

The Data Analytics Platform

Rob Keevil & Krzysztof Adamski

Barcelona 22/05/2019

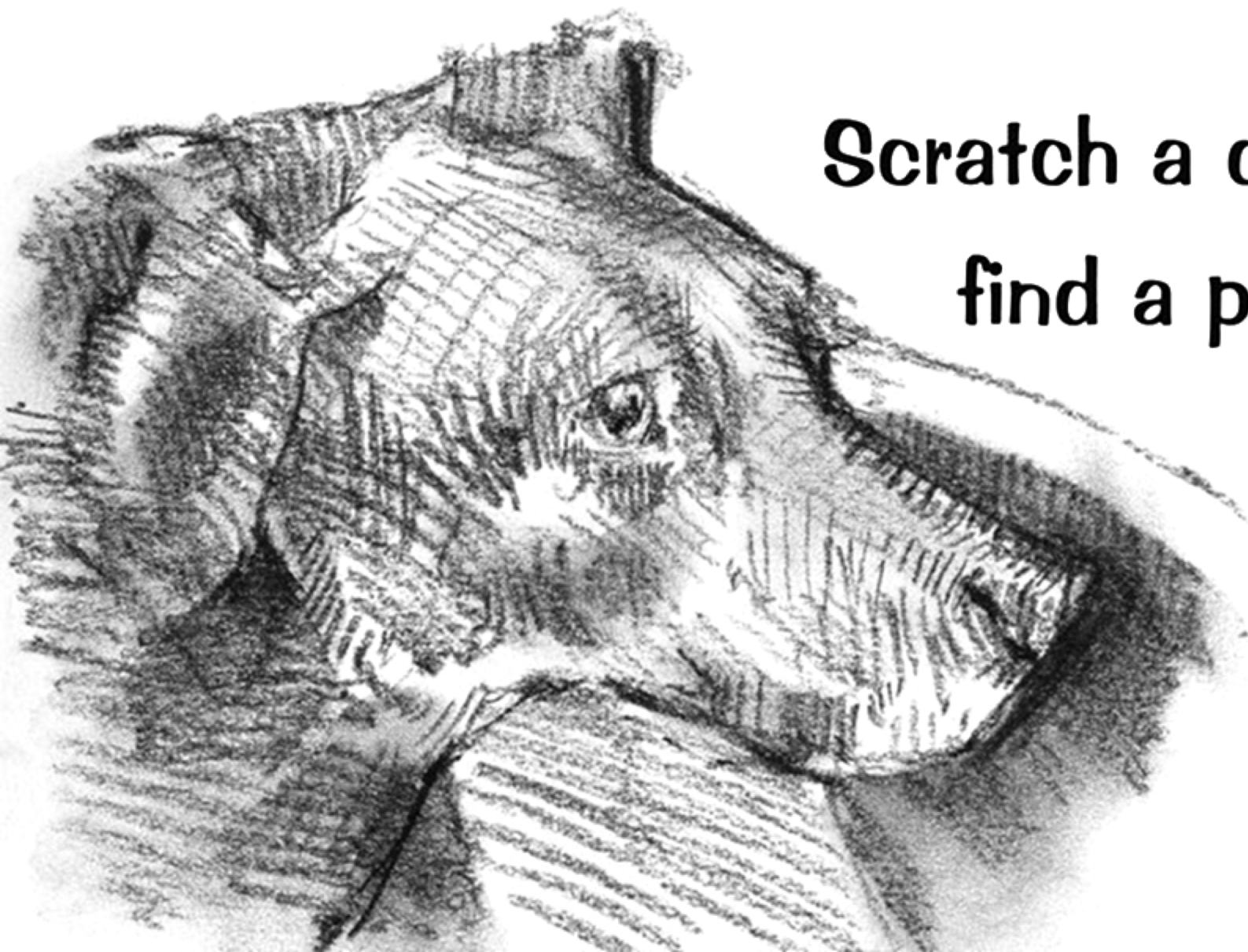


The Journey So Far





The Problem

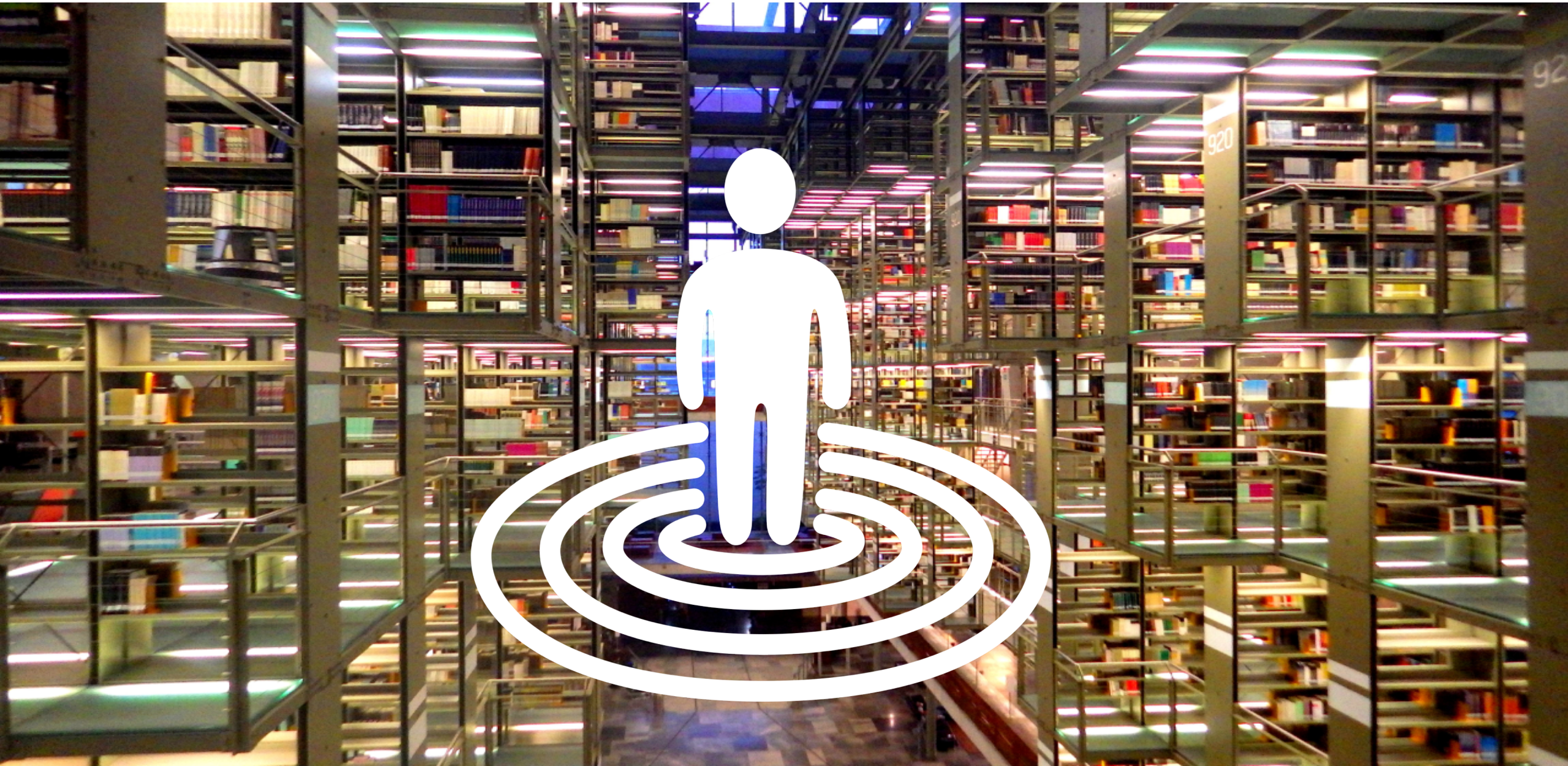


**Scratch a dog and you'll
find a permanent job.**

Franklin P. Jones



Our Solution





The Rise Of Microservices





The Architecture



Front End

Portal



Security + Integration

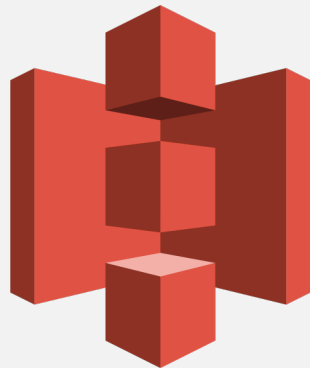
Apache **Atlas**

Compute



Apache Ranger

Storage





The Architecture

Front End

Portal



Security + Integration

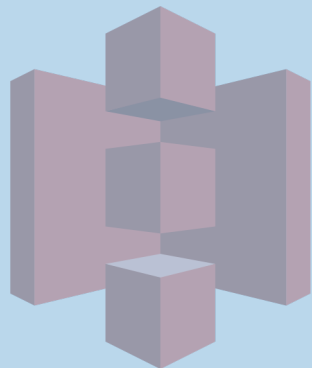
Apache **Atlas**

Compute



Apache Ranger

Storage



Rokku



ceph



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	DATA PII
	Contact Person(s)	Anabel . [redacted] @ing.com
<input type="checkbox"/>	Created	01 January 2018 0:01 AM
	Exploration	True
<input type="checkbox"/>	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

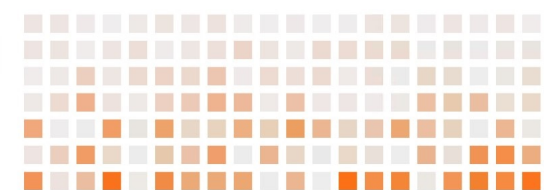
Frequent Users



Data Quality



Usage Frequency



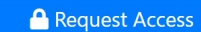




Table Information (grid.x_grid)

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

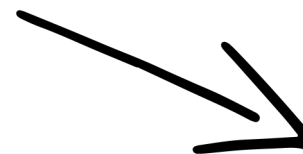
- grid
- companies
- company data
- customers
- clients
- ing customers
- ing clients
- name
- country
- town
- net sales
- ing risk rating
- naics code
- naics desc
- website
- customer type desc
- customer segment desc
- postcode
- is exchange listed



Frontend



DIP





The Architecture

Front End

Portal



Security + Integration

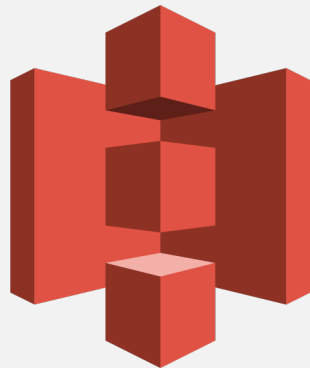
Apache **Atlas**

Compute



Apache Ranger

Storage





The Architecture



Front End

Portal



Security + Integration

Apache **Atlas**

Compute



APACHE **Spark**



druid

presto

Apache Ranger

Storage



Rokku

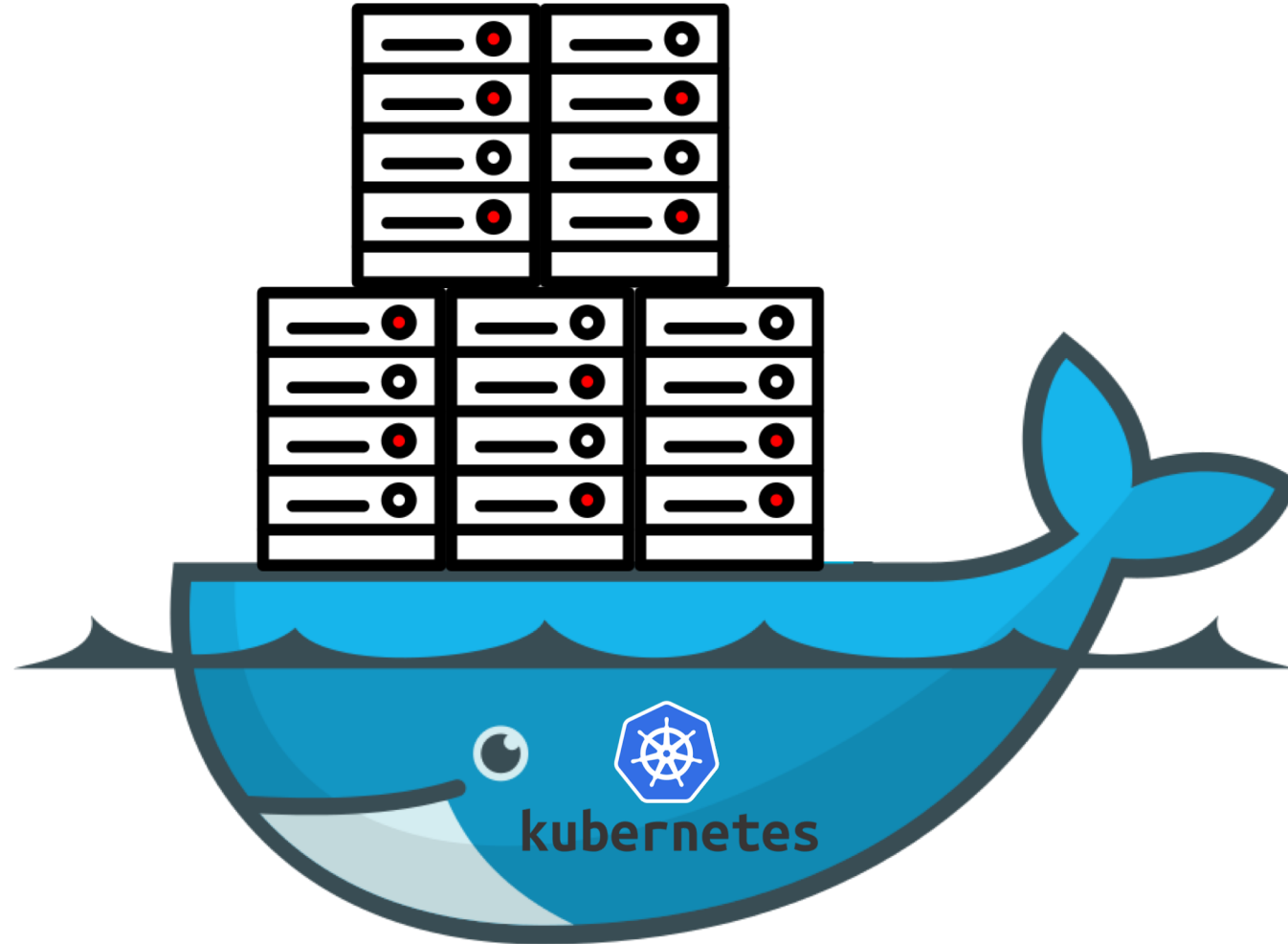
ceph

ODPI
EGERIA

Airflow



Containers as Compute



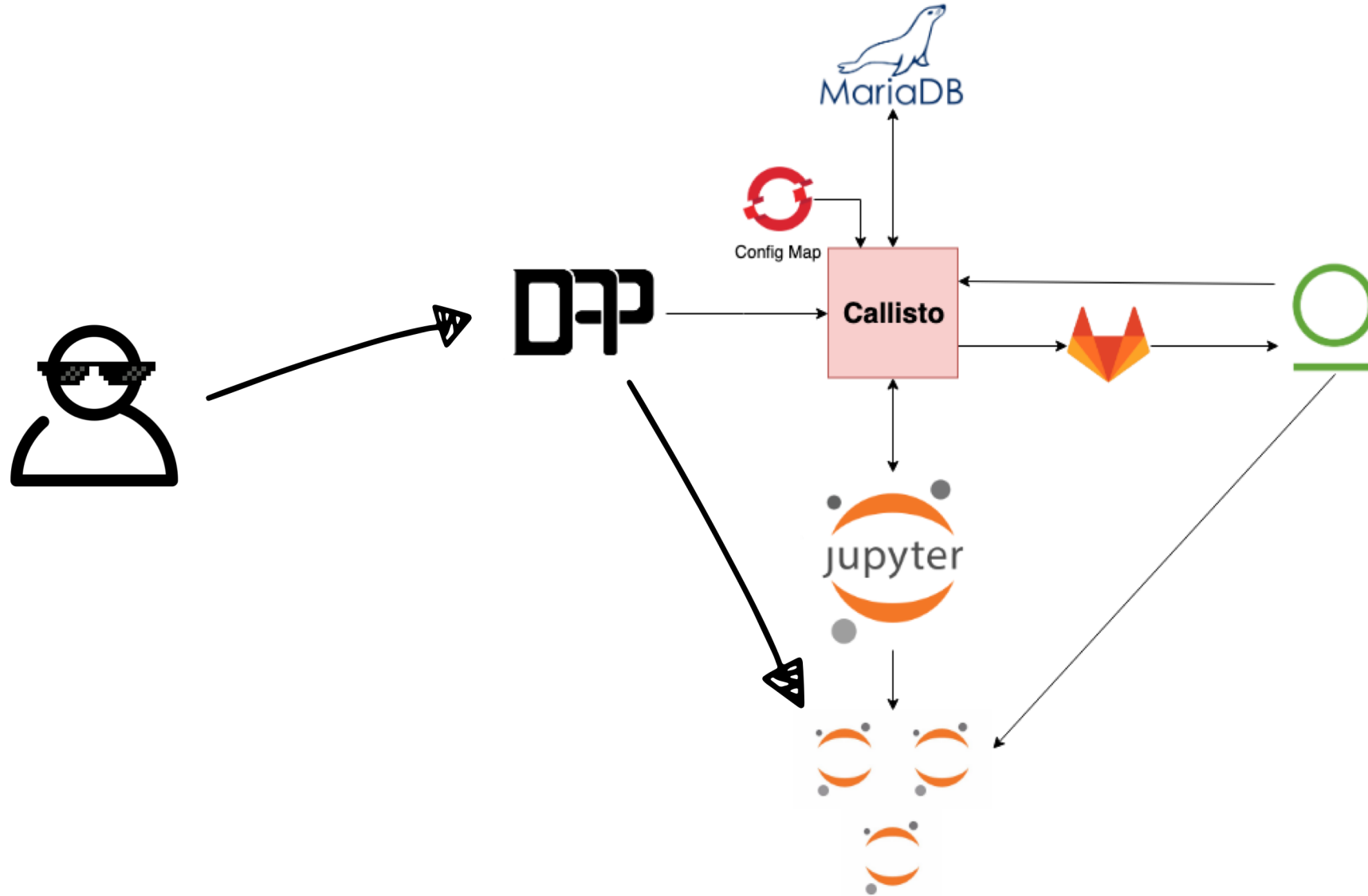


Data Science in a Box





Data Science in a Box





The Architecture

Front End

Portal



Security + Integration

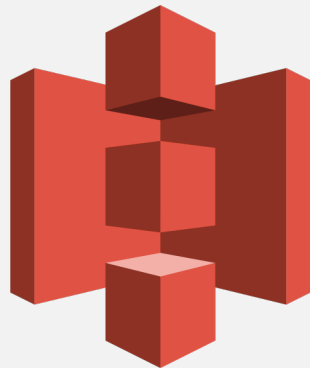
Apache **Atlas**

Compute



Apache Ranger

Storage





The Architecture



Front End

Portal



Security + Integration

Apache **Atlas**

Compute



Apache Ranger

Storage





Unified Storage



Rokku





The Architecture

Front End

Portal



Security + Integration

Apache **Atlas**

Compute



Apache Ranger

Storage





The Architecture



Front End

Portal



Compute



Storage



Security + Integration

Apache **Atlas**



Apache Ranger





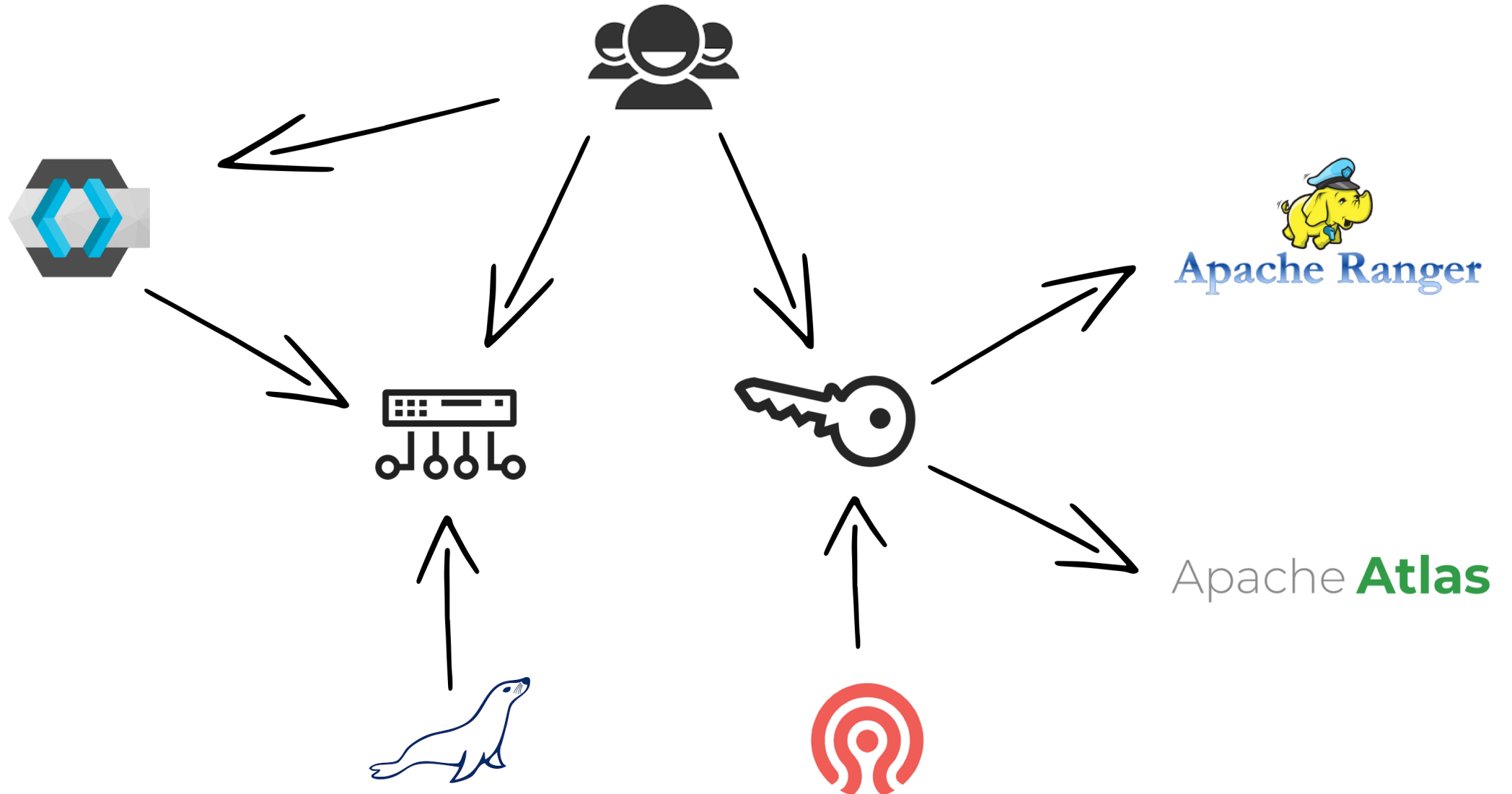
Security, Security, Security.





Authorization layer

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





Security Model



Assume Role: A

Assume Role: B

Kubernetes



Namespace A



Namespace B



Storage Proxy

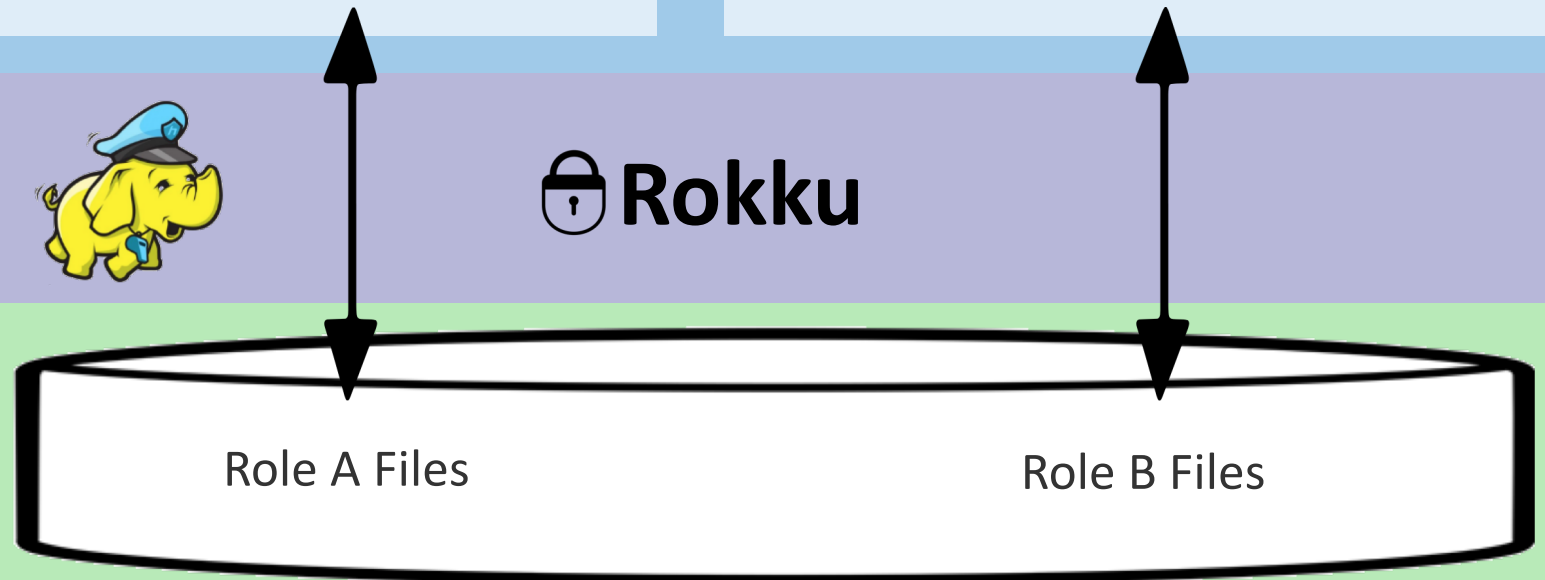


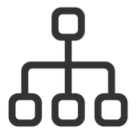
Rokku

Storage

Role A Files

Role B Files



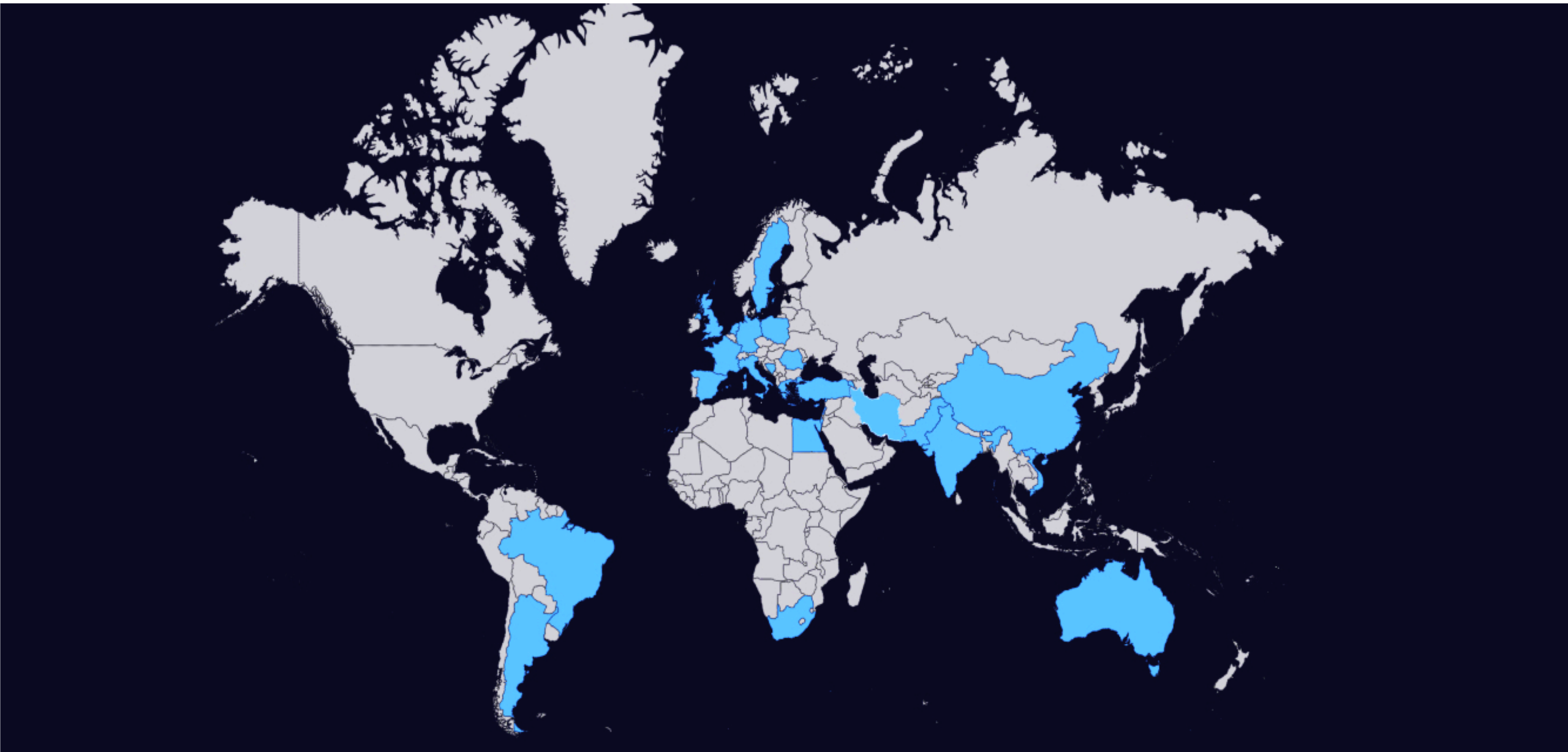


Data Lineage





The Team



**THANK
YOU!**