



# **Tiger Data Fabric**

Aug 2022



# **Tiger Data Fabric - Introduction**



"Tiger's Accelerator for Implementing Data Lake & Data Fabric on Microsoft Azure in a highly scalable manner with powerful Governance Capabilities at it's core".

### **AGILITY WITH SELF-SERVICE**

Cut-down the time it takes for setting up new Data Pipelines from Days & Weeks to Few Minutes.

#### **MEDALLION DATA LAKE**

휳

## SUPPORT MULTIPLE DATA SOURCES

Supports Extracting and Ingesting Massive Volumes of Multi-Structured Data from RDBMS, File System, SFTP & different Cloud Storages.

#### **TRANSFORMATION CAPABILITIES**

Capabilities to perform Transformation **Operations** like data cleansing, validation, quarantine, profile & merge.

**Highly Scalable & Service Oriented** Architecture – developed using different

### **GOVERNANCE CAPABILITIES**

In-built capabilities for metadata & lineage Cataloging, Logging, Notification & Operational Reporting.

# **POWERED BY:** 💀 🔇 🌔 宁 回 🚺 🍦 실

### SUPPORTED INTERFACES:

3. Web UI **REST API (JSON)** Chatbot / VA 4. **EXCEL Upload** 2.

# 



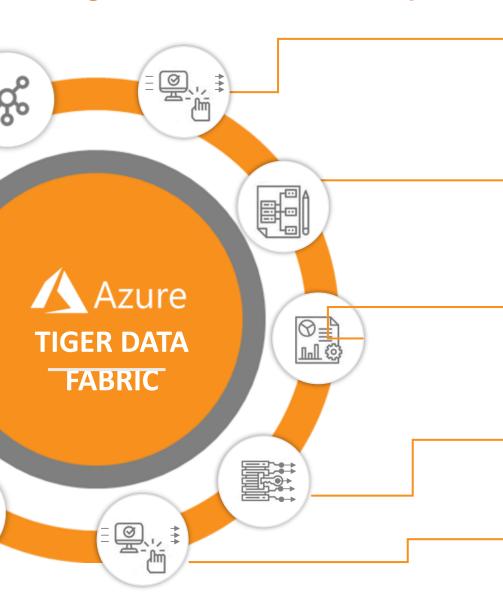
based on Medallion Data Lake Architecture (Bronze 2 Silver 2 Gold).

### SCALABLE CLOUD ARCHITECTURE



Azure PaaS & Open Source Technologies.

# Tiger Data Fabric - Key Features



### SELF SERVICE DATA LAKE MANAGEMENT

- Highly scalable, End to end Data Lake management accelerator developed using Azure PaaS and Open Source technologies.
- Powerful Self-Service capabilities for setting up new data ingestion pipelines & transformation feeds, perform data life cycle management and monitor operations using Chatbot.

### **CONFIGURATION DRIVEN EXTRACTION & INGESTION PIPELINES**

- Configurable extraction and ingestion pipelines , that can be reused for various data sources
- Supports Data extraction from File, RDBMS sources with Full, Incremental modes using Truncate and Reload, Upsert and Append merge strategies
- Support for Event , Schedule triggers

### **AUTOMATED DATA QUALITY & PROFILING**

- Solution for implementing configuration driven data testing and automated data profiling to surface data quality issues and assess data quality confidence.
- Library of more than 50+ testing rules and ability to add more custom rules.
- Integrated DQ Reporting and alerting capabilities.

### SERVICE FABRIC BUILT USING API

- Highly scalable, reusable services that can talk to each other using API
- Flexible, Robust API's that makes integration between various services easier

### **GOVERNANCE CAPABILITIES**

 Provide Data / Platform governance capabilities like Data catalogues, metadata management, Data lineage etc.

# Data Pipeline Development Lifecycle – Without Accelerator

Data Ops

 Understand Dataset Metadata (Format, Channel, Table, Columns, Data type, Primary Key, Business Key, possible acquisition strategy

**Identify** the data

source & dataset.

1.

1. Configure & Test DB Connections with secured connection string credentials.

etc.).

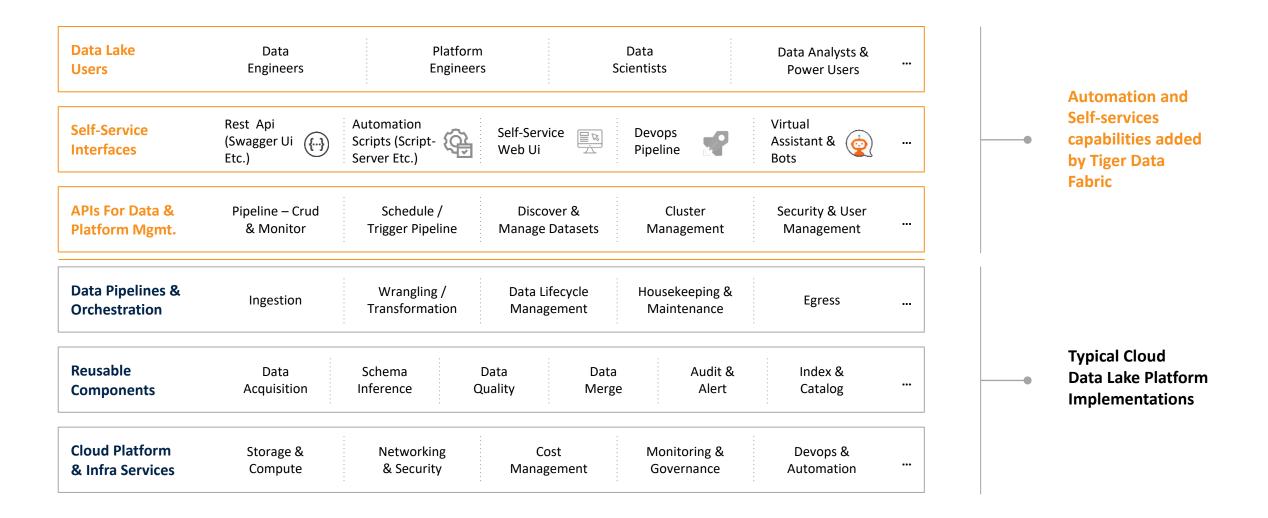
- 4. Prepare **Target** structure in Data Lake (Pipeline Metastore entries, Hive Database and Tables, Data Lake Folder structures etc.)
- Understand consumption pattern to design data partitioning strategy.
- 4. Define transformation logic (data cleansing, validation, profiling, standardization, merge etc. ).

- 7. Implement governance features to improve maintainability & support (logging, alert & notification, checkpointing etc.).
- 7. Unit Test all the Pipeline code (ADF, Databricks, SQL Database, ADLS, Synapse etc.).
- 7. Schedule & run the pipelines in dev environment.

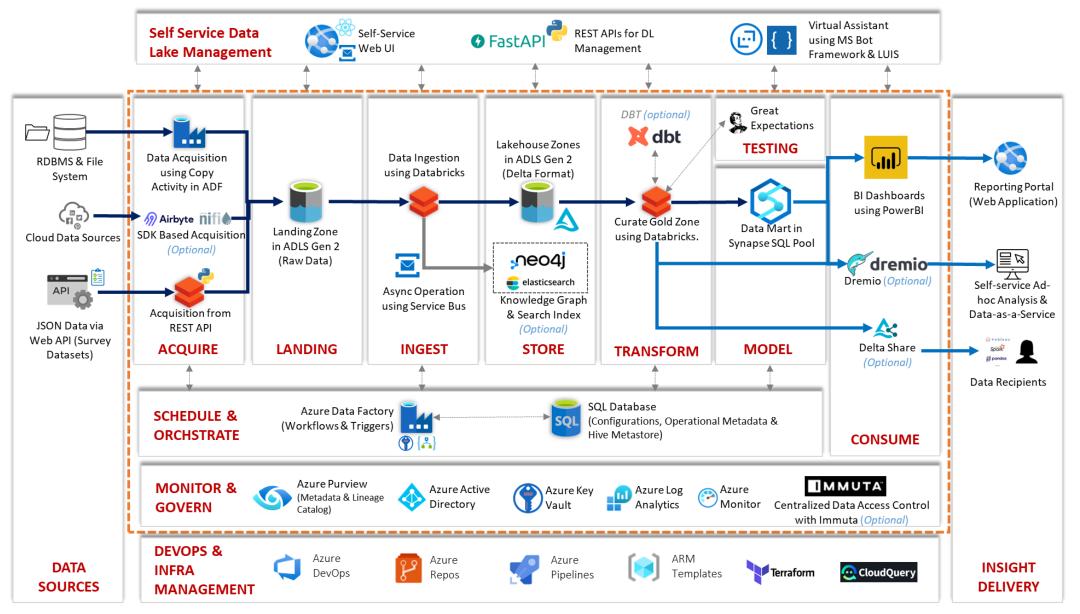
- 10. Deploy, Monitor and Test the pipelines in higher environments like QA and UAT / Staging.
- 10. Deploy to **Production** and monitor.
- Perform various data
   lifecycle management
   & house-keeping
   activities like clean-up
   in case failure, restore,
   archive, reprocess
   error records etc.

#### CAN TAKE FROM DAYS TO WEEKS

# How Self-Service is Enabled in Tiger Data Fabric



# High Level Architecture – Tiger Data Fabric (Azure)



# Integrated Knowledge Graph Based Data Catalog

