MAPPING DATA FLOWS

1

2

3

4

5

INTRODUCTION

Overview of Mapping Data Flows

SCENARIOS

Common Use Cases for Mapping Data Flows

USER INTERFACE

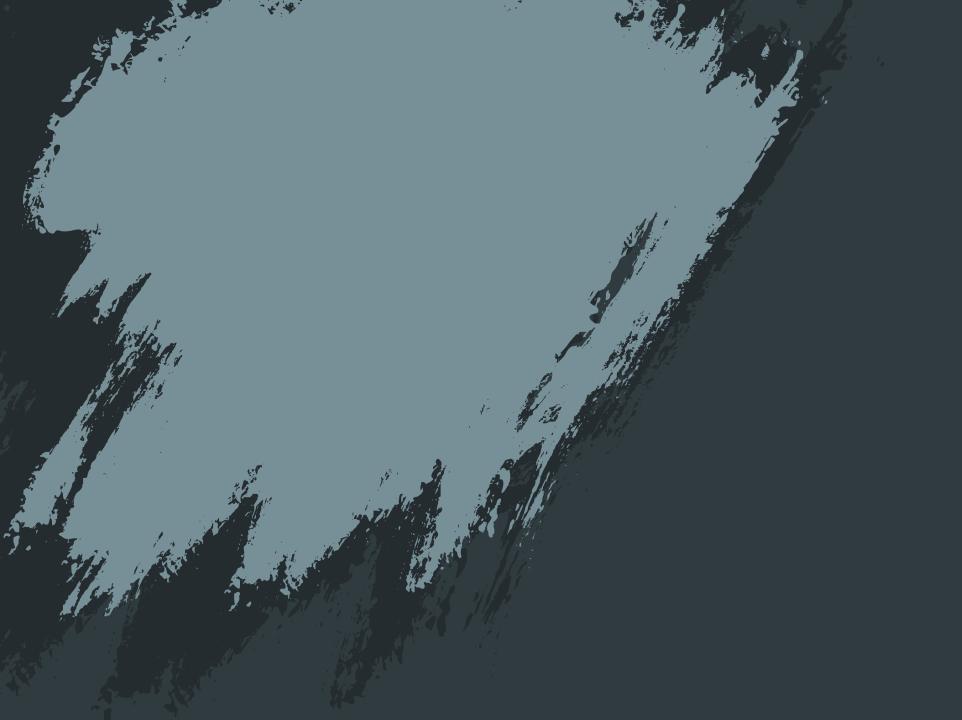
User Interface of Mapping Data Flows

DATA TRANSFORMATION

Implementing a pipeline with mapping data flows

PERFORMANCE

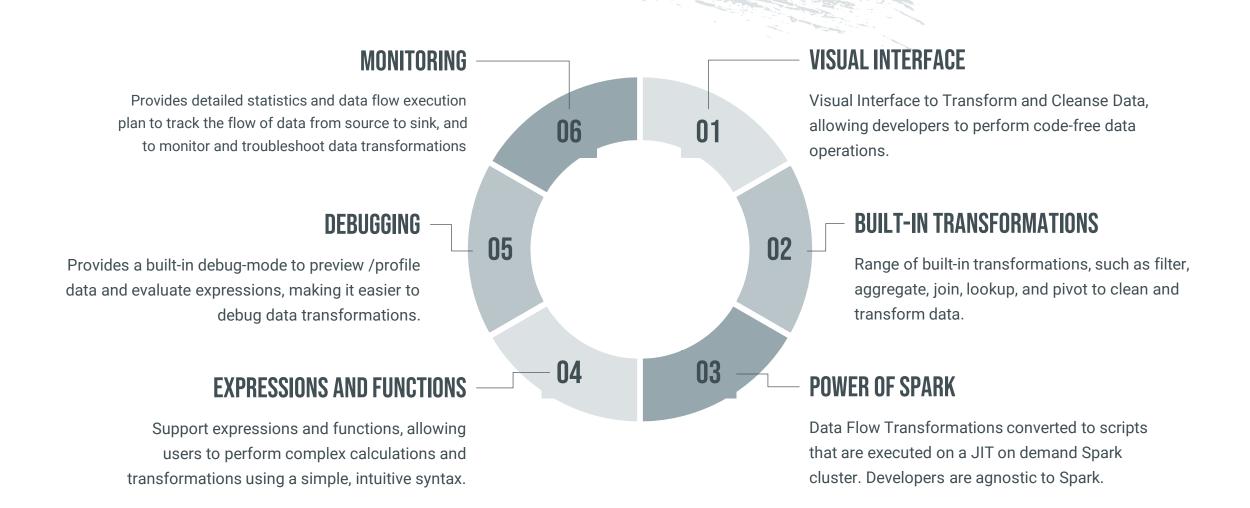
Data Flow Performance

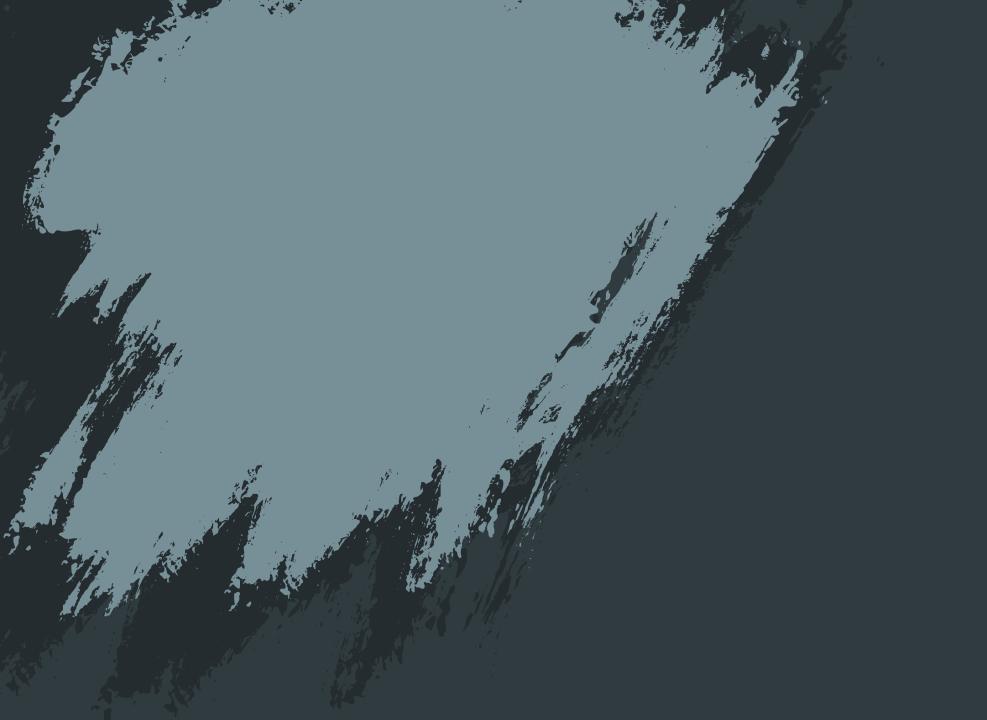


SECTION 1 INTRODUCTION

INTRODUCTION

Overview of Mapping Data Flows

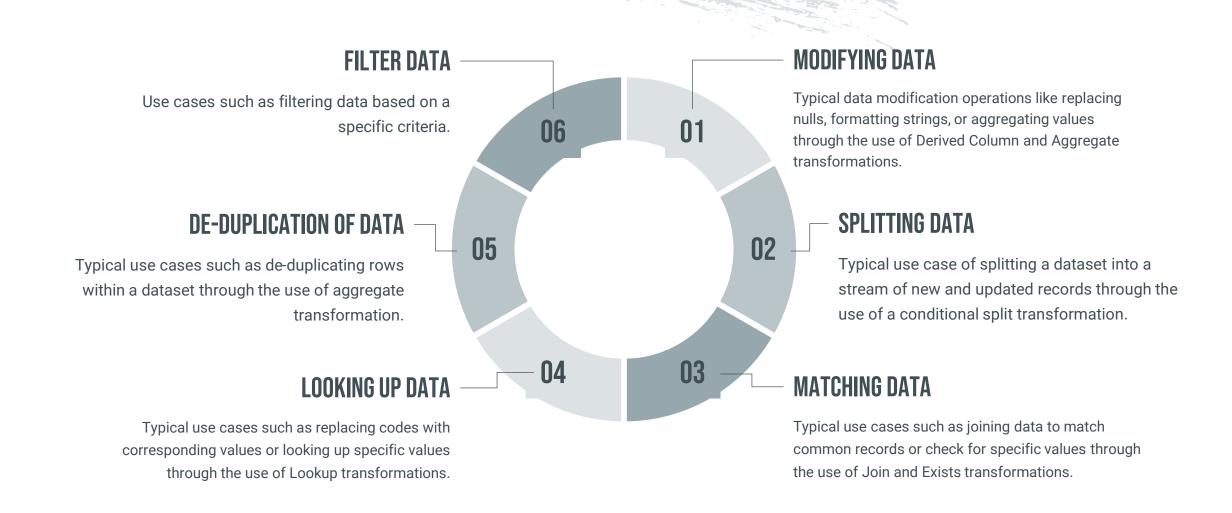




SECTION 2 SCENARIOS

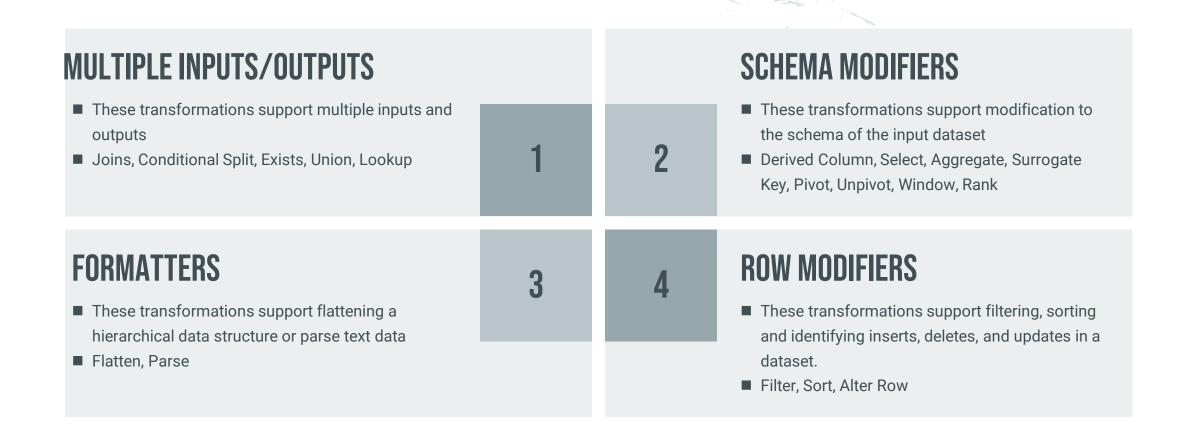
SCENARIOS

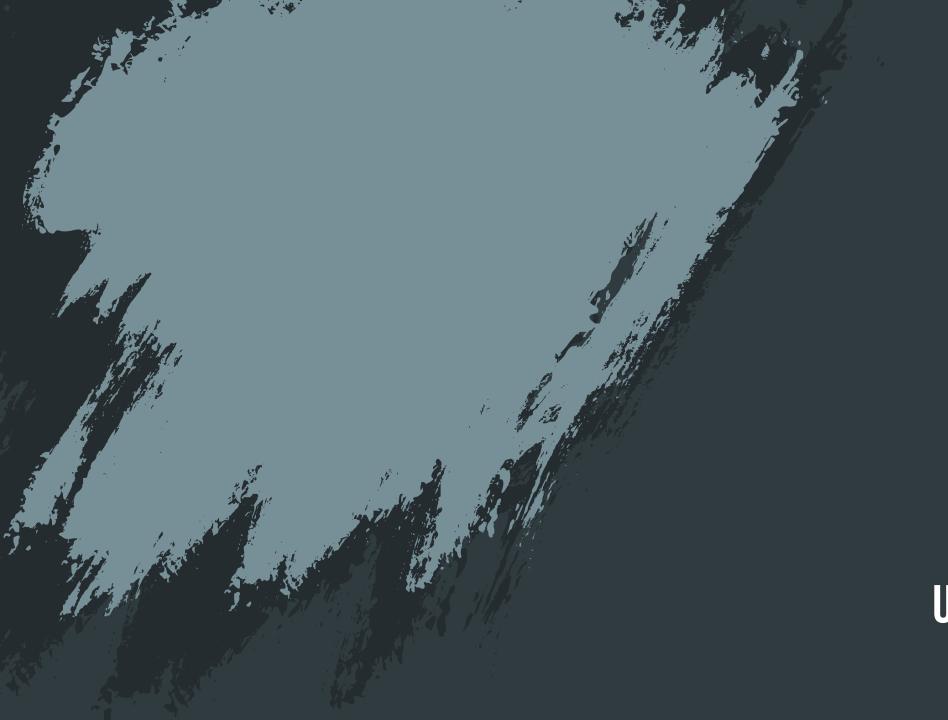
Common Use Cases for Mapping Data Flows



SCENARIOS

Transformations for Common Use Cases in Mapping Data Flows





SECTION 3 USER INTERFACE

USER INTERFACE

Overview of Mapping Data Flows User Interface

Screencast

Show how to add a mapping data flow

Show the different types of transformation

Show the different settings / configuration tabs of a data flow transformation

Show how to change the order of the transformations

Show the debug mode (time to live and data preview settings)

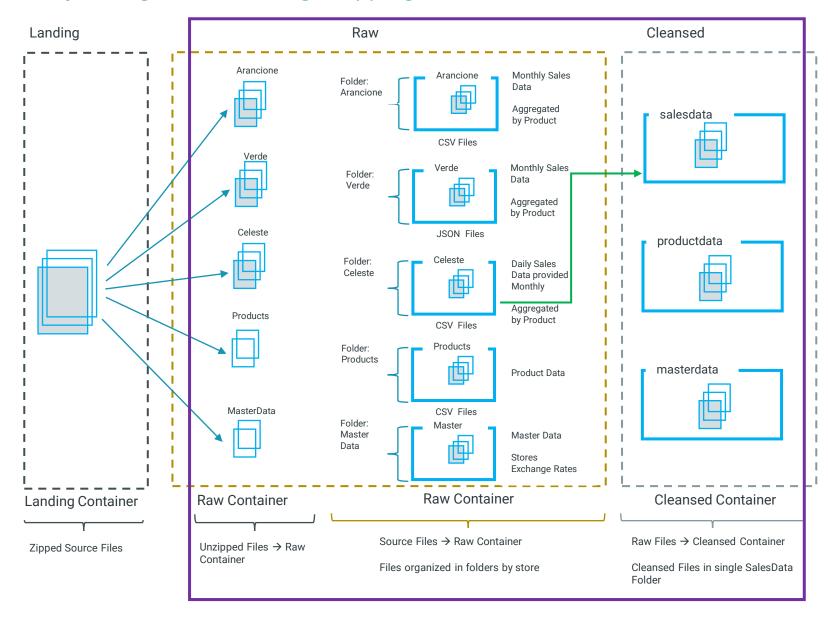
Show the data preview

Show the expression builder



DATA TRANSFORMATION

Transforming the data using Mapping Data Flows



What we will Implement?

- Build a data factory pipeline with data flows
- Copy Celeste CSV files from raw to cleansed container
- Use transformations to transform the data files
- We will review the results and the pipeline
- Optimization of data flows

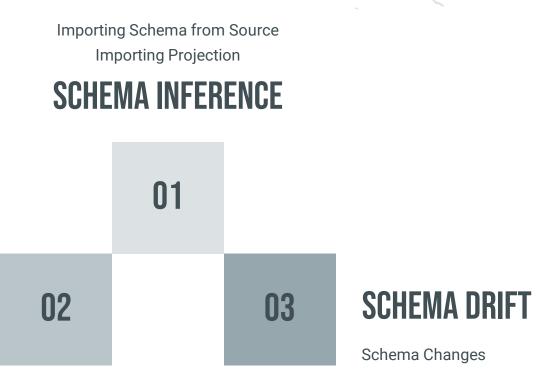
DATA FLOW CONCEPTS

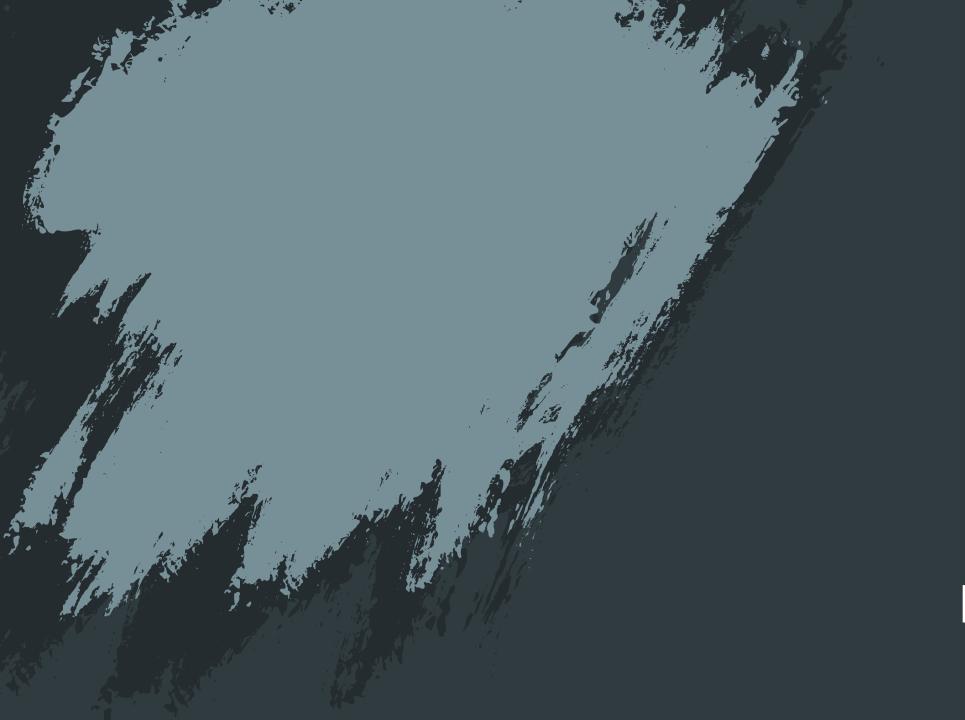
Key Concepts within Mapping Data Flows

SCHEMA BINDING

Early Binding

Late Binding





SECTION 5 PERFORMANCE

DATA FLOW

Integration Runtime Performance

General Purpose Memory Optimized

CLUSTER TYPE



TIME TO LIVE

Time to live specifies the duration a cluster is alive Cold cluster start-up takes time

PERFORMANCE

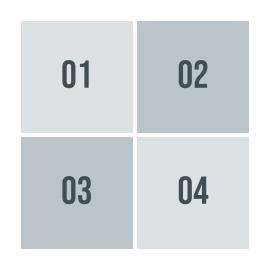
Monitoring Data Flow Performance

Time spin up an Apache Spark cluster Optimized by enabling time to live

CLUSTER STARTUP TIME

TRANSFORMATION TIME

Time to perform transformations in data flow
Optimized by identifying bottlenecks in
transformations



Time taken to read data from source Optimized by partition source strategy

SOURCE READ TIME

SINK WRITE TIME

Total time to write to sink
Optimized by sink partition strategy

REFERENCES

Data Flow Transformations

https://learn.microsoft.com/en-us/azure/data-factory/data-flow-transformation-overview

Data Flow Monitoring

https://learn.microsoft.com/en-us/azure/data-factory/concepts-data-flow-monitoring

Data Flow Performance and Tuning

https://learn.microsoft.com/en-us/azure/data-factory/concepts-data-flow-performance

MODULE SUMMARY

In this module we learnt



OVERVIEW

We got an overview of Mapping Data Flows

We looked at the user interface of Mapping
Data Flows within ADF and some of the
technical capabilities it offers



SCENARIOS

We learnt about the typical use cases where mapping data flows are used

We learnt about the different built-in transformations within mapping data flows



HANDS-ON

We learnt how to use mapping data flows within our data pipeline using the:

Filter Transformation, Aggregate
Transformation, and Derived Transformation