Azure Data Factory for Beginners: A Teacher's Guide

Welcome to a beginner's guide to Azure Data Factory! This presentation will cover the basics of this powerful tool for building and managing data pipelines in the cloud.

by Jafer Negery

What is Azure Data Factory?

Cloud-based platform

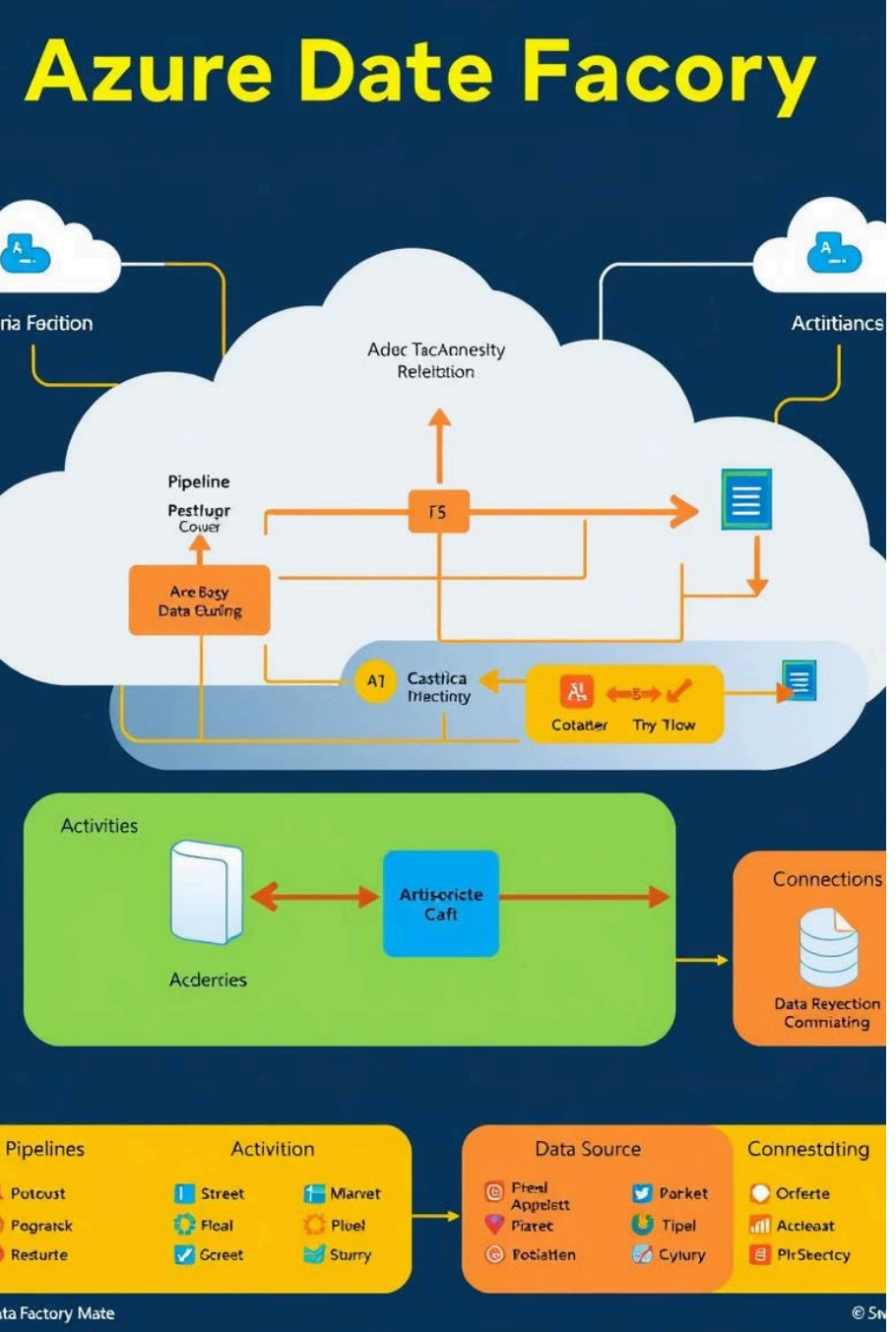
Azure Data Factory is a fully managed cloud service that allows you to create and orchestrate data pipelines.

Data integration

It enables the movement and transformation of data from various sources to target locations, such as data warehouses, data lakes, and databases.

Scalable and reliable

Built for scalability and reliability, it can handle large volumes of data and complex data flows.



Key Components of Azure Data Factory

1 Pipelines

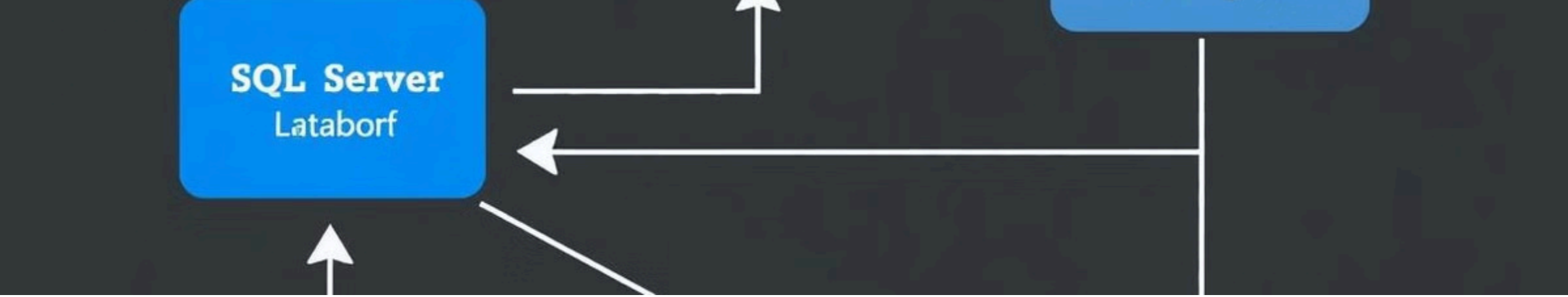
Pipelines are the fundamental building blocks of Azure Data Factory. They define the sequence of activities that execute to move and transform data.

2 Activities

Activities are the building blocks of a pipeline. They perform specific tasks like copying data, transforming data, or validating data.

3 Connections

Connections define how your Azure Data Factory connects to your data sources and sinks, allowing data movement and transformations.



Connecting to Data Sources

Databases

Connect to various databases, including SQL Server, Azure SQL Database, and MySQL.

Cloud Storage

Connect to cloud storage services like Azure Blob Storage and Azure Data Lake Storage.

File Systems

Connect to local file systems and remote file systems like Azure File Share and FTP servers.

Designing Data Pipelines

1

Data Sources

Define the source of your data, specifying the connection, location, and format.

2

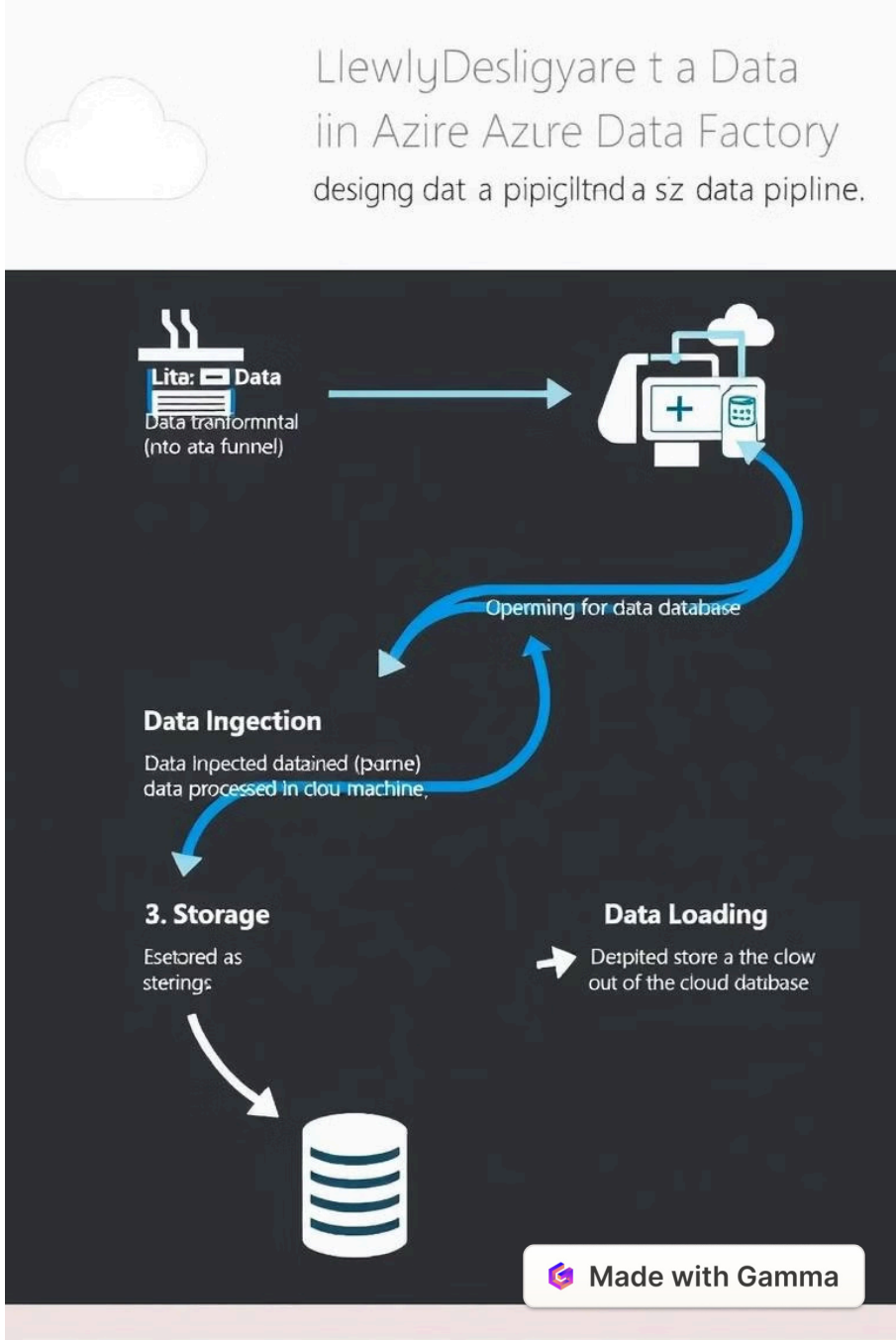
Data Transformations

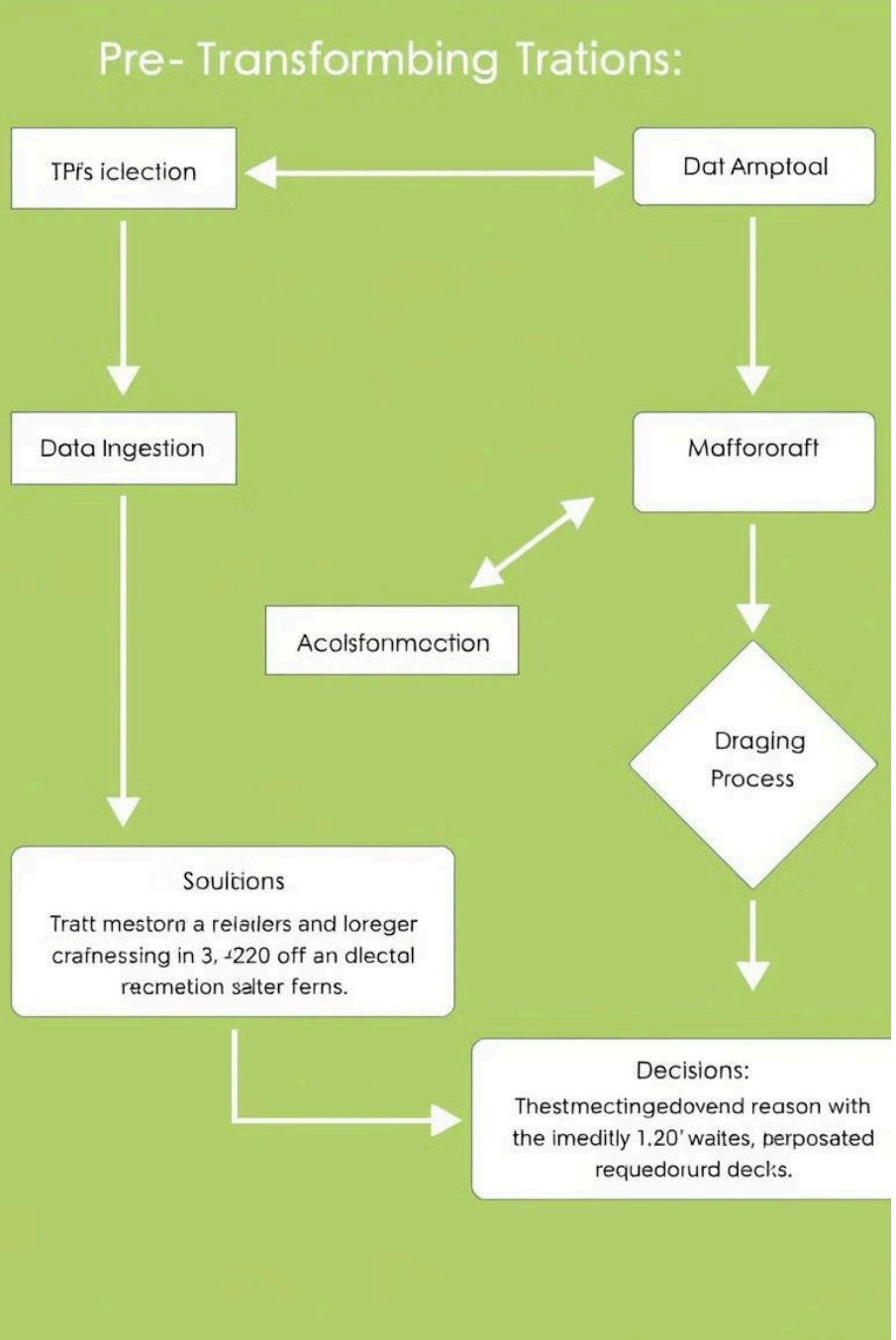
Apply transformations to your data using activities like Copy, Lookup, and Data Flow.

3

Data Targets

Choose the target location for your transformed data, specifying the connection and format.





Transforming Data with Activities

Copy Activity

Moves data from one location to another, supporting various data formats and transformations.

Lookup Activity

Retrieves data from a source and adds it to the pipeline using a query or filter.

Data Flow Activity

A powerful visual tool for transforming data using expressions, functions, and data flow components.

Scheduling and Monitoring Pipelines

1

Triggering Pipelines

Schedule your pipelines to run on a specific schedule, such as daily, hourly, or on demand.

2

Monitoring Performance

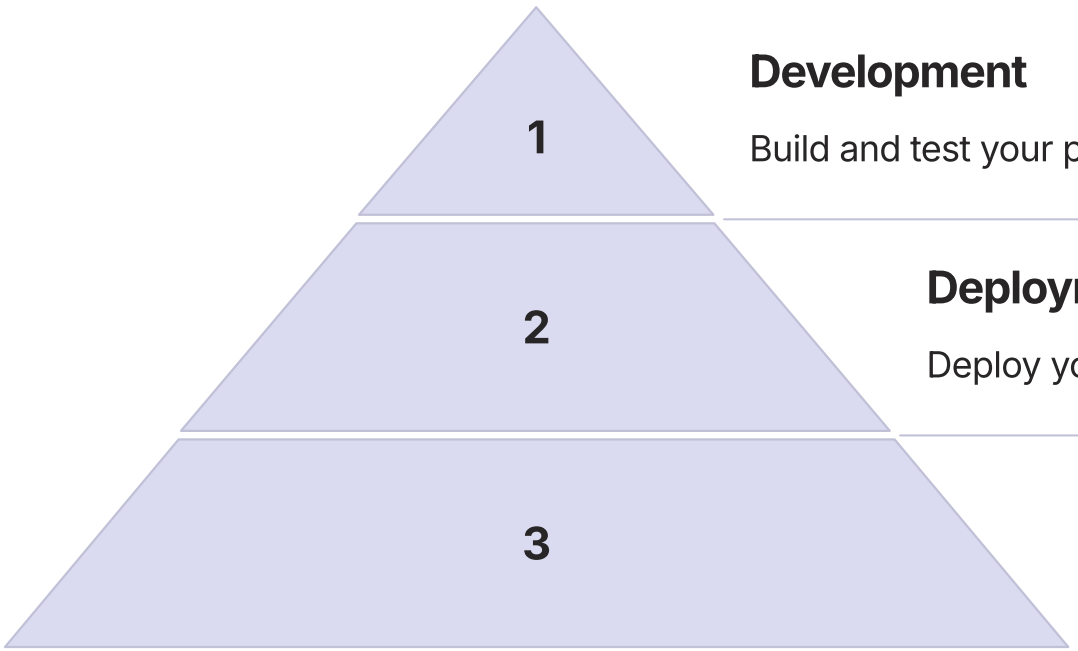
Track the performance of your pipelines, including execution time, data volume, and errors.

3

Alerting and Notifications

Set up alerts to notify you of pipeline failures or performance issues.

Deploying and Managing Pipelines



Development

Build and test your pipelines locally.

Deployment

Deploy your pipelines to Azure Data Factory.

Management

Manage your pipelines in Azure Data Factory.

Security and Governance Considerations

1

Authentication

Secure access to data sources.

2

Authorization

Control user permissions.

3

Data Masking

Protect sensitive data.

4

Auditing and Logging

Track pipeline activities.

Common Use Cases and Best Practices

1

ETL

Extract, transform, and load data.

2

ELT

Extract, load, and transform data.

3

Data Integration

Integrate data from multiple sources.

4

Data Quality

Validate and clean data.