Knowing How Data Flows



Gary Grudzinskas
CLOUD ENGINEER AND AUTHOR
@garygrudzinskas



Objectives



Know hokkndatanohovekafaofhovesufreentoource to destination

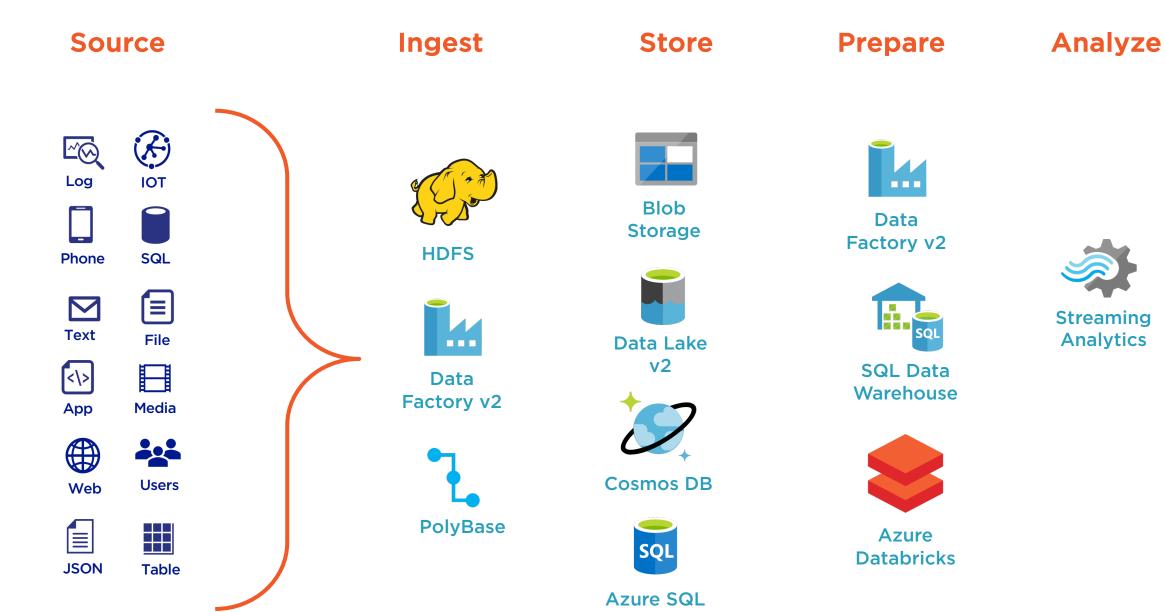
Know hokkned of From the Market of the state of the state

Know hokknewulnew at a ufred at a fear both enable the flow to fed about of data

Examine Encamp Process Pool of Elasse works

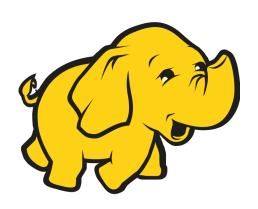
Define the differtence if the model of the the control of the control of the the control of the the control of the con







Hadoop Distributed File System



Distributed across servers so it is tolerant to hardware failure

Designed for large data sets

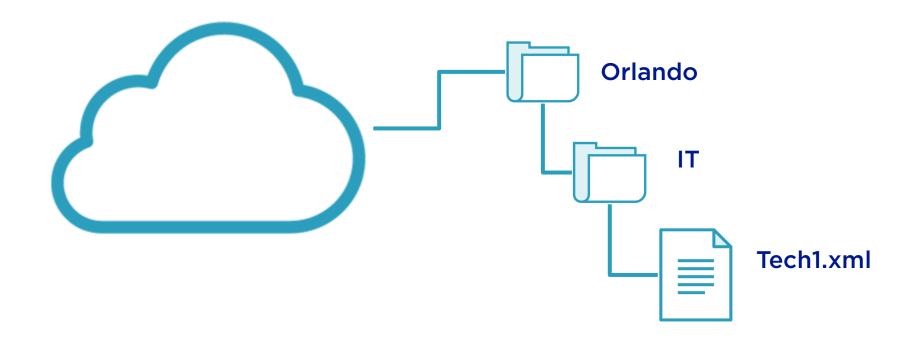
"Moving Computation is Cheaper than Moving Data"

Portable from one platform to another

Uses a universal hierarchical namespace



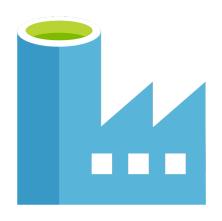
Hierarchical Namespace



wasbs://<containername>@<accountname>.blob.core.windows.net/orlando/it/tech1.xm



Azure Data Factory v2



Manages the flow of data between various data stores

Automate data transformation

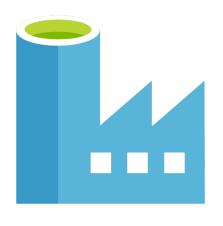
Uses compute services such as Azure HDInsight, Hadoop, Spark, and Azure Machine Learning

Publish output data to Azure SQL Data Warehouse

Use to organize raw data into meaningful data stores and data lakes



Azure Data Factory v2

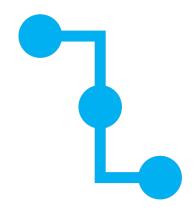


Use When...

- Publish output to data stores
- Organize raw data
- Set up data pipelines
- Need to move data around
- Want to move data into Azure Data Warehouse



PolyBase



Data virtualization

Built into ADF, SQL server, and SQL DW

Access and combine both non-relational and relational data

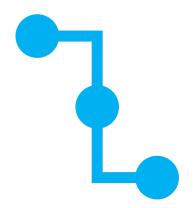
Import, export, and query outside data from Hadoop or Azure blob storage

Integrate with BI tools

Load data into Azure SQL Data Warehouse



PolyBase



Use When...

- Load a large amount of data into Azure SQL Data Warehouse
- Export data out of Azure SQL Data Warehouse
- Archive data to blob storage
- Process Transact-SQL queries from outside data sources



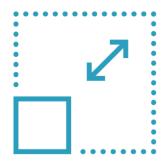
ETL

Extract



Raw data is defined by type and source

Transform



Split, combine, add, remove, aggregate and merge data in order to "fit"

Load



Define the destination, then write the data into the destination



ELT

Extract



Raw data is ingested from the source or virtualized at source

Load



Data is delivered to the destination whole

Transform



Tools are used to sort and organize the data for various entities

