



Exploring Data Integration Solutions with Azure Data Factory

Naveen Kumar M

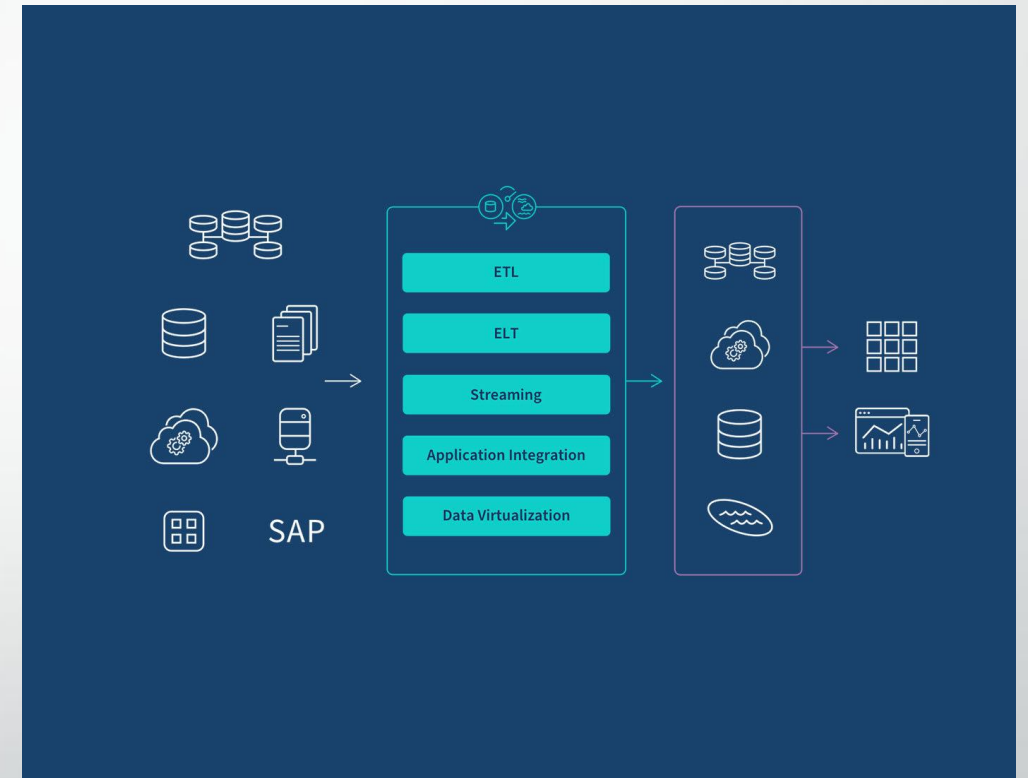
Data Engineer | C# Corner MVP | Blogger

Agenda

- Understanding Data Integration
- Introduction to Azure Data Factory
- Components of Azure Data Factory
- Demo
 - Data Movement and Transformation in Azure Data Factory
 - Monitoring and Management
- Real-world Use Cases
- Q&A

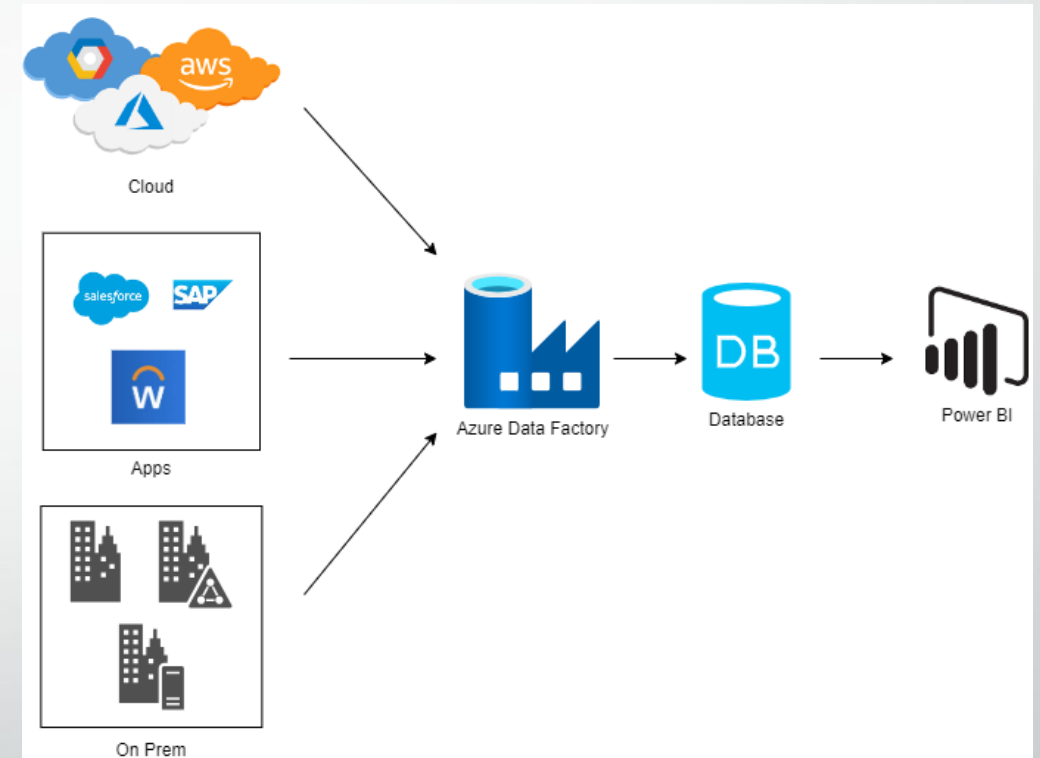
Understanding Data Integration

- Process of combining data from different sources into a unified view.
- Transforming it into a consistent format that can be analyzed and used effectively.
- Making it accessible and valuable for various business purposes.



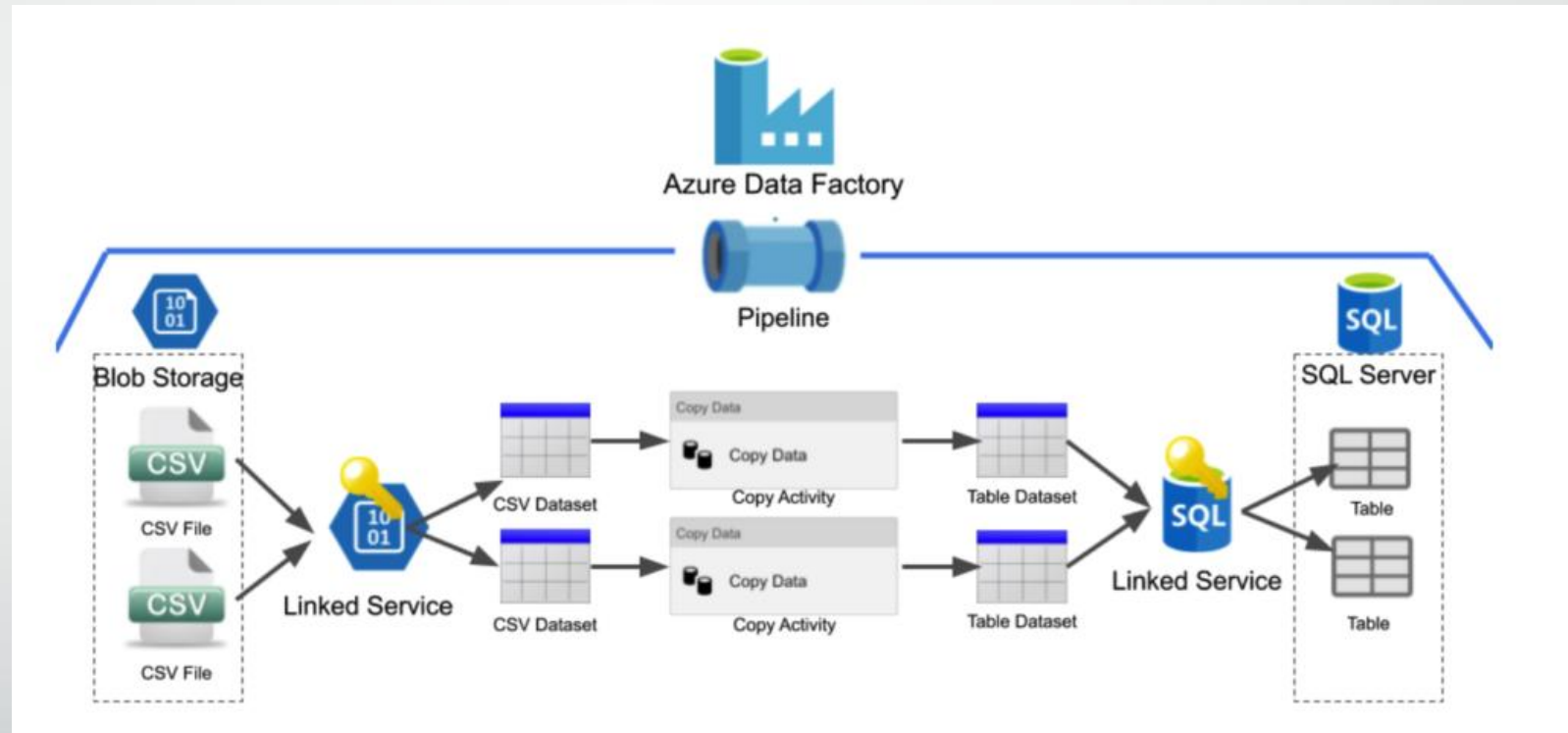
Introduction to Azure Data Factory

- Cloud bases ETL as a Service.
- Allows to create, schedule, and manage data-driven workflows.
- Enables seamless movement of data between various data stores, both on-premises and in the cloud.

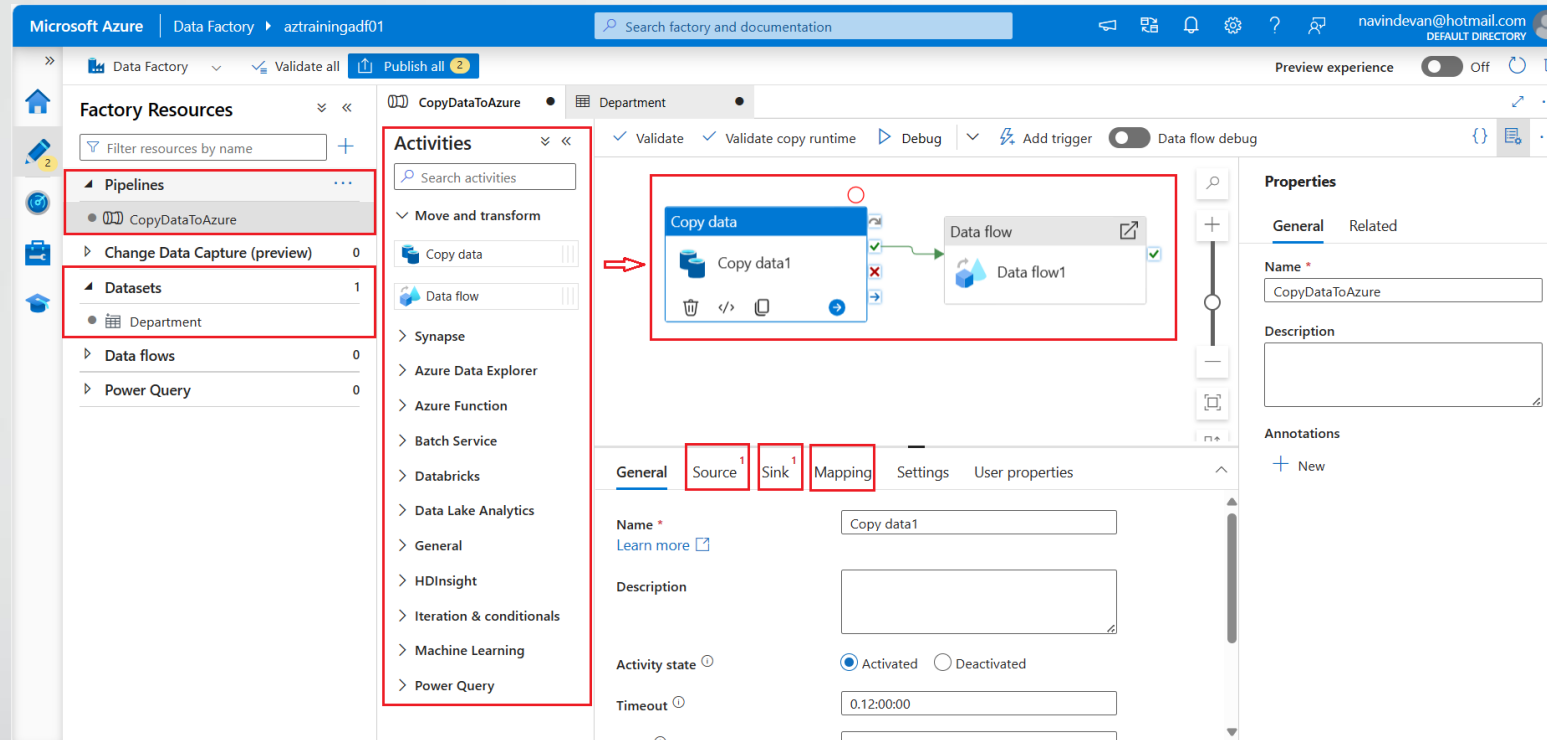


Components of Azure Data Factory

- Pipelines
- Activities
- Linked Services
- Datasets
- Triggers
- Integration Runtime



Components of Azure Data Factory Cont..



- **Pipelines** – It's the logical groups of activities that perform a specific job.
- **Activities** – An activity is just like a logical operation or the action that we perform on our data.
- **Datasets** – Datasets represent designated perspectives of data, serving as pointers or references to the specific data required for your tasks, whether as inputs or outputs.

Components of Azure Data Factory Cont..

- **Linked Services** – Linked services are a very important component to link your data store to the ADF.

The screenshot displays the Azure Data Factory portal interface. On the left, the navigation pane shows the 'Linked services' option highlighted. The main area shows the 'Linked services' page with a '+ New' button and a table of existing services. A red box highlights the '+ New' button and the table. A red arrow points from the '+ New' button to the 'New linked service' dialog on the right. The dialog shows a search bar and a grid of service options, including Amazon Marketplace Web Service, Amazon RDS for Oracle, Amazon RDS for SQL Server, Amazon Redshift, Amazon S3, and Amazon S3 Compatible.

Linked services

Linked service defines the connection information to a data store or compute. [Learn more](#)

[+ New](#)

Filter by name Annotations : Any

Showing 1 - 7 of 7 items

Name	Type
AZFileSystem	Azure Blob Storage
AZMyWorks	Azure SQL Database
AzureBlobStorage1	Azure Blob Storage
AzureBlobStorage2	Azure Blob Storage
AzureBlobStorage3	Azure Blob Storage
AzureSqlDatabase1	Azure SQL Database
OnPermFileSys	File system

New linked service

Data store Compute

Search

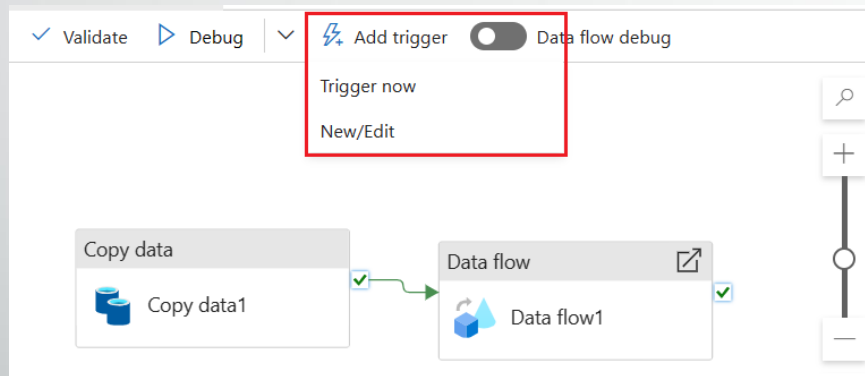
All Azure Database File Generic protocol NoSQL Services and apps

Amazon Marketplace Web Service Amazon RDS for Oracle Amazon RDS for SQL Server Amazon Redshift Amazon S3 Amazon S3 Compatible

Continue Cancel

Components of Azure Data Factory Cont..

- **Triggers** – The triggers within ADF offer an alternative method for initiating a pipeline run.



The 'New trigger' dialog box is shown. It has a 'Name' field with the value 'trigger1' and a 'Description' field. The 'Type' dropdown is set to 'Schedule' and is highlighted with a red box. Below the dropdown, there is a 'Filter...' field. The 'Schedule' section is expanded, showing options for 'Tumbling window', 'Storage events', and 'Custom events'. The 'Every' field is set to '15' and the unit is 'Minute(s)'. There is a checkbox for 'Specify an end date'. The 'Annotations' section has a '+ New' button. The 'Start trigger' section has a checkbox for 'Start trigger on creation' which is checked. At the bottom, there are 'OK' and 'Cancel' buttons.

Components of Azure Data Factory Cont..

- **Integration Runtime** - Integration Runtime in Azure Data Factory is a managed compute infrastructure used for data integration and data transfer scenarios.
 - Azure Integration Runtime
 - Self Hosted

Microsoft Azure | Data Factory | aztrainingadf01

Search factory and documentation

navindevan@hotmail.com
DEFAULT DIRECTORY

» Data Factory | Validate all | Publish all

General

- Factory settings
- Connections
- Linked services
- Integration runtimes**
- Microsoft Purview
- Source control
- Git configuration
- ARM template
- Author
- Triggers
- Global parameters
- Data flow libraries
- Security
- Credentials
- Customer managed key
- Outbound rules
- Managed private endpoi...

Integration runtimes

The integration runtime (IR) is the compute infrastructure to provide the following data integration c

+ New Refresh

Filter by name

Showing 1 - 2 of 2 items

Name	Type	Sub-type	Status
AutoResolveIntegrationR...	Azure	Public	✓ R
OnPermSFIR	Self-Hosted	---	✓ R

Integration runtime setup

Network environment:

Choose the network environment of the data source / destination or external compute to which the integration runtime will connect to for data flows, data movement or dispatch activities:

- Azure**
Use this for running data flows, data movement, external and pipeline activities in a fully managed, serverless compute in Azure.
- Self-Hosted**
Use this for running activities in an on-premises / private network
[View more](#)

External Resources:

You can use an existing self-hosted integration runtime that exists in another resource. This way you can reuse your existing infrastructure where self-hosted integration runtime is setup.

- Linked Self-Hosted**
[Learn more](#)

Continue Back Cancel

Data Movement and Transformation in ADF

- Get data integration with more than 90 built-in connectors
- Copy Activity
- Data Flow Activity
- HDInsight Activities



Monitoring and Management

Pipeline runs

Triggered Debug Rerun Cancel Refresh Edit columns List Gantt

Search by run ID or name Pacific Time (US & C... : Last 7 days Status : All Runs : Latest runs Add filter Copy filters

Showing 1 - 21 items

<input type="checkbox"/> Pipeline name	Run start ↑↓	Run end	Duration	Triggered by	Status	Run	Par
<input type="checkbox"/> S3ToDataLakeCopy	11/5/20, 6:00:18 AM	11/5/20, 6:03:50 AM	00:03:32	12HourTrigger	✓ Succeeded	Original	
<input type="checkbox"/> DatabricksJarPipeline	11/4/20, 6:04:11 PM	11/4/20, 6:10:31 PM	00:06:20	DayTrigger	✓ Succeeded	Original	
<input type="checkbox"/> S3ToDataLakeCopy	11/4/20, 6:00:18 PM	11/4/20, 6:03:59 PM	00:03:40	12HourTrigger	✓ Succeeded	Original	
<input type="checkbox"/> S3ToDataLakeCopy	11/4/20, 6:00:19 AM	11/4/20, 6:04:09 AM	00:03:50	12HourTrigger	✓ Succeeded	Original	
<input type="checkbox"/> DatabricksJarPipeline	11/3/20, 6:03:35 PM	11/3/20, 6:11:14 PM	00:07:39	DayTrigger	✓ Succeeded	Original	

Microsoft Azure | UxDemoFactory

Pipeline runs

Triggered **Debug** Rerun Cancel Refresh Edit columns

End time : Last 24 hours (8/27/20 11:32 AM - 8/28/20 11:32 AM) Time zone : Pacific Time (US & Canada) (UT... Status : All status Runs : Latest runs

Showing 1 - 2 items

<input type="checkbox"/> Pipeline name	Run Start (PST) ↑↓	Duration	Status ↑↓	Triggered by	Error
<input type="checkbox"/> S3ToDataLakeCopy	8/28/20, 11:33:02 AM	00:00:02	🔄 In progress	Manual trigger	
<input type="checkbox"/> SalesAnalyticsMLPipeline	8/28/20, 11:10:08 AM	00:01:03	✗ Failed	Manual trigger	🗨

Real-world Use Cases

- Utilized to facilitate data migration for advanced analytics initiatives.
- Offers the capability to streamline the ETL process, transitioning from SQL Server Integration Services to additional data sources.
- Provides a solution for extracting data from client servers or online sources and transferring it to an Azure Data Lake through pipeline creation to manage data flow.
- Serves as a versatile tool for executing various data integration tasks and is recognized as one of the leading ETL solutions.
- Enables the integration of data from diverse ERP systems into Azure Synapse for reporting purposes, with support for analysis and reporting via Power BI.
- The ADF tools boast seamless integration, enabling rapid development of ETL, big data, data warehousing, and machine learning solutions, while maintaining the flexibility to evolve and adapt to changing or enhanced requirements.



Let's connect on LinkedIn



Naveen Kumar M

Data Engineer | SQL / ETL Lead |
C# Corner MVP

