Erasmus Centre for Data Analytics

Erasmus Data Collaboratory – House of AI Mixed Source Analytics

DWBISummit- March 27, 2024

AI, Data & Digitalisation

Erasmus University Rotterdam

Erasmus University Rotterdam

Erasmus Centre for Data Analytics

Flagship centre for cross-disciplinary insight on Artificial Intelligence ,Data, Digitalization and Immersive Technologies at the Erasmus University Rotterdam.

Erasmus Centre for Data Analytics

ECDA as supporting hub

Erasmus University Rotterdam

Funding

AI, Data & Digitalisation

AI, Data & Digitalisation

Erasmus University Rotterdam

Erasmus University Rotterdam

Erasmus University Rotterdam

Erasmus Trustfonds

Strategic research programmes on Digitalization, AI & immersive tech

Strategic EUR innovation programmes on Digitalization, AI & immersive tech

Transdisciplinary interfaculty initiatives

Supported Programs

Erasmus Data Collaboratory House of AI

Convergence AIDD

Labs & Initiatives

AI-PACT

SSH

influence of digitalization on work, prosperity and entrepre-neurship

(under discussion)

AI-MAPS

NWO

ALGOSOC

NWO

Smart Campus

AI @ EUR

Energy Transition programme

.....

.....

Offering

Facilitate Research & Innovation programs

Facilitate Hands on Education / (re)skilling

Share Thought Leadership XPs

Community Building Connect & Inspire

Activate & Empower Partner organizations

Competencies

ECDA

AI & Data Competency / Marketing & Communications / Programs & Events / Business Development

Erasmus Centre for Data Analytics

Erasmus Data Collaboratory

Erasmus University Rotterdam

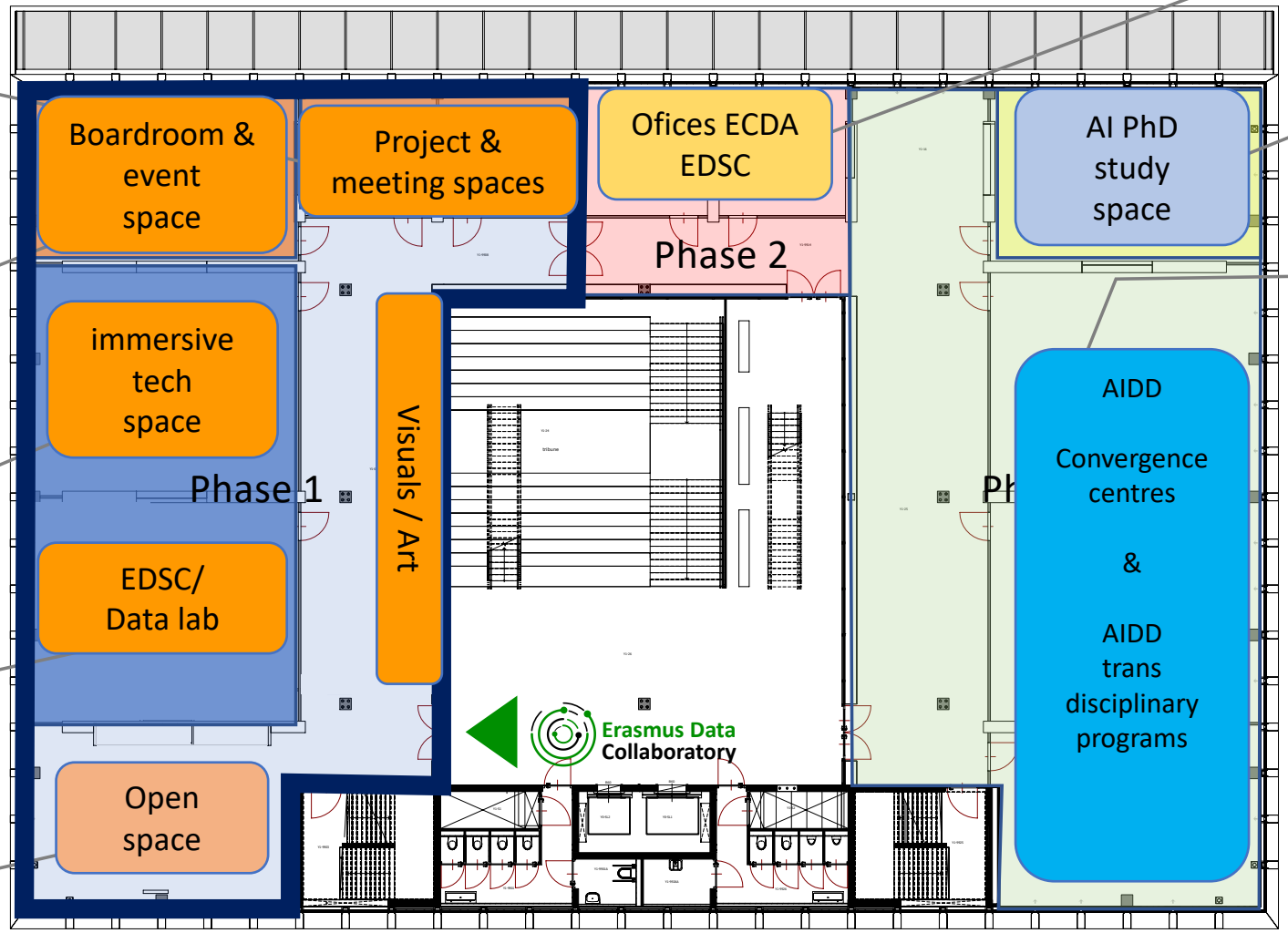
Project room 1 and 2 (multi purpose) ECDA owned

Co-Lab space Meeting room (multi purpose) ECDA owned

Teaching and Visualisation room (from 3.20) EDSC owned

Computer room incl. Bloomberg / Eiken (from 3.18) EDSC owned

Coffee Lounge



EDSC, ECDA, Healthy Start Offices (from 3.14 / 3.15)

PhD trans-faculty study spaces Healthy Start, AI-Pact, AI-maps etc.

General Interfaculty teaching and study area with focus on data analytics & AI

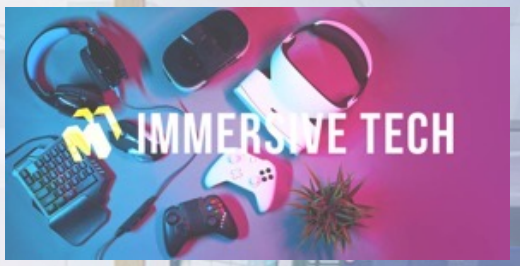
Phase 1 – 29-10-2021 / Phase 2 – a.s.a.p. / Phase 3 – ambition driven

smus Centre for
ta Analytics

Erasmus Data Collaboratory

Visualization room
Immersive Tech Space

Erasmus University Rotterdam



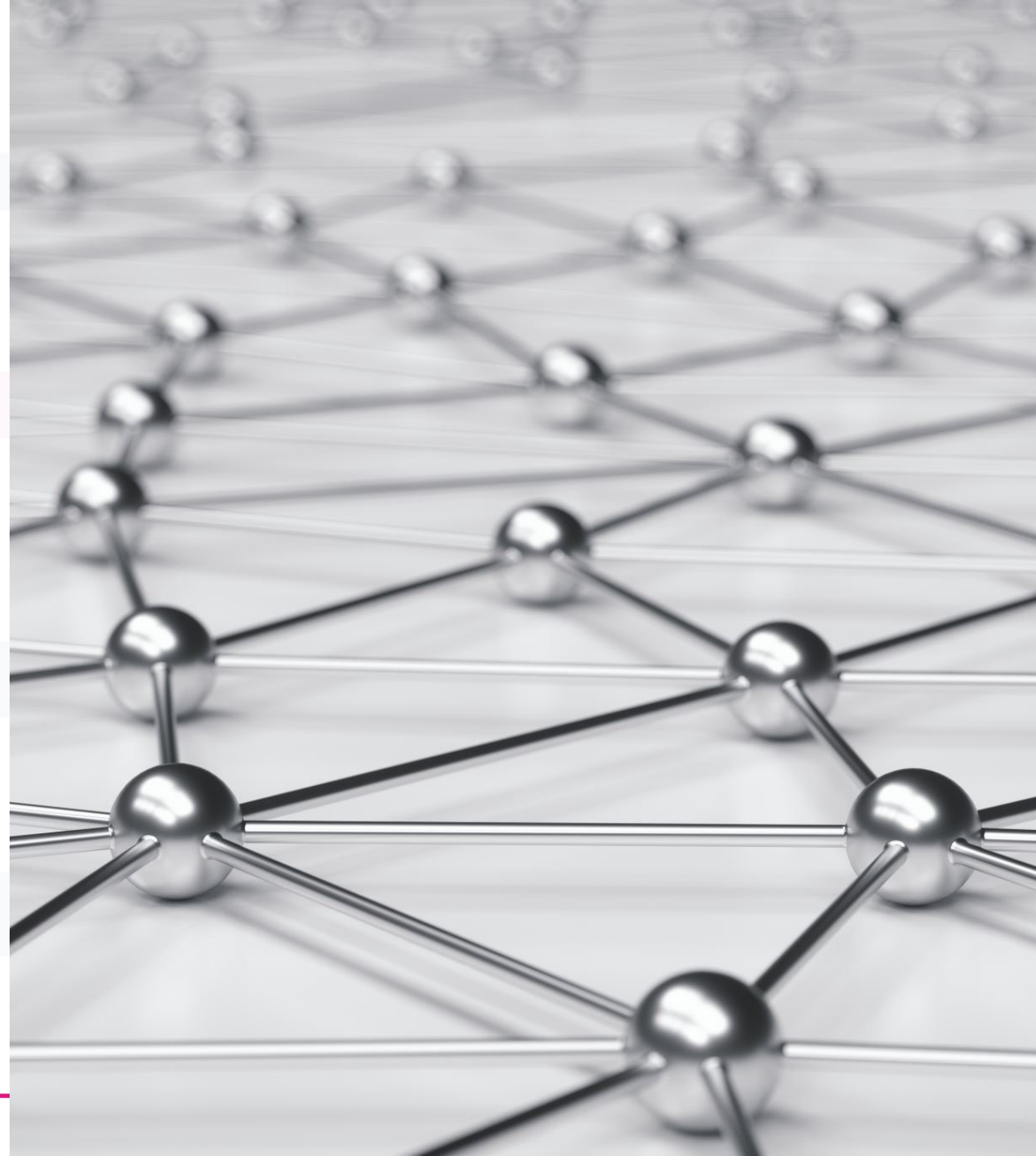
Erasmus Centre for Data Analytics

Erasmus Data Collaboratory

"...support innovative data intensive projects with both internal and external parties at EUR "

If your project contains one or more of the following characteristics, please reach out to us:

- Innovative in nature
- Has a major data, AI or Immersive Technology component;
- Involves collaboration with External Parties with active exchange of data/expertise;
- Needs additional resources such as compute or engineering support.



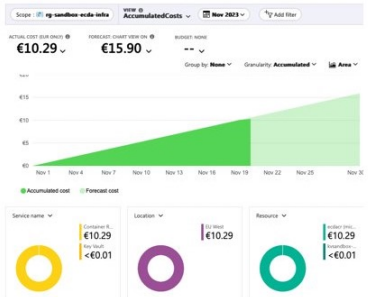
Erasmus Data Collaboratory

Services for qualified research projects

- Access to external network (for data & funding)
- Access to internal resources (for skills & knowledge)
- Qualified data engineering support staff
- Safe and secure online collaboration environment
- Local compute & GPU (€0, limited)
- Access to cloud compute (€-€€€)

Local compute-storage-gpu:

- 104 vCpu
- 640 Gb Ram
- 3 * Nvidia RTX 4090
- 12 TB Nvme SSD
- 25 TB SATA SSD
- 120 TB Harddisk



Microsoft Azure (EDIS)

- Virtually unlimited capacity
- we WILL charge you ☺

Lenovo Thinkstation P620:
1 * Nvidia RTX 6000 Ada

HP Z8 Fury G5:
2 * Nvidia RTX 6000 Ada

Erasmus Data Collaboratory

Meet the Team!


Data Team

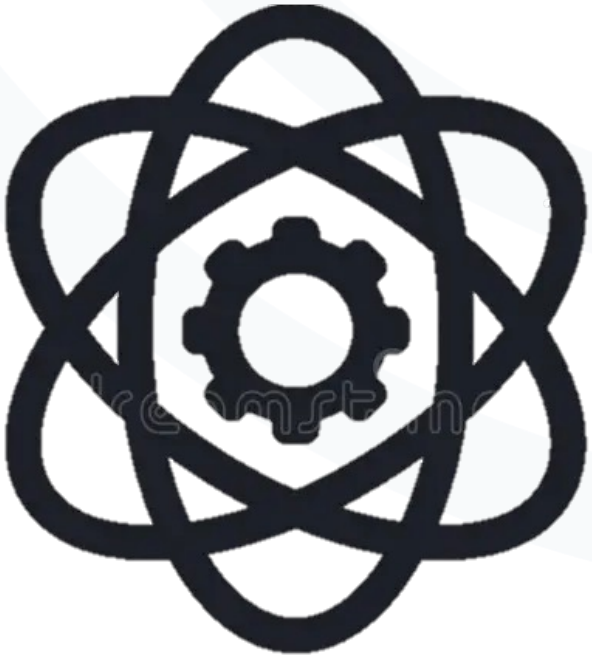

Immersive Tech Team

Platform & Apps

Infrastructure

Erasmus Data Collaboratory

EDC Data Sandbox Core Principles



1. Never compromise on security
2. No to Personal Data (unless…)
3. Data is a strategic asset
4. Don't reinvent the wheel
5 Everything should be reusable
6. Embrace change
7. Stay independent
8. Don’t (try to) boil the ocean
9. Maximize Utilization
10. Minimize Carbon Footprint

EDC Sandbox: powered by Open Source

Ollama

CrateDB

HOP

python™

Apache Airflow

pfsense

PROXMOX

docker

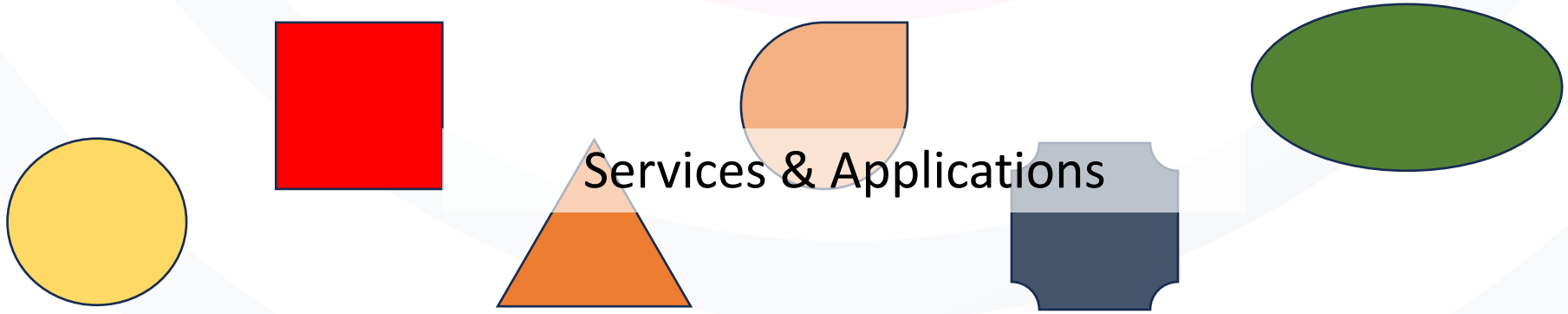
kubernetes

RANCHER

High Level Architecture

Identity & Access Management

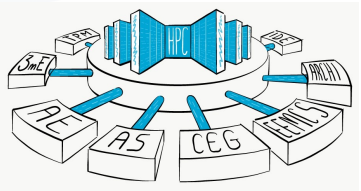
SURF



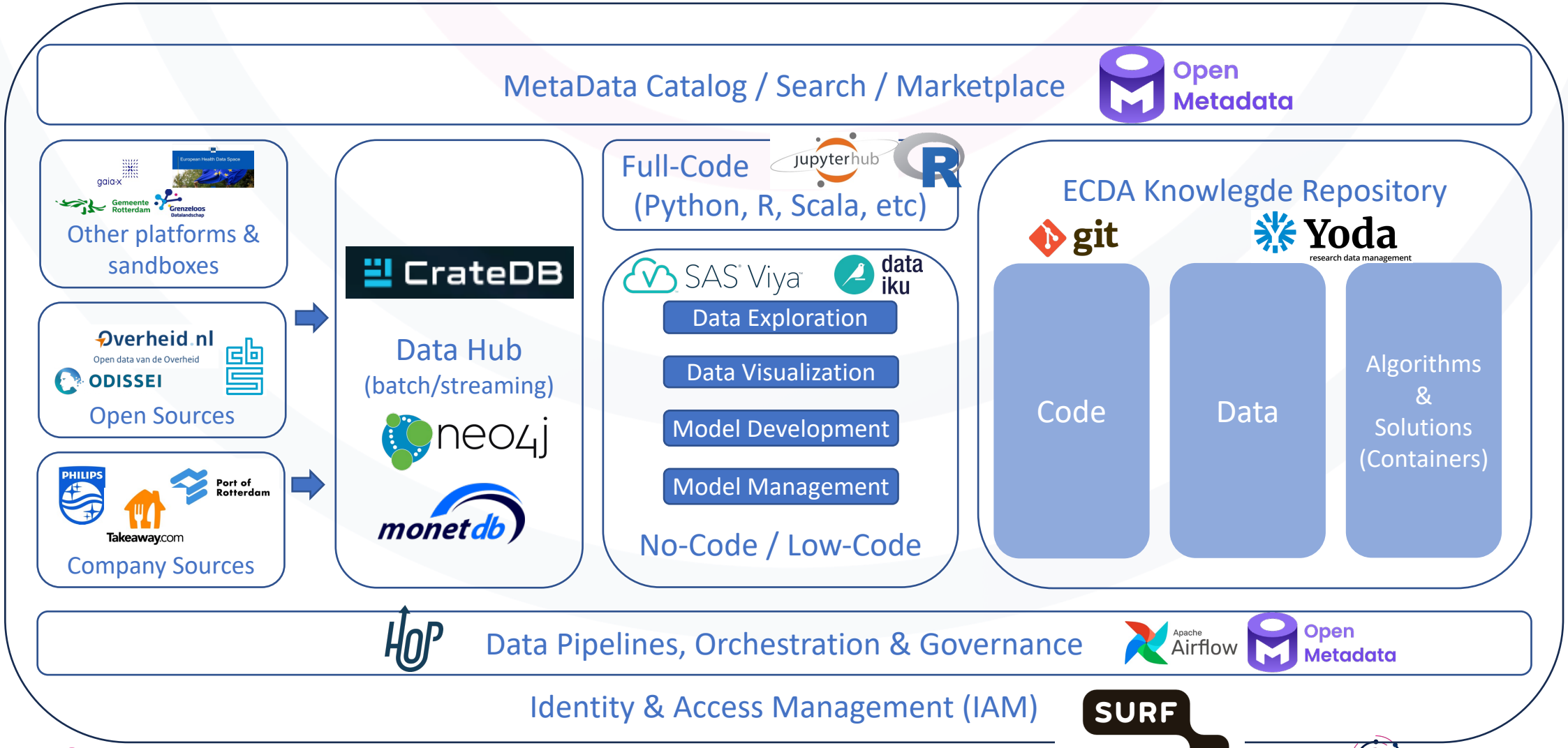
kubernetes kubernetes kubernetes kubernetes

Azure

leafcloud



Erasmus Data Collaboratory Sandbox Tech



Erasmus Centre for Data Analytics

Sandbox = Hybrid 'Cloud'

6 * MinisForum UM 790 Pro

16 vCpu, 64 GB Ram
internal GPU
2*2TB Nvme SSD
10 Gbit interconnects

4 * GPU Workstation

32 vCpu, 192 GB Ram
nVidia RTX 4090
2*1TB + 2*2TB Nvme SSD
25 Gbit interconnects

Backup/ shared storage

8 core cpu, 64 GB Ram
2*2TB Nvme SSD, 2*1.8TB SSD
6 * 20 TB Sata
10 Gbit interconnect

Monthly Cost on Azure would be > € 8.500, excl. GPU's!

EDC - House of AI Local Infrastructure

PROXMOX

kubernetes

EDC – House of AI LeafCloud

EDC - House of AI EDIS-Azure

EDC - House of AI Surf Research Cloud

Total compute capacity

- *112 cores/224 vCpu*
- *1.15 TB Ram*

Total storage capacity (raw/net)

- *50 / 34 TB Nvme SSD*
- *120 / 60 TB Sata*

Total compute/storage capacity

- *Virtually unlimited*

Erasmus Centre for Data Analytics

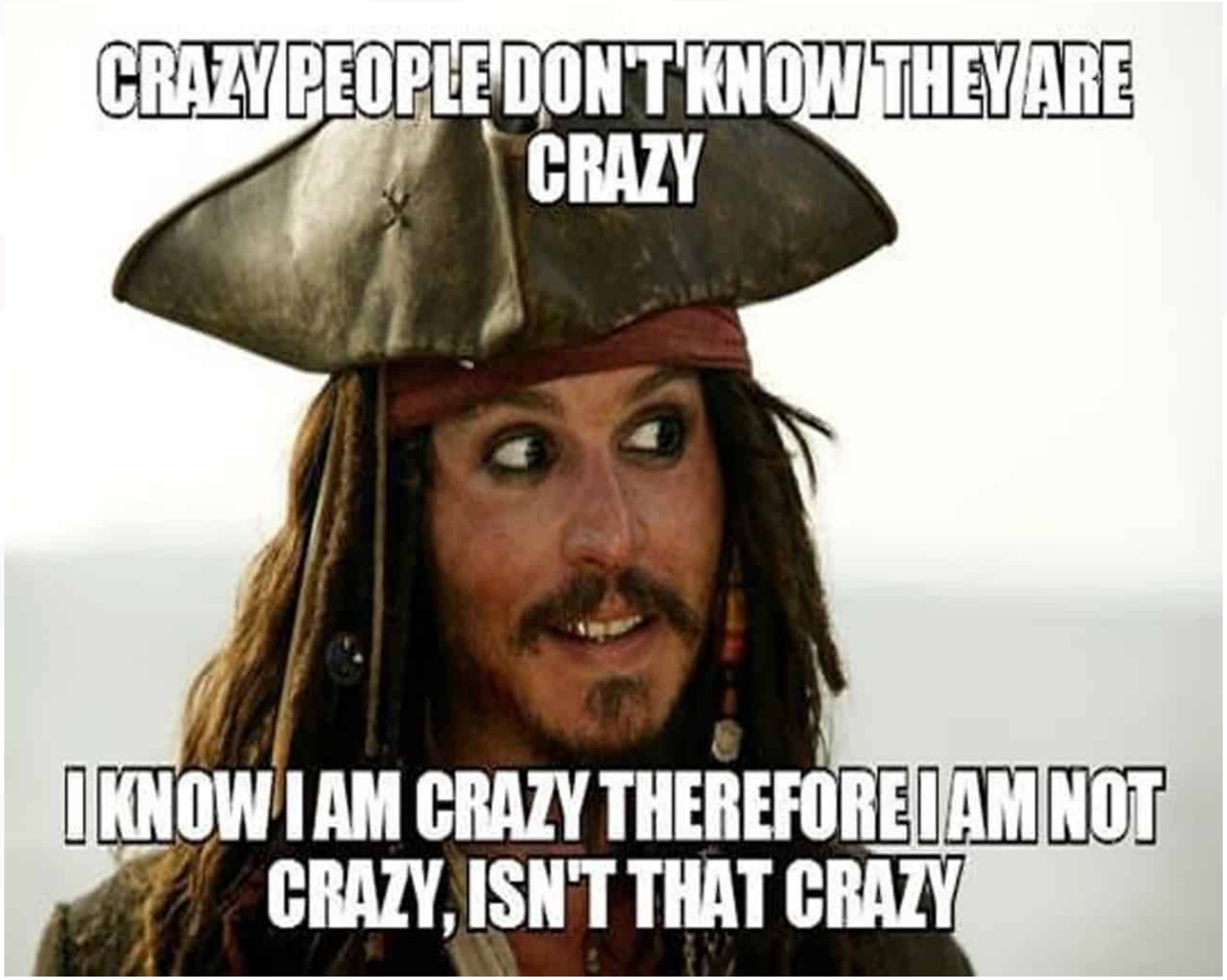
The EDC Compute Cabinet ☺

✅ Fast ✅ Affordable ✅ Reliable ✅ Energy Efficient

Public Cloud vs On Prem

"You're crazy if you don't start in the cloud; you're crazy if you stay on it"

andreessen horowitz



Source: [The Cost of Cloud, a Trillion Dollar Paradox]

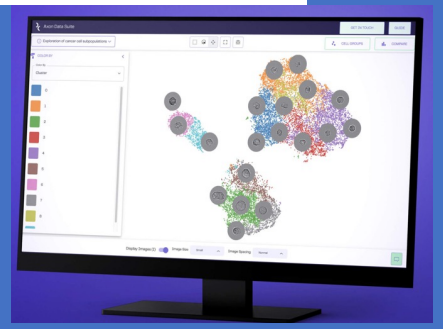
Erasmus Data Collaboratory

Sandbox Experiments / Projects

Erasmus University Rotterdam

Deepcell

Erasmus MC



Other platfor
sandboxe

Open Sources

Company Sou

sitory

Algorithms
&
Solutions
Containers)

FeedbackFruits

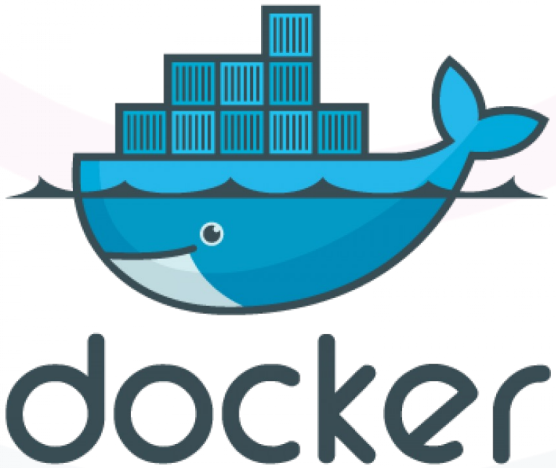
Ollama

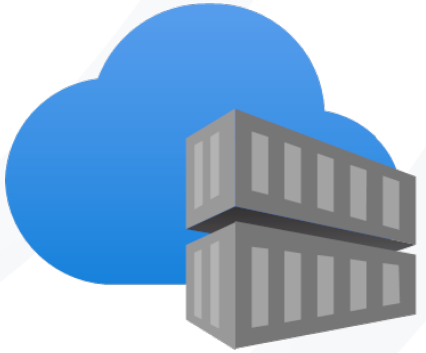
chat.ecda.ai

Erasmus Centre for Data Analytics

Data Engineering @ Erasmus Data Collaboratory

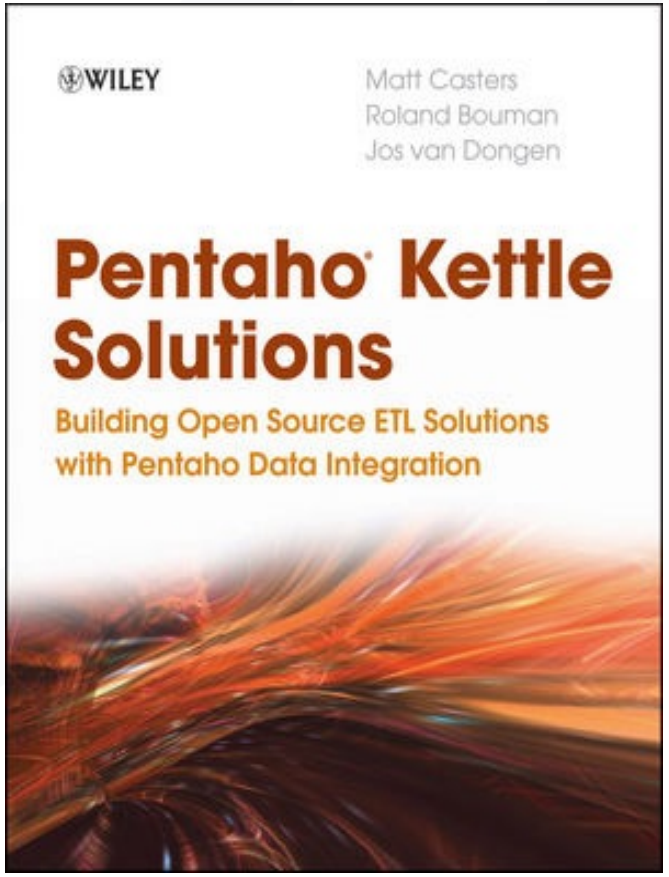
Tools of the Trade

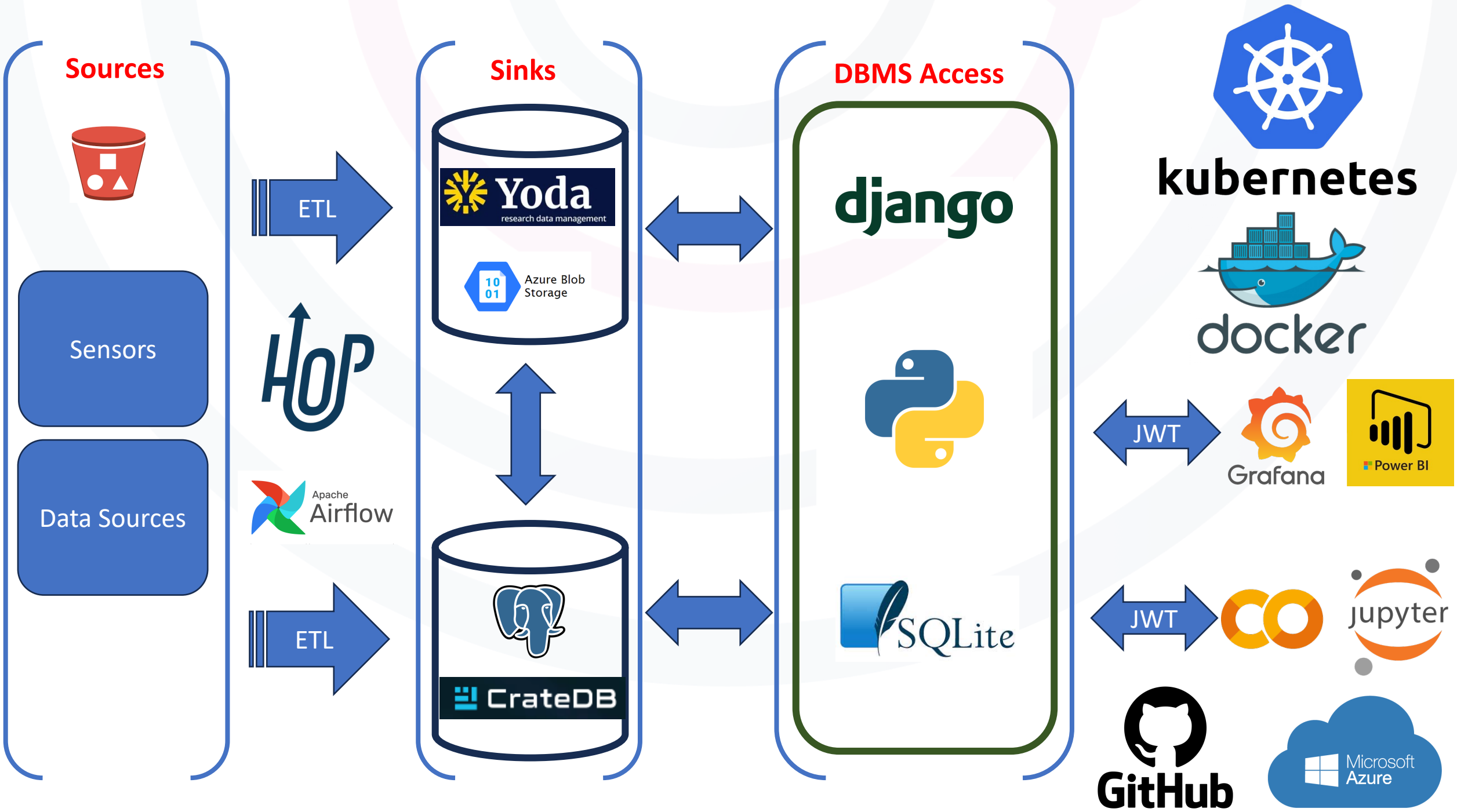




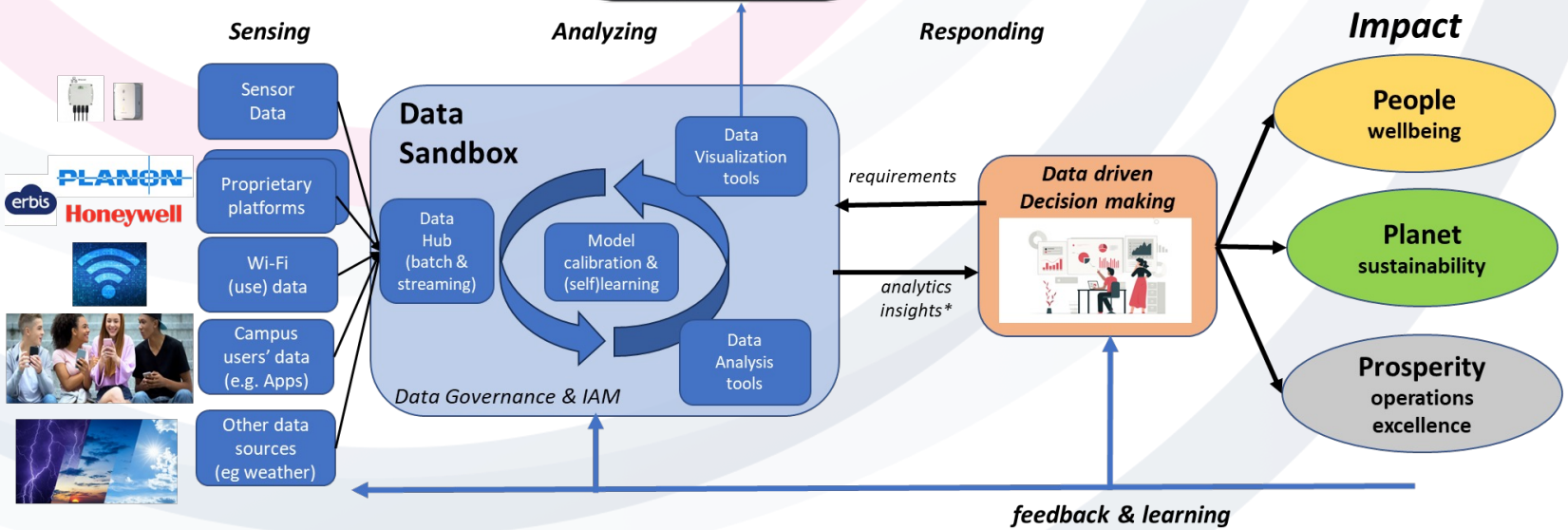
Apache Hop

- A.k.a. Kettle a.k.a. Pentaho Data Integration
- Kettle founded by Matt Casters early 2000's
- 2005: Open Sourced & 'acquired' by Pentaho
- 2015: Pentaho acquired by Hitachi
- 2019: Forked by community
- 2020: Apache Incubator project
- 2022: Apache Top Level project (Jan 18)
- https://hop.apache.org




Sources
Sensors
Data Sources
ETL
HOP
Apache Airflow
ETL
Sinks
Yoda
research data management
Azure Blob Storage
CrateDB
DBMS Access
django
SQLite
kubernetes
docker
JWT
Grafana
Power BI
JWT
jupyter
GitHub
Microsoft Azure

Erasmus University Rotterdam



* Descriptive, Diagnostic, Predictive, Prescriptive, Automated optimization

Erasmus Centre for Data Analytics (ECDA)

Erasmus Digitalisation & Information Services (EDIS)

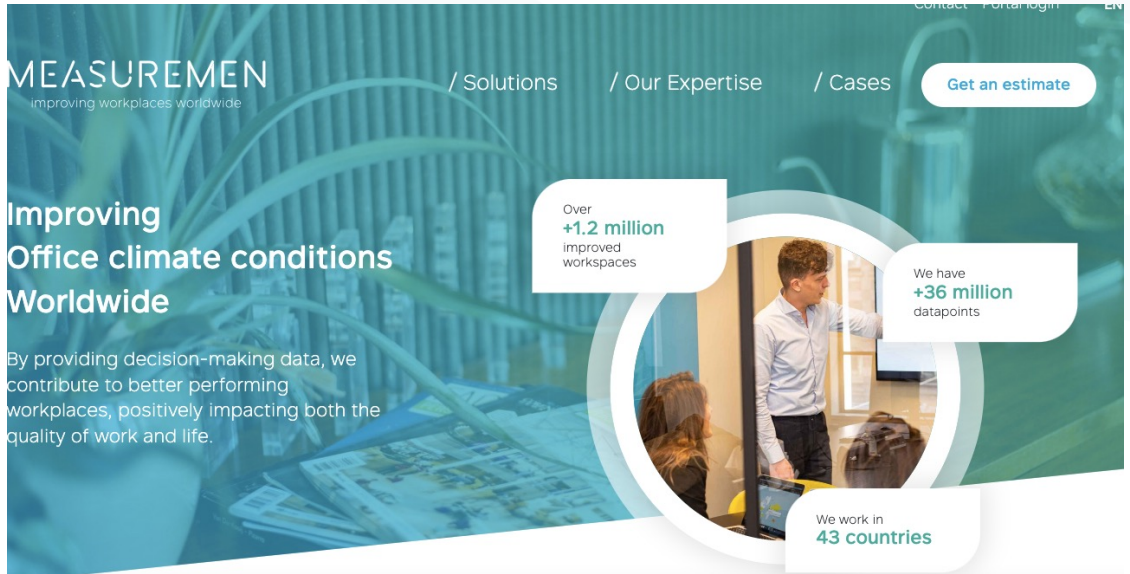
Real Estate & Facilities (RE&F)

Sensor data From API System of the Measurement

Hop

ETL

Apache Airflow



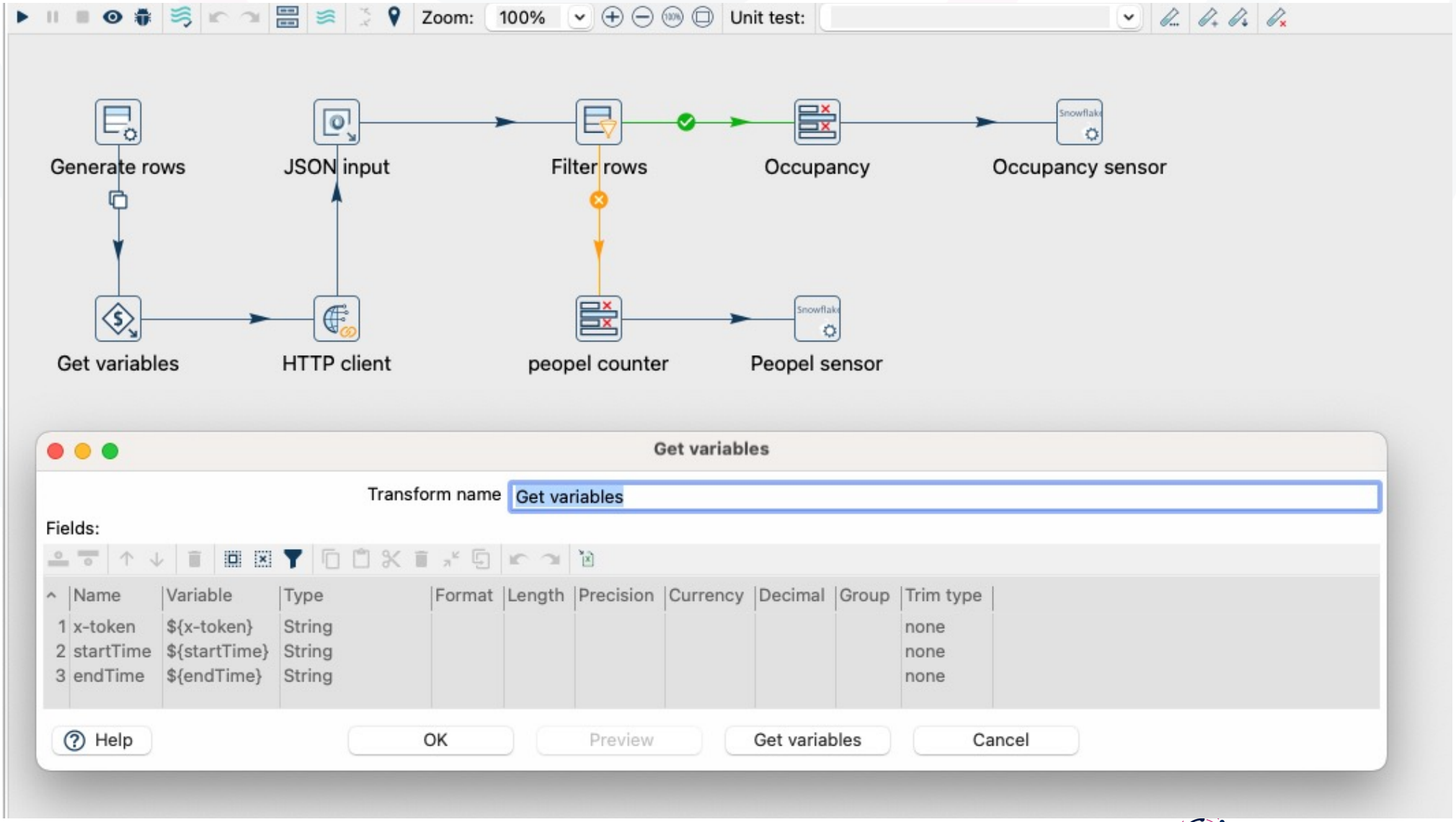
django

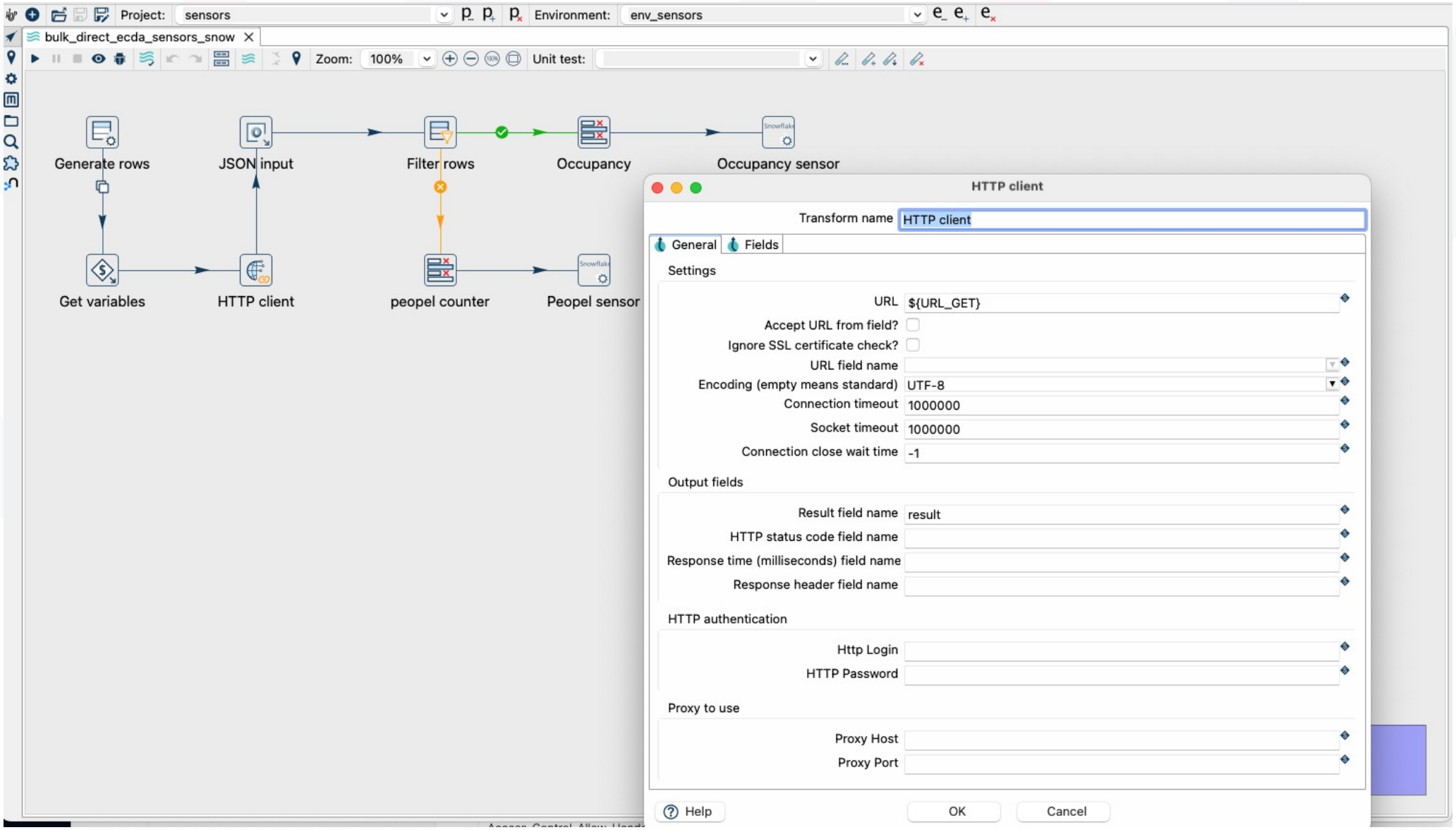
SQLite

JWT

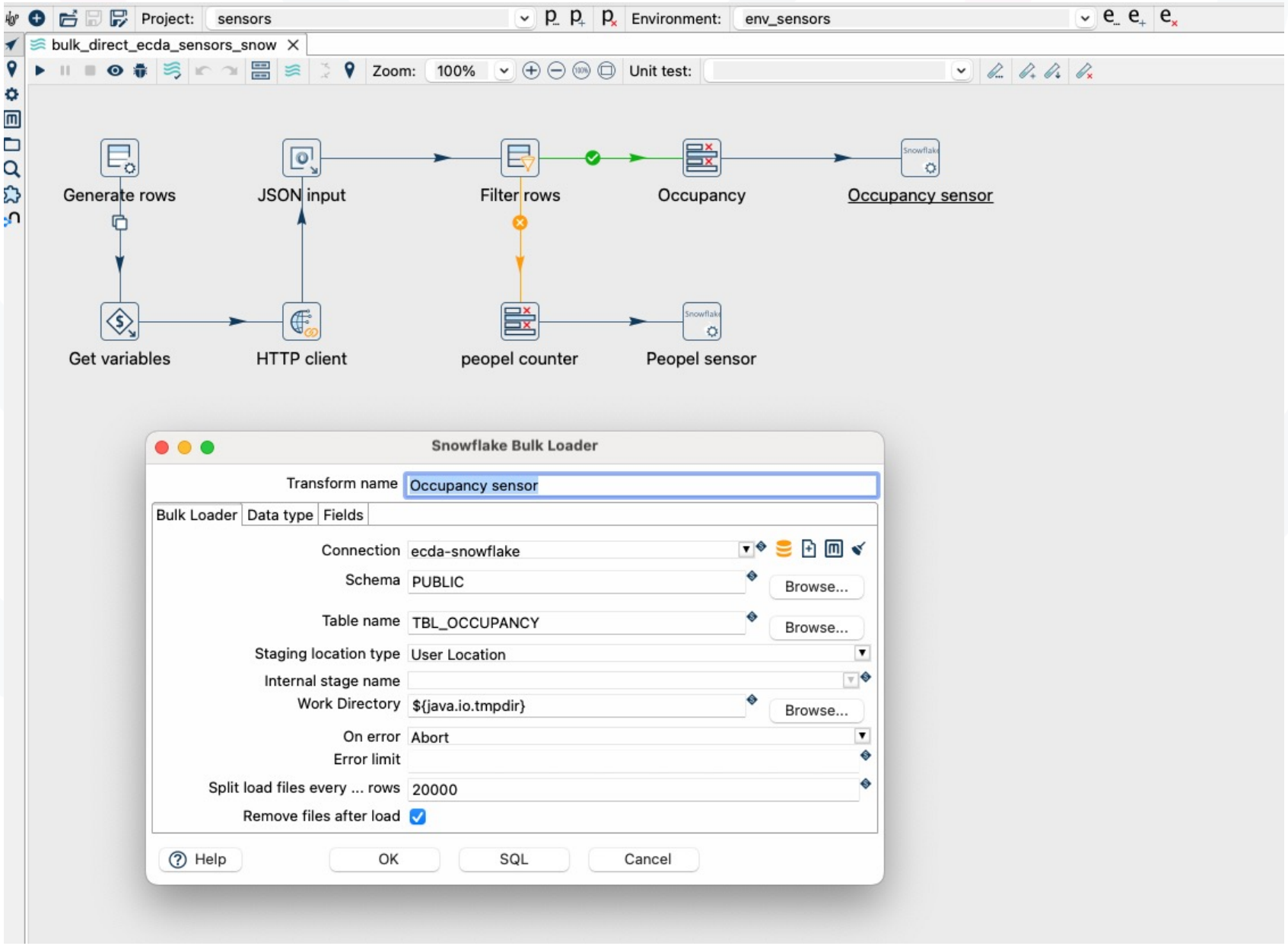
jupyter

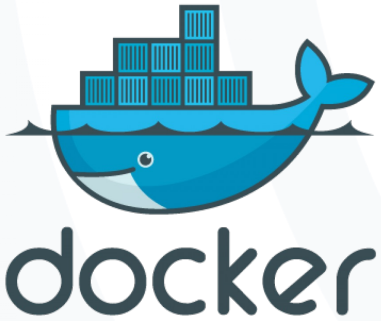
In Apache HOP, the data is extracted from API Swagger (Measurement Company) and transferred to CrateDB





Erasmus Centre for Data Analytics


Project: sensors
Environment: env_sensors
bulk_direct_ecda_sensors_snow
Zoom: 100%
Unit test:
Generate rows
JSON input
Filter rows
Occupancy
Occupancy sensor
Get variables
HTTP client
peopel counter
Peopel sensor
Snowflake Bulk Loader
Transform name
Occupancy sensor
Bulk Loader
Data type
Fields
Connection
ecda-snowflake
Schema
PUBLIC
Browse...
Table name
TBL_OCCUPANCY
Browse...
Staging location type
User Location
Internal stage name
Work Directory
${java.io.tmpdir}
Browse...
On error
Abort
Error limit
Split load files every ... rows
20000
Remove files after load
Help
OK
SQL
Cancel



Dockerfile > ...

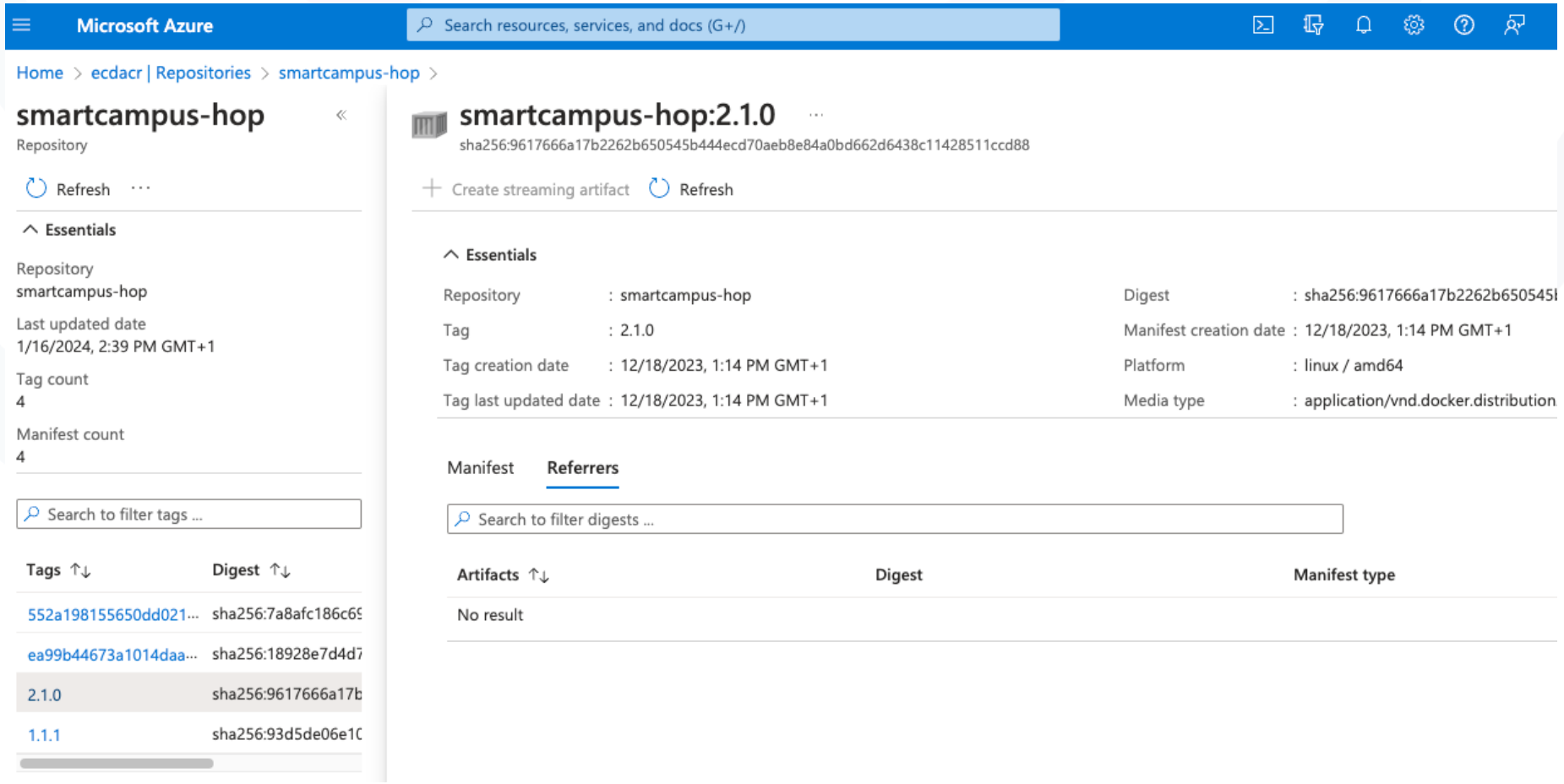
```dockerfile
# Use the base Apache Hop image
FROM apache/hop:latest

# Set the working directory in the container
WORKDIR /project

# Copy your project files into the container.
# Adjust the paths according to where your Hop project and environment files are located on your build context.
COPY /sensors /project
COPY /sensors /project-config

# Set environment variables to match your original DockerOperator configuration
ENV HOP_LOG_LEVEL="Basic" \
    HOP_FILE_PATH="/project/bulk_ecda_sensors_snow_hourly.hwf" \
    HOP_PROJECT_DIRECTORY="/project" \
    HOP_PROJECT_NAME="sensors" \
    HOP_ENVIRONMENT_NAME="env_sensors" \
    HOP_ENVIRONMENT_CONFIG_FILE_NAME_PATHS="/project-config/env_sensors.json" \
    HOP_RUN_CONFIG="local"
```

Az login
Docker build -t ecdacr.azurecr.io/smartcampus-hop:2.1.0
Docker push ecdacr.azurecr.io/smartcampus-hop:2.1.0



Apache
Airflow

```python
with DAG('smartcampus_k8s_hop_hourly',
        default_args=default_args,
        schedule_interval='0 * * * *',
        catchup=False,
        is_paused_upon_creation=False,
        max_active_runs=1) as dag:


    start_dag = EmptyOperator(
        task_id='start_dag'
    )

    end_dag = EmptyOperator(
        task_id='end_dag'
    )

    hop = KubernetesPodOperator(
        task_id='smartcampus_k8s_hop',
        image='ecdacr.azurecr.io/smartcampus-hop:2.1.0', #modify the name, repo, and version of the container accordingly
        namespace='airflow',  # Change to your namespace if different
        name='smartcampus-hop-task',
        get_logs=True,
        image_pull_secrets='my-acr-secret',
    )

    start_dag >> hop >> end_dag
```

kubernetes

override-values.yaml

```yaml
dags:
  gitSync:
    enabled: true
    repo: git@github.com:Erasmus-Data-Collaboratory/eur-smart-campus.git
    branch: main
    subPath: "dags"
    sshKeySecret: airflow-ssh-secret
extraSecrets:
  airflow-ssh-secret:
    data: |
      gitSshKey: ''

```

Code Issues Pull requests Actions Projects 1 Security Insights Settings

Files

main

Go to file t

- .github
- airflow-helm
  - airflow-nginx
  - README.md
  - git-credentials.yaml
  - my-custom-values.yml
  - variable-configmap.yml
- dags
- .gitignore
- README.md

Apache-hop / airflow-helm /

Add file

fhasanabadi add doc for nginx and cert-manager b6da1d4 · 4 months ago History

Name	Last commit message	Last commit date
..		
airflow-nginx	add doc for nginx and cert-manager	4 months ago
README.md	add doc for nginx and cert-manager	4 months ago
git-credentials.yaml	add helm airflow instructions	4 months ago
my-custom-values.yml	add helm airflow instructions	4 months ago
variable-configmap.yml	add helm airflow instructions	4 months ago

README.md

Up and running Airflow on AKS cluster with gitSync

Prepare and create the cluster

Apache-hop Private Edit Pins Watch 1

main 1 Branch 0 Tags Go to file t Add file Code

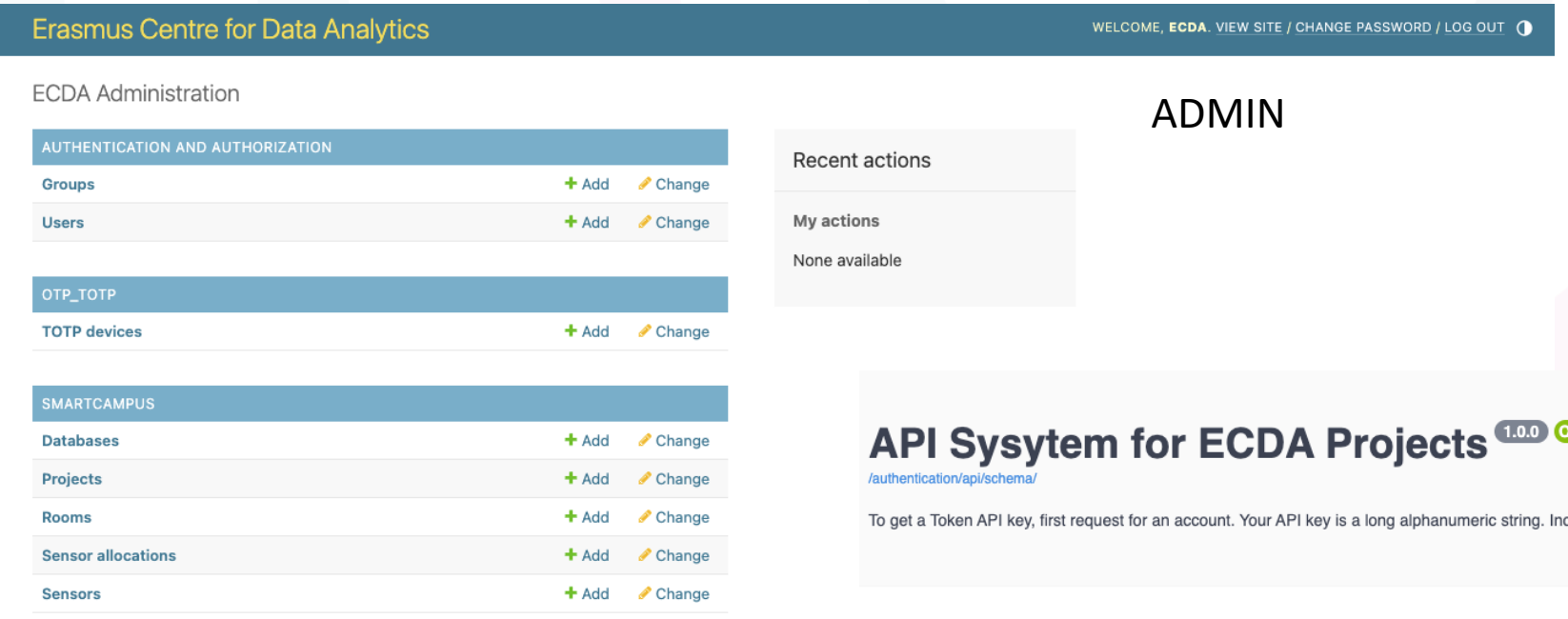
actions-user commit from Erasmus-Data-Collaboratory/eur-smart-campus 698bca8 · 2 weeks ago 119 Commits

.github/workflows	Update image-build-push.yml	4 months ago
airflow-helm	add doc for nginx and cert-manager	4 months ago
dags	commit from Erasmus-Data-Collaboratory/eur-smart-cam...	2 weeks ago
.gitignore	Added .DS_Store to .gitignore	5 months ago
README.md	edit readme	4 months ago

README

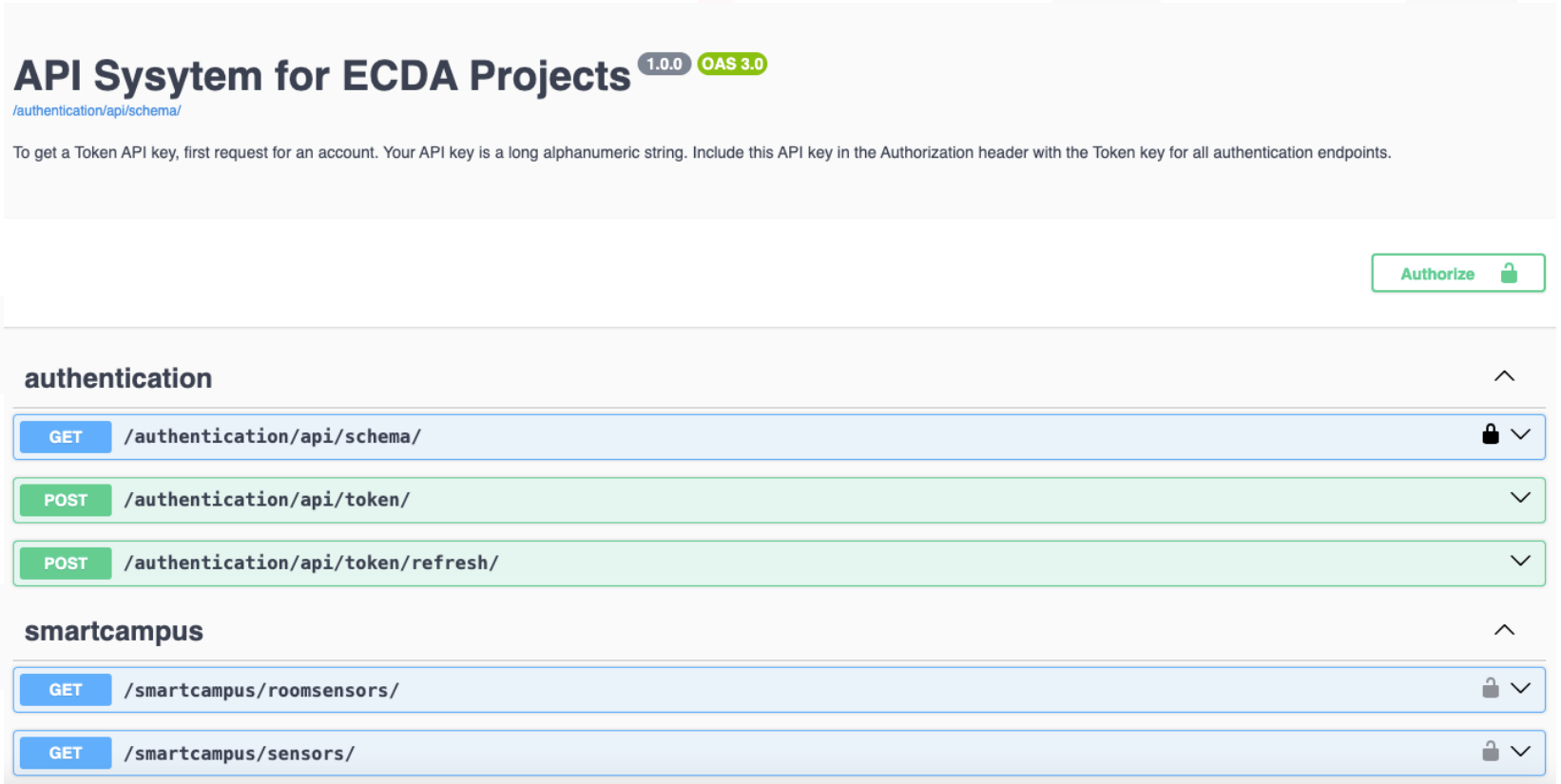
Apache-hop

Repository for Hop data projects
Including Apache HOP and Apache Airflow
The repo contains the Airflow DAGs for syncing with a Airflow instance running on AKS. The DAGs are located in the `dag` folder according to the name of the project. For example, the DAGs belong to the project `smartcampus` are in `/dags/smartcapus` .


Erasmus Centre for Data Analytics
WELCOME, ECDA. VIEW SITE / CHANGE PASSWORD / LOG OUT
ECDA Administration
AUTHENTICATION AND AUTHORIZATION
Groups + Add Change
Users + Add Change
OTP_TOTP
TOTP devices + Add Change
SMARTCAMPUS
Databases + Add Change
Projects + Add Change
Rooms + Add Change
Sensor allocations + Add Change
Sensors + Add Change
Recent actions
My actions
None available

ADMIN

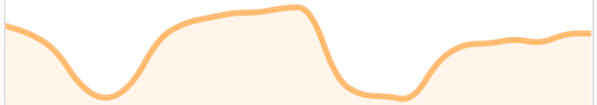
USERS


API Sysytem for ECDA Projects
1.0.0
OAS 3.0
/authentication/api/schema/
To get a Token API key, first request for an account. Your API key is a long alphanumeric string. Include this API key in the Authorization header with the Token key for all authentication endpoints.
Authorize
authentication
GET /authentication/api/schema/
POST /authentication/api/token/
POST /authentication/api/token/refresh/
smartcampus
GET /smartcampus/roomsensors/
GET /smartcampus/sensors/

Y1-08

20.8 °C
Temperature

Temperature Trend
20.8 °C



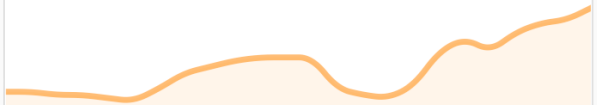
478.0 ppm
Co2 level

Humidity	44.0 %
Presence	Detected
Desk 1	38.8 W
Desk 2	10.6 W
Coffee Machine	0.0 W

Y1-10

21.9 °C
Temperature

Temperature Trend
21.9 °C



740.0 ppm
Co2 level

Humidity	47.0 %
Presence	Clear

Y1-12

22.3 °C
Temperature

Temperature Trend
22.3 °C

513.0 ppm
Co2 level

Humidity	42.0 %
Presence	Detected
Desk 1	73.3 W
Desk 2	18.1 W

Y1-13

21.1 °C
Temperature

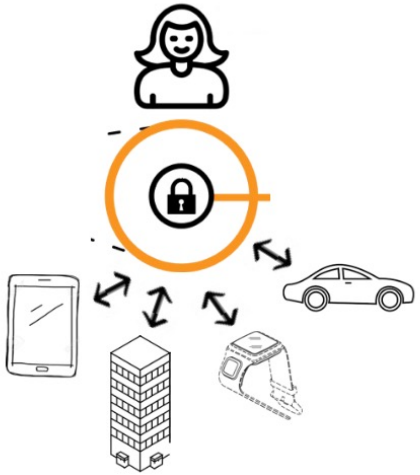
Temperature Trend
21.1 °C

459.0 ppm
Co2 level

Humidity	42.0 %
Presence	Clear

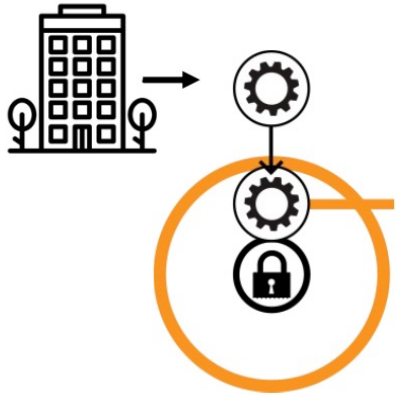
New Pilot: bubl cloud

1 Demonstratable Security & Privacy



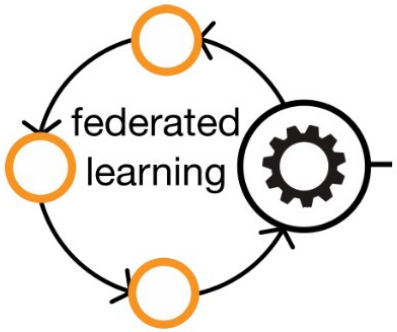
Personal encrypted data vault

2 Use AI on combined sensitive personal data



Sending the service to the data

3 Get detailed insights privacy safe



Privacy safe exchange

Erasmus Centre for
Data Analytics

Questions?

Erasmus
University
Rotterdam