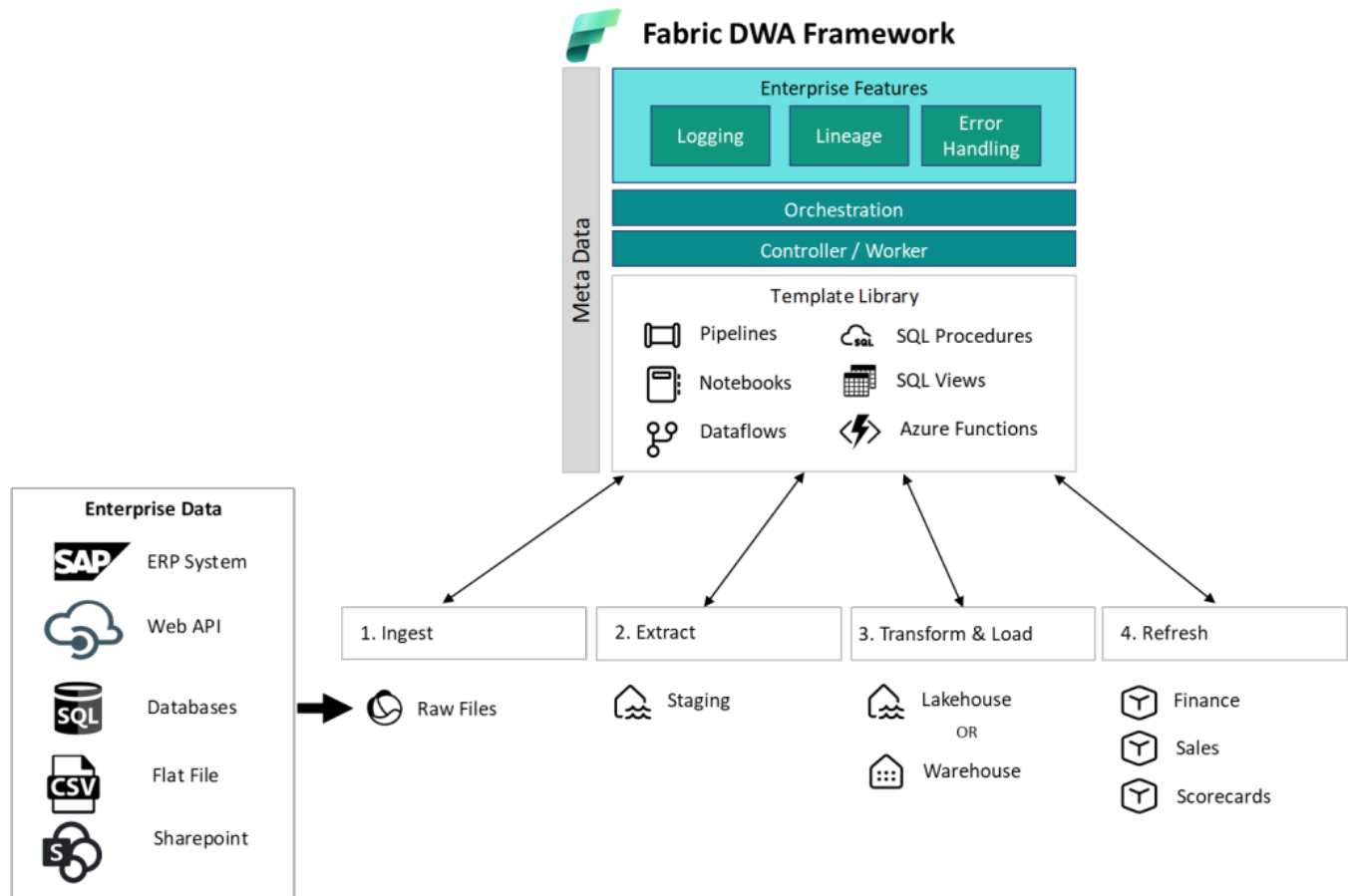


Fabric DWA Framework

On 15th November 2023 we launched our Fabric DWA Framework Accelerator. This is the result of over 3,500 hours of work migrating our Synapse DWA Framework to Fabric and also re-designing the template library system to make it extensible enough to handle any and all Fabric artefacts like Pipelines, Dataflows, Notebooks as well as traditional SQL Procedures and Views.

This framework not only saves months of development time by delivering an out of the box data warehouse accelerator but also is highly extensible and agile due to its re-use of templates to guarantee high supportability and flexibility.



Videos to explore concepts

Fabric DWA Introduction



10 Minute Introduction and Architecture

Fabric DWA 20 Min Demo



20 Minute demo of Templates and concepts

Top Fabric DWA Features

1. Fabric Data Engineering Templates

We supply pre-built and tested python notebook and pipeline templates for the most common formats and end points such as:

- CSV
- Excel
- SharePoint
- Database
- Parquet
- SFTP
- Web API

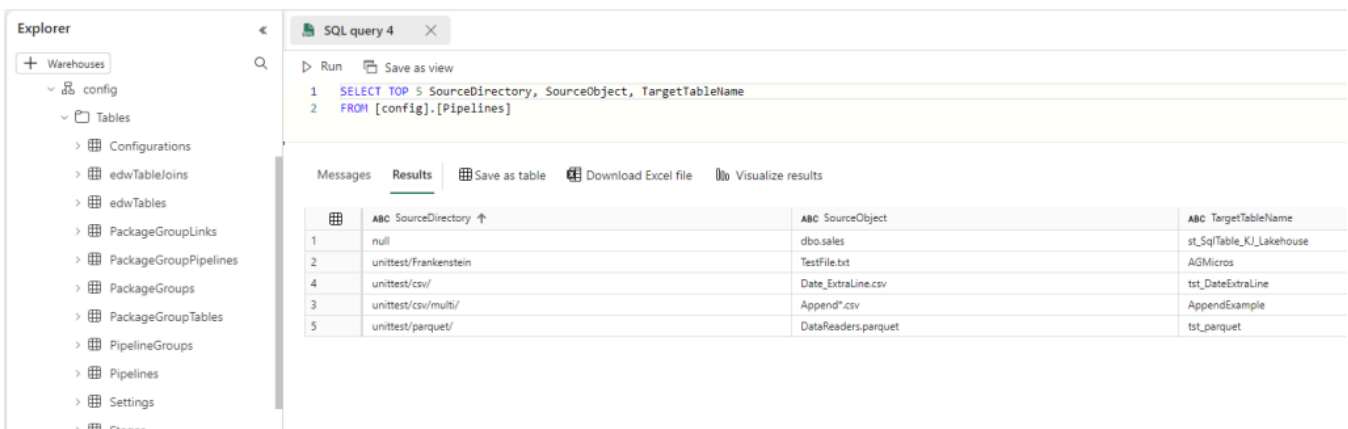
We supply blank starter templates to extend for any other data source supported by Fabric or python. Additionally we have a long list of ERP and other solutions we have integrated with such as:

- SAP ERP
- Infor M3
- Oracle ERP
- Oracle Fusion PPM
- Workday
- Greenhouse
- Exchequer
- Agresso
- QuickBooks

2. Meta Data Driven Declarative Platform

Once you have a declarative template, all development is meta data driven. This allows data movement in the ingest and extract phases to be fully automated without any need for development.

Our table design allow for inheritance, so that common properties such as format and location can be defined for a group of pipelines, helping to eliminate any repetition (or DRY as the Architects would say)



The screenshot shows a SQL query editor with the following query:

```
1 SELECT TOP 5 SourceDirectory, SourceObject, TargetTableName
2 FROM [config].[Pipelines]
```

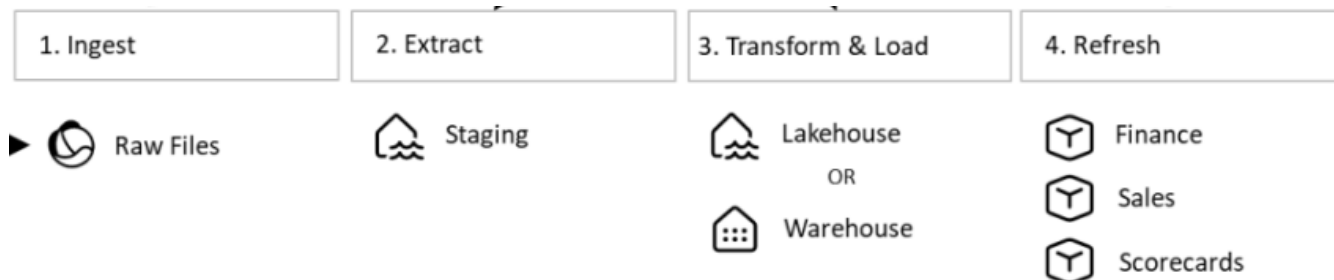
The results are displayed in a table with the following columns: SourceDirectory, SourceObject, and TargetTableName.

	ABC SourceDirectory ↑	ABC SourceObject	ABC TargetTableName
1	null	dbo.sales	st_SqlTable_KJ_Lakehouse
2	unittest/Frankenstein	Testfile.txt	AGMicros
4	unittest/csv/	Date_ExtraLine.csv	tst_DateExtraLine
3	unittest/csv/multi/	Append*.csv	AppendExample
5	unittest/parquet/	DataReaders.parquet	tst_parquet

3. Flexible Architecture

Microsoft Fabric offers a lot of choice in DW Architecture, especially around tools used to transform and query data. We support:

- Both Lakehouse and Datawarehouse engines. Our standard reference architecture uses a hybrid approach with lakehouse for staging data and datawarehouse for presenting and transforming a star schema.
- Full tool support for Notebooks, Pipelines, Dataflows, Azure functions and SQL. We recommend focusing on notebooks and SQL to minimise complexity and productivity.



4. World Class Extensibility

While the framework ships with many standard notebook templates, most projects have some non-standard formats which need a fresh template developed in any of the supported artefacts. These custom templates can be plugged into the Framework simply by registering meta data without any refactoring or code changes.

5. Star Schema Automation with Data Virtualisation

Moving from raw staged data to a clean star schema perfectly suitable for analytics is what we call the T and the L part of ETL. (Transform and Load).

While the framework supports using any template to make star schemas we also supply an out of the box solution for star schema automation from staging tables. This largely comprises using SQL Views to implement business mappings and calculations with automation for core function such as:

- Creating Physical Star Schema Tables
- Schema change and migration
- Business to Surrogate key handling
- Efficient loading using a wide range of approaches

A quick 10 minute demo of this process is shown below

Bob Duffy : Star Schema Automation in EDW/SqlPool (Dedicated SQL Pool)



- **Logging** on key activities. We log meta data activities to SQL tables, and also provide analytical templates for using Fabric logs to monitor pipeline performance.
- **Password Management.** All secrets are stored in Azure KeyVault. Allowing secure audited and policy based access.
- **Data Lineage.** Fabric itself support schema lineage to show how data flows from notebook to data warehouse to Power BI. We support full data lineage which tracks every row back to its source files for full auditability.

7. Dynamic Orchestration

Many DW solutions use hard coded orchestration to sequence activities such as what files to extract and load for a give business unit. Often this is developed in Pipelines.

The DWA Framework supports fully dynamic orchestration by grouping pipelines into units we call **package groups**. A package group can be dynamically edited to include/exclude pipelines without any code changes or code deployments.

8. Scalable and Parallel

All templates can be configured to control whether they run in sequence or parallel via their metadata. This allows for any amount of scale out that Fabric can supply to be harnessed.

9. Build 100% on Microsoft Fabric

The DWA Framework is simply an accelerator that is pre-built using Fabric components to install in your workspace, and brings your our 30 years experience as a head start in your project.

There is no license for any additional technology or or other licensing except for Fabric required.

10. Full EDW Lifecycle coverage

The DWA Framework doesnt just cover ETL it covers all the stages of a DW such as:

- Ingest
- Extract
- Transform
- Load
- Refresh (Power BI)
- Export
- Egress

11. Its Free !

We don't charge for the Prodata DWA Framework. Merely for our consultancy services to help install, configure, and train you in its use, and if required, any build support.

Related Blog Posts on DWA

EDW Automation Architecture

This blog explores a universal architecture for EDW automation. I'm, attempting to be technology agnostic, so little mention of specifics here. Albeit those who know me, know I am a SQL centric fan boy with a software engineering and Finance background. There are many different technologies for building data warehouses. Currently I see three emerging ... Continue reading



0

The need for DW Automation (1/3)

Many folks build star schema data warehouses and the supporting ecosystem of Semantic Models, Business Intelligence and/or ML Analytics. This is a quick note on how to assess how you are with automation, regardless of what tools you are using: SSIS,ADF, SQL, data bricks, Synapse, or non Microsoft tools. I recently ran a twitter thread ... Continue reading



2

A DW Automation Maturity Model (2/3)

In a previous blog I talked about the need for DW Automation in order to achieve that holy grail of Agility, Extensibility and perhaps Robustness with our data architecture. This blog is now focusing on some different levels of maturity. What's different about this maturity model and other ones is that its not necessary better ... Continue reading



1

DW Automation demo for Synapse SQL Pools (3/3)

This is the third blog in a series looking at the need for DW Automation focusing on Synapse SQL Pools (dedicated). In the first blog we looked at why we need DW Automation tools and what they look like:<https://prodata.ie/2021/09/14/the-need-for-dw-automation-1-3/> In the second blog we looked at a maturity model for DW automation with six levels ... Continue reading



1

Related Prodata Videos on DWA

EDW Automation in Synapse SQL Pools



Edit
