



Glue

- ◆ Serviço Gerenciado de ETL
- ◆ Baseado em SPARK

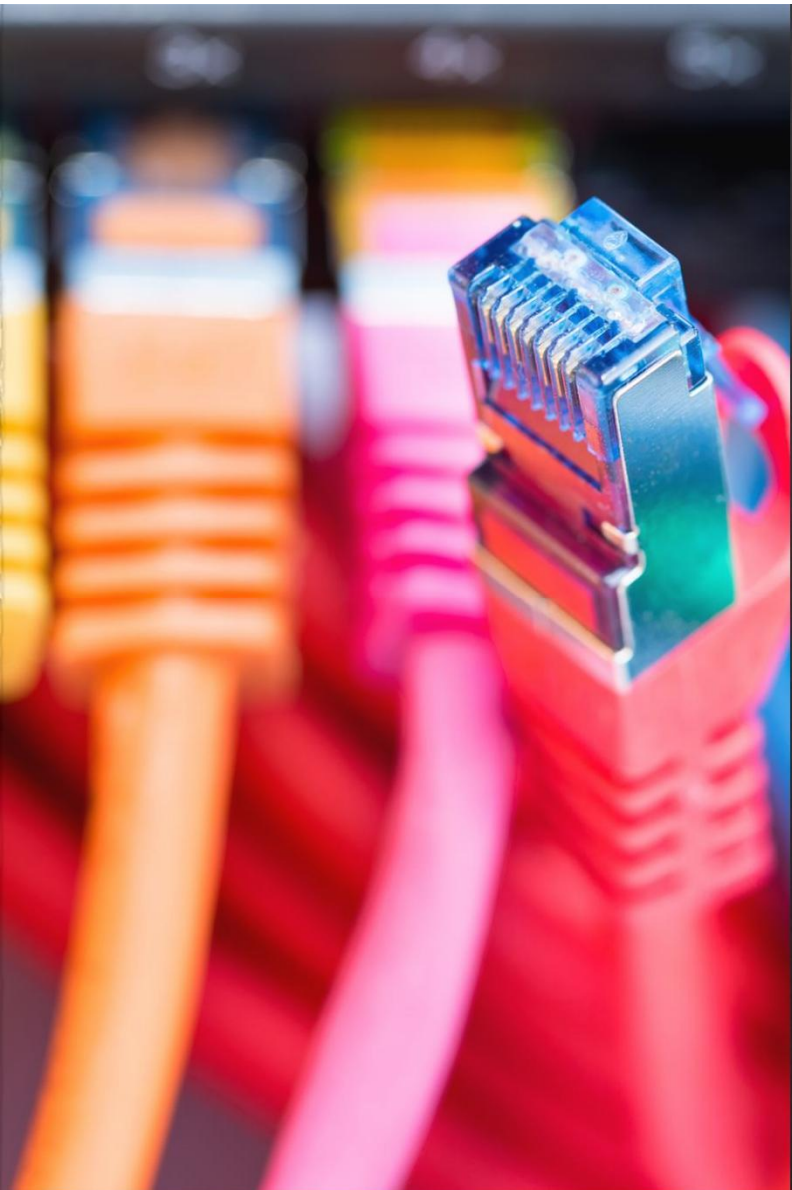


Construir um DataLake

- ◈ Schema de Dados na Origem e Destino
- ◈ Conectar a Origem e Destino
- ◈ Transformar Dados

Crawler

- ◊ Conecta em fontes de dados para definição de schema
- ◊ Cria metadados no Glue “Data Catalog”



Connector

- ◆ Informações necessárias para conexões com fontes de dados

Glue Job

- ◇ Processo de ETL
- ◇ Geralmente:
 - ◇ Origem
 - ◇ Transformações
 - ◇ Destino
- ◇ Executados:
 - ◇ Sob demanda
 - ◇ Agenda
 - ◇ Gatilhos



Data Catalog

- ◆ Data Catalog:
 - ◆ Repositório de Metadados
- ◆ Database:
 - ◆ Agrupa logicamente tabelas
 - ◆ Não cria nem carrega dados
- ◆ Tables
 - ◆ Representações de metadados de dados tabulares
 - ◆ Dados mantem-se em sua forma e fonte original