

Overview of Data Services and Streaming Data Solution with Azure

Tara Mason

Senior Consultant

tmason@impactmakers.com

Platform as a Service Offerings

SQL Server On Premises vs. Azure

SQL Server

- Azure SQL Database
- Azure SQL Database Managed – Preview March

SQL Server Application Platform System (MPP)

- Azure SQL Data Warehouse

SQL Server Analysis Services

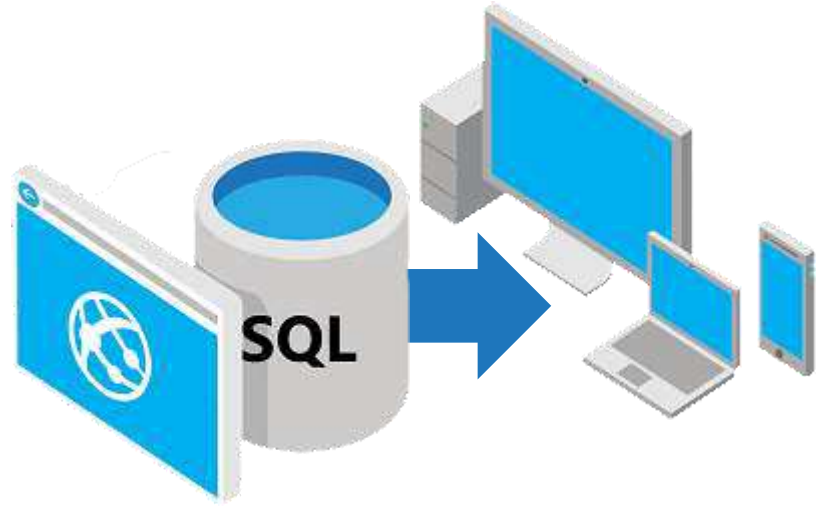
- Azure Analysis Services

SQL Server Integration Services

- Azure Data Factory*

Azure SQL Database

- Azure SQL Database is a cloud-based relational database service, built on the Microsoft SQL Server engine designed to deliver predictable performance and scalability, with virtually no downtime and near-zero administration.
- Up to 1TB
- There are a number of features from SQL Server not available in Azure SQL
 - Some T-SQL functions
 - Cross-database queries
 - Database mirroring
 - SQL Server Agent

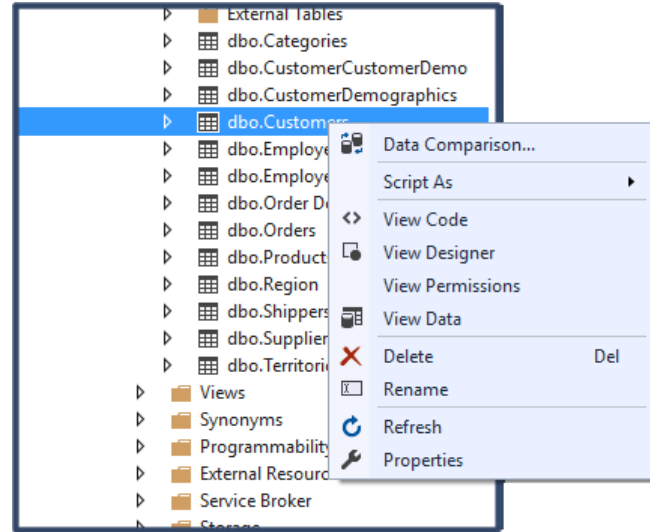


<https://docs.microsoft.com/en-us/azure/sql-database/sql-database-features>

Azure SQL Database Development

Use Visual Studio and SQL Server Management Studio just as you would with SQL Server

- Rich and up-to-date transact-SQL (T-SQL) support
- Stored procedures, user-defined functions, triggers, and views
- Newest features from SQL Server 2016



Azure Managed SQL Database

Microsoft has [announced](#) the public preview of Azure SQL Database Managed Instance.

In summary:

Managed Instance is an expansion of the existing SQL Database service, providing a third deployment option alongside single databases and elastic pools. It is designed to enable database lift-and-shift to a fully-managed service, without re-designing the application. SQL Database Managed Instance provides the broadest SQL Server engine compatibility and native virtual network (VNET) support so you can migrate your SQL Server databases to SQL Database without changing your apps. It combines the rich SQL Server surface area with the operational and financial benefits of an intelligent, fully-managed service.

Azure SQL Data Warehouse – MPP

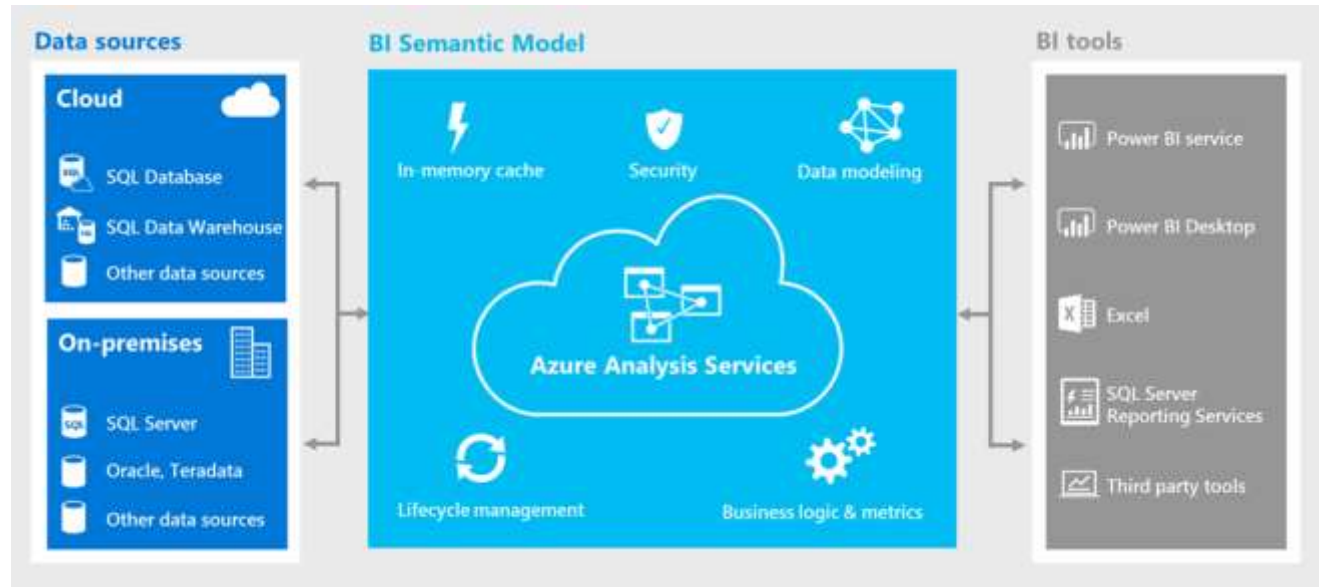
SQL Data Warehouse is a cloud-based Enterprise Data Warehouse (EDW) that leverages Massively Parallel Processing (MPP) to quickly run complex queries across petabytes of data. Use SQL Data Warehouse as a key component of a big data solution. Import big data into SQL Data Warehouse with simple PolyBase T-SQL queries, and then use the power of MPP to run high-performance analytics.

Feature	SQL Database	SQL Data Warehouse
Size	Max of 1TB for a database	No limit
Concurrent Queries	Up to 6,400	Up to 32
Active Connections	Up to 32,000	Up to 1,024
Cross-database Queries	Supported	Not supported
Pause/Resume	Not supported	On-demand pause/resume of resources
Scalability	11 tiers	12 options that can be managed by sliding a bar up and down
Replication	Supported through secondary databases in different regions(up to 4)	Not supported
In-Memory OLTP Tables	Supported	Not supported
<u>Polybase</u>	Not supported	Supported

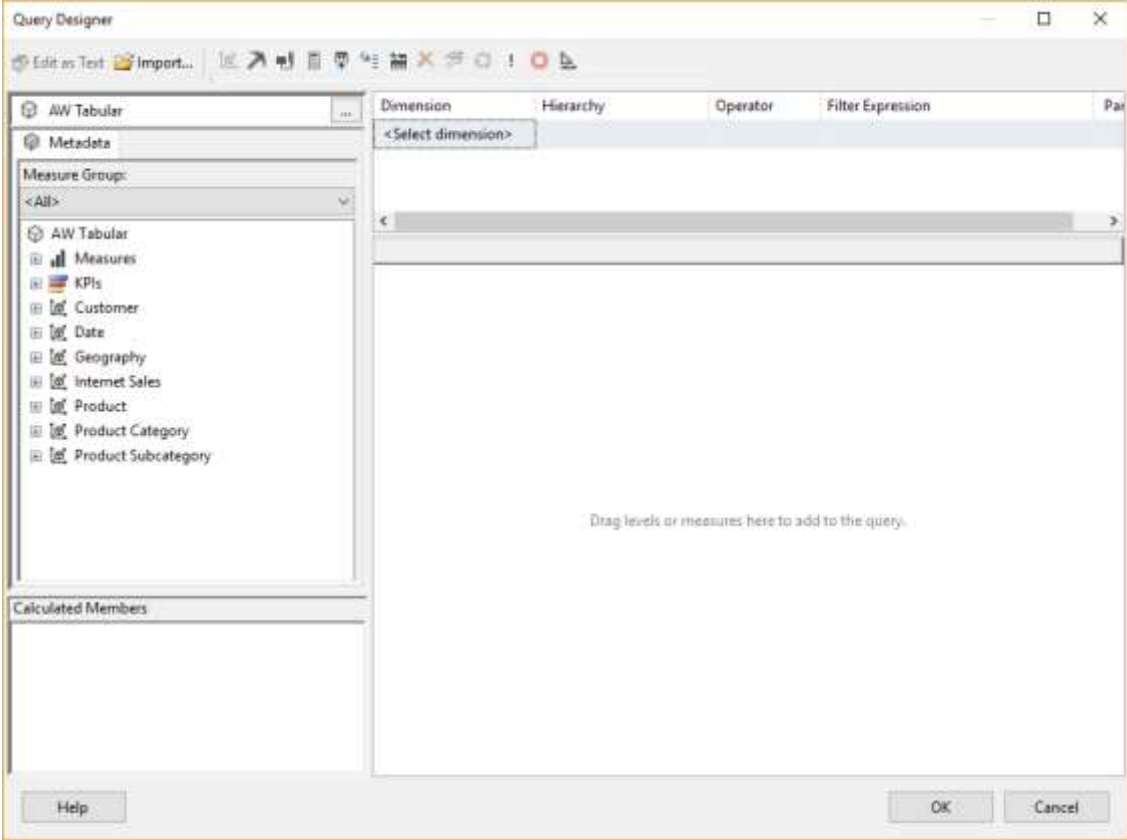
Azure Analysis Services – Bringing Data to the Users

Azure Analysis Services provides enterprise-grade data modeling in the cloud. It is a fully managed platform as a service (PaaS), integrated with Azure data platform services.

With Analysis Services, you can mashup and combine data from multiple sources, define metrics, and secure your data in a single, trusted semantic data model.



Azure Analysis Services



Azure Data Factory

It is a *cloud-based data integration service that allows you to create data-driven workflows in the cloud that orchestrate and automate data movement and data transformation.*

Using Azure Data Factory, you can do the following tasks:

- Create and schedule data-driven workflows (called pipelines) that can ingest data from disparate data stores.
- Process or transform the data by using compute services such as Azure HDInsight Hadoop, Spark, Azure Data Lake Analytics, and Azure Machine Learning.
- Publish output data to data stores such as Azure SQL Data Warehouse for business intelligence (BI) applications to consume.

It's more of an Extract-and-Load (EL) and Transform-and-Load (TL) platform rather than a traditional Extract-Transform-and-Load (ETL) platform.

Announcing Azure Data Factory



Other Data Storage Options

Azure Blob Store

Azure Blob storage is Microsoft's object storage solution for the cloud. Blob storage is optimized for storing massive amounts of unstructured data, such as text or binary data.

Azure Data Lake

Azure Data Lake Store is an enterprise-wide hyper-scale repository for big data analytic workloads. Azure Data Lake enables you to capture data of any size, type, and ingestion speed in one single place for operational and exploratory analytics.

Azure Data Bricks

Azure Databricks is an Apache Spark-based analytics platform optimized for the Microsoft Azure cloud services platform.

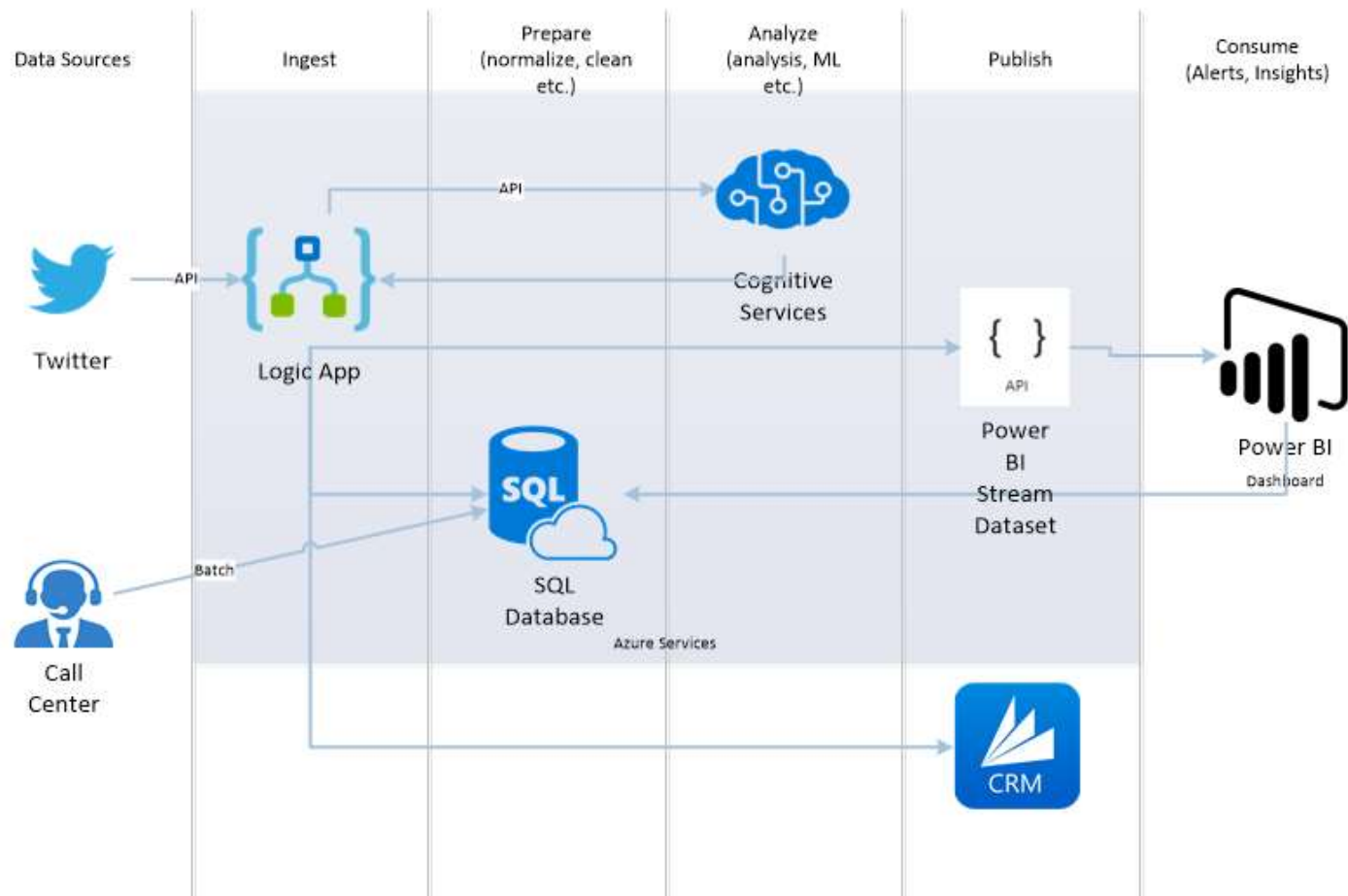
Demo: Social Streaming Data Integrated with Azure SQL Database and Power BI

Streaming Data Solution...

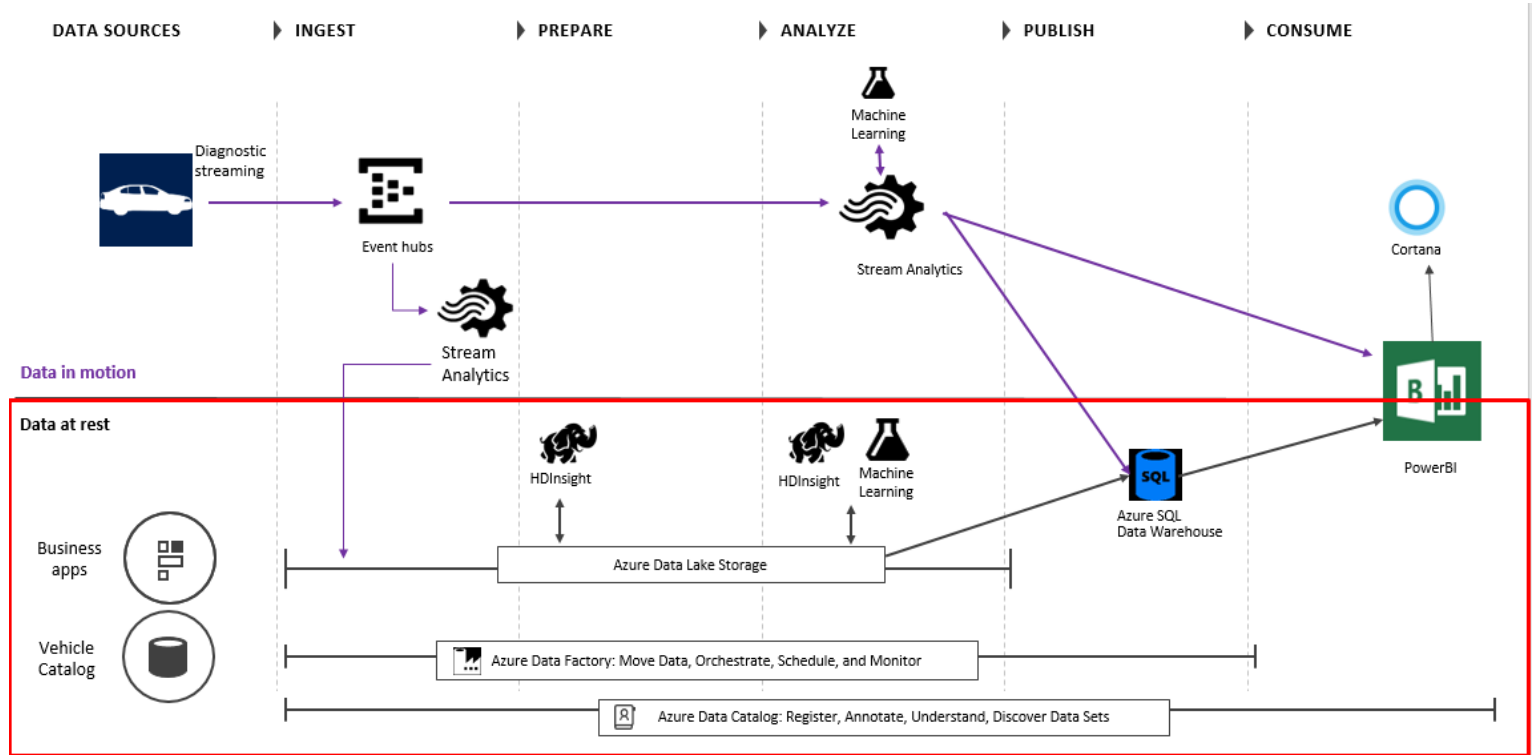
1. Query social media for data about a topic/company
2. Use an analysis engine to determine the sentiment of social media posts in real-time
3. Set a threshold on the sentiment and send an email (or post to CRM system) if a post goes below the threshold
4. Save the social media data so it can be analyzed.
5. Create a dashboard with historical social media data for decision making and analysis

The following services were used:

1. Azure Logic Apps
 - Serverless workflow service
2. Azure Cognitive Services – Text Analytics
 - Assign Sentiment Score (0 to 1)
3. Azure SQL Database
 - Store Twitter data history
4. Power BI Service
 - Consume data and create dashboard



Lambda architecture is a data-processing architecture designed to handle massive quantities of data (i.e. “Big Data”) by using both batch-processing and stream-processing methods. This idea is to balance latency, throughput, scaling, and fault-tolerance by using batch processing to provide comprehensive and accurate views of batch data, while simultaneously using real-time stream processing to provide views of online data.



Data Sources

Ingest

Prepare
(normalize, clean, etc.)

Analyze
(stat analysis, ML, etc.)

Publish
(for programmatic
consumption, BI/visualization)

Consume
(Alerts, Operational Stats,
Insights)

Near Realtime Data Analytics Pipeline

Simulated browsing activity

Queued browsing activity



1



2



3

Aggregated browsing activity



Azure Storage



Browsing activity



Merged customer profile
with engineered features



Enriched customer profile
with predictions

8

Visualize



PowerBI
Dashboard

Profile Enrichment orchestrated by
Azure Data Factory



4

Referential data
(demographics,
products, purchases)



4

Browsing data

5

Engineered
features

6

Data for
scoring

7

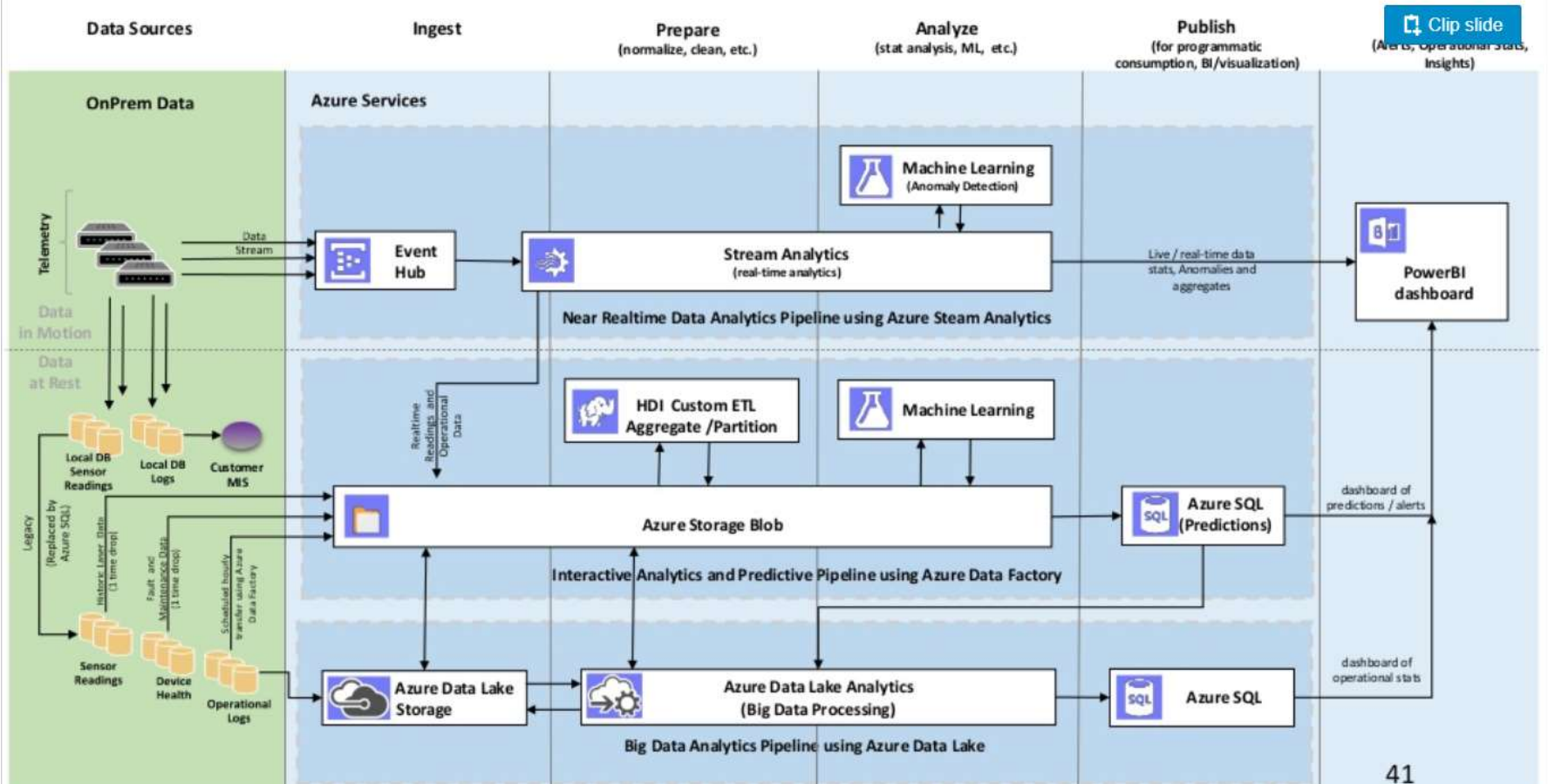
Predictions

Batch ETL and Predictive Pipeline

Enriched customer profile
(external table)

Azure Services

Base Architecture : Big Data Advanced Analytics Pipeline



Resources

Channel 9

<https://channel9.msdn.com/>

<https://channel9.msdn.com/Blogs/Azure/Connect-Twitter-to-Power-BI-and-Beyond-with-Logic-Apps>

Demos, Labs and Technical Content for Azure

<https://github.com/Microsoft/TechnicalCommunityContent>

Other

<https://docs.microsoft.com/en-us/azure/cognitive-services/>

<http://www.jamesserra.com/archive/2016/08/azure-sql-database-vs-sql-data-warehouse/>

<https://soundcloud.com/eckerson-group>

Azure Data Lake

Comprehensive, big-data storage + analytics platform

Purpose-built from real-world experiences

- Office 365, Skype, Bing, etc.

Leverage existing skills and technologies

Benefits of an Azure-hosted service

- Elastic, dynamically provisioned compute resources for varying query needs
- Infinite storage capacity
- Focus on extracting meaning from data, not on infrastructure

Azure Data Lake Store

HDFS-as-a-service

Durable, redundant storage

A variety of data scenarios

- High capacity
- High frequency
- High throughput

Store data in its native format

- Structured, semi-structured, unstructured storage formats

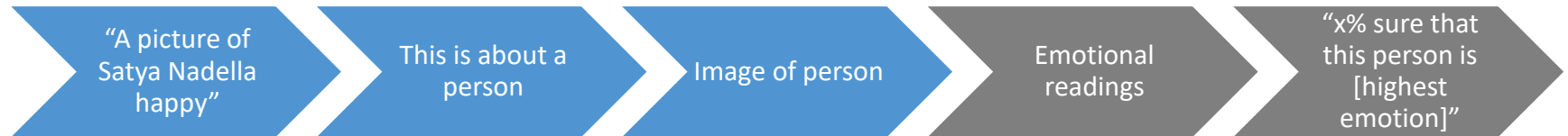


Cognitive Services:

What are Cognitive Services?

Cognitive services are a set of APIs that are designed to democratize artificial intelligence by enabling systems to see, hear, speak, understand and interpret our needs using natural methods of communication.

What these services generally do is bring structured semantic data to human knowledge I/O with a degree of confidence



The Cognitive Services

Vision	Speech	Language	Knowledge	Search
Computer Vision Content Moderator Emotion Face Video	Bing Speech Custom Speech Service Speaker Recognition	Bing Spell Check Language Understanding Linguistic Analysis Text Analytics Translator WebLM	Academic Entity Linking Knowledge Exploration QnA Maker Recommendations	Bing Autosuggest Bing Image Search Bing News Search Bing Video Search Bing Web Search

Cognitive Services: Language

Bing Spell Check

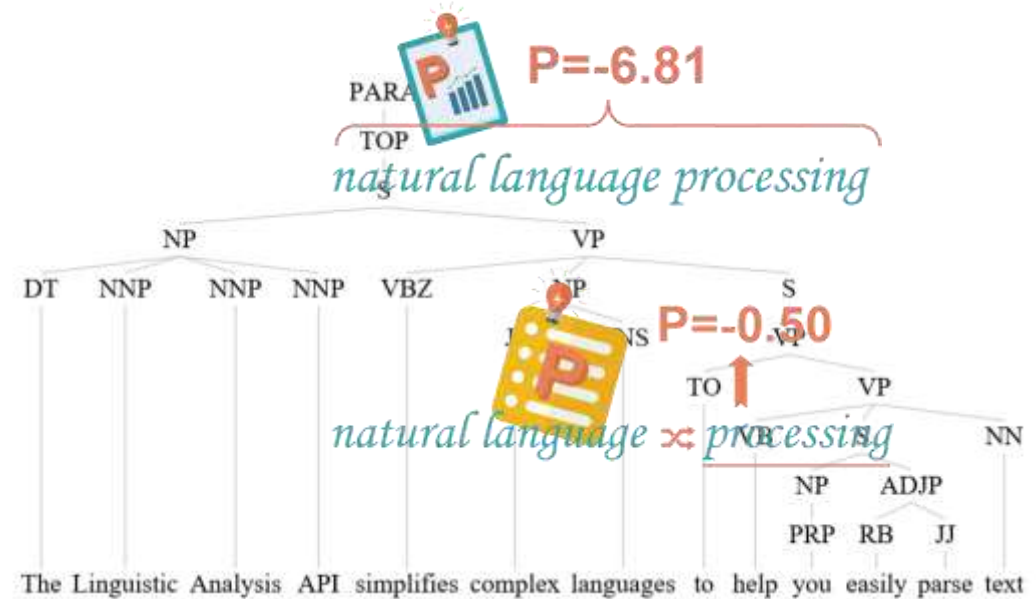
Linguistic Analytics

Text Analysis

Translator

WebLM

Language Understanding



natural language → processing
speech
find
...

Power BI

Power BI is a suite of business analytics tools to analyze data and share insights, with tools for business users to gain access to their most important metrics in a single location across all devices and platforms.



Tools & Platforms

Services	Desktop	Mobile	Embedded
<ul style="list-style-type: none">• Web-based• All platforms• Limited design tools• Limited data sources• Dashboards• Publishing	<ul style="list-style-type: none">• Windows only• Robust design tools• Query and modeling tools	<ul style="list-style-type: none">• Cross-platform• Report viewers	<ul style="list-style-type: none">• Report integration• REST services• Create custom visualizations

Power BI Building Blocks

Visualizations

Datasets

Reports

Dashboards

Tiles

