Configuration
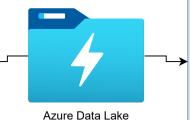
- Compatible with Hadoop HDFS
- Distributed file system (Apache ADLS Gen2/HD Insight)
- Granular security (RBAC and access control list)

Concepts
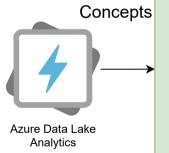
Azure Data Lake

- Large raw data repository (data lake) & structured information (data warehouse)
  - Load data before processing (ELT)
  - Stage ingested data
- Hierarchical directory & file system semantics + security + scale
  - Similar to Azure Blob storage organized near-infinite file system
  - Organize files into directories and sub-directories
  - Portable Operating System Interface (POSIX) for RBAC

Concepts

Azure Data Lake Analytics

- On demand big data analytics
- Framework & tools
  - Transform data and extract insights
- U-SQL (hybrid SQL & C#)
  - Declarative & procedural
  - Describe work performed
  - @ DB_name = EXTRACT Col FROM Table
  - Parse & process data in partitions (MPP)
  - Output to XML/HTML/csv/Data Lake catalog with SQL-like interface