

The Data Analytics Platform

Rob Keevil & Krzysztof Adamski

Barcelona 22/05/2019



The Mission

ING





The Journey So Far

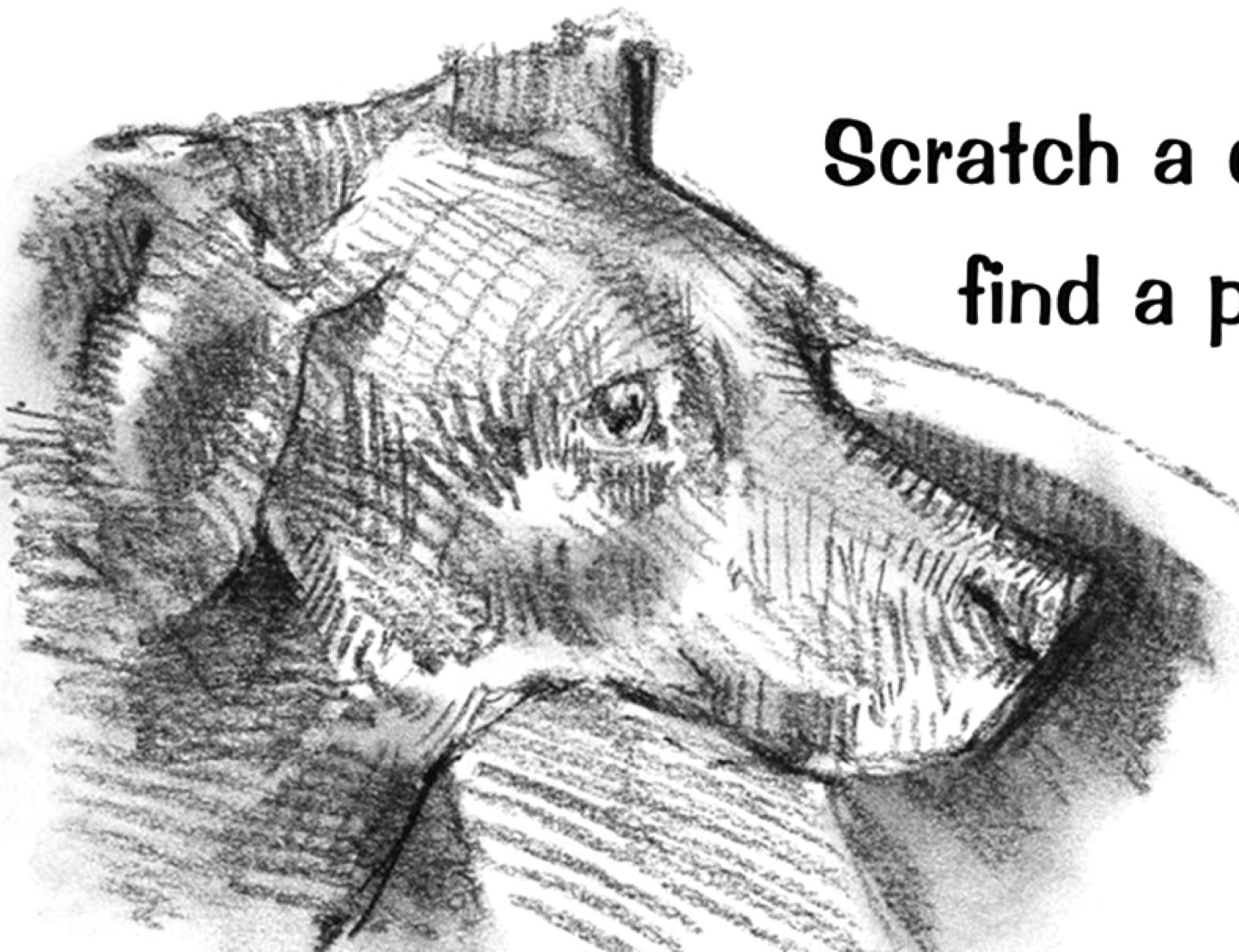
ING





The Problem

ING The ING lion logo, a golden lion standing on its hind legs with its front paws raised, holding a ring in its front paw.



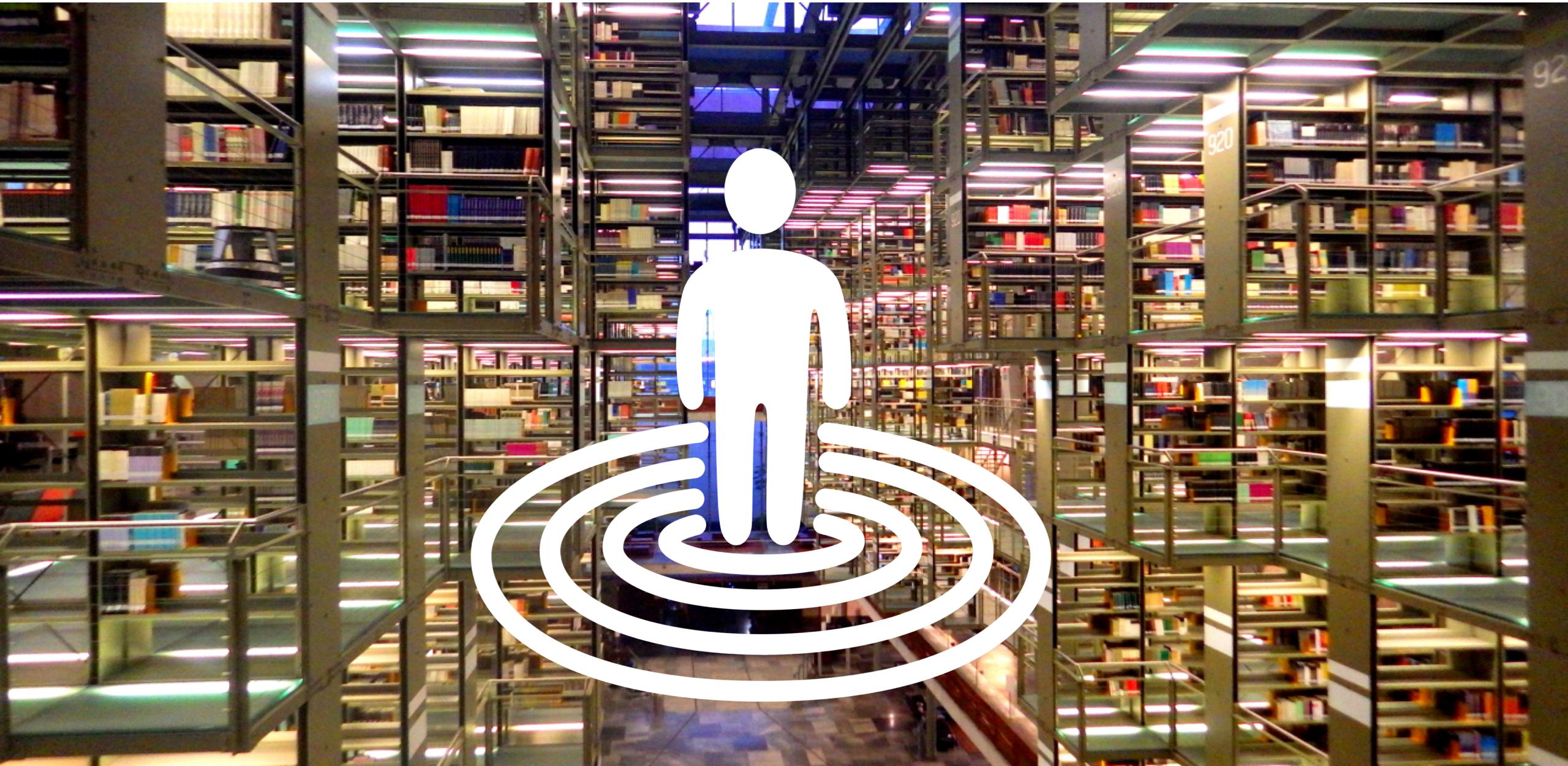
**Scratch a dog and you'll
find a permanent job.**

Franklin P. Jones



Our Solution

ING





The Rise Of Microservices

ING



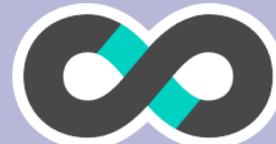


The Architecture

ING

Front End

Portal



Compute



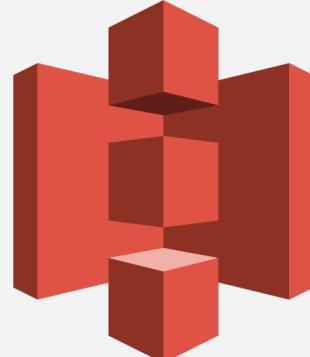
APACHE
Spark™



druid

presto

Storage

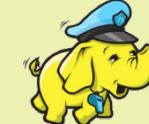


Rokku

ceph

Security + Integration

Apache **Atlas**



Apache **Ranger**

ODPi
EGERIA

Airflow



The Architecture

ING 

Front End

Portal



Compute

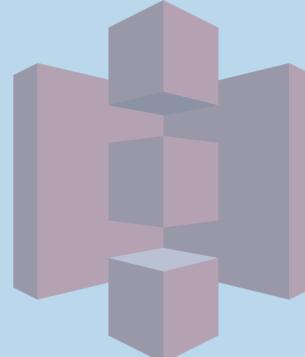


APACHE
Spark™



druid
presto

Storage



Rokku

ceph

Security + Integration

Apache **Atlas**



Apache Ranger

ODPI
EGERIA



Airflow

 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	
 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

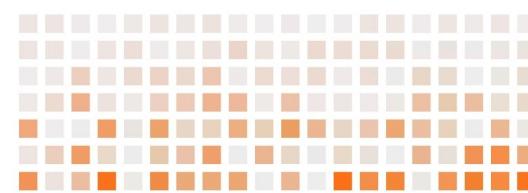
Frequent Users



Data Quality



Usage Frequency



 Request Access

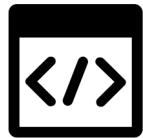
Follow

Table Information (grid.x_grid)

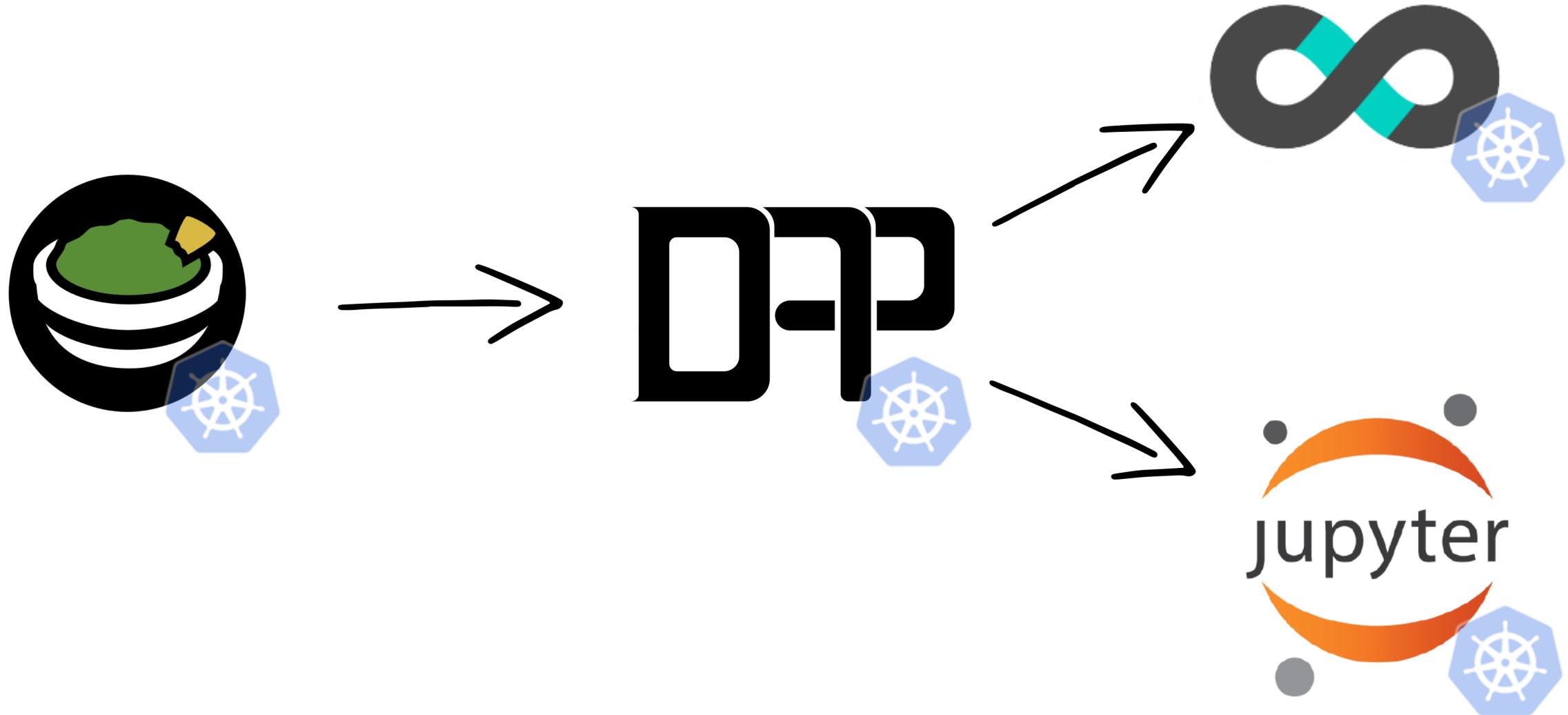
Column Name	Type	Description
grid_id	int	Local ID in GRID of company
grid_leg_parent_id	int	Legal parent entity id
grid_leg_up_id	int	Legal ultimate parent entity id
grid_ecg_parent_id	int	Economic parent entity id
grid_ecg_up_id	int	Economic ultimate parent entity id

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



Frontend



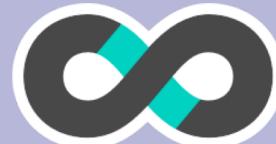


The Architecture

ING 

Front End

Portal



Compute

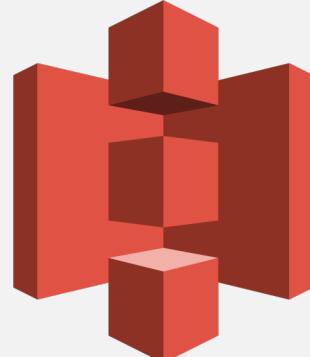


APACHE
Spark™



druid
presto

Storage



Rokku

ceph

Security + Integration

Apache **Atlas**



Apache **Ranger**

ODPi
EGERIA

Airflow



The Architecture

ING 

Front End

Portal



Compute

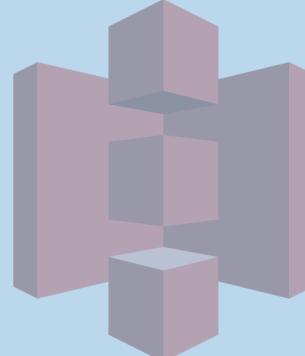


APACHE
Spark™



-d druid
presto

Storage



Rokku

ceph

Security + Integration

Apache **Atlas**

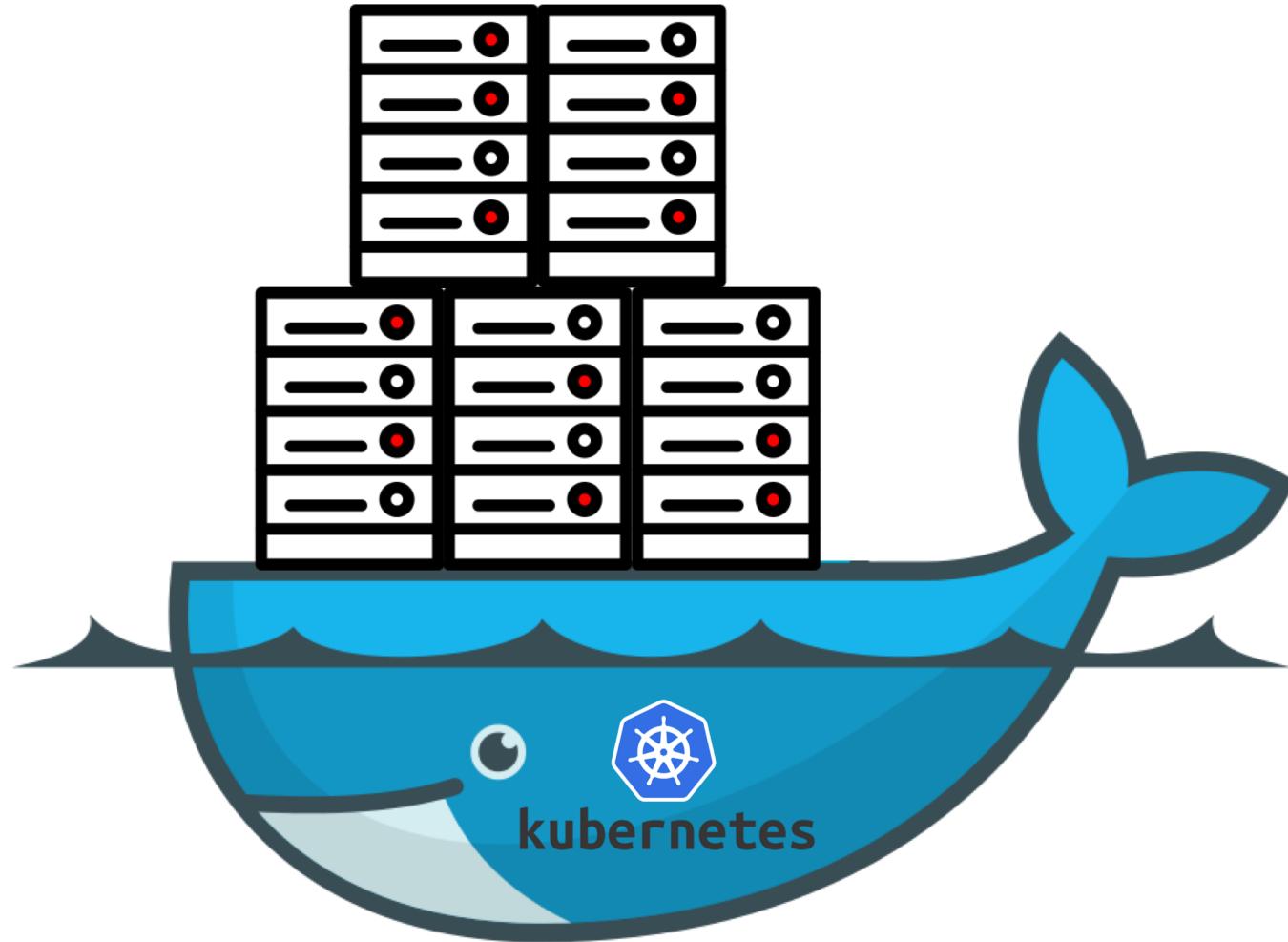
Apache Ranger

ODPI-EGERIA

Airflow



Containers as Compute





Data Science in a Box

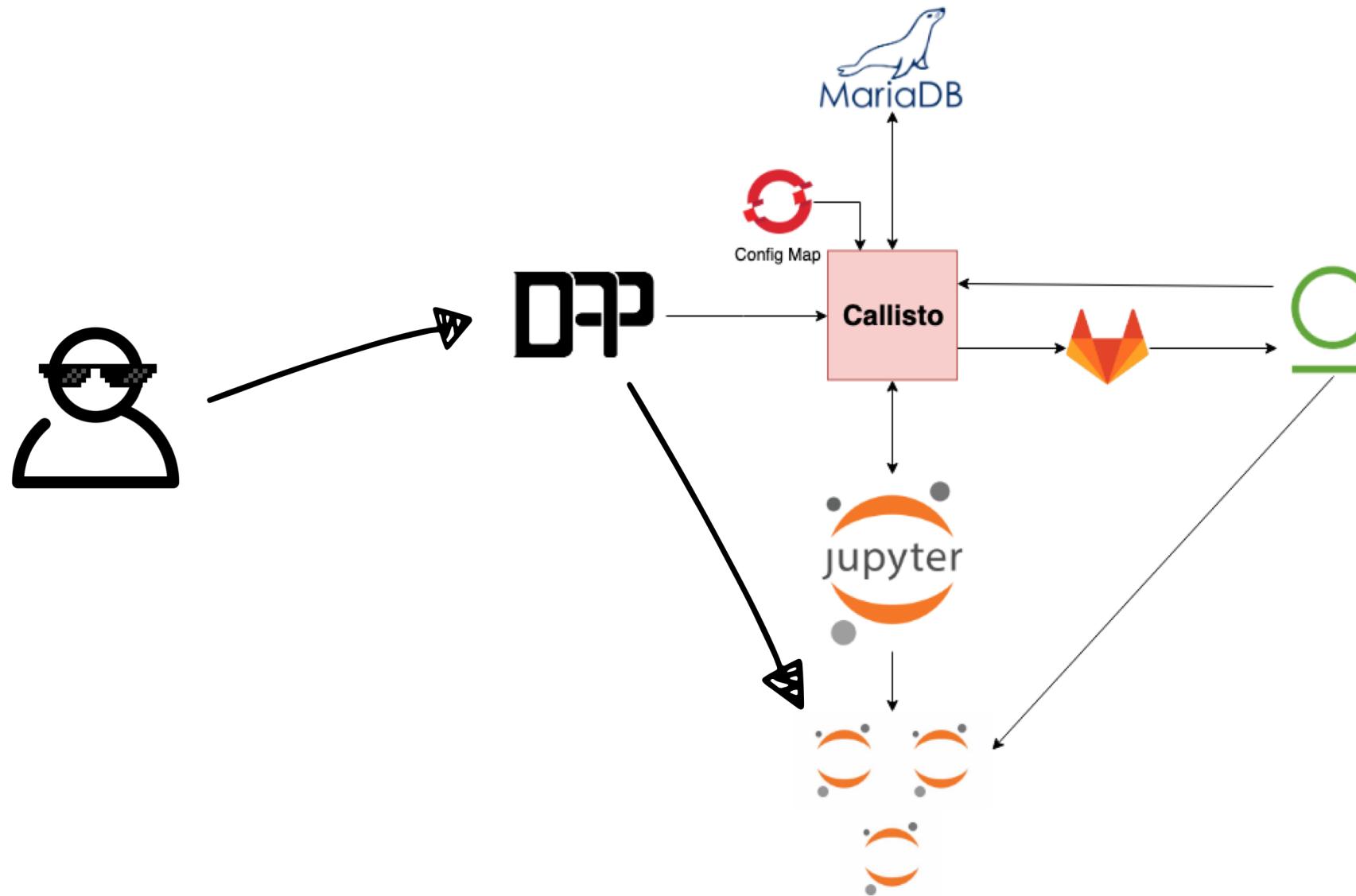
ING





Data Science in a Box

ING





The Architecture

ING 

Front End

Portal



Compute

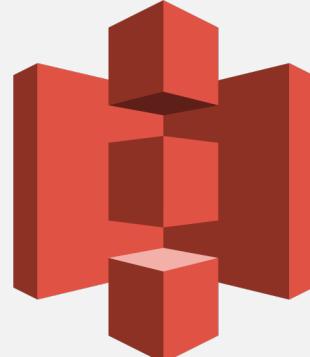


APACHE
Spark™



druid
presto

Storage



Rokku

ceph

Security + Integration

Apache **Atlas**



Apache **Ranger**

ODPi
EGERIA

Airflow



The Architecture

ING 

Front End

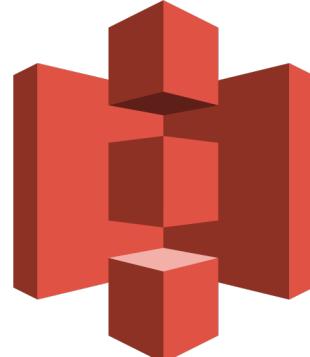
Portal



Compute



Storage



Rokku



Security + Integration

Apache **Atlas**



Apache Ranger





Unified Storage

ING

Rokku



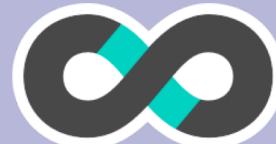


The Architecture

ING 

Front End

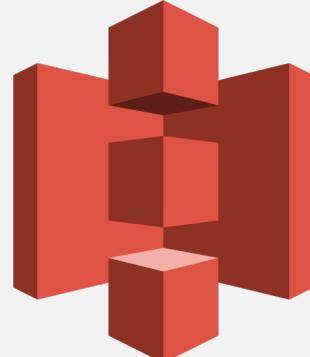
Portal



Compute



Storage



Security + Integration

Apache **Atlas**



Apache **Ranger**



ODPi EGERIA



Airflow



The Architecture

ING 

Front End

Portal



Compute

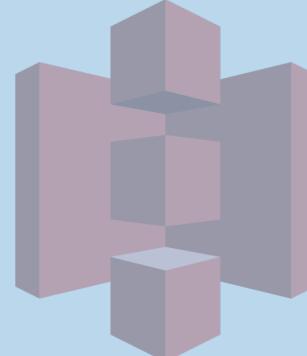


APACHE
Spark™



druid
presto

Storage



Rokku

ceph

Security + Integration

Apache **Atlas**



Apache **Ranger**

ODPi **EGERIA**



Airflow



Security, Security, Security.

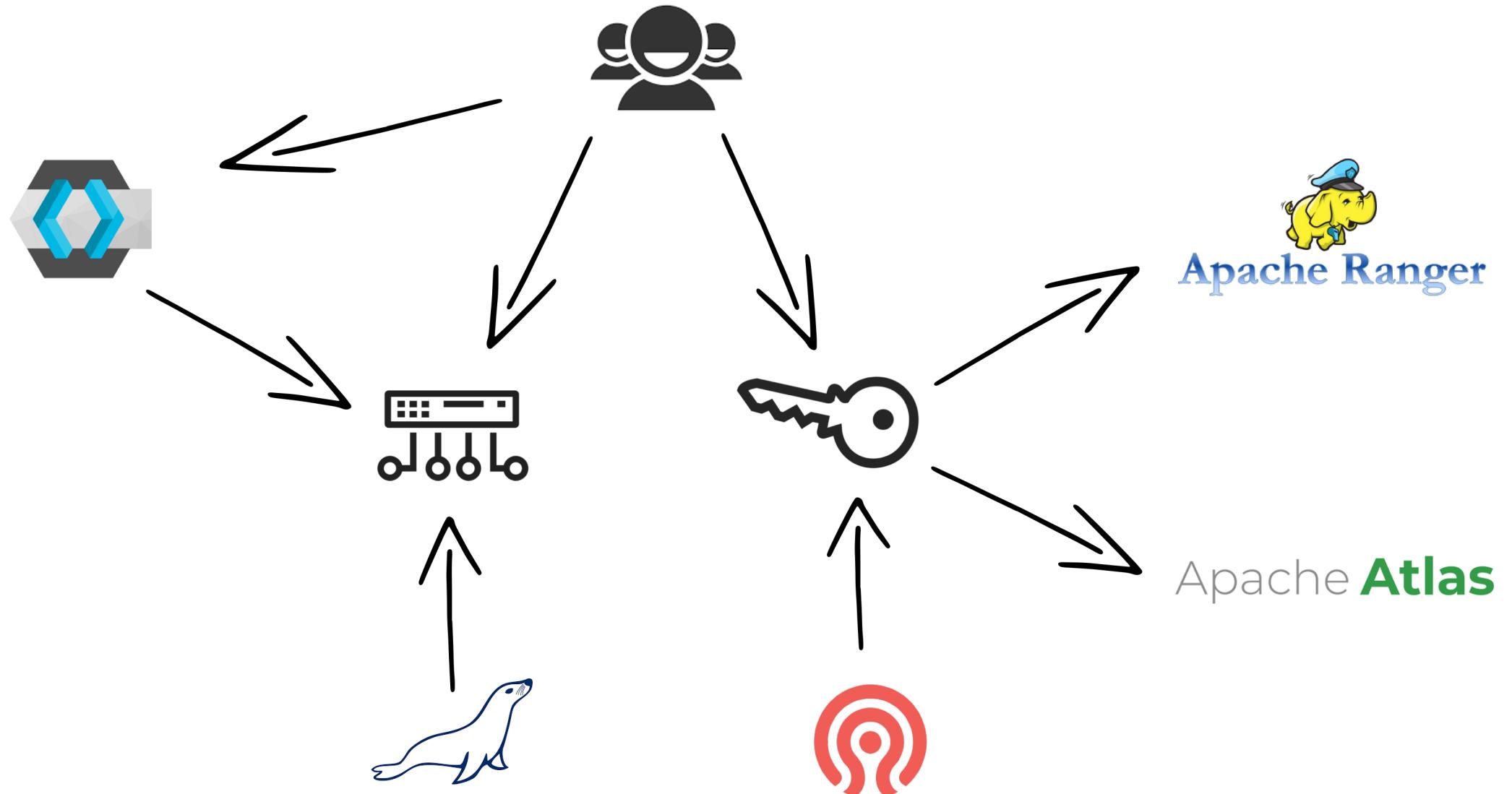
ING 





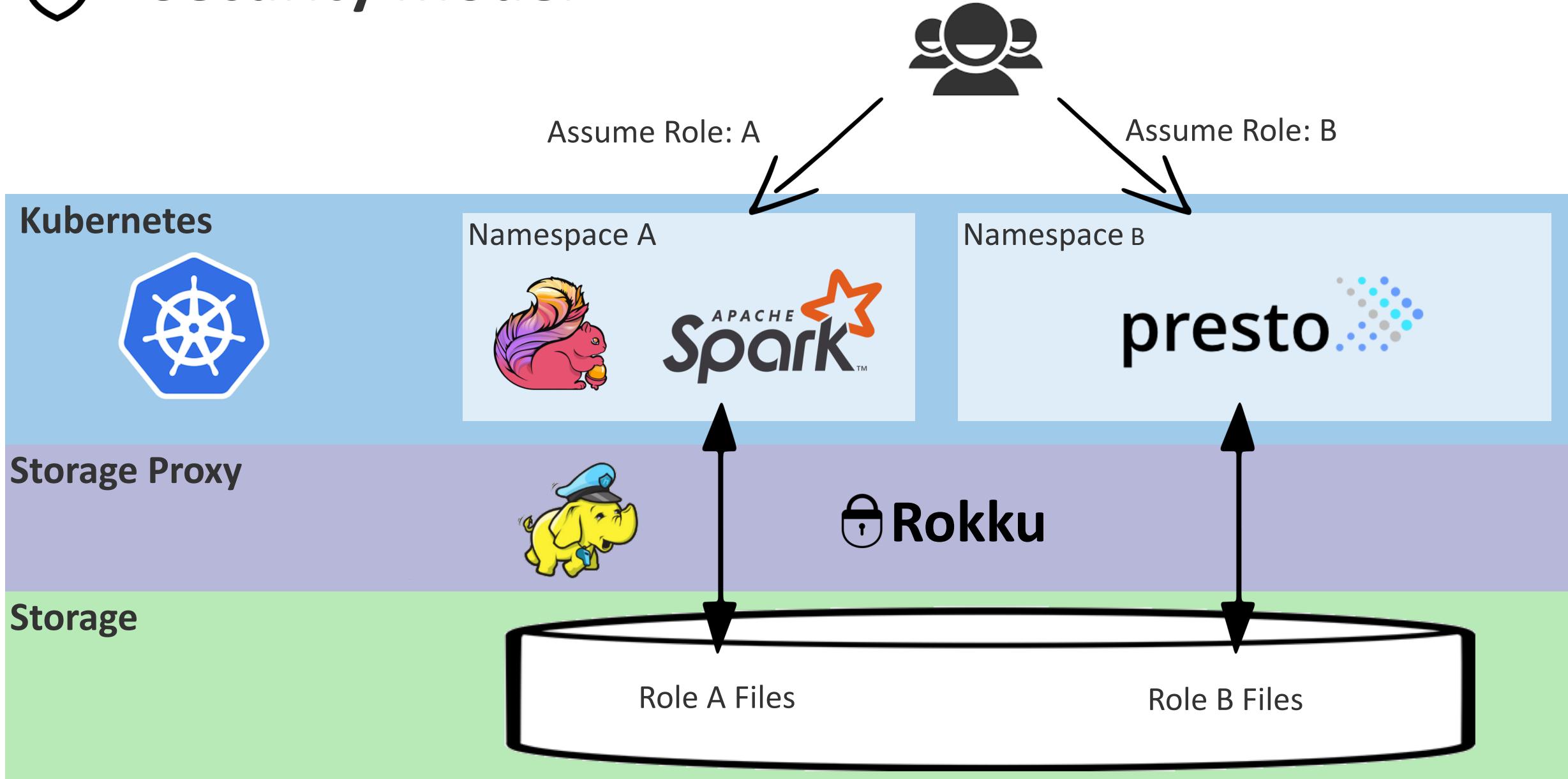
Authorization layer

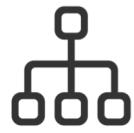
<https://github.com/ing-bank/rokku>





Security Model





Data Lineage

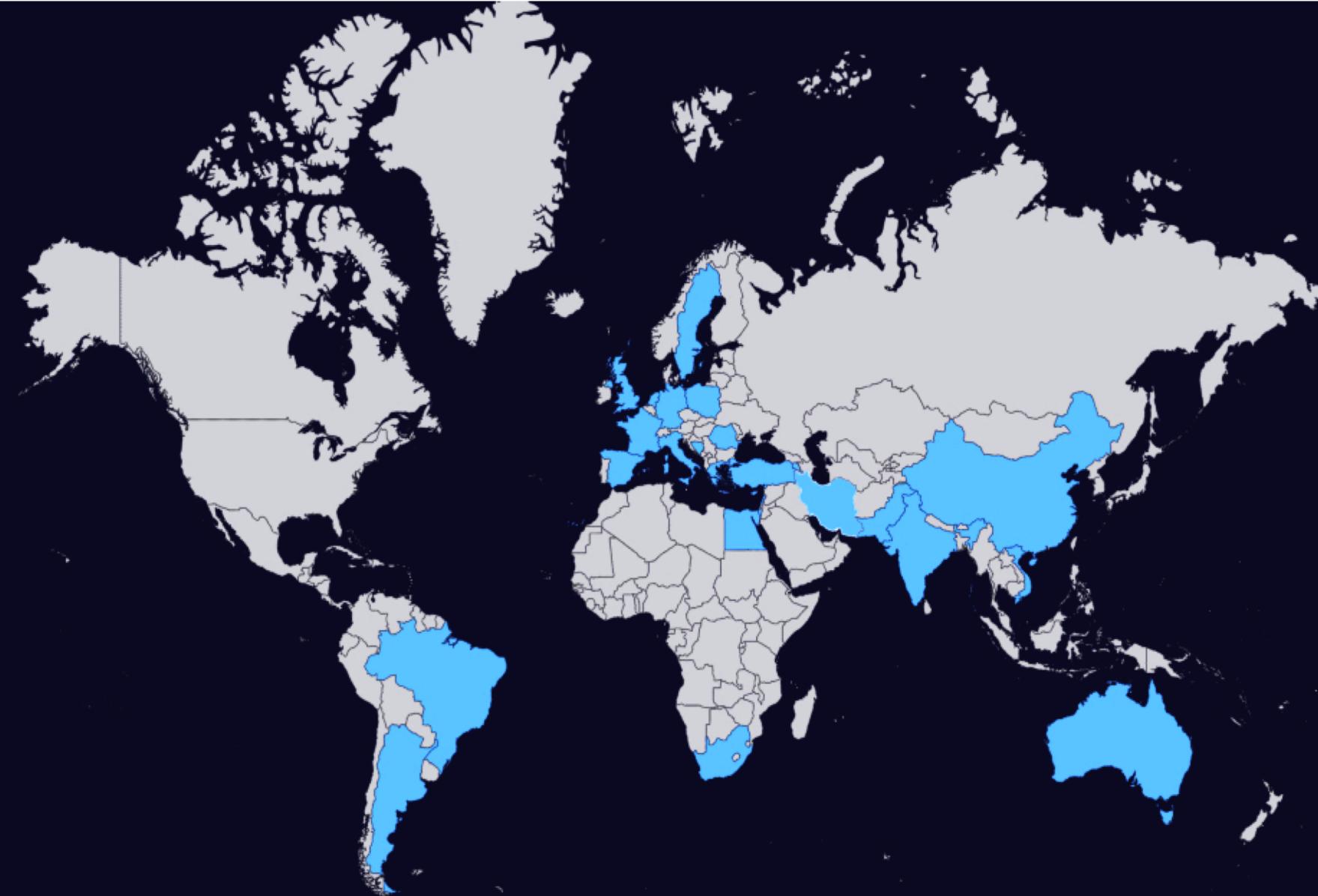
ING 





The Team

ING



**THANK
YOU!**