

Building an Azure Business Intelligence Solution End to End

Hands On Workshop

Paul Andrew | Senior Consultant

Terry McCann | Principal Consultant

Simon Whiteley | Cloud Architect



Gold Data Analytics
Gold Data Platform
Gold Cloud Platform



<https://github.com/Adatis>

ModernDataWarehouseWorkshop

Agenda for the Day

Module 1

Microsoft Azure

Module 2

Storage
Uploading Data
Data Lake

Module 3

Real-time Data
Streaming
Power BI

Module 4

U-SQL - Data
Transformation
Basics

Module 5

USQL - Advanced
Analytics
Cognitive Services

Module 6

Data Factory
Orchestration
Dynamic Pipelines

Module 7

Data Presentation
& Consumption
Power BI Models

Module 8

Other Services
Q&A

Session Agenda

What is Azure
Data Lake?

Storage & Compute

Why use Data
Lake?

The Modern Data
Warehouse

How can we work
with Data Lake?

Development &
Management

Agenda

What is Azure
Data Lake?

Storage & Compute

Why use Data
Lake?

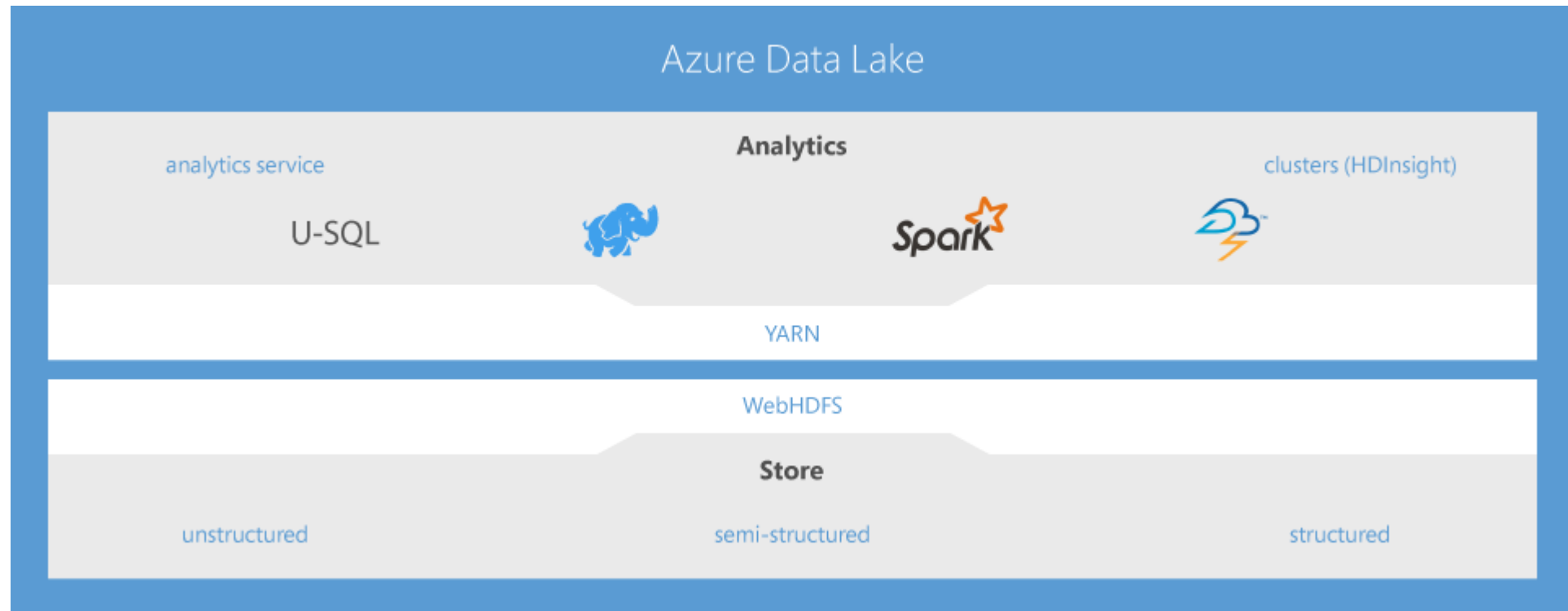
The Modern Data
Warehouse

How can we work
with Data Lake?

Development &
Management

What is Azure Data Lake?

The Microsoft version:



What is Azure Data Lake?

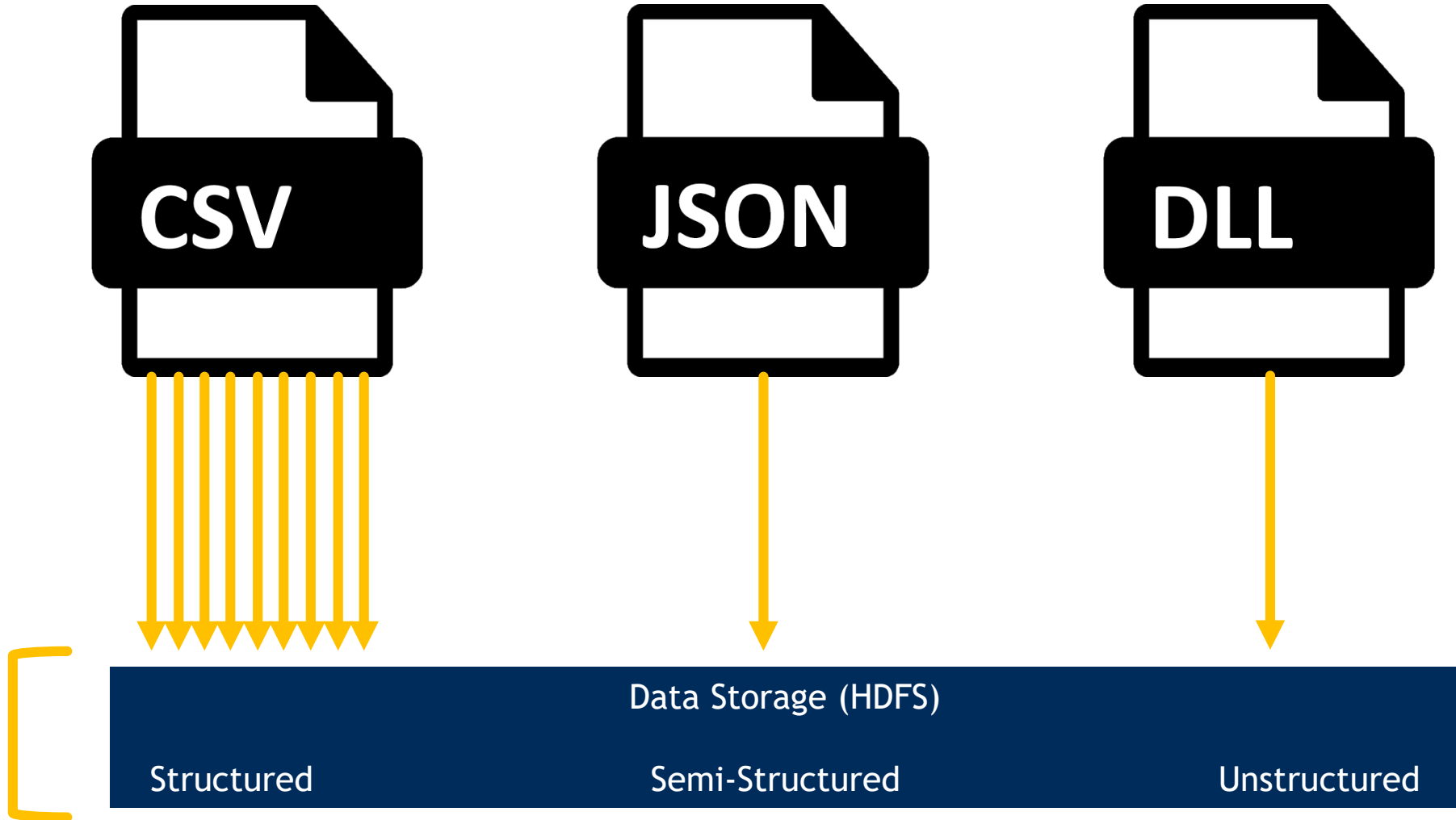
HDFS Extents (MB)

Default: 128

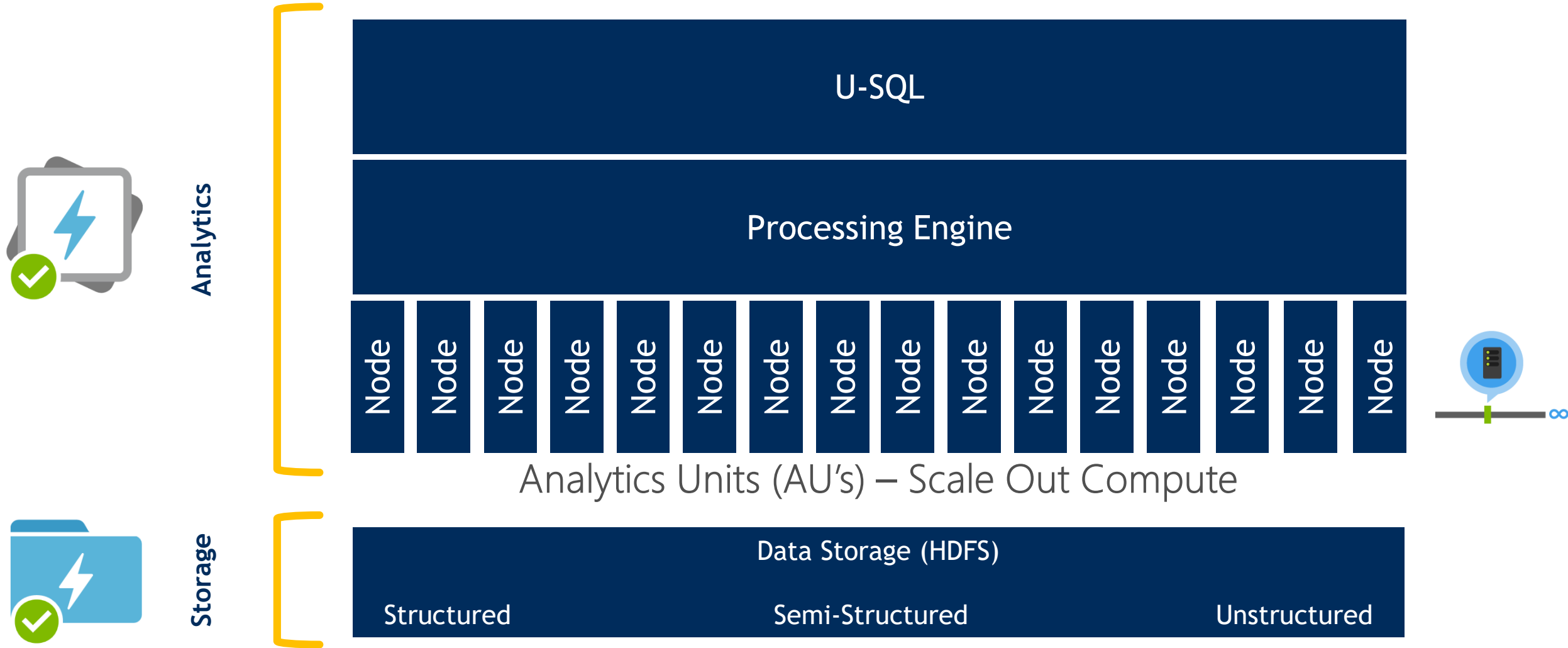
Variable: 4 to 256



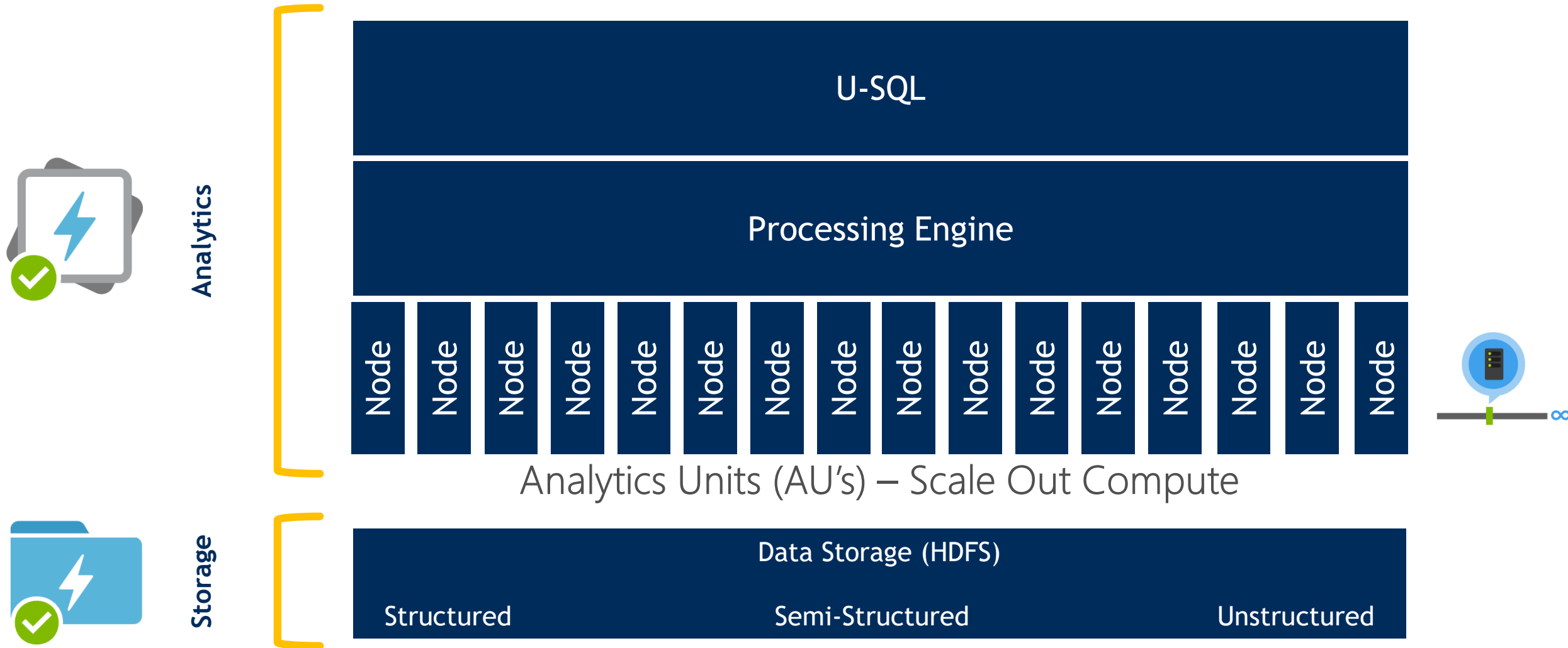
Storage



What is Azure Data Lake?



What is Azure Data Lake?



Azure Data Lake vs Other Data Services

Azure SQL Database (SQLDB)



T-SQL

Compute Node

Database Storage Engine

Database Transaction Units (DTU's) –
Scale Up Compute

Azure SQL Data Warehouse (SQLDW)



D-SQL

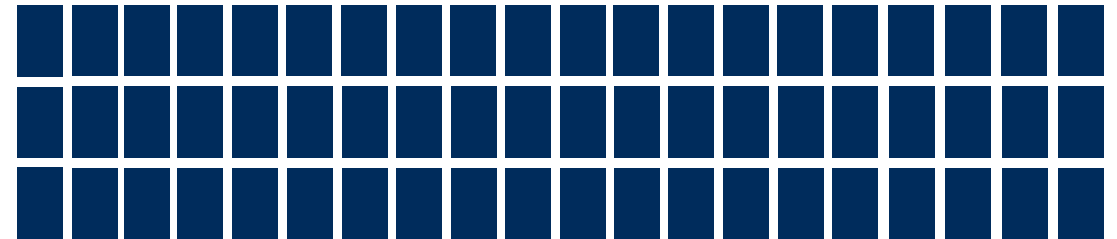
Control Node

Compute Node

Compute Node

Compute Node

Compute Node



Database Warehouse Units (DWU's) –
Scale Out Compute

Agenda

What is Azure
Data Lake?

Storage & Compute

Why use Data
Lake?

The Modern Data
Warehouse

How can we work
with Data Lake?

Development &
Management

Why use Azure Data Lake?

The Microsoft version:

Microsoft Azure

SALES 0800 098 8435 ▼ | MY ACCOUNT | PORTAL | Search 🔍

Why Azure? Solutions Products Documentation Pricing Partners Blog Resources Support

FREE ACCOUNT >

Data Lake

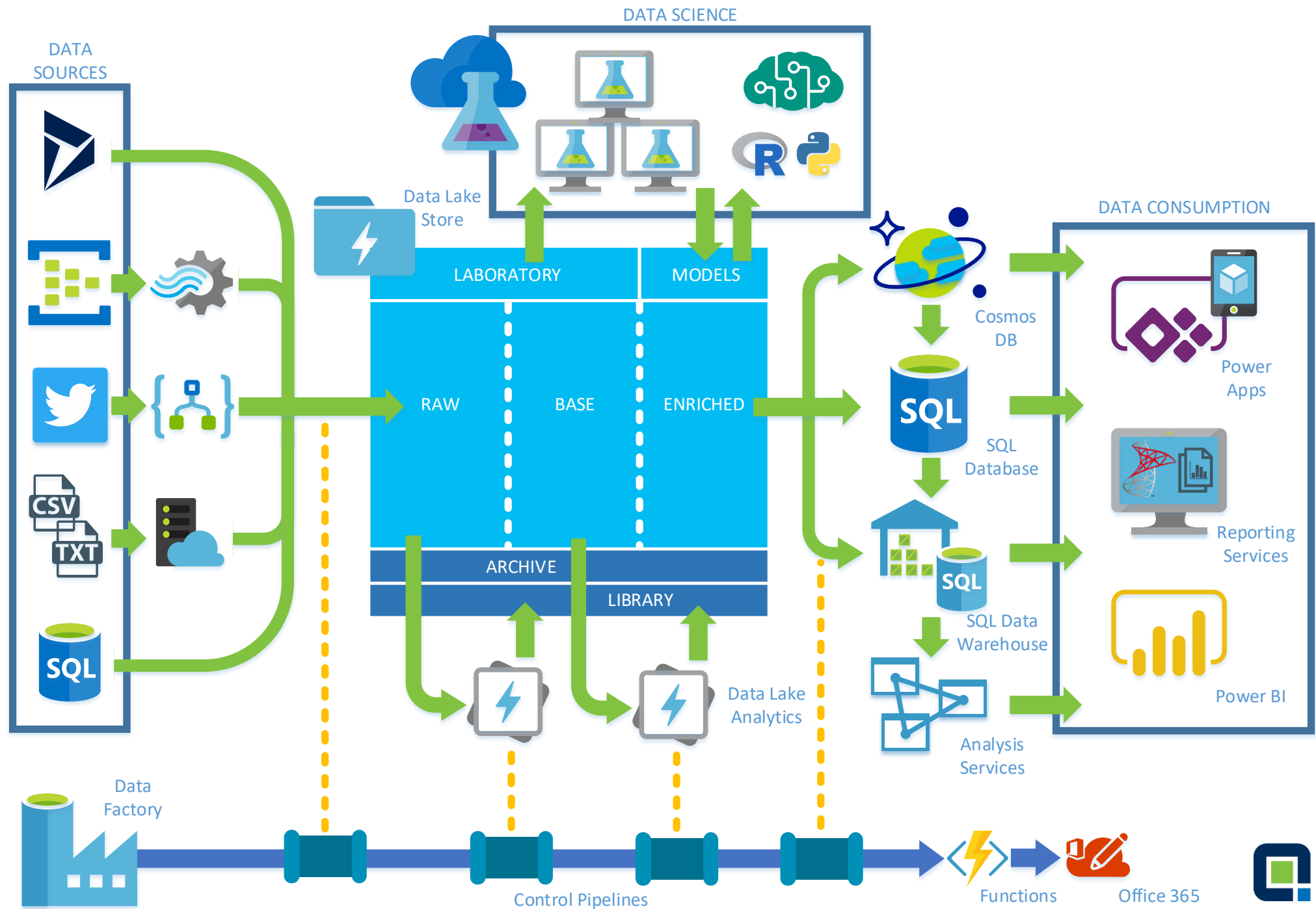
A no-limits data lake to power intelligent action

- ✓ Store and analyse petabyte-size files and trillions of objects
- ✓ Develop massively parallel programs with simplicity
- ✓ Debug and optimise your big data programs with ease
- ✓ Enterprise-grade security, auditing and support
- ✓ Start in seconds, scale instantly and pay per job
- ✓ Built on YARN, designed for the cloud

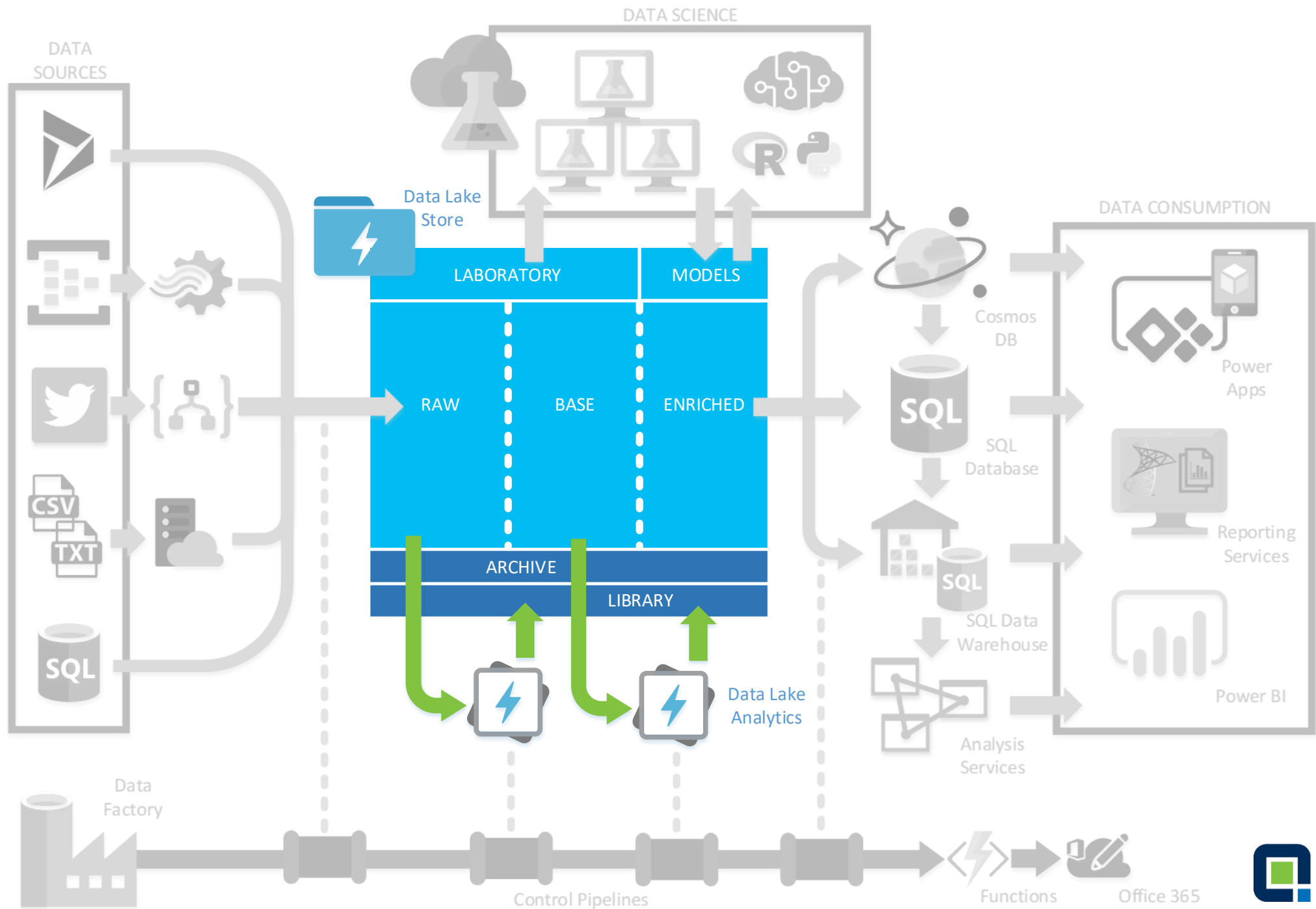
Try it now >

X Geo Redundancy

The Modern Data Warehouse



The Modern Data Warehouse



Agenda

What is Azure
Data Lake?

Storage & Compute

Why use Data
Lake?

The Modern Data
Warehouse

How can we work
with Data Lake?

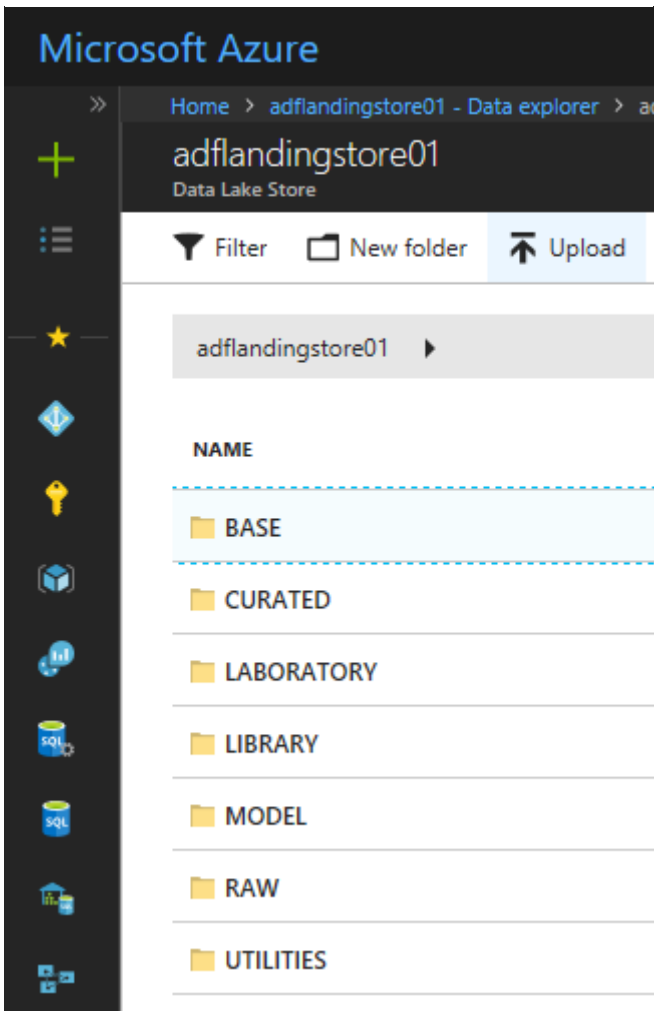
Development &
Management

Working with Azure Data Lake Storage

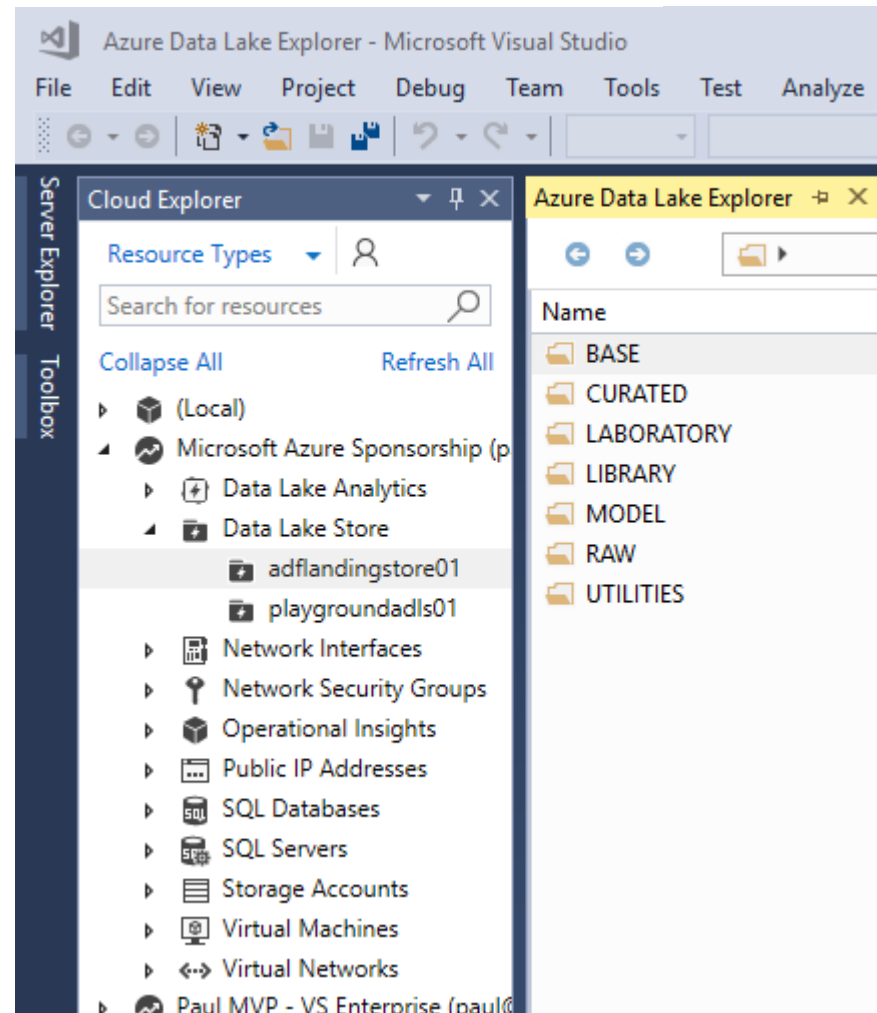


Manual File Uploads

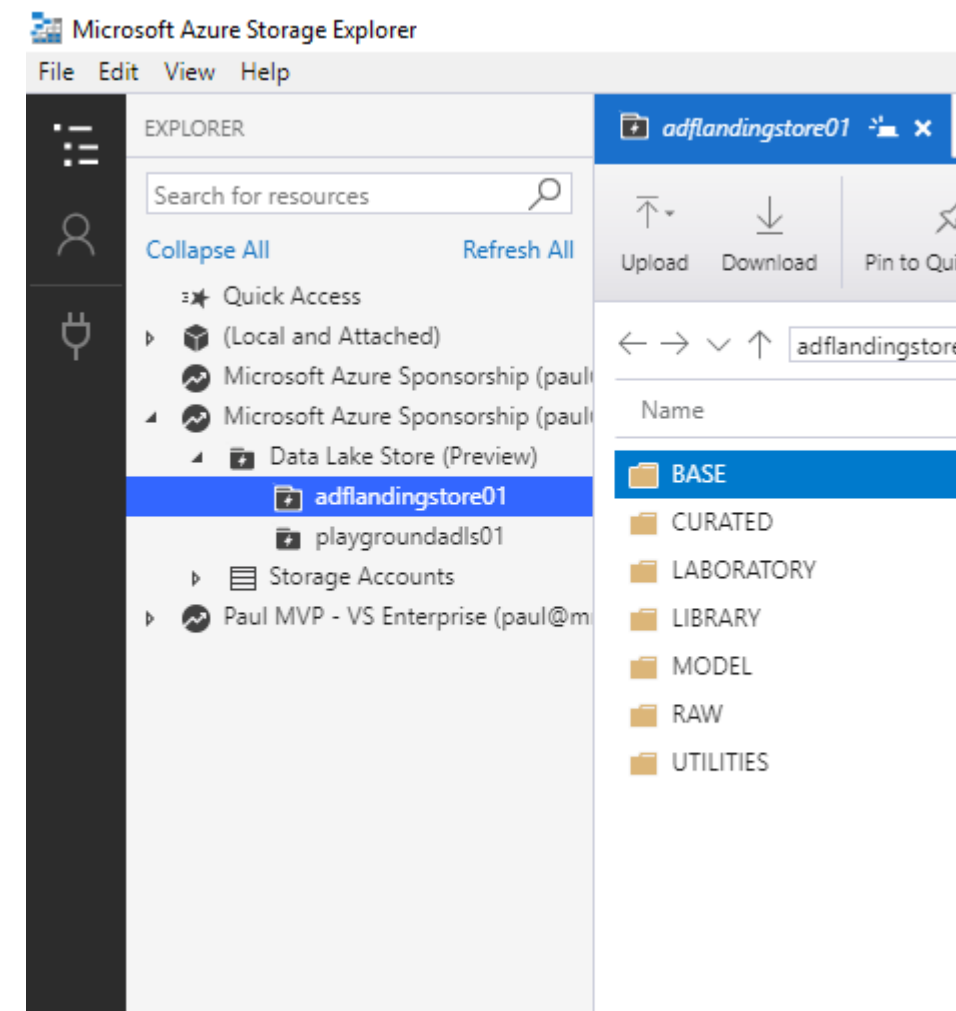
Azure Portal



Visual Studio Cloud Explorer



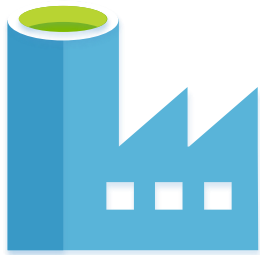
Azure Storage Explorer





Automatic File Uploads

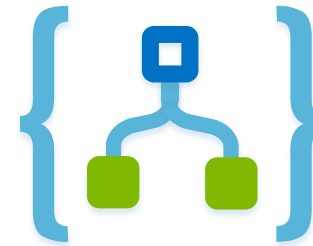
Data Factory



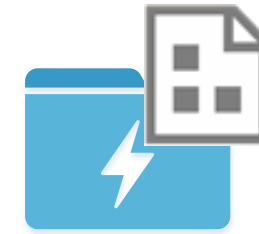
Stream Analytics



Logic Apps



SSIS



.Net SDK



PowerShell



Python



REST API



Working with Azure Data Lake Analytics



Manual U-SQL Job Execution

Azure Portal

The screenshot shows the Microsoft Azure portal interface. The breadcrumb navigation indicates the path: Home > playgroundadla01 > New job. The 'New job' page is displayed, showing the account 'playgroundadla01' and a job name 'New job'. The 'AUs' (Azure Units) are set to 1 Min. The 'More options' section is expanded, showing a list of U-SQL jobs. The first job is selected, and its details are shown in a table.

Job name	Created	Modified	Status	Progress	Errors	Log
1 @Raw = 2 EXTRACT 3 UserId int, 4 Start DateTime, 5 Region string, 6 Query string, 7 Duration int?, 8 Urls string, 9 ClickedUrls string 10 FROM 11 "/Samples/Data/SearchLog.tsv" 12 USING 13 Extractors.Tsv();						

Visual Studio Project

The screenshot shows the Visual Studio IDE with a project named 'USQLSampleApplication'. The 'Server Explorer' pane on the left shows the 'Toolbox' and 'Cloud Explorer' tabs. The 'Toolbox' tab is selected, showing a list of U-SQL jobs. The job 'SearchLog-3b-Adding CSharp Behind.usql' is selected, and its details are shown in a table.

Job name	Created	Modified	Status	Progress	Errors	Log
1 @Raw = 2 @Raw = 3 EXTRACT 4 UserId int, 5 Start DateTime, 6 Region string, 7 Query string, 8 Duration int?, 9 Urls string, 10 ClickedUrls string 11 FROM 12 "/Samples/Data/SearchLog.tsv" 13 USING 14 Extractors.Tsv(); 15 16 17 18 19 20 21						

Visual Studio Code

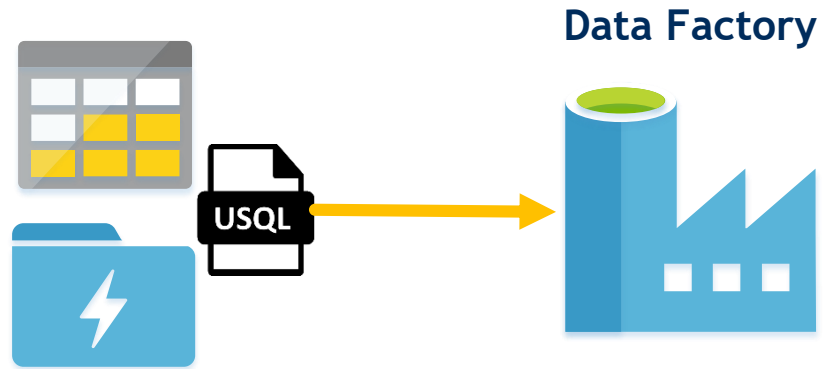
The screenshot shows the Visual Studio Code editor with a file named 'SearchLog-3b-Adding CSharp Behind.usql'. The code is a U-SQL query that extracts data from a TSV file and uses the 'Extractors.Tsv()' function to process it.

```
1
2 @Raw =
3     EXTRACT
4         UserId int,
5         Start DateTime,
6         Region string,
7         Query string,
8         Duration int?,
9         Urls string,
10        ClickedUrls string
11 FROM
12     "/Samples/Data/SearchLog.tsv"
13 USING
14     Extractors.Tsv();
15
16
17
18
19
20
21
22
23
```

Working with Azure Data Lake Analytics



Automatic U-SQL Job Execution

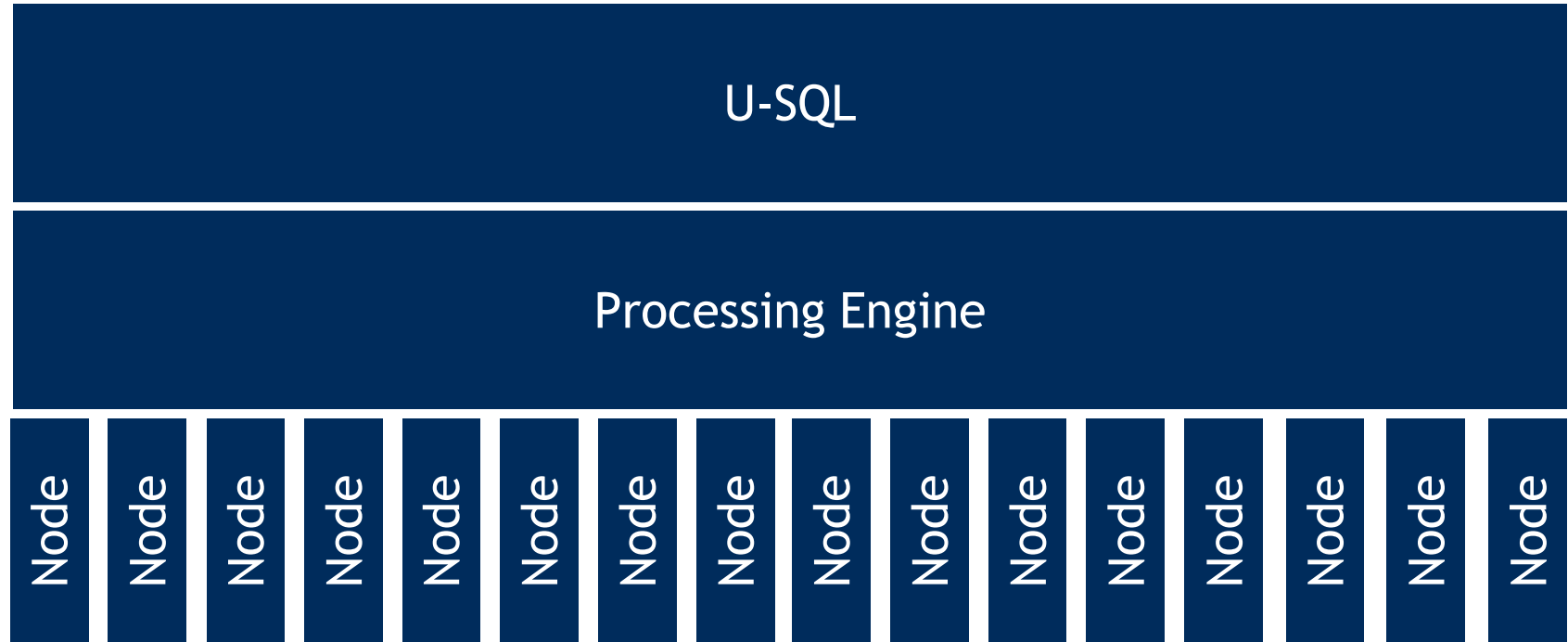


Working with Azure Data Lake Analytics

Job Execution



Analytics



Analytics Units (AU's) – Scale Out Compute

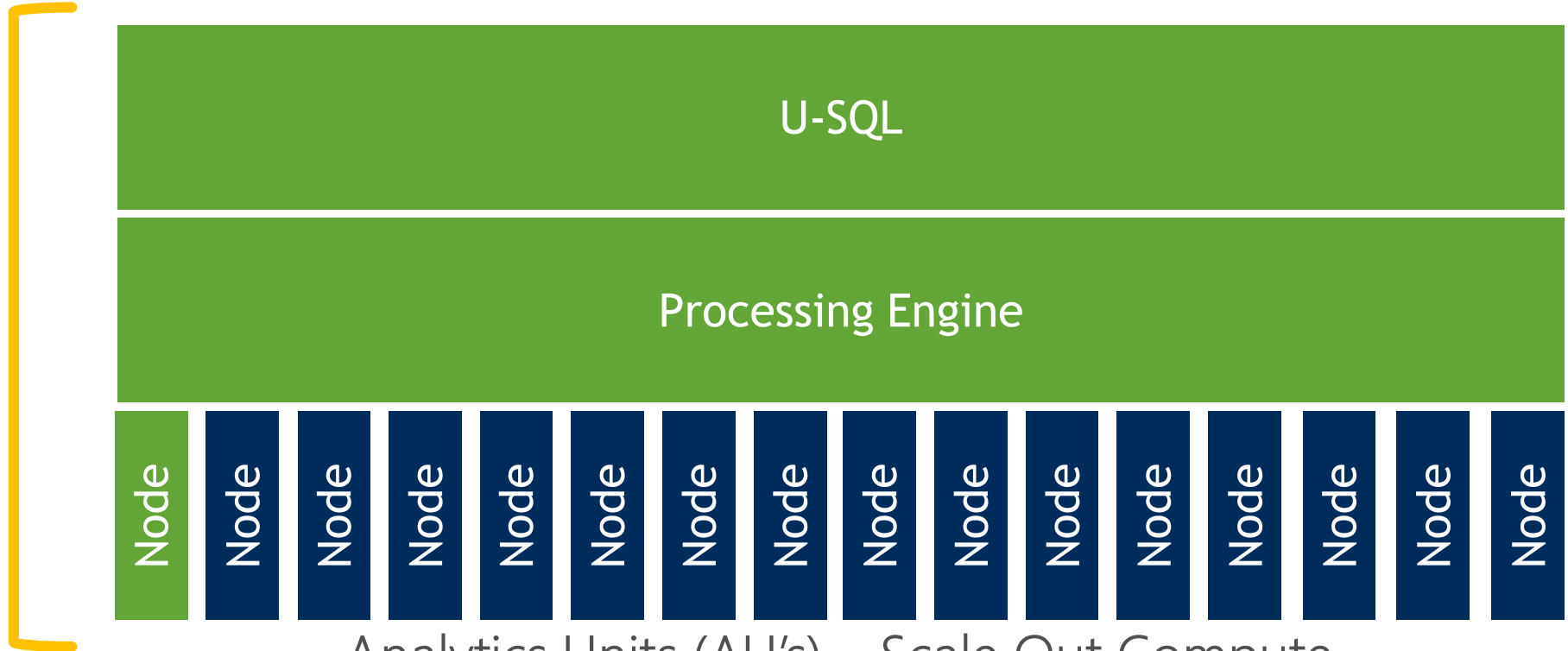
AU/hours

Working with Azure Data Lake Analytics

Job Execution



Analytics



Analytics Units (AU's) – Scale Out Compute

1 x AU/hour = £1.49 *

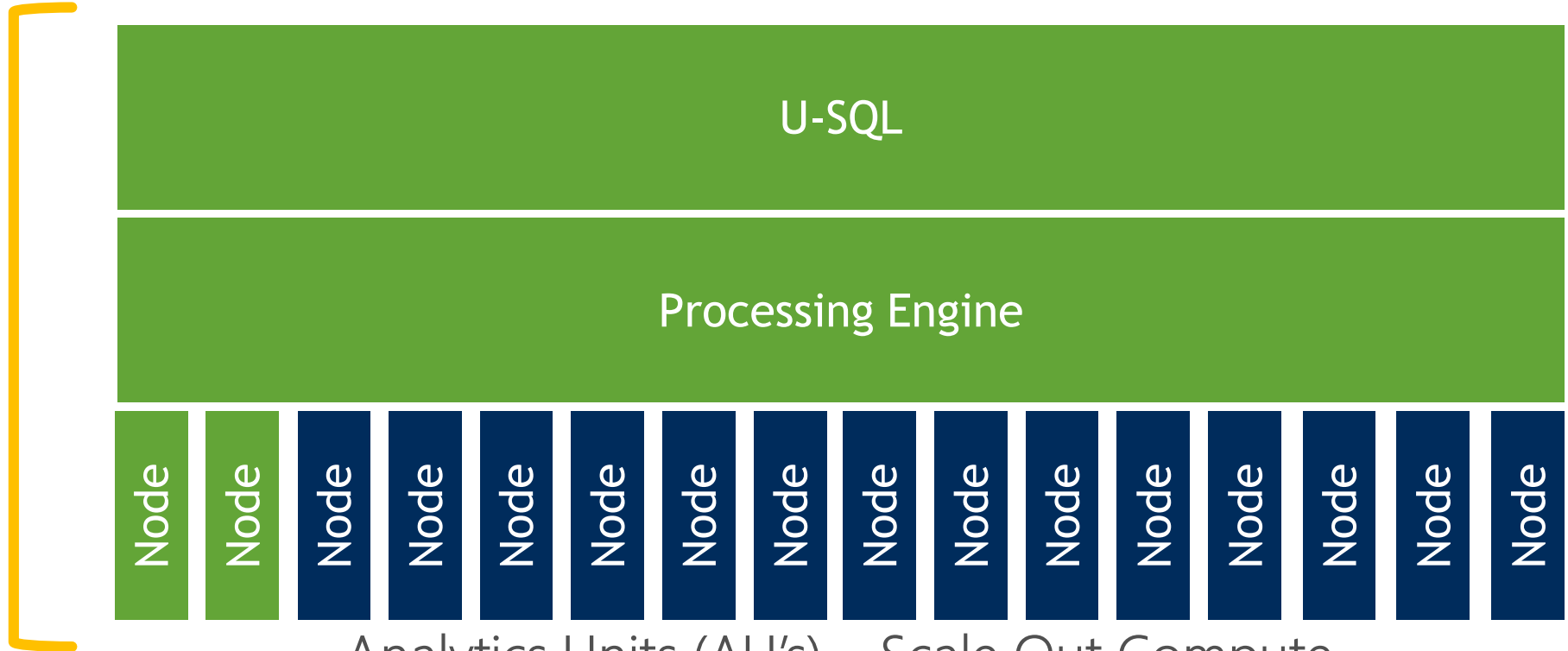
* Price Checked April 2018

Working with Azure Data Lake Analytics

Job Execution



Analytics



Analytics Units (AU's) – Scale Out Compute

2 x AU/hour = £2.98 *

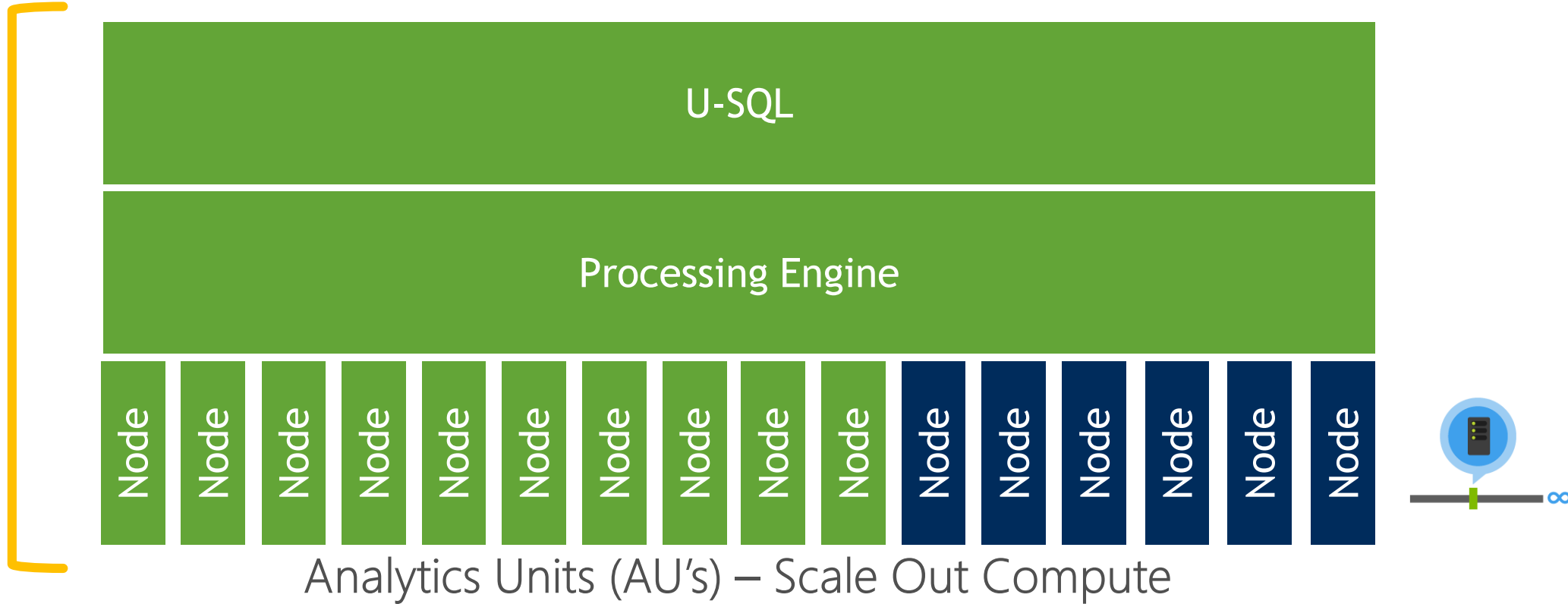
* Price Checked April 2018

Working with Azure Data Lake Analytics

Job Execution



Analytics



10 x AU/hour = £14.90 *

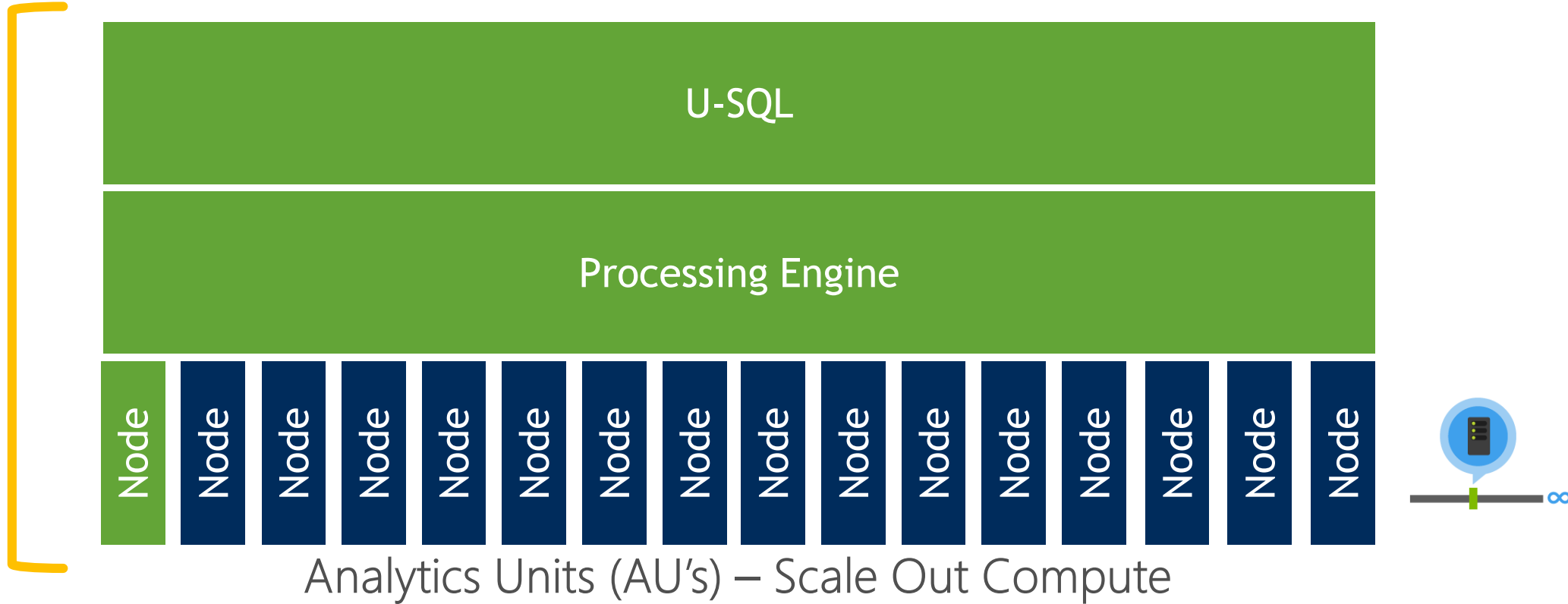
* Price Checked April 2018

Working with Azure Data Lake Analytics

Job Execution



Analytics



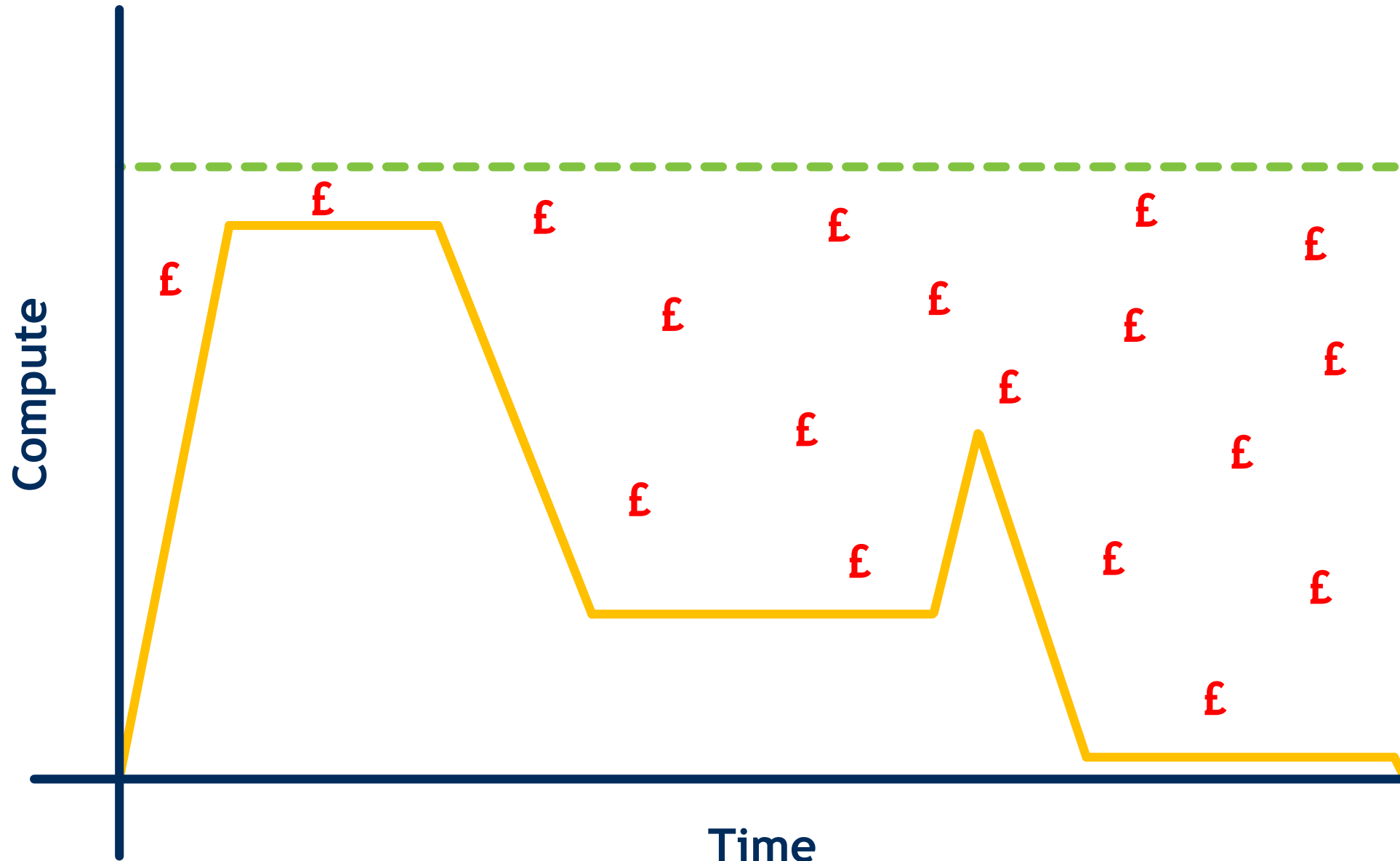
1 minute to complete

1 x AU/hour = ~~£1.49~~ *

£0.02

* Price Checked April 2018

Working with Azure Data Lake Analytics



Agenda for the Day

Module 1

Microsoft Azure

Module 2

Storage
Uploading Data
Data Lake

Module 3

Real-time Data
Streaming
Power BI

Module 4

U-SQL - Data
Transformation
Basics

Module 5

USQL - Advanced
Analytics
Cognitive Services

Module 6

Data Factory
Orchestration
Dynamic Pipelines

Module 7

Data Presentation
& Consumption
Power BI Models

Module 8

Other Services
Q&A