Topic:

Azure Data Factory Design Patterns

Details:

The Azure Data Factory (ADF) services currently supports three offerings that can be used to design and deploy pipelines.

The second version of Azure Data Factory was released in the summer of 2018. Azure Data Factory v2 excels at connecting, orchestrating, delegating and managing, using a combination of legacy and modern environments. Many ELT pipelines have been developed to move data to the cloud. The self hosted integration runtime is a important component when working with a hybrid environment.

Mapping Data Flows for ADF was release in the Fall of 2019. This new offering brought ETL transformations to ADF for cloud born services. The transformations within this product resemble controls that you might see in SSIS. However, a Databricks spark cluster is the horsepower behind this offering.

Wrangling Data Flows in ADF allow you to do code-free data preparation at cloud scale. These data flows translate M code generated by the Power Query Online Mashup Editor into spark code for cloud scale execution. Currently, this service is in public preview for all to use.

In summary, there are many different ways to create Azure Data Factory pipelines that extract, translate and load data. Knowing the pros and cons of each design pattern is very important when designing systems for customers. Please join us online for this presentation on April 22nd, at 7 pm EST.