**Cognite Functions Template for Time Series Analytics - Integration Architecure**

**CoE - Analytics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Version Control** | | | | |
| Version | Date | Change Descritpion | Responsible | Approver |
| 1.0 | 08.09.23 | Initial version | Vetle Nevland |  |
| 2.0 | 21.11.23 | Generalization to multiple time series inputs/outputs | Vetle Nevland |  |
|  |  |  |  |  |
|  |  |  |  |  |

[1. Document Objective 4](#_Toc41383220)

[2. Problem Descrition 4](#_Toc41383221)

[3. General Graphical Overview 4](#_Toc41383222)

[4. Data Extraction 4](#_Toc41383223)

[5. Transformations 4](#_Toc41383224)

[6. Access Control 4](#_Toc41383225)

[CDF DataSet Configuration 4](#_Toc41383226)

[7. Considerations 4](#_Toc41383227)

[7.1. Premisses 4](#_Toc41383228)

[7.2. Restrictions 4](#_Toc41383229)

[8. Roles and Responsabilties 4](#_Toc41383230)

[9. Applications and Use Cases impacted by New Integration/Change 5](#_Toc41383231)

[10. Infrastructure Impact 5](#_Toc41383232)

# Document Objective

*This document aims at describing the overall architecture of a integration one or more source systems to CDF (RAW or CLEAN). As well as describing the extraction step, it will also describe the following pipeline steps (if applicable), i.e. transformations, contextualisation and any RAW to CLEAN activity.*

Extraction step is irrelevant. All transformations are performed using Cognite’s Python SDK, deployed and run on prescribed schedules using Cognite Functions. The contextualized data model constitutes the same elements and relationships as the original time series. Time series’ transformed with this project’s template are written to a dataset in CDF, Center of Excellence – Analytics, designated to basic and advanced time series analytics.

# Problem Descrition

*Here we shall describe what this solution is trying to achieve, source systems, use stories (functional requirements, business rules), CDF resources, etc.*

The goal of the project is to provide a template for running semi-advanced calculations on time series data residing in CDF. Currently, the template only supports time series data, so Time Series is the relevant CDF resource type. The primary purpose of the template is to enable seamless deployment of new time series that are automatically updated by a prescribed schedule, which in turn facilitates streamlined analytics and rapid insight.

The project demonstrate the process by extracting time series data of oil volume percentage in tanks, and transforming the data to calculate daily average drainage rate to help detect leakages.

# General Graphical Overview

*Represent, in a graphical manner, the main components involved in the integration. Here we can use any type of diagram which will help the understanding of the proposed solution: data flow, sequence, etc. Here we should also describe the main interactions (i/o) between the different components.*

The data pipeline comprises a closed loop where time series data is extracted from Cognite Fusion Prod, transformed with Cognite Functions through Python SDK, written to Cognite Fusion Dev, tested and validated in a testing environment, and finally deployed as a governed data product to Cognite Fusion Prod. The data can then be visualized in, e.g., Grafana Dashboards for further insights.

A diagram of a process

Description automatically generated

# Data Extraction

*Describe the data extraction in a more detailed and technical manner: source systems, servers, technology, data and meta data being transported, destination, RAW (DB and Tables), CLEAN (CDF Resources representation).*

The data of interest is already available in CDF Clean as Time Series CDF resource type. The project is thus not dependent on any external source systems or extractors.

# Transformations

*Describe all the transformations from RAW to CLEAN in a simple easy to read manner as well as all the technical aspects involved in the transformation, such as type of platform used (GCP, Azure or Cognite), and the different steps (jobs) taken to make it to CLEAN.*

The data has already been transformed to a CDF time series resource type residing in CLEAN. Herein, Cognite’s Python SDK are used to perform further transformations on time series data residing in CLEAN. We refer to the GitHub documentation ([AkerBP-DataOps/deos-cognite-functions-template (github.com)](https://github.com/AkerBP-DataOps/deos-cognite-functions-template)) for further details of how to use the Cognite Python SDK, together with Cognite Functions, to transform time series’ in CLEAN.

# Access Control

# To deploy data to the Cognite Fusion Dev environment, we need write access to a designated dataset. This is done by submitting an access request form for Cognite Data Fusion. The form is found here: <https://forms.office.com/Pages/ResponsePage.aspx?id=cEF-O0iDpEq_rgaj4YZ0aUVYsXTN0c9Dil0iHGZgj0lUOTBXVFlSWDlMUFk1WUNBS1lKWjZKWko2TyQlQCN0PWcu>. To grant access to a new dataset, the following were submitted for this project:

# In \**Area of the request\** select Cognite Data Fusion (CDF)

# In \**Category*\* select New access or create new dataset

# In \**New access or create new dataset request*\* select Create new dataset

# In \**Justification of dataset*\*, reason why you need to create a new dataset. For this project, we reason that the new dataset will serve as a hub for time series analytics within Center of Excellence, deployed using our Cognite Functions Template.

# Leave \**Source information*\* empty

# Leave \**RAW database and tables information*\* empty

# In \**Tranformations/Pipelines information\** mention Cognite Functions and Python SDK

# In \**Data extraction and its corresponding transformation*\* provide a link to the GitHub repository where the project documentation resides

# Leave \**Service account information*\* empty

# In \**READ access to CDF resources*\* tick off “Time Series READ”, “Files READ”, “Relationships READ” and “Functions READ”

* + (“Files: READ” must be ticked since the Aker BP tenant in CDF requires Cognite Functions to be associated with a dataset, which is connected through a File).

# In \**WRITE access to CDF resources*\* tick off “Time Series WRITE”, “Files WRITE”, “Relationships WRITE” and “Functions WRITE”

# Leave \**Specification of dataset regarding READ/WRITE access*\* empty

# For \**CDF tenant\** select AKER BP DEV

CDF Dataset Configuration:

* There are no specific access restrictions on the produced dataset. General access permissions apply.

**Access groups:**

**DEV/TEST/PROD**

**Group:** UNKNOWN

**Group capabilities** (scoped to the “Center of Excellence – Analytics” dataset)

* *time\_series:read* and *time\_series:write*
* *files:read and files:write*
* *functions:read and functions:write*
* *relationships:read and relationships:write*

# 

# Considerations

## Premisses

## Restrictions

# Roles and Responsabilties

# Applications and Use Cases impacted by New Integration/Change

|  |  |  |
| --- | --- | --- |
| **Application** | **Type of Impact (Dev or Test)** | **Responsbile** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# Infrastructure Impact

*Here we should define if we have any need of infrastructure involvement (servers, firewalls,etc)*