Building Streaming Data Pipelines Using Azure Cloud Services
My Assumptions for Today

1. You know what the **cloud** and what **Azure** is
2. You have **awareness** of the **main Azure services**
3. You have used the **Azure Portal** before
4. You have **knowledge** of **database services**
5. You will **rate** my session **10 / 10**
Basic Program for Today

10 Introduction

20 What exactly is the data platform nowadays?

30 Data pipeline services in Azure

40 Demonstration: Lets see some data pipelines!

50 Where to next? Wrap up and summary

60 goto 10
Introduction
Why is data so important?

Because there’s just so much of it!
On-Prem vs IaaS vs PaaS vs SaaS – Which One?

- On Premises:
  - Applications
  - Data
  - Runtime
  - Middleware
  - O/S
  - Virtualization
  - Servers
  - Storage
  - Networking
  - You manage
  - Managed by Microsoft

- Infrastructure (as a Service):
  - Applications
  - Data
  - Runtime
  - Middleware
  - O/S
  - Virtualization
  - Servers
  - Storage
  - Networking
  - You manage
  - Managed by Microsoft

- Platform (as a Service):
  - Applications
  - Data
  - Runtime
  - Middleware
  - O/S
  - Virtualization
  - Servers
  - Storage
  - Networking
  - Managed by Microsoft

- Software (as a Service):
  - Applications
  - Data
  - Runtime
  - Middleware
  - O/S
  - Virtualization
  - Servers
  - Storage
  - Networking
  - Managed by Microsoft

Serverless
Azure Services – Which One?
http://azureinteractives.azurewebsites.net/Azure101Cards/default.html
And so what exactly is a “data pipeline” anyway?

• Different definitions **depending on which vendor you talk to**
• **Microsoft** have **no formal definition**, *but some I like*...

  “**pipelines** are formed from multiple individual ‘fit for purpose’ services aligned in sequences that perform a set of specific targeted actions on data that is typically in transit.”

  Source: (Rolf Tesmer) 😊

  “**a pipeline** is a set of data processing elements connected in series, where the output of one element is the input of the next one. The elements of a **pipeline** are often executed in parallel or in time-sliced fashion”

  Source: (Wikipedia)

  “**a data pipeline** is the software that consolidates data from multiple sources and makes it available to be used strategically”

  Source: (Unknown)
What is the LAMBDA architecture?

“The Objective of **Lambda Architecture** is to leverage the combined power of both **batch** & **real-time** processing to address the business scenarios where it requires both **historic view of the data** as well as getting insight into the **data in real-time** as business happens.”

https://docs.microsoft.com/en-us/azure/architecture/guide/architecture-styles/big-data
https://azure.microsoft.com/en-us/services/iot-edge/
Where did this come from, and why do we care?

1. Customers are on a **multi-year transformational journey**

2. Many **data sources** are **not static** or **at rest**

3. Solutions **cannot wait** for data to be landed before using it

4. Building pipelines...
   - **Historically** → Complex, costly capital investment, time consuming
   - **Today** → Fast, simple, “fit for purpose” services, same **data platform**

As modern day Data Professionals we have to deal with it
What exactly is the Data Platform Nowadays?
What was the data platform?

Up till ~5 years ago it was typically a relational platform

...and... included relational-like services (OLTP, OLAP, DW, ETL, MDM, +)

...and... often on-prem, or in a hosted DC

...and... rarely hosted in external public cloud providers

Occasionally included special projects (ie Big Data, NoSQL, IoT)

https://mrfoxsql.wordpress.com/2017/04/19/what-exactly-is-the-data-platform-nowadays/
What is the data platform now?

- A mix of... **different data types, shapes, rates & sizes**
- A mix of... **on-prem & public cloud**
- A mix of... **deployment models** (IaaS, PaaS, SaaS)
- A mix of... **specific “fit for purpose”** individual data services

- These services are across a range of uses including:
  1. **Ingestion**
  2. **Transformation**
  3. **Storage**
  4. **Analytics**
  5. **Visualisation**
Data Pipeline Services in Azure
What are some of the Azure pipeline services?

Ridiculous Example Architecture

- **High Speed NoSQL Distributed Data Layer**
- **Serverless Custom Code & Functions**
- **Incoming Data Flow**
- **Cloud Data Ingestion Point AEH | AIH**
- **General Data Archive**

**Event Hubs**

**Stream Analytics**

**Storage blob**

**Logic App**

**Machine Learning**

**Data Lake**

**Data Factory**

**Structured Storage**

**SQL Data Warehouse**

**Structured Warehouse ASDB | ASDW**

**Unstructured Storage**

**CEP In-Stream Analytics**

**Selective Load**

**Real-Time Report**

**Trend Report**

**Intelligence Report**

**Analytics Report**

**Data Movement / Orchestration**

**Reporting / Visualisation**

**Business Workflow / Logic**

**Machine Learning API Calls**

**Selective Load**

**Intelligent Services**

**Cognitive API Calls**

**Operationalised Data Science**

**Real-Time Report**

**Archive**

**Scheduled Pull**

**Full Load**
Demos / Examples:

Let's see some Azure pipelines!
Demonstration → Mobile G-Force Solution - !

[Diagram of the Mobile G-Force Solution]

- **IoT Hub**: Event JSON
- **Mobile**: Event JSON
- **Stream Analytics**: Event Archive JSON
- **Storage blob**: Event Archive JSON
- **Azure Stream Analytics**: SQL Database
- **Machine Learning**: Event
- **JSON G-Force Prediction API**: All Events CSV
- **REAL-TIME Event Telemetry Report**: Streaming Dataset
- **ON-DEMAND Event Trend Report**: SQL Query
- **Twilio Phone Call**: EVENT-DRIVEN
- **New Event Trigger**: JSON
- **Alert Event**: CSV
Demonstration → Mobile G-Force Solution -!
EVENT → AEH → ASA = < 2 sec

Data Archive
Avg: 56GB/day

Logic App
Status Report
~1/hour

Service Bus Queue
~1/hour

Stream Analytics
Telemetry Input 3900/sec

Power BI
Aggregation Path
3900/Sec → 1/min

Azure SQL Database
Max: 3900/sec
Avg: 2300/sec
(5 days = 1b rows)
(1 year = 72b rows)

JSON Events
Stream (HOT Path)

Stream Analytics
SHEventHistory
(Short Term Store)

Azure SQL Database

Power BI
1/min
(200K tumbling window)

Real Time
Dashboards
(troy.earle)

Historical
Reports
(troy.earle)

Other Examples → High Scale Web Search Telemetry
Web Search Telemetry – Total Events (By Day)

**INCOMING MESSAGES**

**AVG Workload** → 1,410,000,000 / week  
= 201,000,000 / day  
= 8,392,000 / hour  
= 139,000 / min  
= 2,330 / sec
Web Search Telemetry – Events/Sec (By Hour)

600% increase over 9 hours
When is scale an issue?

- What do you mean by “scale”?
  - **IoT Device – Streaming Telemetry Workload:**
    - 29,000 / sec → 2,505,600,000 / day → 914,544,000,000 / year

- Lambda principals still apply!!
  - **Ingestion** → handling the “peak” rate without latency/delay/error
  - **Processing/Speed** → need data granularity, or are aggregate windows OK
  - **Storage/Batch** → need adhoc on-demand data engineering, or recurring
  - **Serving** → what granularity is important, what decisions will be made

- **Question:** can you pre-process at “the edge”? 
So where to from here?
Wrap up and summary
What’s next for the data platform?
...and what does this mean for us Data Professionals?

1. **On-prem** hosted/deployed data solutions are **diminishing**

2. **Public cloud** data ecosystem is **mature** and **expanding**

3. **IaaS is popular, PaaS is King** (ie *Serverless world is the future*)

4. **Customer “expectation”**...
   ...This is the “**Domain of the Data Professional**”
Where can I try this out – or learn more?

• **Cortana Intelligence Gallery Pre-Built Solutions (one-click deploy)**
  - Vehicle Telemetry
    https://gallery.cortanaintelligence.com/Solution/Telemetry-Analytics
  - Personalised Offers
    https://gallery.cortanaintelligence.com/Solution/Personalized-Offers-2
  - Energy Demand Forecasting
    https://gallery.cortanaintelligence.com/Solution/Demand-Forecasting-3

• **EdX Self-Paced Courses (3-4 hrs/week for ~4 weeks)**
  - Developing IoT Solutions with Azure IoT
    https://www.edx.org/course/developing-iot-solutions-azure-iot-microsoft-dev225x
  - Processing Real-Time Data Streams in Azure
    https://www.edx.org/course/processing-real-time-data-streams-azure-microsoft-dat223-2x-0
  - Orchestrating Big Data with Azure Data Factory
    https://www.edx.org/course/orchestrating-big-data-azure-data-microsoft-dat223-3x-0
Your Homework → Twitter Social Media Analytics

Social Media Pipeline
Region: Australia SE
Function
.Net (C#)
Azure SQL DB
Sentiment Schema
Call
Tweet Data
Sentiment
Key Phrases

Twitter
Logic App
Check Twitter
Every 3 min

Sentiment Key Phrases

Azure Machine Learning
Region: Southeast Asia

Text Analytic API

Sentiment

Data Connection

Power BI Desktop
On-Prem

New Power BI Reports

Office 365
Power BI

C Level Dashboards
Marketing Dashboards

Executive

Social / Marketing

Tweets
@Handles
#Tags

powerbi.com

C Level Dashboards
Marketing Dashboards

~USD $6 / day
100K tweets / month
Use FREE Azure

This is your homework assignment for today!

[End of Presentation]
Appendix
Appendix and References

- Online Azure Interactive Services Diagram - [http://azureplatform.azurewebsites.net/en-us/](http://azureplatform.azurewebsites.net/en-us/)
- Cortana Intelligence Gallery - [https://gallery.cortanaintelligence.com/](https://gallery.cortanaintelligence.com/)
- EdX - Developing IoT Solutions with Azure - [https://www.edx.org/course/developing-iot-solutions-azure-iot-microsoft-dev225x](https://www.edx.org/course/developing-iot-solutions-azure-iot-microsoft-dev225x)
Other Examples → Business Incident Management

Region: Australia South East
- Azure SQL DW / DB (PaaS) (push)
- Event msg (tabular data)
- External – Business Event Message
- JSON
- Azure Stream Analytics (PaaS)
- (pull) JSON msg ([max 256KB] 1 Event/Msg)
- (push) Original JSON msg
- Azure Event Hub (PaaS)
- Database (SQL)
- SQL Agent Scheduler
- Secure endpoint
- Data
- Azure VM (IaaS)
- SQL SSAS (cubes)
- SQL SSRS
- SQL SSIS
- External Ref Data
- External Data Sources
- HDInsight (on demand analytics) (future option) (PaaS)
- In-Stream Reporting
- Live/Batch ML (R Language API calls)
- SQL Agent Scheduler
- stg tables
dw tables

Region: Southeast Asia
- Azure Machine Learning / R (future option)
- On demand Data Science (future option)
- Users
- New Reports and Datasets
- powerbi.com
- Azure Cloud (SaaS)
- (pull) Live/Batch Reporting (tabular data)
- (pull) Live/Batch ML (R Language API calls)
- Users
- Cortana
- New Reports and Datasets
- Power BI Desktop (on-prem author)
Where can I find even more examples of this stuff?

https://gallery.cortanaintelligence.com/browse?categories=["10"]&orderby=freshness_desc
Azure has more global regions than any other cloud provider—offering the scale needed to bring applications closer to users around the world, preserving data residency, and offering comprehensive compliance and resiliency options for customers.

50 regions worldwide 140 available in 140 countries
### Microsoft Azure Data Services

<table>
<thead>
<tr>
<th>IaaS</th>
<th>PaaS</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL Server, Oracle, Mongo, MySQL, ++ in Azure VM</td>
<td>Azure Relational Databases</td>
</tr>
<tr>
<td></td>
<td>SQL, PostgreSQL, MySQL</td>
</tr>
</tbody>
</table>

- **fully featured RDBMS**
- **transactional processing**
- **rich query**

**Managed as a service**

- **Elastic scale**
- **Schema-free data model**
- **Internet accessible http/rest**
- **Arbitrary data formats**
Azure Relational Database Platform (PaaS)

Scale/Sizing Based on "Throughput Units"
- CosmosDB (NoSQL)
- SQL Data Warehouse
- SQL Database

Scale/Sizing Based on Cores + Storage
- SQL Managed Instance
- PostgreSQL
- MySQL

Database
- Intelligent: Advisors, Tuning, Monitoring

Services
- Flexible: On-demand scaling, Resource governance

Platform
- Trusted: HA/DR, Backup/Restore, Security, Audit, Isolation

Power BI, App Services, Data Factory, Analytics, ML, Cognitive, Bot...

Azure Compute

Azure Storage

Global Azure with 50 Regions

Get Azure...?

Your Enterprise Agreement (EA)
Various options – Currently being setup and configured for MLC
Would be linked to your corporate identity/login/account

Azure 30 day free account up to $260 (time boxed to 30 days)
Would be linked to your personal identity/login/account

MSDN
Thee are free monthly Azure credits within MSDN subscriptions. Rolls over month to month
Would be linked to your corporate identity/login/account

Research Programs and Grants
Free credits available for specific research programs = https://www.microsoft.com/en-us/research/academic-program/microsoft-azure-for-research/
Learn Azure…?  Free Online-Training

edX – Free online courses on Microsoft Azure
45+ Free Azure Courses = https://www.edx.org/course?search_query=azure

Introduction to Azure - https://www.edx.org/course/introduction-microsoft-azure-microsoft-azure201x


Developing Azure Solutions - https://www.edx.org/course/developing-microsoft-azure-solutions-microsoft-dev233-1

Developing Apps and Bots - https://www.edx.org/course/developing-intelligent-apps-bots-microsoft-dat211x-1

Deliver a DW in the Cloud - https://www.edx.org/course/delivering-data-warehouse-cloud-microsoft-dat220x-0

Delivery Big Data Solutions with Machine Learning - https://www.edx.org/course/developing-big-data-solutions-azure-microsoft-dat228x

Provision SQL Databases in Azure - https://www.edx.org/course/provisioning-databases-azure-sql-server-microsoft-dat219x-0
Learn Azure...?  Free 1x day In-Person Training

https://azure.microsoft.com/en-us/community/events/?Country=Australia&query=azure+discovery+day

Azure Discovery Days

Reserve your spot. Accelerate your learning.

Microsoft Azure is a growing collection of integrated cloud services – analytics, computing, database, storage, mobile, networking, the web – to help you move faster, work smarter and achieve more.

Its potential is unlimited if you know how to unlock it. Which is why we offer Azure Discovery Days: free sessions led by Microsoft Azure experts and built to help you extract extra value out of Azure.

Think big, then delve deeper

Each Discovery Day offers an Azure overview, plus primers on security, billing, identity, storage, networking and compute. We’ll also cover these core topics.

- Administer Azure via the Azure Portal & PowerShell
- Create & configure VMs in Azure to connect securely to On-Premises
- Monitoring & Recovery Services
- Monitoring and operational Insights
- Automation and Azure Quick Start Templates
- MyApps portal & Self Service capabilities
- Multi Factor Authentication
- Protecting, encrypting, and tracking documents

Pick your city. Start your journey.
Learn Azure...? Patterns, Blogs and Feedback

Need Azure Patterns and Guidance? Check out the Azure Architecture Centre - !

- [https://docs.microsoft.com/en-au/azure/architecture/](https://docs.microsoft.com/en-au/azure/architecture/)

Need Some Azure Updates? Subscribe to the Global Azure Blog & Update Feed - !


You Have a Cool New Azure Idea? Submit it to Azure Ideas & Feedback - !

- [https://feedback.azure.com/forums/34192--general-feedback](https://feedback.azure.com/forums/34192--general-feedback)
Certify in Azure...?

Azure Certifications


![Diagram of Microsoft Certification Paths]

**Recommended**

- Mobility
- Cloud Platform & Infrastructure
- Productivity
- Data Management & Analytics
- App Builder
- Business Applications

**Recommended**

- MCSE Mobility (Earned 2018)
- MCSE Cloud Platform & Infrastructure (Earned 2018)
- MCSE Productivity (Earned 2018)
- MCSE Data Management & Analytics (Earned 2018)
- MCSD App Builder (Earned 2018)
- MCSE Business Applications (Earned 2018)
Certify in Azure...?

MCSE Cloud Platform & Infrastructure

Reflects updated list of exams for MCSA: Cloud Platform and the Elective Pool

Select one elective to complete your MCSE. Choose a different elective next year or retake a previously taken cloud-based exam to stay relevant and prove your expanded skills.
Certify in Azure...?

MCSE Data Management & Analytics

DATA MANAGEMENT & ANALYTICS

461: Querying Microsoft SQL Server 2012/2014
462: Administering Microsoft SQL Server 2012/2014 Databases
468: Implementing a Data Warehouse with Microsoft SQL Server 2012/2014

761: Querying Data with T-SQL
762: Developing SQL Databases
765: Provisioning SQL Databases

764: Administering a SQL Database Infrastructure
766: Developing SQL Data Models
769: Analyzing and Visualizing Data with Power BI

774: Perform Cloud Data Science with Azure Machine Learning
779: Analyzing Data with Excel

776: Engineering Data with Microsoft Cloud Services
777: Engineering Data with Microsoft Cloud Services

MCSE SQL Server 2012/2014
MCES SQL 2016 Database Admin
MCES SQL 2016 Database Admin
MCES SQL 2016 BI Development

MCSE Data Management & Analytics Earned: 2018

Elective

ELECTIVE EXAM POOL
(CHOSE ONE*)

- 464: Developing Microsoft SQL Server Databases
- 465: Designing Database Solutions for SQL Server
- 466: Implementing Data Models and Reports with Microsoft SQL Server
- 467: Designing Business Intelligence Solutions with Microsoft SQL Server
- 473: Designing and Implementing Cloud Data Platform Solutions
- 475: Designing and Implementing Big Data Analytics Solutions
- 760: Developing SQL Databases
- 767: Implementing a SQL Data Warehouse
- 768: Developing SQL Data Models
- 773: Analyzing Big Data with Microsoft R
- 774: Perform Cloud Data Science with Azure Machine Learning
- 775: Perform Data Engineering on Microsoft Azure HDInsight

*Select one elective to complete your MCSE. Choose a different elective next year or retake a previously taken cloud-based exam to stay relevant and prove your expanded skills.

Reflects removal of planned path