



Unleash the power of Azure Data Factory

Sergio Zenatti Filho,
Associate Director Data & Analytics,
Satalyst



Sergio Zenatti Filho

Associate Director Data & Analytics at Satalyst

I am Data and Analytics Director with over 16 years experience in the delivery of Business Intelligence and Analytics Solutions. I worked internationally around Australia, New Zealand and Brazil, in sectors that include Mining, Oil & Gas, Government, Healthcare, Financial Services, Telecom, Automotive and dairy. I enjoy learning new technology and help people to learn.

[in /sergiozenatti](#)

[@SergioZenatti](#)

<https://zenatti.net/>

SQL Saturday Perth - 2018

- 25th of August from 8:30 to 5pm;
- Tracks on Analytics, Big Data, SQL 2017, Power BI and more;
- We are looking for Speakers, Sponsors and Volunteers;
- Limited seats;
- Register today <http://www.sqlsaturday.com/761>

Session objectives

- Azure Data Factory (ADF) Overview
- Understand how ADF Works
- ADF v1, v2 and new features
- Understand ADF components
- Be able to deploy, build and manage a simple ADF implementation

Azure Data Factory

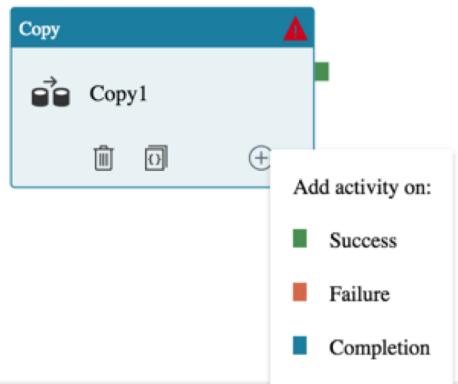
- Cloud-based data integration service. Currently ADF v1 (GA) and ADF v2 (Preview).
- Used to automate the movement and transformation of data.
- Connect to on-premises and cloud data sources.
- Integration with HDInsight, Azure Batch, Azure Functions, Logic Apps and more.
- Globally deployed data movement as a service (Integration Runtime).



ADF v1, v2 and new Features

- ADF v1 is designed for the batch data processing of time series data.
- ADF v2 is general-purpose hybrid data integration service with very flexible execution patterns.

ADF v2 – New Features



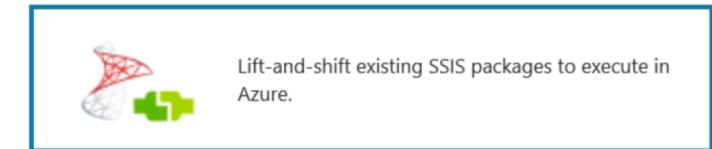
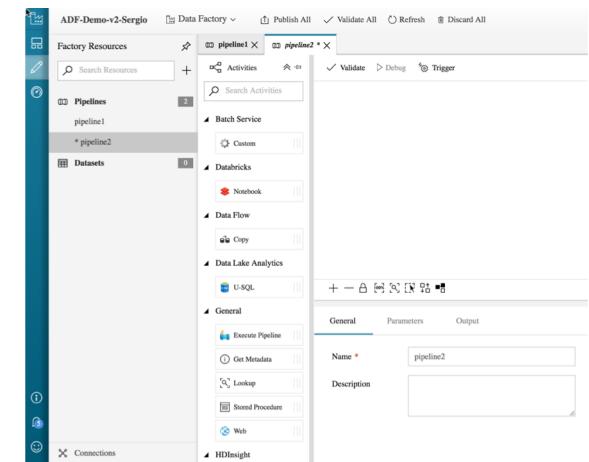
Use [expressions](#), [functions](#) or refer to [system variables](#).

▲ System Variables

- Data Factory Name
Name of the data factory the pipeline run is running within
- Pipeline Name
Name of the pipeline
- Pipeline run ID
ID of the specific pipeline run
- Pipeline trigger ID
ID of the trigger that invokes the pipeline
- Pipeline trigger name
Name of the trigger that invokes the pipeline
- Pipeline trigger time
Time when the trigger that invoked the pipeline. The trigger time is the actual fired...
- Pipeline trigger type
Type of the trigger that invoked the pipeline (Manual, Scheduler)

▲ Functions

- ▼ Expand All
- ▶ Collection Functions
- ▶ Conversion Functions



ADF Components

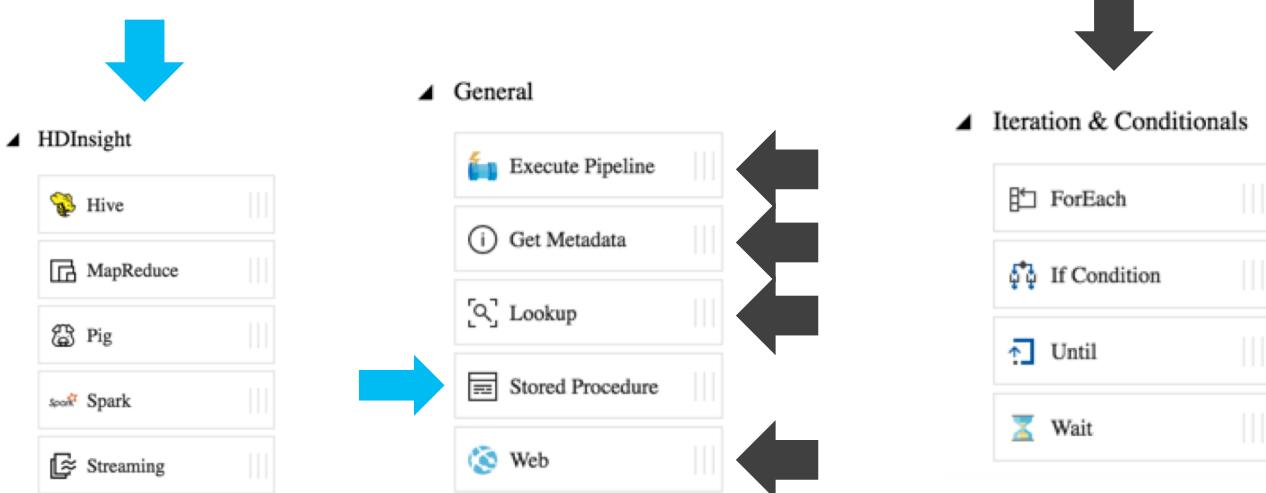


Activities Type:

- Data Movement
- Data Transformation
- Control

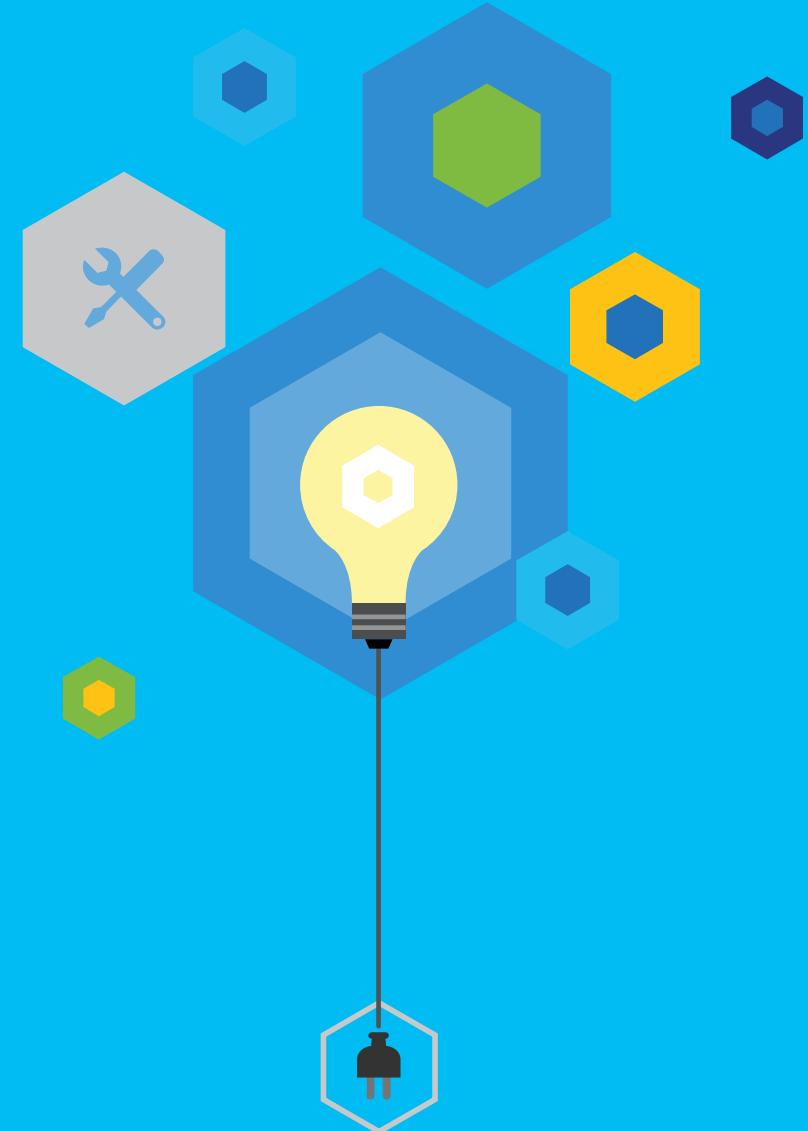
ADF Activities Options

- Data Movement
- Data Transformation
- Control



Create Data Factory & Pipeline

**Portal, PowerShell,
.NET, Python, REST
and ARM**



ADF – Real Life Scenario - Components

- Azure SQL Data Warehouse – Expensive to run if it is not using all the time



POST https://management.azure.com/subscriptions/{subscription-id}/resourceGroups/{resource-group-name}/providers/Microsoft.Sql/servers/{server-name}/databases/{database-name}/**pause**?api-version=2014-04-01-preview HTTP/1.1

GET https://management.azure.com/subscriptions/{subscription-id}/resourceGroups/{resource-group-name}/providers/Microsoft.Sql/servers/{server-name}/databases/{database-name}?api-version=2014-04-01 HTTP/1.1

POST https://management.azure.com/subscriptions/{subscription-id}/resourceGroups/{resource-group-name}/providers/Microsoft.Sql/servers/{server-name}/databases/{database-name}/**resume**?api-version=2014-04-01-preview
HTTP/1.1

ADF – Real Life Scenario - Components

- Azure Analysis Services – Process data model after data load

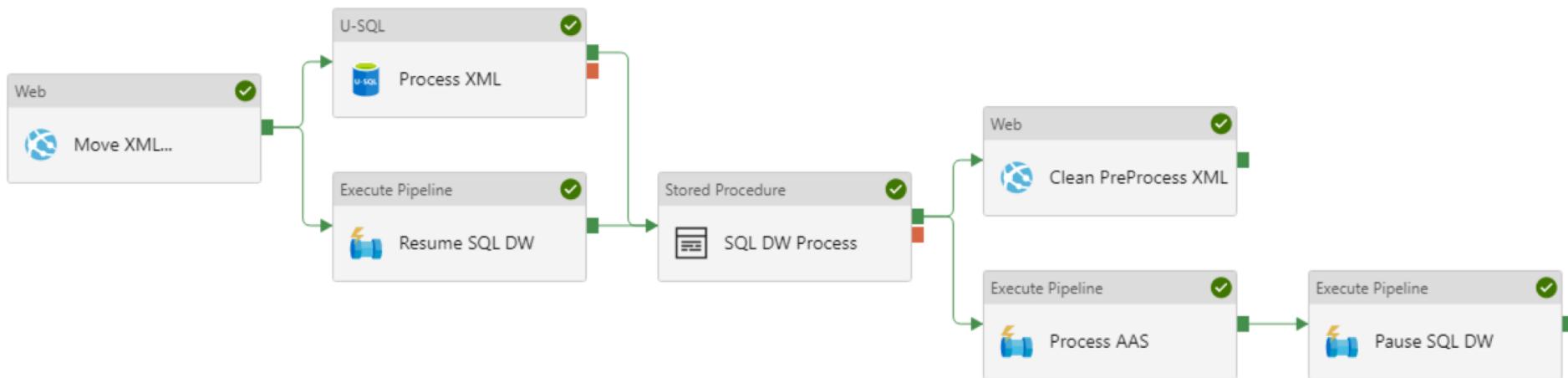


POST https://<region>.asazure.windows.net/servers/<my server>/models/<my db>/refreshes

GET https://<region>.asazure.windows.net/servers/<my server>/models/<my db>/refreshes/<run id>

ADF – Data Integration Scenario

- Process around 15k XML files a day with total size around 6-7 GB
- Infrastructure cost: max \$5k a month
- Process files once every hour
- ADF, Logic Apps, Azure Data Lake, Azure SQL DW and Azure Analysis Services



ADF – Learning

New Azure Data Factory self-paced hands-on lab for UI V2

<https://azure.microsoft.com/en-us/blog/new-azure-data-factory-self-paced-hands-on-lab-for-ui/>

Orchestrating Big Data with Azure Data Factory V1

<https://www.edx.org/course/orchestrating-big-data-azure-data-microsoft-dat223-3x-3>

Learning path for Azure Data Factory V1 and V2

<https://azure.microsoft.com/en-us/documentation/learning-paths/data-factory/>

ADF - Overview

- Azure Data Factory (ADF) Overview
- Understand how ADF Works
- ADF v1, v2 and new features
- Understand ADF components
- Be able to deploy, build and manage a simple ADF implementation
- Next step Start Using it ☺



 GLOBAL INTEGRATION BOOTCAMP

Thank you for your time!

Sergio Zenatti Filho
Data & Analytics Architect

 \sergiozenatti

 @SergioZenatti

zenatti.net

 Azure