The Data Analytics Platform

Rob Keevil & Krzysztof Adamski

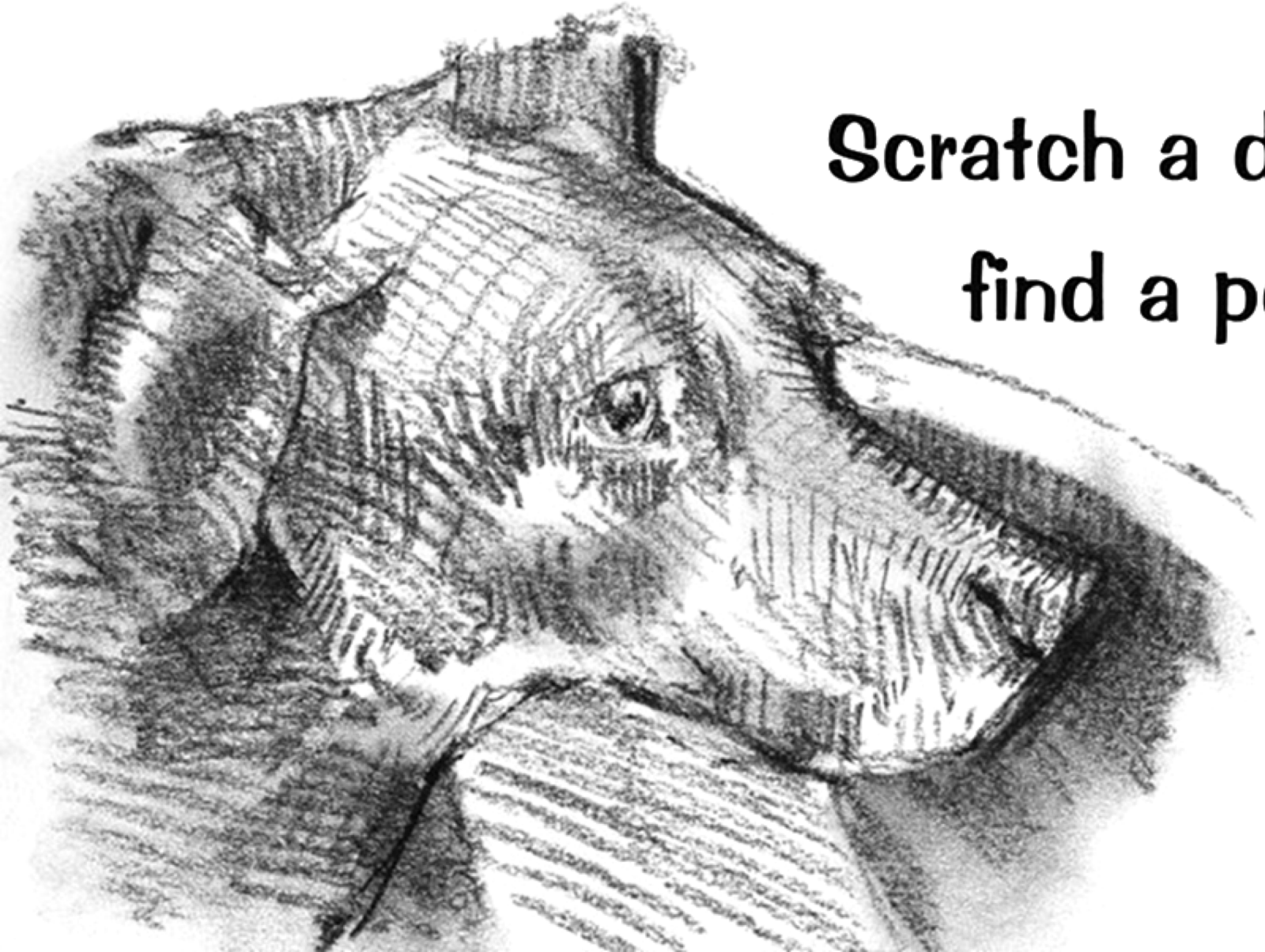
Barcelona 22/05/2019

ING

The Mission

The Journey So Far

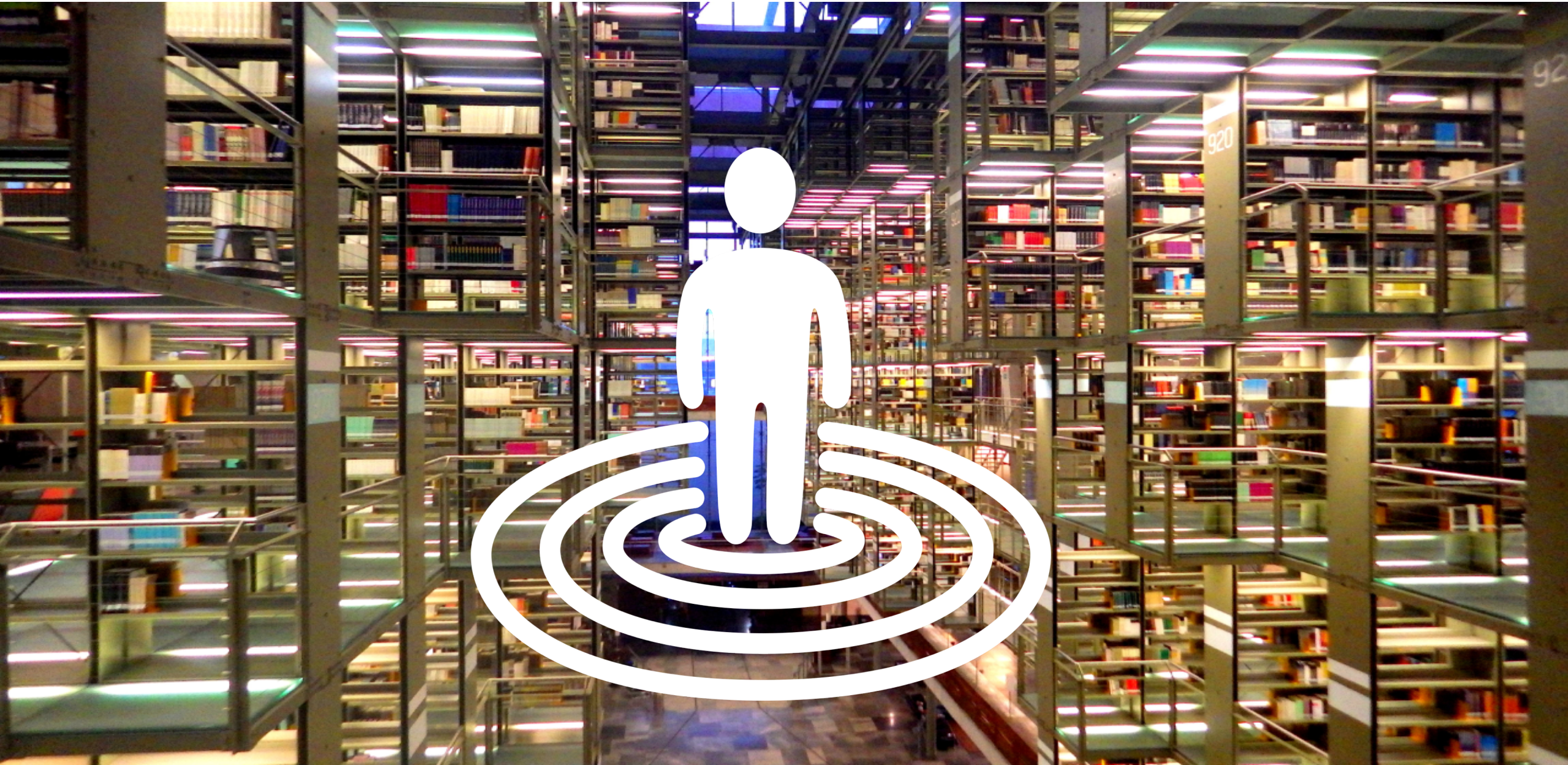
⚠ The Problem



Scratch a dog and you'll find a permanent job.

Franklin P. Jones

Our Solution



The Rise Of Microservices

ING

The Architecture

ING

Front End

Portal

Compute

Storage

Security + Integration

Front End

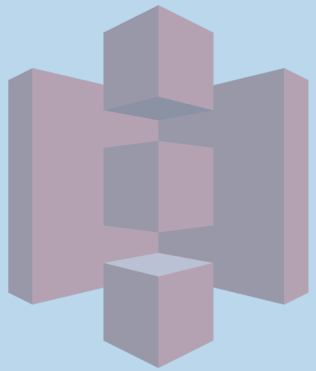
ING

Front End

Portal

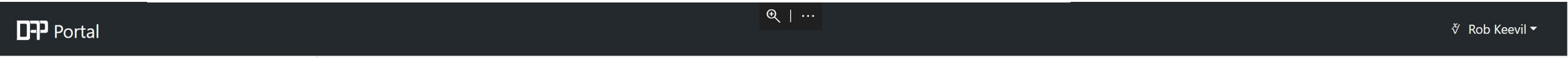
Compute

Storage



Security + Integration

DAP Portal



- Service Catalog

DATA

- Data Discovery

STORAGE

- Buckets

ENVIRONMENTS

- Data Science in a Box

UTILITIES

- Create HDFS Home
- XRDP

GRID

GRID contains company data for ING customers. It contains information about the business clients of ING, such as name, risk rating, sector, industry

Basic Information

Access Policies	DATA_PII
Contact Person(s)	Anabel [redacted]@ing.com)
Created	01 January 2018 0:01 AM
Exploration	True
Last Modified	10/27/2018 2:53 AM
Lineage	atlas
Location	hdfs://datalab/apps/hive/warehouse/data_pii/grid.db/x_grid
Records	12757010
Size	410.51 MB

Table Information (grid.x_grid)

Column Name	Type	Description
grid_id	int	Local ID in GRID of company
grid_leg_parent_id	int	Legal parent entity id
grid_leg_up_id	int	Legal ultimate parent entity id
grid_ecg_parent_id	int	Economic parent entity id
grid_ecg_up_id	int	Economic ultimate parent entity id

Frequent Users

Data Quality

★ ★ ★ ★ ☆

Usage Frequency

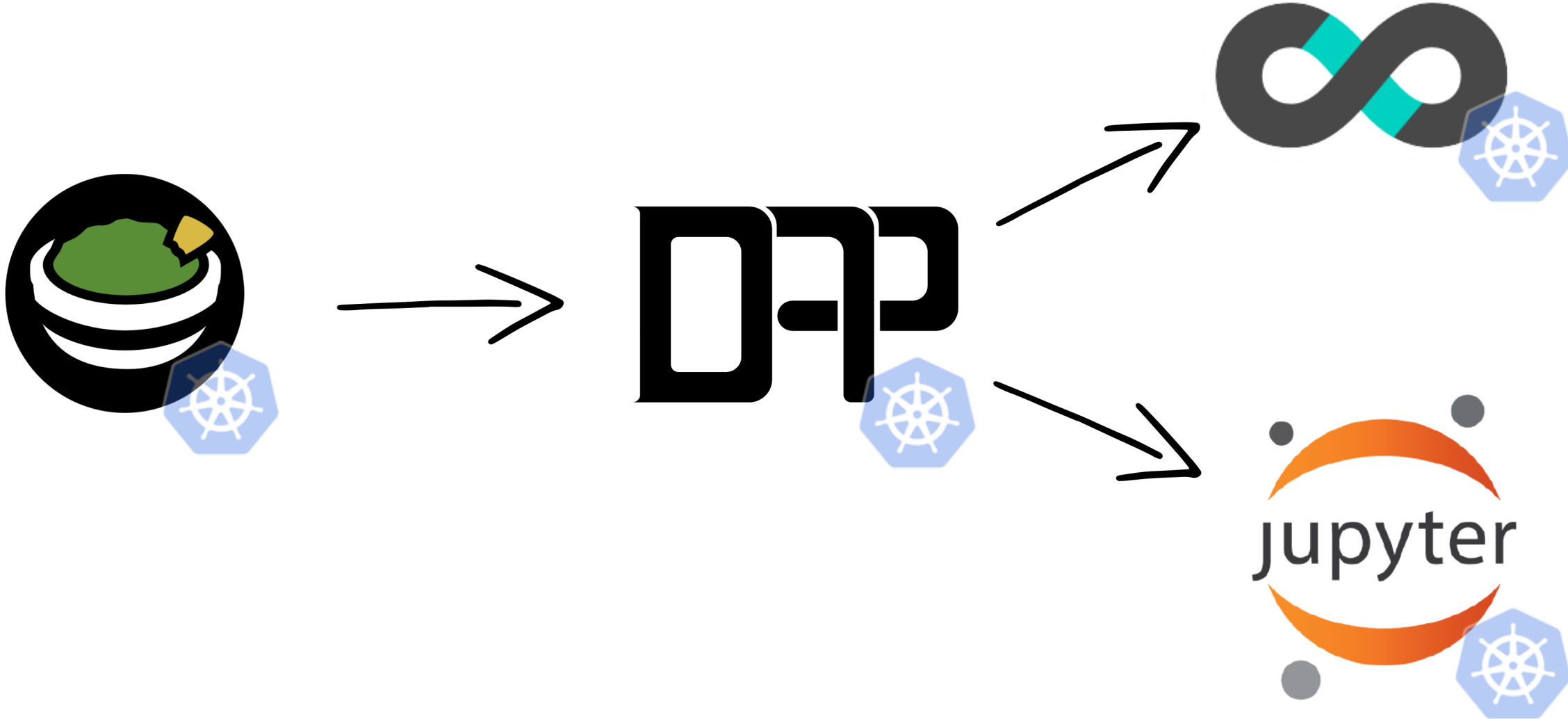
Request Access | Follow

Tags

grid · companies · company data · customers · clients · ing customers · ing clients · name · country · town · net sales · ing risk rating · naics code · naics desc · website · customer type desc · customer segment desc · postcode · is exchange listed

Front End

ING



Compute

ING

Front End

Portal

Compute

Storage

Security + Integration

Compute

ING

Front End

Portal

jupyter

Compute

Apache Spark

druid

presto

Storage

Rokku

ceph

Security + Integration

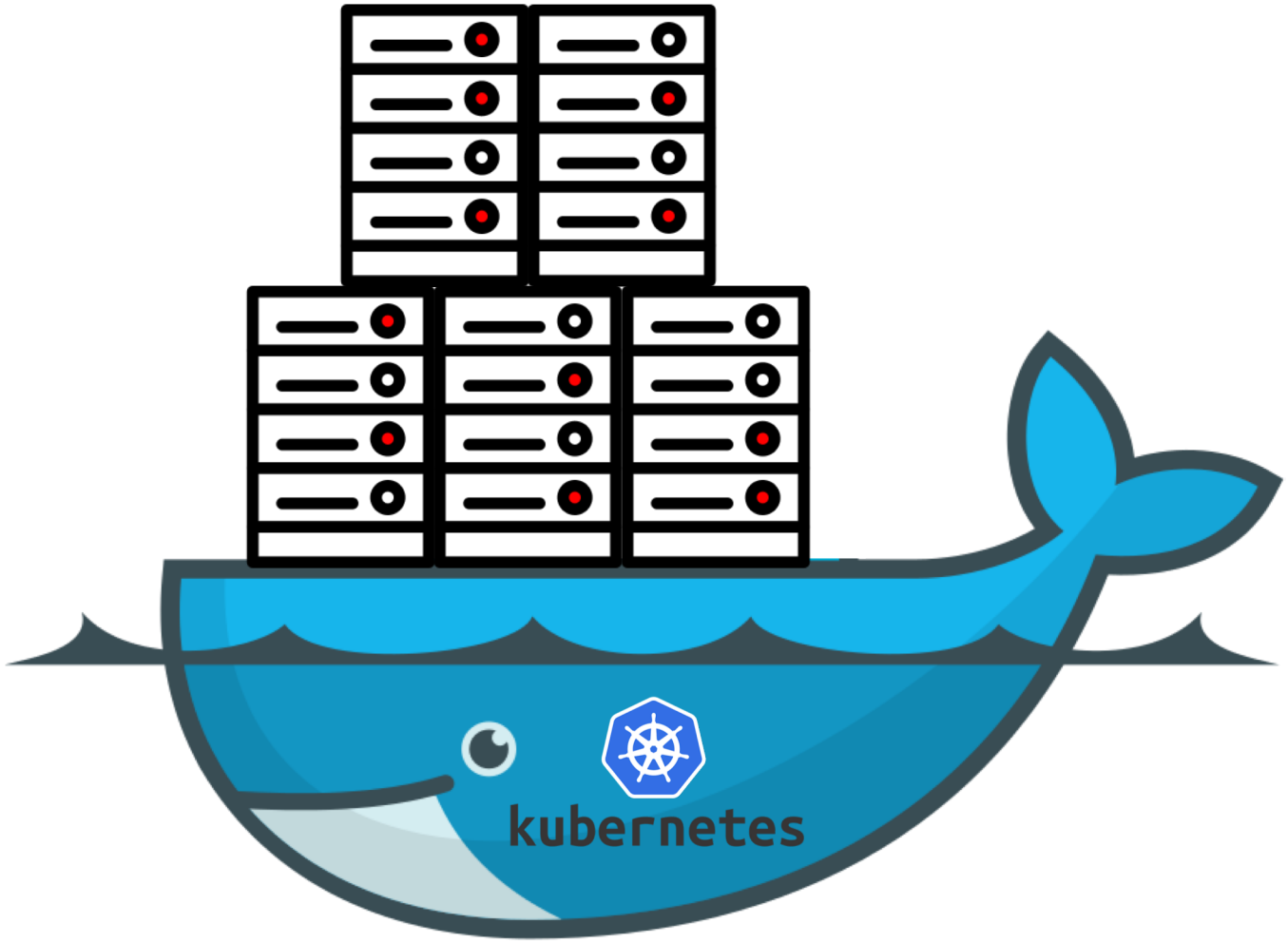
Apache Atlas

Apache Ranger

ODPi EGERIA

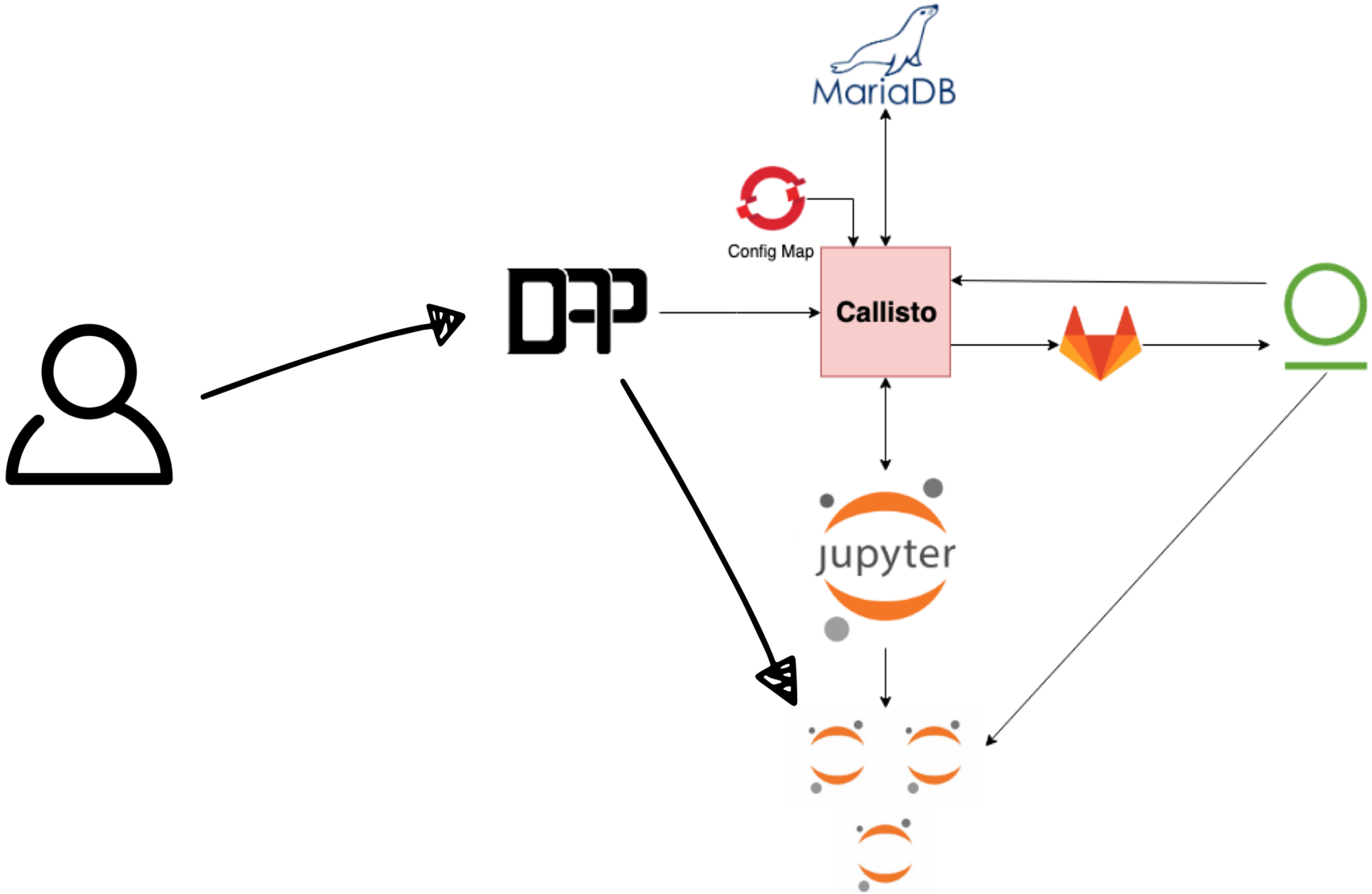
Airflow

Containers as Compute



Data Science in a Box

Data Science in a Box



Storage

ING

Front End

Portal

Compute

Storage

Security + Integration

Storage

ING

Front End

Portal

jupyter

Superset

Compute

APACHE Spark

druid

presto

Storage

Rokku

ceph

Security + Integration

Apache Atlas

Apache Ranger

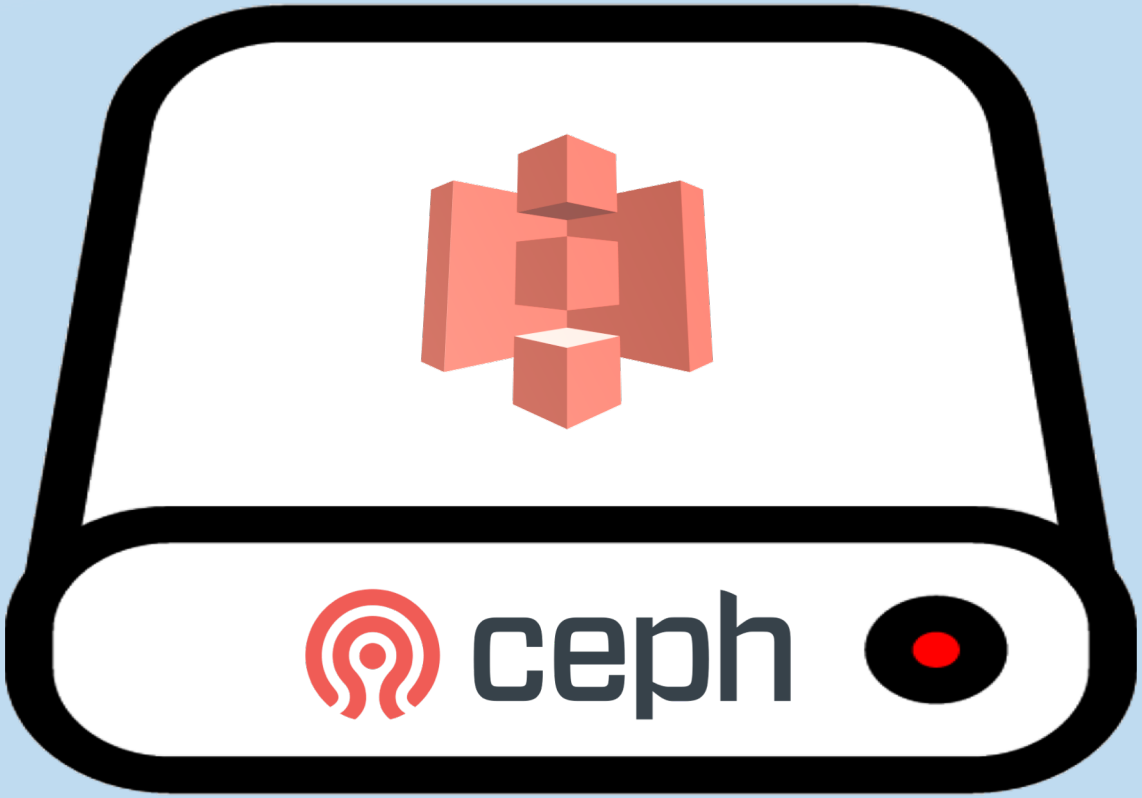
ODPi EGERIA

Airflow

Unified Storage

ING

Rokku



Security & Integration

ING

Front End

Portal

jupyter

Compute

APACHE Spark™

druid

presto

Storage

Rokku

ceph

Security & Integration

Apache Atlas

Apache Ranger

ODPi EGERIA

Airflow

Security & Integration

ING

Front End

Portal

jupyter

Compute

Spark

druid

presto

Storage

Rokku

ceph

Security & Integration

Apache Atlas

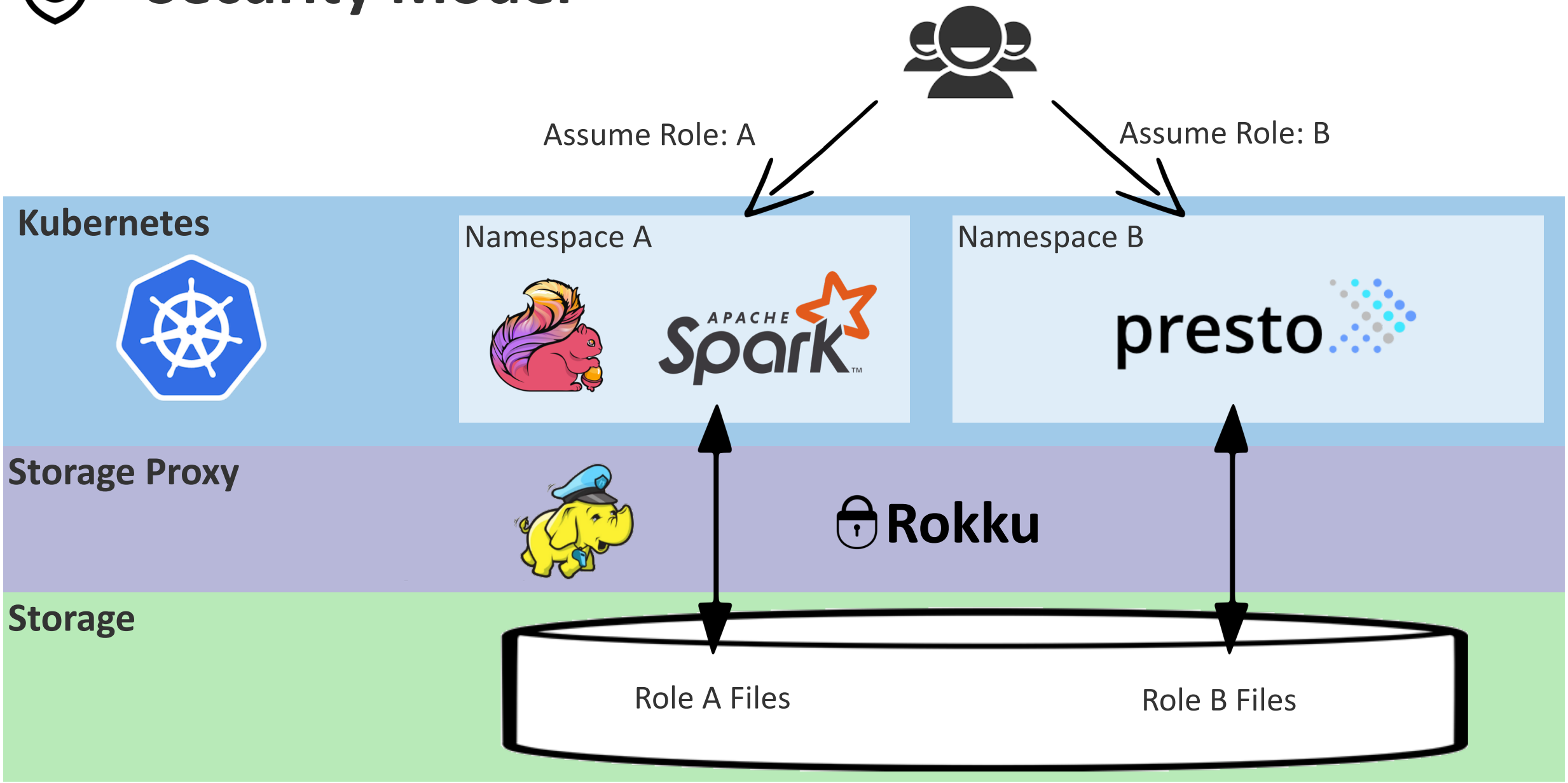
Apache Ranger

ODPi EGERIA

Airflow

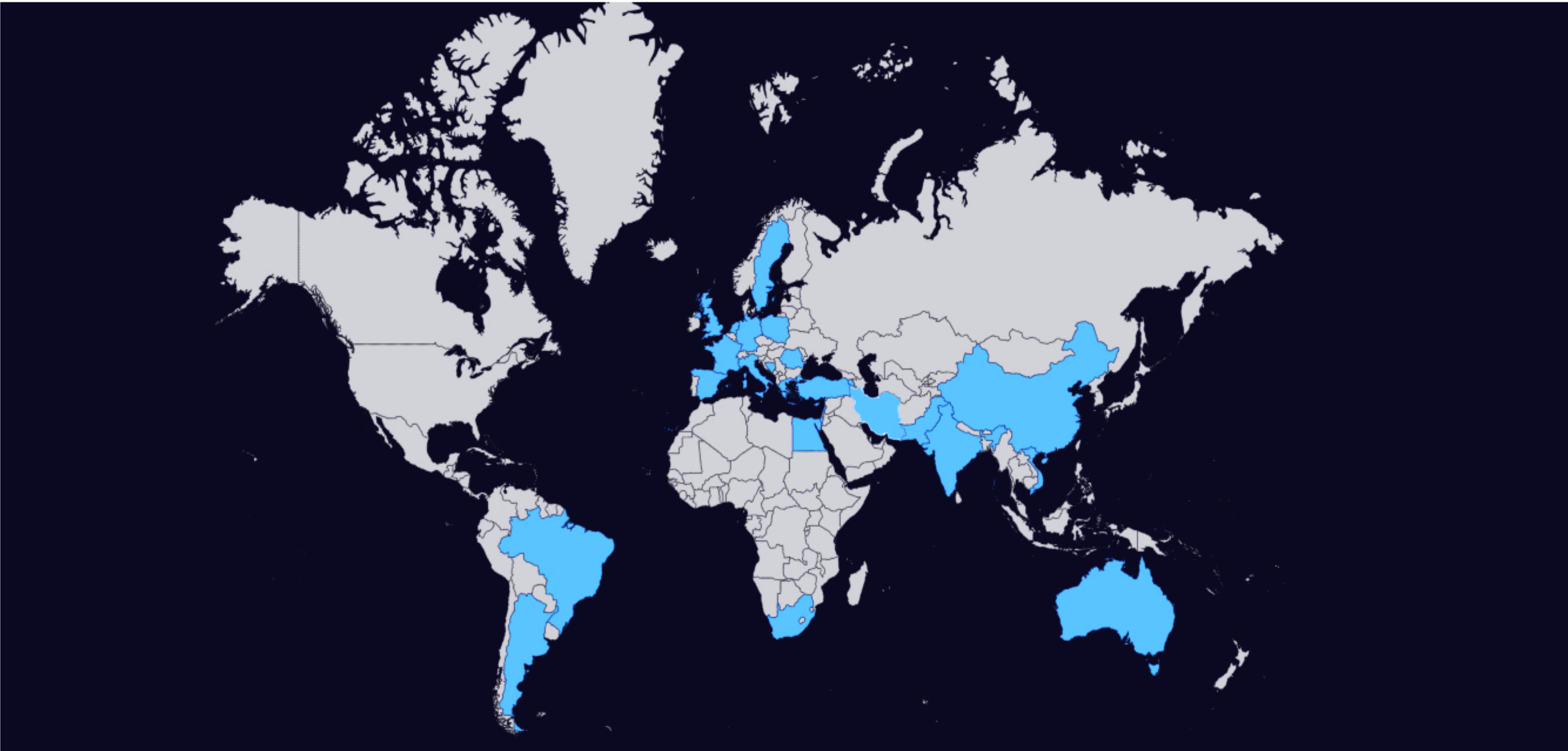
Security, Security, Security.

Security Model



Data Lineage

The Team



THANK
YOU!