SWE for large scale science

Amy Krause

EPCC, The University of Edinburgh

VERCE project

The VERCE project provides a framework to the seismological community to exploit the increasingly large volume of seismological data

- Support data-intensive and HPC applications
- e-Science Gateway for submitting applications
- Distributed and diversified data sources
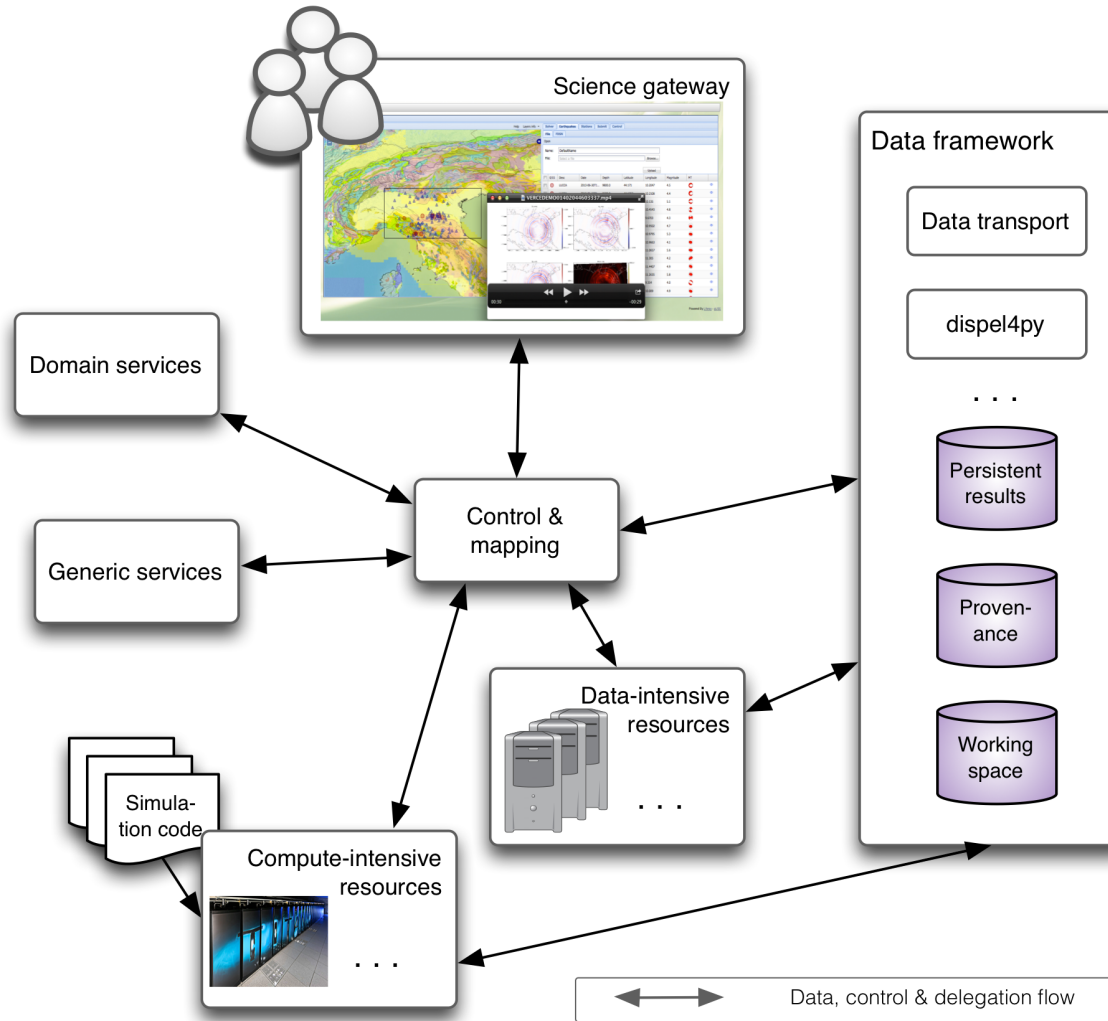- Distributed HPC resources on Grid, Cloud and HPC clusters

VERCE architecture

Science gateway

Domain services

Generic services

Control & mapping

Simula-tion code

Compute-intensive resources

. . .

Data-intensive resources

. . .

Data framework

Data transport

dispel4py

. . .

Persistent results

Proven-ance

Working space

Data, control & delegation flow

Experiences and challenges

- Large distributed team of software engineers
- Scientists (seismologists) as users
- A lot of expertise with various technologies
- Good support for multi-site software development
- Regular technical meetings
  - F2F meetings
  - Skype calls
- One-to-one interactions
  - Everyone was approachable

- Project needed a technical manager
  - Lack of direction
- Lack of communication between technical staff and researchers/project lead
- Seismologists had experience with using large parallel applications
- Less interest in streaming applications
- Struggling to find a large enough dataset