Building Blocks of Cortana Intelligence Suite

Melissa Coates
Analytics Architect
SentryOne

Blog: sqlchick.com
Twitter: @sqlchick

Presentation content last updated: April 21, 2017
Building Blocks of Cortana Intelligence Suite

Agenda

1. Azure Overview
2. Tour of Cortana Intelligence Suite: Purpose, Use Cases, Building Blocks
3. Resources for Samples & Tutorials
Cortana Intelligence Suite

Azure Overview
Microsoft Azure Platform

Microsoft Azure is a cloud computing platform and infrastructure for building, deploying, and managing applications and services through a global network of Microsoft-managed datacenters.

Azure provides services and supports many different programming languages, tools and frameworks, including both Microsoft-specific and third-party software and systems.

http://azureplatform.azurewebsites.net/
Azure Objectives

- Extend data center infrastructure
- Shared infrastructure = reduced cost of ownership
- Simplified management = reduced cost of ownership
- Built-in high availability & disaster recovery
- Separate compute from storage
- Simplified development = faster time to value
- Open source interoperability
- Integration of services
- Scalability
- Shared code base with (some) on-premises resources
- Self-service provisioning
- Built-in high availability & disaster recovery
Types of Cloud Deployments

IaaS
- Azure Virtual Machines

PaaS
- Azure SQL Data Warehouse
- Azure SQL Database
- Azure HDInsight
- Azure Data Lake Store
- Azure Stream Analytics

SaaS
- Power BI
- Office 365
- Azure Data Catalog
- Azure Data Factory
- Azure Machine Learning
- SharePoint Online
- Exchange Online
- Azure SQL Data Warehouse
- Azure SQL Database
- Azure HDInsight
- Azure Data Lake Store
- Azure Stream Analytics
- Power BI
- Office 365
- Azure Data Catalog
- Azure Data Factory
- Azure Machine Learning
- SharePoint Online
- Exchange Online
Types of Cloud Deployments

On-Premises

- Shared Infrastructure (Lower Cost)
  - More Control (Higher Administration Effort)

- Dedicated Infrastructure (Higher Cost of Ownership)
  - Less Control (Lower Administration Effort)

Cloud

- High Scalability
  - Software as a Service (SaaS)
  - Platform as a Service (PaaS)
  - Infrastructure as a Service (IaaS)
  - Virtual Server
  - Physical Server

Limits to Scalability
Cortana Intelligence Suite

Formerly known as Cortana Analytics Suite

- **Data Sources**
  - Data Management
  - Data Factory
  - Data Catalog
  - Event Hub

- **Big Data Stores**
  - Data Lake Store
  - SQL Data Warehouse

- **Machine Learning and Analytics**
  - Machine Learning
  - Data Lake Analytics
  - HDInsight (Hadoop and Spark)
  - Stream Analytics

- **Intelligence**
  - Cognitive Services
  - Bot Framework
  - Cortana
  - Power BI

- **Dashboards & Visualizations**

Data → Intelligence → Action
Cortana Intelligence in a Sentence

“Cortana Intelligence is a platform and a process to perform advanced analytics from start to finish”

Source: Chris Testa O’Neill
Data Science Team at Microsoft
Cortana Intelligence Suite Objectives

Big Data and Advanced Analytics with less cost and effort

“Intelligent action” from people or automated systems

Enable opportunities for automation and innovation:
✓ Templates
✓ Preconfigured solutions
✓ Interoperability
✓ Easier to operationalize solutions
✓ Open standards

Cortana Intelligence Suite = a marketing term for a bundle of integrated services
## Cortana Intelligence Suite Sample Scenarios

<table>
<thead>
<tr>
<th>Industry</th>
<th>Sales &amp; marketing</th>
<th>Finance &amp; risk</th>
<th>Customer &amp; channel</th>
<th>Operations &amp; workforce</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retail</td>
<td>Demand forecasting</td>
<td>Fraud detection</td>
<td>Personalization</td>
<td>Store location demographics</td>
</tr>
<tr>
<td></td>
<td>Loyalty programs</td>
<td>Pricing strategy</td>
<td>Lifetime customer value</td>
<td>Supply chain management</td>
</tr>
<tr>
<td></td>
<td>Cross-sell &amp; upsell</td>
<td></td>
<td>Product segmentation</td>
<td>Inventory management</td>
</tr>
<tr>
<td></td>
<td>Customer acquisition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Financial services</td>
<td>Customer churn</td>
<td>Fraud detection</td>
<td>Personalization</td>
<td>Call center optimization</td>
</tr>
<tr>
<td></td>
<td>Loyalty programs</td>
<td>Risk &amp; compliance</td>
<td>Lifetime customer value</td>
<td>Pay for performance</td>
</tr>
<tr>
<td></td>
<td>Cross-sell &amp; upsell</td>
<td>Loan defaults</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Customer acquisition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthcare</td>
<td>Marketing mix optimization</td>
<td>Fraud detection</td>
<td>Population health</td>
<td>Operational efficiency</td>
</tr>
<tr>
<td></td>
<td>Patient acquisition</td>
<td>Bill collection</td>
<td>Patient demographics</td>
<td>Pay for performance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacturing</td>
<td>Demand forecasting</td>
<td>Pricing strategy</td>
<td>Supply chain optimization</td>
<td>Remote monitoring</td>
</tr>
<tr>
<td></td>
<td>Marketing mix optimization</td>
<td>Performance risk management</td>
<td>Personalization</td>
<td>Predictive maintenance</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Asset management</td>
</tr>
</tbody>
</table>
Cortana Intelligence Components

Information Management
- Data Factory
- Data Catalog
- Event Hub

Big Data Stores
- SQL Data Warehouse
- Data Lake Store

Machine Learning & Analytics
- HDInsight
- Machine Learning
- Data Lake Analytics
- Stream Analytics

Visualization
- Power BI

Intelligence
- Cortana Assistant Framework
- Bot Framework
- Cognitive Services

Plus other solution components such as:
- IoT Hub
- Blob Storage
- Document DB
- Azure SQL Database
- Azure Virtual Machine
- Applications
- Azure Automation
- Azure Active Directory
- Express Route
- Virtual Network
- Azure Backup
- Azure Key Vault
Storage Layers vs. Compute Services

**Storage**
- SQL Data Warehouse
- Azure SQL Database
- Document DB
- SQL Server
- Data Lake Store
- Blob Storage

**Compute**
- SQL Data Warehouse
- Azure SQL Database
- SQL Server
- HDInsight
- Machine Learning
- Data Lake Analytics
- Azure Batch
- Stream Analytics
- Event Hub
- IoT Hub
Scenario:

Loan Repayments
Loan Repayment Scenario

Current State:

Organizational data
- Customer application (income, assets)
- Loan history
- Payment activity

Third party data
- Credit bureau history
Loan Repayment Scenario

**Desired State:**

**Predictive Analytics**
model to predict repayment ability

**Fraud alerts** re: line of credit withdrawals

**Sentiment analysis** for phone records

**Text analytics** for e-mail records

Analyze comments made on social media
Tour of Cortana Intelligence Suite
Cortana Intelligence Suite

Azure Data Lake

ADLS + ADLA: Generally Available as of November 2016
HDInsight: Generally Available as of October 2013
Azure Data Lake

A collection of 3 services:

- **Data Lake Analytics**
  - Big Data queries-as-a-service

- **HDInsight**
  - Big Data cluster-as-a-service

- **Data Lake Store**
  - Big Data storage as-a-service

A repository for analyzing large quantities of disparate sources of data in its native format

One architectural platform to house all types of data:

- Machine-generated data (ex: IoT, logs)
- Human-generated data (ex: tweets, e-mail)
- Traditional operational data (ex: sales, inventory)
Purpose

Big Data Storage
✓ Storage to support analytic applications
✓ HDFS (Hadoop Distributed File System) for the cloud, with no size limitations
✓ Stores data in its native format: objective is to not reformat

File System Optimized for Analytics
✓ An alternative to general purpose Azure Blob Storage
  ✓ Parallel read scans
  ✓ Scaled out over multiple machines
  ✓ Low latency writes
  ✓ Large file sizes

WebHDFS-Compatible
✓ Accessible to all HDFS-compliant projects (If integrated with HDInsight)
Common Use Cases

Big Data Analytic Workloads
✓ Agility: reduce up-front effort for ingestion of data (defer work to ‘schematize’)
✓ Optimized to work with ADL Analytics
✓ Also supports HDInsight (Hadoop)

Influx of Data
✓ Acquire multi-structured data
✓ Persist data in its native format
✓ No limits on account/file sizes

Active Archive
✓ Rarely used data

Data Science
✓ Analytic sandbox
✓ Raw & curated data zones

Data Warehouse Support
✓ Staging area for DW
Structuring a Data Lake

Azure Data Lake Store

- Transient/Temp Zone
- Raw Data/Staging Zone
- Analytics Sandbox
- Curated Data Zone

Metadata | Security | Governance | Information Management
Building Blocks
Repository for Ingestion of New Types of Data

1. Ingest web logs into the data lake for purpose of analyzing website visits

2. Ingest social media data into the data lake for purpose of analyzing comments
Big Data ‘queries-as-a-service’
Big Data Processing
✓ Ability to process any data, regardless of size or structure
✓ Query scalability: resources allocated for each query
✓ YARN application built on open standards
✓ Optimized to work with Azure Data Lake Store

Simplification
✓ Abstracts away the cluster nodes→focuses on convenience, efficiency, and scalability
✓ U-SQL = familiar SQL and C# to reduce learning curve
✓ Separation of ADL Analytics from ADL Store: easier to manage, debug, and optimize
U-SQL

SQL + C#
✓ New big data query processing language
✓ Applies ‘schema on read’ logic
✓ Mix of multiple SQL dialects (T-SQL and ANSI SQL)
✓ Native extensibility of user code written in C#
✓ Full C# expressions
✓ Reuse in assemblies
✓ Define custom types, functions, etc.
✓ Automatically scales and parallelizes across nodes

Read the input, write it directly to output (just a simple copy)

```csharp
@orders =
EXTRACT
OrderId int,
Customer string,
Date DateTime,
Amount float
FROM "/input/orders.txt"
USING Extractor.Tsv();

OUTPUT @orders
TO "/output/orders_copy.txt"
USING Outputters.Tsv();
```

Common Use Cases

**Focus on Business Logic**
- Focus on jobs rather than on infrastructure for a cluster
- Abstracts away the cluster nodes and focuses on convenience, efficiency, and scalability

**File Management**
- Scheduled batch processes to manage ADLS or Blob storage
  
  *(U-SQL is *not* currently suitable for ad hoc query workloads)*

**Various Size Workloads**
- Scalability on an individual job basis
- Objective is to not reserve capacity that’s not needed

**U-SQL is a Fit**
- Skillsets and preferences are a fit *(U-SQL = SQL + C#)*
Building Blocks
Batch Analysis of Data Stored in the Data Lake

1. U-SQL: execute federated query from SQL DW + ADL Store to analyze data
2. U-SQL: write results back to ADL Store
Hadoop-based distribution for ‘Big Data’ solutions

Source Data

Azure Blob Storage

General Purpose Data Storage

Container

Container

Container

Azure Data Lake Store

Optional Storage

Azure HDInsight

Distributed Data Processing, Scaling, and Querying

Hadoop Cluster

Head Node

Worker Nodes

Storm Cluster \((\text{Streaming})\)

Spark Cluster \((\text{In-Memory})\)

HBase Cluster \((\text{NoSQL})\)

Hive

Pig

MapReduce

...
Purpose

Big Data Processing
✓ A Big Data ‘cluster-as-a-service’ for distributed data processing, scaling, and querying capabilities
✓ Supports the Apache Hadoop open source ecosystem: Hive, Spark, R, Solr, Storm, etc
✓ Considered a ‘compute’ service
✓ Linux or Windows

3 Ways to Manage HDInsight
✓ As a service
✓ On-demand (ADF)
✓ Inside a virtual machine

Hortonworks Partnership
✓ Based on Hortonworks Data Platform (HDP) distribution
✓ Microsoft + Hortonworks joint engineering team
# Common Use Cases

## Big Data Scenarios  
**Volume | Variety | Velocity**  
- Leveraging the Hadoop ecosystem  
- You want to manage a cluster & go beyond what U-SQL can easily do with ADL Analytics  
- Integration with other open source projects

## Development/POC  
- Inexpensive way to test out proof of concept before investing in an on-prem big data cluster

## Data Processing Engine  
- Computations, transformations and data movement for data sent to DW and analytics systems

## Data Exploration  
- Part of data scientist’s toolbox

## One-Time Data Loads  
- Large batch jobs (ex: Sqoop)
# Deciding on ADLA vs HDInsight

<table>
<thead>
<tr>
<th></th>
<th>Azure Data Lake Analytics (ADLA)</th>
<th>HDInsight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Storage</td>
<td>Azure Data Lake Store</td>
<td>Azure Data Lake Store or Azure Blob Storage</td>
</tr>
<tr>
<td>Analytical Capability</td>
<td>U-SQL batch processing jobs</td>
<td>Supports all open source projects (Hive, Pig, Spark, SQL-on-Hadoop, etc)</td>
</tr>
<tr>
<td>Open Source</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Pricing</td>
<td>Pay-as-you-go Scale per U-SQL job</td>
<td>Big data cluster Permanent &amp; on-demand</td>
</tr>
</tbody>
</table>
Cortana Intelligence Suite

Azure SQL Data Warehouse

Generally Available as of July 2016
Relational Data Warehousing in Azure*

SQL Server in a Virtual Machine (IaaS)

Run SQL Server workloads (including SSAS, SSRS, SSIS, etc) in an Azure Virtual Machine.

Best for:
✓ Migrating/extending existing database
✓ Administer all aspects
✓ Bring your own license
✓ Dev/test scenarios

Azure SQL Database (PaaS)


Best for:
✓ < 1TB data volume (sharding across DBs is not suitable for DW workloads)
✓ OLTP with scaling & pooling needs (unpredictable workloads)
✓ Reduced administration of DB, O/S, and hardware

Azure SQL Data Warehouse (PaaS)

A DW-as-a-service (DWaaS) optimized for performance and large scale, distributed workloads.

Best for:
✓ Larger data volumes on MPP architecture
✓ Ability to scale up/down/pause on-demand
✓ Combining relational + nonrelational data

*Excluding: Other technologies in a VM such as Oracle, and Big Data technologies like Hive, etc
Azure SQL DW

- Large-scale DW workloads
- Massively parallel processing (MPP) architecture across distributions
- Part of the SQL Server family (with some key differences)

Purpose

MPP Scale-Out Query Engine
✓ Cloud-based, multi-tenant, platform-as-a-service (PaaS)
✓ Massively parallel processing (MPP)
✓ Built on SQL Server (with some differences & limits)
✓ Clustered columnstore indexes used by default

Elastic Scale
✓ Scale up/down on-demand or on schedule

PolyBase
✓ T-SQL for Hadoop queries & data loads

Storage + Compute
✓ Storage and compute is decoupled
✓ Separate billing & scaling for storage vs. compute
✓ Data Warehouse Units (DWUs) controls compute billing
✓ Increase/decrease/pause compute ability independently of data storage
Common Use Cases

Analytical and Ad Hoc Workloads
✓ Batch inserts and updates
✓ OLTP workloads are *not* suitable for SQL DW

Varying Workloads
✓ Large workloads which suit the ability to scale ‘compute’ up/down (ex: data loads or intensive analytical operations)

Large Scale Cloud DW
✓ Easier to provision large-scale environments in cloud than on-premises

Data Variety
✓ Integration with various data source types and data structures (i.e., takes advantage of PolyBase)
Two Ways of Using PolyBase

1. **Querying** of relational + semi/unstructured data in a single consolidated query. Objective: avoid data movement from where the data currently resides.

<table>
<thead>
<tr>
<th>Driver Name</th>
<th>Car Model</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amanda Green</td>
<td>Civic</td>
<td>WA</td>
</tr>
<tr>
<td>Joe Brown</td>
<td>Escort</td>
<td>WA</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Driver Name</th>
<th>Car Model</th>
<th>State</th>
<th>Avg Speed</th>
<th>Miles Driven</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amanda Green</td>
<td>Civic</td>
<td>WA</td>
<td>58</td>
<td>91</td>
</tr>
<tr>
<td>Joe Brown</td>
<td>Escort</td>
<td>WA</td>
<td>25</td>
<td>10</td>
</tr>
</tbody>
</table>
Two Ways of Using PolyBase

Data movement from source to target.

Currently PolyBase is the recommended method for loading SQL DW due to its parallel processing behavior.
Building Blocks
Modern Data Warehouse with a Data Lake

1. Load staging data from on-premises source systems to the data lake

2. PolyBase used for loading from the data lake to the data warehouse

3. PolyBase used for federated queries between the DW & ADL Store

*Performance is not yet optimized for this use case, but it works*
A service for building and operating data pipelines

Factory analogy:

- Raw Materials
- Acquire Raw Materials
- Integration and Preparation
- Finished Goods
- Deliver Finished Goods
Purpose

Data Orchestration
✓ Automation of data ingestion, orchestration, and data processing
✓ Serves as the ‘glue’ for stitching together services
Components of a Factory

Pipeline
Grouping of activities
- Activity
  Actions performed on the data
- Activity
- Activity

Dataset
Tables, files, folders, documents

Linked Service
Data source and/or compute resource for execution of activity

Source
- Ex1: Azure SQL Database
- Ex2: Azure Data Lake Store

Compute Environment
- Azure SQL Database
- Azure Data Lake Analytics

Activity
- Stored Procedure
- U-SQL Script

Sink
- Azure SQL Data Warehouse
- Azure Data Lake Store
How Does ADF compare to Integration Services?

Integration Services
- Control Flow
- Data source connection
- Data Flow
  - Source / destination
  - Built-in transformations
  - Custom transformations
- SQL Server Agent
  - Scheduling

Azure Data Factory
- Linked service
- Dataset
- Pipeline
  - Activities
  - Monitoring
  - Scheduling

Hive, Pig, C# scripts
-----
Stored proc
-----
Copy
Key Differences from SSIS

- Transformation capabilities: Hive, Pig, C#, or SQL DB stored procs, Copy activity
- Pipeline intent and scheduling are all combined together in ADF (not modular)

Planning for One Data Factory vs. Multiple Factories

- An ADF data management gateway can be used for only one ADF
- ADF diagram can get large
  - Grouping activities within pipelines can help with volume
  - Data lineage can be tracked across one ADF diagram
- Alignment of each factory with a Visual Studio project

JSON

- Most ADF elements are hand-written JSON scripts
Azure Data Factory

SELECT TC.SessionID, TC.TimestampUTC, TC.EventData.ControlName, ?WAV\ AS ControlName
FROM TC
WHERE TC.TimestampUTC = '1901-01-01 00:00:00 UTC'
AND TC.TimestampUTC < '1901-01-01 00:00:00 UTC'
WindowStart, WindowEnd
### Common Use Cases

#### Big Data Processing
- HDInsight & big data stores are its strength
- Provide the script you want to run (Hive, Pig, etc) & it will spin up/tear down an HDInsight cluster on-demand

#### Operationalize Solutions
- Scheduling of data processing scripts

#### Stage Into Supported Data Source
- Certain Azure services only support cloud data sources

#### Data is in the Cloud
- Source and/or destination is in the cloud
- Comfortable with scripting
1 PolyBase is executed within a Data Factory copy activity for movement of data; Activity windows handle incremental data loads

2 Data Factory copy activity is used for movement of data; Activity windows handle incremental data loads
A service for building predictive analytics solutions

Custom R or Python

Data pre-processing modules

Execution of algorithms
Purpose

Predictive Analytics
✓ Build predictive models using statistical techniques
✓ Learn from existing data to forecast future behaviors, outcomes & trends
✓ Minimize learning curve with predefined algorithms and drag & drop authoring environment
✓ Extensible with R and Python
Common Use Cases

Finding Anomalies
✓ Examining patterns for detection of fraud
✓ Locating unusual or abnormal equipment readings

Descriptive Analytics
✓ Analysis of returns
✓ Customer segmentation (ex: by buying habits or age group) to improve customer service
✓ Personalized offer recommendations

Predictive Analytics
✓ Credit risk
✓ Product demand & revenue predictions
✓ Customer retention
✓ Weather predictions
✓ Machine maintenance & smart buildings
✓ Hospital readmissions
✓ Student dropouts
Common Use Cases

Getting Started with Machine Learning
✓ Lowers the learning curve
✓ Extensible with custom R or Python

Operationalizing a Solution
✓ Deploy as a web service

Image from: https://powerbi.microsoft.com/en-us/blog/power-bi-azure-ml/
Building Blocks
Operationalizing an ML Model

1. Tested ML model is published as a web service
2. Execution of the ML model is invoked by calling the web service
3. Integrate scored results to the data warehouse for further analysis, using Data Factory
Cortana Intelligence Suite

Power BI

Generally Available as of July 2015
Data analysis and visualization tools
Purpose

Data Preparation, Data Modeling, Data Visualization
✓ Set of desktop, web, and mobile tools
Common Use Cases

Self-Service BI
✓ Mashup of data into a small data model which is imported & refreshed in Power BI Service
✓ Data preparation & cleansing
✓ Data visualizations

Front-End Reporting
✓ Reports and dashboards from Corporate BI sources via live queries

Prototyping
✓ Test out ideas for data structure, calculations, and reports

Third Party Reporting
✓ 3rd party content packs quick start for isolated reporting scenarios

Embed in Custom App
✓ Power BI Embedded (separate Azure Service)
Building Blocks
Self-Service and Corporate BI Scenarios

1. Connect Power BI to the data warehouse in DirectQuery mode to facilitate corporate reporting scenarios.

2. ML scored results can be returned directly to Power BI.

3. Utilize Power BI to facilitate self-service BI.

See next page
Building Blocks
Bimodal Business Intelligence

1. User runs Report A
2. Data is requested via gateway to present in report

A. Data mashup prepared
B. Data model and reports published
C. Data refresh schedule created for imported dataset
D. User runs Report B & results are returned from the imported dataset
Cortana Intelligence Suite

Azure Data Catalog

Generally Available as of April 2016
Azure Data Catalog

Register, manage, search, and explore organizational data sources

Microsoft Azure Data Catalog

Search by: Tag, Object Type, Source Type, Expert Name

Data Preview (first x rows)

Data Profiling Info
Register, manage, search, and explore organizational data sources.

- Descriptions for each column
- Tags for each column
- How to connect & how to request access
Azure Data Catalog

Alternative to ADC interface:
APIs in conjunction with custom portal

Azure Portal
Security users & roles;
Audit logs;
Resource management
https://portal.azure.com/

Client Tools
Power BI Desktop
Excel
SQL Server Data Tools

Azure Data Catalog Portal
Metadata Repository
Search + Enrich Metadata
https://azuredatalakecatalog.com

On-Premises & Cloud Source Systems

Databases
SharePoint
Data Lake
Web data, OData
Hadoop, Hive, HDFS
File System + Storage

Publishing Application
Registration of Data Assets in Data Catalog
Click once application

Subject Matter Expert

Alternative to ADC interface: APIs in conjunction with custom portal
Purpose

Data Documentation & Discovery
✓ Enterprise-wide metadata catalog for data assets
✓ Simplified data source discovery via search
✓ Enrich & understand assets with tags & annotations
✓ Collaboration between data producers & data consumers

Important to Know
✓ One Data Catalog per organization
  (not one per subscription)
✓ Authentication only accepts an organizational account
  (cannot use a Microsoft account)
# Common Use Cases

## Documentation for Centralized Data Sources
- ✓ Line of business systems
- ✓ Data warehouse, marts
- ✓ Analytic systems
- ✓ Reporting Services
- ✓ File system
- ✓ Data lake

## Facilitate Self-Service BI
- ✓ Data dictionary
- ✓ Assist combining data from multiple sources
- ✓ Reduce duplication of effort

## Capture ‘Tribal Knowledge’
- ✓ Documentation about the data is maintained by subject matter experts
- ✓ Enhance understanding

## Data Discovery & Provisioning
- ✓ Users search to discover data assets
- ✓ Who to contact to request access
Building Blocks
Data discovery

1. Register data from the data warehouse
2. Selectively register curated data from the data lake store
3. View metadata; view data from a table or view using Power BI or Excel
Cortana Intelligence Suite

Azure Stream Analytics + Event Hub

Generally Available as of April 2015
Real-time analytics for high velocity streaming events
# Azure Stream Analytics

**Purpose**

### Stream Analytics
- Analytic processing engine for streaming events
- Internet of Things (IoT) solutions for data in motion from devices, sensors, social media, etc

### Event Hub
- A publish-subscribe service that handles high volume & high velocity data streams
- Allows events to be ingested into Azure from many platforms & devices (another option: IoT Hub)
- The preferred method of event ingestion for Stream Analytics

---

**Simplification**
- Alternative to batch loading processes
- Lower bar to entry for developers by using SQL-like language
- More straightforward than an HDInsight Storm cluster
Common Use Cases

Device Monitoring & Telemetry
✓ Level beyond acceptable threshold
✓ Business continuity

Web Logs
✓ Clickstream analysis
✓ A/B testing
✓ Errors or degraded experience

Traffic
✓ Accident & traffic conditions

Demand-Based Pricing
✓ Bookings in past x minutes

Identity Protection
✓ Real-time fraud alerts
✓ Identity theft scenarios

Social Media
✓ Real-time sentiment analysis

Inventory Levels
✓ Shelf volume vs. register checkouts
Building Blocks
Fraud Alerts for Unusual Activity

1. Events ingested to the raw data event queue
2. Consume & process data for a window of time
3. Invoke ML for fraud predictions
4. Real-time reporting and/or alerts
5. Persist data for historical reporting & analysis
Cortana Intelligence Suite

Cognitive Services

General Availability Dates Vary Per API
Cognitive Services

Set of APIs, SDKs, and cloud services to build intelligent systems

<table>
<thead>
<tr>
<th>Vision</th>
<th>Speech</th>
<th>Language</th>
<th>Knowledge</th>
<th>Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer Vision</td>
<td>Speaker Recognition</td>
<td>Text Analytics</td>
<td>Academic Knowledge</td>
<td>Bing Search API</td>
</tr>
<tr>
<td>Face</td>
<td>Speech</td>
<td>Bing Speller</td>
<td>Entity Linking Service</td>
<td>Bing Image Search API</td>
</tr>
<tr>
<td>Emotion</td>
<td>CRIS</td>
<td>Web Language Model</td>
<td>Knowledge Exploration Service</td>
<td>Bing Video Search API</td>
</tr>
<tr>
<td>Video</td>
<td></td>
<td>Linguistic Analysis</td>
<td>Recommendations</td>
<td>Bing News Search API</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Language Understanding Intelligent Service</td>
<td></td>
<td>Bing Auto Suggest API</td>
</tr>
</tbody>
</table>
Give Applications a “Human Side”

✓ Designed to make applications more personalized, intelligent, and engaging

✓ Set of APIs, SDKs, and cloud services to see, hear, and interpret
  ✓ Vision ✓ Emotion
  ✓ Speech ✓ Face Recognition
  ✓ Language ✓ etc...

✓ Some APIs supported by Azure Machine Learning or Bing services

✓ Integration with U-SQL in Azure Data Lake Analytics
Cognitive Services

Common Use Cases

**Sentiment Analysis**
- Detect key phrases & topics being discussed
- Identify feedback posted to website
- Convert speech to text
- Emotion recognition

**Security Systems**
- Facial detection and recognition
- Video intelligence

**Search & Auto-Suggestions**
- Completion of partial search queries

**Personalized Shopping Experience**
- Recommend items likely to be purchased together
1. Perform text analytics on customer e-mail records using cognitive capabilities integrated in U-SQL.

2. Perform sentiment analysis on customer phone records.

3. Integrate results in the data warehouse for further analysis.
Cortana Intelligence Suite

Bot Framework

In Public Preview
Purpose

Conversation Agent
✓ Automated, yet contextual & natural, interactions with users
✓ Enable applications and services to have a conversational user interface (CUI)
Common Use Cases

**Messaging Bots**
- Web chats
- Text/SMS conversation

**Content Bots**
- Share relevant content, such as news or weather, with you

**Watcher Bots for Alerting**
- Flight is delayed
- Dental appointment reminder

**E-Commerce Bots**
- Order food
- Book a flight
- Check inventory for a product
Building Blocks
Web-based customer service

1. Conversation agent on website to handle common customer questions via chat

2. Integrate bot with Cognitive Services to enhance capabilities & perform text analytics

3. Integrate results in the data warehouse for further analysis
Why is it Called “Cortana Intelligence Suite” Anyway?

Halo
Cortana is a fictional synthetic intelligence character in the Halo novel & video game series. Cortana is “smart AI” which can learn and adapt.

Windows Digital Assistant
Cortana was the inspiration for the Windows Digital Assistant. Initially Cortana was a codename, but got such a strong response that Microsoft kept the name.

Cortana Intelligence Suite
According to Joseph Sirosh, Cortana symbolizes the contextualized intelligence they intend to achieve with the suite of tools.
Cortana Assistant

Personal Digital Assistant

Mark, I'll be ready to roll in a few seconds.

Check the weather
Add a reminder
Set an alarm
Make a call
Schedule

I've gathered a look at the day for you.

Today's Calendar

Meeting with Mike
17:00 - 17:30

American Airlines 97
On time - Lands 10 hrs 45 min
SEA --- HND

Ask me anything
Power BI integration not currently available
Purpose

Apply Language More Pervasively
✓ Virtual personal assistant for:
  ✓ Asking questions
  ✓ Finding things on PC
  ✓ Managing calendar
  ✓ Task completion
  ✓ Monitoring & alerts
  ✓ Tracking packages

Integration
✓ Reminders are integrated between Windows devices
✓ Search within other apps, such as a calendar
✓ Interact with bots to make requests
Common Use Cases

Reminders
✓ Help with remembering appointments

Work With Bots
✓ Interact with a bot to place an order or schedule a meeting
Building Blocks
Voice-based customer assistance

1. Cortana for voice-based customer assistance, integrated with the Bot Framework
2. Render a Power BI report using Cortana
Cortana Intelligence Suite

Wrap-Up
Cortana Intelligence Components

Information Management
- Data Factory
- Data Catalog
- Event Hub

Big Data Stores
- SQL Data Warehouse
- Data Lake Store

Machine Learning & Analytics
- HDInsight
- Data Lake Analytics
- Stream Analytics

Visualization
- Power BI
  - Intelligence
  - Cortana Assistant
  - Bot Framework
  - Cognitive Services

Plus other solution components such as:
- IoT Hub
- Blob Storage
- Document DB
- Azure SQL Database
- Azure Virtual Machine
- R Analytics
- Excel
- Applications
- Azure Automation
- Azure Active Directory
- Express Route
- Virtual Network
Quick Reference of Cortana Intelligence Components

Data Catalog
- Enterprise metadata catalog & data dictionary

Data Lake Store
- Big data storage

Data Lake Analytics
- Query service for big data processing

Event Hub
- Event ingestion for streaming & IoT

Event Hub
- Event ingestion for streaming & IoT

Power BI
- Data visualization + self-service data prep & modeling

Cortana Assistant
- Personal digital assistant

Bot Framework
- Agents which automate processes

Machine Learning
- Predictive analytics service

Cognitive Services
- Extends applications to see, hear & interpret

Data Factory
- Data ingestion, orchestration, and batch processing

HDInsight
- Big data clusters with support for Hadoop open source projects

SQL Data Warehouse
- Relational data warehousing at scale integrated with big data

Stream Analytics
- Processing engine for streaming & IoT

Big data analytics at scale integrated with big data
Cortana Intelligence Suite – Current State

Young Set of Services
✓ Many services have become generally available relatively recently
✓ Functionality is still maturing

Integration Still Evolving
✓ The ultimate goal is deep integration between many Azure services
Resources for Samples & Tutorials
Team Data Science Process

The 'Process' part of Cortana Intelligence Suite

The TDSP lifecycle is composed of five major stages that are executed iteratively. These include:

- **Business Understanding**
- **Data Acquisition and Understanding**
- **Modeling**
- **Deployment**
- **Customer Acceptance**

For each stage, we provide the following information:

- **Goals**: the specific objectives itemized.
- **How to do it**: the specific tasks outlined and guidance provided on completing them.
- **Artifacts**: the deliverables and the support for producing them.

Cortana Intelligence Gallery

https://gallery.cortanaintelligence.com
Cortana Intelligence Gallery – Solutions

Data Warehousing and Modern BI on Azure

By AzureWI Team for Microsoft • April 17, 2017

Summary

This one-click deploy tutorial creates a fact and dimension generation pipeline fronted with tabular models to showcase data warehousing on the cloud using the AdventureWorks dataset.

Description

Overview

With the advent of Big Data infrastructure and Cloud, enterprises have started to collect, persist, process and analyze large amount of structured and unstructured data from various internal and external data sources. Furthermore, traditional appliance-based on-premises Enterprise Data Warehouse (EDW) systems have started to demand a cost effective, flexible, and scalable alternative for cloud computing.

In this solution, we demonstrate how a hybrid EDW scenario can be implemented on Azure using:

1. Azure SQL Data Warehouse as a Data mart to vend business-line specific data.
2. Azure Analysis Services as an analytical engine to drive reporting.
3. Azure Blob Storage as a Data Lake to store raw data in their native format until needed in a flat architecture.
4. Azure HDInsight as a processing layer to transform, sanitize and load raw data into a de-normalized format suitable for analytics.
5. Azure Data Factory as our orchestration layer to move, transform, monitor and process data in a scheduled time-slice based manner.

https://caqs.azure.net/
Cortana Intelligence Gallery - Solutions

Data Warehousing and Modern BI on Azure

Prerequisites:
- This pattern requires 12 Azure Data Factory 'reserve cores' to operate. Ensure adequate cores are available before provisioning.
- At the time of deployment, please select a region that supports creation of V12 SQL Server instance for Azure SQL Data Warehouse.

Estimated Provisioning Time: 40 Minutes
Not ready to deploy or need more information on Cortana Intelligence Solutions? Contact us.

Deployment name:
dwbtest

Subscription:
Microsoft Azure Sponsorship

Locations:
East US 2

Description (optional):
Data Warehousing and Modern BI tutorial from Gallery

Create
Info on what has been provisioned & what steps need to be taken next.
LearnAnalytics@MS

Start Learning Today
Dive into Webinars, On-Demand Videos, and Classroom Training to quickly master big data and advanced analytics techniques with Microsoft.

Data Science Virtual Machine – A Walkthrough of end-to-end Analytics Scenarios
View the webinar to learn how the Data Science Virtual Machine (DSVM) in Microsoft Azure conveniently enables key end-to-end data analytics scenarios. See a variety of popular scenarios which will cover best and development training and scoring for deep-learning on GPU-based instances of the DSVM.

Tutorial: Using R for Scalable Data Analytics: From single machines to Hadoop Spark clusters
Learn about Spark data-science workflow using Sparklyr, scalable R/Spark workflows in the hands on sections, and play with Microsoft R Server to operationalize, parallelize using R and Spark other CRAN packages. (These tutorials were presented at Strata, San Jose - Mar 14, 2017)

http://learnanalytics.microsoft.com/
Azure Learning Paths

Learning path for Azure Data Factory

Data Factory is a cloud-based data integration service that orchestrates and automates the movement and transformation of data. Follow the guidance on this page for an effective learning path through Data Factory content.
Visual Studio Projects

New Project

Recent

Installed

Templates
- Business Intelligence
- Visual C#
- Visual Basic
- Visual F#
- Visual C++
- SQL Server
- Python
- Azure Data Lake
- HIVE
- Pig (HDInsight)
- Storm (HDInsight)
- U-SQL (ADLA)

- JavaScript
- PowerShell
- TypeScript
- Game
- Build Accelerator

- Data Factory

- U-SQL Project
- Class Library (For U-SQL Application)
- U-SQL Unit Test Project
- U-SQL Unit Test Sample

- Azure Data Lake

- Data Factory Templates
- Empty Data Factory Project

- Data Factory

Type: Data Factory
A project for creating a data integration solution based on a template.

Type: Azure Data Lake
A project for creating a U-SQL Sample Project.
Channel 9 Videos

Azure SQL DW
Part 2 – working with tables in Azure SQL Data Warehouse
Sep 30, 2016 at 3:22PM
by sushmay, ktlele
In this video, you will be able to create a database and tables within an Azure SQL Data Warehouse. The content will also explore the key considerations such as partitions, indexes and statistics. You...

7 ratings  0 comments
view episode

https://channel9.msdn.com/Series/Azure-SQL-DW
Additional Reading

Setting up a PC for Azure Cortana Intelligence Suite Development

What is the Cortana Intelligence Suite?

Should You Use a SQL Server Marketplace Image for an Azure Virtual Machine?

How to Build a Demo/Test Environment for Azure Data Catalog

Overview of Azure Data Catalog in the Cortana Intelligence Suite
Thank You for Attending!

To download a copy of this presentation:
SQLChick.com “Presentations & Downloads” page

Melissa Coates
Analytics Architect
SentryOne

Blog: sqlchick.com
Twitter: @sqlchick

Creative Commons License:
Attribution-NonCommercial-NoDerivative Works 3.0