Data Architectures in Azure for Analytics & Big Data

October 20, 2018

Melissa Coates

Solution Architect, BlueGranite
Microsoft Data Platform MVP

Blog: www.sqlchick.com
Twitter: @sqlchick
Data Architecture

A set of *rules, policies, standards, & models* that govern and define the type of *data collected* & how it is *used, stored, managed, & integrated* within an organization & its database systems.

Source: Technopedia
Data Architecture Components

Azure

A public cloud computing platform and infrastructure for building, deploying, and managing MSFT-specific and third party software and services through a global network of Microsoft-managed datacenters.

Source: [http://azureplatform.azurewebsites.net/](http://azureplatform.azurewebsites.net/)
Azure Technologies
Relational database of your choice in a virtual machine

- Azure SQL Database
- Azure Database for MySQL
- Azure Database for PostgreSQL
- Azure Database for MariaDB

IaaS
Infrastructure as a Service

PaaS
Platform as a Service

SMP
Symmetric Multi-Processing

MPP
Massively Parallel Processing

Managed Database

Managed Instance

Azure SQL Data Warehouse

MPP
Massively Parallel Processing

SMP
Symmetric Multi-Processing

MPP
Massively Parallel Processing
Data Storage: Big Data

Object Storage (Flat)
- Azure Blob Storage

Hierarchical Storage
- Azure Data Lake Store (Gen1)

Multi-Modal
- Azure Data Lake Storage (Gen2) (PREVIEW)
Data Storage: NoSQL

PaaS
Platform as a Service

HDInsight
HBase
Cluster

Azure
Table Storage

Multi-Model

Azure
CosmosDB

Key
Value

Column
Family

JSON
Documents

Graph
Data Storage: Analytical & OLAP

IaaS
Infrastructure as a Service

SQL Server Analysis Services

PaaS
Platform as a Service

Azure Analysis Services

Power BI
Compute: Big Data

**IaaS**
Infrastructure as a Service
- HDInsight in a VM
- Hortonworks
- cloudera
- MAPR

**PaaS**
Platform as a Service
- HDInsight Spark Cluster
- HDInsight Interactive Query Cluster (Hive LLAP)
- Azure Databricks

**SaaS**
Software as a Service
- Azure Data Lake Analytics

- HDInsight Spark Cluster
- HDInsight Interactive Query Cluster (Hive LLAP)
- Azure Databricks
- Azure Data Lake Analytics
Compute: Streaming & Event Processing

PaaS
Platform as a Service

- HDInsight Kafka Cluster
- HDInsight Storm Cluster
- HDInsight Spark Streaming Cluster
- Azure IoT Hub
- Azure Event Hub
- Azure Stream Analytics
Compute: Data Integration

IaaS
Infrastructure as a Service

Tool of your choice in a virtual machine

PaaS
Platform as a Service

Azure Data Factory

Azure Databricks

Serverless

HDInsight:
Spark, Hive, Pig, Scoop, Oozie

Azure Functions

Azure Automation

Microsoft SQL Server

Oracle

Informatica

talend
Reference Architectures

Following are examples only!

There are many variations & opportunities to exchange one Azure service for another.
Small/Medium Data Warehousing

Source Data → Multi-Structured Data

Azure Blob Storage

Multi-Structured Data → DW: Structured Data

Azure SQL Database

DW: Structured Data → Semantic Layer

Azure Analysis Services

Semantic Layer → Reporting & Analysis Tools

Power BI

Excel

Reporting & Analysis Tools
Enterprise Data Warehousing and BI

Multi-Structured Data
- Azure Data Lake Storage

Data Mart(s)
- Azure SQL Database

DW: Structured Data
- Azure SQL Data Warehouse
  - Control Node
  - Compute Node
  - Compute Node
  - Compute Node
  - Distributed Data Storage

Semantic Layer
- Azure Analysis Services

Power BI
- Excel
Unified Data Science & Data Engineering

Data Lake: Multi-Structured Data

- Raw Data
- Curated Data
- Data Lake Storage

Data Science Sandbox

- Exploratory Analytics
- Operationalized Analytics

Structured Data

- Azure SQL Database
- Azure Databricks
- Spark
- SQL
- Python
- SCALA

Scheduled Notebook Job

Structured Data
Big Data Interactive Querying (SQL on Hadoop)

- Data Lake
  - Azure Data Lake Storage
- Hive Metastore
- Hive Data Warehouse
  - HDInsight Interactive Query Cluster (Hive LLAP)
- Azure SQL Database

HiveQL

External Tools:
Big Data Batch Processing

Data Lake: Multi-Structured Data

Azure Data Lake Storage

Azure Data Lake Analytics

U-SQL Job Processing

Job 1

Job 2

ADLA Catalog

Database

Tables
Views
Schemas

Procedures
Functions
Assemblies

External Data Sources

Azure SQL DB

SQL Server in Azure VM

Cognitive Services

Python

U-SQL Extensions
IoT + Batch Data (Lambda Architecture)

Speed Layer
- Azure Event Hub
- Azure Stream Analytics

Batch Layer
- Azure Data Lake Storage
- Azure SQL Data Warehouse

Serving Layer
- Power BI
- Azure Analysis Services
- Streaming Dashboard
- Excel
- Power BI
Web Application

Azure SQL Database

- Diagnostics
- Backups

Storage Account

- 1001

Web App

- Cache

App Service Plan

Storage Account

- 1001

Web Page
Wrapper: Wrap-Up
Azure Solution Architectures:
More Info

Azure Data Architecture Guide:
Thanks!

Download latest version of slides:
SQLChick.com > Presentations & Downloads page

Creative Commons License 3.0

Attribute to me as original author if you share this material

No usage of this material for commercial purposes

No derivatives or changes to this material