



PASS
SUMMIT 2015

Azure Data Factory

VS. SSIS

Reza Rad, Consultant, RADACAD



Please silence
cell phones

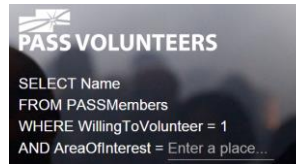
Explore Everything PASS Has to Offer



FREE ONLINE WEBINAR EVENTS



FREE 1-DAY LOCAL TRAINING EVENTS



VOLUNTEERING OPPORTUNITIES



LOCAL USER GROUPS
AROUND THE WORLD



ONLINE SPECIAL INTEREST
USER GROUPS



CommunityCONNECTOR

PASS COMMUNITY NEWSLETTER

PASS BLOG

WHITE PAPERS

SESSION RECORDINGS



FREE ONLINE RESOURCES



BUSINESS ANALYTICS TRAINING



BA INSIGHTS NEWSLETTER

About Me: Reza Rad

DW/BI Architect, Consultant

SQL Server MVP

Author of SQL Server and BI books

Author of SSIS webcast series

Microsoft Certified Trainer

Microsoft Certified Professional



[@Rad_Reza](#)



a.raad.g@gmail.com



www.radacad.com/blog



www.linkedin.com/in/rezarad



Microsoft
CERTIFIED
Trainer

RADACAD

Microsoft
CERTIFIED
Professional

Agenda

Azure Data Factory Overview

SSIS Overview

Comparison

Why This Session?



I was working with SSIS



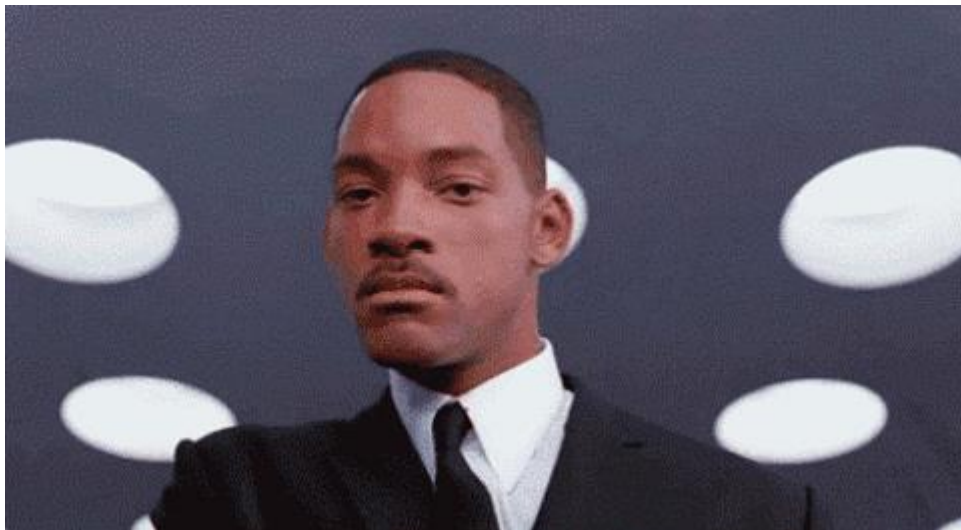
I've heard about Azure Data Factory



Panic!



Give it a go



Found it as Great tool



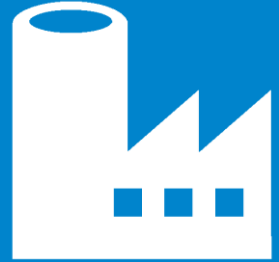
Tell to others about it



Compare



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**.

General Availability August 2015

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, Stored Procedure, and C#.

Monitoring

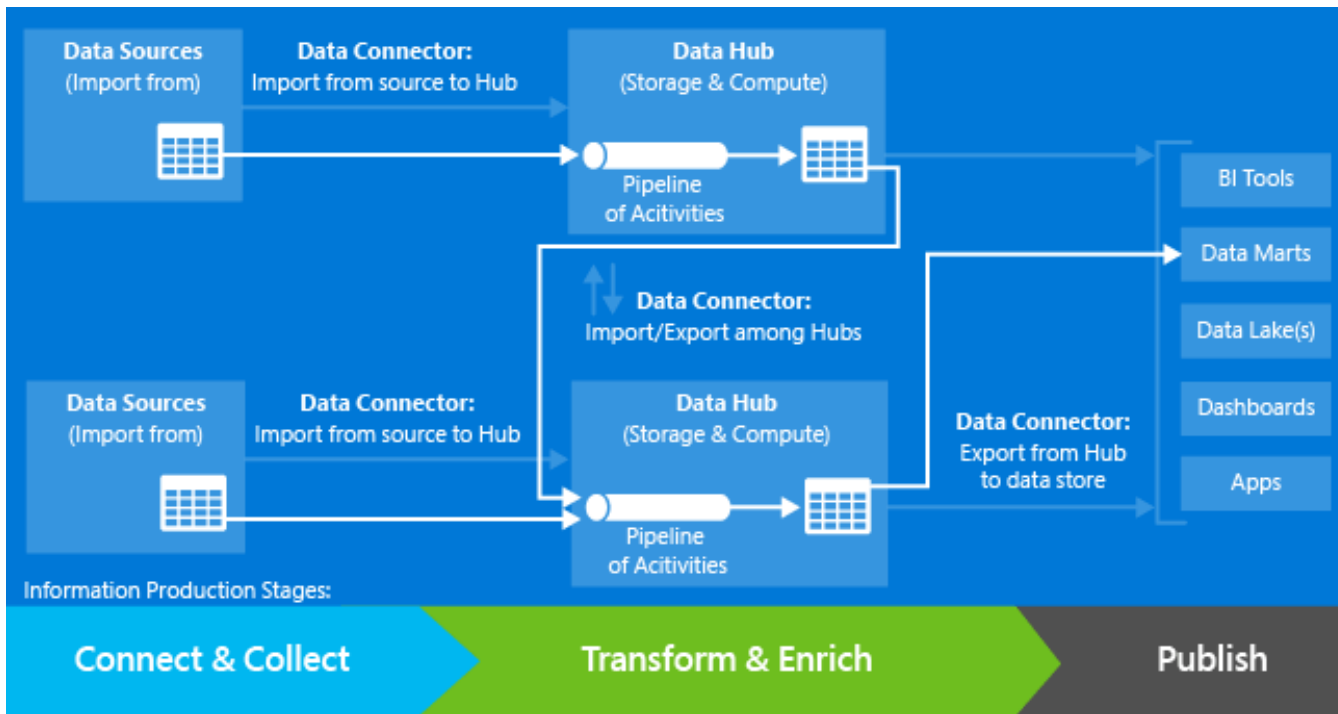
- the pipeline of data, validation and execution of scheduled jobs

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



The screenshot shows the Administration Panel for a Data Factory named 'Test1DF'. The interface is divided into several sections:

- Summary:** A grid of tiles for 'Diagram', 'Quick start', 'Properties', and 'Tags' (0 tags).
- Datasets:** 2 datasets, with 0 having errors.
- Pipelines:** 1 pipeline named 'ADFTutorialPipeline'.
- Linked services:** 4 linked services, including 1 Data Gateway and 3 Data Stores.
- Operations:** A bar chart for 'Events in the past week' and 0 alert rules.

The right-hand pane shows a detailed view of 'Linked services' with a search bar and a table of 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			

ADF Editor

Write JSON

Edit/Create/Delete Objects

Deploy



The screenshot displays the Azure Data Factory (ADF) Editor interface. The left sidebar shows the navigation menu with options: HOME, NOTIFICATIONS, and BROWSE. The main area is divided into two panes. The left pane, titled 'Test34DF', shows a tree view of objects: Linked services (AzureSqlTestDB, AzureStorageCustomerCSV), Tables (No datasets in this data factory), Pipelines, and Drafts (Draft-1). The right pane, titled 'Drafts/Draft-1', shows the JSON definition for the selected draft object.

```
{  
  "name": "AzureBlobDatasetTemplate",  
  "properties": {  
    "structure": [  
      {  
        "name": "<columnname>",  
        "type": "<columntype>"  
      }  
    ],  
    "location": {  
      "type": "AzureBlobLocation",  
      "folderPath": "<Path to the container and folder in the blob storage. Example: myblobconta",  
      "partitionedBy": [  
        {  
          "name": "Date",  
          "type": "Date",  
          "format": "yyyy-MM-dd",  
          "valueFrom": "pipelineStartTime" }  
        ]  
      }  
    }  
  }  
}
```

Main Benefits

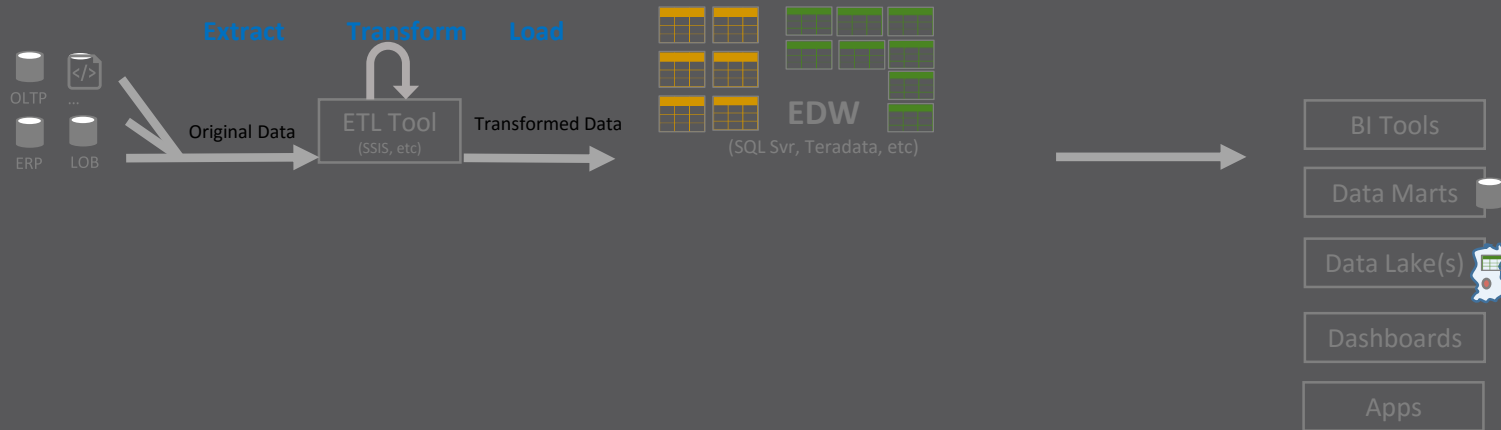
Pay as Usage

Cloud Based

HDInsight compatibility

Demo: Azure Data Factory

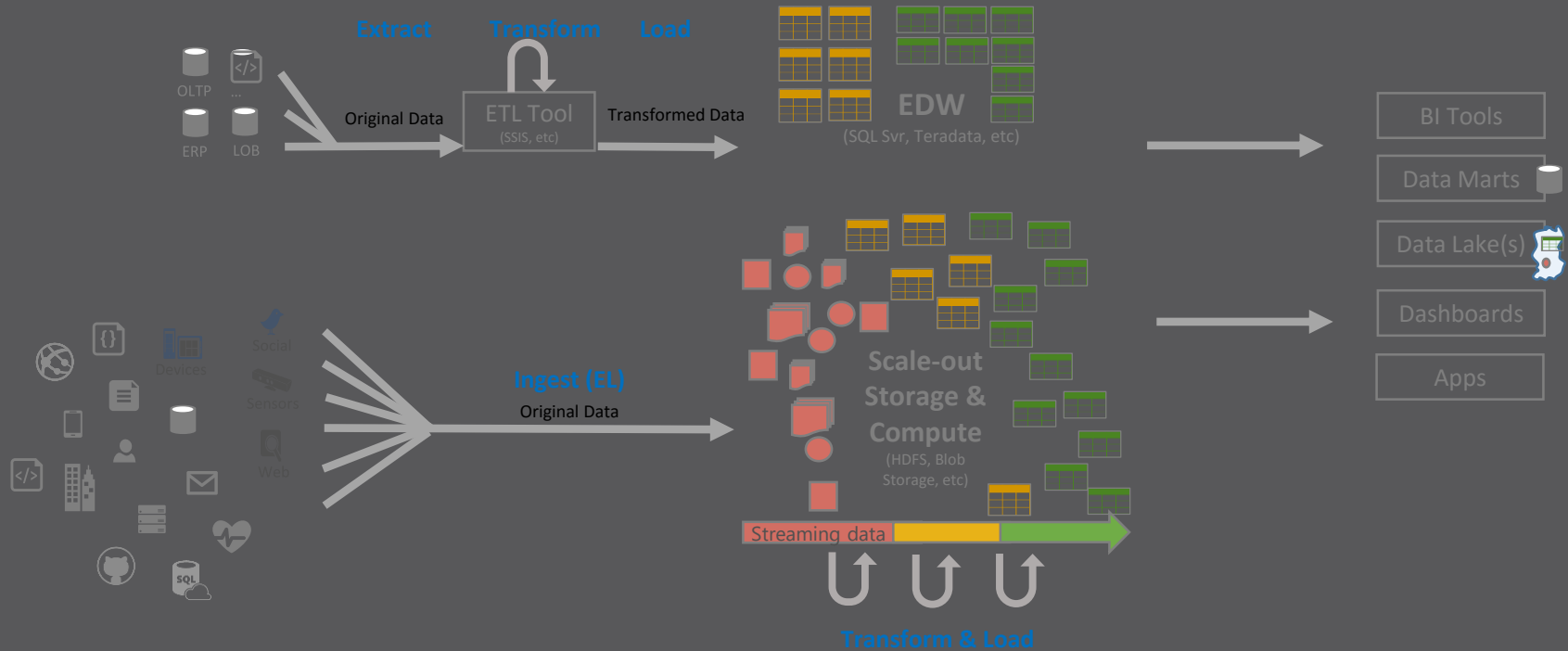
Evolving Approaches to Analytics



Evolving Approaches to Analytics



Evolving Approaches to Analytics



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

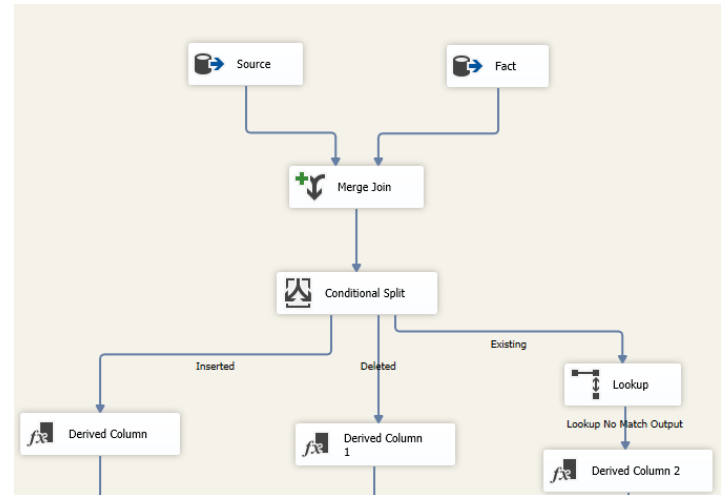
What you can do with SSIS

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

Comparison Factors

Comparison Factors

Development Tool

Administration Tool

User Type

Data
Sources/Destinations

Data Transformations

Pricing

Error Handling

Deployment

Monitoring

Security

Technology

Environment
Requirements

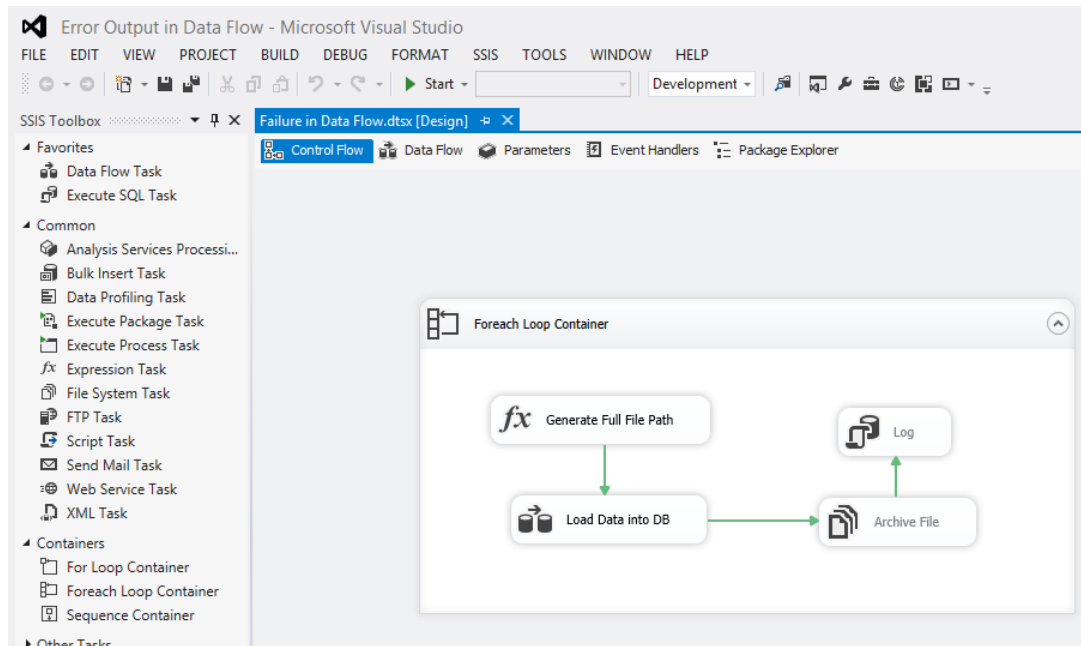
HDInsight
Compatibility

Azure Analytics
Integration

Development Tools: SSIS

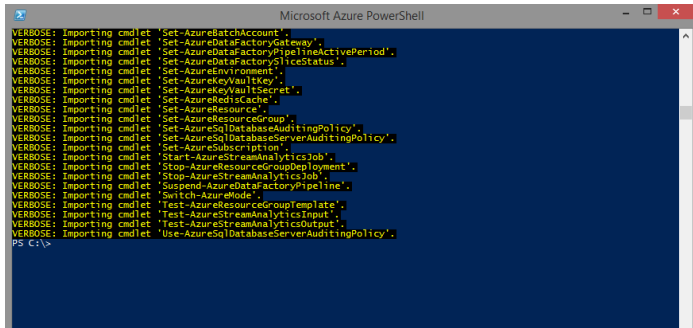
SSDT / BIDS

Rich and Powerful
Available Free

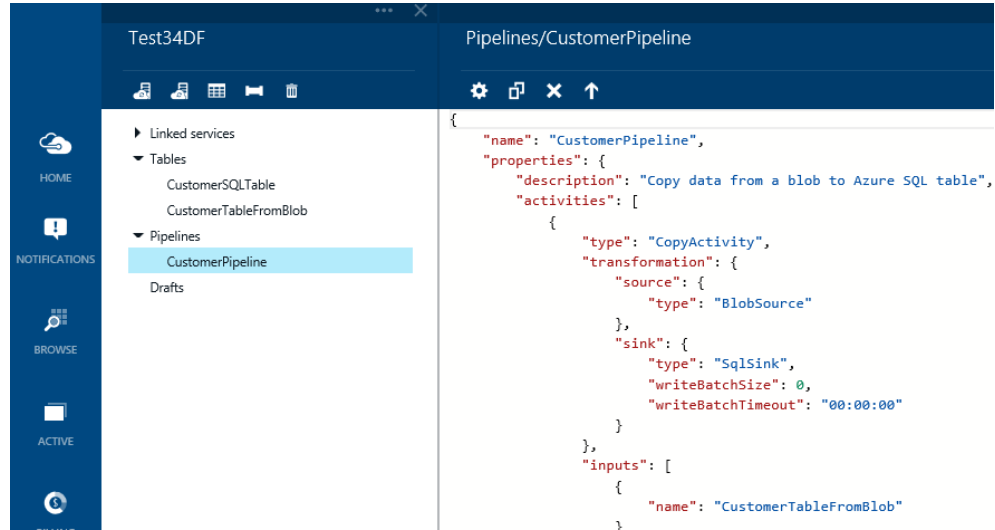


Development Tools: ADF

Azure Portal
New ADF Editor
Azure Power Shell
JSON scripts



```
Microsoft Azure PowerShell
VERBOSE: Importing cmdlet 'Set-AzureBatchAccount'.
VERBOSE: Importing cmdlet 'Set-AzureDataFactoryGateway'.
VERBOSE: Importing cmdlet 'Set-AzureDataFactoryPipelineEffectivePeriod'.
VERBOSE: Importing cmdlet 'Set-AzureDataFactorySyncStatus'.
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-AzureRole'.
VERBOSE: Importing cmdlet 'Test-AzureResourceGroupTemplate'.
VERBOSE: Importing cmdlet 'Test-AzureStreamAnalyticsInput'.
VERBOSE: Importing cmdlet 'Test-AzureStreamAnalyticsOutput'.
VERBOSE: Importing cmdlet 'Use-AzureSQLDatabaseServerAuditingPolicy'.
PS C:\>
```



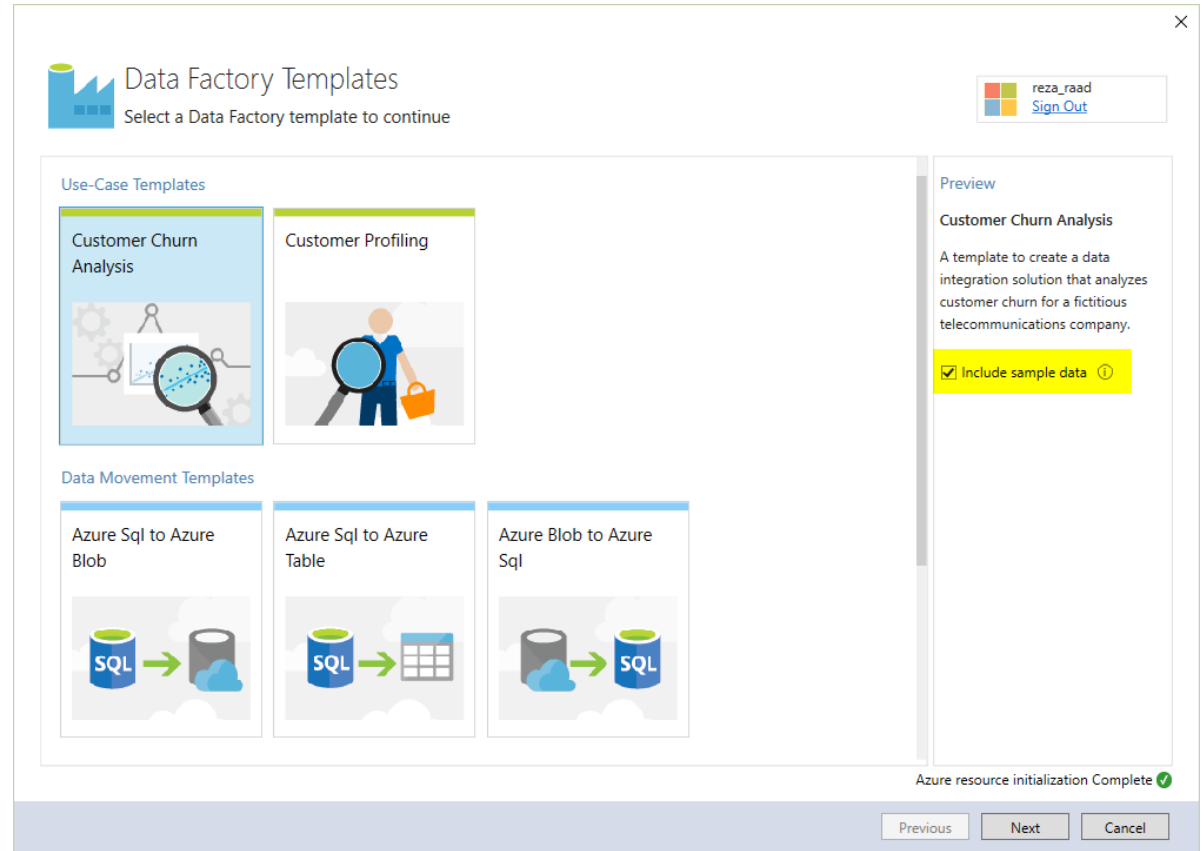
```
Test34DF Pipelines/CustomPipeline
{
  "name": "CustomerPipeline",
  "properties": {
    "description": "Copy data from a blob to Azure SQL table",
    "activities": [
      {
        "type": "CopyActivity",
        "transformation": {
          "source": {
            "type": "BlobSource"
          },
          "sink": {
            "type": "SqlSink",
            "writeBatchSize": 0,
            "writeBatchTimeout": "00:00:00"
          }
        },
        "inputs": [
          {
            "name": "CustomerTableFromBlob"
          }
        ]
      }
    ]
  }
}
```

Visual Studio Templates for ADF

Awesome!
But way far from
Rich Editor

No GUI
No Intellisense
No Debugging

....



Comparison: Development Tool

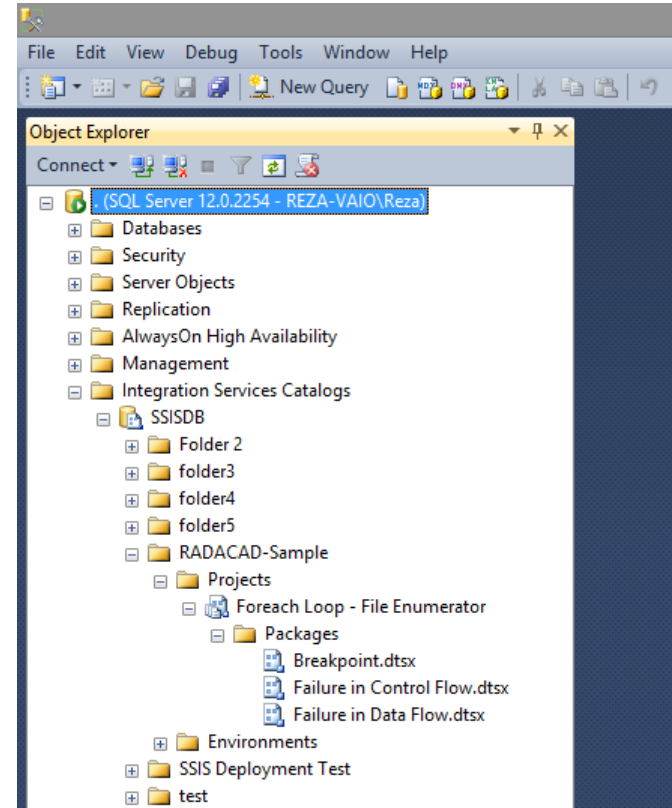
	SSIS	ADF
Standalone Tool	Yes	Yes
Powerful GUI	Yes	No
Availability	Free	Free
Prerequisite	SQL Server License	Azure Subscription



Administration Tool: SSIS

SSMS

Available Free (New Versions)



Administration Tool: ADF

Azure Portal

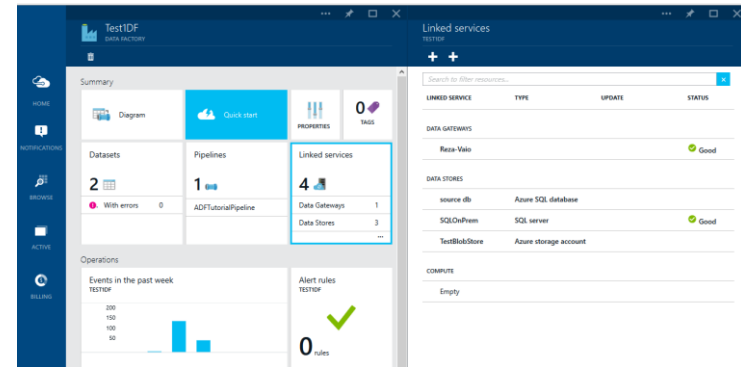
Available under Azure subscription

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-AzureDataFactoryPipelineStatus'.
VERBOSE: Importing cmdlet 'Set-AzureEnvironment'.
VERBOSE: Importing cmdlet 'Set-AzureKeyVaultKey'.
VERBOSE: Importing cmdlet 'Set-AzureKeyVaultSecrets'.
VERBOSE: Importing cmdlet 'Set-AzureRedisCache'.
VERBOSE: Importing cmdlet 'Set-AzureResource'.
VERBOSE: Importing cmdlet 'Set-AzureResourceGroup'.
VERBOSE: Importing cmdlet 'Set-AzureSQLDatabaseServerAuditingPolicy'.
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: 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



Not a Power User, But Developer



Data Src/Dest: SSIS

Wide Range

- Other Sources
 - ADO NET Source
 - CDC Source
 - Excel Source
 - Flat File Source
 - ODBC Source
 - OLE DB Source
 - Raw File Source
 - XML Source
- Other Destinations
 - ADO NET Destination
 - Data Mining Model Training
 - DataReader Destination
 - Dimension Processing
 - Excel Destination
 - Flat File Destination
 - ODBC Destination
 - OLE DB Destination
 - Partition Processing
 - Raw File Destination
 - Recordset Destination
 - SQL Server Compact Desti...
 - SQL Server Destination

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

Data Store Types Supported so far:

- + Azure SQL Data Warehouse
- + Azure Data Lake Store



Azure Storage



Azure SQL Database



SQL Server



File System



Oracle Database



PostgreSQL



MySQL



Teradata



DB2



Sybase

Comparison: Data Src/Dest

	SSIS	ADF
SQL Server On Premises	Yes	Yes
SQL Azure	Yes	Yes
Azure Storage	Yes	Yes
Oracle	Yes	Yes
SAP	Yes	
MySQL	Yes	Yes
DB2	Yes	Yes
CSVs	Yes	Yes
Web Services	Yes	
Many other Data src/dest	Yes	
Azure Data Lake Store		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

- Common
 - Aggregate
 - Conditional Split
 - Data Conversion
 - Derived Column
 - Lookup
 - Merge
 - Merge Join
 - Multicast
 - OLE DB Command
 - Row Count
 - Script Component
 - Slowly Changing Dimension
 - Sort
 - Union All
- Other Transforms
 - Audit
 - Cache Transform
 - CDC Splitter
 - Character Map
 - Copy Column
 - Data Mining Query
 - DQS Cleansing
 - Export Column
 - Fuzzy Grouping
 - Fuzzy Lookup

Data Transformations: ADF

Copy Activity

Stored Procedure (SPROC) Activity

Hive and Pig Activity

- Support wide range of transformations through HDInsight

Custom C# Activity

- Write your own transformation with C#

Azure ML Batch Scoring Activity

Azure Data Lake Analytics U-SQL Activity

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
Azure ML Batch Scoring	With Scripting	Yes
Stored Procedure	Yes	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

	USAGE	PRICE
Cloud	First 5 activities/month	Free
	6 – 100 activities/month	\$0.60 per activity
	> 100 activities/month	\$0.48 per activity
On-Premises	First 5 activities/month	Free
	6 – 100 activities/month	\$1.50 per activity
	> 100 activities/month	\$1.20 per activity

High frequency

	USAGE	PRICE
Cloud	1 – 100 activities/month	\$1 per activity
	> 100 activities/month	\$0.80 per activity
On-Premises	1 – 100 activities/month	\$2.50 per activity
	> 100 activities/month	\$2 per activity

Data Movement

Cloud	\$0.25 per hour
On-Premises	\$0.10 per hour

Inactive Pipelines

\$0.80/month

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

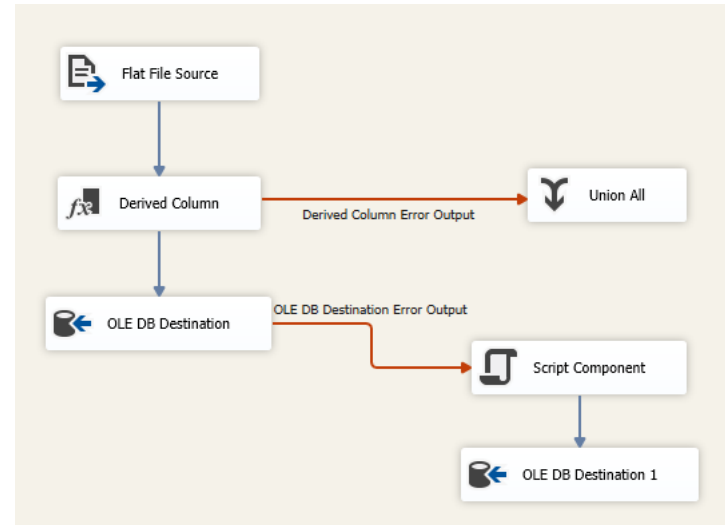
	SSIS	ADF
Hardware Setup	Yes	
Software Setup	Yes	
Administration Costs	Yes	
Azure Environment Usage		Yes

Error Handling: SSIS

Error Handling through Event Handlers (OnError)

Error Output in Data Flow

Failure Precedence Constraint



Error Handling: ADF

Error Message Logging

Alert Rules

No Error Handling Events

The screenshot displays the Azure Data Factory (ADF) monitoring interface. The main dashboard area is divided into several sections:

- Summary:** Includes a 'Diagram' button, a 'Quick start' button, and 'PROPERTIES' and 'TAGS' sections.
- Datasets:** Shows 2 datasets, with 1 having errors.
- Pipelines:** Shows 1 pipeline named 'ADFTutorialPipelineOnPr...'. Below it, 'Data Gateways' are listed as 1 and 'Data Stores' as 2.
- Operations:** Features a bar chart titled 'Events in the past week' for 'ADFONPREM' and an 'Alert rules' section showing 0 rules.
- Monitoring:** Includes a 'Data factory metrics' section.

On the right side, the 'Events' table provides a detailed log of operations:

OPERATION	STATUS	TIME	CALLER
RunFinished	Failed	2 d ago	--
RunStarted	Started	2 d ago	--
RunFinished	Failed	2 d ago	--
RunFinished	Succeeded	2 d ago	--
RunStarted	Started	2 d ago	--
RunStarted	Started	2 d ago	--
RunFinished	Succeeded	2 d ago	--
RunStarted	Started	2 d ago	--
RunFinished	Succeeded	2 d ago	--
RunStarted	Started	2 d ago	--
RunFinished	Succeeded	2 d ago	--
RunStarted	Started	2 d ago	--
RunFinished	Succeeded	2 d ago	--
RunStarted	Started	2 d ago	--
RunFinished	Succeeded	2 d ago	--

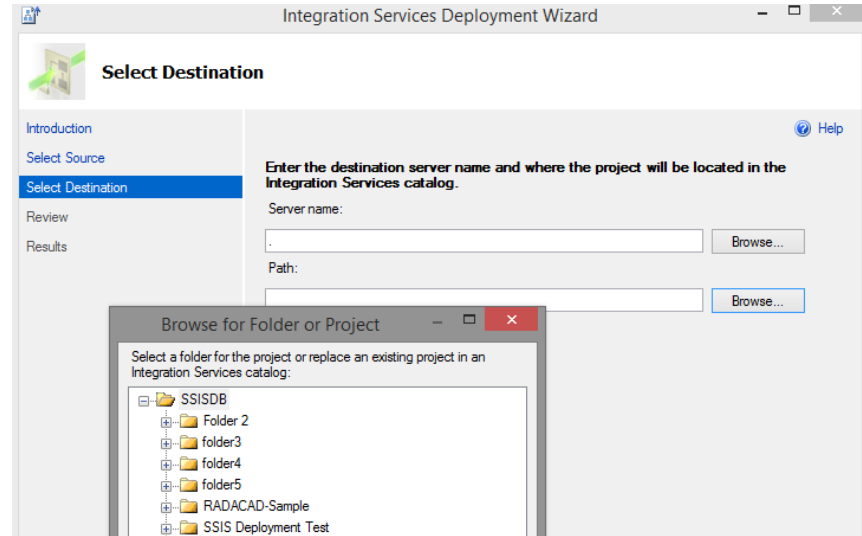
Comparison: Error Handling

	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



Deployment through SSDT: ADF

Data Factory Configuration
Create new or select an existing data factory

reza_raad

Configure data factory

Configure data stores

Configure compute

Summary

Deployment Status

Use existing Data Factory

Name: ADFOnPrem

Create New Data Factory

Name: Customer Churn Analysis

Tags:

Description: Customer Churn Analysis

Subscription: 71f18220-46d3-4870-ad7e-97aa201c3948

Resource Group: ADF

Region: West US

Publish to Data Factory

Previous Next Cancel

Comparison: Deployment

	SSIS	ADF
Deployment GUI	Yes	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

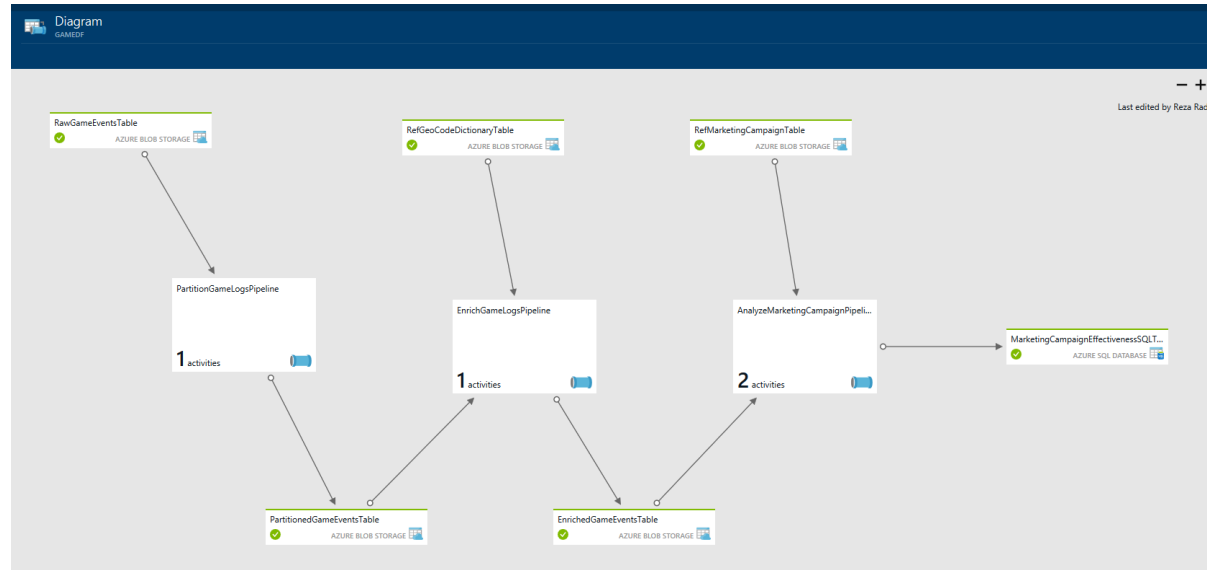
Drill through monitoring features

Powerful GUI

Data Slice execution

Data Lineage

- Very Important Feature



Comparison: Monitoring

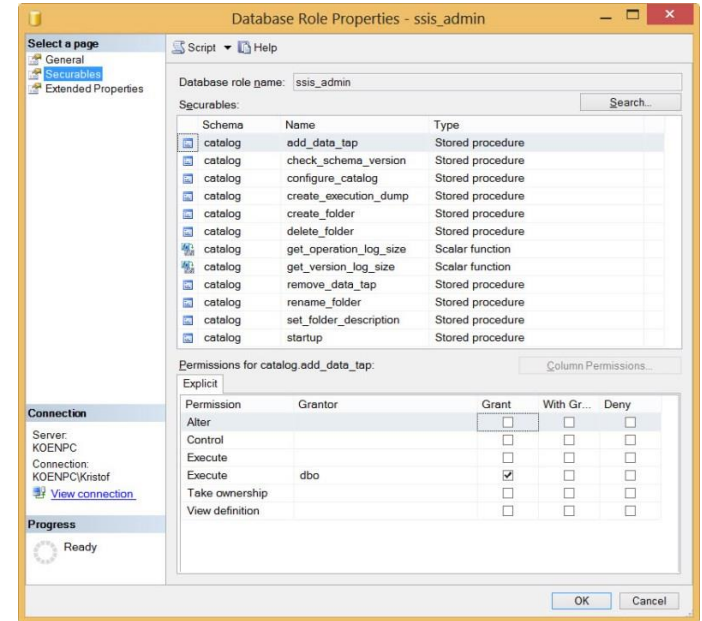


	SSIS	ADF
Monitoring GUI	Yes	Yes
Drill through Monitoring	Yes	Yes
Data Slice		Yes
Data Lineage		Yes

Security: SSIS

Role Based

Roles for Deploy, Execute, Monitor...



Security: ADF

Role Based

Roles:

- Owner
- Contributor
- Reader
- Data Factory Contributor
- User Access Administrator

The screenshot shows the 'Roles' page for the application 'TESTIDF'. It lists five roles with their respective user and group counts and a right-pointing arrow for each role.

Role	Users	Groups
OWNER	0	1
CONTRIBUTOR	0	0
READER	0	0
DATA FACTORY CONTRIBUTOR	0	0
USER ACCESS ADMINISTRATOR	0	0

Comparison: Security

	SSIS	ADF
Role Based Security	Yes	Yes

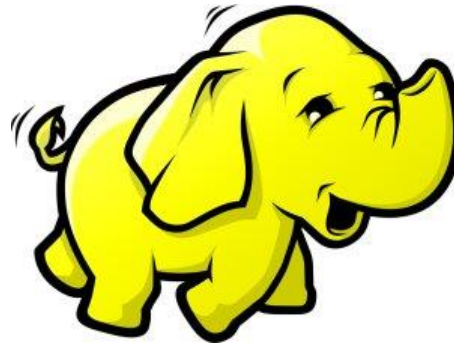


HDInsight Compatibility

Azure Data Factory can work with existing HDInsight Clusters

ADF Can create HDInsight cluster on demand

ADF HDInsight Activity run Pig and Hive scripts



Azure Analytics Integration

Azure ML Batch Scoring Activity

Data Lake Analytics U-SQL Activity



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	No Costs
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
Azure Analytics Integration		High

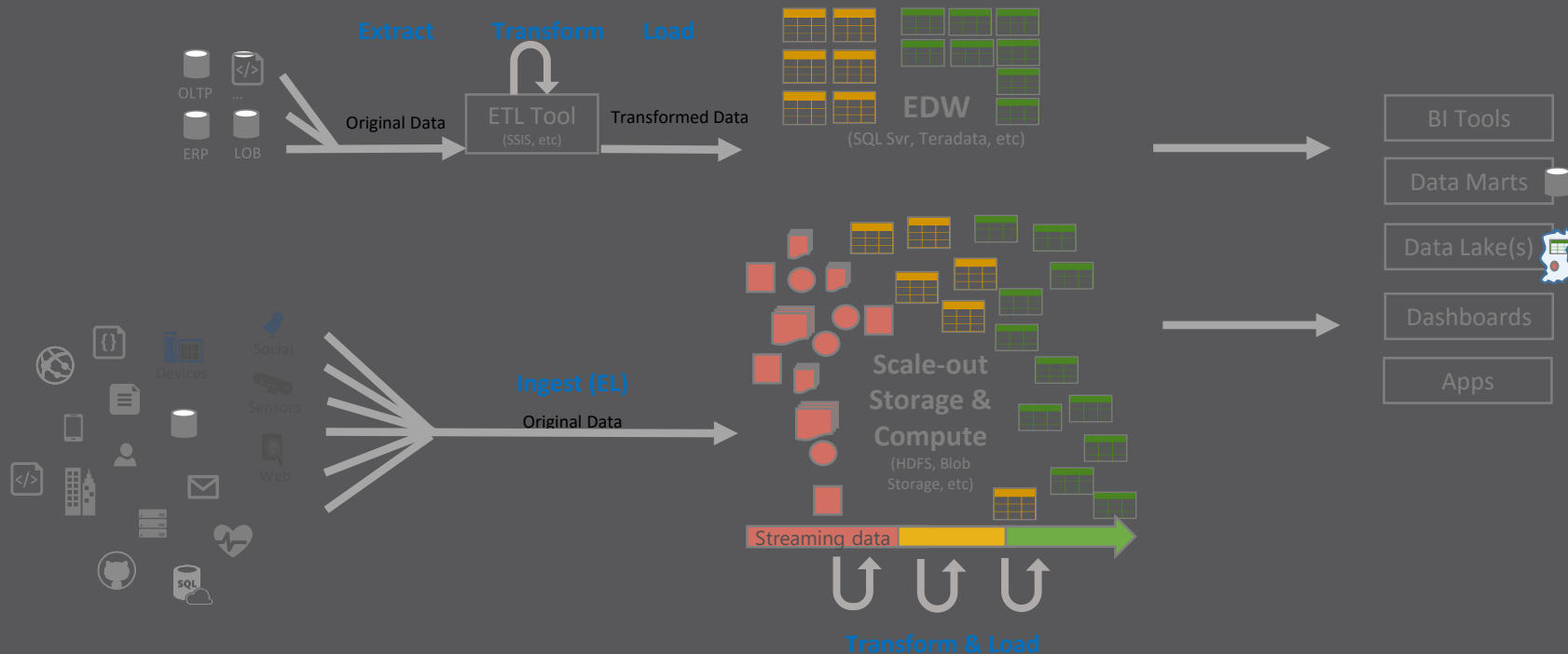
Hybrid

Purpose

These tools/services didn't built for the same purpose

They are complement of each other

Evolving Approaches to Analytics



Get Benefits of Both

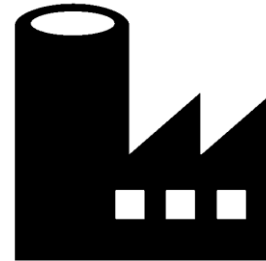
Hybrid SSIS and ADF

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

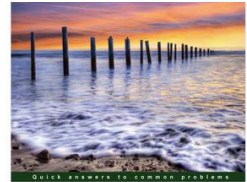
<http://azure.microsoft.com/en-us/documentation/services/data-factory/>

Azure Data Factory posts on my Blog:

<http://www.radacad.com/tag/azure-data-factory>

SSIS Book

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



Microsoft SQL Server 2012
Integration Services:
An Expert Cookbook

Over 80 expert recipes to design, create, and deploy
SSIS packages

Foreword by Phil Beaman, Microsoft MVP - SQL Server

Reza Rad Pedro Perfeito [PACKT] enterprise

Questions?



[@Rad Reza](#)



a.raad.g@gmail.com



www.radacad.com/blog



www.linkedin.com/in/rezarad

Session Evaluations

Your feedback is important and valuable.

3
ways to access

Submit by 5pm
Friday November 6th to
WIN prizes



Go to passSummit.com



Download the GuideBook App
and search: **PASS Summit 2015**



Follow the QR code link displayed
on session signage throughout the
conference venue and in the
program guide



Thank You