Performing ETL (Extract, Transform, Load) Operations with Azure Databricks

Michael Bender

AUTHOR EVANGELIST, PLURALSIGHT

@MichaelBender

Overview

Basics of ETL

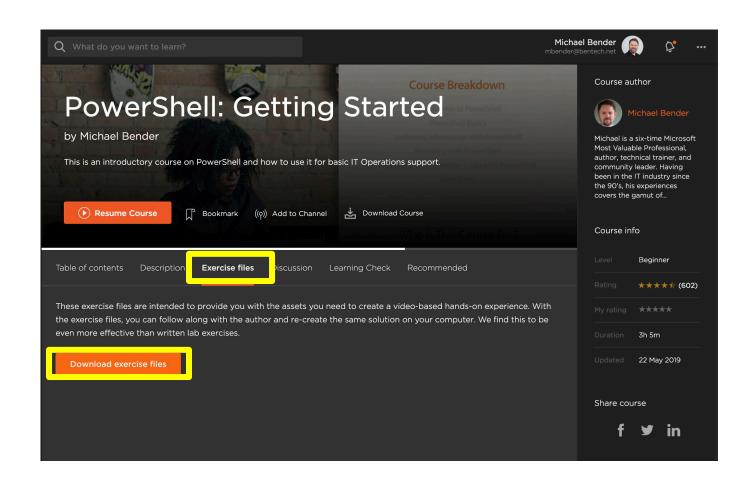**Performing ETL with Azure Databricks**

Exercise Files

Slides
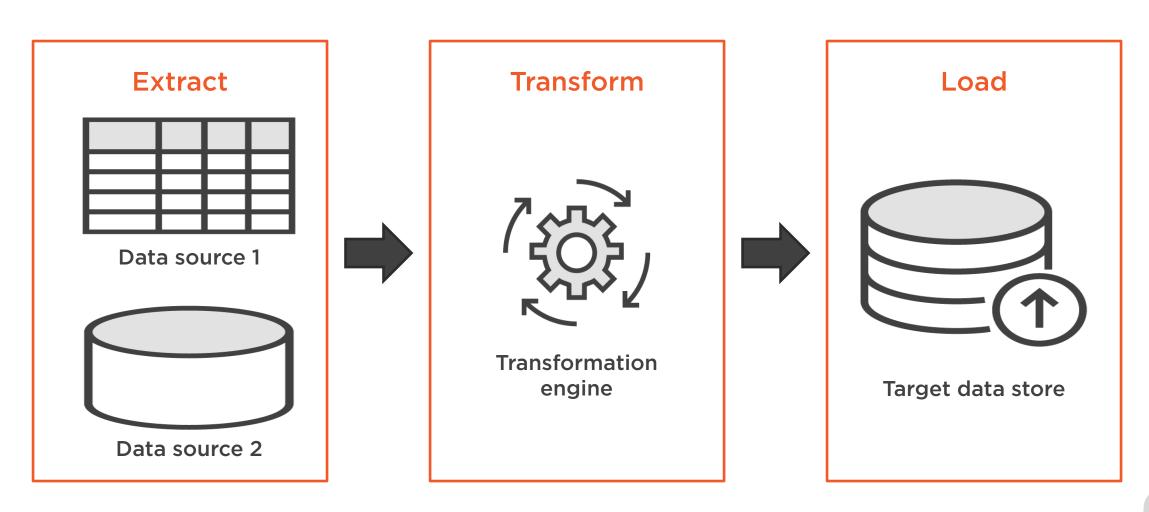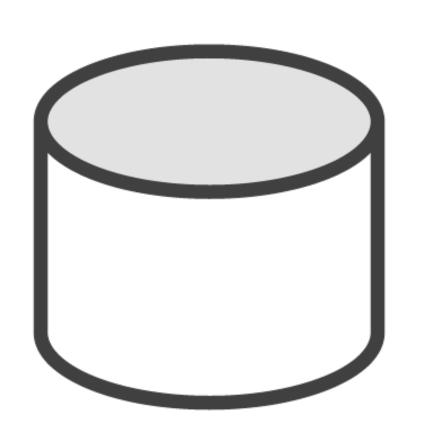
Code

Links to Resources

Basics of Extract, Transform, and Load (ETL) Process

Extract

Data source 1

Data source 2

Transform

Transformation engine

Load

Target data store

Extracting Data from Storage
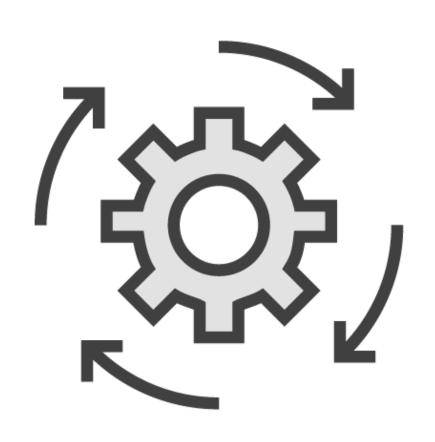
Raw storage including

- Azure Blob Storage

- Azure Data Lake Storage

- Hadoop Storage

Driven by Notebook calls

Secure access to data stores

Transforming Data with Azure Databricks

Refining data into predictions and insights

Azure Databricks clusters do the work

Data is refined to what you need

Use your language of choice in Notebooks
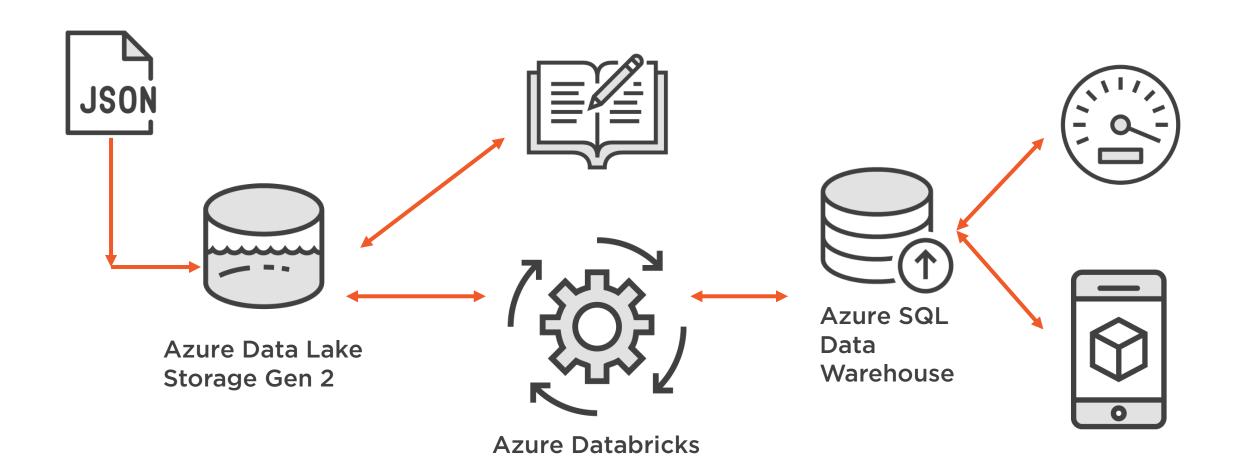
Loading Data in Azure SQL Data Warehouse

Choose the database service that meets your application needs

Storage for your intelligent applications

Data refresh is easy

Scenario – Working with Audience Information

JSON

Azure Data Lake
Storage Gen 2

Azure Databricks

Azure SQL
Data
Warehouse

Pre-requisites for Demo

Azure SQL data warehouse

Azure Blob storage and Data Lake storage account

Service principal

http://bit.ly/ETLdemo

Demo

Performing ETL operation in Azure Databricks

Summary

Everything you need to perform ETL with your data

Notebooks are where the magic happens

Simply modify your code and re-run to see changes

For Further Learning

Azure Databricks documentation at **docs.microsoft.com**
https://docs.microsoft.com/en-us/azure/azure-databricks/

Azure Databricks documentation at docs.azuredatabricks.net
https://docs.azuredatabricks.net/user-guide/index.html

Remember the module exercise files

Questions? Join on the conversation at pluralsight.com

Next Up:
Batch Scoring of Apache Spark ML Models with Azure Databricks