Building cloud-based data services to enable earth-science workflows across HPC centres

ECMWF

Milana Vuckovic, John Hanley, Stephan Siemen, James Hawkes, Tiago Quintino and Florian Pappenberger

Forecast Department, ECMWF, Reading, United Kingdom

About HiDALGO

HPC and Big Data Technologies for Global Systems – European project funded by the Horizon 2020 Framework Programme of the European Union carried out by 13 institutions from seven countries.

HiDALGO

<u>The Vision:</u> To advance technology to master global challenges
<u>The Mission:</u> To develop novel methods, algorithms and software for HPC & HPDA to accurately model and simulate the complex processes which arise in connection with major global challenges.

Pilot Test Cases using ECMWF data

Air Pollution pilot - Széchenyi István University, Hungary

Simulation of Air Pollution Dispersion with Computational Fluid Dynamics (CFD) based using:
- ECMWF weather forecast as one of the inputs and
- CAMS Air pollution data for boundary conditions
When processed, the data that is fed into the model is a fraction of what is transferred.
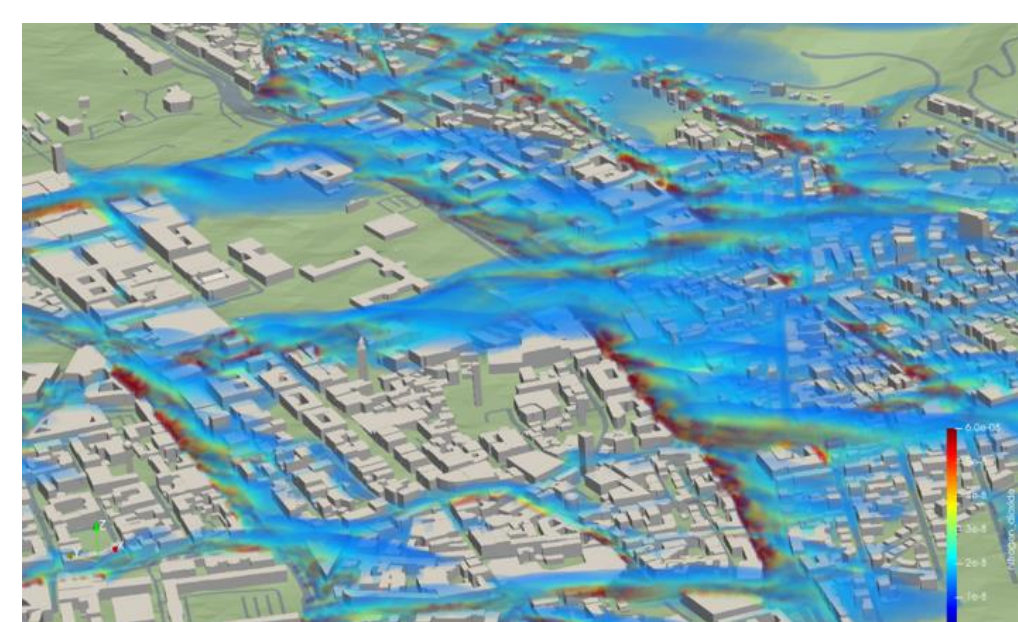
Migration pilot - Brunel University, UK

- Agent-based models (ABMs) based simulation framework performing Refugee Movement Simulations
- The effects of weather on movement of refugees are being investigated.
- Using Machine Learning to determine the available routes between locations with GloFAS dataset to incorporate the effects of river Nile on these routes in South Sudan.
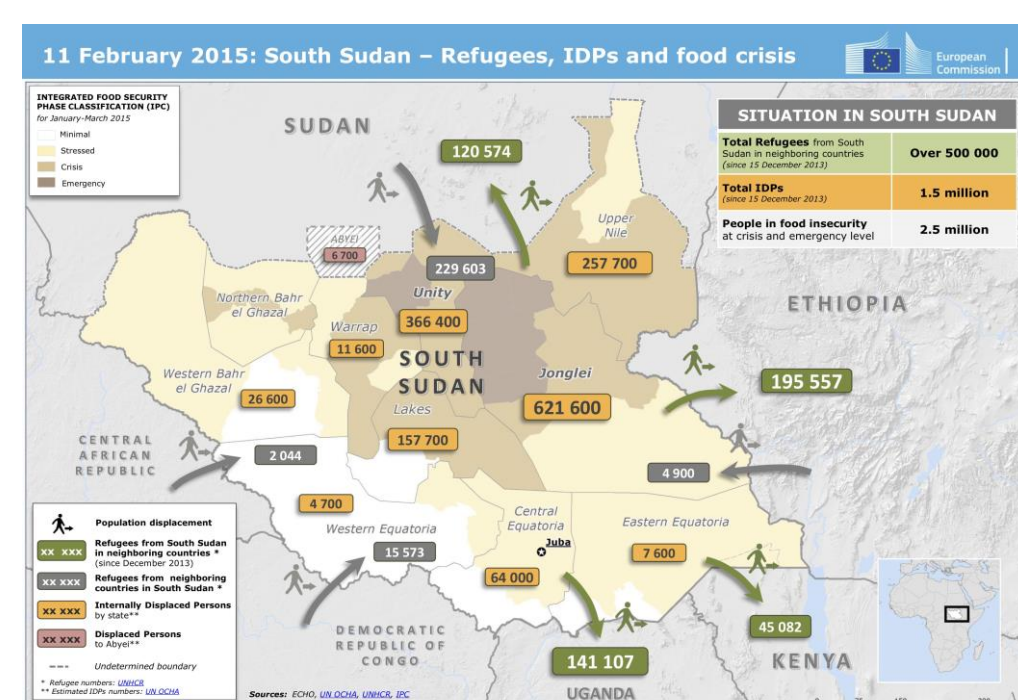
Covid-19 simulation - Brunel University, UK

Codebase from Migration model is modified to construct simulations of COVID19 on London borough level to study how properties of the epidemic change when measures are lifted.
When the base simulation is completed, the effect of weather conditions and air pollution will be investigated.
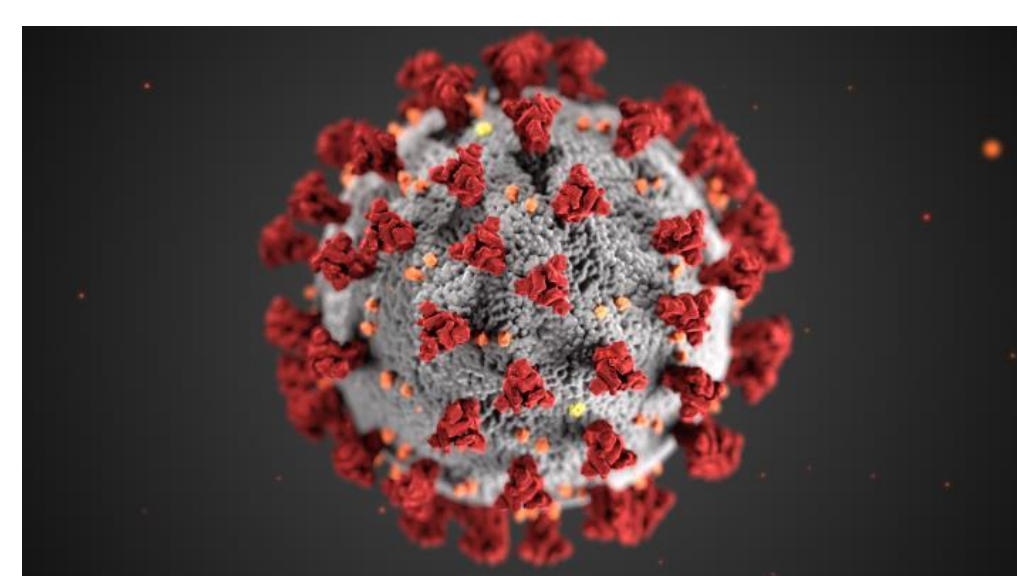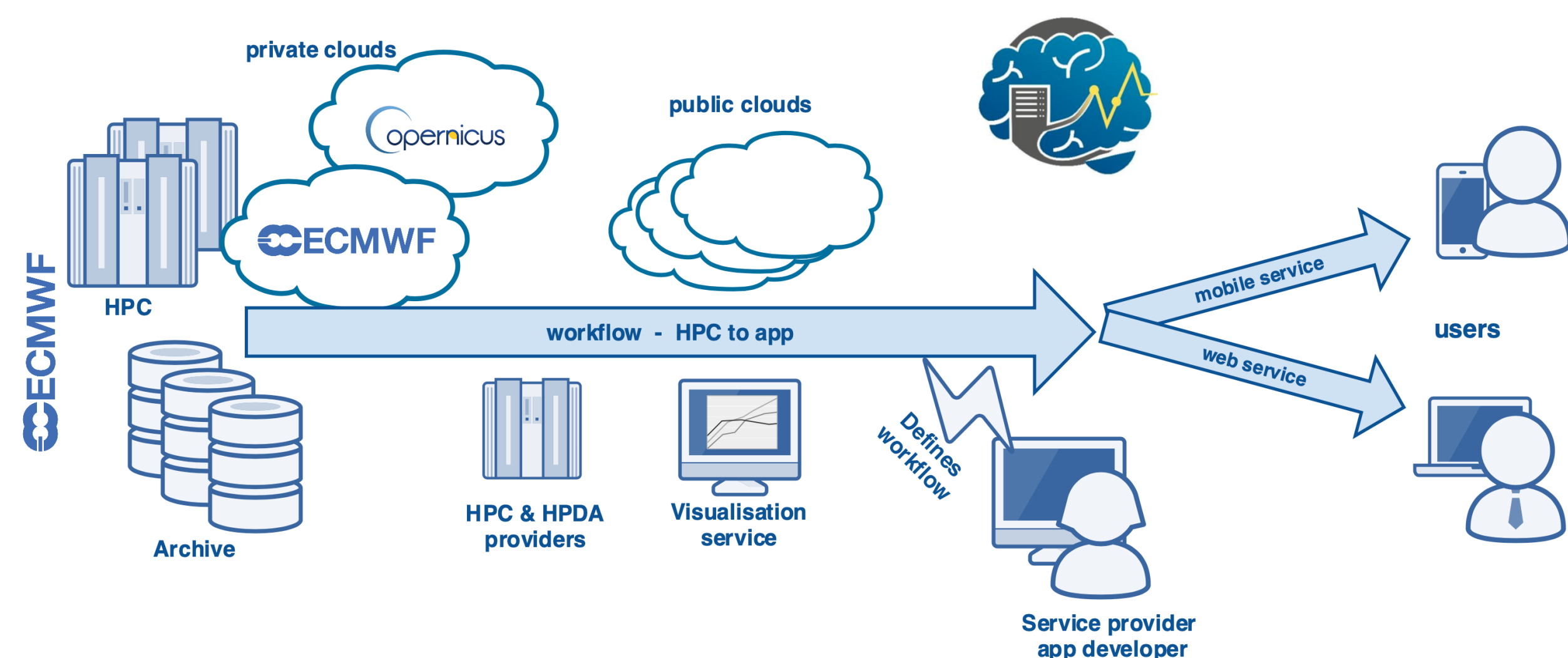
Photo credit: CDC/ Alissa Eckert, MS; Dan Higgins, MAMS

ECMWF's role in HiDALGO

- Weather forecasts produced by ECMWF and environment services by the Copernicus programme act as a vital input for many downstream simulations and applications.

- Transferring, storing and locally modifying large volumes of such data prior to integration currently presents a significant challenge to users.

- The key aim for ECMWF within the H2020 HiDALGO project is to migrate these tasks to the cloud, to enable fast and seamless application integration.

- The required cloud infrastructure development is also being feed into ECMWF's contribution to the European Weather Cloud pilot, a collaborative project between ECMWF and EUMETSAT.

- The two HiDALGO pilot applications (Migration and Air Pollution pilot) will be among first users who will benefit of receiving already pre-processed data

The Data Challenge

Historical Growth of Generated Products
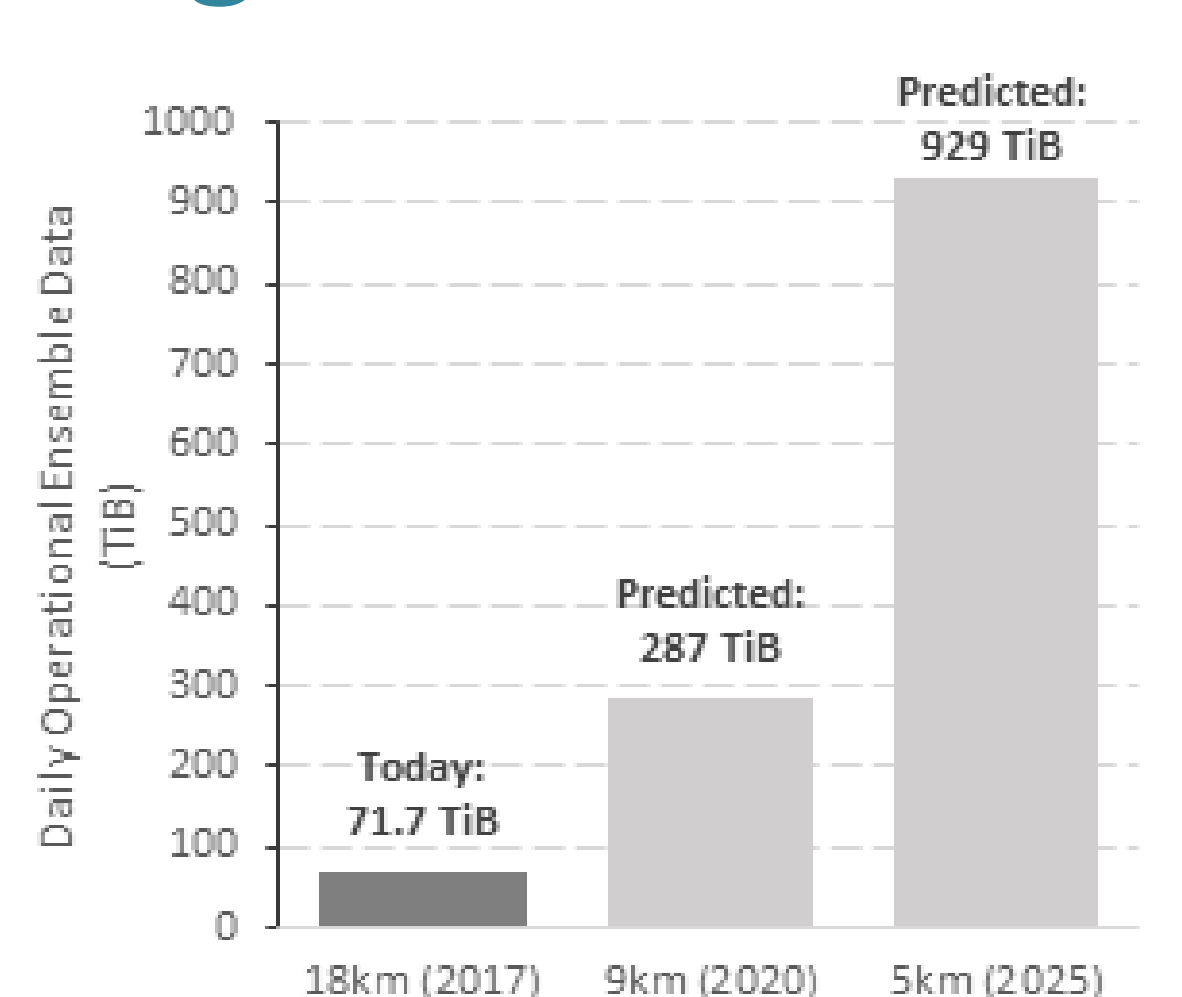
Model Output Projected Growth

- The real-time data produced will increase by a factor of 12 by 2025
- Ensemble (ENS) forecasts, widely used in decision support systems, are the biggest contributors.
- Assuming current demand and delivery mechanism, the data volume distributed will increase by the same factor. Is this sustainable?

The Key Challenge:

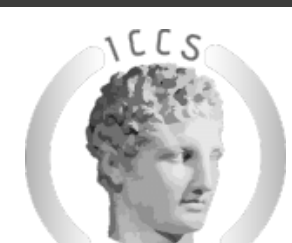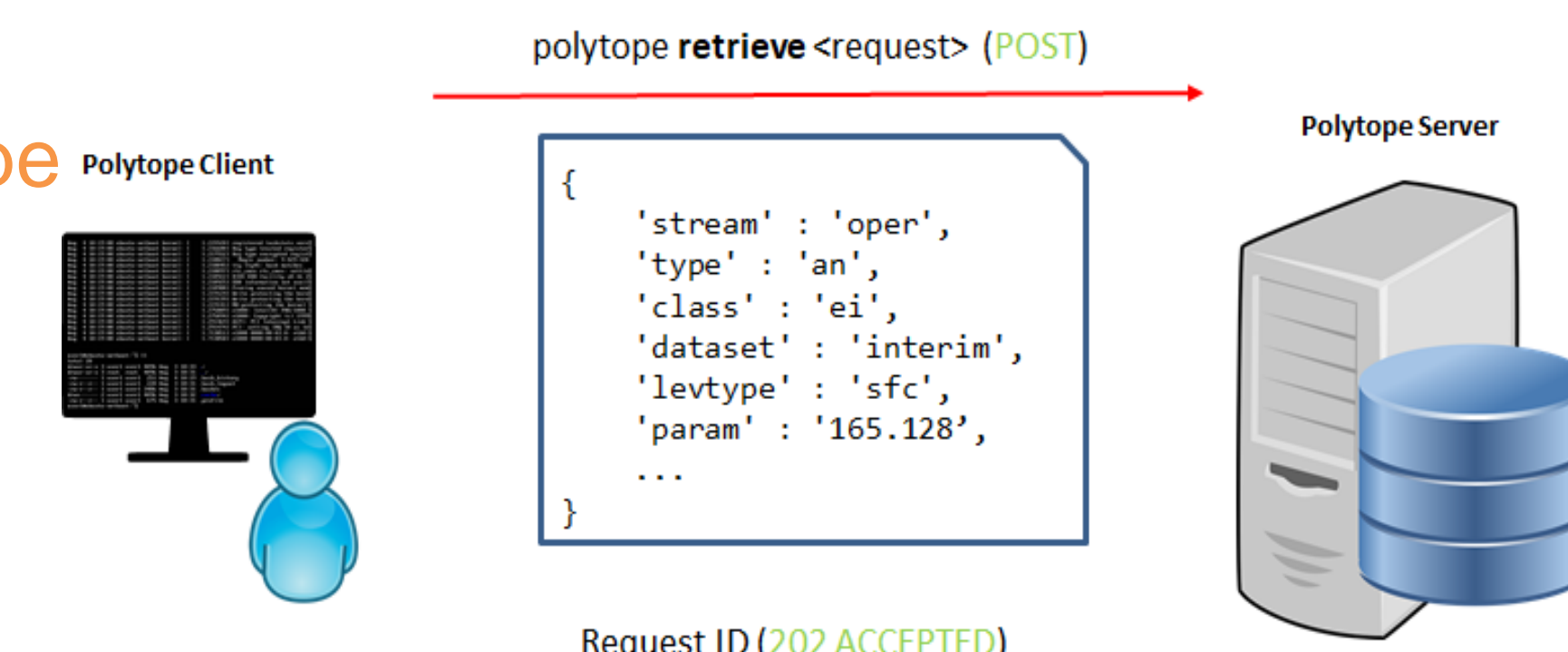**How do we improve user access to such volumes of data?**

Addressing the Key Challenge:

1. Bring users to the data and avoid moving the data out of the data centre.

2. Provide users with computing resources collocated directly with data.

3. Align with data-centric approach of "move the compute, not the data".

ECMWF + EUMETSAT
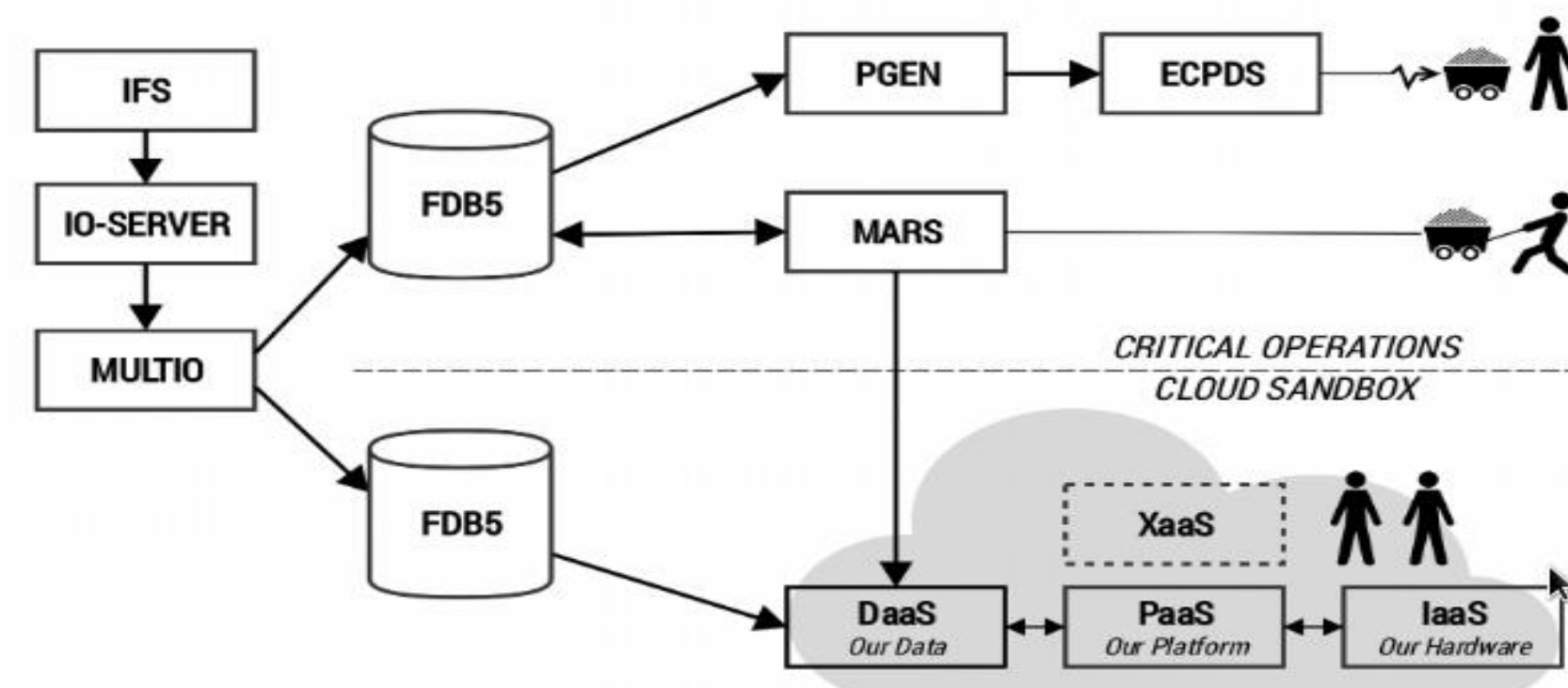
Cloud Data-as-a-Service: Polytope

Service designed for efficient provisioning of meteorological data to Cloud and HPC applications.