

# Transformação de dados com Data Factory e Data Mapping da Azure

Aluno: Renato Gomes Marcacini

Docente: Cristina Dutra de Aguiar Ciferri





# Roteiro

- Introdução
- Arquitetura básica de Data Warehousing
- Data Factory
- Pipeline
- Mapping Data Flow
- Experimento



# Introdução



Azure é a Nuvem da Microsoft



Azure é o Datacenter da Microsoft



Azure é um Serviço da Microsoft



# Introdução

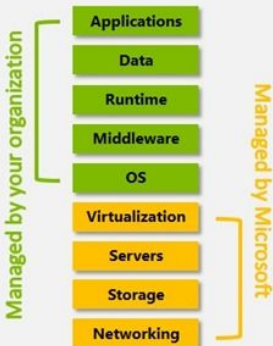
O Microsoft Azure é uma coleção crescente de serviços de nuvem integrados.

## On Premises



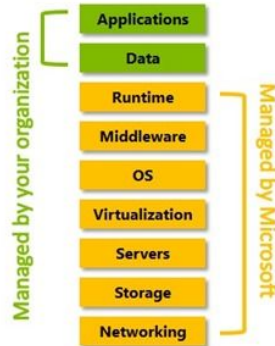
## IaaS

(Infrastructure as a Service)



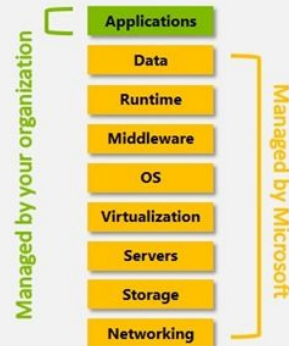
## PaaS

(Platform as a Service)

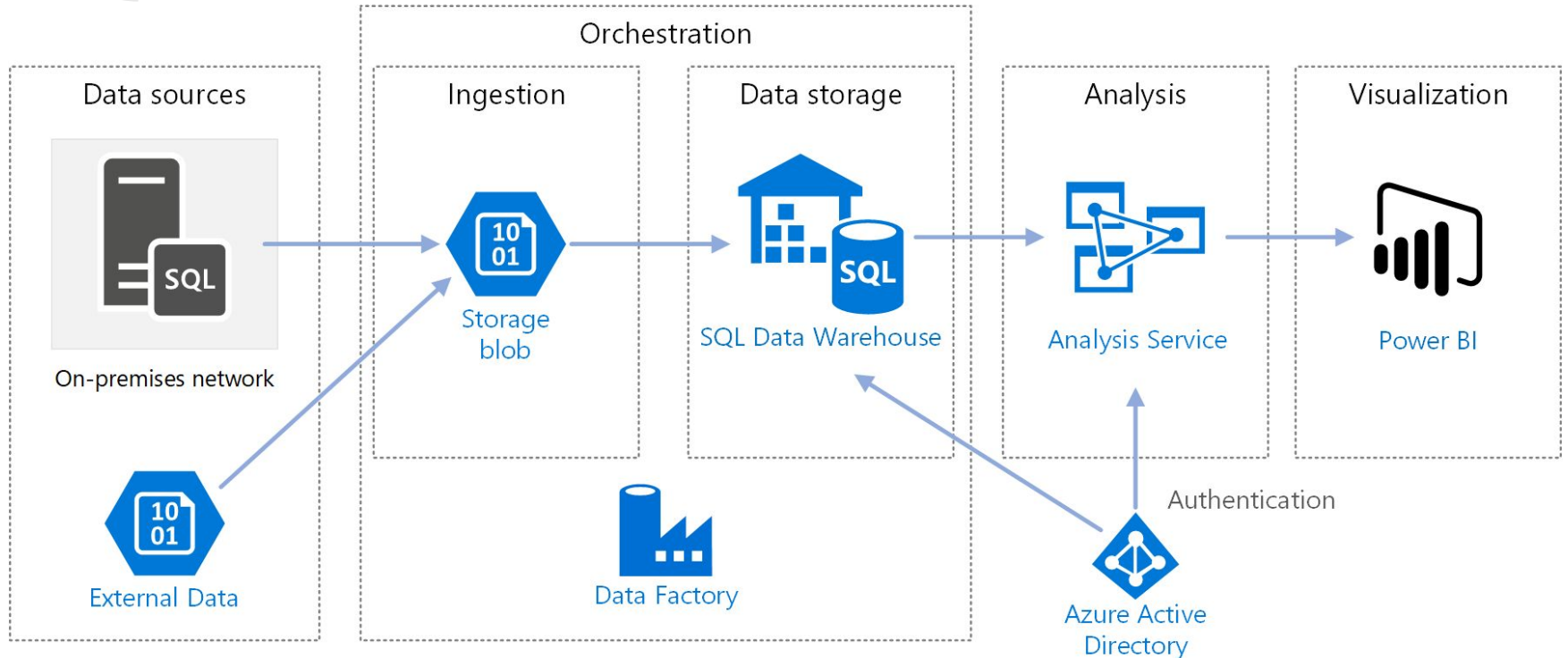


## SaaS

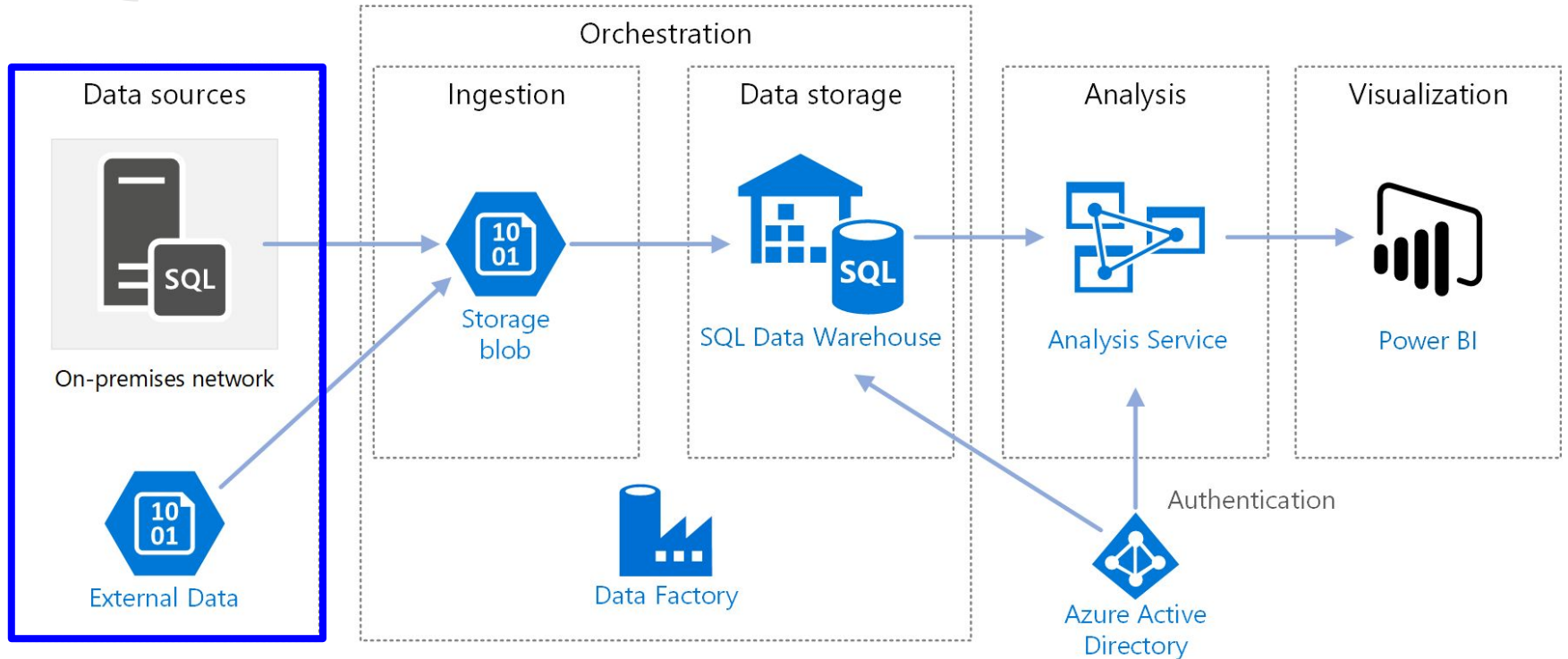
(Software as a Service)



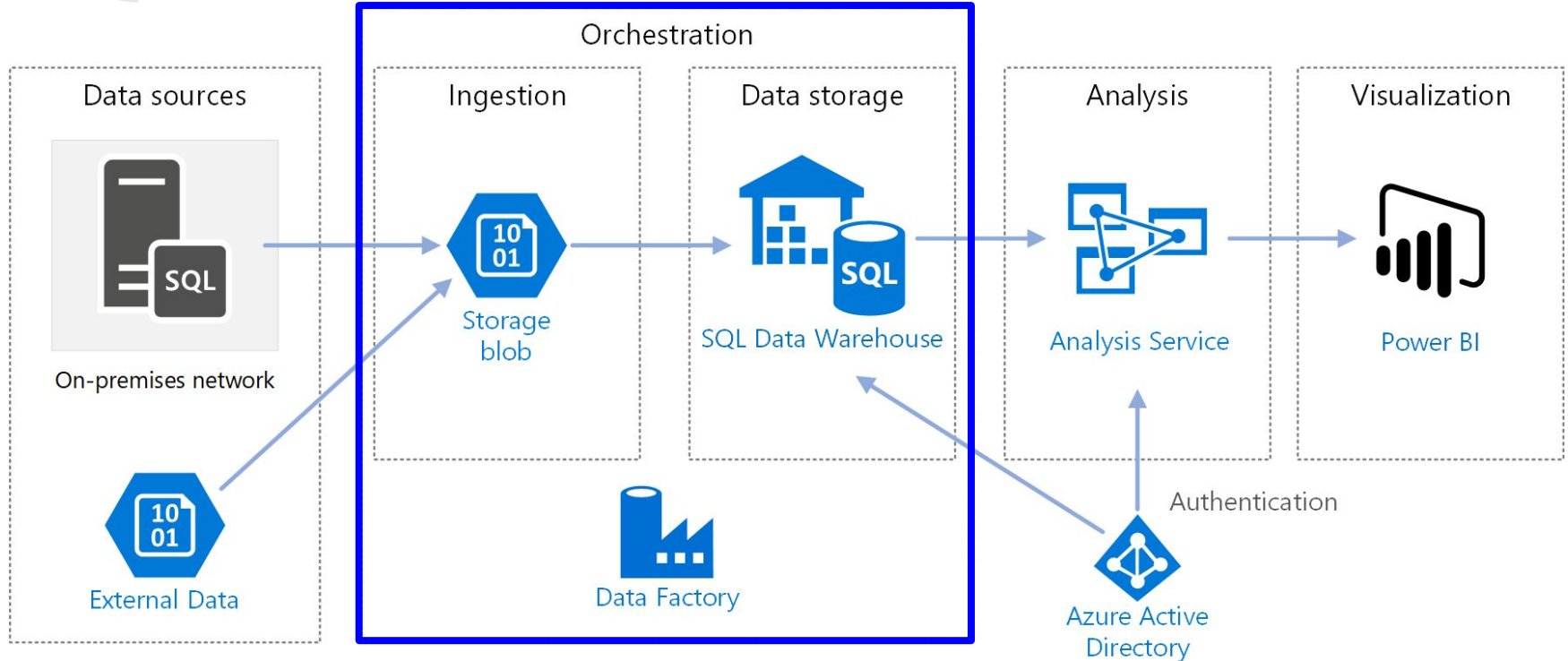
# Arquitetura Data Warehousing



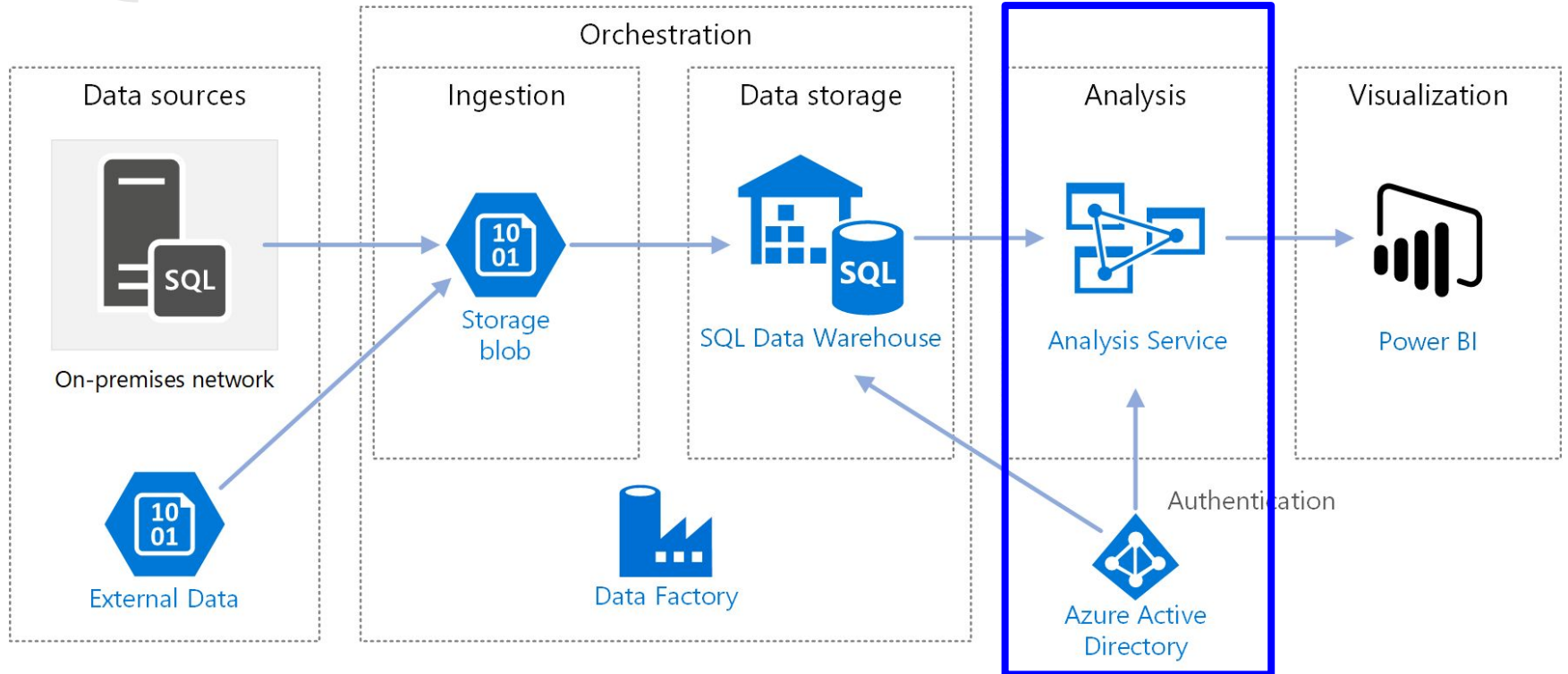
# Arquitetura Data Warehousing



# Arquitetura Data Warehousing

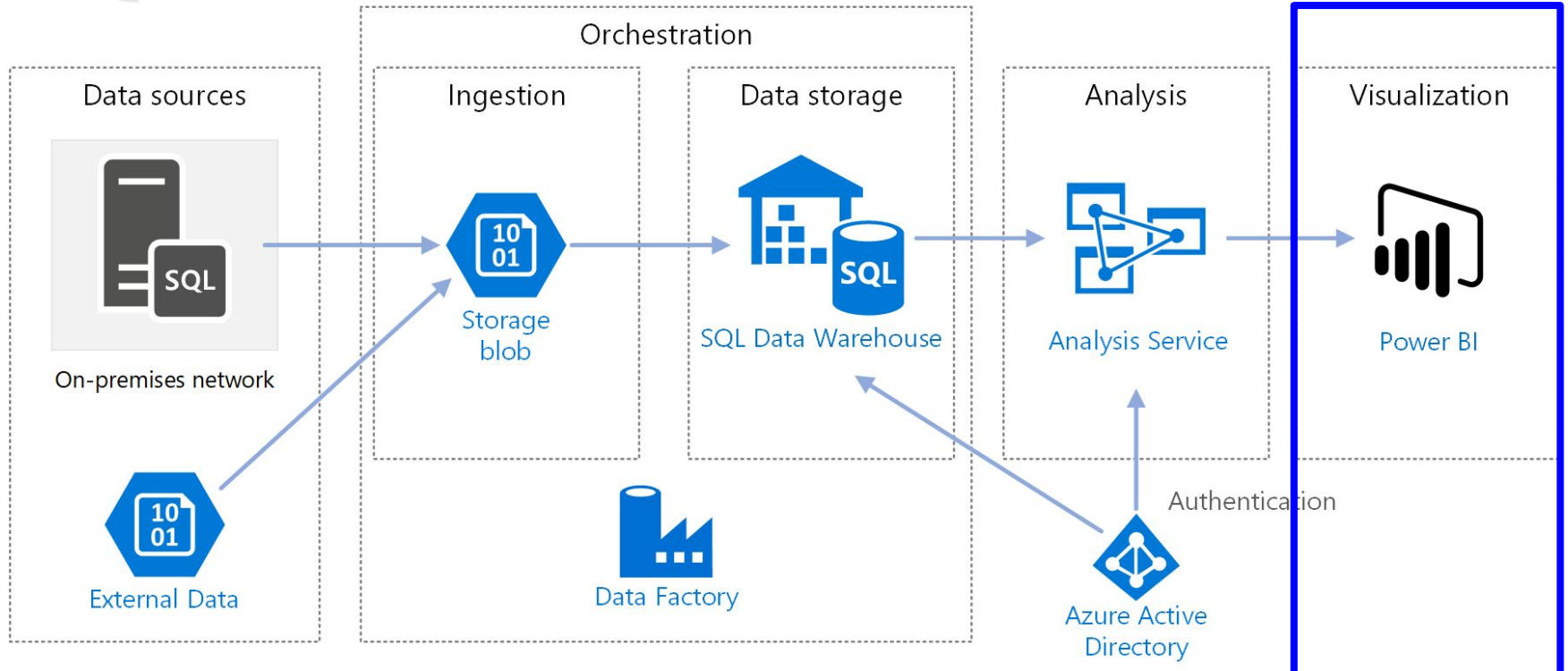


# Arquitetura Data Warehousing



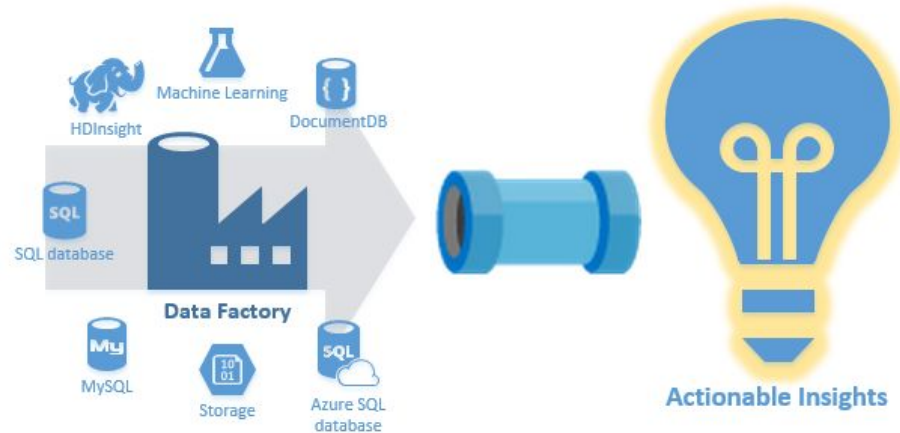


# Arquitetura Data Warehousing



# Data Factory

O Data Factory é um serviço de ETL de integração de dados baseado em nuvem, totalmente gerenciado, que automatiza a movimentação e a transformação de dados.





# Data Factory

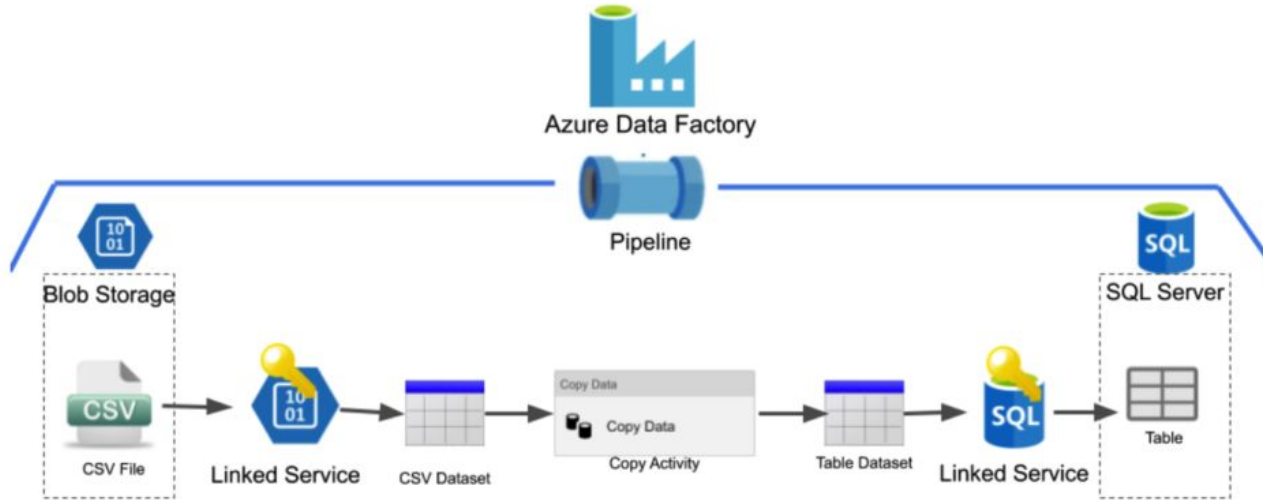
Vantagens:

- Ele oferece uma interface de **drag and drop**. Pode ser usado para fazer **build, debug**, e operacionalizar iterativamente seus pipelines de Big Data.
- Integração com os Azure Services.
- Podem ingerir dados de fontes distintas e podem transformar dados em percepções significativas usando aplicativos de BI, como o Power BI.

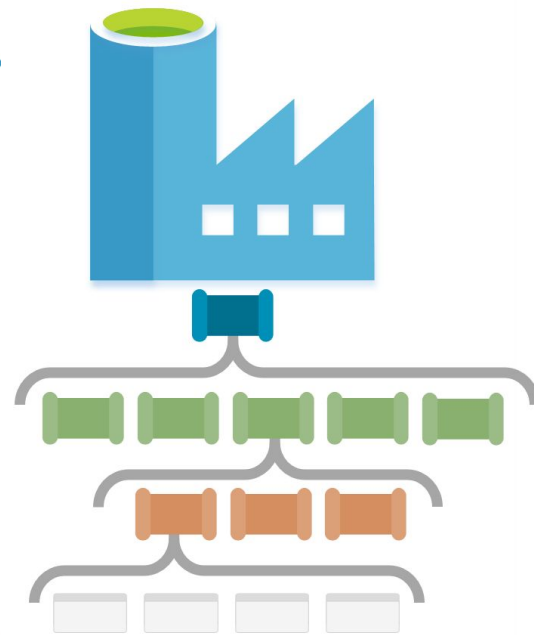
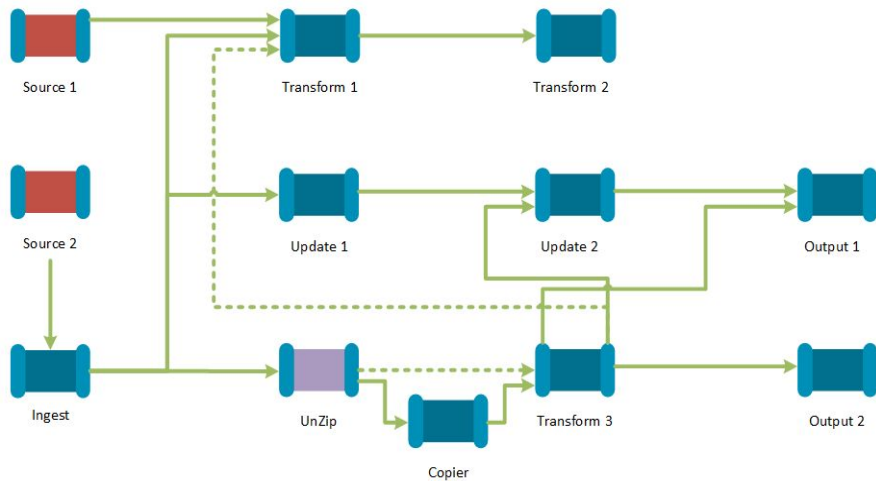


# Pipelines

Pipelines são os agrupamentos lógicos de operações que executam uma determinada atividade. Um Data Factory pode ter N pipelines.



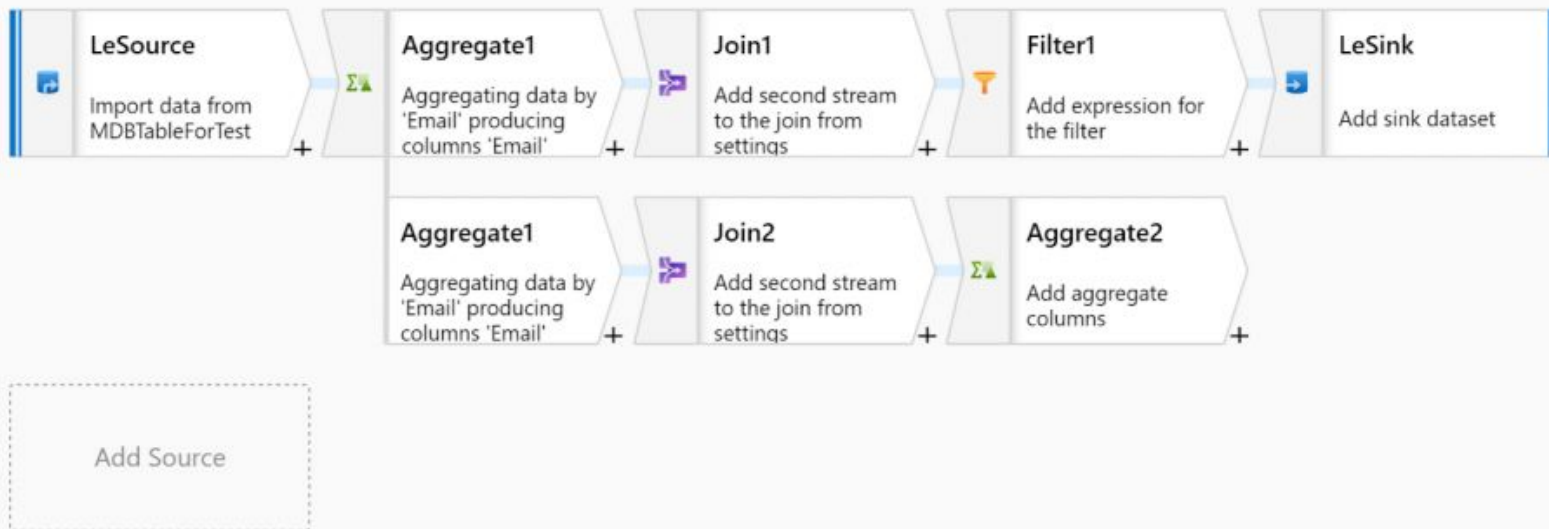
# Pipelines



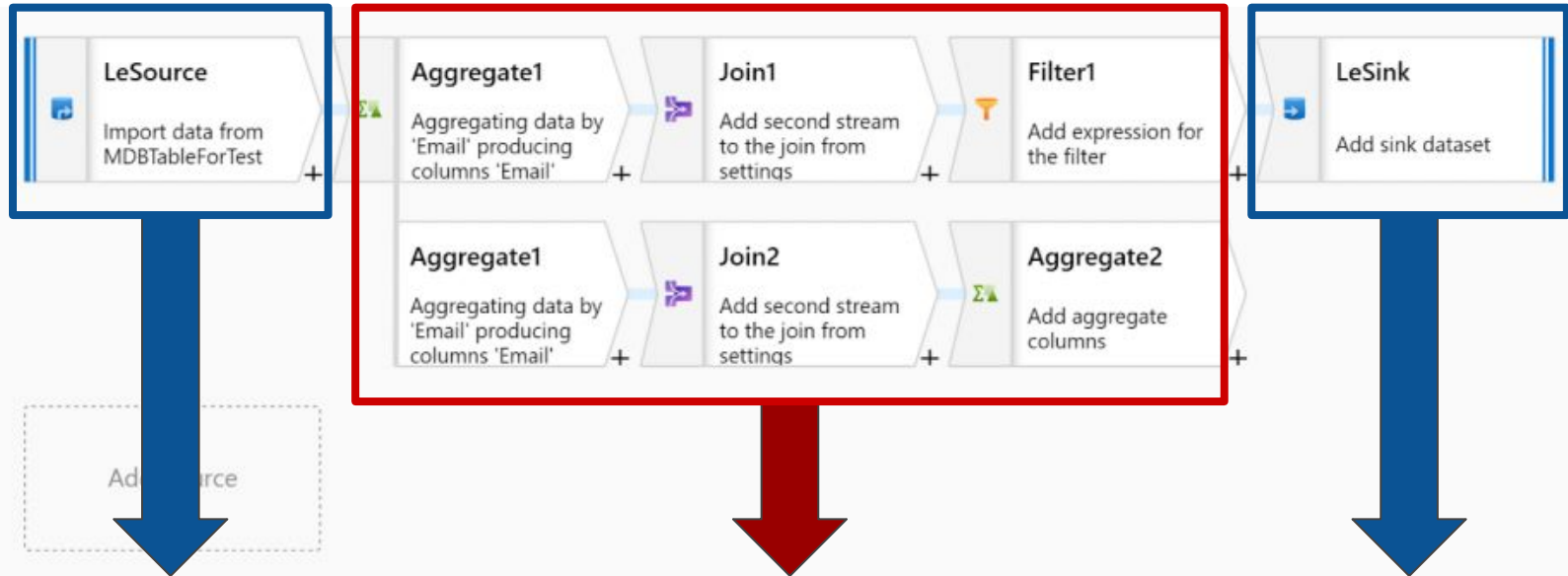


# Mapping Data Flow

O mapeamento de fluxos de dados fornece uma experiência totalmente visual sem a necessidade de codificação.



# Mapping Data Flow



Entradas de dados

Transformação de dados






Carga de dados



# Mapping Data Flow







## Multiple inputs/outputs

---

-  New Branch
-  Join
-  Conditional Split
-  Union
-  Lookup





## Schema modifier

---

-  Derived Column
-  Aggregate
-  Surrogate Key
-  Pivot
-  Unpivot
-  Window

## Row modifier

---

-  Exists
-  Select
-  Filter
-  Sort





# Experimentos

- Listar todos os filmes com avaliação Rotten Tomatoes acima 75%.
- Listar todos os filmes acima do ano de 1960 e abaixo de 2020.
- Listar média de avaliação Rotten Tomatoes dos filmes por ano.
- Listar filmes ordenados por Ano e Avaliação Rotten Tomatoes.
- Listar quantidade de filmes por Gênero Principal e Ano.
- Listar quantidade de Filmes por Gênero Principal.
- Listar Ranking dos melhores filmes que possui gênero 'Adventure'.
- Listar Ranking dos melhores filmes que possui gênero 'Western'.
- Listar melhor/pior filme, média das avaliações Rotten Tomatoes e quantidades de filmes por gênero pelo ano.

**Muito obrigado!**  
**\(^0^)/**

Aluno: Renato Gomes Marcacini  
Docente: Cristina Dutra de Aguiar Ciferri





# Referências

**Transformar dados usando fluxos de dados de mapeamento**, disponível em:

<https://docs.microsoft.com/pt-br/azure/data-factory/tutorial-data-flow>

**Visão geral da transformação fluxo de dados de mapeamento**, disponível em:

<https://docs.microsoft.com/pt-br/azure/data-factory/data-flow-transformation-overview>

**Introdução ao Data Factory, um serviço de integração de dados**, disponível em:

<https://opdhsblobprod01.blob.core.windows.net/contents/4a6d75bb3af747de838e6ccc97c5d978/68ac7b63a23f4ccb87ba36eda3f3e82a?sv=2018-03-28&sr=b&si=ReadPolicy&sig=kyqCAJNJxU0qxYD4e2QCymgiyb3n%2F1ss3bE64v4GKQ0%3D&st=2020-11-27T16%3A57%3A26Z&se=2020-11-28T17%3A07%3A26Z>

**Portal Azure**, disponível em: <https://portal.azure.com/>