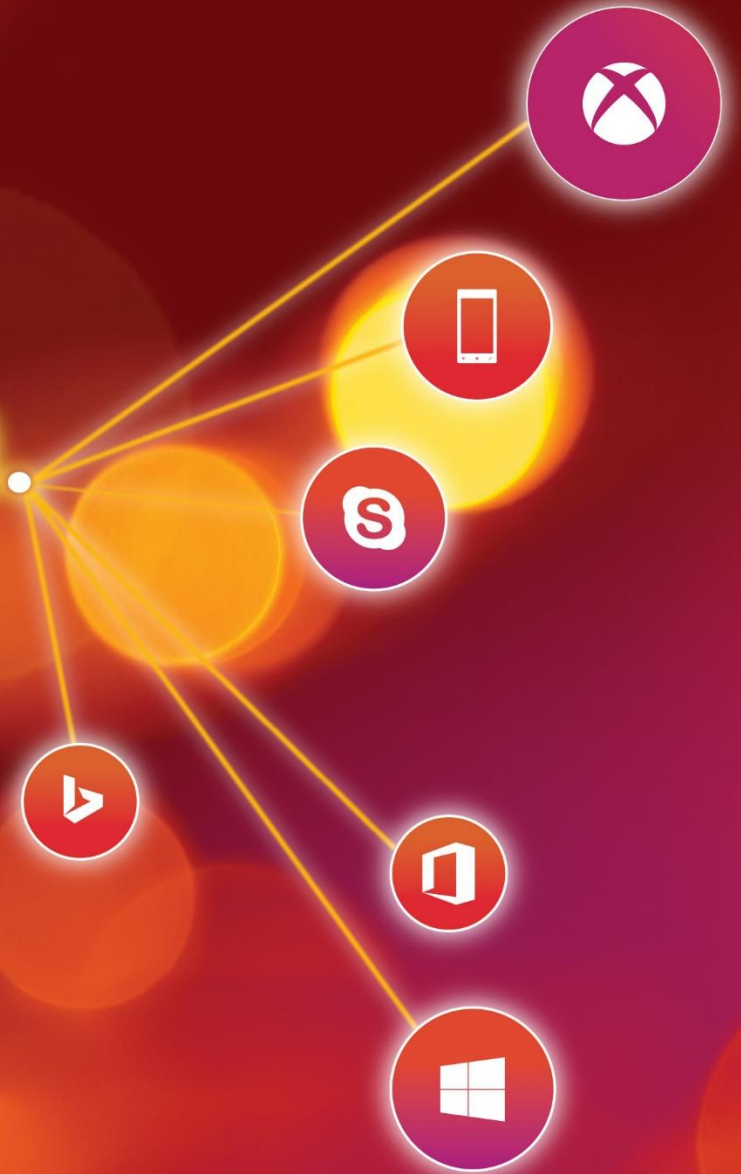


Spark your Innovation.

February 11-13, 2015

Hong Kong Convention and Exhibition Centre

Microsoft®
tech·days
Hong Kong | 2015



SSIS Vs Azure Data Factory

Reza Rad

RADACAD

About Me: Reza Rad

DW/BI Architect, Consultant

SQL Server MVP

Author of SQL Server and BI books

Author of MSBI webcast series on RADACAD

Microsoft Certified Trainer

Microsoft Certified Professional



RADACAD



[@Rad_Reza](https://twitter.com/Rad_Reza)



a.raad.g@gmail.com



www.rad.pasfu.com



www.linkedin.com/in/rezarad

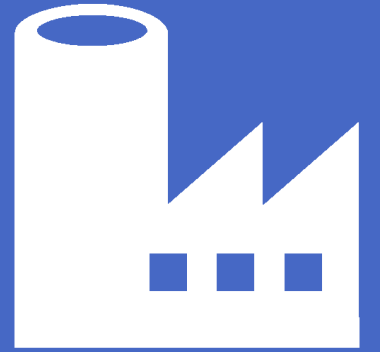
Agenda

Azure Data Factory Overview

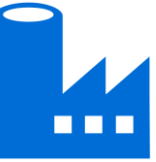
SSIS Overview

Comparison

Azure Data Factory



Azure Data Factory

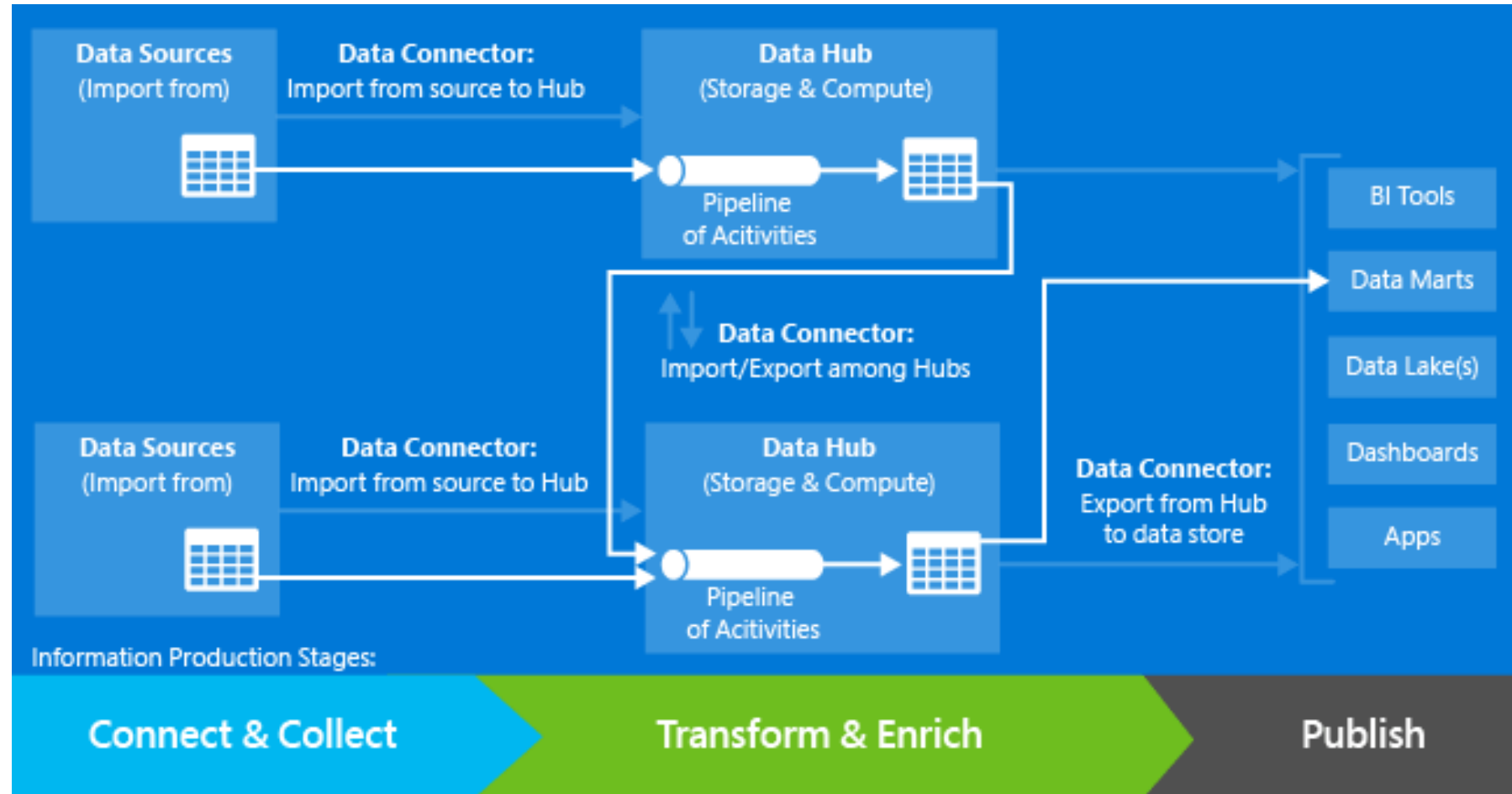


- Fully Managed Service for Composing Data Storages, Processing, and Movement Services into Streamlined, Scalable, and Reliable Data Production Pipelines.

What You can do with Azure Data Factory

- Access to data sources
 - such as SQL Server On premises, SQL Azure, and Azure Blob storage
- Data transformation
 - through Hive, Pig, and C#.
- Monitoring
 - the pipeline of data, validation and execution of scheduled jobs
- Load it into desired Destinations
 - such as SQL Server On premises, SQL Azure, and Azure Blob storage
- And last but not least; This is Cloud based service.

Azure Data Factory Architecture



Administration Panel



Test1DF
DATA FACTORY

...

✦

□

✕

☁

HOME

!

NOTIFICATIONS

🔍

BROWSE

📄

ACTIVE

💰

BILLING

🗑️

Summary

Diagram

2

With errors 0

Quick start

1

ADFTutorialPipeline

Linked services

4

Data Gateways 1

Data Stores 3

...

Operations

Events in the past week
TEST1DF

200

150

100

50

0

ADFTutorialPipeline

1

Alert rules
TEST1DF

0

rules

Linked services
TEST1DF

+

+

Search to filter resources...

LINKED SERVICE	TYPE	UPDATE	STATUS
DATA GATEWAYS			
Reza-Vaio			✓ Good
DATA STORES			
source db	Azure SQL database		
SQLOnPrem	SQL server		✓ Good
TestBlobStore	Azure storage account		
COMPUTE			
Empty			

Main Benefits

- Pay as Usage
- Cloud Based
- HDInsight compatibility

Demo

Azure Data Factory

Date goes here

SSIS 101

SSIS

- Data Transformation and Consolidation Tool
- Part of SQL Server Installation
- Next generation of legacy DTS technology
- Introduced in 2005, Mature, but still new features coming

What you can do with SSIS

- Defining Flow of Activities
- Get data from many types of data sources
- Transform data
- Load data into many types of destinations

Main Benefits

- Rich Development Tool: SSDT
- Many Built-in Transformations

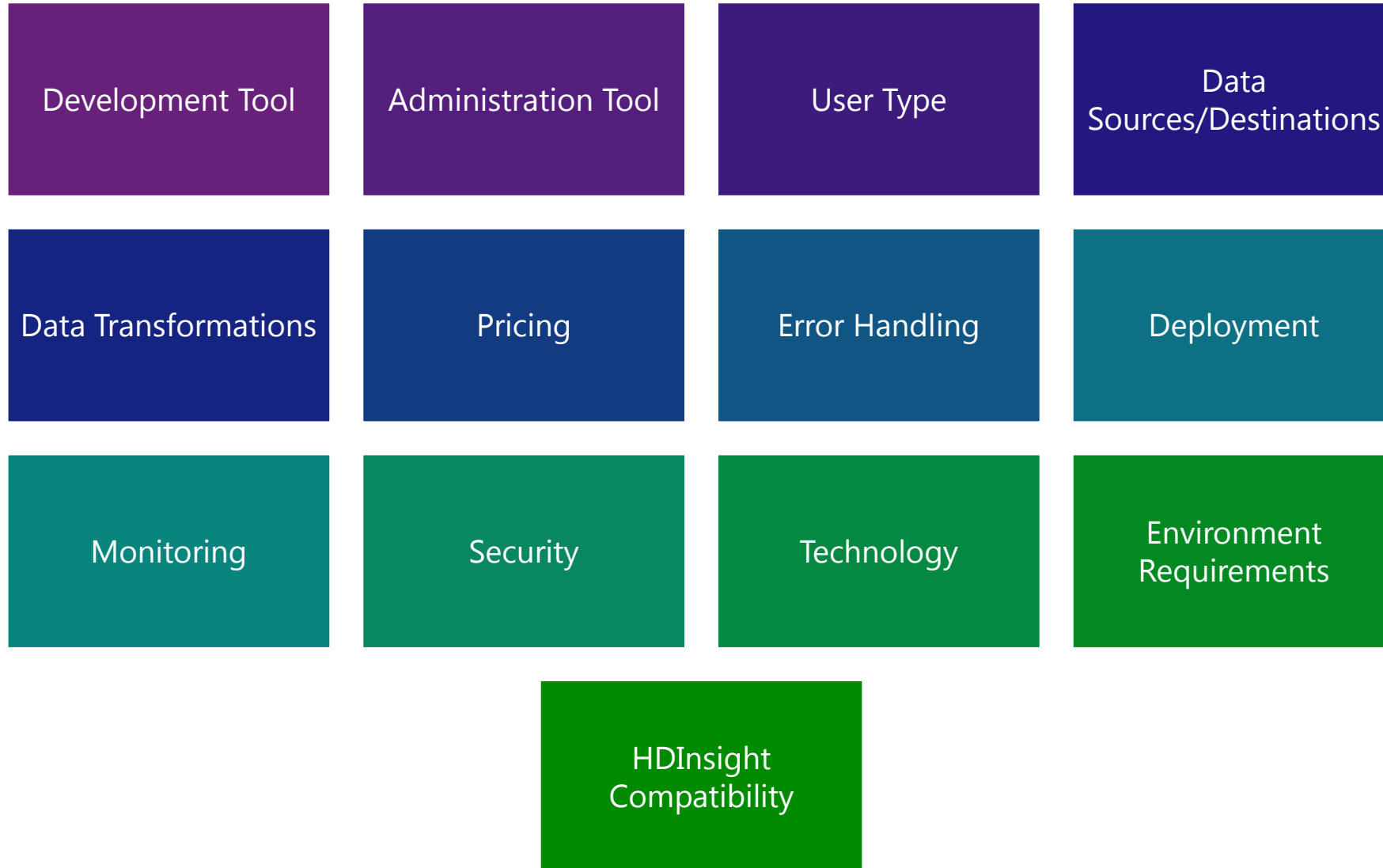
Demo

SSIS

Date goes here

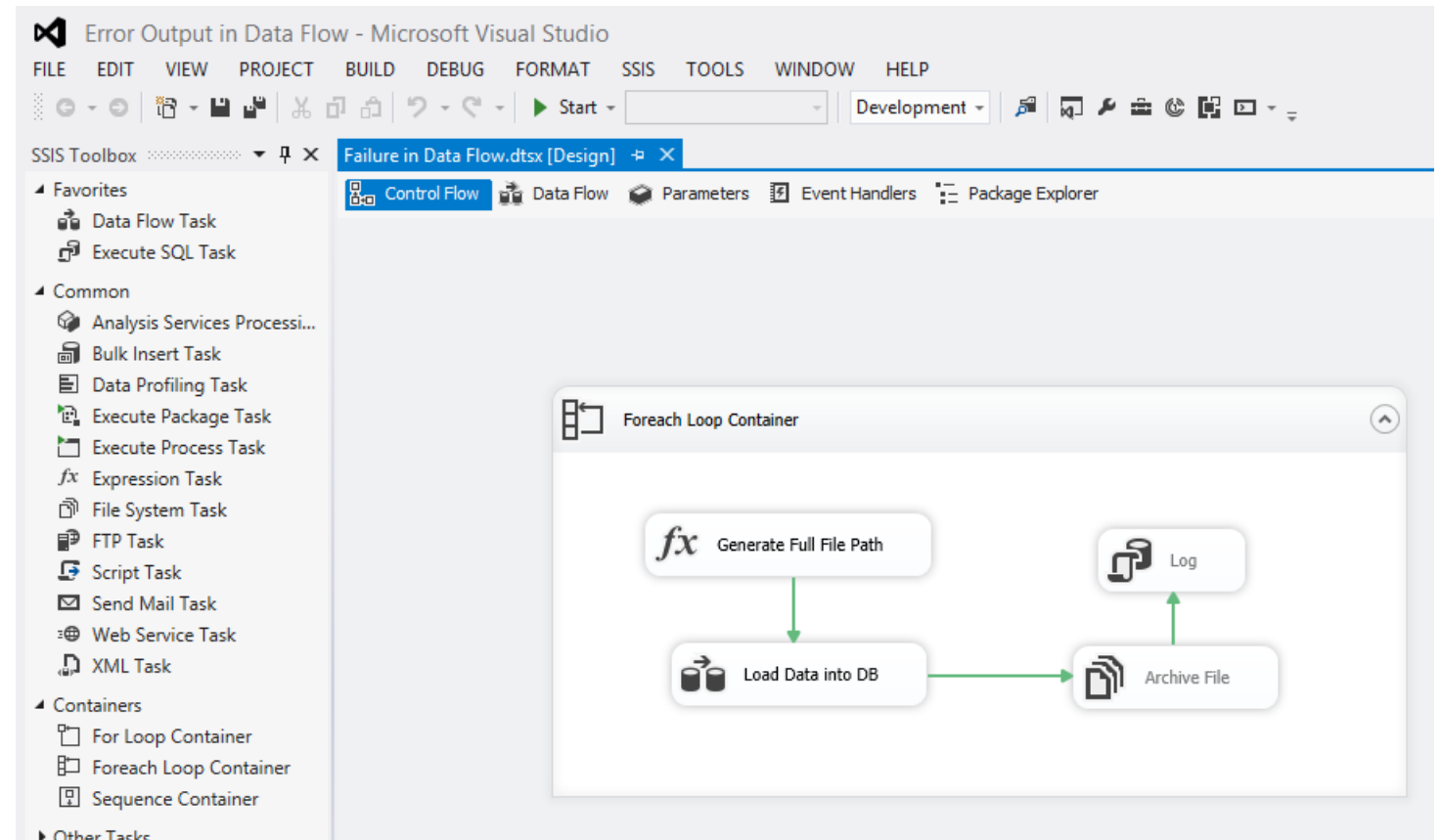
Comparison Factors

Comparison Factors



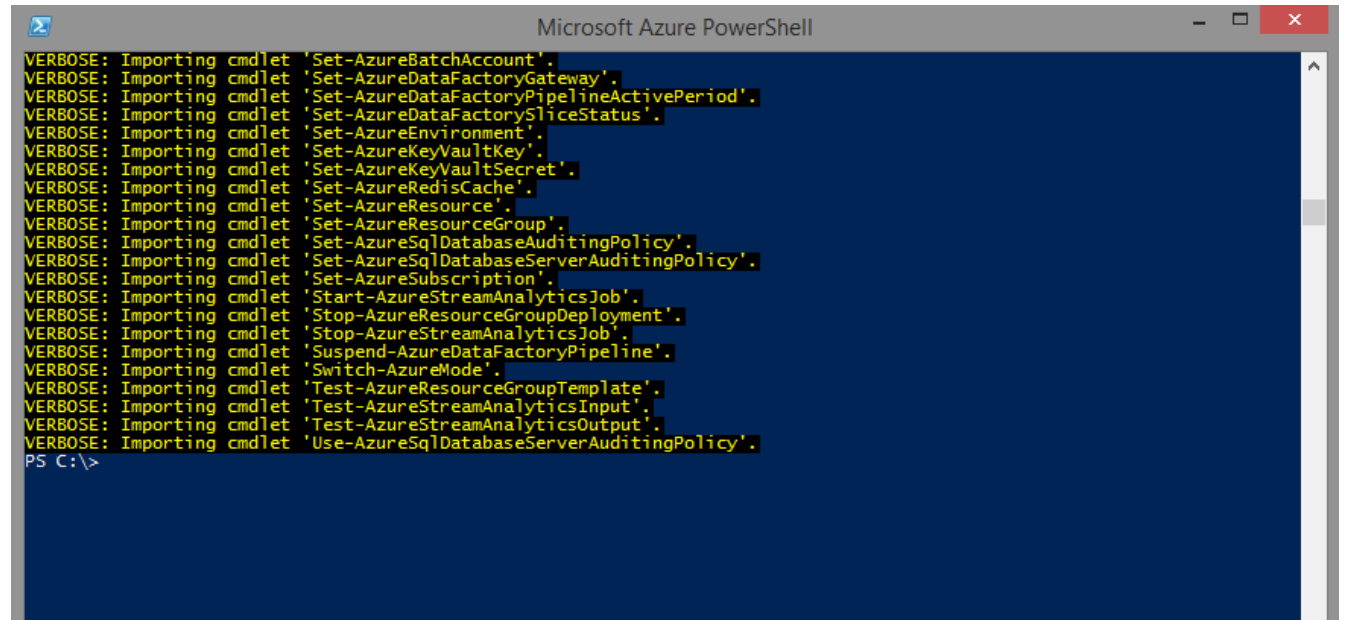
Development Tools: SSIS

- SSDT / BIDS
- Rich and Powerful
- Available Free



Development Tools: ADF

- Azure Portal
- Azure Power Shell
- JSON scripts



```
Microsoft Azure PowerShell

VERBOSE: Importing cmdlet 'Set-AzureBatchAccount'.
VERBOSE: Importing cmdlet 'Set-AzureDataFactoryGateway'.
VERBOSE: Importing cmdlet 'Set-AzureDataFactoryPipelineActivePeriod'.
VERBOSE: Importing cmdlet 'Set-AzureDataFactorySliceStatus'.
VERBOSE: Importing cmdlet 'Set-AzureEnvironment'.
VERBOSE: Importing cmdlet 'Set-AzureKeyVaultKey'.
VERBOSE: Importing cmdlet 'Set-AzureKeyVaultSecret'.
VERBOSE: Importing cmdlet 'Set-AzureRedisCache'.
VERBOSE: Importing cmdlet 'Set-AzureResource'.
VERBOSE: Importing cmdlet 'Set-AzureResourceGroup'.
VERBOSE: Importing cmdlet 'Set-AzureSqlDatabaseAuditingPolicy'.
VERBOSE: Importing cmdlet 'Set-AzureSqlDatabaseServerAuditingPolicy'.
VERBOSE: Importing cmdlet 'Set-AzureSubscription'.
VERBOSE: Importing cmdlet 'Start-AzureStreamAnalyticsJob'.
VERBOSE: Importing cmdlet 'Stop-AzureResourceGroupDeployment'.
VERBOSE: Importing cmdlet 'Stop-AzureStreamAnalyticsJob'.
VERBOSE: Importing cmdlet 'Suspend-AzureDataFactoryPipeline'.
VERBOSE: Importing cmdlet 'Switch-AzureMode'.
VERBOSE: Importing cmdlet 'Test-AzureResourceGroupTemplate'.
VERBOSE: Importing cmdlet 'Test-AzureStreamAnalyticsInput'.
VERBOSE: Importing cmdlet 'Test-AzureStreamAnalyticsOutput'.
VERBOSE: Importing cmdlet 'Use-AzureSqlDatabaseServerAuditingPolicy'.

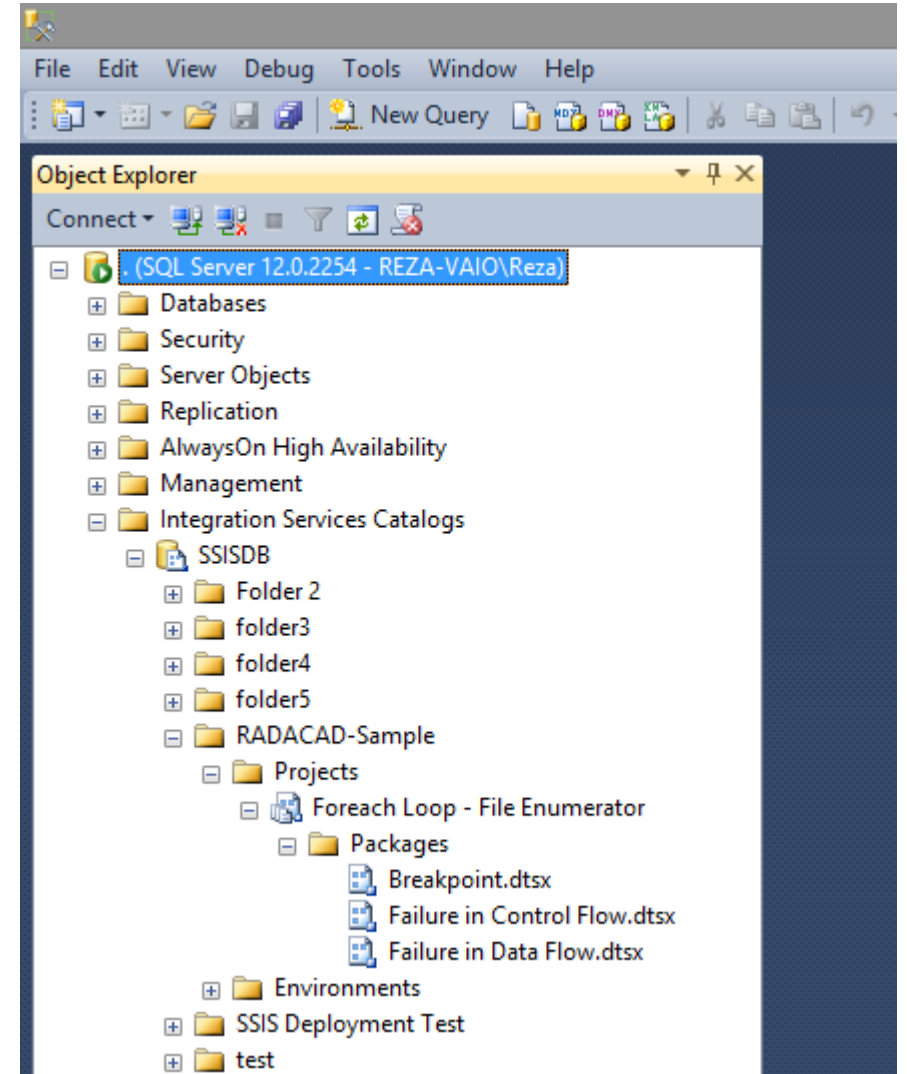
PS C:\>
```

Comparison: Development Tool

	SSIS	ADF
Standalone Tool	Yes	Yes
Powerful GUI	Yes	No (Power Shell)
Availability	Free	Free
Prerequisite	SQL Server License	Azure Subscription

- SSMS
- Available Free (New Versions)

- SSMS
- Available Free (New Versions)

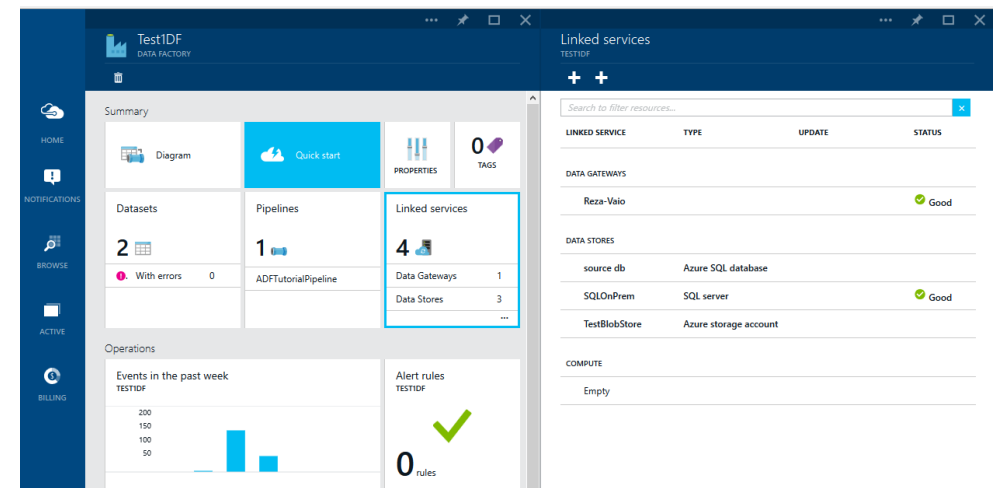


Administration Tool: ADF

- Azure Portal
- Available under Azure subscription
- Azure Power Shell; Scheduling Pipelines...
- Azure Data Factory Cmdlets

```
Microsoft Azure PowerShell

VERBOSE: Importing cmdlet 'Set-AzureBatchAccount'.
VERBOSE: Importing cmdlet 'Set-AzureDataFactoryGateway'.
VERBOSE: Importing cmdlet 'Set-AzureDataFactoryPipelineActivePeriod'.
VERBOSE: Importing cmdlet 'Set-AzureDataFactorySliceStatus'.
VERBOSE: Importing cmdlet 'Set-AzureEnvironment'.
VERBOSE: Importing cmdlet 'Set-AzureKeyVaultKey'.
VERBOSE: Importing cmdlet 'Set-AzureKeyVaultSecret'.
VERBOSE: Importing cmdlet 'Set-AzureRedisCache'.
VERBOSE: Importing cmdlet 'Set-AzureResource'.
VERBOSE: Importing cmdlet 'Set-AzureResourceGroup'.
VERBOSE: Importing cmdlet 'Set-AzureSqlDatabaseAuditingPolicy'.
VERBOSE: Importing cmdlet 'Set-AzureSqlDatabaseServerAuditingPolicy'.
VERBOSE: Importing cmdlet 'Set-AzureSubscription'.
VERBOSE: Importing cmdlet 'Start-AzureStreamAnalyticsJob'.
VERBOSE: Importing cmdlet 'Stop-AzureResourceGroupDeployment'.
VERBOSE: Importing cmdlet 'Stop-AzureStreamAnalyticsJob'.
VERBOSE: Importing cmdlet 'Suspend-AzureDataFactoryPipeline'.
VERBOSE: Importing cmdlet 'Switch-AzureNode'.
VERBOSE: Importing cmdlet 'Test-AzureResourceGroupTemplate'.
VERBOSE: Importing cmdlet 'Test-AzureStreamAnalyticsInput'.
VERBOSE: Importing cmdlet 'Test-AzureStreamAnalyticsOutput'.
VERBOSE: Importing cmdlet 'Use-AzureSqlDatabaseServerAuditingPolicy'.
PS C:\>
```



Comparison: Administration Tool

	SSIS	ADF
Standalone Tool	Yes	Yes
Powerful GUI	Yes	No (Power Shell)
Availability	Free	Under subscription
Prerequisite	SQL Server License	Azure Subscription

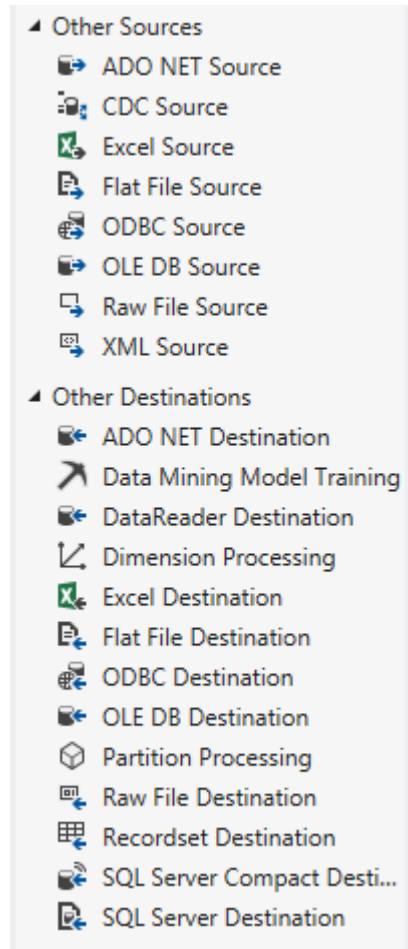
User Type

- Both SSIS and ADF are Developer Tools and Services



Data Src/Dest: SSIS

■ Wide Range



SQL Server

Oracle

SAP

Salesforce

Access

Sybase

Informix

Teradata

MySQL

PostgreSQL

FoxPro

CSVs

Web Services

Many other Data Sources and Destinations types

Data Src/Dest: ADF

- Three Data Store Types Supported so far



Azure Storage



Azure SQL Database



SQL Server

Comparison: Data Src/Dest

	SSIS	ADF
SQL Server On Premises	Yes	Yes
SQL Azure	Yes	Yes
Azure Storage	Yes	Yes
Oracle	Yes	
SAP	Yes	
MySQL	Yes	
DB2	Yes	
CSVs	Yes	Yes (When located on Azure Storage)
Web Services	Yes	

Data Transformations: SSIS

- Wide range of Built-in Transformation
- Ability to extend more with Script Component

Merge Join

Aggregate

Union

Lookup

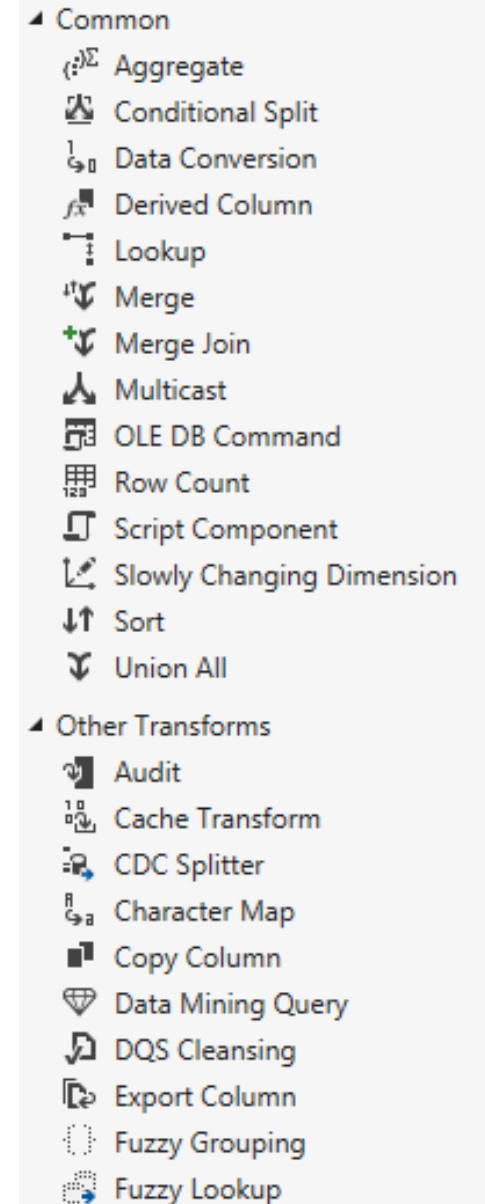
Fuzzy Lookup/Grouping

Conditional Split

Derived Column

OLE DB Command

Many other Data Transformations



Data Transformations: ADF

- Copy Activity
- Hive and Pig Activity
 - Support wide range of transformations through HDInsight
- Custom C# Activity
 - Write your own transformation with C#

Comparison: Data Transformation

- More Built-in Transformation will come for ADF

	SSIS	ADF
Copy	Yes	Yes
C# Custom Transformations	Yes	Yes
Pig and Hive	With Scripting	Yes
Built-in Data Transformations	Yes	

Pricing: SSIS

- Basic SSIS for Free (Import and Export Wizard)
 - Express Edition
- Standard SSIS features
 - Standard and BI Edition
- Advanced SSIS Features
 - Such as CDC and advanced adapters
 - Enterprise Edition



Pricing: ADF

- Pay for Usage

Low Frequency	\$0.60	\$0.48	\$1.50	\$1.20
High Frequency	\$1.00	\$0.80	\$2.50	\$2.00
	0-100 activities	100+ activities	0-100 activities	100+ activities
	Cloud		On Premises	



Comparison: Pricing

	SSIS	ADF
Licensing	Yes	
Pay for Features	Yes	
Pay for Usage		Yes



Environment: SSIS

- You need usually a physical good spec Server
 - Example SSIS practice for 1TB data movement in 30 Minutes!
- Hardware Requirements
- Software Requirements
- Administration Efforts



Environment: ADF

- Azure will take care of Environment
- Cloud based benefits
- HDInsight for ADF is also supported
- No Hardware, Software Administration efforts

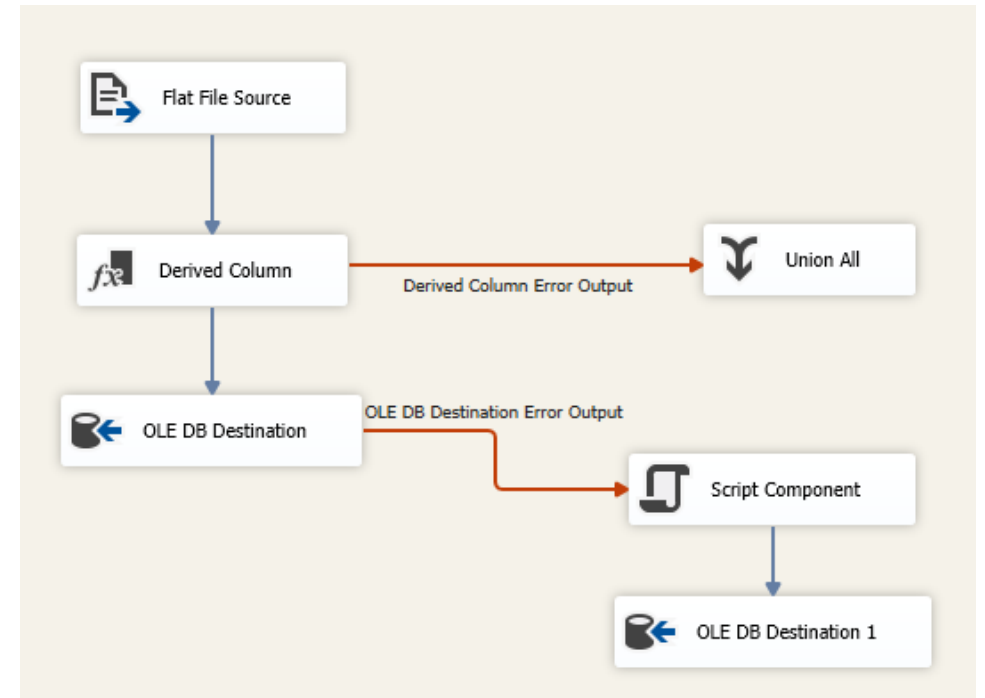


Comparison: Environment

	SSIS	ADF
Hardware Setup	Yes	
Software Setup	Yes	
Administration Costs	Yes	
Azure Environment Usage		Yes

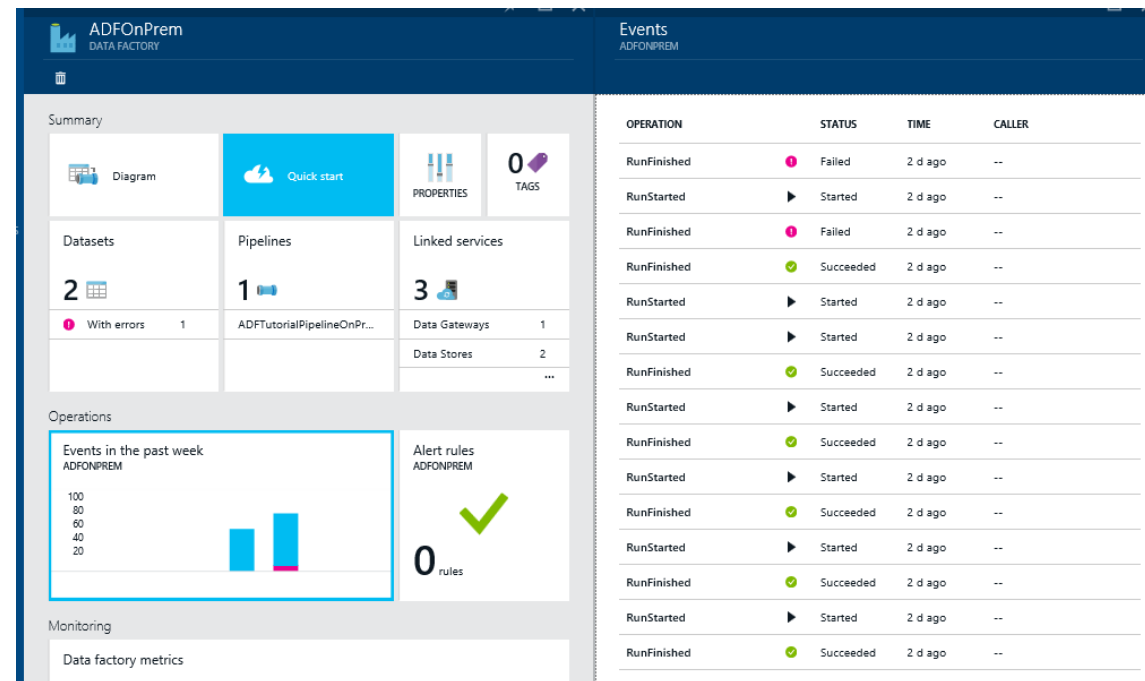
Error Handling: SSIS

- Error Handling through Event Handlers (OnError)
- Error Output in Data Flow
- Failure Precedence Constraint



- Error Message Logging
- Alert Rules
- No Error Handling Events

- Error Message Logging
- Alert Rules
- No Error Handling Events

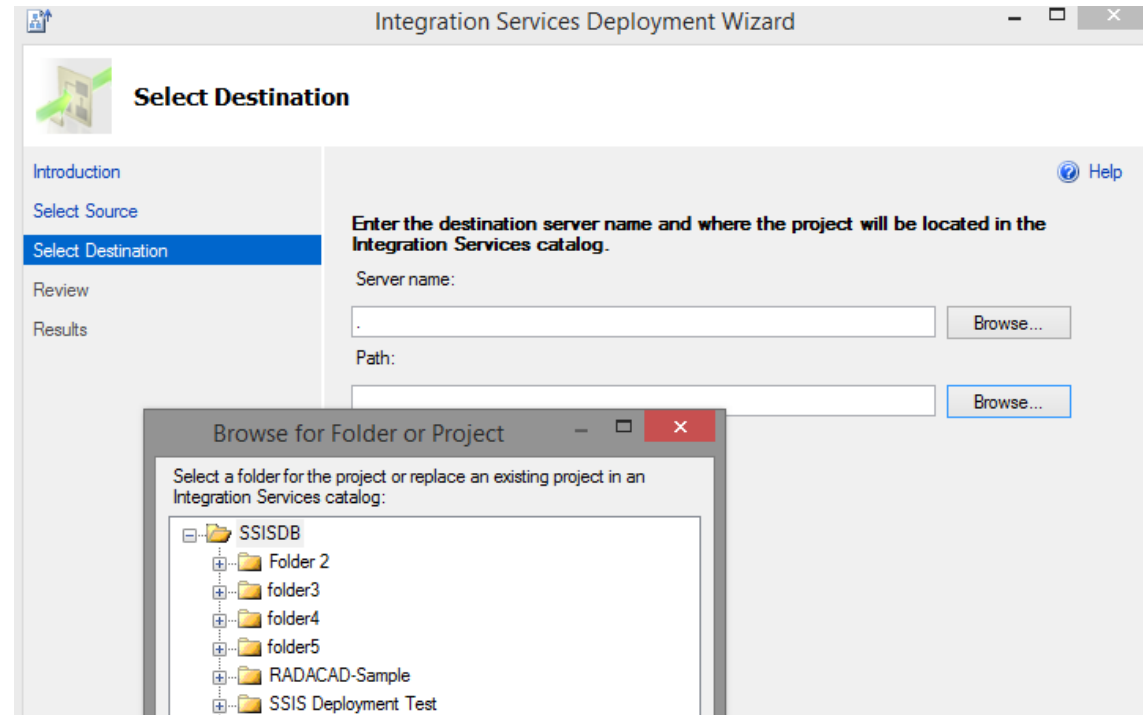


Comparison: Error Handling

	SSIS	ADF
Alerts	Yes	Yes
Error Logging	Yes	Yes
Error Handling	Yes	

Deployment: SSIS

- Deployment Wizard
- DTSPAC files to export



Deployment: ADF

- Power Shell scripts
- All steps can be automated through a Power Shell Script



Comparison: Deployment

	SSIS	ADF
Deployment GUI	Yes	
Deployment Scripts	Yes	Yes
Corporate Deployment Scenario with automated scripting	Yes	Yes


Monitoring: SSIS

- SSIS Logging
- SSIS Catalog Reports

All Executions

on W7X64S12ENT at 8/12/2012 8:22:37 PM

This report provides information about the Integration Services package executions that have been performed on the connected SQL Server instance.

 **Filter:** Start time range: 8/6/2012 - 8/12/2012; Status: All; (4 more)

Execution Information

0
Failed

0
Running

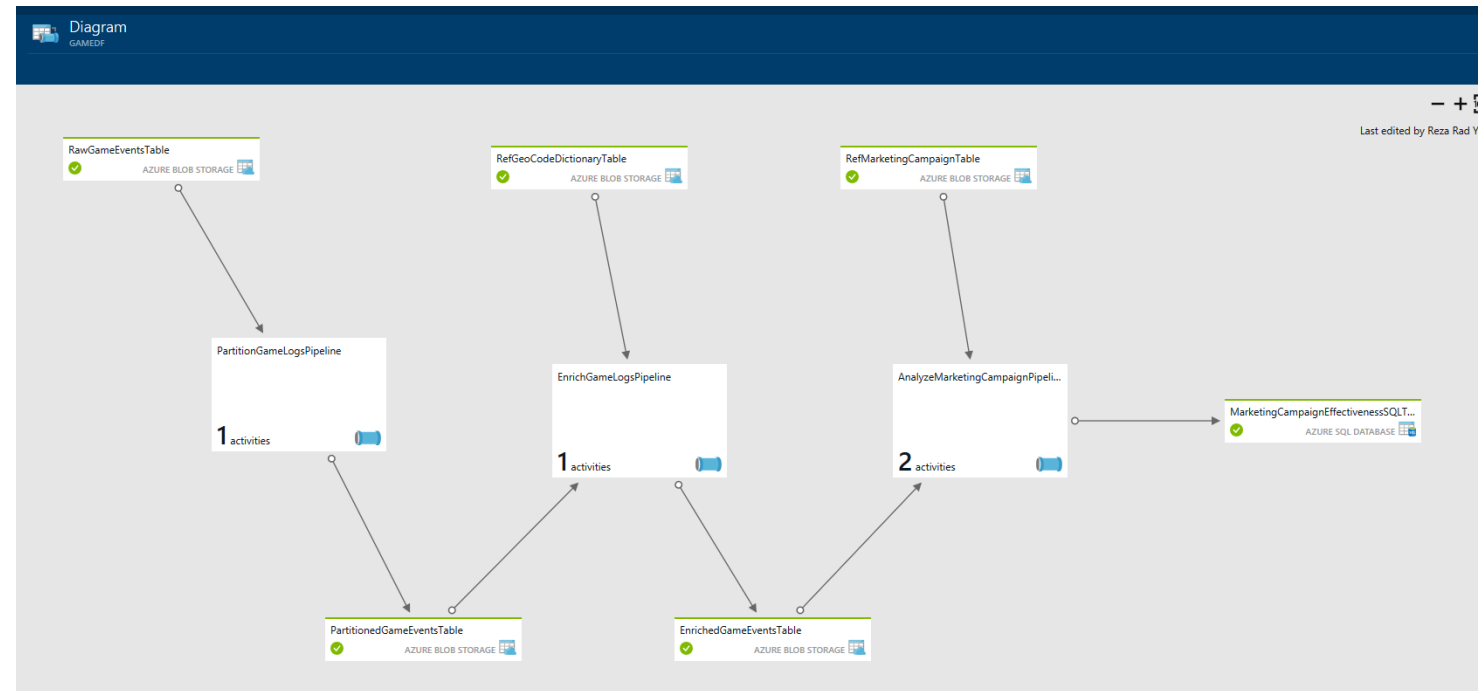
5
Succeeded

0
Others

ID	Status	Report			Folder Name	Project Name	Package Name
20036	Succeeded	Overview	All Messages	Execution Performance	DWBIVC	R01_Parameters	P1_Master.dtsx
20035	Succeeded	Overview	All Messages	Execution Performance	DWBIVC	R01_Parameters	P1_Master.dtsx
20033	Succeeded	Overview	All Messages	Execution Performance	Test	R01_Parameters	P1_Master.dtsx
20030	Succeeded	Overview	All Messages	Execution Performance	Test	R01_Parameters	P1_Master.dtsx
20029	Succeeded	Overview	All Messages	Execution Performance	Test	R01_Parameters	P1_Master.dtsx

Monitoring: ADF

- Diagram View
- Drill through monitoring features
- Powerful GUI
- Data Slice execution
- Data Lineage
 - Very Important Feature



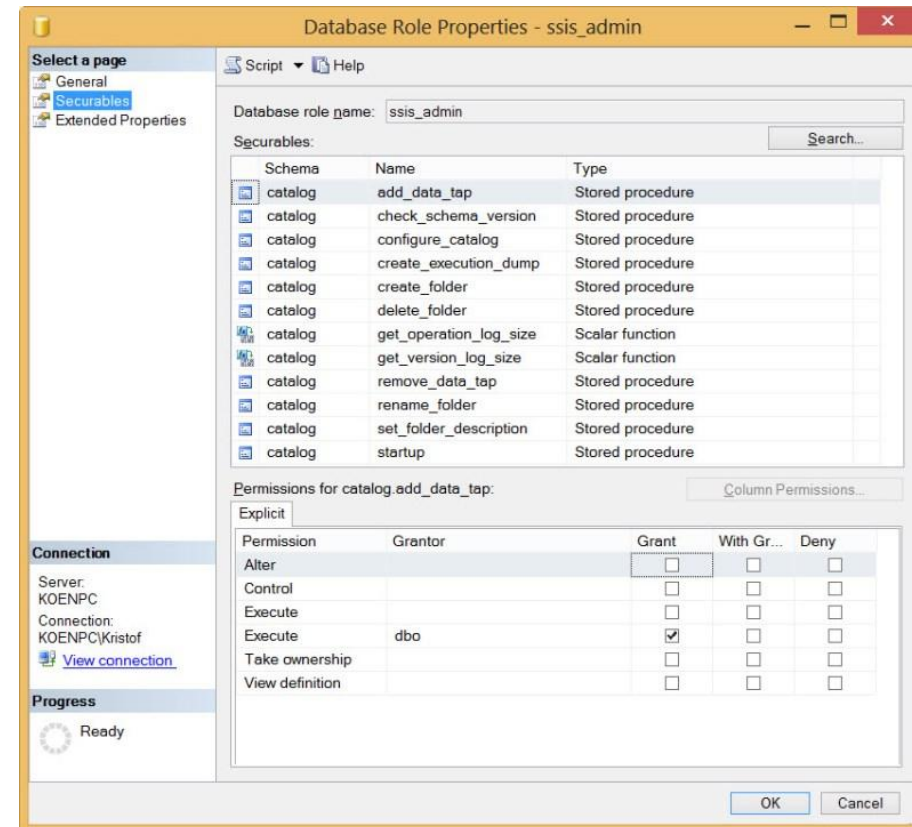
Comparison: Monitoring

	SSIS	ADF
Monitoring GUI	Yes	Yes
Drill through Monitoring	Yes	Yes
Data Slice		Yes
Data Lineage		Yes








Security: SSIS

- Role Based
- Roles for Deploy, Execute, Monitor...



Security: ADF

- Role Based
- Roles:
 - Owner
 - Contributor
 - Reader
 - Data Factory Contributor
 - User Access Administrator

Roles		
TESTIDF		
	OWNER ⓘ 0 users, 1 group	>
	CONTRIBUTOR ⓘ 0 users, 0 groups	>
	READER ⓘ 0 users, 0 groups	>
	DATA FACTORY CONTRIBUTOR ⓘ 0 users, 0 groups	>
	USER ACCESS ADMINISTRATOR ⓘ 0 users, 0 groups	>

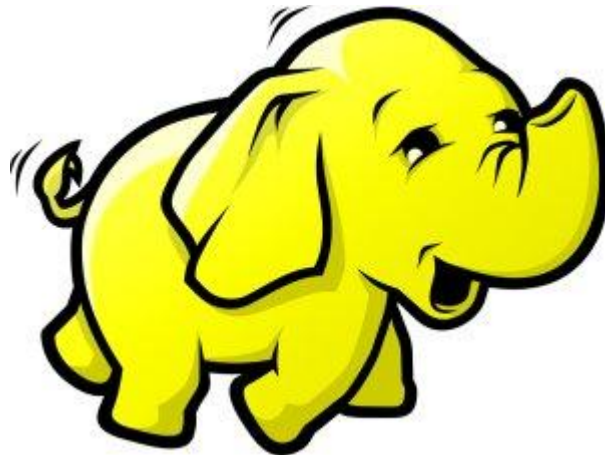
Comparison: Security

	SSIS	ADF
Role Based Security	Yes	Yes



HDInsight Compatibility

- Azure Data Factory can work with existing HDInsight Clusters
- ADF Can create HDInsight cluster on demand
- ADF HDInsight Activity run Pig and Hive scripts



Comparison Summary

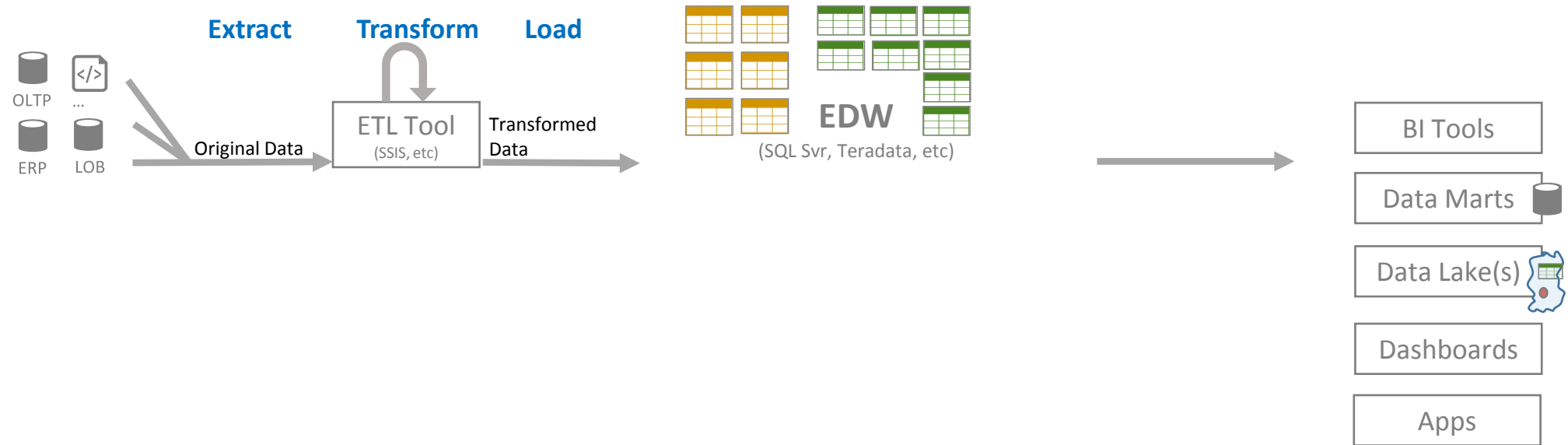
	SSIS	ADF
Development Tool	Yes	Yes (More Scripting Required)
Administration Tool	Yes	Yes (More Scripting Required)
User Type	Developer	Developer
Data Source and Destinations	Wide Range	Azure Storage, SQL Azure, SQL Server On Premises
Data Transformations	Many	Custom Code required
Environment	Setup Required	Azure Supported
Pricing	Pay for Features	Pay for Usage
Environmental and Administration Costs	Yes	
Error Handling	Event Handlers, Alerts	Alerts
Logging	Yes	Yes
Deployment	Yes	Yes
Monitoring	Yes	Yes
Data Lineage		Yes
Security	Role Based	Role Based
HDInsight Compatibility	Normal	High

Hybrid

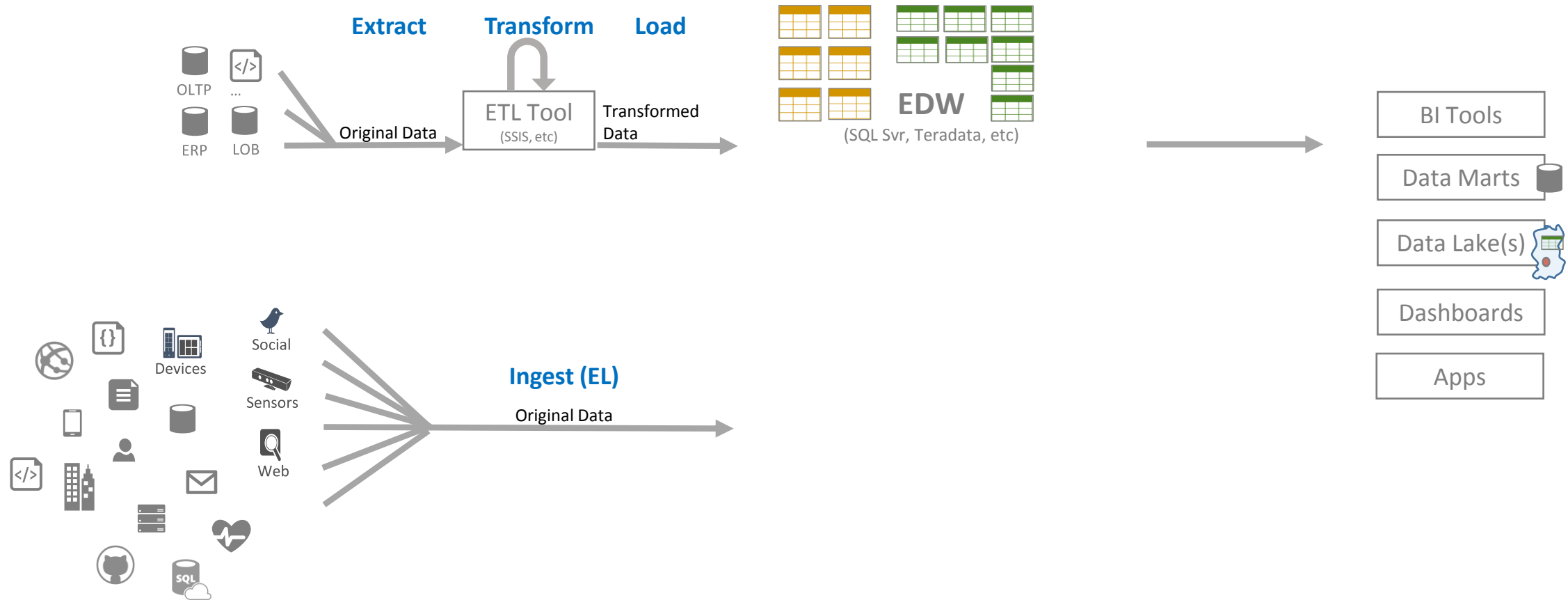
Purpose

- These tools/services didn't built for the same purpose
- They are complement of each other

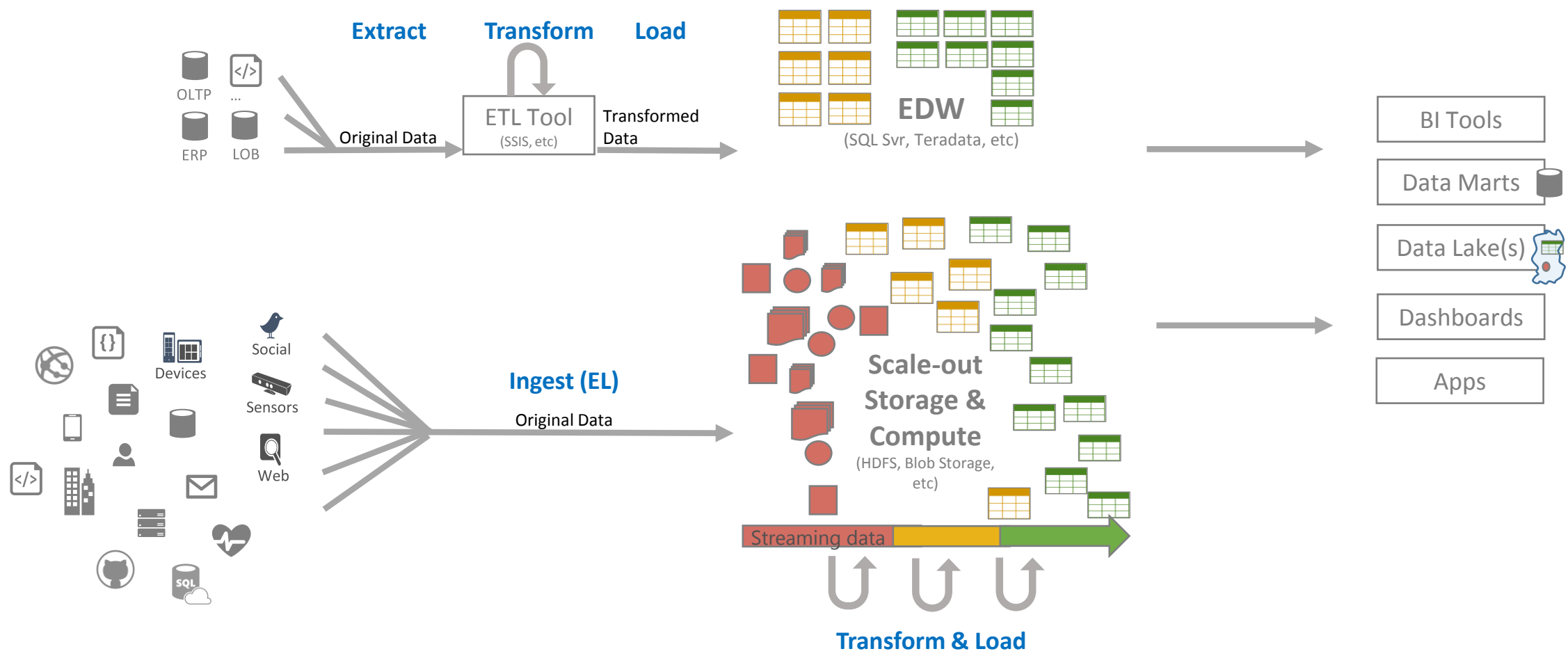
Evolving Approaches to Analytics



Evolving Approaches to Analytics



Evolving Approaches to Analytics

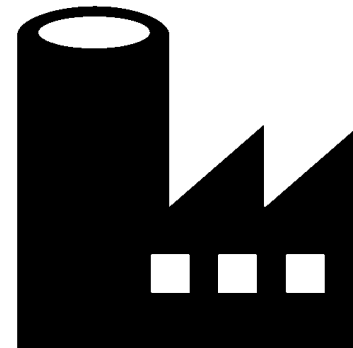


Get Benefits of Both

- Hybrid SSIS and ADF
- Cloud based data movement, Computing and monitoring
- Rich On Premises Data Transformations



Microsoft®
SQL Server®
Integration Services



Summary

- If you are dealing with Big Data
 - If source or destination of data is on cloud
 - If you are dealing with Azure Storage
 - If you want to cut down environmental costs
 - If you want to cut down administration costs
 - If you want to cut down Pricing
 - If Azure is one side of the data
- Then ADF Can be good response
-
- But always Consider Hybrid Scenarios

Summary

- ADF comes as a Complementary service



- Not to Compete with SSIS



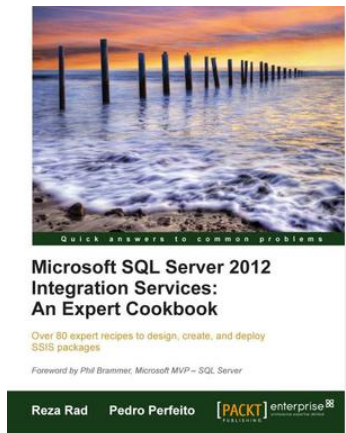
Reference To Study

- Azure Data Factory Documentation

<http://azure.microsoft.com/en-us/documentation/services/data-factory/>

- SSIS Book

<https://www.packtpub.com/networking-and-servers/microsoft-sql-server-2012-integration-services-expert-cookbook>



Questions?



[@Rad Reza](#)



a.raad.g@gmail.com



www.rad.pasfu.com



www.linkedin.com/in/rezarad

Session Evaluation



<http://aka.ms/DBI387>



Microsoft