PIPELINE ACTIVITIES AND PARAMETERS

MODULE 6

1 ACTIVITIES

Overview of Activities

2 EXPRESSIONS AND VARIABLES

Overview of Expressions and Variables

3 PARAMETERS

Overview of Parameters

4 AZURE KEY VAULT

Setting up Azure Key Vault

5 IMPORTING DATA

Implementing pipelines with loops and metadata activities

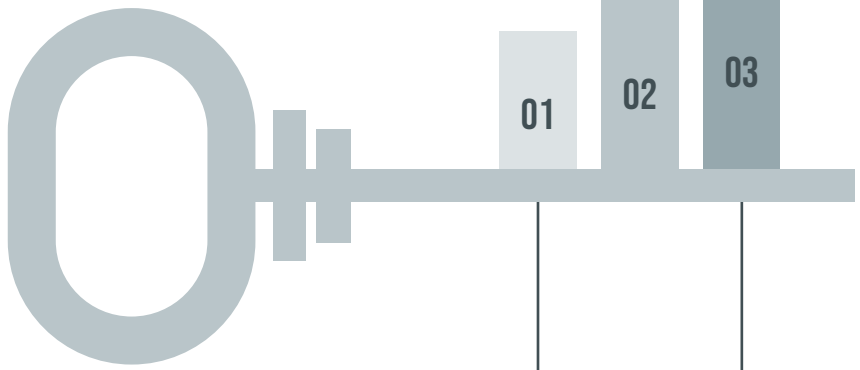
SECTION 1

ACTIVITIES

ACTIVITIES

Overview of Activities

AN ACTIVITY IS THE SINGLE TASK WITHIN A PIPELINE



DATA TRANSFORMATION ACTIVITIES

Task to change data e.g. Data Flow Activity

Mostly external ADF activities such as – Databrick Notebooks, Azure Functions, Stored Procedures

DATA MOVEMENT ACTIVITIES

Task to move data from a source to a sink e.g. Copy Activity

Multiple types of data stores supported as source / sink in ADF

CONTROL FLOW ACTIVITIES

Task for logical control flow e.g. If condition, Foreach Loop

Mostly native ADF activities to loop, to branch or to set variables

ACTIVITIES

Overview of Activity chaining

Create a pipeline from a chain of Activities

PIPELINE

01

02

03

MANY TO ONE

Link multiple activities to a single activity

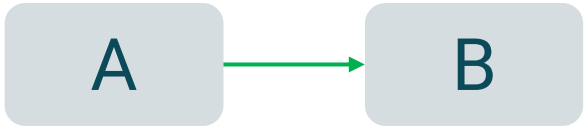
DEPENDENCY

Success, Failure, Completion or Skipped

ACTIVITIES

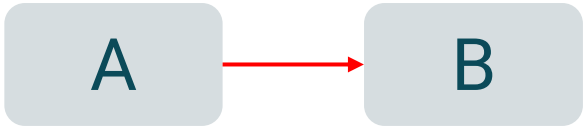
Overview of Activity Dependencies

Success



Activity B starts only if Activity A has completed successfully

Failure

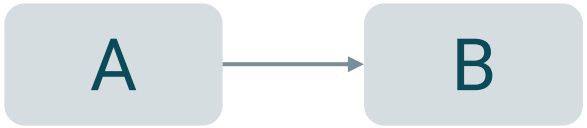


Activity B starts only if Activity A has failed

Completion

Activity B starts as long as Activity A has completed

Skipped



Activity B starts if Activity A was skipped

COPY ACTIVITY

Some key concepts of Copy Activity

SCHEMA INFERENCE

Ability to infer schema

Configure schema mapping

TYPE CONVERSION

3-step type conversion

ADF Interim Data Types (IDTs)

DEBUG MODE

Ability to view data at each stage

DATA INTEGRATION UNIT

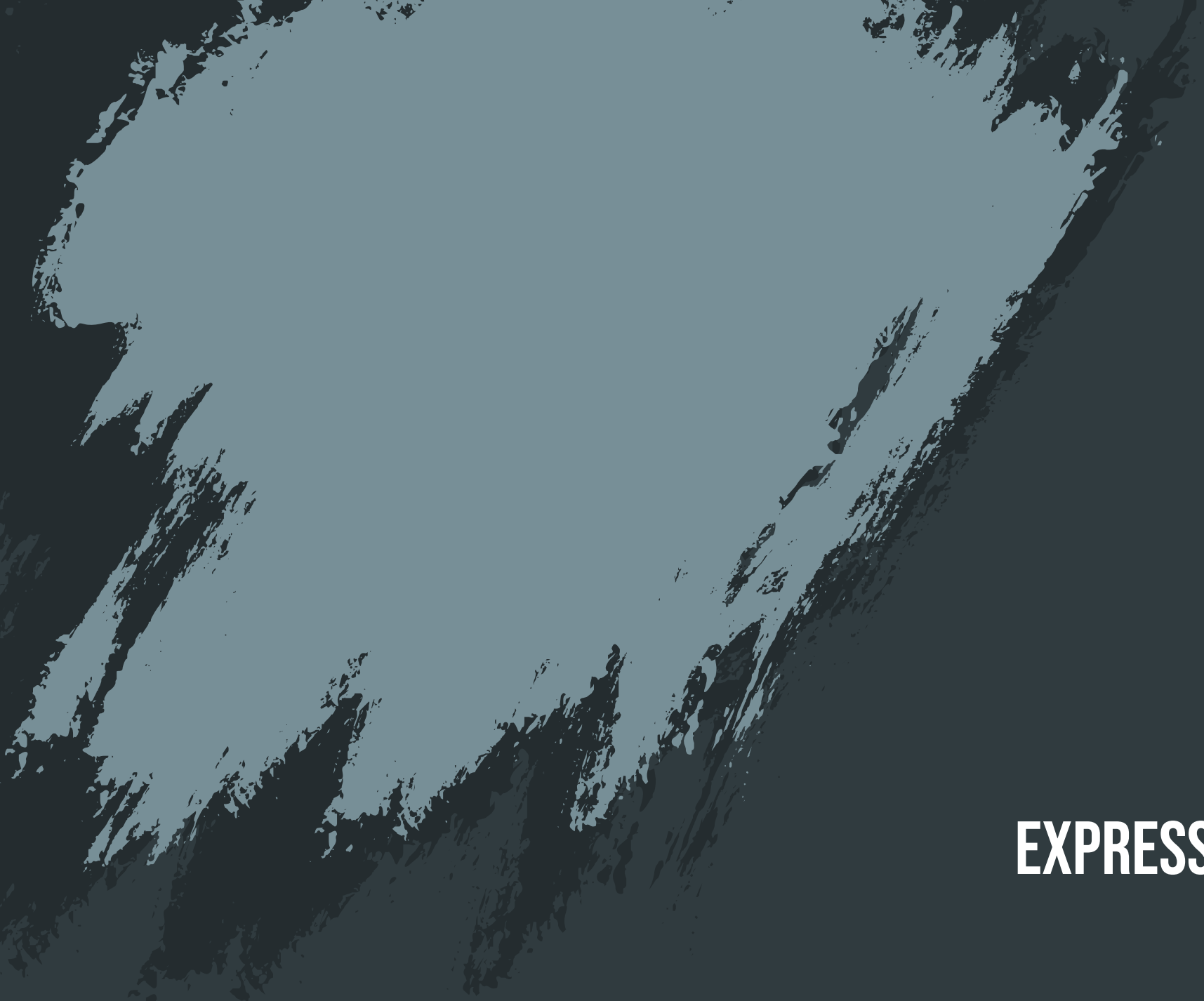
Copy Activity power measured in DIUs

Combined measure of CPU, Memory and Network Usage

DEGREE OF PARALLELISM

Parallel execution of copy activity

Define number of threads to read multiple files simultaneously



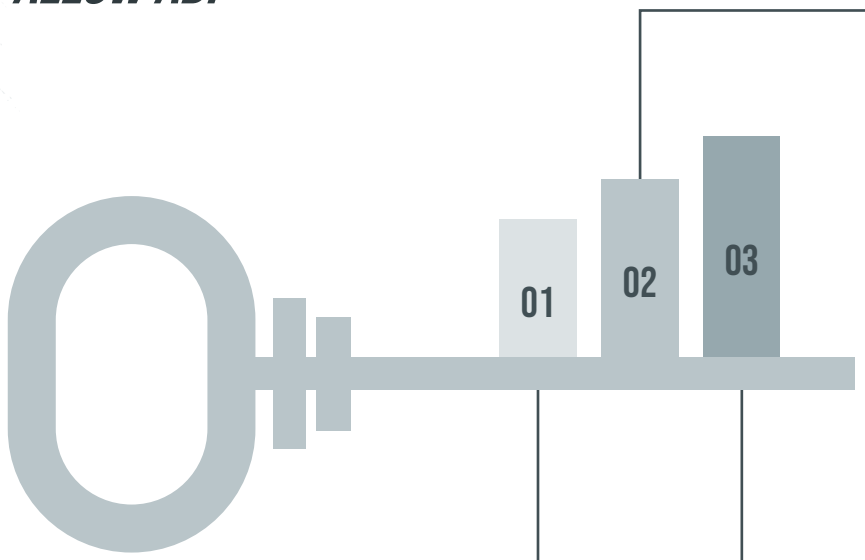
SECTION 2

EXPRESSIONS AND VARIABLES

EXPRESSIONS AND VARIABLES

Overview of Expressions and Variables

EXPRESSIONS AND VARIABLES ALLOW ADF PIPELINES TO BE DYNAMIC



VARIABLES

Properties that can be set and modified during a pipeline run

Expressions can be specified for a variable

User Variables scoped at pipeline level accessible to all activities in the pipeline

System variables provide access to runtime values of various system properties

EXPRESSIONS

Property value evaluated during pipeline execution

Expressions can be created via the Expression builder in ADF
Note: ADF expressions are case-sensitive

Expressions can be string literals, variables or functions

EXPRESSION LANGUAGE

ADF expression language has no infix operators, only functions e.g. add(1, 2) and not 1 + 2

Expression Functions – library of functions available for use in expressions

Placeholder expressions evaluated at runtime e.g. @{pipeline().DataFactory}

SECTION 3

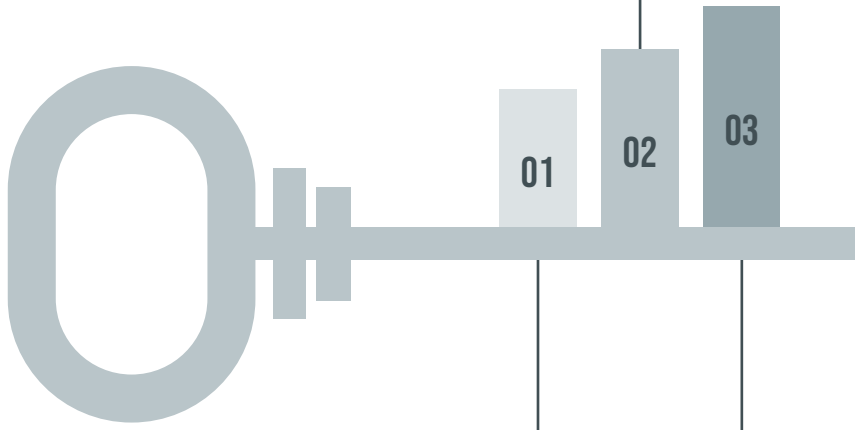
PARAMETERS

PARAMETERS

Overview of Parameters

VALUES PROVIDED TO A PIPELINE AND CANNOT BE MODIFIED DURING A PIPELINE RUN.

CONTROLS THE BEHAVIOUR OF A PIPELINE AND ITS ACTIVITIES



OPTIONAL PARAMETER

Runtime parameter that can be made optional by defining a default value

Default value used when no parameter value is provided

RUNTIME PARAMETER

Pipeline parameter – runtime parameter for an ADF pipeline

Dataset parameter – runtime parameter for an ADF dataset

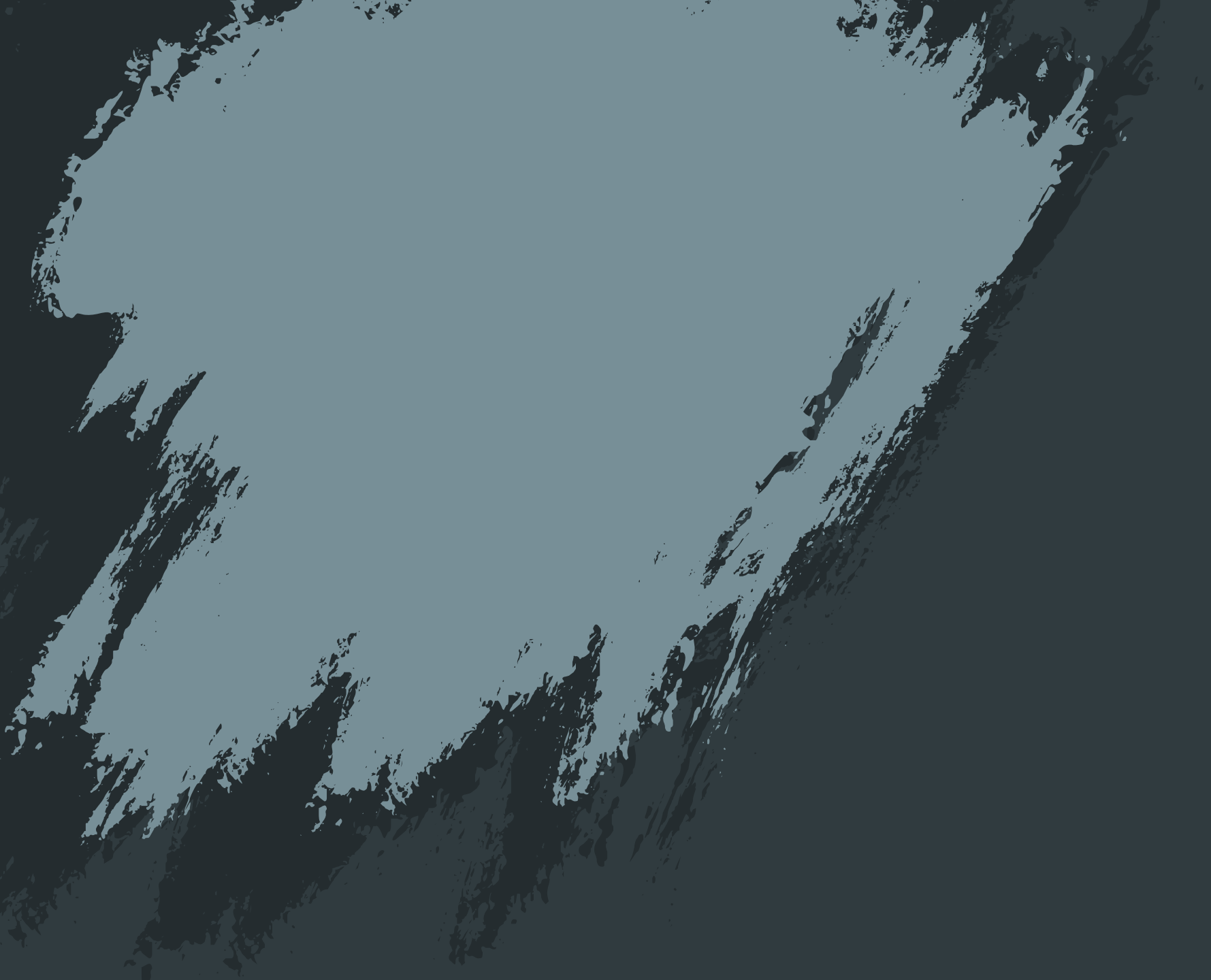
Linked Service parameter – runtime parameter for an ADF linked service

GLOBAL PARAMETER

Constant value scoped at the data factory instance

Shared by all pipelines

Do not enable runtime value substitution and treated as constants

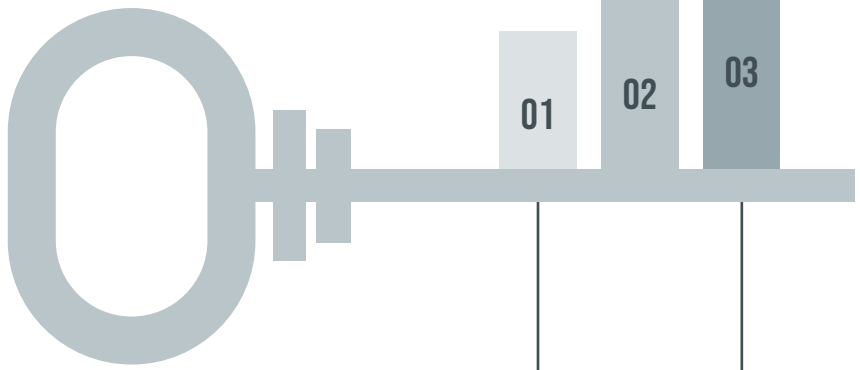


SECTION 4

AZURE KEY VAULT

AZURE KEY VAULT

Overview of Azure Key Vault



KEY MANAGEMENT

Used as a Key Management solution, so we can create and control the encryption keys used to encrypt your data.

SECRET MANAGEMENT

Securely store and tightly control access to tokens, passwords, certificates, API keys, and other secrets

CERTIFICATE MANAGEMENT

Provision, manage, and deploy certificates for use with Azure.

AZURE KEY VAULT

Why Azure Key Vault

01

CENTRALIZE STORAGE

It allows us to centralize the storage of application secrets

02

APPLICATION PARAMETERS

Developers no longer need to embed security information inside their applications

03

ACTIVE DIRECTORY

Access to the secrets within a Key Vault can be managed via Azure Active Directory

04

MONITORING

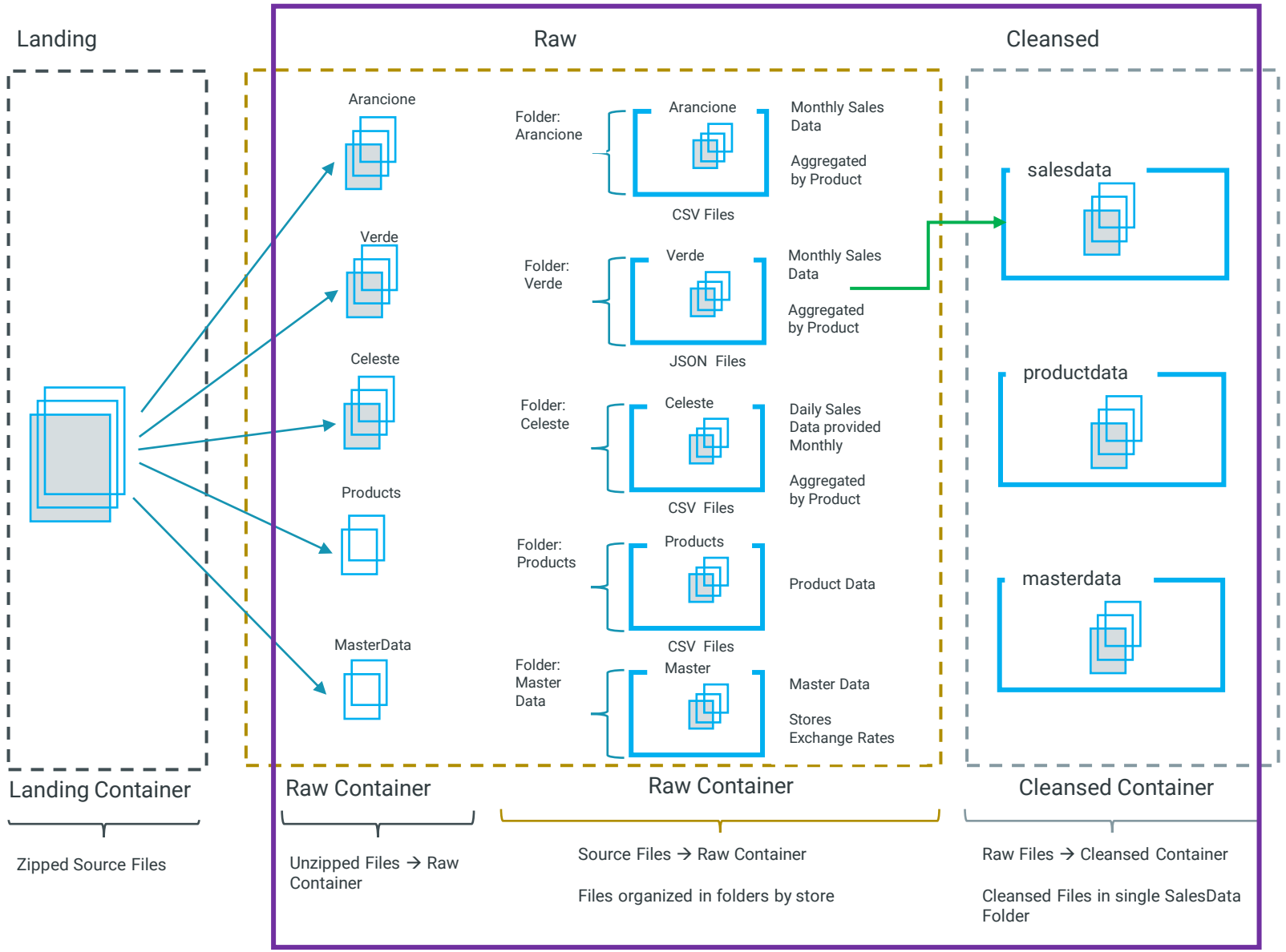
Monitor who accessed the Key Vault and what they accessed and when they accessed by enabling logging

SECTION 5

IMPORTING DATA

IMPORTING DATA

Importing Semi-structured data



What we will Implement?

- *Build a data factory pipeline*
- *Copy Verde JSON files from raw to cleansed container and transform to CSV*
- *We will use the ForEach and the Metadata Activities*
- *We will use parameters and use Azure Key Vault*
- *We will review the results and the pipeline*

MODULE SUMMARY

In this module we learnt

OVERVIEW

We got an overview of the different types of Activities, activity dependencies

We learnt about expressions, variables and parameters

INTEGRATION

We learnt about Azure Key Vault

We learnt how to setup Azure Key Vault

HANDS-ON

We learnt to build a pipeline using the:

ForEach Activity, Metadata Activity

Integrating with Azure Key Vault

And how to use parameters

REFERENCES

Activities

https://learn.microsoft.com/en-us/azure/data-factory/concepts-pipelines-activities?tabs=data-factory

Parameters and Variables

https://learn.microsoft.com/en-us/azure/data-factory/concepts-parameters-variables

Azure Key Vault

https://learn.microsoft.com/en-us/azure/data-factory/store-credentials-in-key-vault