

Azure Data Factory vs SSIS

REZA RAD

sqlbits

About Reza Rad

DW/BI Consultant, Mentor, Trainer

SQL Server MVP

Author of SQL Server and BI books

Author of MSBI webcast series on RADACAD

Microsoft Certified Trainer

Microsoft Certified Professional

Co-Leader of NZ BI User Group



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



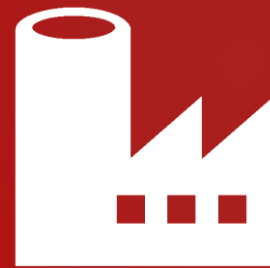
Found it as a Great tool



Compare



Azure Data Factory



Azure Data Factory

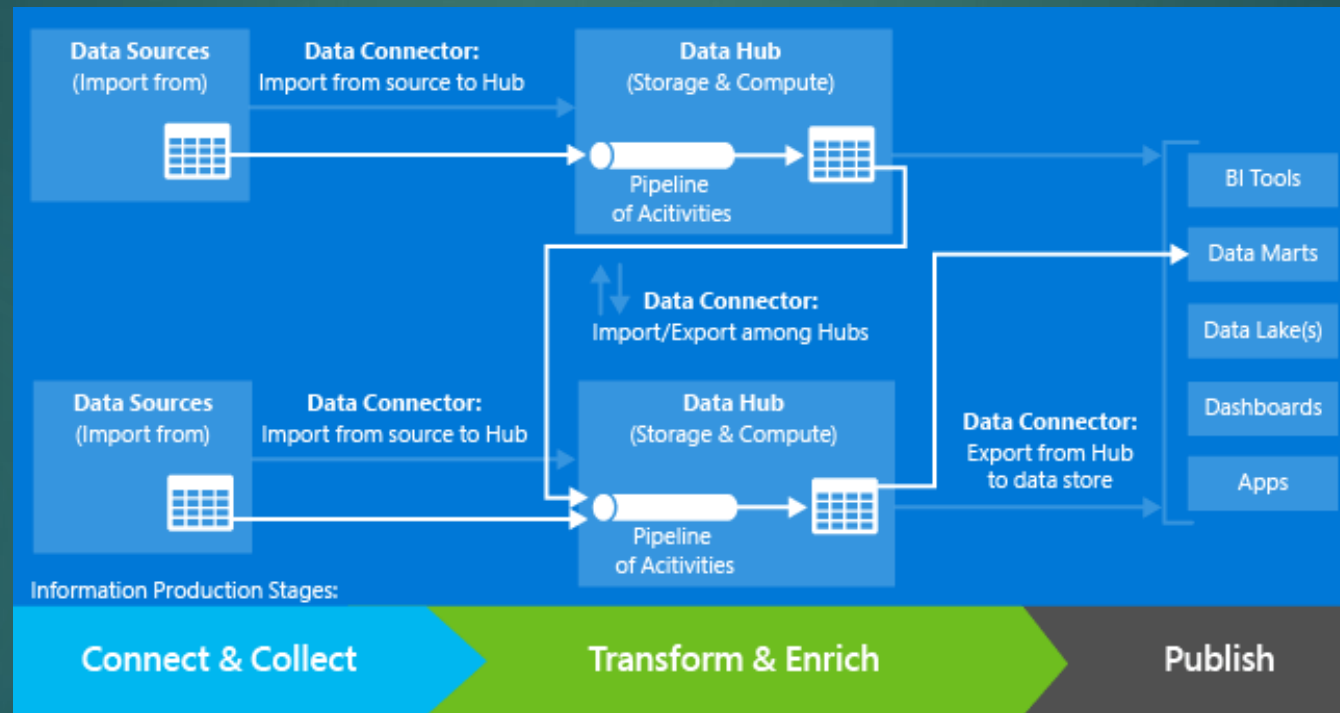


Fully Managed Service for **Composing** Data Storages, Processing, and Movement Services into Streamlined, Scalable, and Reliable **Data Production Pipelines**.

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 displays the Azure Data Factory Administration Panel for a workspace named 'Test1DF'. The interface is divided into several sections:

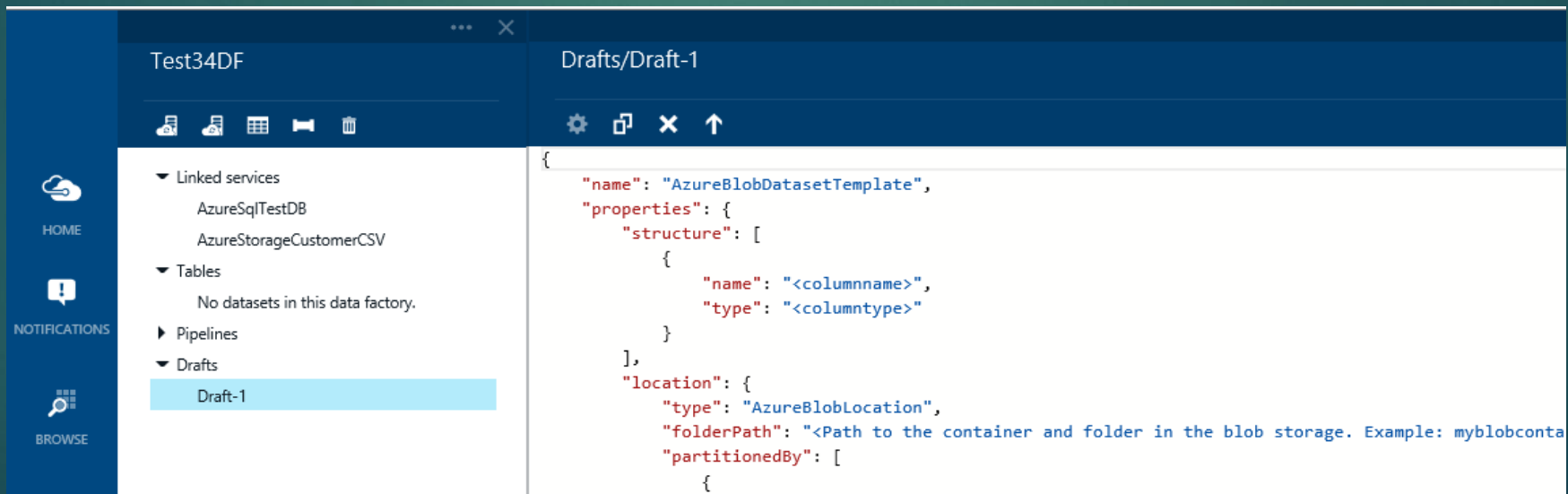
- Summary:** A dashboard with four main tiles: 'Diagram', 'Quick start' (highlighted in blue), 'PROPERTIES', and 'TAGS' (showing 0 tags).
- Datasets:** Shows 2 datasets, with 0 having errors.
- Pipelines:** Shows 1 pipeline named 'ADFTutorialPipeline'.
- Linked services:** Shows 4 linked services, including 1 Data Gateway and 3 Data Stores.
- Operations:** Includes a bar chart for 'Events in the past week' and a section for 'Alert rules' showing 0 rules.

The right-hand pane is titled 'Linked services' and contains 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



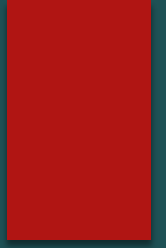
The screenshot displays the Azure Data Factory Editor interface. On the left, a navigation pane shows the hierarchy: Test34DF > Drafts > Draft-1. The main area shows the JSON definition for the selected draft object:

```
{  
  "name": "AzureBlobDatasetTemplate",  
  "properties": {  
    "structure": [  
      {  
        "name": "<columnname>",  
        "type": "<columnntype>"  
      }  
    ],  
    "location": {  
      "type": "AzureBlobLocation",  
      "folderPath": "<Path to the container and folder in the blob storage. Example: myblobconta",  
      "partitionedBy": [  
        {
```


Main Benefits

- ▶ Pay as Usage
- ▶ Cloud Based
- ▶ HDInsight compatibility
- ▶ Built for Massive Data Movement

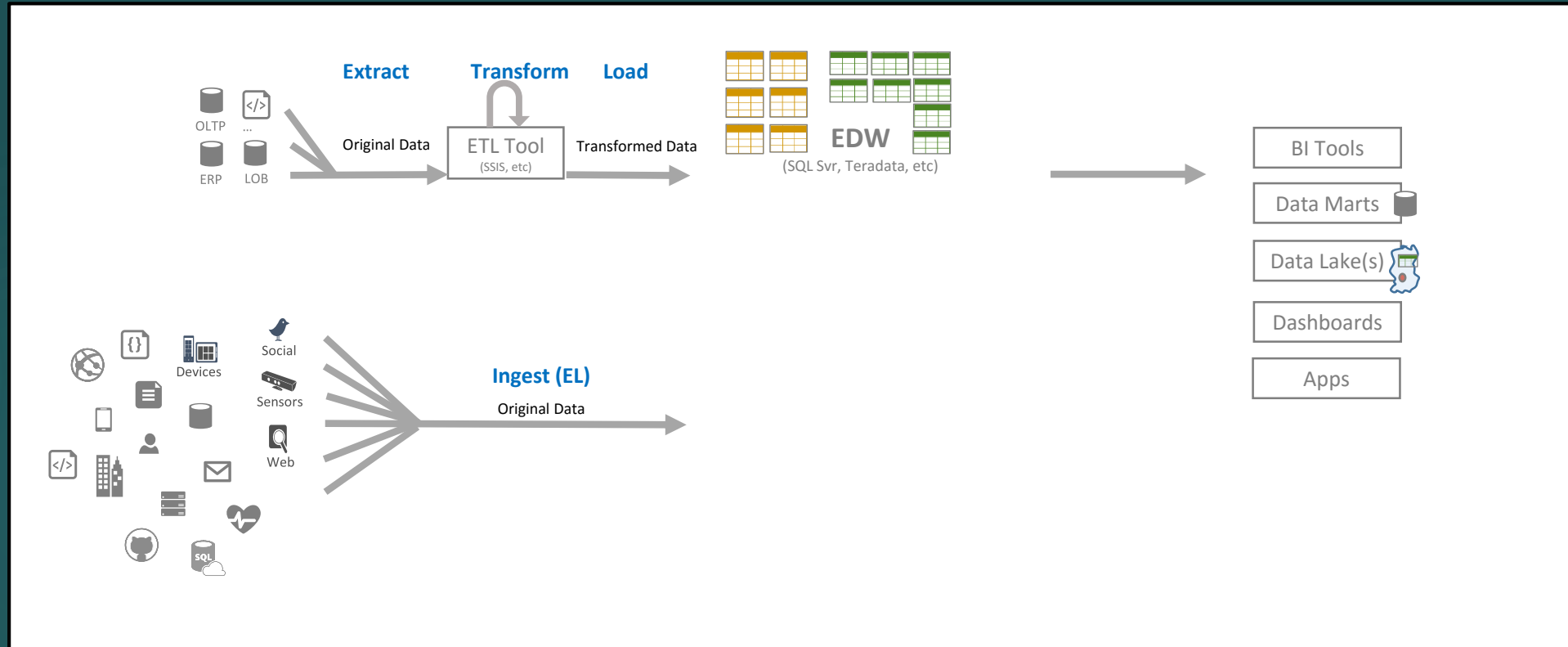
Demo: Azure Data Factory



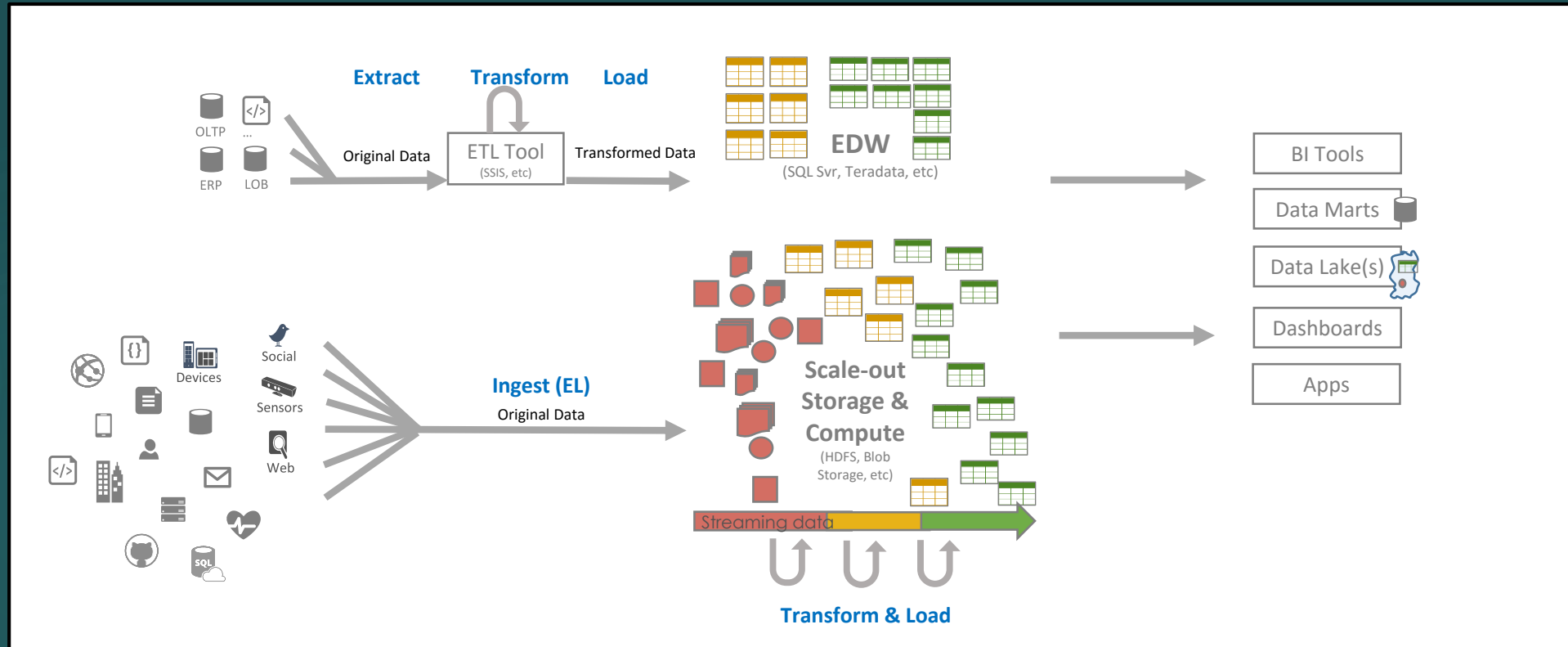
Evolving Approaches to Analytics



Evolving Approaches to Analytics



Evolving Approaches to Analytics



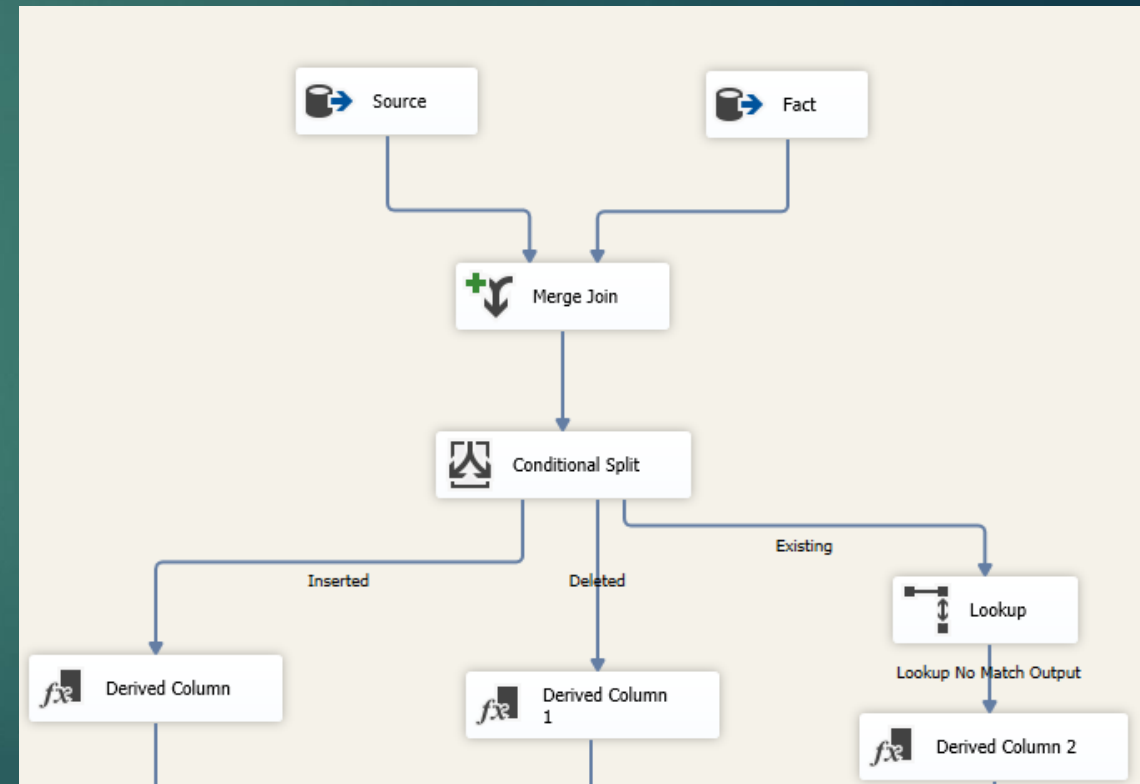
SSIS 101

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

Comparison Factors

Comparison Factors

Development Tool

Administration Tool

User Type

Data Sources/Destinations

Data Transformations

Extension & Automation

Pricing

Error Handling

Deployment

Monitoring

Security

Technology

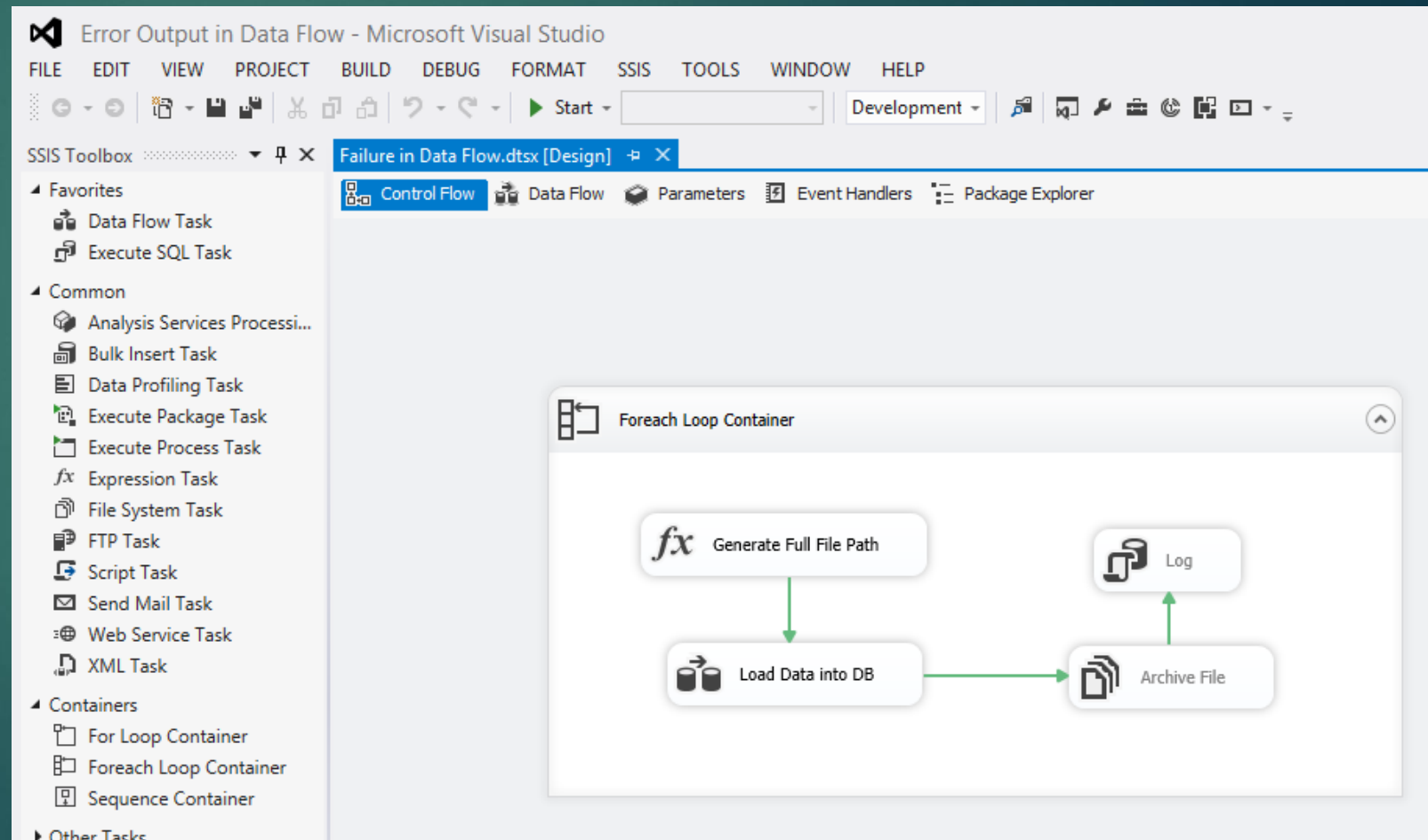
Environment Requirements

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

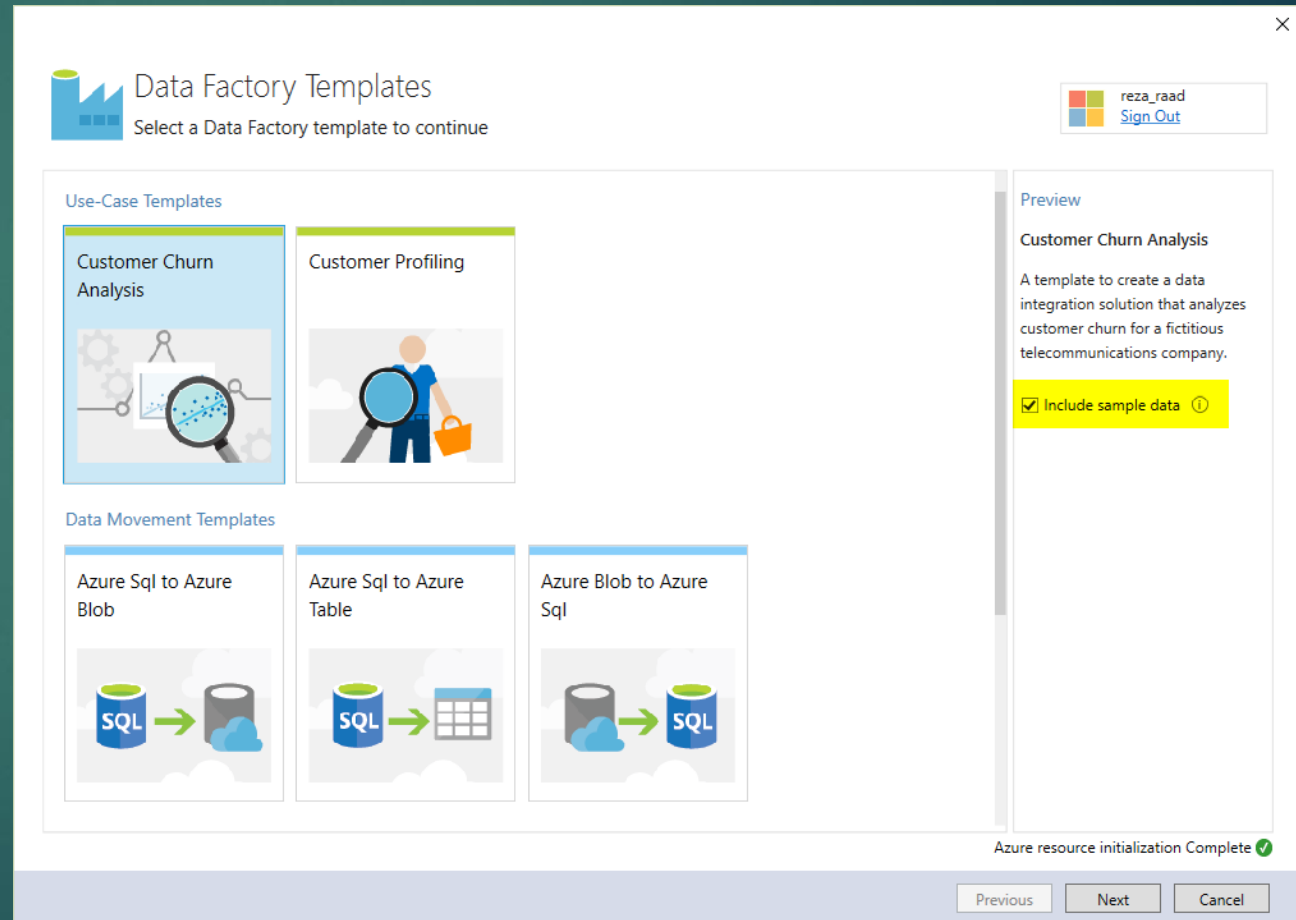
```
Microsoft Azure PowerShell
VERBOSE: Importing cmdlet 'Set-AzureBatchAccount'.
VERBOSE: Importing cmdlet 'Set-AzureDataFactoryGateway'.
VERBOSE: Importing cmdlet 'Set-AzureDataFactoryPipelineActivePeriod'.
VERBOSE: Importing cmdlet 'Set-AzureDataFactorySliceStatus'.
VERBOSE: Importing cmdlet 'Set-AzureEnvironment'.
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:\>
```

The screenshot displays the Azure Portal interface for an Azure Data Factory (ADF) instance named 'Test34DF'. The left-hand navigation pane shows a tree view with categories: Linked services, Tables (containing 'CustomerSQLTable' and 'CustomerTableFromBlob'), Pipelines (containing 'CustomerPipeline' and 'Drafts'), and Drafts. The 'CustomerPipeline' item is selected and highlighted. The main content area on the right shows the JSON definition for the 'CustomerPipeline' activity. The JSON is as follows:

```
{
  "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
- ▶



Copy Data - Preview

- ▶ Wizard for Copy Data
- ▶ Still Preview
- ▶ Not for Edit, only for Create
- ▶ Good Start

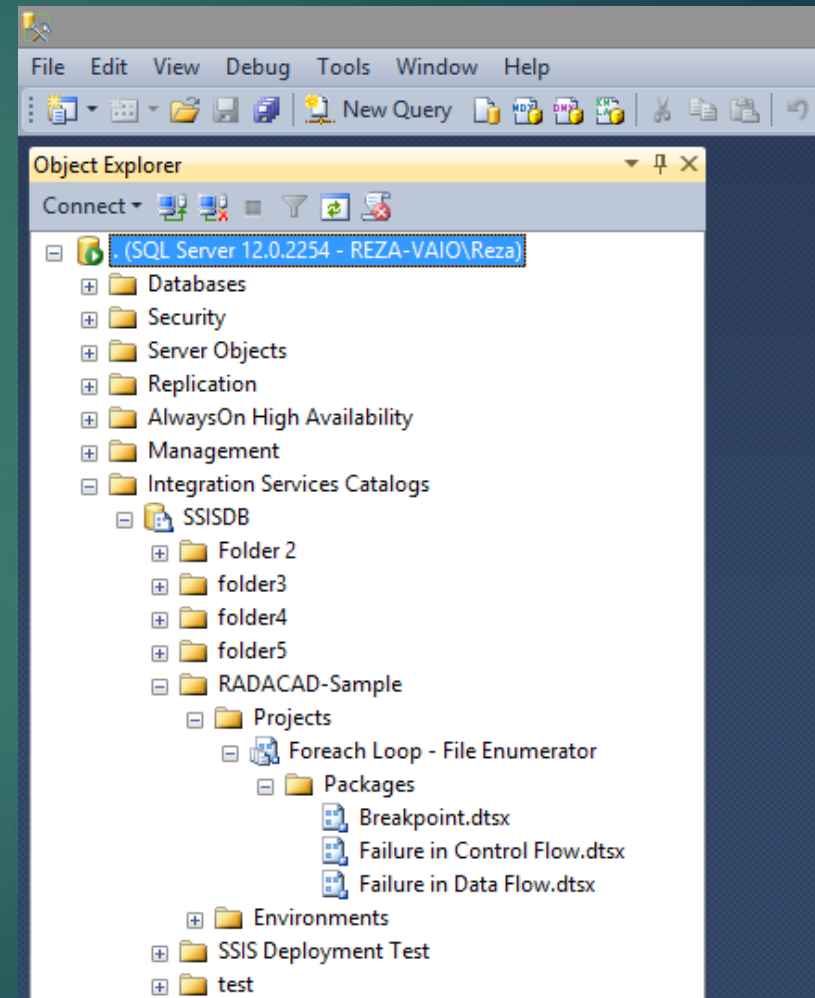
Comparison: Development Tool

| | 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-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:\>
```

The screenshot shows the Azure Data Factory (ADF) portal interface for a workspace named 'Test1DF'. The interface is divided into several sections:

- Summary:** A dashboard with four main tiles: 'Diagram', 'Quick start', 'Properties', and 'Tags'. Below these are 'Datasets' (2), 'Pipelines' (1, including 'ADFTutorialPipeline'), and 'Linked services' (4). The 'Linked services' tile is highlighted with a blue border and shows a sub-summary: 'Data Gateways' (1) and 'Data Stores' (3).
- Operations:** A section with a bar chart titled 'Events in the past week' and an 'Alert rules' section showing '0 rules' with a green checkmark.
- Linked services:** A table on the right side of the interface listing various linked services.

| 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 | | | |

The New Monitoring App

- ▶ Data Factory Monitoring App released 10 days ago!

The screenshot displays the Data Factory Monitoring App interface. The main window shows a pipeline diagram with several activities: PartitionedProductsUsageTable, InputProductsUsageTable, ProductSimilarityPipeline, and ProductRecommendationPipeline. A calendar view is overlaid on the PartitionedProductsUsageTable activity, showing the month of 2016 with days Jan through Dec. Below the pipeline diagram is a table of activity windows.

| Pipeline | Activity | Window Start | Window End | Status | Type | Attempt Start | Attempt End | Duration | Attempts |
|----------------------|--------------------|---------------------|---------------------|--------|------------------------|----------------------|----------------------|----------|----------|
| MaySimilarProduct... | MayRecommende... | 05/01/2016 12:00 A. | 05/01/2016 12:00 A. | Ready | HD Insight Hive Act... | 05/28/2016 9:01 PM. | 05/28/2016 9:01 PM. | 00:03:54 | 1 |
| MayRecommende... | MayRecommende... | 05/01/2016 12:00 A. | 05/01/2016 12:00 A. | Ready | HD Insight Hive Act... | 05/28/2016 9:01 PM. | 05/28/2016 9:04 P. | 00:02:52 | 1 |
| PartitionedProdu... | ProdPartitioned... | 05/01/2016 12:00 A. | 05/01/2016 12:00 A. | Ready | HD Insight Hive Act... | 05/28/2016 8:45 P. | 05/28/2016 8:48 P. | 00:03:03 | 2 |
| PrepareUsagePip... | PrepareUsageAct... | 05/01/2016 12:00 A. | 05/01/2016 12:00 A. | Ready | HD Insight Hive Act... | 05/28/2016 9:09 PM. | 05/28/2016 9:12 PM. | 00:02:29 | 1 |
| ProductRecommen... | MaloutActivity | 05/01/2016 12:00 A. | 05/01/2016 12:00 A. | Ready | HD Insight Hive Act... | 05/28/2016 10:32 P. | 05/28/2016 10:40 P. | 00:08:29 | 1 |
| ProductRecommen... | MaloutActivity | 05/01/2016 12:00 A. | 05/01/2016 12:00 A. | Ready | HD Insight Hive Act... | 05/28/2016 10:41 P. | 05/28/2016 10:49 P. | 00:08:03 | 1 |
| ProductRecommen... | ProductRecommen... | 05/01/2016 12:00 A. | 05/01/2016 12:00 A. | Ready | HD Insight Hive Act... | 05/28/2016 10:53 PM. | 05/28/2016 11:00 PM. | 00:06:53 | 1 |

The interface also includes a Resource Explorer on the left, an Activity Window Explorer on the right, and a detailed Activity Window view at the bottom right showing the start and end time, status (Ready), and activity name (ProdPartitionedActivity).

Comparison: Administration Tool

| | SSIS | ADF |
|-----------------|--------------------|---------------------------|
| Standalone Tool | Yes | Yes |
| Powerful GUI | Yes | Yes, few with Power Shell |
| Availability | Free | Under subscription |
| Prerequisite | SQL Server License | Azure Subscription |

User Type

- ▶ Both SSIS and ADF are Developer Tools and Services, Not for Power User



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
- ▶ +Web
- ▶ +HDFS
- ▶ +ODBC
- ▶ +OData

| | |
|---|--------------------|
|  | Azure Storage |
|  | Azure SQL Database |
|  | SQL Server |
|  | File System |
|  | Oracle Database |
|  | PostgreSQL |
|  | MySQL |
|  | Teradata |
|  | DB2 |
|  | Sybase |

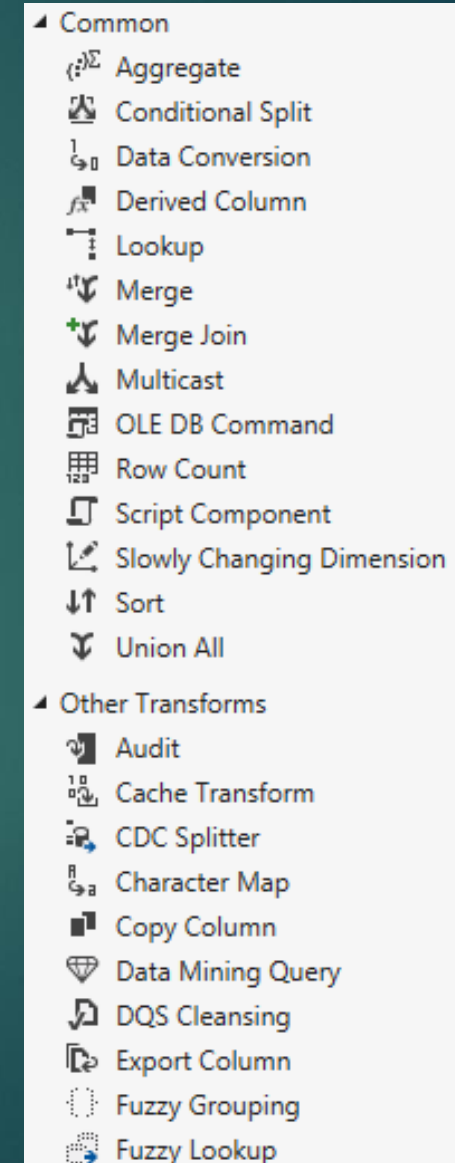
Comparison: Data Src/Dest

| | 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 | Fewer |
| 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 |



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 | |

Extension & Automation: SSIS

- ▶ Programing SDK
- ▶ BIML; Automation
- ▶ Third-party tasks, and components: Pragmatic Works, CozyRoc....

Extension & Automation: ADF

- ▶ Power Shell
- ▶ No Programming SDK
- ▶ No third-party components YET!

Comparison: Extension & Automation

| | SSIS | ADF |
|------------------------|-----------|------------------|
| Programing SDK | Yes | No |
| Automation | Yes, BIML | Yes, Power Shell |
| Third-Party components | Yes | No |



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 | HIGH FREQUENCY |
|-------------|--------------------------------|--------------------------------|
| Cloud | £0.3665 per activity per month | £0.6109 per activity per month |
| On-premises | £0.9164 per activity per month | £1.5273 per activity per month |

Usage beyond 100 activities/month will receive a 20% discount for both low-frequency and high-frequency scenarios. The first 5 low-frequency activities in a month are free of charge in both cloud and on-premises variants.

Data movement

| | |
|-------------|------------------|
| Cloud | £0.1527 per hour |
| On-premises | £0.0611 per hour |

Inactive pipelines

£0.4887/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

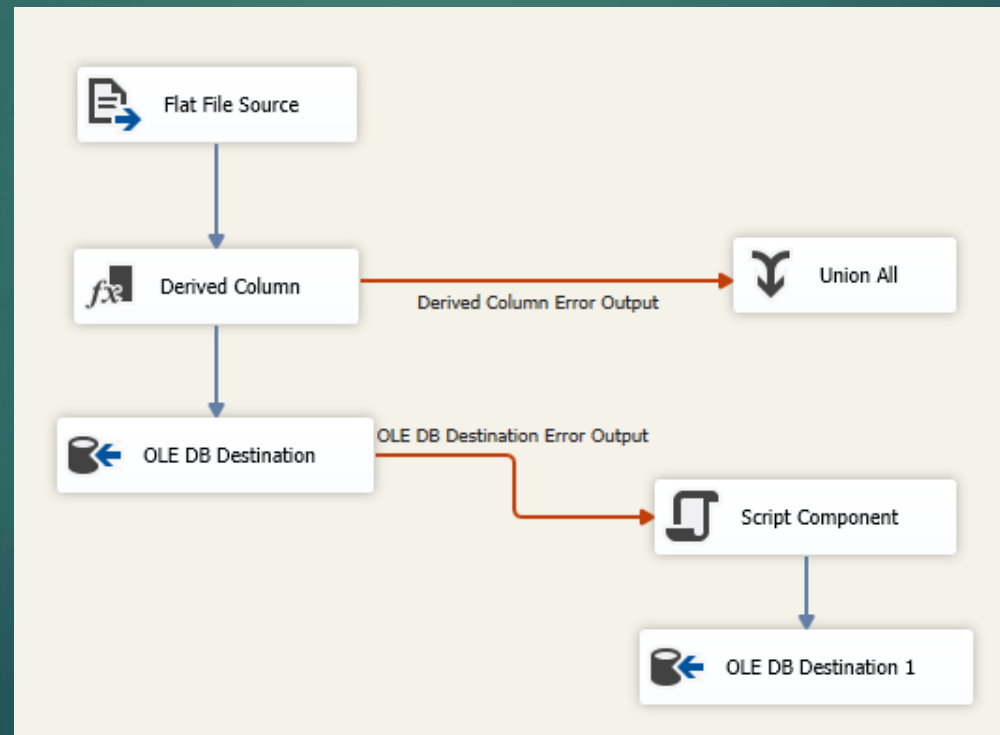


Comparison: Environment

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

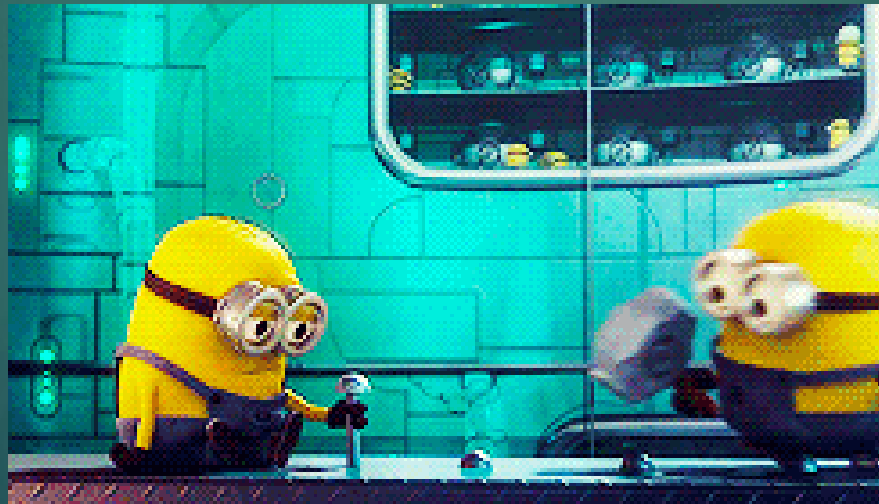
Error Handling: SSIS

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



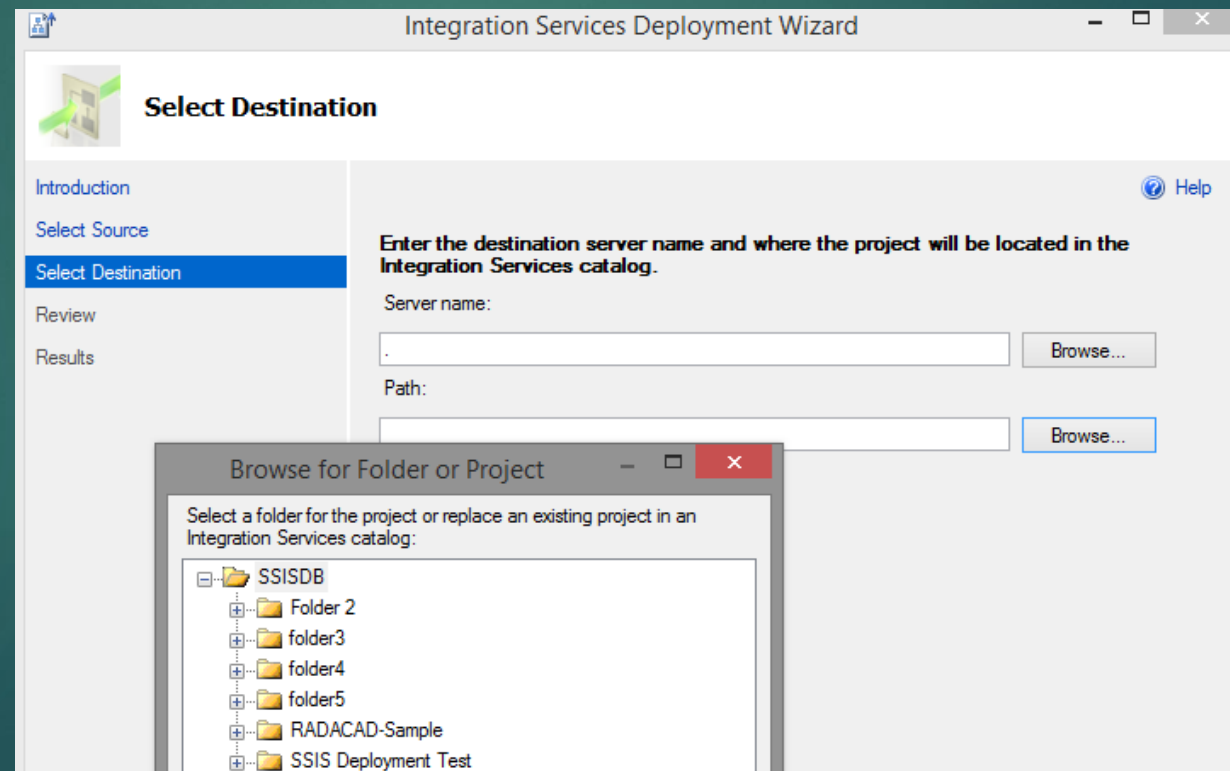
Comparison: Error Handling

| | 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: 77782222-1422-4672-4274-874222112244

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)

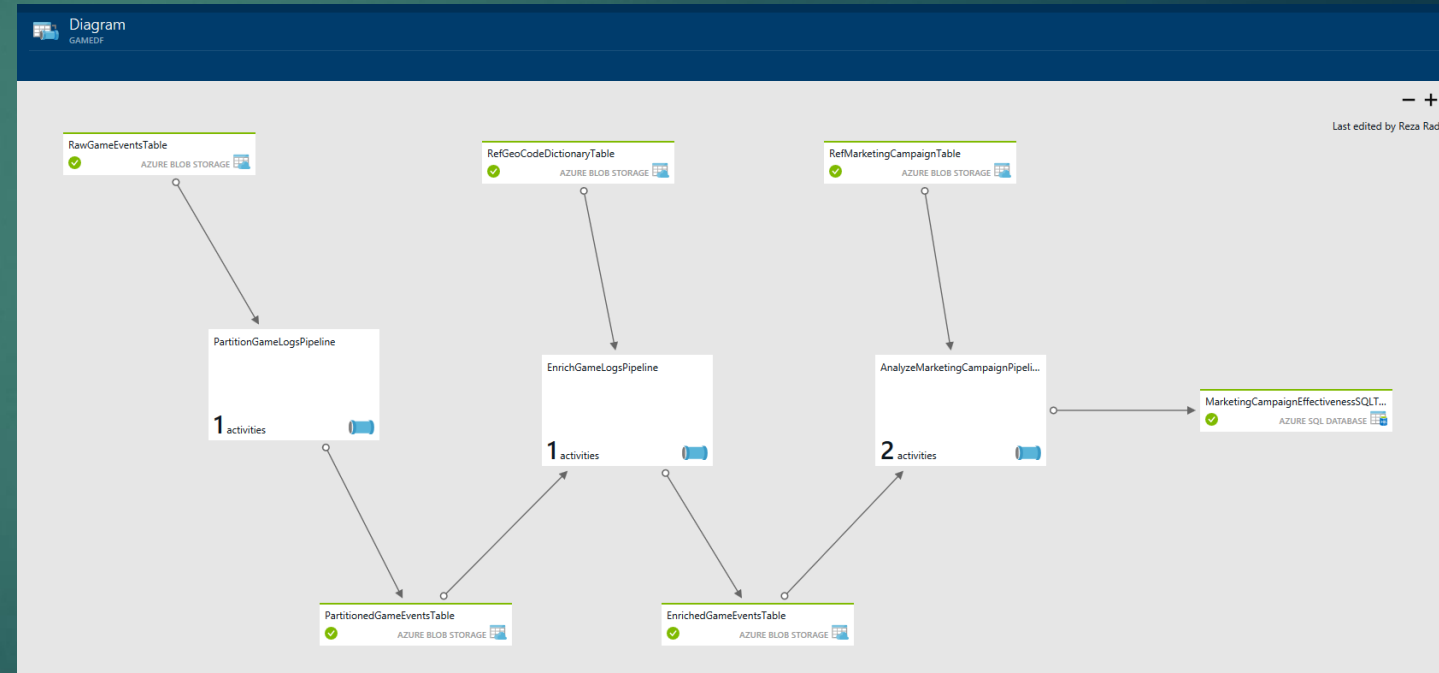
Execution Information

| | | | |
|----------|----------|-----------|----------|
| 0 | 0 | 5 | 0 |
| Failed | Running | Succeeded | 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



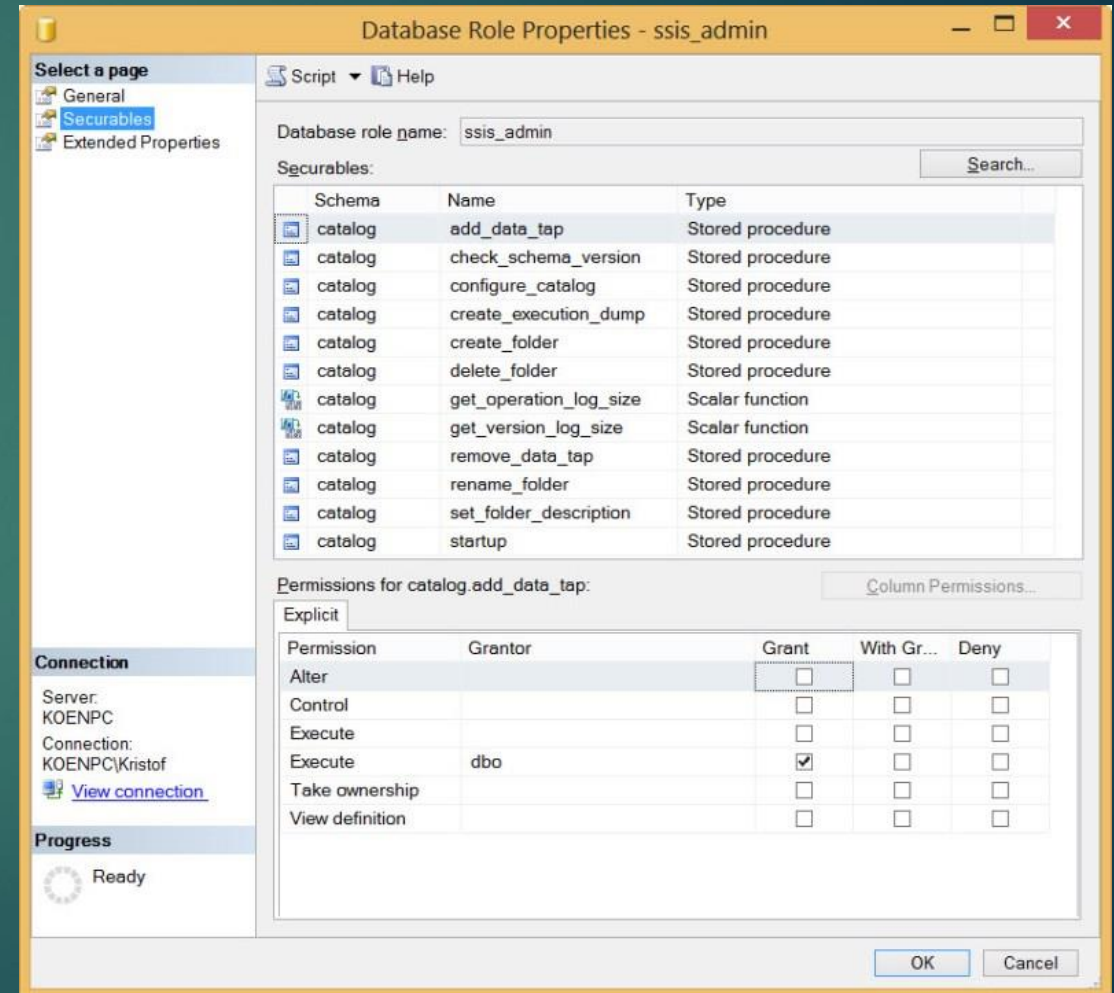
Comparison: Monitoring



| | 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...



Database Role Properties - ssis_admin

Select a page: General, **Securables**, Extended Properties

Database role name: ssis_admin

Securables:

| Schema | Name | Type |
|---------|------------------------|------------------|
| catalog | add_data_tap | Stored procedure |
| catalog | check_schema_version | Stored procedure |
| catalog | configure_catalog | Stored procedure |
| catalog | create_execution_dump | Stored procedure |
| catalog | create_folder | Stored procedure |
| catalog | delete_folder | Stored procedure |
| catalog | get_operation_log_size | Scalar function |
| catalog | get_version_log_size | Scalar function |
| catalog | remove_data_tap | Stored procedure |
| catalog | rename_folder | Stored procedure |
| catalog | set_folder_description | Stored procedure |
| catalog | startup | Stored procedure |

Permissions for catalog.add_data_tap:

Explicit

| Permission | Grantor | Grant | With Gr... | Deny |
|-----------------|---------|-------------------------------------|--------------------------|--------------------------|
| Alter | | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Control | | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Execute | | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Execute | dbo | <input checked="" type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| Take ownership | | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |
| View definition | | <input type="checkbox"/> | <input type="checkbox"/> | <input type="checkbox"/> |






Connection: Server: KOENPC, Connection: KOENPC\Kristof, View connection

Progress: Ready

OK Cancel

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 | |

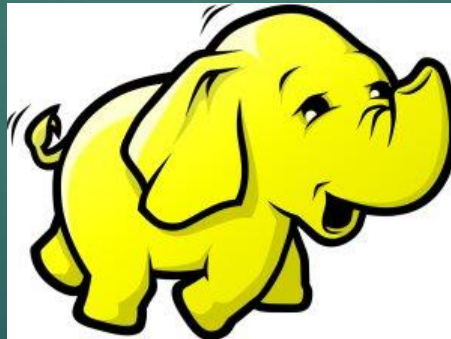
Comparison: Security

| | 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



Azure Analytics Integration

- ▶ Azure ML Batch Scoring Activity
- ▶ Data Lake Analytics U-SQL Activity



Comparison Summary

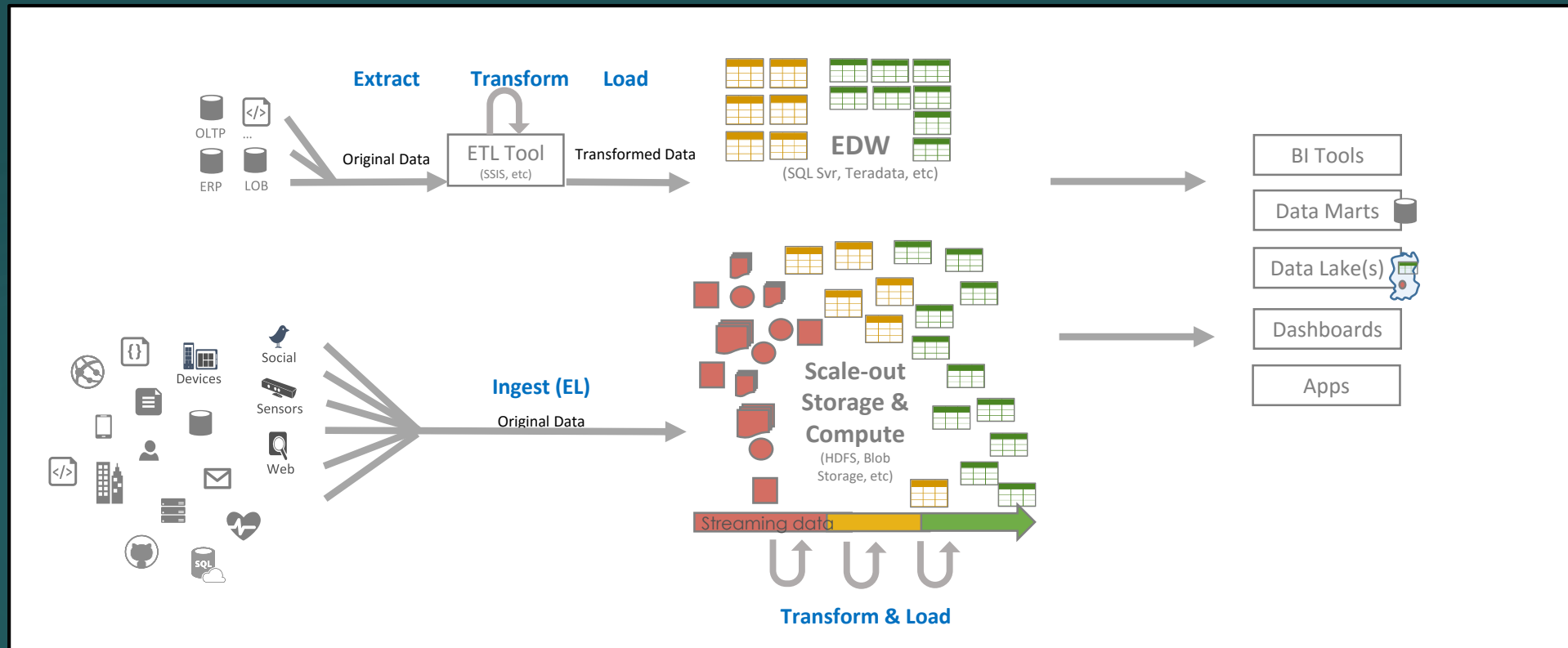
| | SSIS | ADF |
|--|------------------------|--|
| Development Tool | Yes | Yes (More Scripting Required) |
| Administration Tool | Yes | Yes (More Scripting Required) |
| User Type | Developer | Developer |
| Data Source and Destinations | Wide Range | Azure Storage, SQL Azure, SQL Server On Premises |
| 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

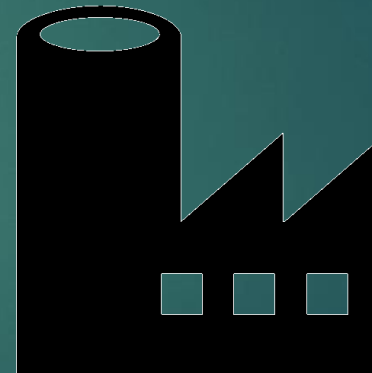
- ▶ These tools/services didn't built for the same purpose
- ▶ They are complement of each other

Evolving Approaches to Analytics



Get Benefits of Both

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



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

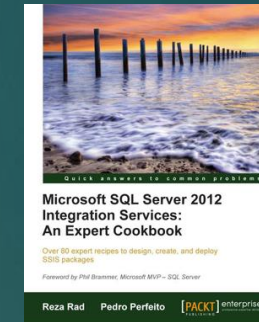
<http://azure.microsoft.com/en-us/documentation/services/data-factory/>

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>



Questions?

| | |
|---|---|
|  | @Rad_Reza |
| reza@radacad.com |  |
|  | www.radacad.com |
| www.linkedin.com/in/rezarad |  |