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

@Rad_Reza
reza@radacad.com
www.radacad.com/blog
www.linkedin.com/in/rezarad
Agenda

Azure Data Factory Overview
SSIS Overview
Comparison
Why This Session?
I was working with SSIS
I’ve heard about Azure Data Factory
Panic!

WE'RE ALL

GONNA DIE
Give it a go
Found it as a Great tool
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
ADF Editor

- Write JSON
- Edit/Create/Delete Objects
- Deploy
Main Benefits

- Pay as Usage
- Cloud Based
- HDInsight compatibility
- Built for Massive Data Movement
Demo: Azure Data Factory
Evolving Approaches to Analytics

- Extract
  - Original Data
  - ETL Tool (SSIS, etc)
- Transform
  - Transformed Data
- Load
  - EDW (SQL Svr, Teradata, etc)

- BI Tools
- Data Marts
- Data Lake(s)
- Dashboards
- Apps
Evolving Approaches to Analytics
Evolving Approaches to Analytics

Extract Original Data → Transform Original Data Using ETL Tool → Load Transformed Data to EDW (SQL Server, Teradata, etc.)

Ingest (EL) Original Data → Scale-out Storage & Compute (HDFS, Blob Storage, etc.) → Transform & Load

Final Output: BI Tools, Data Marts, Data Lake(s), Dashboards, Apps
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
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

<table>
<thead>
<tr>
<th></th>
<th>SSIS</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standalone Tool</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Powerful GUI</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Availability</td>
<td>Free</td>
<td>Free</td>
</tr>
<tr>
<td>Prerequisite</td>
<td>SQL Server License</td>
<td>Azure Subscription</td>
</tr>
</tbody>
</table>
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
The New Monitoring App

- Data Factory Monitoring App released 10 days ago!
## Comparison: Administration Tool

<table>
<thead>
<tr>
<th></th>
<th>SSIS</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standalone Tool</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Powerful GUI</td>
<td>Yes</td>
<td>Yes, few with Power Shell</td>
</tr>
<tr>
<td>Availability</td>
<td>Free</td>
<td>Under subscription</td>
</tr>
<tr>
<td>Prerequisite</td>
<td>SQL Server License</td>
<td>Azure Subscription</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>SQL Server</th>
<th>Oracle</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP</td>
<td>Salesforce</td>
</tr>
<tr>
<td>Access</td>
<td>Sybase</td>
</tr>
<tr>
<td>Informix</td>
<td>Teradata</td>
</tr>
<tr>
<td>MySQL</td>
<td>PostgreSQL</td>
</tr>
<tr>
<td>FoxPro</td>
<td>CSVs</td>
</tr>
<tr>
<td>Web Services</td>
<td></td>
</tr>
<tr>
<td>Many other Data Sources and Destinations types</td>
<td></td>
</tr>
</tbody>
</table>
Data Src/Dest: ADF

- Data Store Types Supported so far:
  - + Azure SQL Data Warehouse
  - + Azure Data Lake Store
  - +Web
  - +HDFS
  - +ODBC
  - +OData
# Comparison: Data Src/Dest

<table>
<thead>
<tr>
<th></th>
<th>SSIS</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL Server On Premises</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SQL Azure</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Azure Storage</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Oracle</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SAP</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>MySQL</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>DB2</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>CSVs</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Web Services</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Many other Data src/dest</td>
<td>Yes</td>
<td>Fewer</td>
</tr>
<tr>
<td>Azure Data Lake Store</td>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>
Data Transformations: SSIS

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

<table>
<thead>
<tr>
<th>Merge Join</th>
<th>Aggregate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Union</td>
<td></td>
</tr>
<tr>
<td>Lookup</td>
<td></td>
</tr>
<tr>
<td>Fuzzy Lookup/Grouping</td>
<td></td>
</tr>
<tr>
<td>Conditional Split</td>
<td></td>
</tr>
<tr>
<td>Derived Column</td>
<td></td>
</tr>
<tr>
<td>OLE DB Command</td>
<td></td>
</tr>
<tr>
<td>Many other Data Transformations</td>
<td></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>SSIS</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copy</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>C# Custom Transformations</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Pig and Hive</td>
<td>With Scripting</td>
<td>Yes</td>
</tr>
<tr>
<td>Azure ML Batch Scoring</td>
<td>With Scripting</td>
<td>Yes</td>
</tr>
<tr>
<td>Stored Procedure</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Built-in Data Transformations</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>
Extension & Automation: SSIS

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

- Power Shell
- No Programing SDK
- No third-party components YET!
Comparison: Extension & Automation

<table>
<thead>
<tr>
<th></th>
<th>SSIS</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programing SDK</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Automation</td>
<td>Yes, BIML</td>
<td>Yes, Power Shell</td>
</tr>
<tr>
<td>Third-Party components</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>REGION</th>
<th>LOW FREQUENCY</th>
<th>HIGH FREQUENCY</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Europe</td>
<td><strong>£0.3665 per activity per month</strong></td>
<td><strong>£0.6109 per activity per month</strong></td>
</tr>
<tr>
<td></td>
<td><strong>On-premises £0.9164 per activity per month</strong></td>
<td><strong>£1.5273 per activity per month</strong></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>REGION</th>
<th><strong>£0.1527 per hour</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud</td>
<td></td>
</tr>
<tr>
<td>On-premises</td>
<td><strong>£0.0611 per hour</strong></td>
</tr>
</tbody>
</table>

Inactive pipelines

**£0.4887/month**
## Comparison: Pricing

<table>
<thead>
<tr>
<th></th>
<th>SSIS</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Licensing</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Pay for Features</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Pay for Usage</td>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>SSIS</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware Setup</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Software Setup</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Administration Costs</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Azure Environment Usage</td>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>
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
Comparison: Error Handling

<table>
<thead>
<tr>
<th></th>
<th>SSIS</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alerts</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Error Logging</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Error Handling</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>
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
## Comparison: Deployment

<table>
<thead>
<tr>
<th></th>
<th>SSIS</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deployment GUI</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Deployment Scripts</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Corporate Deployment Scenario with automated scripting</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Monitoring: SSIS

- SSIS Logging
- SSIS Catalog Reports
Monitoring: ADF

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

<table>
<thead>
<tr>
<th>Feature</th>
<th>SSIS</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitoring GUI</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Drill through Monitoring</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Slice</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Data Lineage</td>
<td></td>
<td>Yes</td>
</tr>
</tbody>
</table>
Security: SSIS

- Role Based
- Roles for Deploy, Execute, Monitor...
Security: ADF

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

<table>
<thead>
<tr>
<th>Role Based Security</th>
<th>SSIS</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Feature</th>
<th>SSIS</th>
<th>ADF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Development Tool</td>
<td>Yes</td>
<td>Yes (More Scripting Required)</td>
</tr>
<tr>
<td>Administration Tool</td>
<td>Yes</td>
<td>Yes (More Scripting Required)</td>
</tr>
<tr>
<td>User Type</td>
<td>Developer</td>
<td>Developer</td>
</tr>
<tr>
<td>Data Source and Destinations</td>
<td>Wide Range</td>
<td>Azure Storage, SQL Azure, SQL Server On Premises</td>
</tr>
<tr>
<td>Data Transformations</td>
<td>Many</td>
<td>Custom Code required</td>
</tr>
<tr>
<td>Environment</td>
<td>Setup Required</td>
<td>Azure Supported</td>
</tr>
<tr>
<td>Pricing</td>
<td>Pay for Features</td>
<td>Pay for Usage</td>
</tr>
<tr>
<td>Environmental and Administration Costs</td>
<td>Yes</td>
<td>No Costs</td>
</tr>
<tr>
<td>Error Handling</td>
<td>Event Handers, Alerts</td>
<td>Alerts</td>
</tr>
<tr>
<td>Logging</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Deployment</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Monitoring</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Data Lineage</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Security</td>
<td>Role Based</td>
<td>Role Based</td>
</tr>
<tr>
<td>HDInsight Compatibility</td>
<td>Normal</td>
<td>High</td>
</tr>
<tr>
<td>Azure Analytics Integration</td>
<td></td>
<td>High</td>
</tr>
</tbody>
</table>
Hybrid
Purpose

- These tools/services didn’t built for the same purpose
- They are complement of each other
Evolving Approaches to Analytics

- Extract
- Transform
- Load

EDW (SQL Svr, Teradata, etc)

BI Tools

Data Marts

Data Lake(s)

Dashboards

Apps

Original Data

ETL Tool (SSIS, etc)

Transformed Data

Scale-out Storage & Compute

(HDFS, Blob Storage, etc)

Transform & Load

Streaming Data

Social

Sensors

Web

Devices

OLTP

LDB

OLAP
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
ADF comes as a Complementary service

Not to Compete with SSIS
Reference To Study

- Azure Data Factory Documentation

Azure Data Factory posts on my Blog:

- SSIS Book
Questions?

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