AIRFLOW SUMMIT

How Airflow Runs the Weather 3.0

Eloi Codina-Torras

About me

- Born near Barcelona
- Studied: Industrial Engineering
- Currently: Product Owner @ Meteosim
- 3rd Airflow Summit!

My family still doesn't understand what I do...

What we do

AIRFLOW SUMMIT

We provide **innovative environmental solutions**

20 +
years of experience

600 +
numerical simulations
run daily

80 +
countries

METEO SIM

What we do: some examples

Risk Management

Plan operations and activate emergency plans based on weather forecasts.

Regulatory compliance

Ensure compliance with local air quality regulations, by predicting, monitoring and analyzing data.

Climatological studies

Simulate past and future meteorological scenarios for various custom applications.

What we do: who uses our data

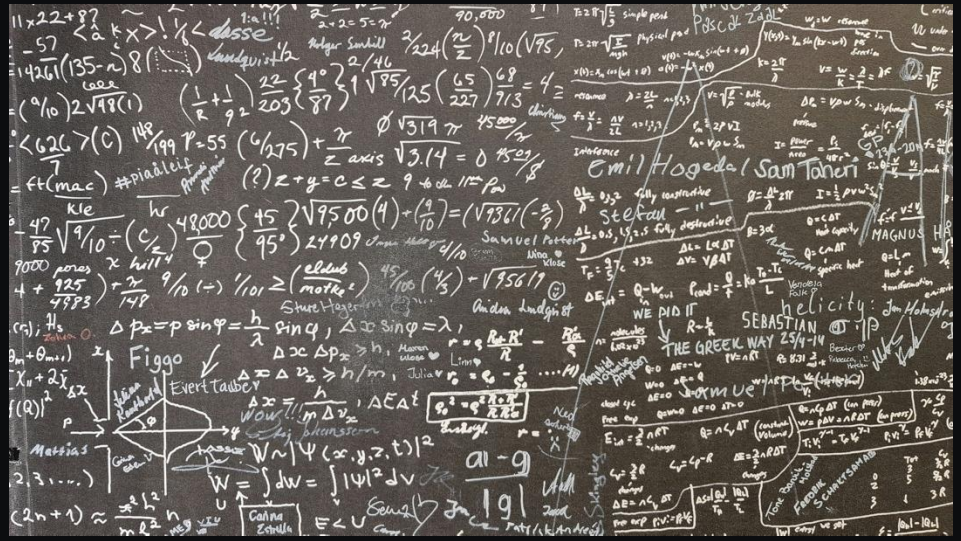
AIRFLOW SUMMIT

Insurance companies	Assess risk for extreme weather conditions
Oil & Gas companies	Manage the environmental impact of their operations and react to emergencies
Mining companies	Manage the environmental impact of their operations
Waste & water companies	Manage smell complaints
Public institutions	Assess climate change impact

METEO SIM

METEO SIM

AIRFLOW SUMMIT



The base of our services is solving complex differential equations using various numerical simulations.

Weather simulations: the basics

- We use WRF, an open-source NWP System developed by NCAR (USA)
- 3D high resolution (up to 333 m and 10 minutes) local weather forecasts

Global models:

- GFS (NOAA)
- IFS (ECMWF)
- ICON (Germany)
- ...

28 km resolution
Hourly data

Weather simulations: the basics

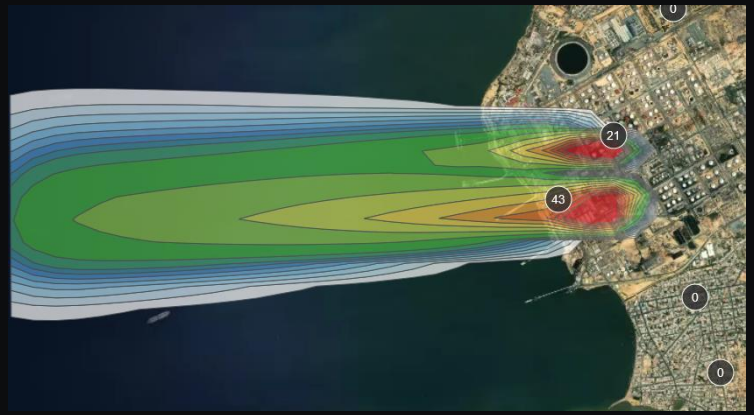
- We use WRF, an open-source NWP System developed by NCAR (USA)
- 3D high resolution (up to 333 m and 10 minutes) local weather forecasts

Local models: what we simulate

Based on global models.
Improves the geographic and temporal resolution
Needs real observation data to validate it.

Air quality simulations: the basics

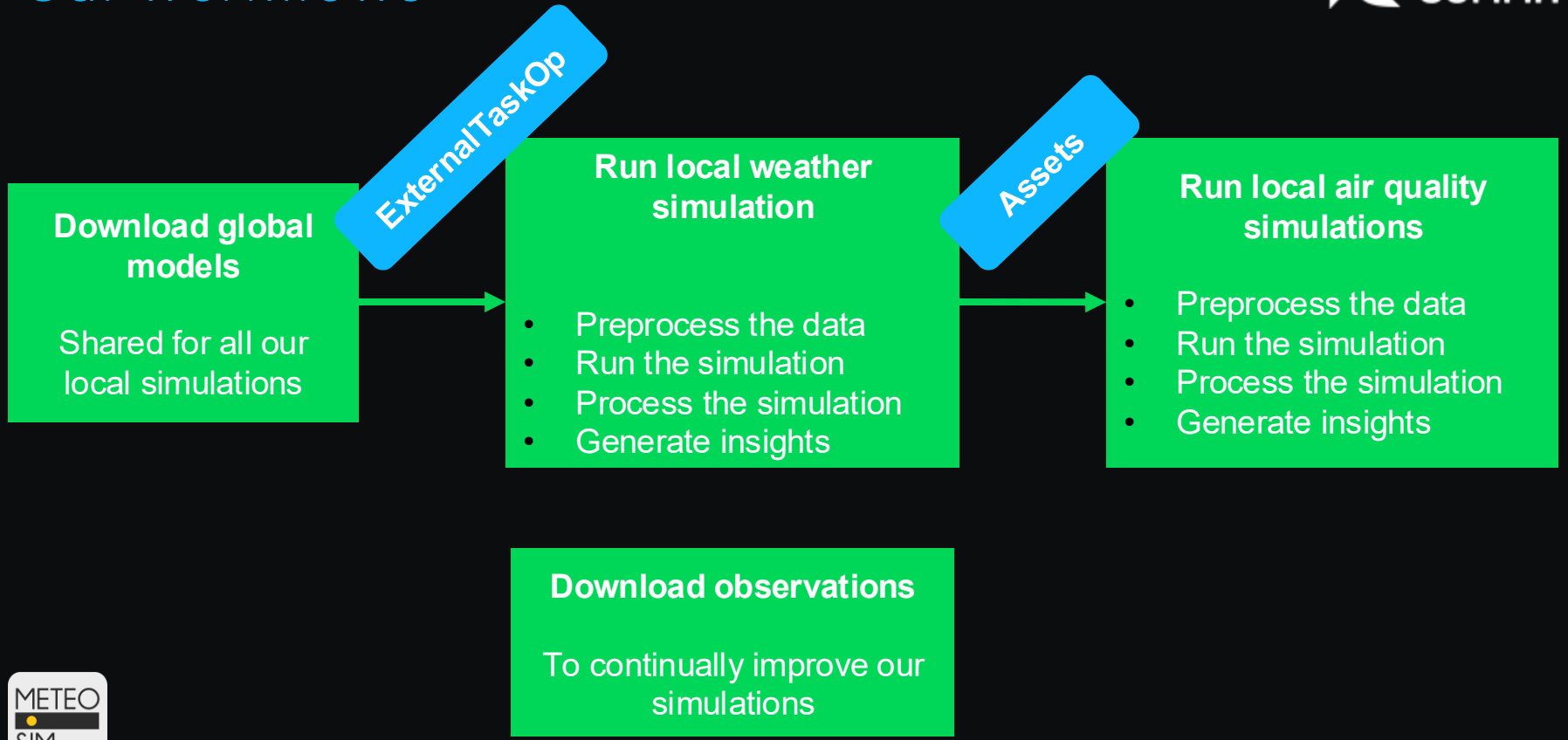
- Based on our weather forecasts and real / estimated emissions
- We generate 3D air quality simulations of up to 100 m resolution
- We use several models tailored to different requirements (chemical emergencies, city pollution, …)



We use Airflow to orchestrate our simulation models

We used crontab many years ago. It was not very scalable...

Our workflows



How we run Airflow

- Our simulations are power hungry
- An area such as Washington State would require:

Parameters

- 1 simulation per day
- 72 hours
- 1 km resolution

Cluster

- 256 CPUs
- 1 TB memory

Compute time

10 hours

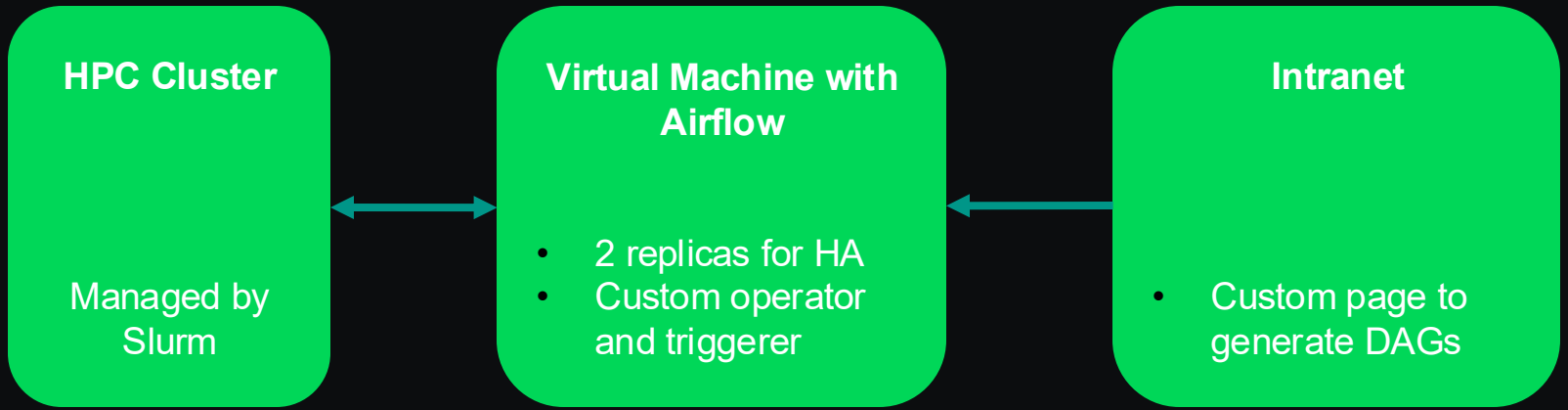
Cloud cost

$6000 /month

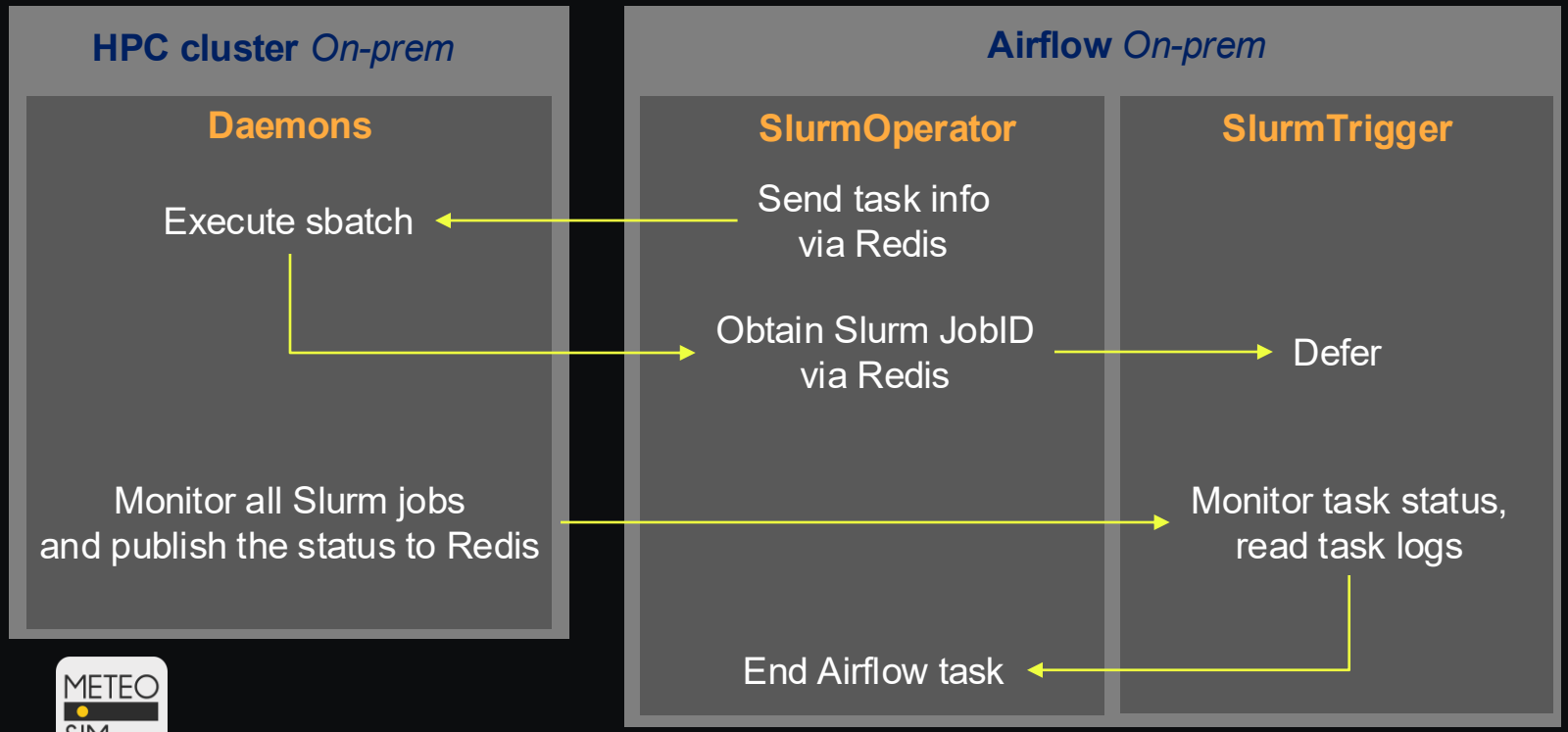
METEO SIM

How we run Airflow

- This is why we have our own HPC
- Managed by Slurm (manages resources and allocation in HPCs)



How we run Airflow



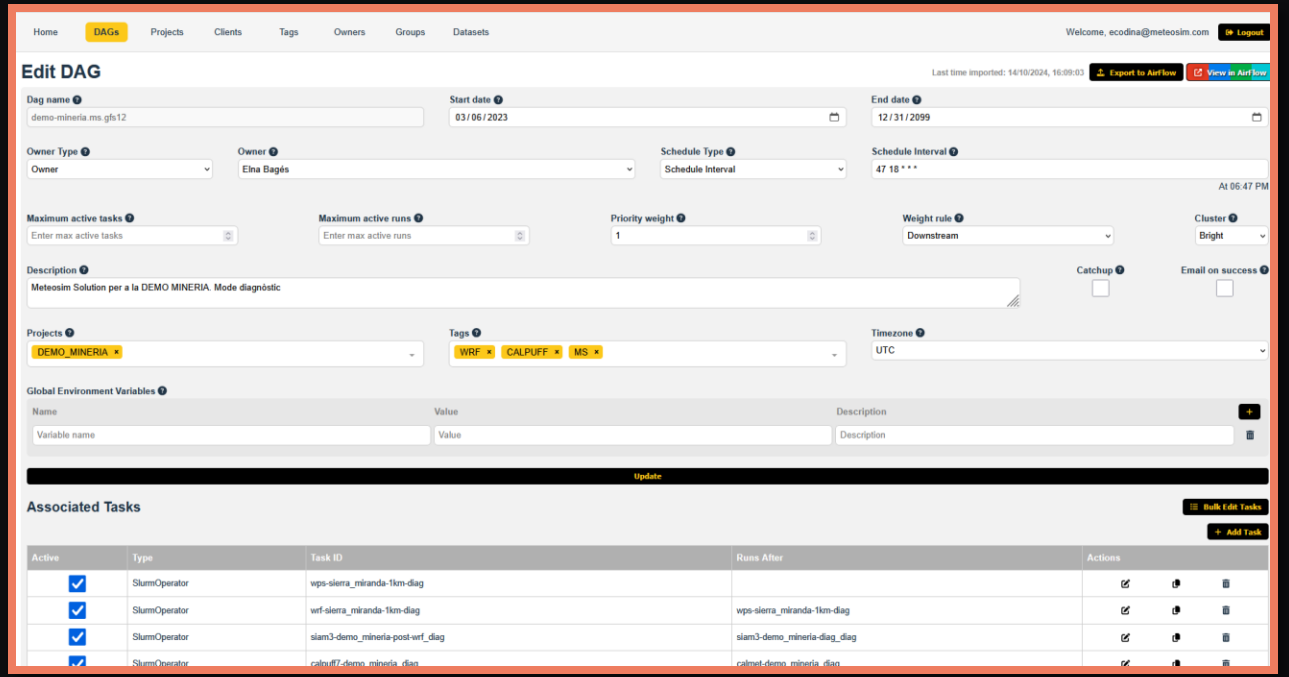
Last year's presentation about Airflow + Slurm

How we run Airflow

We use an Intranet to generate Dags in YAML

They are parsed using dag-factory

This way we can have templates



How Airflow enables our operations

AIRFLOW SUMMIT

Ease of use

- Intuitive web interface
- No need to remember Slurm commands
- Automatic retries for failed tasks
- Automatic notifications for failed tasks
- Support for YAML Dags

Visibility

- Run history
- Real-time logs on the web interface
- Charts

METEO SIM

How Airflow enables our operations

- We use Airflow **a lot**
- Other use cases: simulation validation, reporting, notifications, clean-up, ...
- Our instance:

+600	2.2M	96%
active Dags	*annual Dag runs*	*success rate*

METEO SIM

What's next for us

Upgrade	EdgeExecutor	Implement assets
Upgrade to **Airflow 3.1** (currently in 2.10) • New UI + i18n • Online backfills • Lower resources • Dag versions	Test the new **EdgeExecutor**. Try to remove our custom SlurmOperator and SlurmTriggerer	Use assets **everywhere**. *We have to investigate how to have visibility on unscheduled Dags.*

METEO
SIM

METEO SIM

AIRFLOW SUMMIT

Questions?

eloi-codina