baufest

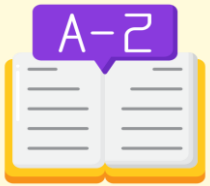
Azure Databricks Implementation

databricks

EVOLVING . BUSINESS . TOGETHER

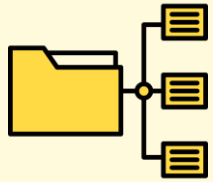
BENEFITS OF AN INDUSTRY MODEL

baufest



Unique business glossary and metadata

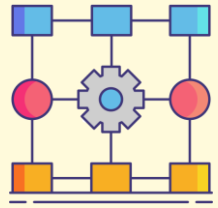
Models with integrated, normalized and consistent data



Single source for developing operational and analytical models

Agreement on definitions of KPIs and business metrics

Complete traceability from data sources to each data layer defined in the data architecture



Integrated Model to normalize any information and consolidate the global vision

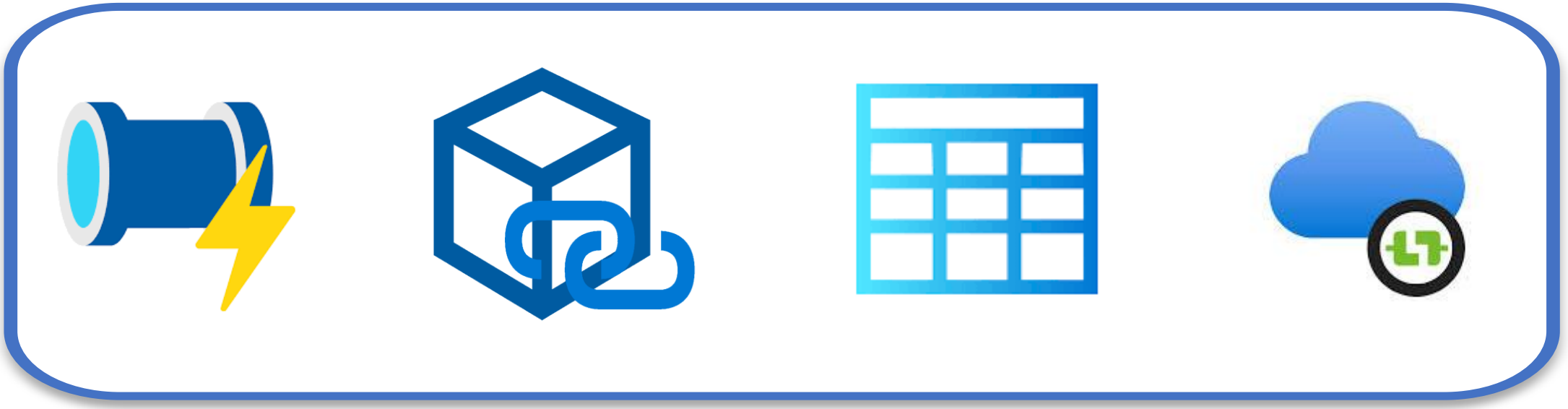
Single global data model

Comparisons

ADF	Databricks	ADF/Databricks
▪ General management of E2E. ▪ E2E parameterization (e.g. Servers, DBs, tables, schemas, queries, secrets, among others). ▪ Reprocessing on demand ▪ Management of sending the E2E results. ▪ Parameterization of dynamic values in triggers. (Parameterization of loading date, etc.) ▪ Automatic query assignment	• Record validation process in the queries to be processed. • Dynamic allocation of the Computing Resource based on the volume of the query (Small, Medium, Large).	▪ Automatic reprocessing in case of error identification (e.g. Concurrency, Structure Changes, among others). ▪ Storage of statistical data. ▪ Data storage that allows process traceability. ▪ Error management that allows tracking them.

Data Factory - Orquestador

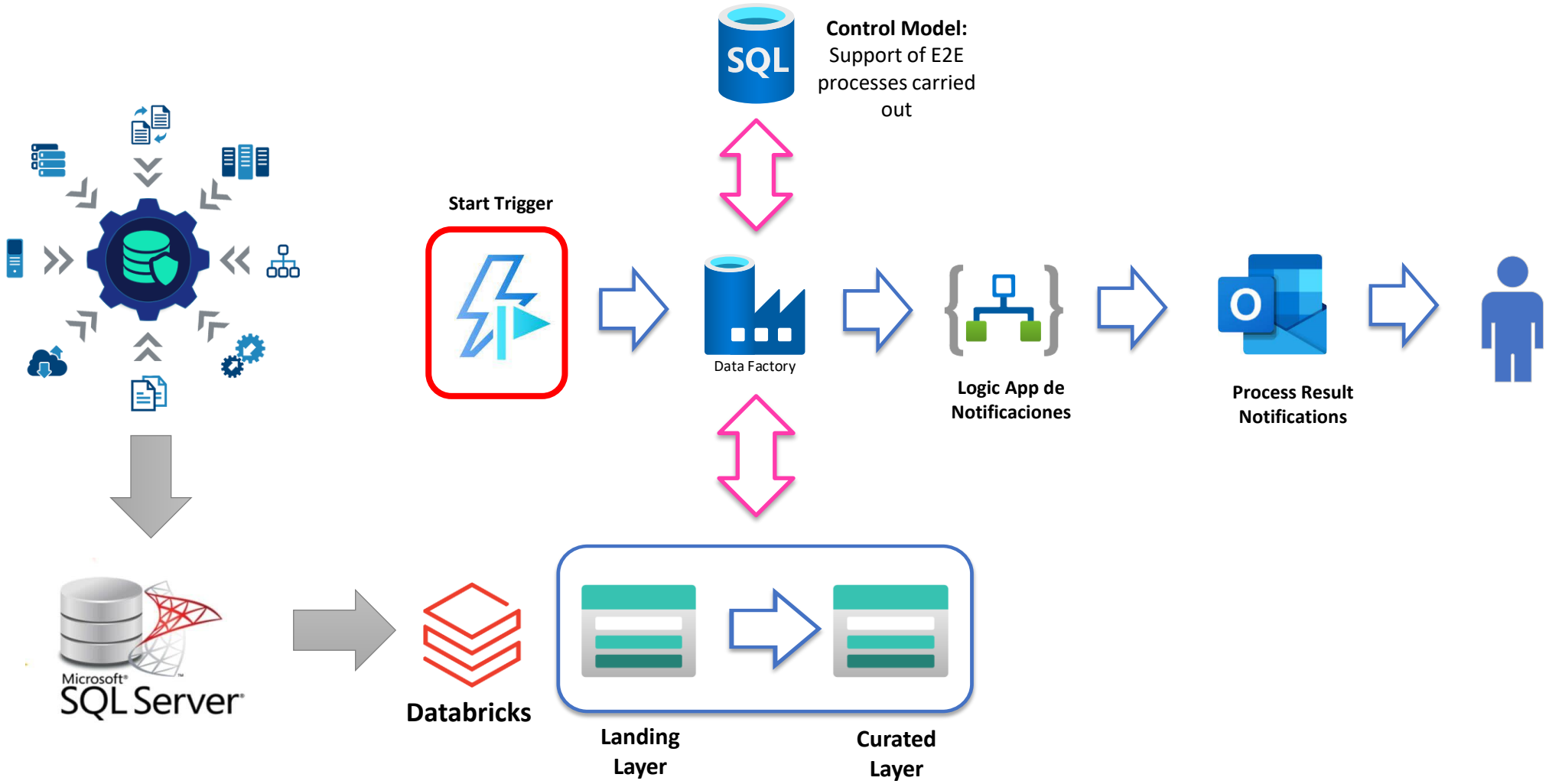
Azure Data Factory allows data loading orchestration processes to be carried out through the following components: pipelines, Linked Services, DataSets and associated Integration Runtimes



Databricks - Orchestrator

Azure Databricks allows E2E execution of data loading through Workflows (commonly called Jobs), Notebooks and associated Clusters.

Prototype service diagram



CI/CD DATABRICKS

CI/CD best practices have been incorporated based on our experience