

Open Edu Analytics (OEA)

Guia de Solução

Publicação: Agosto, 2021

Introdução	1
1) Setup da Arquitetura do OEA	2
2) Percorrendo o exemplo incluído	5
3) Mais sobre Synapse Studio	6
4) Exemplos de Dashboards no Power BI	7
5) Conectando com Workspaces no Power BI	9
6) Privacidade e Segurança	10

Introdução

Este documento fornece instruções passo a passo para a configuração da arquitetura de referência do OEA, bem como informações sobre como implantar módulos e pacotes na arquitetura base.

A OEA é uma solução de propriedade de dados moderna de código aberto para educação, construída no [Synapse Analytics](#) e no poderoso conjunto de serviços de dados da plataforma Azure.

Para um conjunto de breves vídeos introdutórios no Synapse Analytics, consulte: [Azure Synapse Analytics demo videos](#).

Para um guia passo-a-passo através da Synapse Analytics, consulte: [Comece com a Azure Synapse Analytics](#).

Para um e-book detalhado sobre análises no Azure, consulte: [Cloud Analytics com o Microsoft Azure](#)

Todos os scripts e documentação para OEA podem ser encontrados em: <https://github.com/microsoft/OpenEduAnalytics>

O quadro da OEA é composto por:

1. A arquitetura central da estrutura OEA – uma conta de armazenamento Azure, um espaço de trabalho Synapse, uma instância de Application Insights, uma instância do Azure Key Vault, um Apache Spark Pool e um conjunto de grupos de segurança AAD
2. Módulos – Notebooks Apache Spark para o processamento de dados de origem de um sistema de origem específico. Os módulos de dados podem ser vistos como silos de dados, trazendo dados de um único sistema, sem dependências.
3. Pacotes – um pacote de ativos, como notebooks Apache Spark para provisionar uma visão abrangente sobre vários conjuntos de dados, relatórios Power BI e modelos de Machine Learning. Um pacote de soluções utiliza um ou mais módulos de dados para fornecer os dados de origem utilizados na solução.

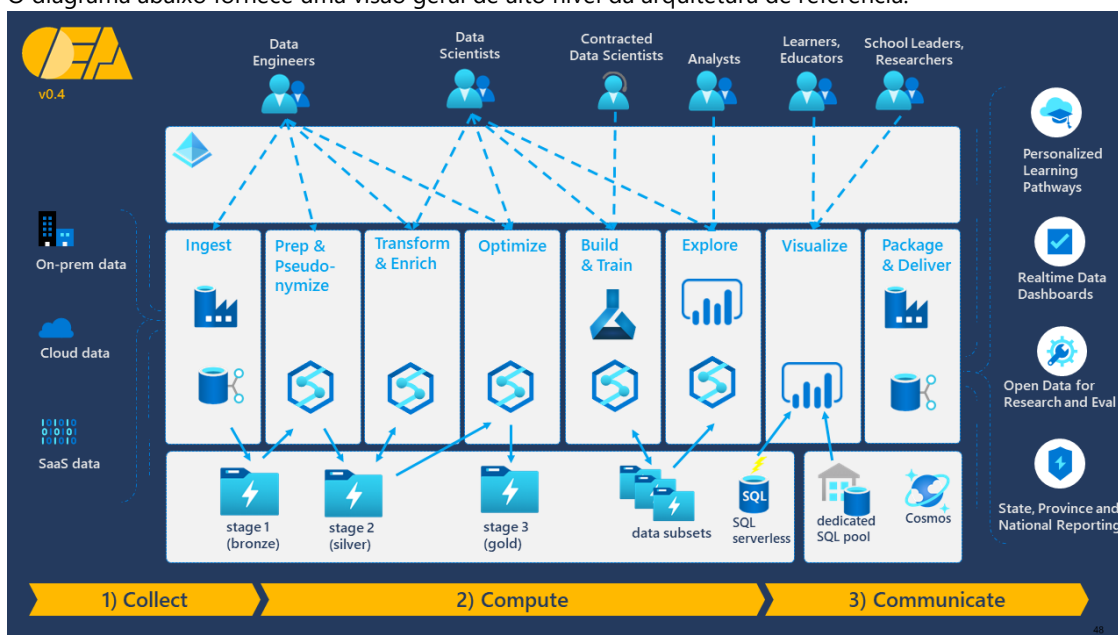
Módulos e pacotes na OEA podem conter o mesmo conjunto de ativos – a principal distinção entre os dois é que os módulos são independentes enquanto os pacotes têm dependências de um ou mais módulos.

Módulos e pacotes têm a seguinte estrutura padrão:

1. um readme.md para documentação básica
2. um script setup.sh a ser usado para implantação automatizada a partir de [cloud shell](#)
3. uma pasta de notebook para notebooks Synapse
4. uma pasta PowerBI para ativos Power BI (isso é opcional)

Para iniciar a configuração da solução OEA, tudo o que você precisa é de uma [assinatura do Azure](#). Consulte a seção a seguir para obter instruções detalhadas de configuração.

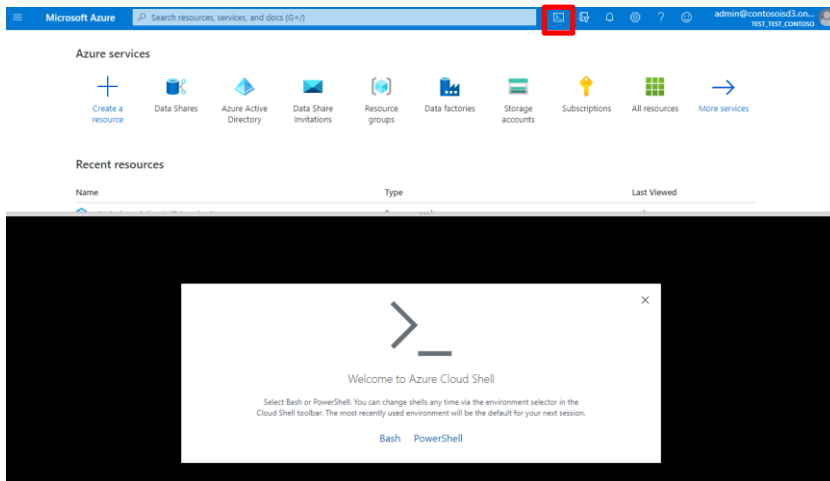
O diagrama abaixo fornece uma visão geral de alto nível da arquitetura de referência.



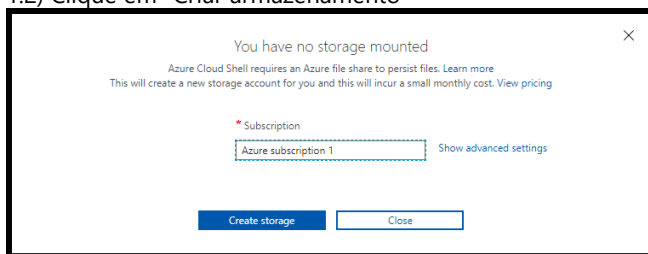
1) Setup da Arquitetura Base OEA

Nesta seção, você usará um script para fornecer os recursos do Azure que compõem o núcleo desta solução, bem como um pacote de soluções de exemplo que fornece conjuntos de dados e notebooks de exemplo para usar para explorar melhor os recursos do Synapse Analytics.

1.1) No portal do Azure, clique no ícone Cloud Shell e selecione "Bash".

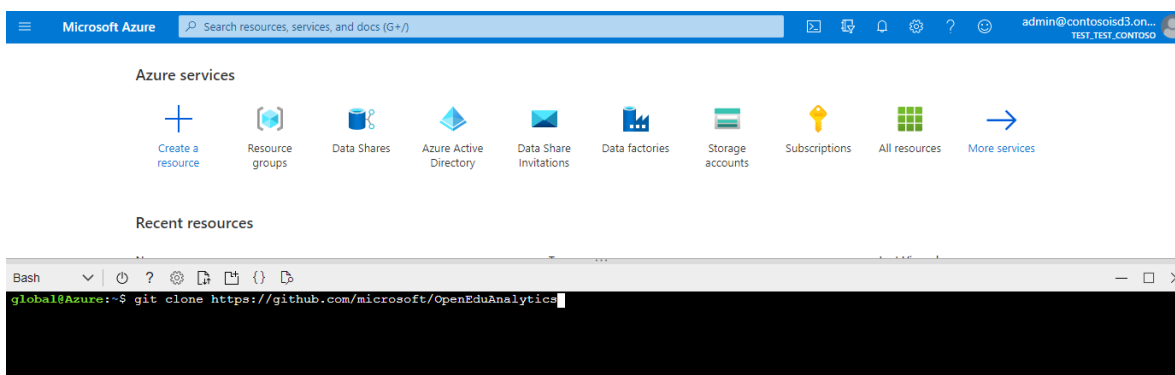


1.2) Clique em "Criar armazenamento"



1.3) No prompt do bash shell, insira os seguintes comandos para baixar o conteúdo do repositório OpenEduAnalytics para a sua unidade de nuvem do Azure.

```
cd clouddrive
git clone https://github.com/microsoft/OpenEduAnalytics
```



1.4) Agora execute o script de configuração encontrado no diretório raiz do OpenEduAnalytics executando os seguintes comandos. Observe que para <unique_suffix> no comando abaixo, você deve inserir um ID para a sua organização que será usado como um sufixo de recursos do Azure que deve ser único. Por exemplo, um distrito escolar chamado Contoso Independent School District pode escolher um ID org de "CISD" ou "ContosoISD".

```
cd OpenEduAnalytics
```

```
./setup.sh <unique_suffix>
```

Os recursos do Azure serão criados na região leste dos EUA por padrão. Para ter os recursos criados em um local diferente, especifique o local desejado como o segundo argumento para o script:

```
./setup.sh <unique_suffix> <location>
```

Para obter uma lista de locais disponíveis, execute este comando em sua linha de comando bash: `az account list-locations`

O script de instalação levará então vários minutos para ser concluído, pois fornece os seguintes recursos do Azure:

1. um grupo de recursos (que serve como um recipiente lógico para o resto dos recursos criados)
2. uma conta de armazenamento com 7 containers de armazenamento (stage1np, stage2np, stage2p, stage3np, stage3p, oea-framework, sinapse)
3. um espaço de trabalho Azure Synapse
4. uma piscina Apache Spark
5. um cofre chave
6. uma instância de Insights de aplicativos

Na foto abaixo estão capturas de tela dos recursos criados:

The screenshot shows the Azure portal interface for the resource group 'rg-oea-cisdggv04k'. The left sidebar contains navigation options like Overview, Activity log, Access control (IAM), Tags, Resource visualizer, Events, Settings, Deployments, Security, Policies, Properties, Locks, and Cost Management. The main area shows the 'Essentials' section with details about the subscription (Azure subscription 1) and location (West US). Below this, a table lists the resources created:

Name	Type	Location
appi-oea-cisdggv04k	Application Insights	West US
kv-oea-cisdggv04k	Key vault	West US
spark3p0sm (syn-oea-cisdggv04k/spark3p0sm)	Apache Spark pool	West US
stoeacisdggv04k	Storage account	West US
syn-oea-cisdggv04k	Synapse workspace	West US

The screenshot shows the Azure portal interface for the storage account 'stoeacisdggv04k'. The left sidebar contains navigation options like Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, Storage Explorer (preview), Data storage, and Containers. The main area shows the 'Containers' section with a table listing the containers created:

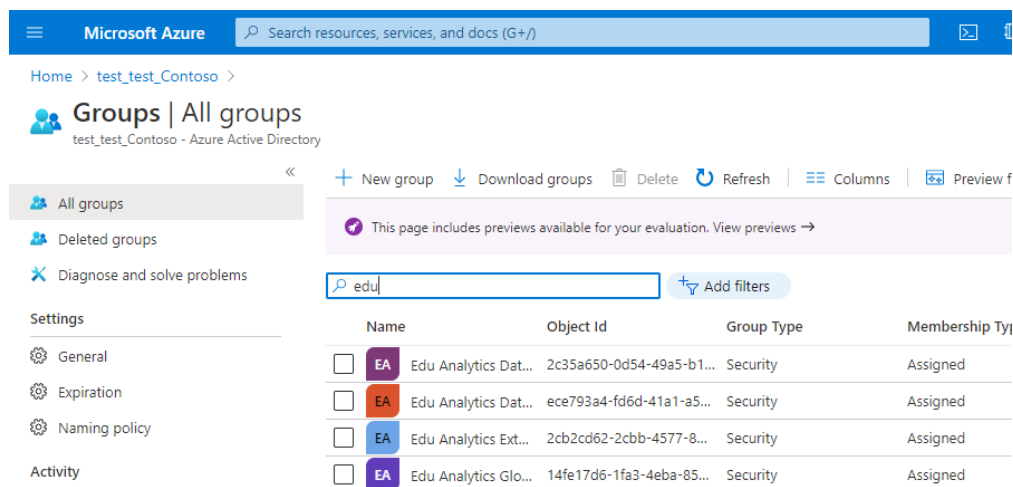
Name	Last modified	Public access level	Lease state
oea-framework	8/23/2021, 3:26:57 PM	Private	Available
stage1np	8/23/2021, 3:26:58 PM	Private	Available
stage2np	8/23/2021, 3:26:59 PM	Private	Available
stage2p	8/23/2021, 3:27:00 PM	Private	Available
stage3np	8/23/2021, 3:27:01 PM	Private	Available
stage3p	8/23/2021, 3:27:02 PM	Private	Available
synapse-workspace	8/23/2021, 3:26:56 PM	Private	Available

Observe também que a configuração automatizada instala conjuntos de dados de teste no contêiner de armazenamento chamado oea-framework. Consulte a seção "Caminhando pelo exemplo" para obter mais detalhes sobre como usar esses dados para executar notebooks de exemplo e aprender mais sobre o Synapse Analytics.

Você também pode optar para que o script crie grupos de segurança facilitando o uso do controle de acesso baseado em função ao lago de dados.

Se você estiver executando a configuração para um ambiente no qual você tem permissões de administração global no inquilino e deseja ter grupos de segurança provisionados, você pode invocar o script de configuração assim:

```
./setup.sh <unique_suffix> <location> true
```



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

Home > test_test_Contoso >

Groups | All groups

test_test_Contoso - Azure Active Directory

« + New group Download groups Delete Refresh Columns Preview f

This page includes previews available for your evaluation. View previews →

edu Add filters

	Name	Object Id	Group Type	Membership Ty
<input type="checkbox"/>	EA Edu Analytics Dat...	2c35a650-0d54-49a5-b1...	Security	Assigned
<input type="checkbox"/>	EA Edu Analytics Dat...	ece793a4-fd6d-41a1-a5...	Security	Assigned
<input type="checkbox"/>	EA Edu Analytics Ext...	2cb2cd62-2cbb-4577-8...	Security	Assigned
<input type="checkbox"/>	EA Edu Analytics Glo...	14fe17d6-1fa3-4eba-85...	Security	Assigned

Depuração

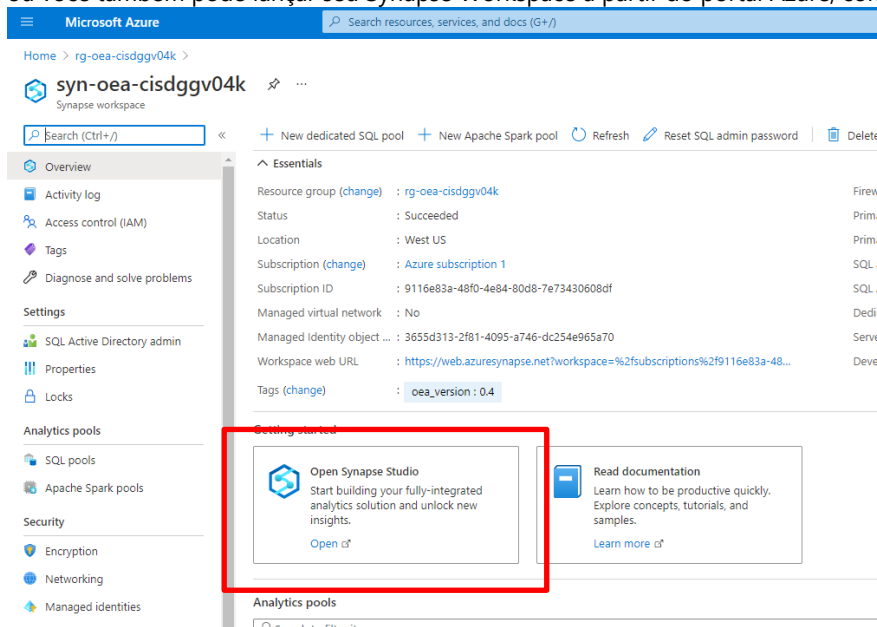
Consulte a wiki OEA para obter as informações de depuração mais [atualizadas: Depuração de problemas na OEA](#)

2) Percorrendo o exemplo incluído

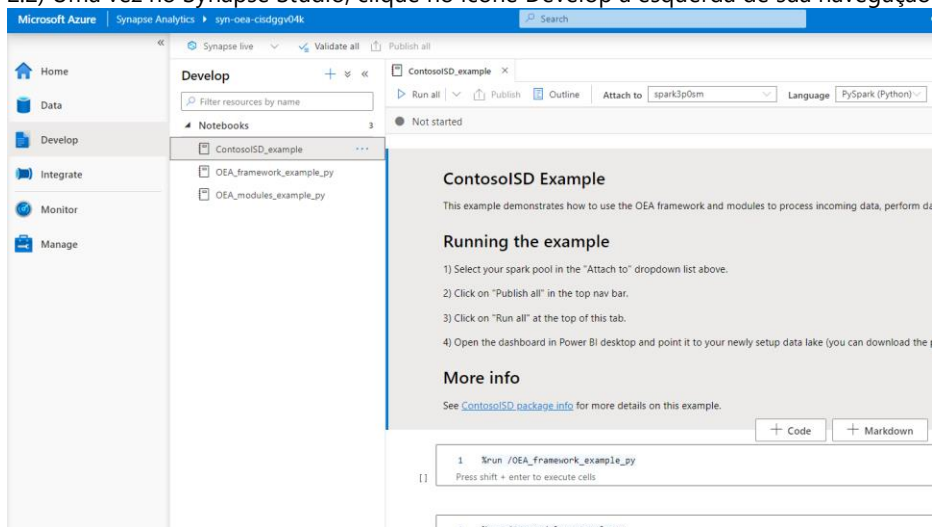
2.1) Abra seu novo Espaço de Trabalho Sinapse clicando na url no final do script de configuração

```
--> Setting up OEA (logging detailed setup messages to oea_setup_20210823_192620.log)
--> Setting up the OEA base architecture.
--> 1) Creating resource group: rg-oea-cisdggv04k
--> 2) Creating storage account: stoacisdggv04k
--> 3) Creating Synapse Workspace: syn-oea-cisdggv04k (this is usually the longest step - it may take 5 to 10 minutes to complete)
--> 4) Creating key vault: kv-oea-cisdggv04k
--> 5) Creating security groups in Azure Active Directory.
--> Setting up the example OEA package.
--> OEA setup is complete. Click on this url to work with your new Synapse workspace (via Synapse Studio): https://web.azuresynapse.net?workspace%2fworkspaces%2fsyn-oea-cisdggv04k
Once in Synapse Studio, click on Develop, select the notebook called ContosoISD_example, and follow the directions shown there.
```

ou você também pode lançar seu Synapse Workspace a partir do portal Azure, como mostrar aqui:



2.2) Uma vez no Synapse Studio, clique no ícone Develop a esquerda de sua navegação e selecione o ContosoISD_example notebook



2.3) Agora siga as instruções naquele notebook para colocar os dados de teste no seu Data Lake, processar esses dados e visualizar os dados no Power BI da área de trabalho.

3) Mais sobre Synapse Studio

OEA foi construído para alavancar o poder do Azure Synapse, e a ferramenta central para trabalhar no Azure Synapse é o Synapse Studio.

Aqui está uma breve lição online sobre o básico do Synapse Studio: [Explore Azure Synapse Studio](#)



✓ 1100 XP

Explore Azure Synapse Studio

43 min • Module • 10 Units

★★★★☆ 4.6 (194)

Rate it

Beginner

Data Engineer

Synapse Analytics

Take a tour of the core application used to interact with the various components of Azure Synapse Analytics.

Learning objectives

In this module, you will:

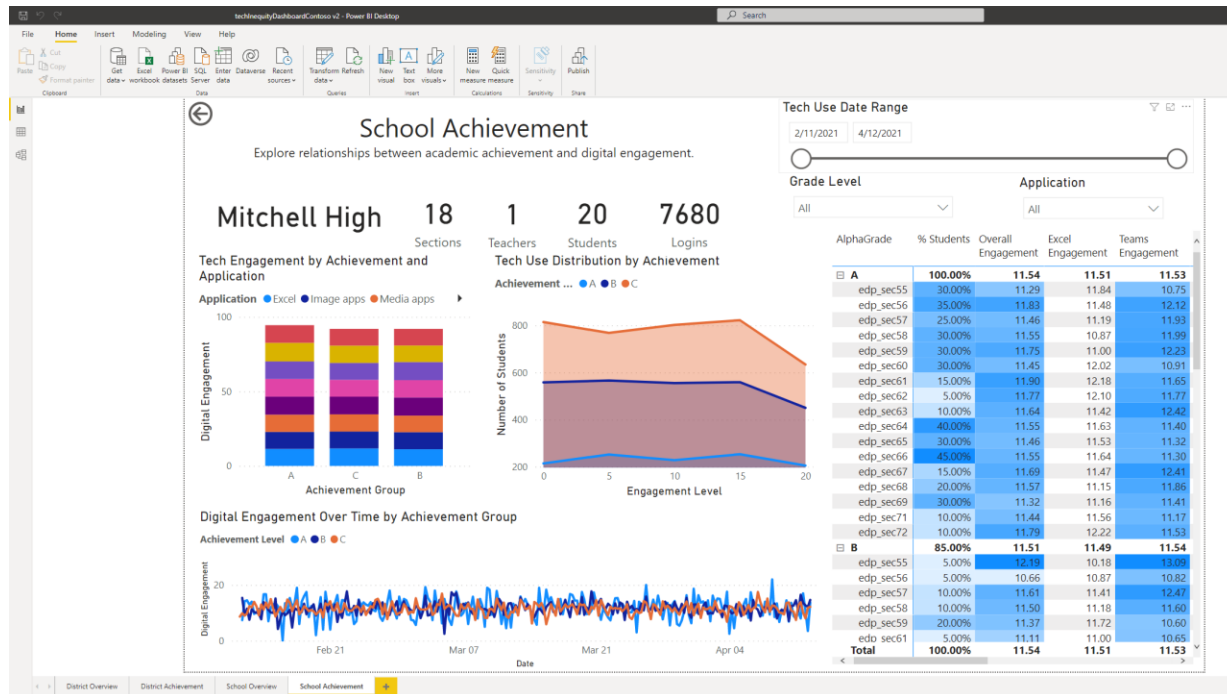
- Use Azure Synapse Studio
- Understand the Azure Synapse Analytical processes
- Explore the Data hub
- Explore the Develop hub
- Explore the Integrate hub
- Explore the Monitor hub
- Explore the Manage hub

4) Exemplos de Dashboards no Power BI

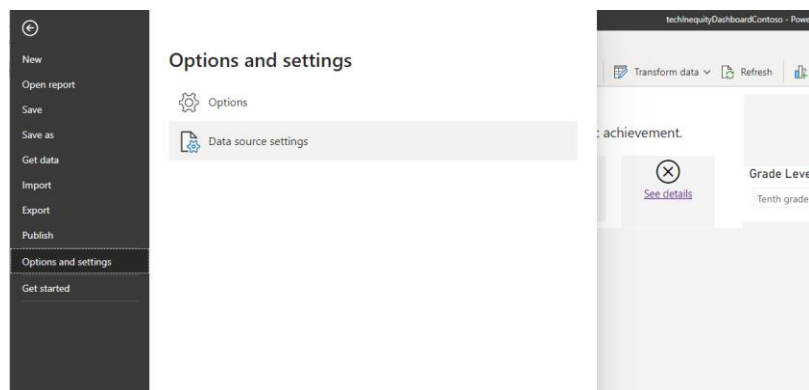
A seção anterior demonstrou as etapas necessárias para uma configuração completa com um ambiente de teste e dados de teste.

Esta seção demonstrará como abrir os exemplos de Dashboard do Power BI, utilizando Power BI Desktop e conectar-se ao Data Lake em seu ambiente de teste via SQL On-Demand. Você precisará ter o Power BI Desktop instalado no seu computador para concluir esta seção (Power BI Desktop é gratuito para download e livre para usar – ele pode ser [baixado a partir daqui](#)).

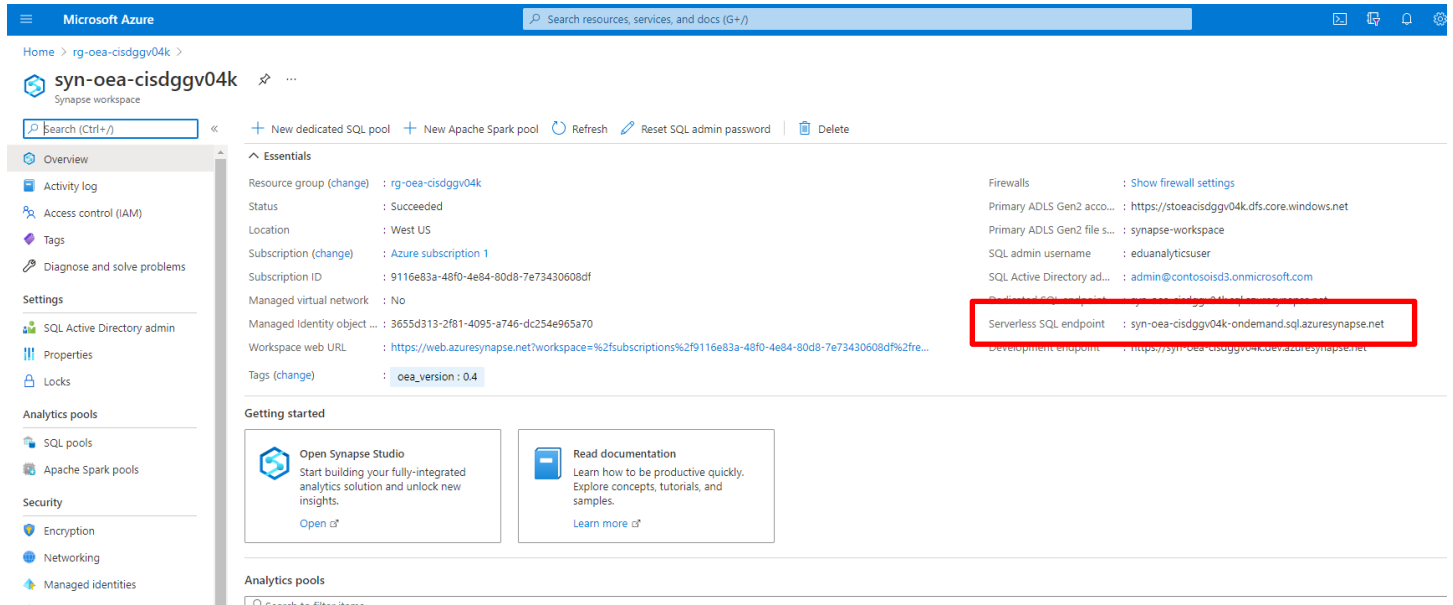
3.1) Baixe e, em seguida, abra este arquivo pbix do OEA GitHub Report: [techInequityDashboardContoso v2.pbix](#)



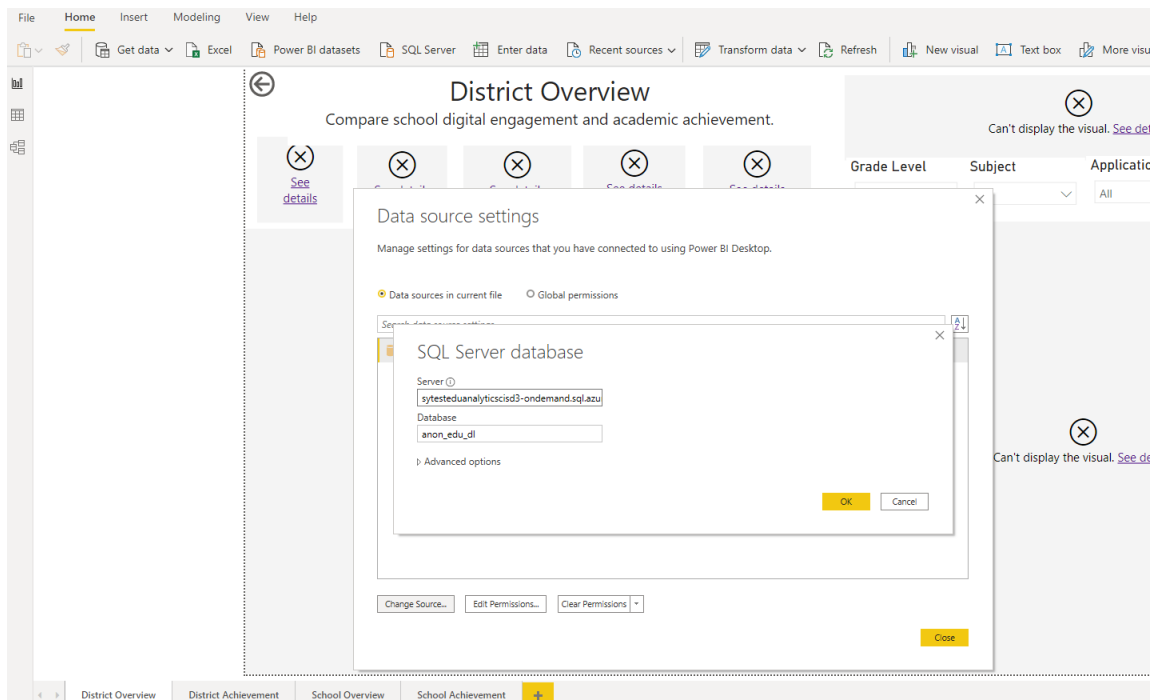
3.2) Os dados que você vê foram incluídos no relatório. Agora clique em File -> Options and Settings -> Data source settings e na próxima tela clique em "Change Source"



3.3) Para obter a url certa para o seu servidor, vá para Portal.azure.com e navegue até a sua instância do Synapse. Você precisa copiar o valor para " Serverless SQL endpoint".



3.5) Digite o valor recuperado na etapa anterior na caixa de texto para "Server" e para "Database" digite "s2_contosoisd", em seguida, clique em " Edit Permissions"



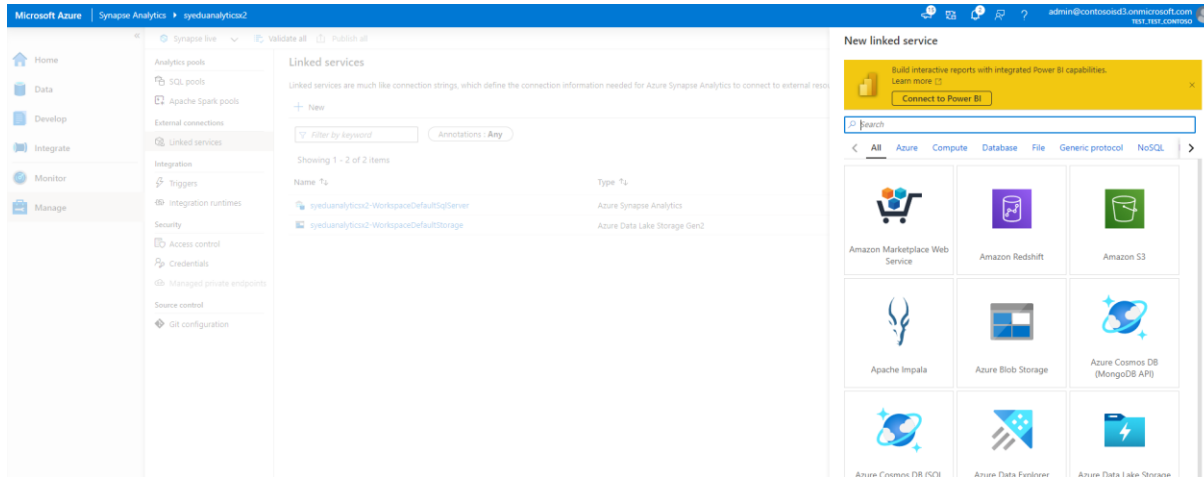
3.6) Clique em "Edit", e na próxima janela clique em " Microsoft account", em seguida, clique em "Sign In", e complete o sinal em processo com as credenciais para o usuário que tem acesso ao espaço de trabalho Sinapse.

Em seguida, clique em "Save", seguido de "OK", seguido de "Close" e clique em " Apply changes".

5) Conectar com Workspace do POWER BI

Caso você possua uma licença Power BI Premium, você terá a opção de se conectar a Nuvem do Workspace do PowerBI ao Synapse.

A conexão do seu WorkSpace do Power BI ao Synapse, deverá ser realizada por meio Synapse studio. Para isso, realize o Login no Synapse studio e clique em “Manage”, então selecione “Linked services”, depois clique em “Connect to Power BI” e complete o formulário com a informação de conexão ao seu WorkSpace do Power BI.



Para obter mais detalhes, consulte: [Vinculando um Power BI Workspace em um Synapse Workspace](#)

Para obter mais informações sobre quais licenças de Power BI são necessárias para um determinado cenário, consulte: [Power BI Premium FAQ - Power BI | Microsoft Docs](#)

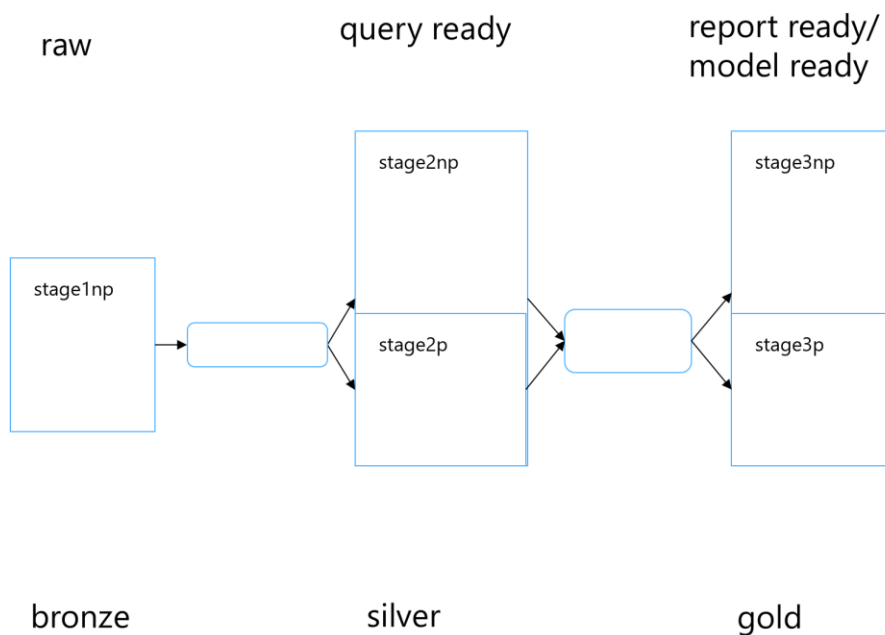
6) Privacidade e Segurança

No nível de armazenamento, a proteção de dados vem da [criptografia automática de dados do Azure](#). Os dados no Data Lake são automaticamente criptografados à medida que são gravados para armazenamento usando criptografia AES de 256 bits e automaticamente descriptografados à medida que são lidos a partir do armazenamento de uma fonte autorizada.

A segurança do acesso é fornecida no nível de armazenamento através do uso de Grupos de Segurança dentro do [Azure Active Directory](#), permitindo que o Administrador Global conceda o acesso mínimo necessário para grupos específicos de usuários, a zonas específicas dentro do Data Lake, com base no acesso necessário para um determinado caso de uso. Através desse uso do RBAC (Role-based Access Control, Controle de Acesso Baseado em Papéis) no nível de armazenamento, o acesso é controlado independentemente das ferramentas usadas para consultar ou analisar os dados. Além disso, o uso do RBAC para configurar o acesso mínimo necessário para um novo grupo de usuários ou para um caso de uso específico é simples e facilmente mantido. Permissões adicionais podem ser definidas no nível SQL para um controle mais fino sobre o acesso. Consulte [garantir acesso a arquivos ADLS usando o modelo de permissão Synapse SQL](#) para obter mais informações.

Na arquitetura OEA, a privacidade dos dados é protegida pela primeira vez reduzindo os dados disponibilizados - ou seja, reduzindo o conjunto de dados ao que é necessário para um determinado caso de uso. Além disso, os dados são pseudônimos para proteger informações pessoalmente identificáveis (PII). A pseudônimo é [definida no GDPR](#) como "o processamento de dados pessoais de tal forma que os dados pessoais não podem mais ser atribuídos a um determinado sujeito de dados sem o uso de informações adicionais, desde que tais informações adicionais sejam mantidas separadamente". Informações adicionais sobre as melhores práticas sobre o uso de pseudônimo em um Data Lake podem ser acessadas em: [Melhores práticas: conformidade com GDPR e CCPA usando Delta Lake](#)

Com a estrutura da OEA, o Data Lake é estruturado para ter 3 etapas conceituais – que refletem uma arquitetura comum do Data Lake em que o primeiro estágio (Stage) é para dados brutos, o segundo é para dados prontos para consulta, e o terceiro é para dados prontos para relatórios. Cada estágio do Data Lake é composta por containers de armazenamento Azure. Os recipientes com o sufixo "p" significam que contêm dados pseudonimizados e recipientes com o sufixo "np" significam que contêm dados não pseudônimos.



O processo de pseudônimo é realizado através do método de pseudônimo no âmbito da OEA.

Este é um documento preliminar e pode ser alterado substancialmente antes da versão comercial final do software aqui descrito. As informações contidas neste documento representam a visão atual da Microsoft Corporation sobre os assuntos discutidos a partir da data de publicação. Como a Microsoft deve responder às mudanças nas condições do mercado, não deve ser interpretado como um compromisso por parte da Microsoft, e a Microsoft não pode garantir a exatidão de qualquer informação apresentada após a data de publicação.

Este white paper é apenas para fins informativos. A Microsoft não faz garantias, expressas ou implícita, neste documento. O cumprimento de todas as leis de direitos autorais aplicáveis é de responsabilidade do usuário. Sem limitar os direitos sob direitos autorais, nenhuma parte deste documento pode ser reproduzida, armazenada ou introduzida em um sistema de recuperação, ou transmitida de qualquer forma ou por qualquer meio (eletrônico, mecânico, fotocópia, gravação ou não), ou para qualquer outro propósito, sem a permissão expressa por escrito da Microsoft Corporation.

A Microsoft pode ter patentes, pedidos de patentes, marcas comerciais, direitos autorais ou outros direitos de propriedade intelectual que cobrem o assunto neste documento. Exceto como expressamente fornecido em qualquer contrato de licença por escrito da Microsoft, o fornecimento deste documento não lhe dá nenhuma licença para essas patentes, marcas comerciais, direitos autorais ou outra propriedade intelectual.

© Microsoft Corporation 2020. Todos os direitos reservados.