVA OLIVEIRA

CPF 00296691208

## Explorar o Azure Synapse Analytics

O Azure Synapse Analytics fornece um conjunto de ferramentas para analisar e processar os dados de uma organização. Ele incorpora tecnologias SQL, recursos de consulta Transact-SQL e ferramentas de código aberto do Spark para permitir que você processe rapidamente grandes quantidades de dados.

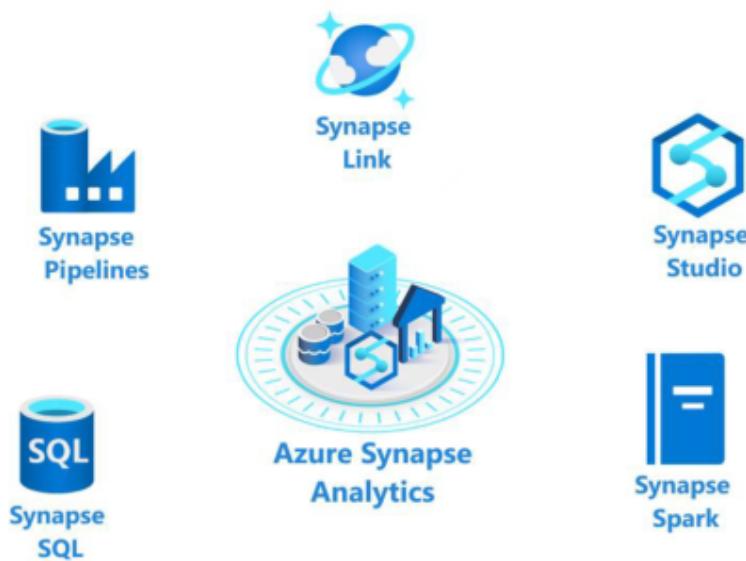
Nesta unidade, você verá mais de perto os recursos do Synapse Analytics e quando deve considerá-lo.

### Quais são os componentes do Azure Synapse Analytics?

O Azure Synapse Analytics é um serviço de análise integrado que permite que as organizações obtenham insights rapidamente de todos os seus dados em qualquer hiperescala, de data warehouses e sistemas de análise de big data.

A Sinapse do Azure é composta pelos seguintes elementos:

- **Pool de sinapse SQL:** Este é um conjunto de servidores que executam Transact-SQL. Transact-SQL é o dialeto do SQL usado pelo Banco de Dados SQL do Azure e Microsoft SQL Server. Você escreve sua lógica de processamento de dados usando Transact-SQL.
- **Piscina Synapse Spark:** Este é um cluster de servidores executando o Apache Spark para processar dados. Você escreve sua lógica de processamento de dados usando uma das quatro linguagens suportadas: Python, Scala, SQL e C # (via .NET para Apache Spark). Os pools do Spark oferecem suporte ao Azure Machine Learning por meio da integração com os pacotes SparkML e AzureML.
- **Pipelines de sinapse:** Um pipeline de sinapse é um agrupamento lógico de atividades que, juntas, executam uma tarefa. As atividades em um pipeline definem ações a serem executadas em seus dados. Por exemplo, você pode usar uma atividade de cópia para transformar dados de um conjunto de dados de origem em um conjunto de dados de destino. Você pode incluir atividades que transformam os dados à medida que são transferidos ou pode combinar dados de várias fontes.
- **Link de sinapse:** Este componente permite que você se conecte ao Cosmos DB. Você pode usá-lo para realizar análises quase em tempo real sobre os dados operacionais armazenados em um banco de dados Cosmos DB.
- **Synapse Studio:** Esta é uma interface de usuário da web que permite aos engenheiros de dados acessar todas as ferramentas do Synapse Analytics. Você pode usar o Synapse Studio para criar pools SQL e Spark, definir e executar pipelines e configurar links para fontes de dados externas.



**NOTA:** Quaisquer dados armazenados no Azure Synapse Analytics podem ser usados para criar e treinar modelos com o Azure Machine Learning.

As seções a seguir descrevem cada um desses elementos com mais detalhes.

## O que são pools SQL?

Quando você usa o Synapse SQL, sua carga de trabalho analítica é executada usando um pool SQL. Em um pool SQL, os nós de controle e computação no cluster executam uma versão do Banco de Dados SQL do Azure que oferece suporte a consultas distribuídas. Você define sua lógica usando instruções Transact-SQL. Você envia suas instruções Transact-SQL para o nó de controle, que divide o trabalho em consultas que operam em um subconjunto de dados e, em seguida, envia essas consultas menores para os nós de computação. Os dados são divididos em blocos chamados distribuições. Uma distribuição é a unidade básica de armazenamento e processamento para consultas paralelas executadas em dados distribuídos. Cada uma das consultas menores é executada em uma das distribuições de dados.

Os nós de controle e computação usam o Data Movement Service (DMS) para mover dados entre os nós conforme necessário para executar consultas em paralelo e retornar resultados precisos.

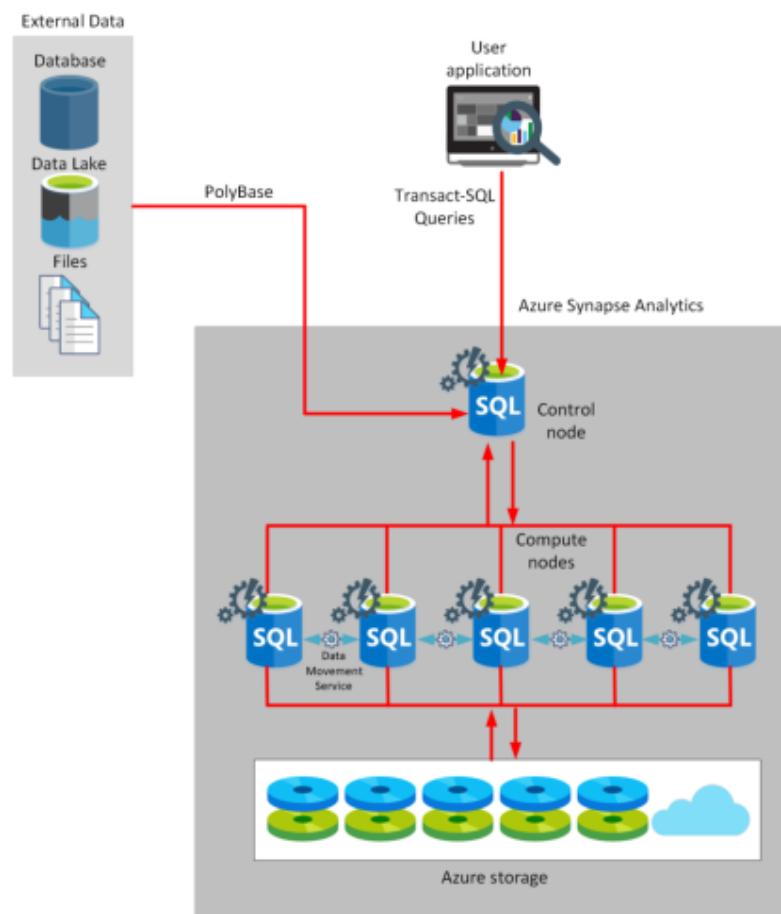
Synapse Analytics usa uma tecnologia chamada PolyBase para fazer com que os dados externos pareçam tabelas SQL. Você pode executar consultas diretamente a essas tabelas ou pode transferir os dados para uma série de tabelas SQL gerenciadas pelo Synapse Analytics para consulta posterior. O Synapse usa o Armazenamento do Azure para gerenciar seus dados enquanto eles estão sendo processados.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Por padrão, um pool SQL sob demanda é criado em cada espaço de trabalho do Azure Synapse Analytics. Você pode provisionar pools adicionais, sob demanda ou provisionados.

**NOTA:** Os pools sob demanda permitem apenas que você consulte dados mantidos em arquivos externos. Se você deseja ingerir e carregar os dados no Synapse Analytics, deve criar seu próprio pool SQL.

O Azure Synapse Analytics foi projetado para executar consultas em grandes conjuntos de dados. Você pode dimensionar manualmente o pool SQL para até 60 nós. Você também pode pausar um pool SQL se você não precisar dele por um tempo. A pausa libera os recursos associados ao pool. Você não é cobrado por esses recursos até que manualmente retome a piscina. No entanto, você não pode executar nenhuma consulta até que o pool seja retomado. A retomada de uma piscina pode levar vários minutos.



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Use pools SQL no Synapse Analytics para os seguintes cenários:

- Relatórios complexos. Você pode usar todo o poder do Transact-SQL para executar instruções SQL complexas que resumem e agregam dados.
- Ingestão de dados. O PolyBase permite que você recupere dados de muitas fontes externas e os converta em um formato tabular. Você pode reformatar esses dados e salvá-los como tabelas e exibições materializadas no Azure Synapse.

## O que são Spark Pools?

O Synapse Spark executa clusters com base no Apache Spark em vez do Banco de Dados SQL do Azure. Você escreve seus trabalhos de análise como blocos de anotações, usando código escrito em Python, Scala, C # ou Spark SQL (este é um dialeto diferente do Transact-SQL). Você pode combinar código escrito em vários idiomas no mesmo bloco de notas.

**NOTA:** Piscinas Spark e pools SQL podem coexistir na mesma instância do Azure Synapse Analytics.

Notebooks também permitem que você visualize dados por meio de gráficos e transforme os dados à medida que são carregados. Os dados podem ser usados por Spark Machine Learning (SparkML) e Azure Machine Learning (AzureML) para treinar modelos de aprendizado de máquina compatíveis com inteligência artificial.

Os Spark pools permitem que você processe dados mantidos em vários formatos, como csv, json, xml, parquet, orc e avro. O Spark pode ser estendido para oferecer suporte a muitos mais formatos com fontes de dados externas.

Os pools Spark fornecem os blocos de construção básicos para executar a computação de cluster na memória. Um trabalho do Spark pode carregar e armazenar dados em cache na memória e consultá-los repetidamente. A computação in-memory é muito mais rápida do que os aplicativos baseados em disco. Os pools do Spark no Azure Synapse são compatíveis com o Armazenamento do Azure e o Azure Data Lake Storage, portanto, você pode usar os pools do Spark para processar seus dados armazenados no Azure.

Os Spark pools podem ter escalonamento automático habilitado, para que os pools sejam escalonados adicionando ou removendo nós conforme necessário. Além disso, os pools Spark podem ser desligados sem perda de dados, já que todos os dados são armazenados no Armazenamento do Azure ou no Data Lake Storage.

Os Spark pools no Synapse Analytics são especialmente adequados para os seguintes cenários:

- Engenharia de Dados / Preparação de Dados. O Apache Spark inclui muitos recursos de linguagem para dar suporte à preparação e ao processamento de

grandes volumes de dados para que possam ser mais valiosos e consumidos por outros serviços no Synapse Analytics. Isso é ativado por meio das bibliotecas do Spark que oferecem suporte ao processamento e à conectividade.

- Aprendizado de Máquina. O Apache Spark vem com MLlib, uma biblioteca de aprendizado de máquina construída sobre o Spark que você pode usar de um pool Spark no Synapse Analytics. Piscinas Spark no Synapse Analytics também incluem Anaconda, uma distribuição Python com uma variedade de pacotes para ciência de dados, incluindo aprendizado de máquina. Quando combinado com o suporte integrado para notebooks, você tem um ambiente para criar aplicativos de aprendizado de máquina.

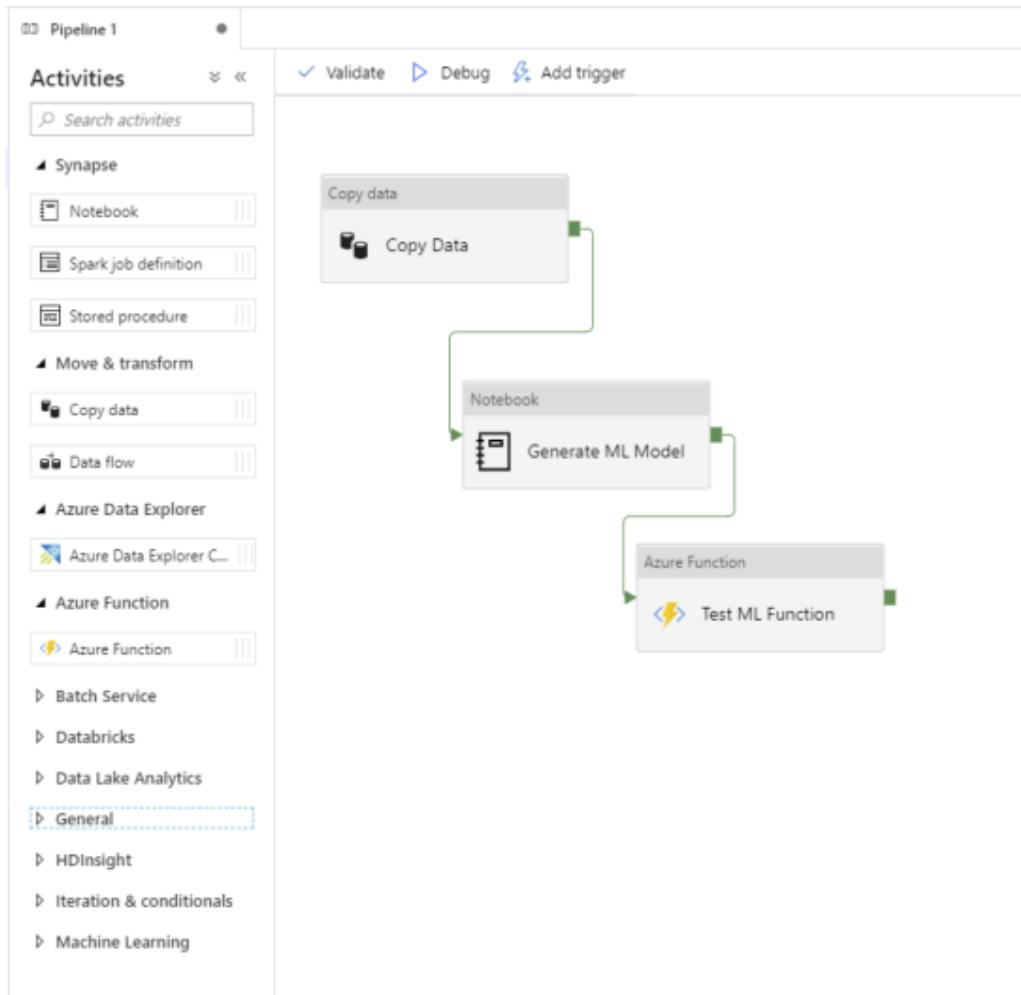
## O que são pipelines de sinapse?

Um pipeline é um agrupamento lógico de atividades que, juntas, executam uma tarefa. Por exemplo, um pipeline pode conter um conjunto de atividades que ingere e limpa dados de log e, em seguida, inicia um fluxo de dados de mapeamento para analisar os dados de log. O pipeline permite que você gerencie as atividades como um conjunto, em vez de cada uma individualmente. Você implanta e programa o pipeline em vez das atividades de forma independente.

As atividades em um pipeline definem ações a serem executadas em seus dados. Por exemplo, você pode usar uma atividade de cópia para copiar dados do Armazenamento de Blob do Azure para a Sinapse do Azure usando um pool SQL. Em seguida, use uma atividade de fluxo de dados ou uma atividade de notebook usando um pool Spark para processar e gerar um modelo de aprendizado de máquina.

Os pipelines de sinapse usam o mesmo mecanismo de integração de dados usado pelo Azure Data Factory. Isso dá a você o poder no Synapse Studio de criar pipelines que podem se conectar a mais de 90 fontes de arquivos simples, bancos de dados ou serviços online. Você pode criar fluxos de dados sem código que permitem fazer mapeamentos e transformações complexos nos dados conforme eles fluem para suas soluções analíticas. O exemplo abaixo mostra um pipeline com três atividades. O pipeline ingere dados e, em seguida, usa um bloco de notas Spark para gerar um modelo de aprendizado de máquina. A função do Azure no final do pipeline testa o modelo de aprendizado de máquina para validá-lo.

CPF: 00296691208



THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## O que é o link Synapse?

O Azure Synapse Link para o Azure Cosmos DB é um recurso de processamento analítico e transacional híbrido nativo da nuvem (HTAP) que permite executar análises quase em tempo real sobre os dados operacionais armazenados no Azure Cosmos DB.

O link de sinapse usa um recurso do Cosmos DB denominado Cosmos DB Analytical Store. O Cosmos DB Analytical Store contém uma cópia dos dados em um contêiner do Cosmos DB, mas organizado como um armazenamento de coluna. A coluna armazena os dados do grupo por coluna em vez de por linha. Os armazenamentos de colunas são um formato mais ideal para a execução de cargas de trabalho analíticas que precisam agregar dados em uma coluna em vez de em uma linha, como gerar somas totais, médias, valores máximos ou mínimos para uma coluna. O Cosmos DB mantém automaticamente os dados em seus contêineres sincronizados com as cópias no armazenamento de coluna.

O Azure Synapse Link permite que você execute cargas de trabalho que recuperam dados diretamente do Cosmos DB e execute cargas de trabalho de análise usando o Azure Synapse Analytics. Os dados não precisam passar por um processo ETL (extrair, transformar e carregar) porque os dados não são copiados para o Synapse Analytics; ele permanece no armazenamento analítico do Cosmos DB.

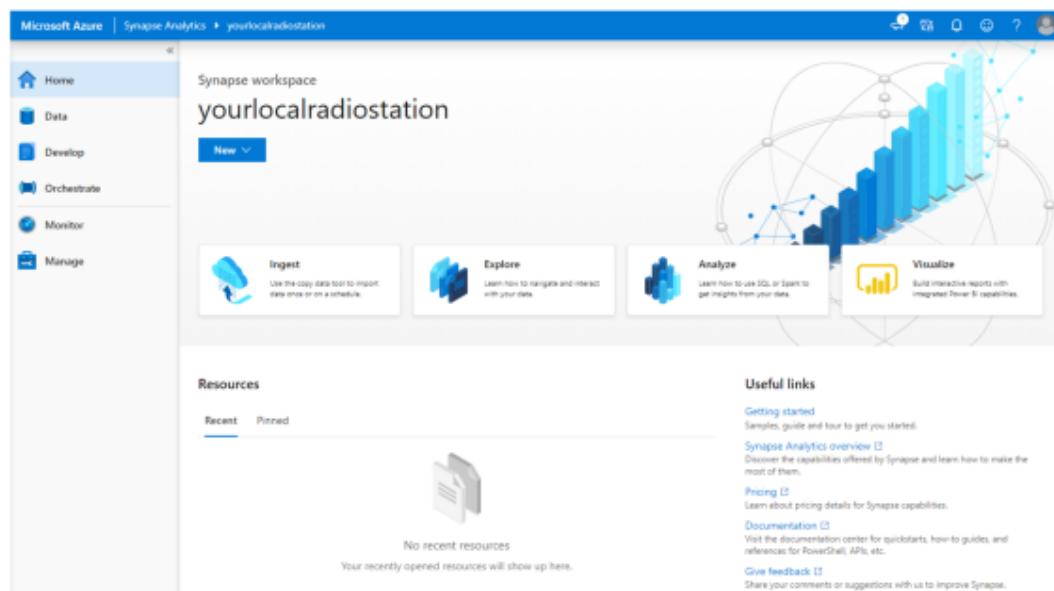
Analistas de negócios, engenheiros de dados e cientistas de dados agora podem usar pools Synapse Spark ou pools Synapse SQL para executar inteligência de negócios, análises e pipelines de aprendizado de máquina quase em tempo real. Você pode conseguir isso sem afetar o desempenho de suas cargas de trabalho transacionais no Azure Cosmos DB.

O link de sinapse tem uma ampla variedade de usos, incluindo:

- Análise e previsão da cadeia de suprimentos. Você pode consultar dados operacionais diretamente e usá-los para construir modelos de aprendizado de máquina. Você pode usar os resultados gerados por esses modelos de volta no Cosmos DB para uma pontuação quase em tempo real. Você pode usar essas avaliações para refinar sucessivamente os modelos e gerar previsões mais precisas.
- Relatório operacional. Você pode usar o Synapse Analytics para consultar dados operacionais usando o Transact-SQL em execução em um pool SQL. Você pode publicar os resultados em painéis usando o suporte fornecido para ferramentas familiares, como o Microsoft Power BI.
- Integração e orquestração de dados em lote. Com as cadeias de suprimentos cada vez mais complexas, as plataformas de dados da cadeia de suprimentos precisam se integrar a uma variedade de fontes e formatos de dados. O mecanismo de integração de dados do Azure Synapse permite que os engenheiros de dados criem pipelines de dados ricos sem exigir um mecanismo de orquestração separado.
- Personalização em tempo real. Você pode criar soluções de comércio eletrônico envolventes que permitem aos varejistas gerar recomendações personalizadas e ofertas especiais para clientes em tempo real.
- Manutenção IoT. As inovações da IoT industrial reduziram drasticamente os tempos de paralisação das máquinas e aumentaram a eficiência geral em todos os campos da indústria. Uma dessas inovações é a análise de manutenção preditiva para máquinas na borda da nuvem. Os dados operacionais históricos dos sensores do dispositivo IoT podem ser usados para treinar modelos preditivos, como detectores de anomalias. Esses detectores de anomalias são então implantados de volta na borda para monitoramento em tempo real. O loopback permite um novo treinamento contínuo dos modelos preditivos.

## O que é Synapse Studio?

Synapse Studio é uma interface da web que permite criar pools e pipelines interativamente. Com o Synapse Studio, você pode desenvolver, testar e depurar blocos de anotações Spark e trabalhos Transact-SQL. Você pode monitorar o desempenho das operações em execução no momento e pode gerenciar os recursos sem servidor ou provisionados. Todos esses recursos são acessados por meio do Synapse Studio nativo da web, que permite gerenciamento, monitoramento, codificação e segurança de modelos.



Você pode acessar o Synapse Studio diretamente do portal do Azure.

THIAGO FERNANDES DA SILVA OLIVEIRA

CNPJ 000296691208

## Teste de conhecimento

### Questão 1

Você tem uma grande quantidade de dados mantidos em arquivos no armazenamento do Azure Data Lake. Você deseja recuperar os dados nesses arquivos e usá-los para preencher as tabelas mantidas no Azure Synapse Analytics. Qual opção de processamento é mais apropriada?

- a) Use o link do Azure Synapse para se conectar ao armazenamento do Azure Data Lake e baixar os dados
- b) **Pool de sinapse SQL**
- c) Piscina Synapse Spark

Explicação: está correto. Você pode usar o PolyBase de um pool SQL para se conectar aos arquivos no Azure Data Lake como tabelas externas e, em seguida, ingerir os dados.

### Questão 2

Qual dos componentes do Azure Synapse Analytics permite que você treine modelos de IA usando o AzureML?

- a) Synapse Studio
- b) Pipelines de sinapse
- c) **Synapse Spark**

Explicação: está correto. Você usaria um bloco de notas para ingerir e modelar dados e, em seguida, usar SparkML e AzureML para treinar modelos com ele.

### Questão 3

No Azure Databricks, como você altera o idioma que uma célula usa?

- a) **A primeira linha da célula é% language. Por exemplo,% scala.**
- b) Altere o idioma do notebook antes de escrever os comandos
- c) Envolva o comando na célula com ## language ##.

Explicação: está correto. Cada célula pode começar com uma definição de idioma.

## Resumo

Nesta lição, você aprendeu sobre:

- Opções de processamento de dados para realizar análises no Azure
- Azure Synapse Analytics

## Saber mais

- O que é Azure Databricks?
- O que é o Azure Synapse Analytics?
- O que é Azure HDInsight?
- Azure Machine Learning (AzureML)
- Pipelines e atividades no Azure Data Factory
- Tutorial: Extrair, transformar e carregar dados usando Azure Databricks
- Link de sinapse do Azure para Azure Cosmos DB: casos de uso de análise quase em tempo real

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Introdução

Microsoft Power BI é uma coleção de serviços de software, aplicativos e conectores que trabalham juntos para transformar suas fontes de dados não relacionadas em percepções coerentes, visualmente imersivas e interativas. Quer seus dados sejam uma pasta de trabalho simples do Microsoft Excel ou uma coleção de data warehouses híbridos baseados em nuvem e no local, Power BI permite que você se conecte facilmente às suas fontes de dados, visualize (ou descubra) o que é importante e compartilhe isso com quem você quiser.

Power BI pode ser simples e rápido, capaz de criar percepções rápidas a partir de uma pasta de trabalho do Excel ou de um banco de dados local. Mas Power BI também é robusto e de nível empresarial, pronto não apenas para modelagem extensiva e análises em tempo real, mas também para desenvolvimento personalizado. Portanto, pode ser seu relatório pessoal e ferramenta de visualização, mas também pode servir como mecanismo de análise e decisão por trás de projetos de grupo, divisões ou corporações inteiras.

Se você é um principiante com o Power BI, esta lição irá ajudá-lo. Se você é um Power BI veterano, esta lição ligará os conceitos e preencherá as lacunas.

## As partes do Power BI

O Power BI consiste em um aplicativo de desktop Microsoft Windows chamado Power BI Desktop, um SaaS online (Software as a Service) serviço chamado de Serviço Power BI e Power BI móvel apps que estão disponíveis em qualquer dispositivo, com aplicativos nativos de BI móvel para Windows, iOS e Android.

Esses três elementos—Área de Trabalho, a serviço, e Móvel aplicativos - são projetados para permitir que as pessoas criem, compartilhem e consumam insights de negócios da maneira que lhes sirva, ou sua função, de forma mais eficaz.

## Como o Power BI corresponde à sua função

O modo como você usa o Power BI pode depender de sua função em um projeto ou equipe. E outras pessoas, em outras funções, podem usar o Power BI de maneira diferente, o que é ótimo.

Por exemplo, você pode visualizar relatórios e painéis no Serviço Power BI, e isso pode ser tudo o que você faz com o Power BI. Mas seu colega de trabalho que analisa números e cria relatórios de negócios pode fazer uso extensivo de Power BI Desktop (e publicar relatórios do Power BI Desktop para o serviço do Power BI, que você usa para exibi-los). E outro colega de trabalho, em vendas, pode usar principalmente seu aplicativo de telefone Power BI para monitorar o progresso em suas cotas de vendas e detalhar novos leads de vendas.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Você também pode usar cada elemento de Power BI em momentos diferentes, dependendo do que você está tentando alcançar ou de qual é sua função em um determinado projeto ou esforço.

Talvez você visualize o inventário e o progresso da fabricação em um painel em tempo real no serviço e também use Power BI Desktop para criar relatórios para sua própria equipe sobre estatísticas de envolvimento do cliente. Como você usa o Power BI pode depender de qual recurso ou serviço do Power BI é a melhor ferramenta para sua situação. Mas cada parte do Power BI está disponível para você, por isso é tão flexível e atraente.

### Baixe o Power BI Desktop

Você pode baixar o Power BI Desktop da web ou como um aplicativo da Microsoft Store na guia Windows.

| Estratégia de download      | Ligações                     | Notas                                     |
|-----------------------------|------------------------------|---|
| Aplicativo da Windows Store | <a href="#">Loja Windows</a> | Permanecerá atualizado automaticamente    |
| Baixar da web               | <a href="#">Baxe o .msi</a>  | Deve atualizar manualmente periodicamente |

### Faça login no serviço Power BI

Antes de entrar no Power BI, você precisará de uma conta. Para obter um teste gratuito, vá para [app.powerbi.com](#) e inscreva-se com seu endereço de e-mail.

Para obter etapas detalhadas sobre como configurar uma conta, consulte "Faça login no serviço Power BI".

### O fluxo de trabalho no Power BI

Um fluxo comum de trabalho no Power BI começa em Power BI Desktop, onde um relatório é criado. Esse relatório é então publicado no Serviço Power BI e finalmente compartilhado, para que os usuários de Power BI Mobile os aplicativos podem consumir as informações.

Nem sempre acontece assim, e tudo bem. Mas usaremos esse fluxo para ajudá-lo a aprender as diferentes partes do Power BI e como elas se complementam.

Ok, agora que temos uma visão geral deste módulo, o que é Power BI e seus três elementos principais, vamos dar uma olhada em como é usar Power BI.

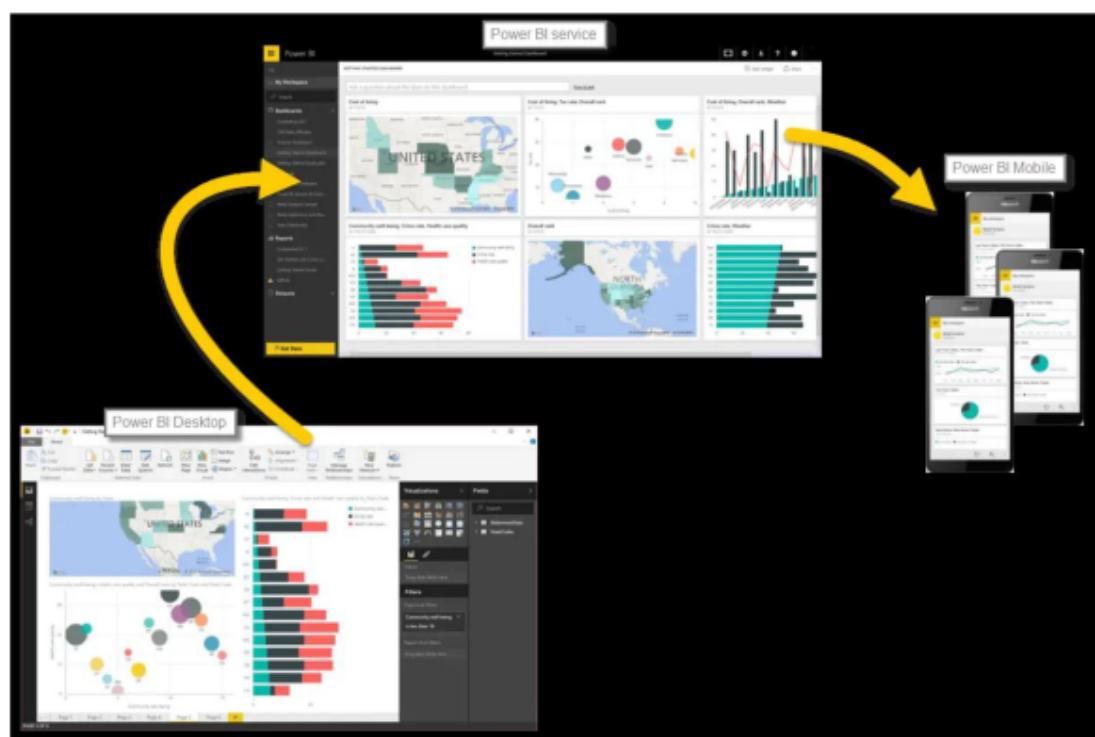
THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Use o Power BI

Agora que apresentamos o básico do Microsoft Power BI, vamos pular para algumas experiências práticas e um tour guiado.

As atividades e análises que você aprenderá com o Power BI geralmente seguem um fluxo comum. O fluxo comum de atividade se parece com isto:

1. Traga dados para o Power BI Desktop e crie um relatório.
2. Publique no serviço Power BI, onde você pode criar novas visualizações ou construir painéis.
3. Compartilhe painéis com outras pessoas, especialmente pessoas que estão em trânsito.
4. Visualize e interaja com painéis e relatórios compartilhados em aplicativos Power BI Mobile.



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 000296691208

Como mencionado anteriormente, você pode passar todo o seu tempo no Serviço Power BI, exibindo recursos visuais e relatórios que foram criados por outros. E tudo bem. Outra pessoa da sua equipe pode passar o tempo Power BI Desktop, o que também é bom. Para ajudá-lo a entender todo o continuum do Power BI e o que ele pode fazer, mostraremos tudo isso. Então você pode decidir como usá-lo da melhor maneira possível.

Então, vamos pular e percorrer a experiência. Sua primeira tarefa é aprender os blocos de construção básicos do Power BI, que fornecerão uma base sólida para transformar dados em relatórios e visuais interessantes.

### Blocos de construção do Power BI

Tudo o que você faz no Microsoft Power BI pode ser dividido em alguns blocos de construção. Depois de entender esses blocos de construção, você pode expandir cada um deles e começar a criar relatórios elaborados e complexos. Afinal, mesmo coisas aparentemente complexas são construídas a partir de blocos de construção básicos. Por exemplo, edifícios são criados com madeira, aço, concreto e vidro, e carros são feitos de metal, tecido e borracha. Obviamente, edifícios e carros também podem ser básicos ou elaborados, dependendo de como esses blocos de construção básicos são organizados.

Vamos dar uma olhada nesses blocos de construção básicos, discutir algumas coisas simples que podem ser construídas com eles e, em seguida, dar uma olhada em como coisas complexas também podem ser criadas.

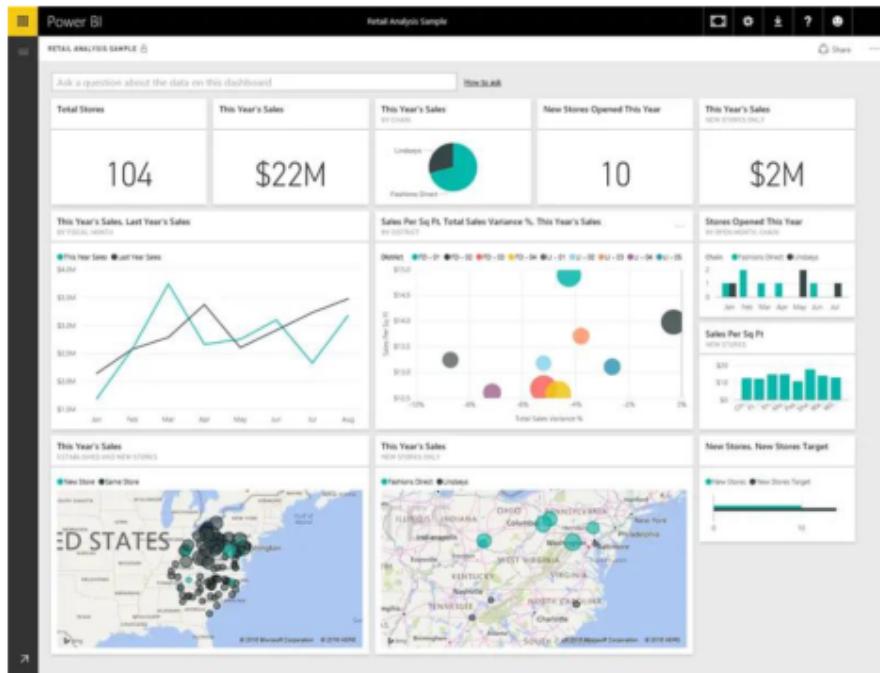
Aqui estão os blocos de construção básicos no Power BI:

- Visualizações
- Conjuntos de dados
- Relatórios
- Dashboards
- Azulejos

### Visualizações

A visualização (às vezes também conhecido como visual) é uma representação visual de dados, como um gráfico, um mapa codificado por cores ou outras coisas interessantes que você pode criar para representar seus dados visualmente. O Power BI tem todos os tipos de tipos de visualização, e mais estão surgindo o tempo todo. A imagem a seguir mostra uma coleção de diferentes visualizações que foram criadas no serviço Power BI.

CPF 00296691208



As visualizações podem ser simples, como um único número que representa algo significativo, ou podem ser visualmente complexas, como um mapa de cor gradiente que mostra o sentimento do eleitor sobre uma determinada questão ou preocupação social. O objetivo de um visual é apresentar os dados de uma forma que forneça contexto e percepções, os quais provavelmente seriam difíceis de discernir de uma tabela bruta de números ou texto.

## Conjuntos de dados

A conjunto de dados é uma coleção de dados que o Power BI usa para criar suas visualizações.

Você pode ter um conjunto de dados simples baseado em uma única tabela de uma pasta de trabalho do Microsoft Excel, semelhante ao que é mostrado na imagem a seguir.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

|      | B    | C     | D          | E              | F      | G              | H                   |
|------|------|-------|------------|----------------|--------|----------------|---------------------|
| 1    | Year | Month | Month Name | Calendar Month | Births | Births Per Day | Births (Normalized) |
| 2119 | 2004 | 1     | January    | 1/1/2004       | 2,937  | 94.7           | 2842                |
| 2120 | 2004 | 2     | February   | 2/1/2004       | 2,824  | 97.4           | 2921                |
| 2121 | 2004 | 3     | March      | 3/1/2004       | 3,128  | 100.9          | 3027                |
| 2122 | 2004 | 4     | April      | 4/1/2004       | 2,896  | 96.5           | 2896                |
| 2123 | 2004 | 5     | May        | 5/1/2004       | 3,008  | 97.0           | 2911                |
| 2124 | 2004 | 6     | June       | 6/1/2004       | 3,047  | 101.6          | 3047                |
| 2125 | 2004 | 7     | July       | 7/1/2004       | 2,981  | 96.2           | 2885                |
| 2126 | 2004 | 8     | August     | 8/1/2004       | 3,079  | 99.3           | 2980                |
| 2127 | 2004 | 9     | September  | 9/1/2004       | 3,219  | 107.3          | 3219                |
| 2128 | 2004 | 10    | October    | 10/1/2004      | 3,547  | 114.4          | 3433                |
| 2129 | 2004 | 11    | November   | 11/1/2004      | 3,365  | 112.2          | 3365                |
| 2130 | 2004 | 12    | December   | 12/1/2004      | 3,143  | 101.4          | 3042                |
| 2131 | 2005 | 1     | January    | 1/1/2005       | 2,921  | 94.2           | 2827                |
| 2132 | 2005 | 2     | February   | 2/1/2005       | 2,699  | 96.4           | 2892                |
| 2133 | 2005 | 3     | March      | 3/1/2005       | 3,024  | 97.5           | 2926                |
| 2134 | 2005 | 4     | April      | 4/1/2005       | 3,037  | 101.2          | 3037                |
| 2135 | 2005 | 5     | May        | 5/1/2005       | 3,231  | 104.2          | 3127                |
| 2136 | 2005 | 6     | June       | 6/1/2005       | 3,163  | 105.4          | 3163                |
| 2137 | 2005 | 7     | July       | 7/1/2005       | 3,119  | 100.6          | 3018                |
| 2138 | 2005 | 8     | August     | 8/1/2005       | 3,156  | 101.8          | 3054                |
| 2139 | 2005 | 9     | September  | 9/1/2005       | 3,439  | 114.6          | 3439                |

Conjuntos de dados também pode ser uma combinação de muitas fontes diferentes, que você pode filtrar e combinar para fornecer uma coleção exclusiva de dados (um conjunto de dados) para uso no Power BI.

Por exemplo, você pode criar um conjunto de dados a partir de três campos de banco de dados, uma tabela de site, uma tabela do Excel e resultados online de uma campanha de marketing por email. Essa combinação única ainda é considerada um único conjunto de dados, embora tenha sido reunido de muitas fontes diferentes.

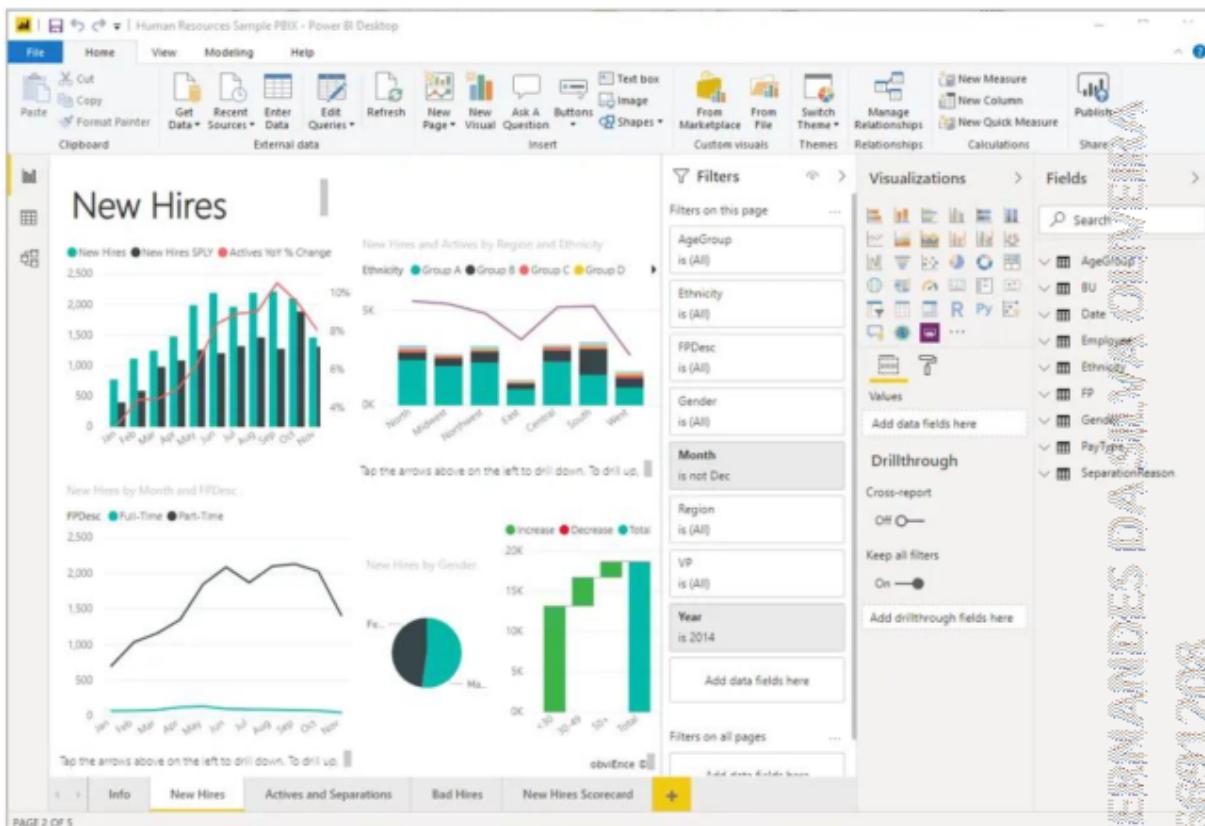
Filtrar dados antes de colocá-los no Power BI permite que você se concentre nos dados que são importantes para você. Por exemplo, você pode filtrar seu banco de dados de contatos para que apenas os clientes que receberam e-mails da campanha de marketing sejam incluídos no conjunto de dados. Você pode então criar recursos visuais com base nesse subconjunto (a coleção filtrada) de clientes que foram incluídos na campanha. A filtragem ajuda você a concentrar seus dados - e seus esforços.

Uma parte importante e habilitadora do Power BI é a infinidade de dados conectores que estão incluídos. Quer os dados que você deseja estejam no Excel ou em um banco de dados Microsoft SQL Server, no Azure ou Oracle, ou em um serviço como o Facebook, Salesforce ou MailChimp, o Power BI tem conectores de dados integrados que permitem que você se conecte facilmente a esses dados, filtrar se necessário, e coloque-o em seu conjunto de dados.

Depois de ter um conjunto de dados, você pode começar a criar visualizações que mostram diferentes partes dele de maneiras diferentes e obter insights com base no que você vê. É aí que entram os relatórios.

## Relatórios

No Power BI, um relatório é uma coleção de visualizações que aparecem juntas em uma ou mais páginas. Assim como qualquer outro relatório que você possa criar para uma apresentação de vendas ou escrever para um trabalho escolar, um relatório no Power BI é uma coleção de itens relacionados entre si. A imagem a seguir mostra um relatório no Power BI Desktop - neste caso, é a segunda página em um relatório de cinco páginas. Você também pode criar relatórios no serviço Power BI.



Os relatórios permitem que você crie muitas visualizações, em várias páginas, se necessário, e organiza essas visualizações da maneira que melhor conta sua história.

Você pode ter um relatório sobre vendas trimestrais, crescimento do produto em um segmento específico ou padrões de migração de ursos polares. Seja qual for o seu assunto, os relatórios permitem que você reúna e organize suas visualizações em uma página (ou mais).

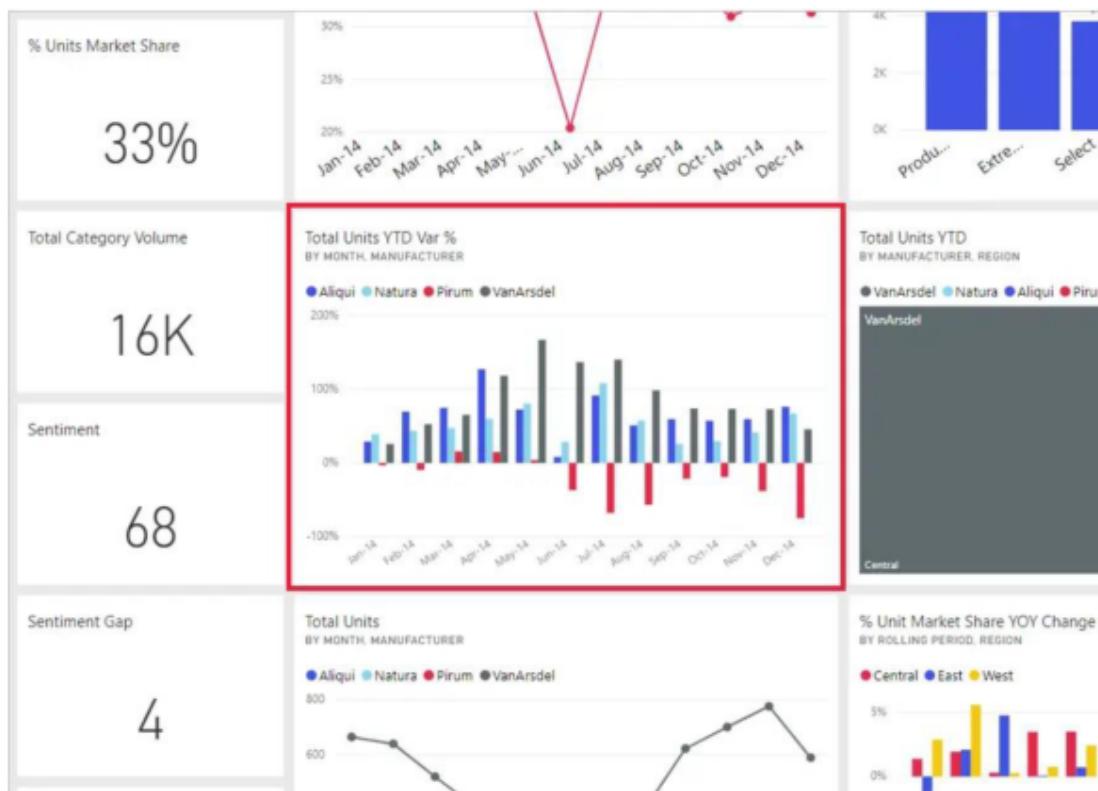
## Dashboards

Quando estiver pronto para compartilhar uma única página de um relatório ou uma coleção de visualizações, você cria um painel de controle. Muito parecido com o painel de um carro, um painel de controle é uma coleção de recursos visuais de uma única página que você pode compartilhar com outras pessoas. Frequentemente, é um grupo selecionado de recursos visuais que fornecem uma visão rápida dos dados ou da história que você está tentando apresentar.

Um painel deve caber em uma única página, geralmente chamada de tela (a tela é o fundo em branco no Power BI Desktop ou o serviço, onde você coloca visualizações). Pense nisso como a tela que um artista ou pintor usa - um espaço de trabalho onde você cria, combina e retrabalha visuais interessantes e atraentes. Você pode compartilhar painéis com outros usuários ou grupos, que podem interagir com seus painéis quando estiverem no serviço Power BI ou em seus dispositivos móveis.

## Blocos

No Power BI, blocos são uma visualização única em um relatório ou painel. É a caixa retangular que contém um visual individual. Na imagem a seguir, você vê um bloco, que também está cercado por outros blocos.



Quando é criado um relatório ou painel no Power BI, você pode mover ou organizar os blocos como quiser. Você pode torná-los maiores, alterar sua altura ou largura e acomodá-los em outras peças.

## Todos juntos agora

Esses são os princípios básicos do Power BI e seus blocos de construção. Vamos fazer uma revisão.

O Power BI é uma coleção de serviços, aplicativos e conectores que permite que você se conecte aos seus dados, onde quer que residam, filtre-os se necessário e, em seguida, leve-os para o Power BI para criar visualizações atraentes que você pode compartilhar com outras pessoas.

Agora que você aprendeu sobre os vários blocos de construção básicos do Power BI, deve ficar claro que você pode criar conjuntos de dados que fazem sentido para você e crie relatórios visualmente atraentes que contam sua história. As histórias contadas com o Power BI não precisam ser complexas ou complicadas para serem atraentes.

Para algumas pessoas, usar uma única tabela do Excel em um conjunto de dados e, em seguida, compartilhar um painel com sua equipe será uma forma incrivelmente valiosa de usar o Power BI.

Para outros, o valor do Power BI estará no uso de tabelas do Azure SQL Data Warehouse em tempo real que se combinam com outros bancos de dados e fontes em tempo real para construir um conjunto de dados momento a momento.

Para ambos os grupos, o processo é o mesmo: criar conjuntos de dados, construir visuais atraentes e compartilhá-los com outras pessoas. E o resultado também é o mesmo para os dois grupos: aproveite seu mundo cada vez maior de dados e transforme-o em percepções açãoáveis.

Quer seus insights de dados exijam conjuntos de dados simples ou complexos, o Power BI ajuda você a começar rapidamente e pode expandir com suas necessidades para ser tão complexo quanto o seu mundo de dados exigir. E como o Power BI é um produto da Microsoft, você pode contar com ele sendo robusto, extensível, compatível com o Microsoft Office e pronto para empresas.

Agora vamos ver como isso funciona. Começaremos dando uma olhada rápida no serviço Power BI.

TITULO DO CAPÍTULO

CNPJ 000296691208

## Faça um tour e use o Power-BI

Como acabamos de aprender, o fluxo comum de trabalho no Microsoft Power BI é criar um relatório no Power BI Desktop, publicá-lo no serviço Power BI e, em seguida, compartilhá-lo com outras pessoas, para que possam visualizá-lo no serviço ou no um aplicativo móvel.

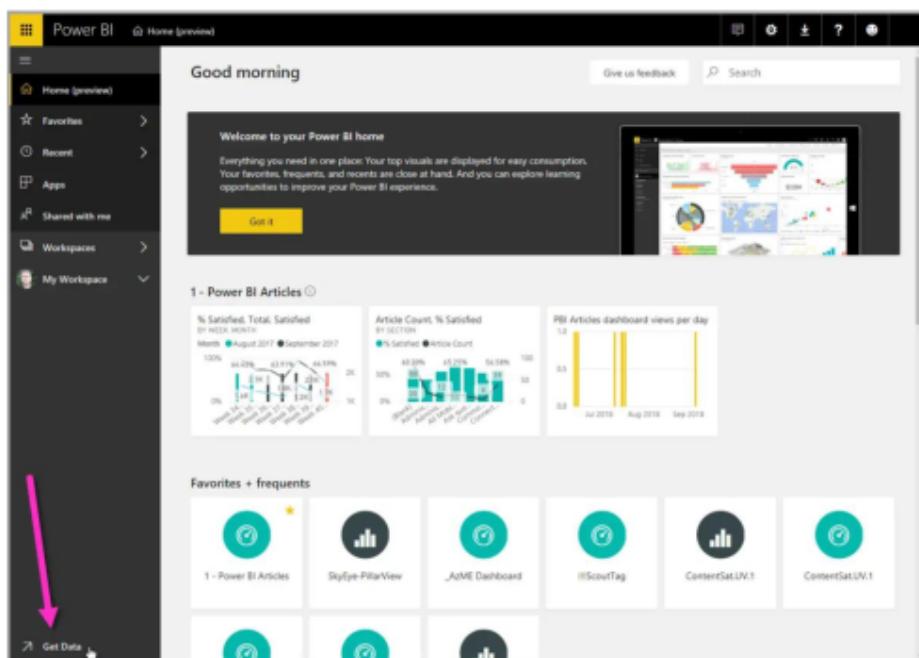
Mas, como algumas pessoas começam no serviço do Power BI, vamos dar uma olhada nisso primeiro e aprender sobre uma maneira fácil e popular de criar visuais rapidamente no Power BI: **apps**.

A aplicativo é uma coleção de visuais e relatórios predefinidos e prontos que são compartilhados com toda a organização. Usar um aplicativo é como colocar um jantar na TV no micro-ondas ou pedir uma refeição de valor fast-food: você só precisa pressionar alguns botões ou fazer alguns comentários, e rapidamente é servido uma coleção de pratos feitos para combinar, todos apresentados em um pacote arrumado e pronto para consumir.

Então, vamos dar uma olhada rápida nos aplicativos, no serviço e como ele funciona. Você pode pensar nisso como um sabor para abrir o apetite.

## Crie painéis prontos para uso com serviços em nuvem

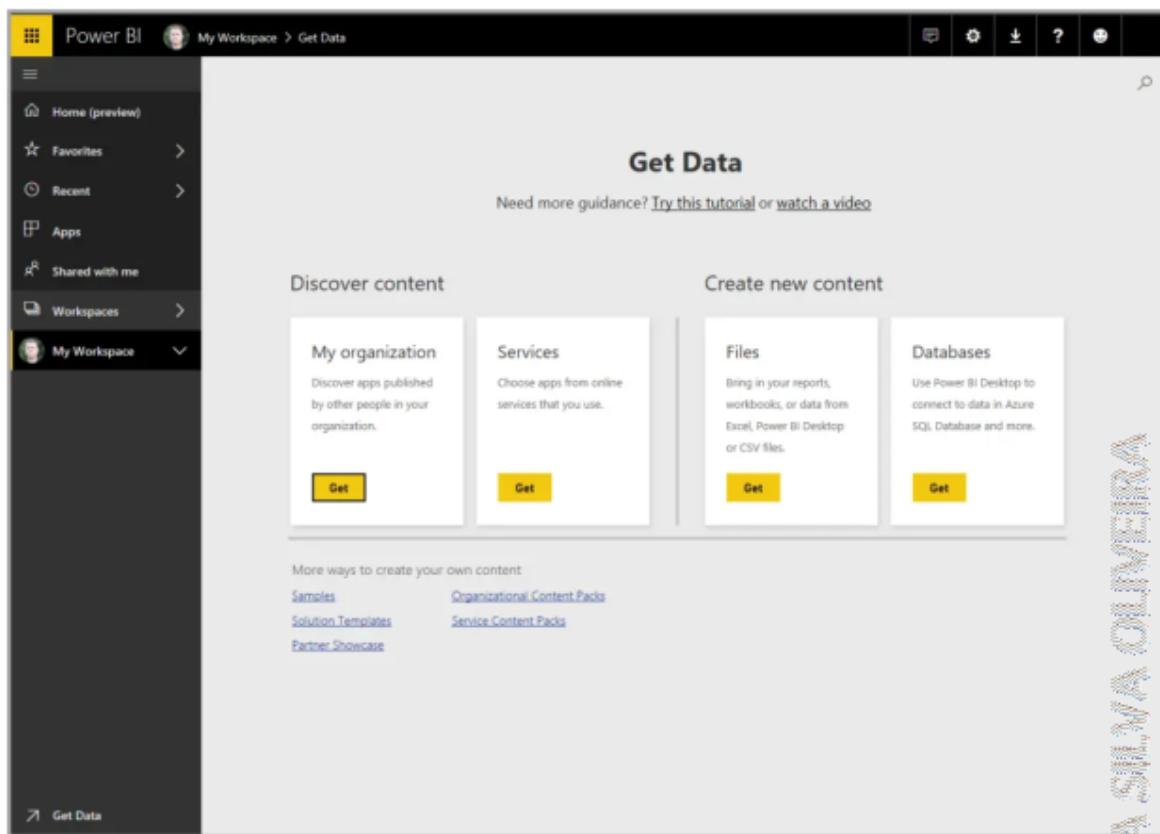
Com o Power BI, é fácil conectar-se aos dados. No serviço Power BI, você pode apenas selecionar **Obter dados** no canto inferior esquerdo da página inicial.



THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

o tela de pintura(a área no centro do serviço Power BI) mostra as fontes de dados disponíveis no serviço Power BI. Além de fontes de dados comuns como arquivos do Microsoft Excel, bancos de dados ou dados do Microsoft Azure, o Power BI pode se conectar facilmente a uma variedade de **serviços de software** (também chamados de provedores SaaS ou serviços em nuvem): Salesforce, Facebook, Google Analytics e mais.

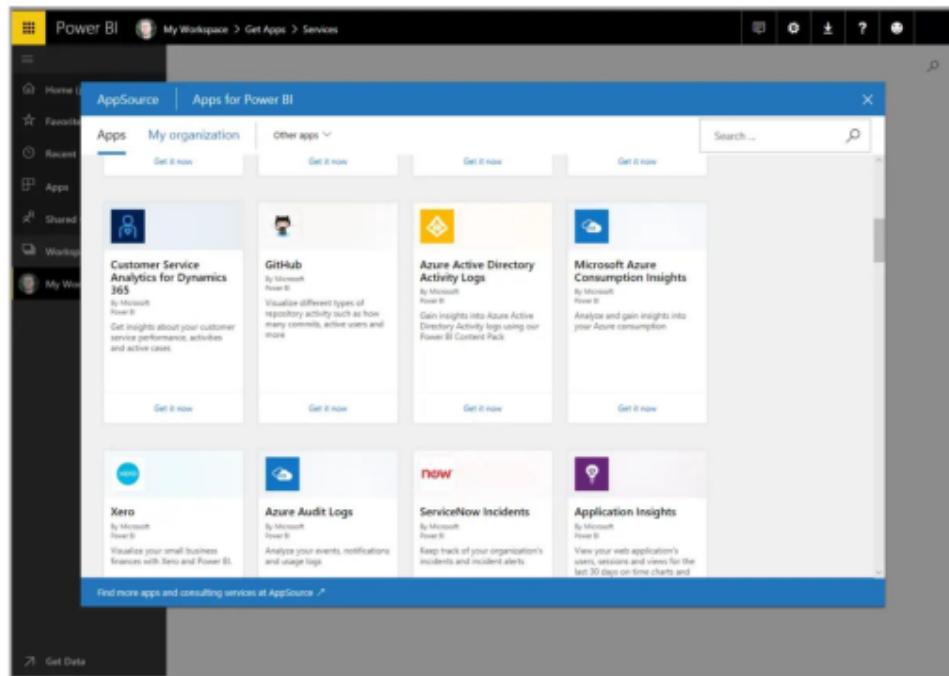


TITULO: FUNDAMENTOS DE DADOS DO MICROSOFT AZURE

CNPJ: 00296691208

Para esses serviços de software, o Serviço Power BI fornece uma coleção de recursos visuais prontos que são pré-organizados em painéis e relatórios para sua organização. Esta coleção de recursos visuais é chamada de aplicativo. Os aplicativos colocam você em funcionamento rapidamente, com dados e painéis que sua organização criou para você. Por exemplo, quando você usa o aplicativo GitHub, o Power BI se conecta à sua conta GitHub (depois de fornecer suas credenciais) e, em seguida, preenche uma coleção predefinida de visuais e painéis no Power BI.

Existem aplicativos para todos os tipos de serviços online. A imagem a seguir mostra uma página de aplicativos que estão disponíveis para diferentes serviços online, em ordem alfabética. Esta página é mostrada quando você seleciona Obter em Serviços (mostrado na imagem anterior). Como você pode ver na imagem a seguir, existem muitos aplicativos para escolher.



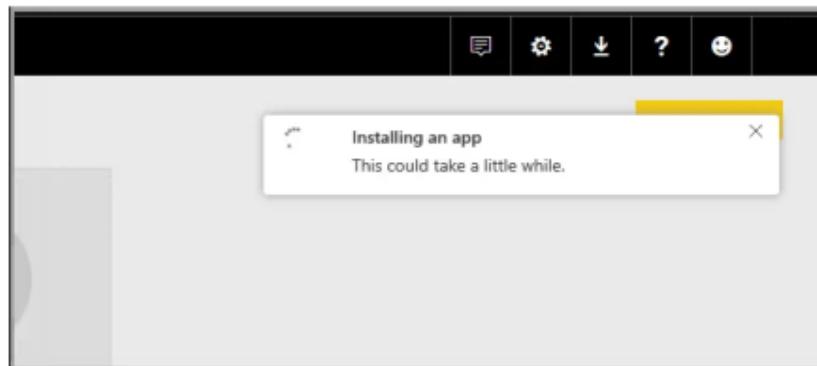
Para nossos propósitos, vamos escolher GitHub. GitHub é um aplicativo para controle de código-fonte online. Quando você seleciona Obtê-lo agora na caixa do aplicativo GitHub, o Conecte-se ao GitHub a caixa de diálogo aparece. Observe que o Github não oferece suporte ao Internet Explorer, portanto, certifique-se de trabalhar em outro navegador.



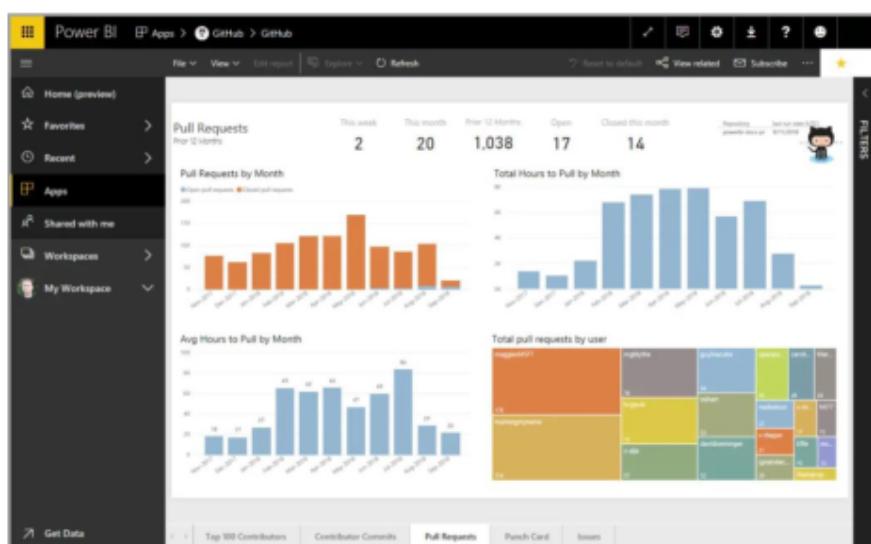
THIAGO FERNANDES DA SILVA OLIVEIRA

CPF: 00296691208

Depois de inserir as informações e credenciais para o aplicativo GitHub, a instalação do aplicativo começa.



Depois que os dados são carregados, o painel predefinido do aplicativo GitHub é exibido.



Além do app **painel de controle**, o **relatório** que foi gerado (como parte do aplicativo GitHub) e usado para criar o painel está disponível, assim como o **conjunto de dados** (a coleção de dados extraídos do GitHub) que foi criado durante a importação de dados e usado para criar o relatório do GitHub.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

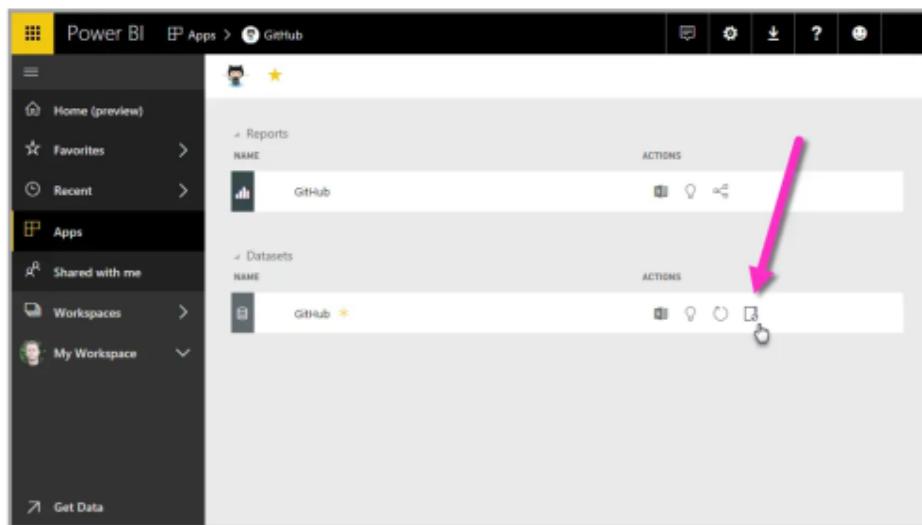
The screenshot shows the Power BI interface with the GitHub app selected. On the left, the navigation pane includes Home (preview), Favorites, Recent, Apps (selected), Shared with me, Workspaces, and My Workspace. The main area displays two sections: 'Reports' and 'Datasets'. Under 'Reports', there is one item named 'GitHub'. Under 'Datasets', there is one item also named 'GitHub'. Two pink arrows point from the text 'Reports' and 'Datasets' towards their respective sections in the interface.

No painel, você pode selecionar qualquer um dos visuais e interagir com eles. Ao fazer isso, todos os outros recursos visuais na página responderão. Por exemplo, quando Maio de 2018 é selecionado no Solicitações pull (por mês), os outros recursos visuais na página se ajustam para refletir essa seleção.

The screenshot shows a detailed Power BI report for GitHub. The left sidebar has the same navigation as the previous screenshot. The main area contains several visualizations: a summary card with metrics for This week, This month, Prior 12 Months, Open, and Closed this month; a bar chart titled 'Pull Requests by Month' comparing open and closed pull requests across months; a bar chart titled 'Total Hours to Pull by Month'; a bar chart titled 'Avg Hours to Pull by Month'; and a treemap titled 'Total pull requests by user' showing the distribution of pull requests among GitHub users. A vertical sidebar on the right is labeled 'FILTERS' and shows a date range from 'May 2018' to 'May 2018'. A large vertical watermark on the right side of the screen reads 'CÓDIGO FEVEREIRO 2018'.

## Atualizar dados no serviço Power BI

Você também pode escolher **atualizar** o conjunto de dados para um aplicativo ou outros dados que você usa no Power BI. Para definir as configurações de atualização, selecione o ícone de atualização de agendamento para o conjunto de dados a ser atualizado e use o menu que aparece. Você também pode selecionar o ícone de atualização (o círculo com uma seta) ao lado do ícone de atualização de programação para atualizar o conjunto de dados imediatamente.



O Conjuntos de dados é selecionado em Configurações. No painel direito, selecione a seta ao lado de Atualização para expandir essa seção. As Configurações aparecem na tela, permitindo definir as configurações de atualização que atendem às suas necessidades.

Settings for GitHub

Next refresh: Wed Sep 12 2018 04:56:35 GMT-0700 (Pacific Daylight Time)  
Refresh history

Gateway connection

Data source credentials

Parameters

Scheduled refresh

Keep your data up to date

On

Refresh frequency: Daily

Time zone: (UTC) Coordinated Universal Time

Add another time

Send refresh failure notification email to me

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

Isso é o suficiente para uma rápida olhada no serviço Power BI. Há muito mais coisas que você pode fazer com o serviço e muitos tipos de dados aos quais você pode se conectar, e todos os tipos de aplicativos.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208

## Teste de conhecimento

### Questão 1

**Qual é o fluxo comum de atividade no Power BI?**

- a) Crie um relatório no Power BI móvel, compartilhe-o com o Power BI Desktop, visualize e interaja no serviço Power BI.
- b) Crie um relatório no serviço Power BI, compartilhe-o com o Power BI móvel, interaja com ele no Power BI Desktop.
- c) Traga dados para o Power BI Desktop e crie um relatório, compartilhe-o com o serviço Power BI, visualize e interaja com relatórios e painéis no serviço e no Power BI móvel.
- d) Traga os dados para o Power BI móvel, crie um relatório e compartilhe-o com o Power BI Desktop.

Explicação: está correto. O serviço Power BI permite exibir e interagir com relatórios e painéis, mas não permite modelar dados.

### Questão 2

**Quais das opções a seguir são blocos de construção do Power BI?**

- a) Blocos, painéis, bancos de dados, dispositivos móveis.
- b) Visualizações, conjuntos de dados, relatórios, painéis, blocos.
- c) Arquivos Visual Studio, C # e JSON.

Explicação: está correto. Os blocos de construção do Power BI são visualizações, conjuntos de dados, relatórios, painéis, blocos.

### Questão 3

**Uma coleção de elementos visuais prontos, pré-organizados em painéis e relatórios, é chamada de o que no Power BI?**

- a) A tela.
- b) Atualização programada.
- c) Um aplicativo.

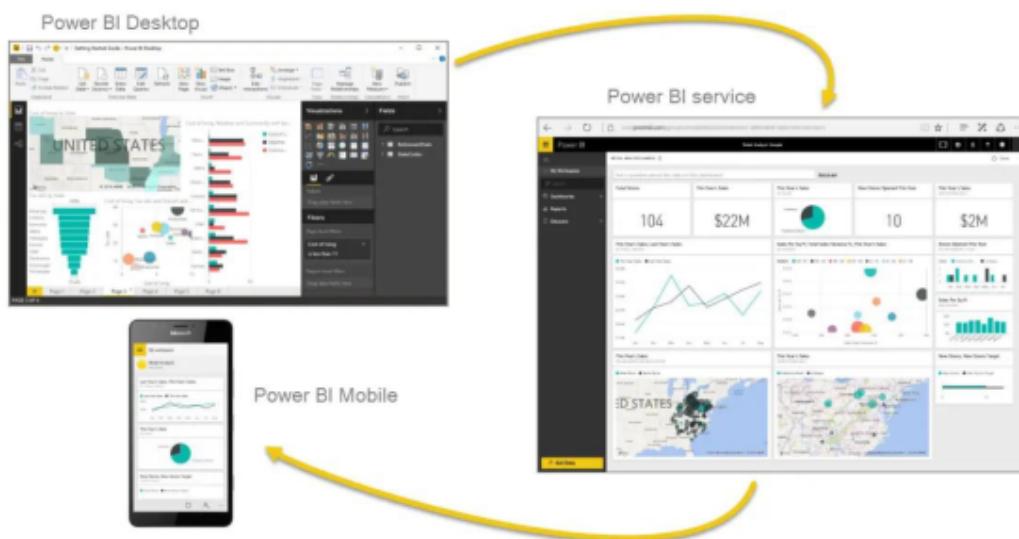
Explicação: está correto. Um aplicativo é uma coleção de recursos visuais prontos, pré-organizados em painéis e relatórios. Você pode obter aplicativos que se conectam a muitos serviços online do AppSource.

## Resumo

Vamos fazer uma revisão rápida do que abordamos nesta lição.

Microsoft Power BI é uma coleção de serviços de software, aplicativos e conectores que funcionam juntos para transformar seus dados em percepções interativas. Você pode usar dados de fontes básicas únicas, como uma pasta de trabalho do Microsoft Excel, ou extrair dados de vários bancos de dados e fontes de nuvem para criar conjuntos de dados e relatórios complexos. O Power BI pode ser tão simples quanto você deseja ou tão pronto para a empresa quanto seus complexos negócios globais exigem.

O Power BI consiste em três elementos principais — Power BI Desktop, o Serviço Power BI, e Power BI Mobile — que trabalham juntos para permitir que você crie, interaja, compartilhe e consuma seus dados da maneira que desejar.



Também discutimos os blocos de construção básicos no Power BI:

- Visualizações - Uma representação visual de dados, às vezes apenas chamada de visual
- Conjuntos de dados - Uma coleção de dados que o Power BI usa para criar visualizações
- Relatórios - Uma coleção de elementos visuais de um conjunto de dados, abrangendo uma ou mais páginas
- Dashboards - Uma coleção de recursos visuais de uma única página criada a partir de um relatório
- Blocos - Uma única visualização em um relatório ou painel

CNPJ 00296691208  
TITULO FERNANDES DA SILVA OLIVEIRA

No Serviço Power BI, nós instalamos um aplicativo em apenas alguns cliques. O aplicativo, uma coleção pronta de recursos visuais e relatórios, nos permite conectar facilmente a um serviço de software para preencher o aplicativo e dar vida aos dados.

Finalmente, configuramos um cronograma de atualização para nossos dados, para termos a certeza de que os dados estarão atualizados quando voltarmos para o serviço do Power BI.

THIAGO FERNANDES DA SILVA OLIVEIRA  
CPF 00296691208