Database Answers

Ruxley Towers,UK

A Look at Azure Data Management

Barry Williams
barryw@databaseanswers.org

1. Management Summary

1.1 Introduction

This document is intended to be a Tutorial to Microsoft's Azure for people (like myself) who come from a traditional Data Management background and want something suitable for a Data Analyst and not a Developer.

It can provide an approach that can be adapted by anyone coming from a traditional Data Management background who wants to apply their expertise in an Azure environment.

If you have any questions or comments we would be very happy to hear from you on barryw@databaseanswers.org.

There seems to be some duplication in the features offered by the Data Factory, the Events Hub and Stream Analytics.

I have included a 'Duplication of Azure Facilities' in Chapter 3 that you can tailor to suit your own requirements.

On the basis that Stream Analytics offers an 'SQL-like' language, we have opted to adopt that as our choice of preferred technique for our Proof-of-Concept for Azure Data Management in Chapter 5.

1.2 BI and Data Modelling

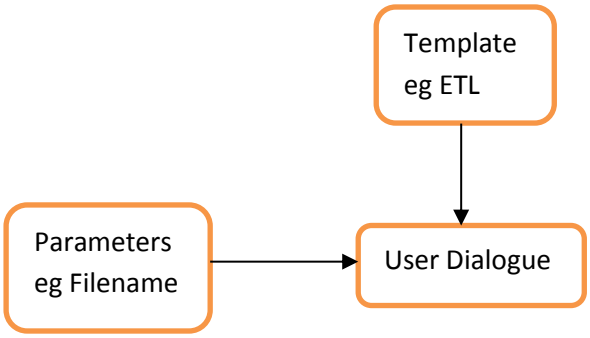
Here is a quote that I like from the IT Director's Guide to Self-Service BI published by Blue Granite (https://www.blue-granite.com/) :-

> "Data Modelling plays an important role in ensuring good quality data.
> The most successful self-service BI implementations have well-defined data models as a semantic layer".

1.3 A User-friendly Interface

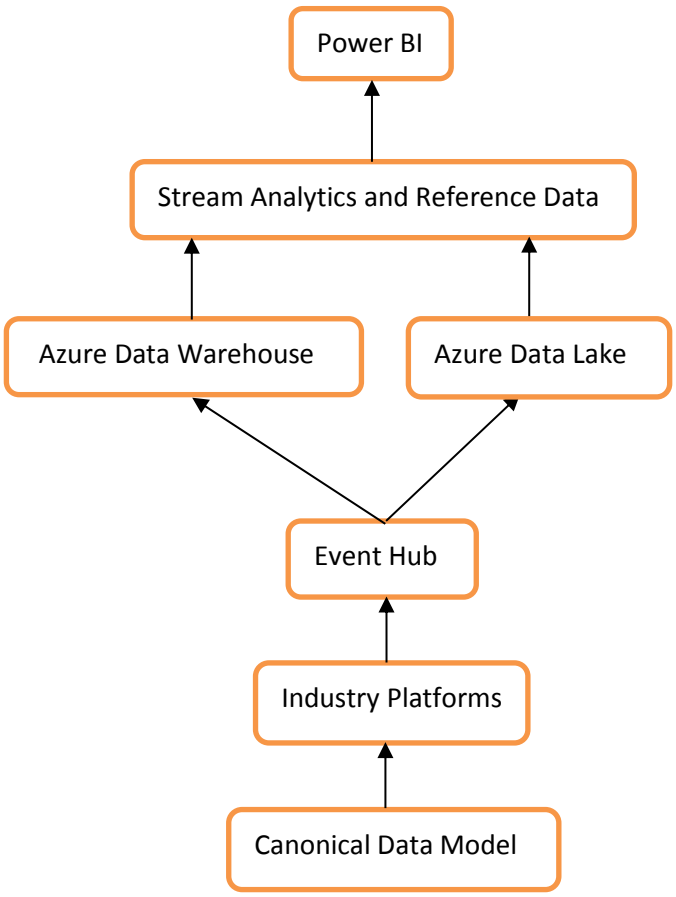
My plan is to develop a range of Self-Service Azure facilities, such as Asset Management and Insurance.
I want to make them user-friendly
In order to achieve this, I want to avoid the lengthy text-based specifications that characterises the present Azure user interface and instead, to adopt an interactive approach.

In Chapter 4, I show examples of ETL parameters.



1.4 Our POC Azure Architecture

Here we show the Architecture that we are planning for our Proof-of-Concept Projects.
It is based on our Canonical Data Model with our Reference Data Architecture and Azure's Stream Analytics and Power BI.
We discuss this in more detail in Chapter 4.



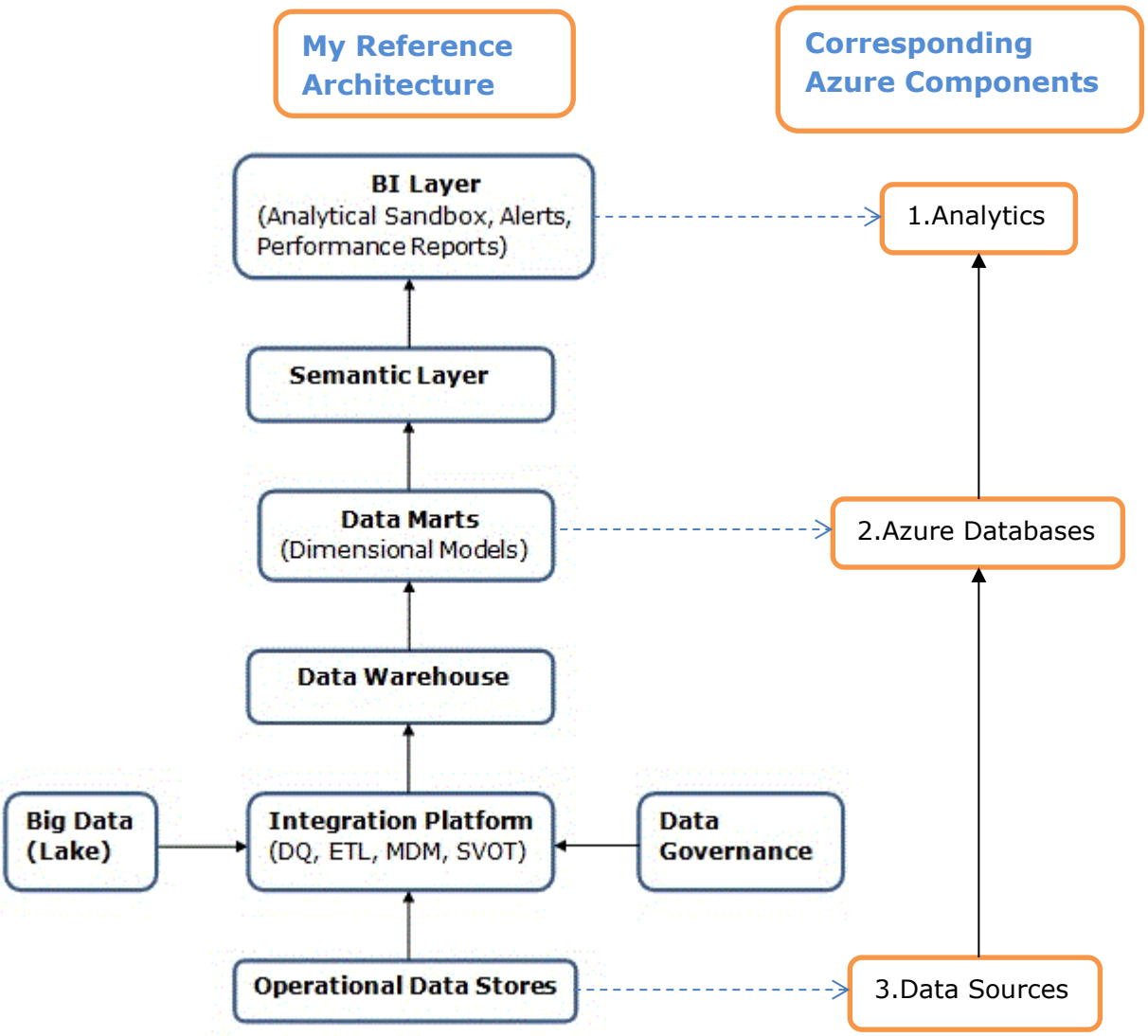
2. Traditional and Azure Data Architectures

2.1 Comparison

We have established 'Best Practice' Data Architecture that we use in all our consulting assignments. It appears on this page of our Database Answers Web Site :-

- http://www.databaseanswers.org/reference_data_architecture.htm

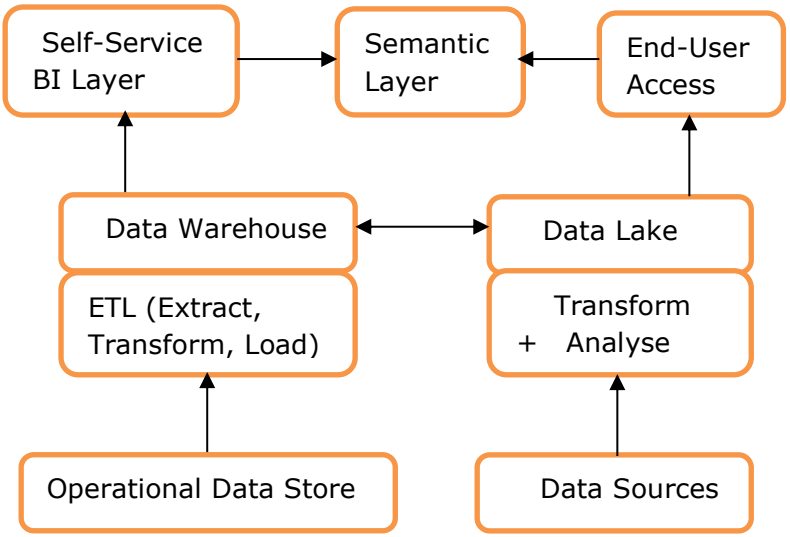
Here we show how the Azure Components map to my Traditional Data Architecture which demonstrates that there is a correspondence between the two that can inform all our thinking and analysis :-



2.2 Our Self-Service Data Platform

This Architecture shows how a Data Warehouse and a Data Lake can be used together to provide an integrated view of data to the end-user.

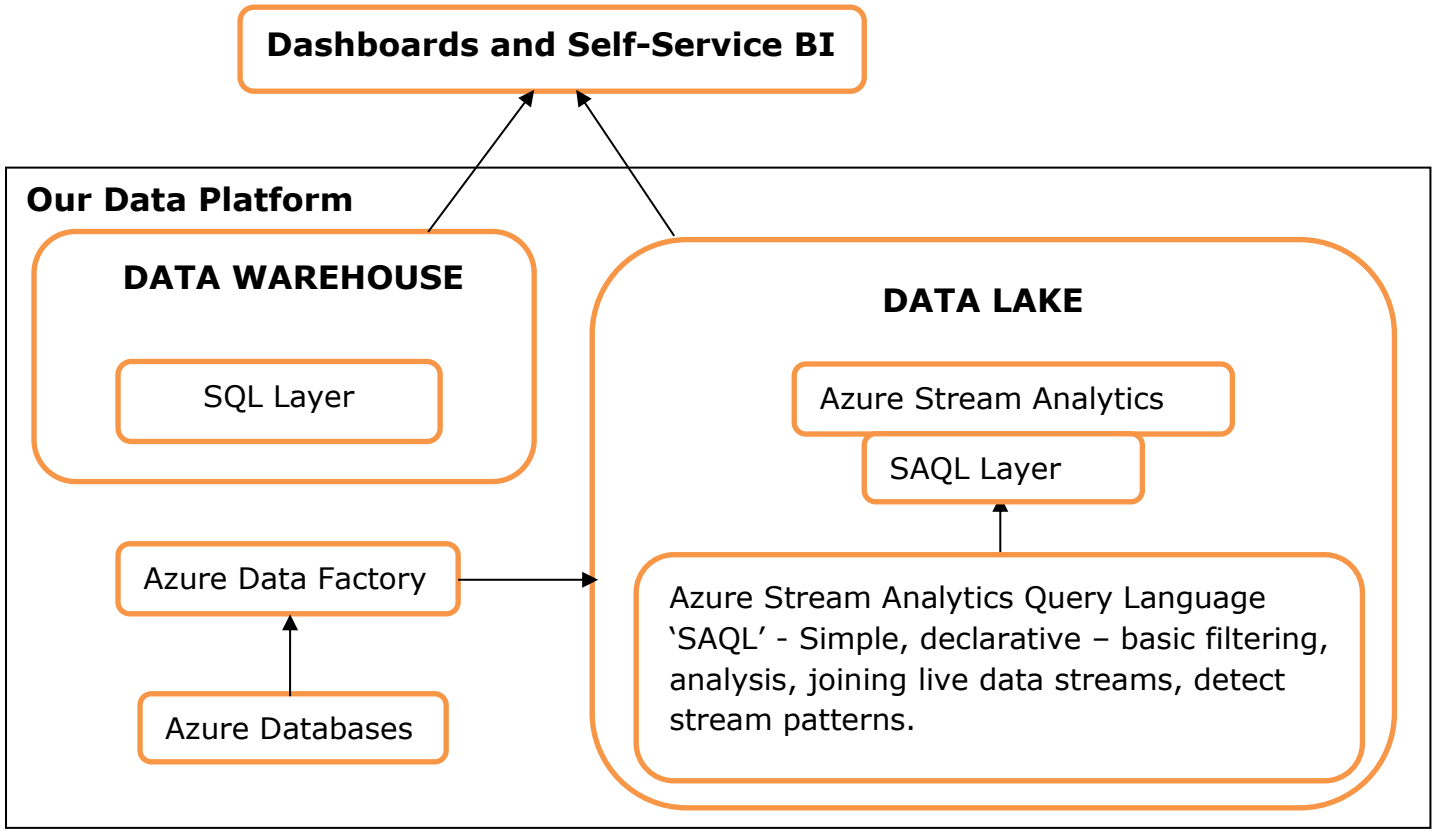
The Semantic Layer provides a way for us to give a 'user-friendly' interface where database terminology is replaced by business terms – for example, a 'Party' replaced by 'Customer'.



2.3 Our Azure Data Platform

Here we show how a Data Warehouse and a Data Lake can co-exist.

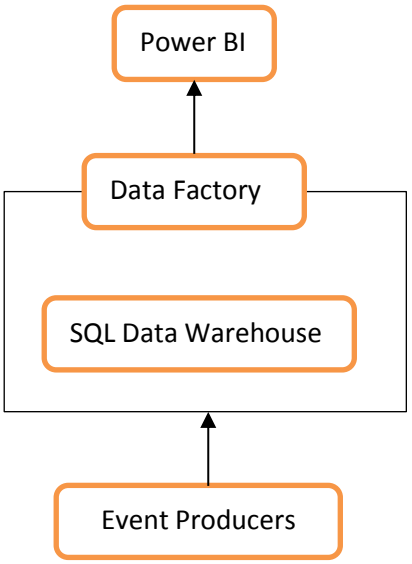
The SQL Layer for the Data Warehouse and the 'SQL' Layer for the Data Lake provide us with a very convenient ability to establish a consistent approach.



2.4 Azure Data Factory

This makes Azure Services User-Friendly for 'Data people' as an introduction to Azure
This is how the Architecture maps on to our Reference Data Architecture.
The Steps are as follows :-

3. If we start with CSV data, then our first step is to load the data into the SQL Data Warehouse.
4. Next we calculate aggregate data for a Data Mart.
5. Finally we use Power BI to build Dashboards to display KPIs and other results.



Here is a good Microsoft introduction to the Data Factory –

- https://blogs.msdn.microsoft.com/data_insights_global_practice/2015/10/23/end-to-end-data-processing-using-azure-data-factory-overview/

2.5 ETL and ELT

ETL stands for 'Extract, Transform and Load' and is fundamental to Traditional Enterprise Data Management.

Wikipedia has an entry for it at :-

- https://en.wikipedia.org/wiki/Extract,_transform,_load

where it says

"**Extract, Transform, Load** (**ETL**) refers to a process in data warehousing. The ETL process became a popular concept in the 1970s - data is E(xtracted) from data sources; and then T(ransformed) for storing in the proper structure for querying and analysis; finally, the data is L(loaded) into the target database or data warehouse.

Wikipedia also has an entry for ELT at :-

- https://en.wikipedia.org/wiki/Extract,_load,_transform

and says "ELT is an alternative to ETL used with Data Lake implementations."

For ELT, the data is transformed after being loaded into the Database.

3. Duplication of Azure Facilities

Here we analyse Features of important categories. It seems that multiple facilities have been developed.

Here we analyse Features of important categories. It seems that two Query Languages have been developed, both (perhaps) modelled on the industry-standard SQL.

Our choice is shown in what we call 'Our Azure Data Platform' and is shown in Section 2.3.

Here we can see that Azure offers :-

- two Query Languages based on SQL
- Three kinds of Data Storage
- Three kinds of Analytics

CATEGORY	COMPONENT	QUERY LANGUAGE	Unique Selling Points (USP)
Analytics	Data Lake Analytics		
	Power BI		
	Stream Analytics	SAQL	Provides compatibility with SQL
Data Storage	Azure Database		
	Azure Data Lakes	U-SQL	
	Azure Data Warehouse		No Primary or Foreign Keys

4. Azure Design Pattern Examples

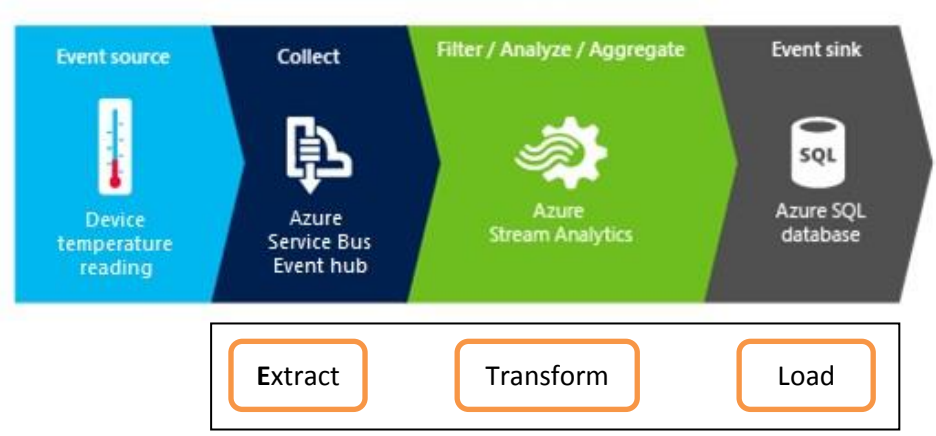
Here we map each Azure Component to the corresponding function of an Extract, Transform and Load process.

4.1 Number 1 – Stream Analytics + Event Hubs

From a Blog by Microsoft's Kirk Evans :-

- https://blogs.msdn.microsoft.com/kaevans/2015/02/26/using-stream-analytics-with-event-hubs/

This post is a demonstration using the walkthrough document Get started using Azure Stream Analytics.



This is the Link above that says 'Get started using Azure Stream Analytics' :-

- https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-real-time-fraud-detection

When we look at the four stages shown in this simple diagram, we can see that they correspond to what is called 'Extract, Transform and Load' (or ELT) in traditional Enterprise Data Management. ELT has been in use since the 70's and here is the Wikipedia entry :-

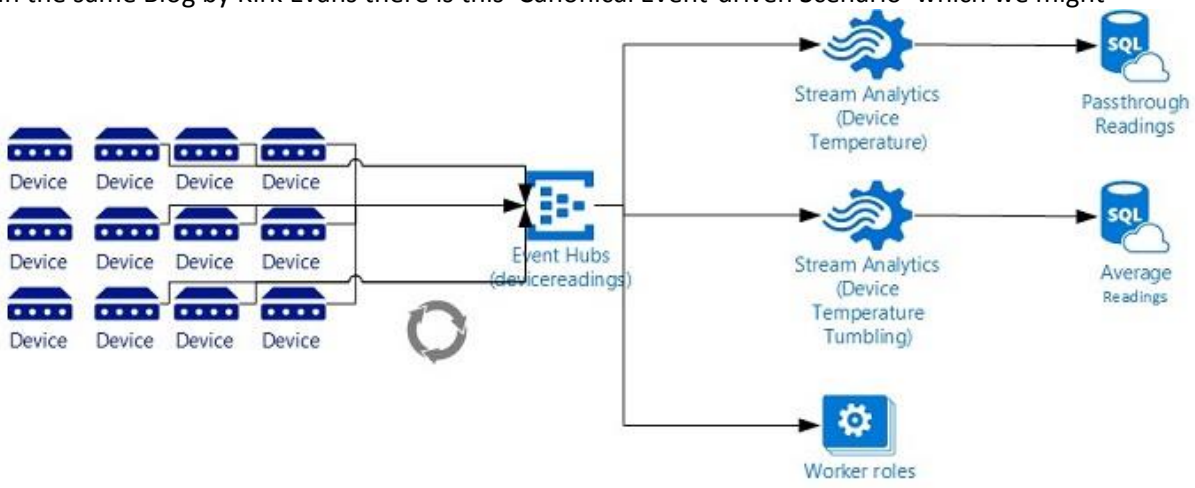
- https://en.wikipedia.org/wiki/Extract,_transform,_load

In passing, we might note that a variation of ETL has emerged for Data Lakes, called ELT :-

- https://en.wikipedia.org/wiki/Extract,_load,_transform

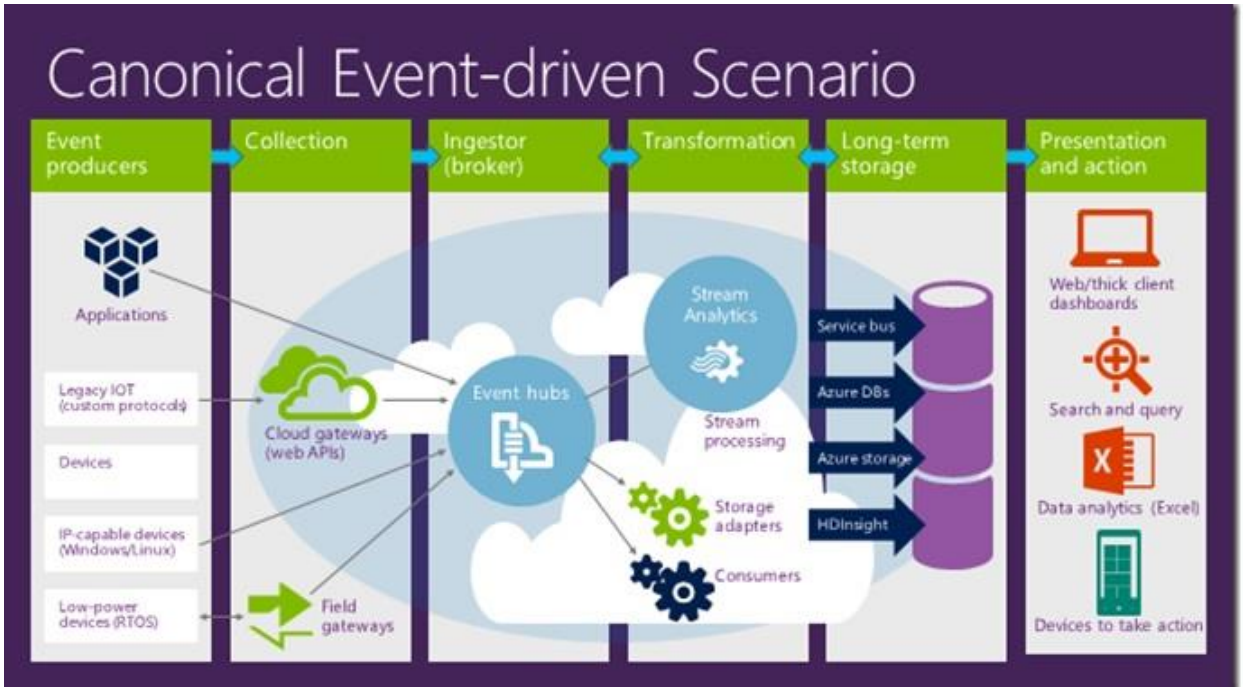
4.2 Number 2 – Simple Event Hubs

In the same Blog by Kirk Evans there is this ‘Canonical Event-driven Scenario’ which we might



4.3 Number 3 – Complex Event Hubs

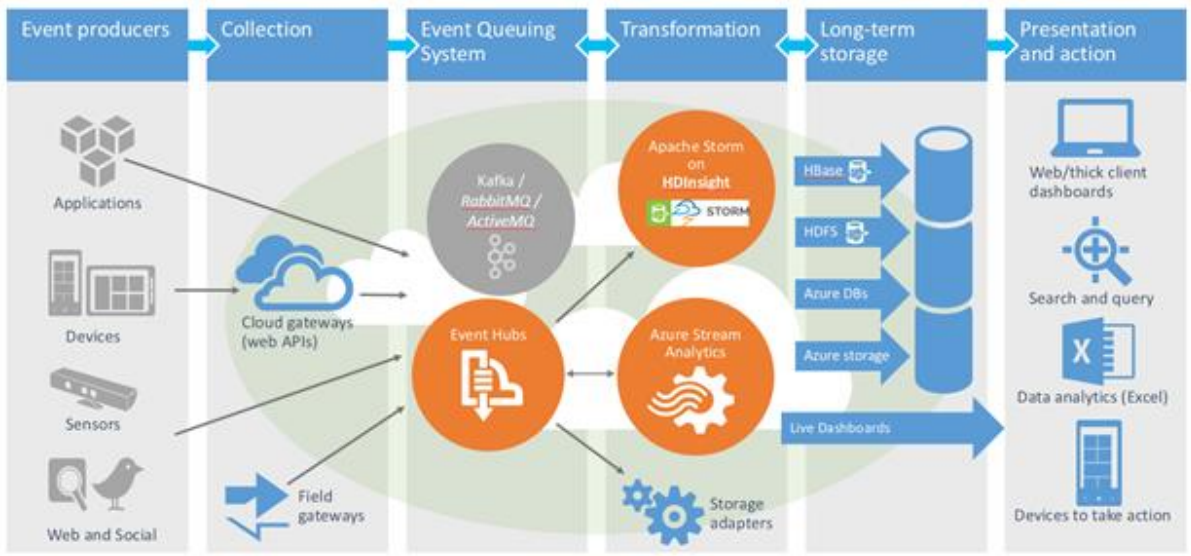
In the same Blog by Kirk Evans there is this ‘Canonical Event-driven Scenario’ which we might consider a more complex example of the same Design Patter n:-



4.4 Number 4 – Data Factory

This diagram appears in Scott's Blog –

- https://weblogs.asp.net/scottgu/azure-announcing-new-real-time-data-streaming-and-data-factory-services



We can see clearly that this is another example of the ETL Design Pattern with

1. Extract (Event producers, Collection and the Event Queuing Systems
2. Transform (Transformation)
3. Load (Long-term storage)

and is therefore a good candidate for our first Design Pattern.

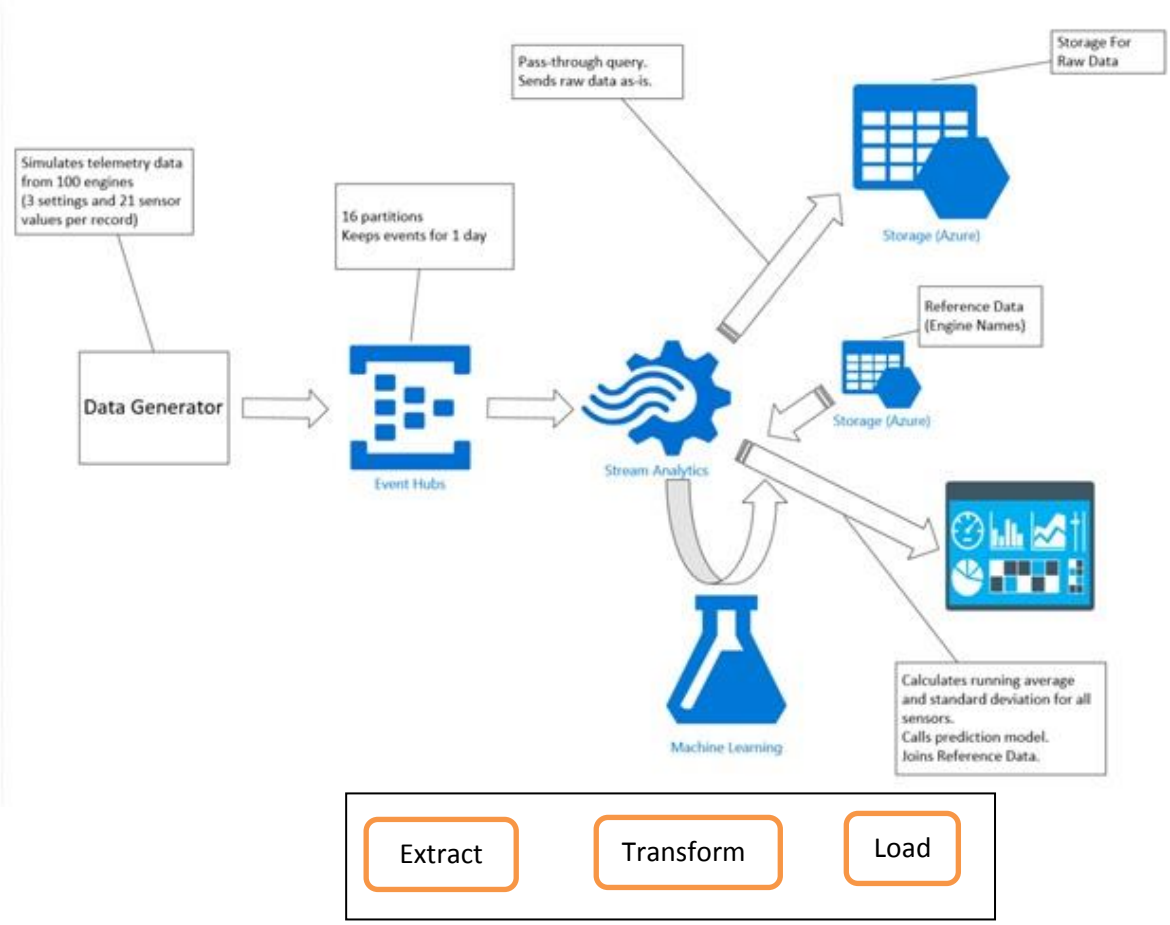
4.5 Number 5 – Stream Analytics and Power BI

This has a good architecture view :-

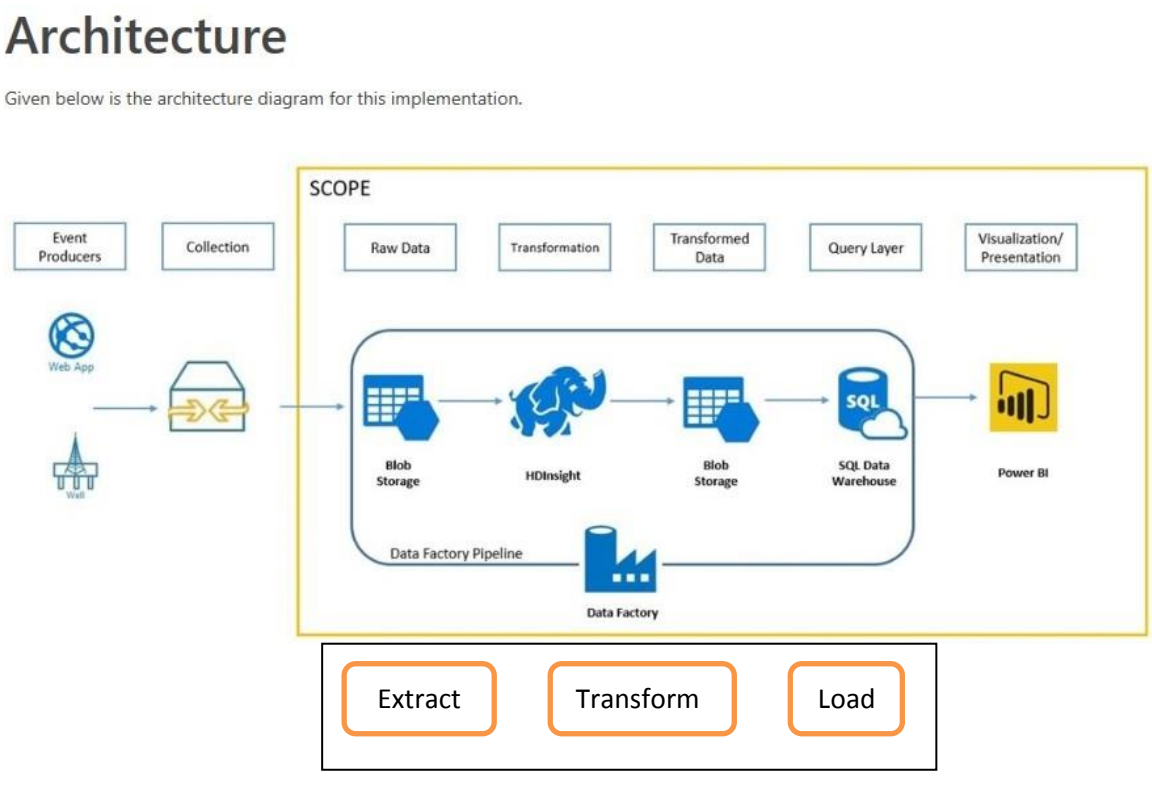
- https://blogs.msdn.microsoft.com/data_insights_global_practice/2015/09/16/event-hubs-stream-analytics-azureml-powerbi-end-to-end-demo-part-i-data-ingestion-and-preparation/

which looks like this :-

This is a high level architecture view of the demo setup.

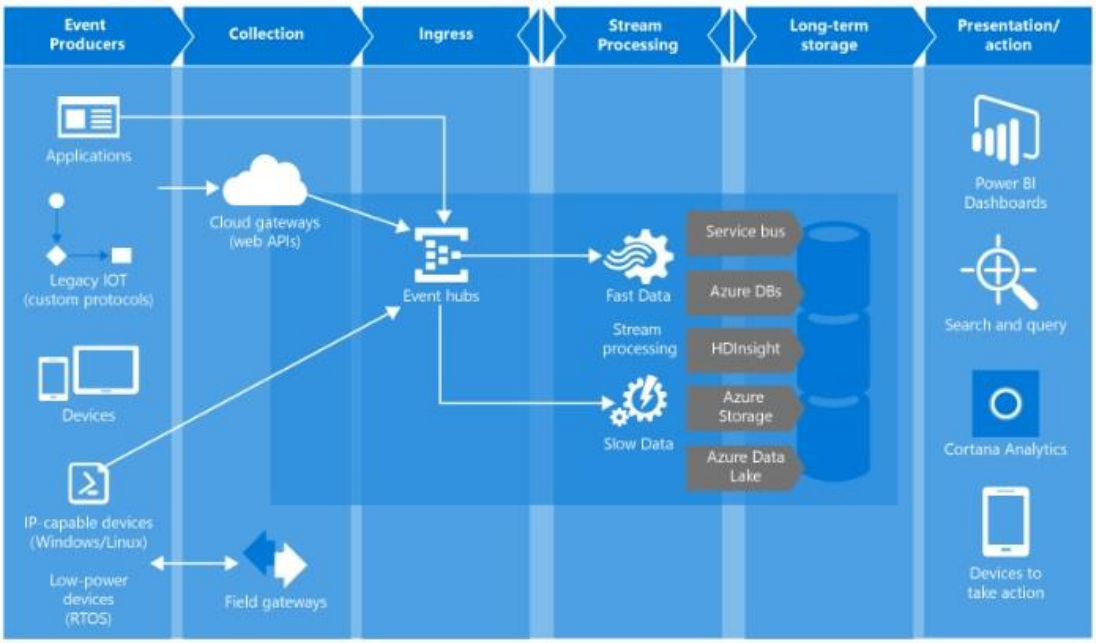


4.6 Number 6 – Data Warehouse and SQL Data Warehouse



4.7 Number 7 - Events Hub Overview

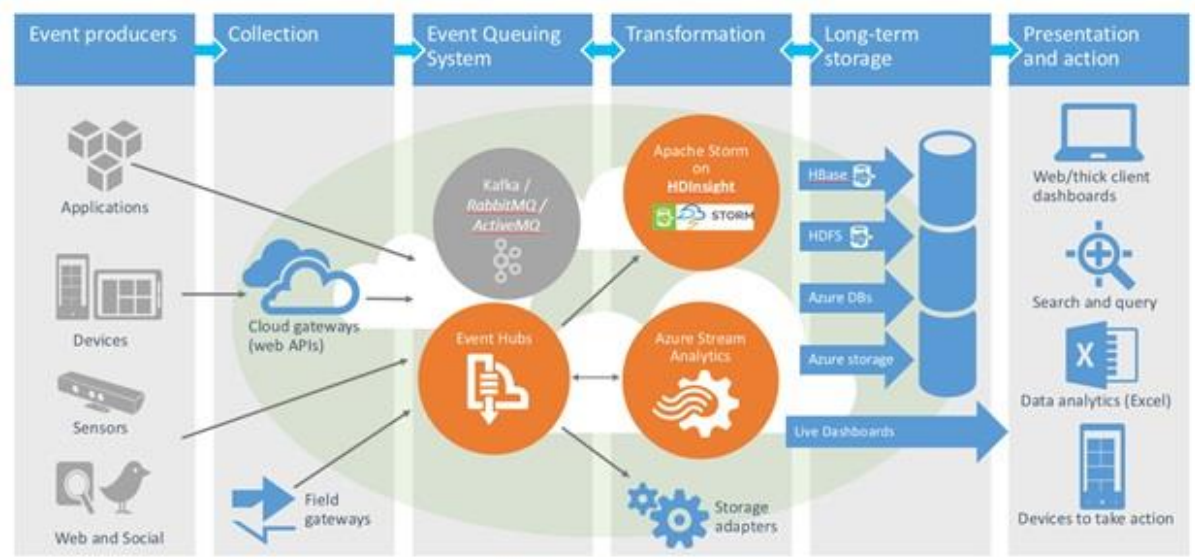
From this page - https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-what-is-event-hubs



4.8 Number 8 - Event Hubs from Scott Guthrie's Blog

In this Chapter I have drawn on material from a Blog by the highly-respected Microsoft Guru called Scott Guthrie discussing Stream Analytics and the Data Factory :-

- https://weblogs.asp.net/scottgu/azure-announcing-new-real-time-data-streaming-and-data-factory-services



5.My POC Short-Term Plans

5.1 Proof-of-Concept Projects

My Plans are to develop three Azure-based applications to prove that my approach is valid.

I have chosen these applications based on these three from my over 1,500 Data Models :-

- Air Transport Platform
  - http://www.databaseanswers.org/data_models/POC_air_transport_platform_2020/index.htm
- Asset Management
  - http://www.databaseanswers.org/data_models/asset_management/Asset_Mgt_Platform_POC.htm
- United Nations
  - http://www.databaseanswers.org/data_models/un_global_compact_platforms_for_2017/index.htm

My target is to complete three by the end of May, and a further 21 by October 6th. (my birthday !!!) :

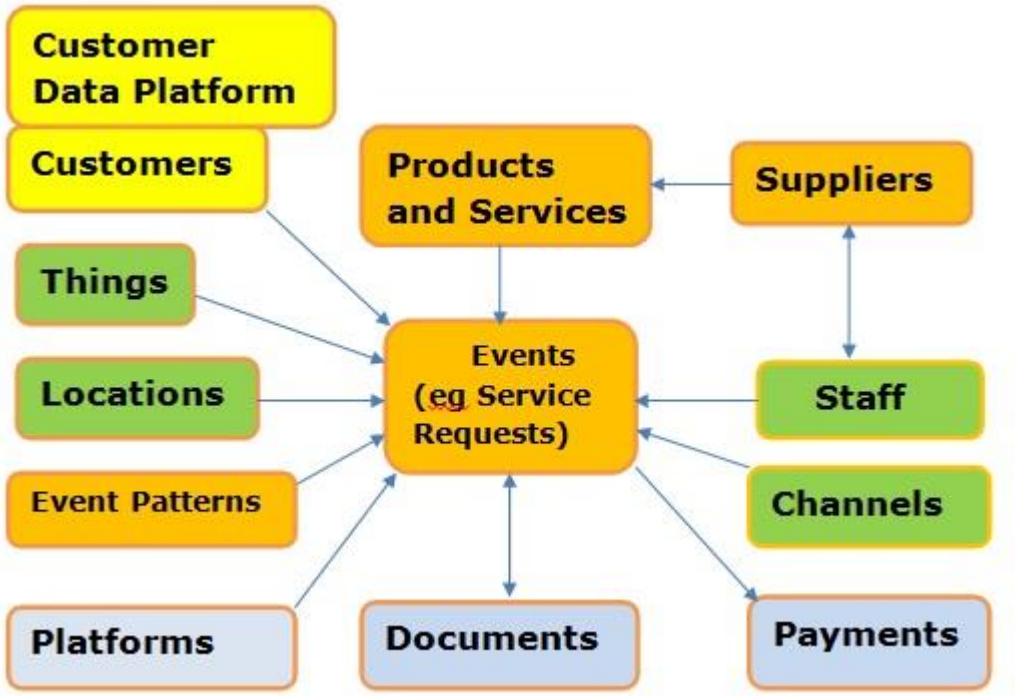
  - http://www.databaseanswers.org/data_models/POC_Cloud_Services.htm

5.2 Our Canonical Data Model

It is on this page - http://www.databaseanswers.org/data_models/canonical_data_models/index.htm

And is show here.

As you can see, the dominant Entity is 'Events' so we can integrate it neatly with Azure's Event Hubs.



5.3 Our Industry-specific Data Platforms

I have designed three Industry-specific Data Platforms based on my Generic Platform shown on this page –

- http://www.databaseanswers.org/data_models/generic_platform/index.htm

Appendix A. Useful Azure Links

A.1 Get Started

- https://blogs.msdn.microsoft.com/data_insights_global_practice/2015/09/16/event-hubs-stream-analytics-azureml-powerbi-end-to-end-demo-part-i-data-ingestion-and-preparation/

Good introduction

- https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-what-is-event-hubs

Programming details

- https://docs.microsoft.com/en-us/azure/event-hubs/event-hubs-programming-guide

Kirk Evans Blog

  - https://blogs.msdn.microsoft.com/kaevans/2015/02/26/using-stream-analytics-with-event-hubs/

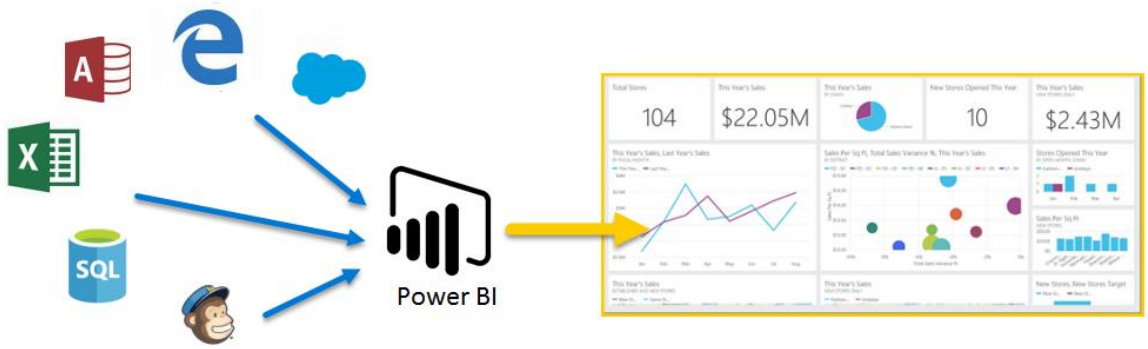
Kirk Evans at LinkedIn

- https://www.linkedin.com/in/kaevans/

A.2 Power BI

- https://blogs.msdn.microsoft.com/data_insights_global_practice/2015/09/16/event-hubs-stream-analytics-azureml-powerbi-end-to-end-demo-part-i-data-ingestion-and-preparation/
- https://mva.microsoft.com/en-US/training-courses/faster-insights-to-data-with-power-bi-jump-start-8291?l=FGLOu8Xy_2504984382
- What is Power BI -
  - https://powerbi.microsoft.com/en-us/guided-learning/powerbi-learning-0-0-what-is-power-bi/

"**Power BI** lets you easily connect to your data sources, visualize (or discover) what's important, and share that with anyone or everyone you want.



Learning Power BI –

- https://powerbi.microsoft.com/en-us/guided-learning/powerbi-learning-0-1-intro-using-power-bi/

The Parts of Power BI
Power BI consists of an **Power BI Desktop**, an online SaaS (Software as a Service) service called the **Power BI service**, and mobile Power BI **apps** available on Windows phones and tablets, as well as for iOS and Android devices."

For our POC, we will be using the **Power BI service.**

A.3 No Keys in an SQL Data Warehouse

In this article –

- http://biinsight.com/azure-sql-data-warehouse-and-power-bi/

I found this statement :-
" But, it is important to keep in mind that there are some features like primary keys and foreign keys that are NOT supported in Azure SQL Data Warehouse which affect the way we use Power BI as a data visualisation tool over Azure SQL Data Warehouse.
There is a workaround for this that we can create some SQL views in Azure side to make it work. This can be an expensive solution. The other way is to load the data warehouse into a Power BI Desktop model which can detect the relationships automatically."

It makes me wonder about this strange state of affairs.