



How Airflow Runs the Weather

Eloi Codina-Torras

3.0

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



We provide innovative environmental solutions

20 +

years of experience

600 +

numerical simulations
run daily

80 +

countries



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



Insurance companies

Assess risk for extreme weather conditions

Oil & Gas companies

Manage the environmental impact of their operations and react to emergencies

Mining companies

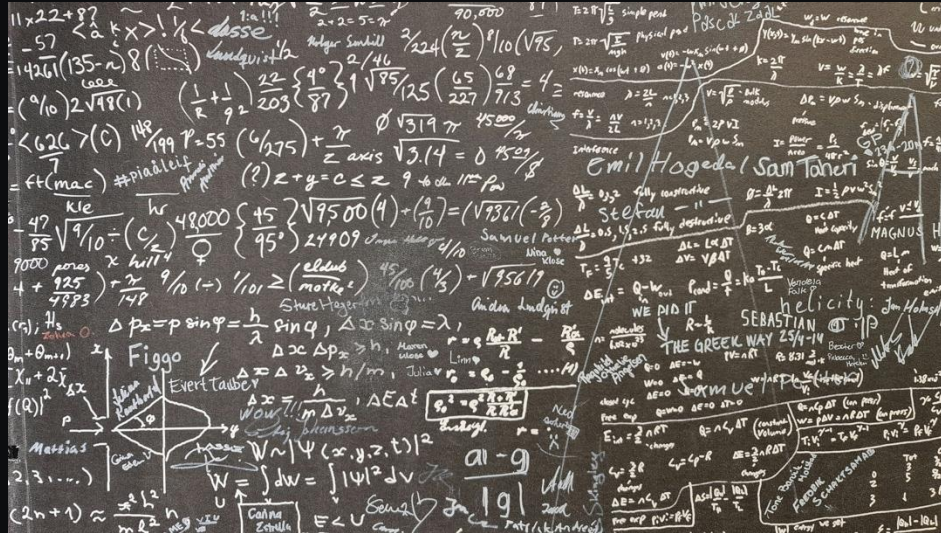
Manage the environmental impact of their operations

Waste & water companies

Manage smell complaints

Public institutions

Assess climate change impact



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

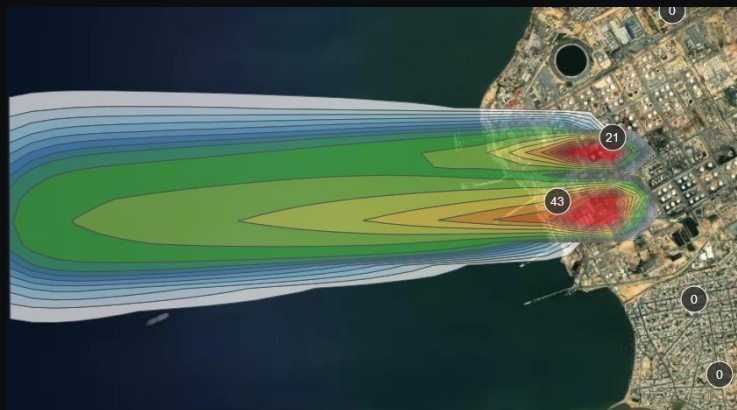


Local models: what we simulate

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, ...)

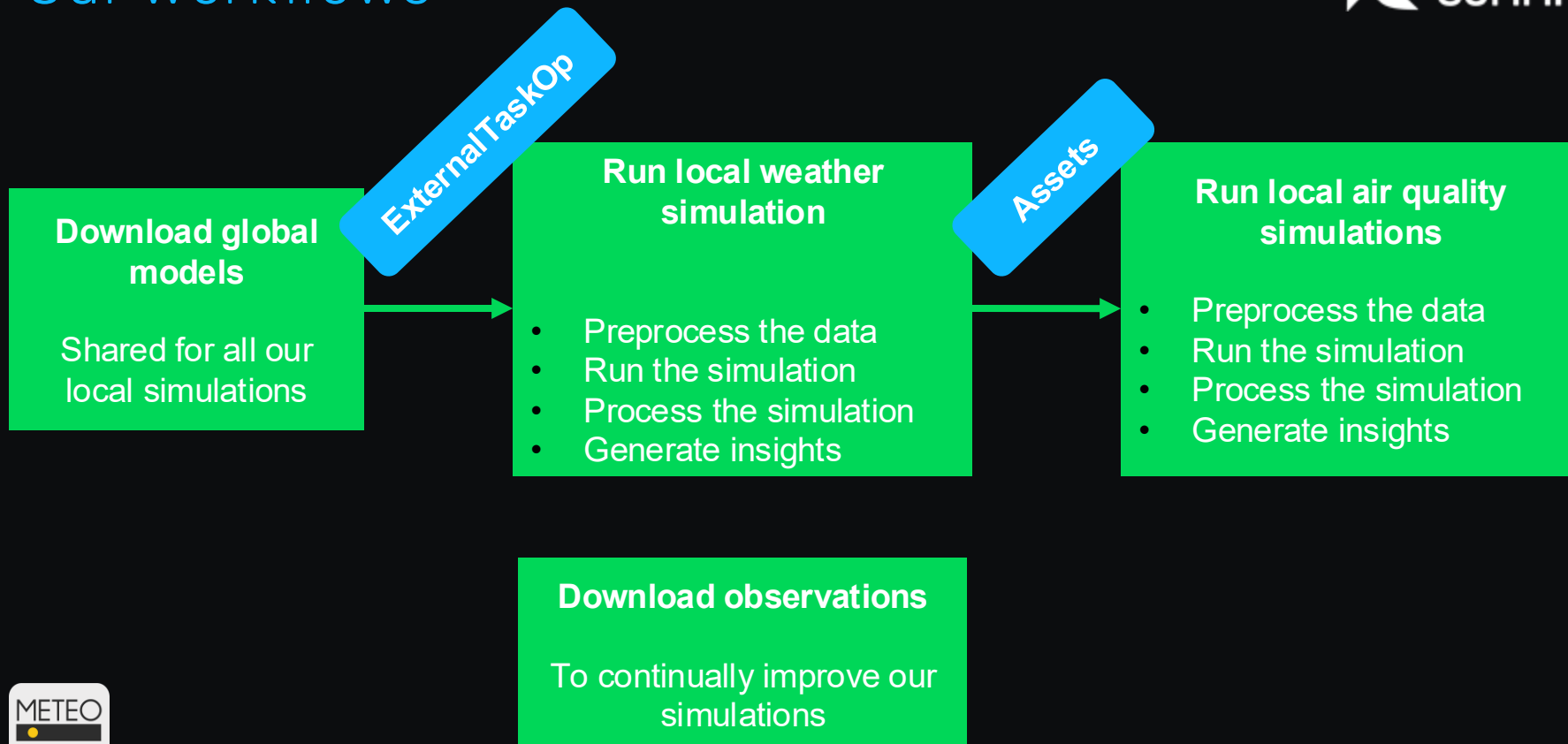




We use Airflow to orchestrate our simulation models

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

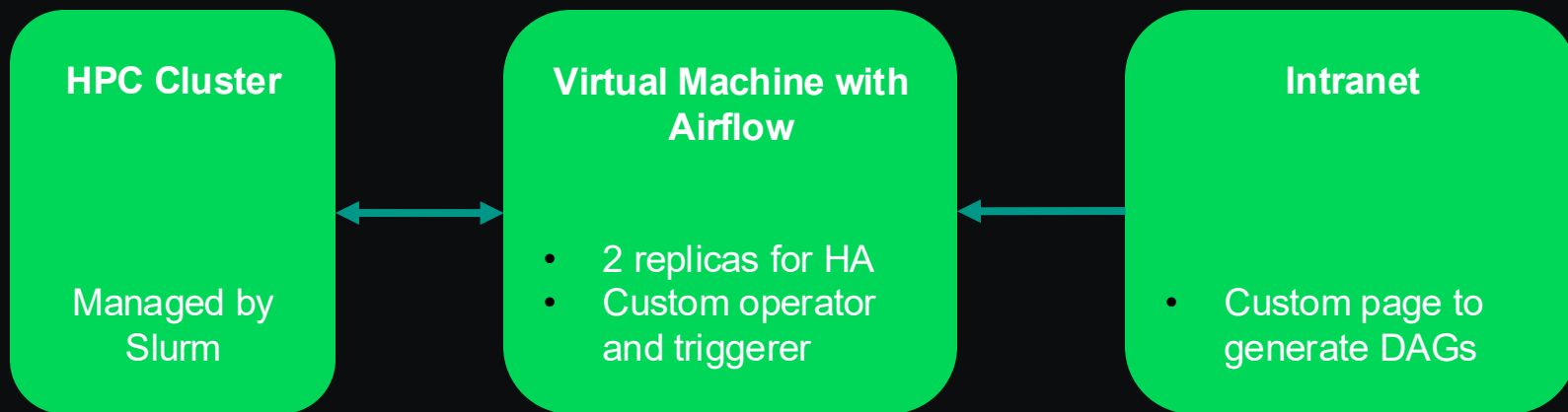
Cloud cost

\$6000 /month

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

HPC cluster On-prem

Daemons

Execute sbatch

Monitor all Slurm jobs
and publish the status to Redis

Airflow On-prem

SlurmOperator

Send task info
via Redis

Obtain Slurm JobID
via Redis

End Airflow task

SlurmTrigger

Defer

Monitor task status,
read task logs



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

The screenshot shows the 'Edit DAG' interface in the Airflow web console. The form includes fields for DAG name, start and end dates, owner, schedule type, and interval. It also has sections for maximum active tasks and runs, priority weight, weight rule, cluster, and description. A table at the bottom lists associated tasks.

Edit DAG

DAG name: demo-mineria.ms.gfs12 | Start date: 03/06/2023 | End date: 12/31/2099

Owner Type: Owner | Owner: Elia Bagés | Schedule Type: Schedule Interval | Schedule Interval: 47 18 ***

Maximum active tasks: Enter max active tasks | Maximum active runs: Enter max active runs | Priority weight: 1 | Weight rule: Downstream | Cluster: Bright

Description: Meteosim Solution per a la DEMO MINERIA. Mode diagnostic

Projects: DEMO_MINERIA | Tags: WRF, CALPUFF, MS | Timezone: UTC

Global Environment Variables

Name	Value	Description
Variable name	Value	Description

Associated Tasks

Active	Type	Task ID	Runs After	Actions
<input checked="" type="checkbox"/>	SlurmOperator	wps-sierra_miranda-1km-dag		🔗 🔍 🗑️
<input checked="" type="checkbox"/>	SlurmOperator	wrf-sierra_miranda-1km-dag	wps-sierra_miranda-1km-dag	🔗 🔍 🗑️
<input checked="" type="checkbox"/>	SlurmOperator	siam3-demo_mineria-post-wrf-dag	siam3-demo_mineria-dag-dag	🔗 🔍 🗑️
<input checked="" type="checkbox"/>	SlurmOperator	calmet-demo_mineria-dag	calmet-demo_mineria-dag	🔗 🔍 🗑️

How Airflow enables our operations



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



How Airflow enables our operations

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

+600

active Dags

2.2M

annual Dag runs

96%

success rate

What's next for us



Upgrade

Upgrade to **Airflow 3.1**
(currently in 2.10)

- New UI + i18n
- Online backfills
- Lower resources
- Dag versions

EdgeExecutor

Test the new
EdgeExecutor.

Try to remove our
custom SlurmOperator
and SlurmTriggerer

Implement assets

Use assets
everywhere.

*We have to investigate
how to have visibility
on unscheduled Dags.*



Questions?

