# Cloud Computing in Ground Segments

# Earth Observation processing campaigns

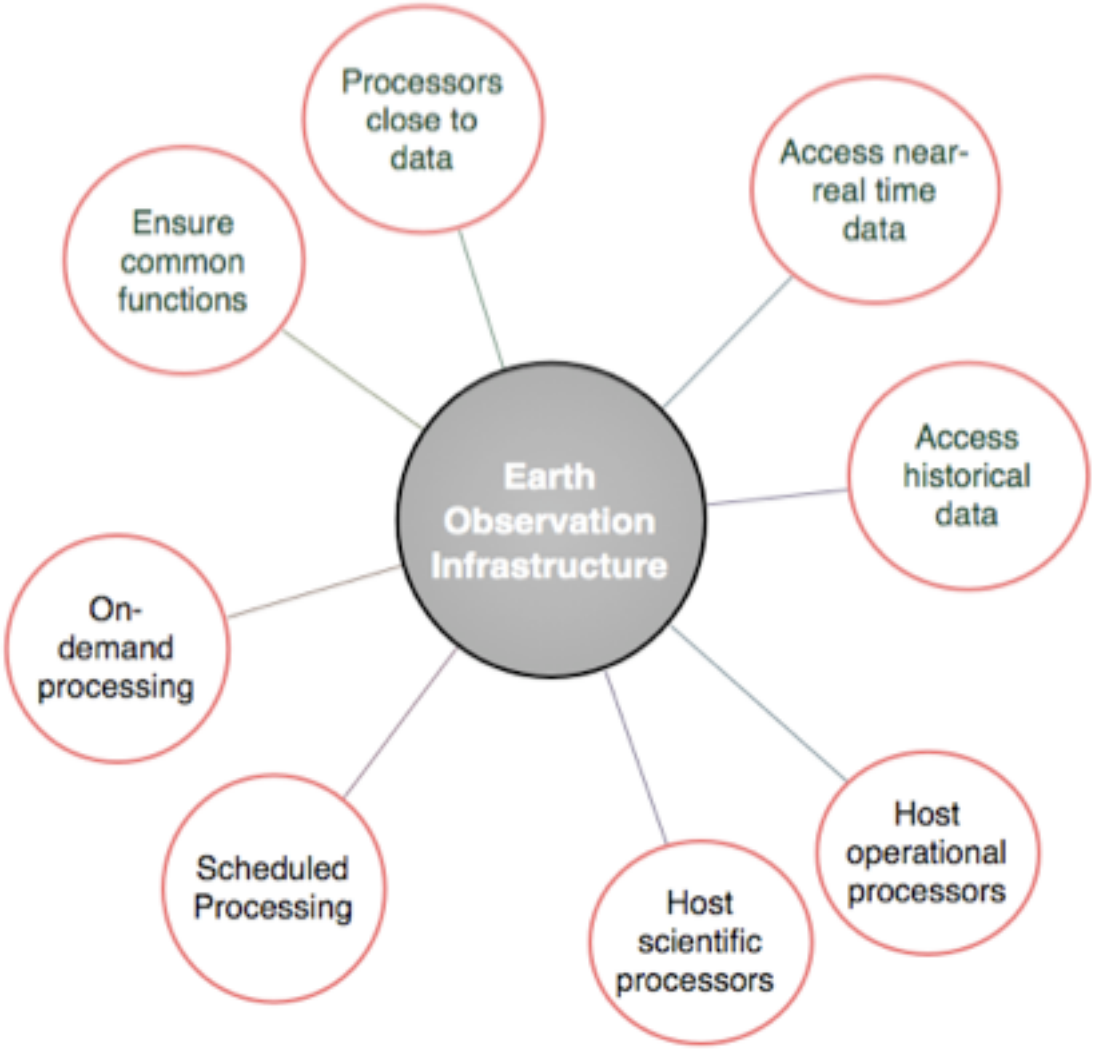
Fabrice Brito – Terradue Srl

GSAW – 2010 Geospatial Clouds

terradue 2.0

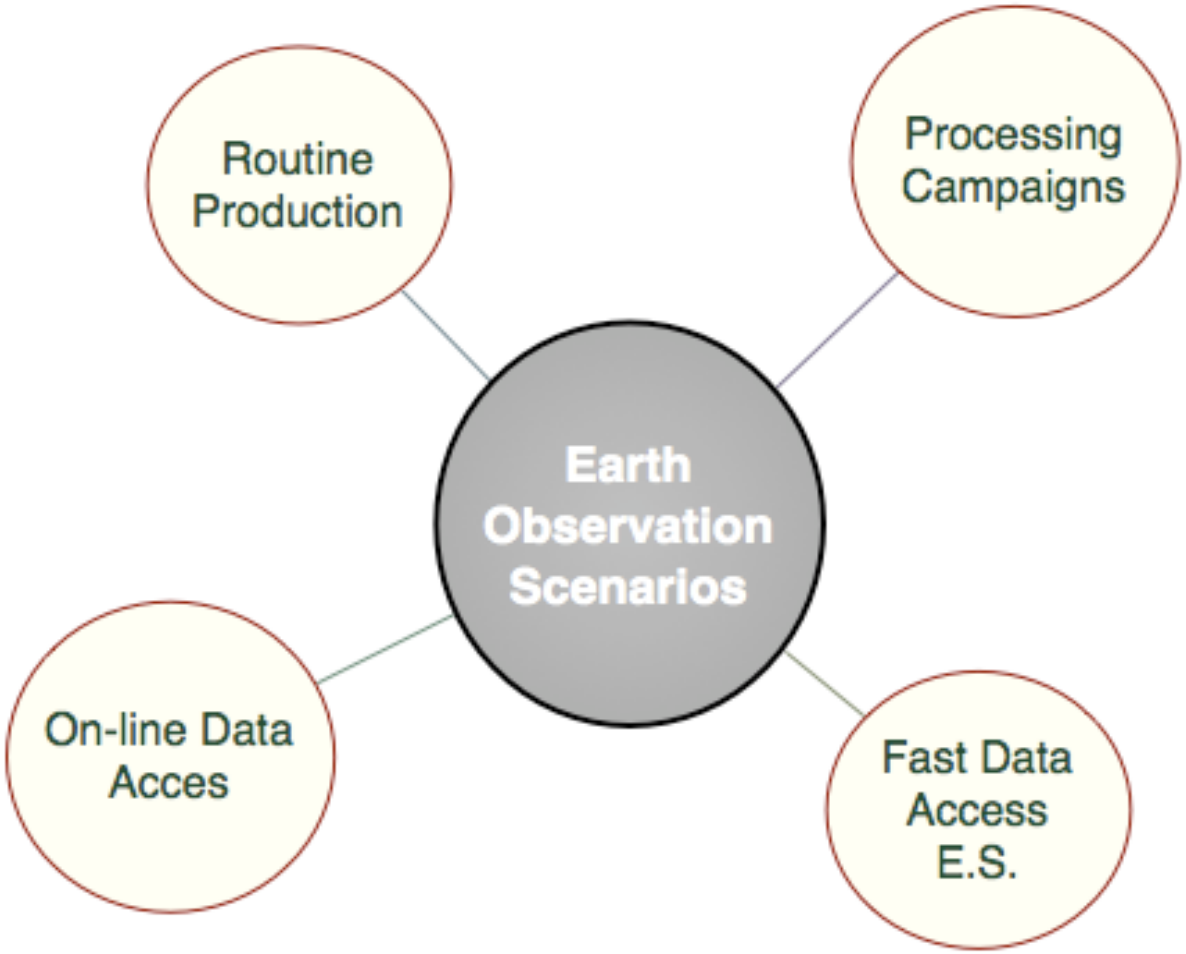
# Earth Observation IT

## EO computing infrastructures have particular requirements



# Earth Observation Scenarios

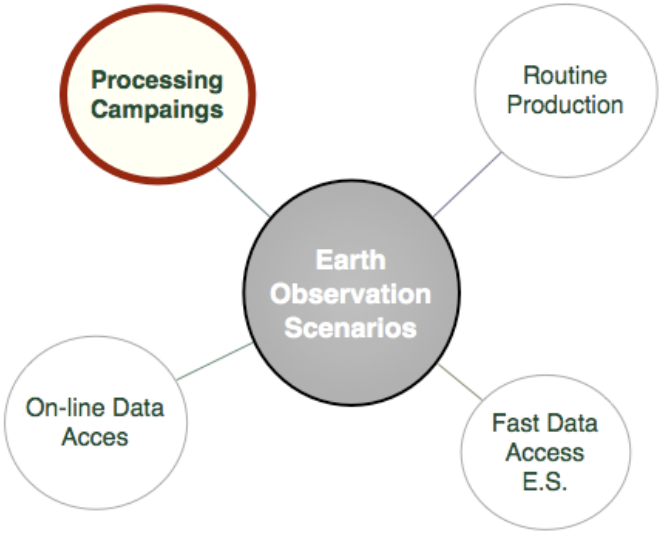
EO computing infrastructures deal with several scenarios



# Processing Campaigns

Earth Observation mission re-processing to improve data quality

- Algorithms evolve
- Instruments' calibration change



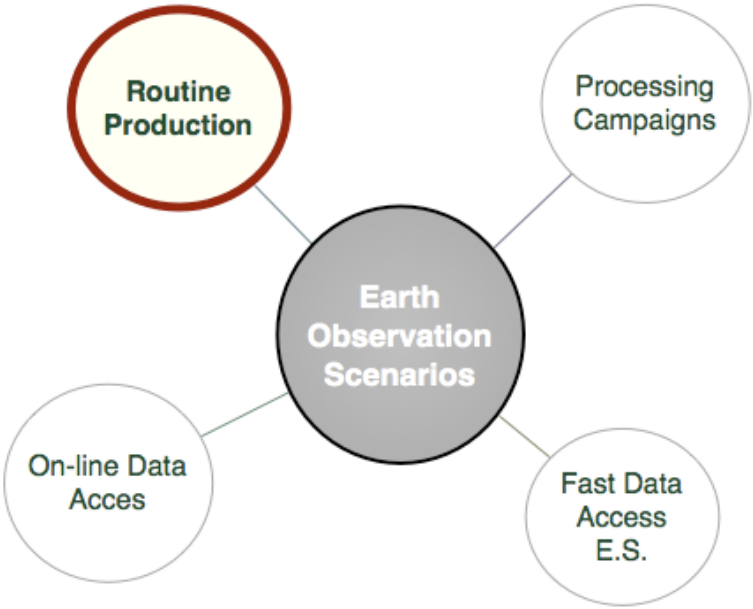

Needs:

- Storage for the input products
- Storage for the re-processed products
- Possible on-line data access for the reprocessed products
- CPU

# Routine Production

Earth Observation routine production:

- generation, archiving and distribution of higher-level products following data or date driven scheduled services




Needs:

- Storage for the input products
- Storage for the higher level products
- Possible on-line data access for the generated products
- CPU

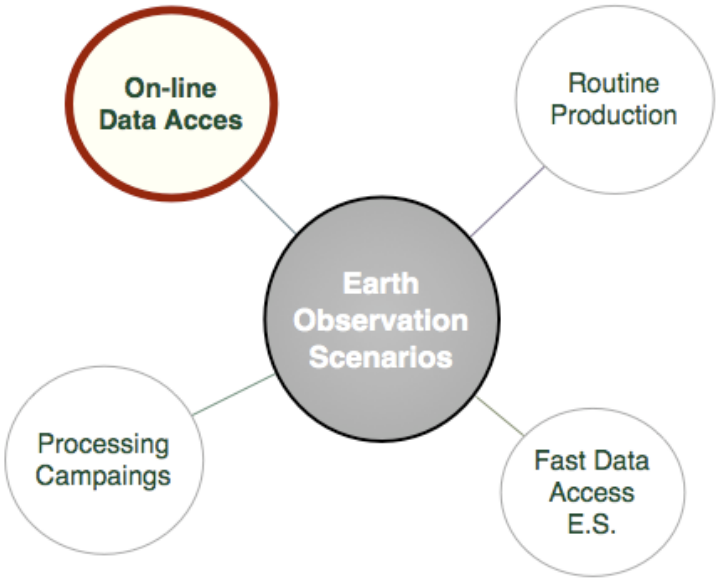
# On-line data access

Earth Observation on-line data access:

- Media are changing, users no longer want DVDs or tapes
- Rolling archives with NRT data
- Historical archives

Needs:

- Storage for the products
- Fast network, simple protocols
- Replicas around the world
- Data policy, access rights enforcement

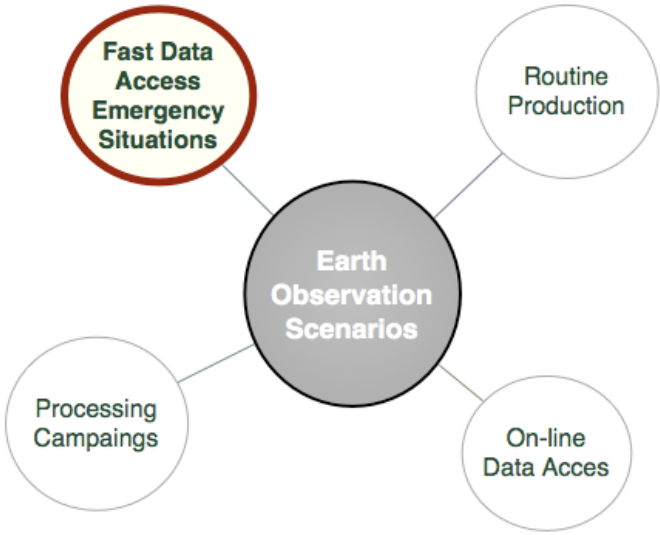


# Fast data access for Emergencies

Earth Observation is an excellent source of information for damage assessment for both natural and man-made disasters

Needs:

- Storage for the input products
- Storage for the crisis generated products
- On-line data access
- CPU



# What can Cloud do?

terradue 2.0

Cloud Computing stored EO data separates on-line data access from archiving and preservation of the data

EO data can be stored in Cloud Computing storage

Cloud Computing stored EO data can be processed over and over

Security and reliability are issues addressed by the service provider

Cloud Computing stored EO data can be accessed and distributed

# How do we implement this?

- G-POD is an ESA-initiated project in 2002
- Industrialized by Terradue Srl in 2006
- G-POD provides a “user-segment” putting EO data and processors closer together

http://gpod.eo.esa.int/

# G-POD objectives

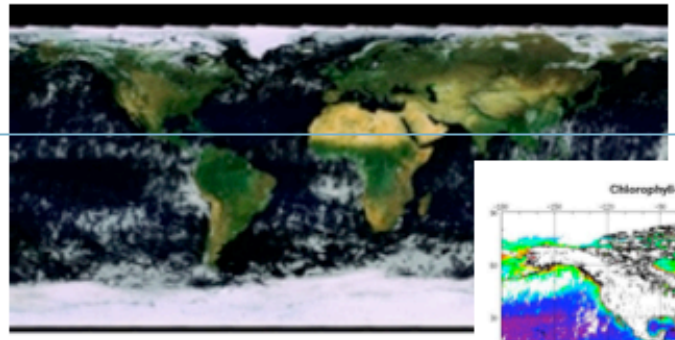
- Access and use of EO mission data available at ESA
- Provide computing infrastructure and tools to assist the generation of “scientific added value products”
- Supporting the development of earth science applications requiring significant data and processing resources
- Fostering of new partnerships between ESA and entities in earth science

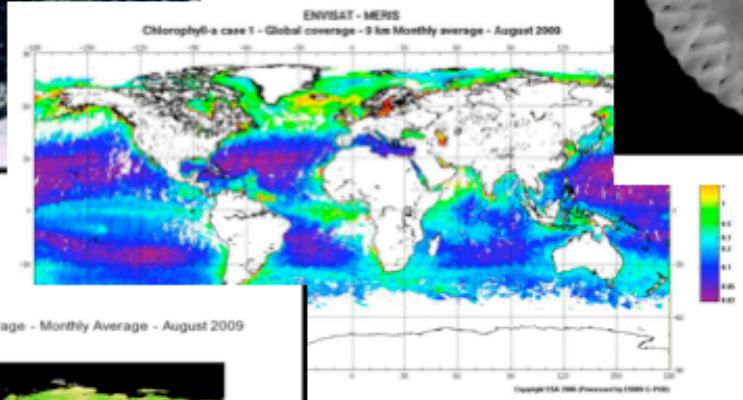
http://gpod.eo.esa.int/

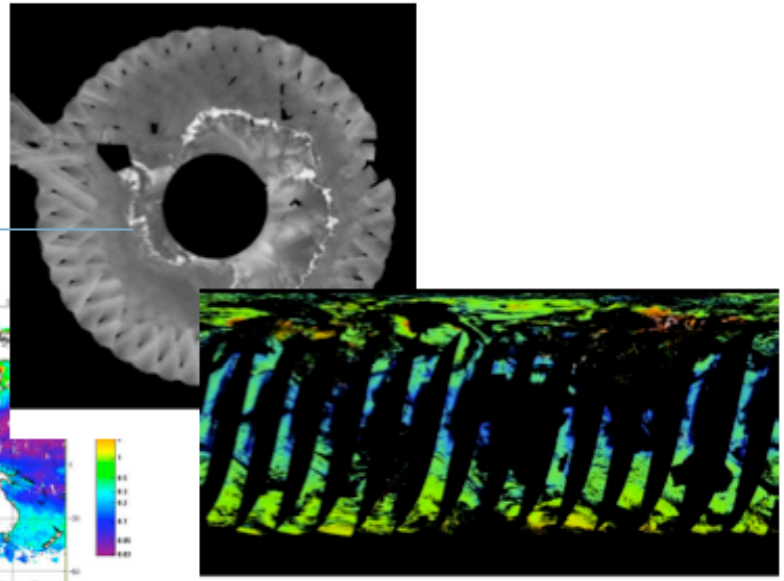
# G-POD Products

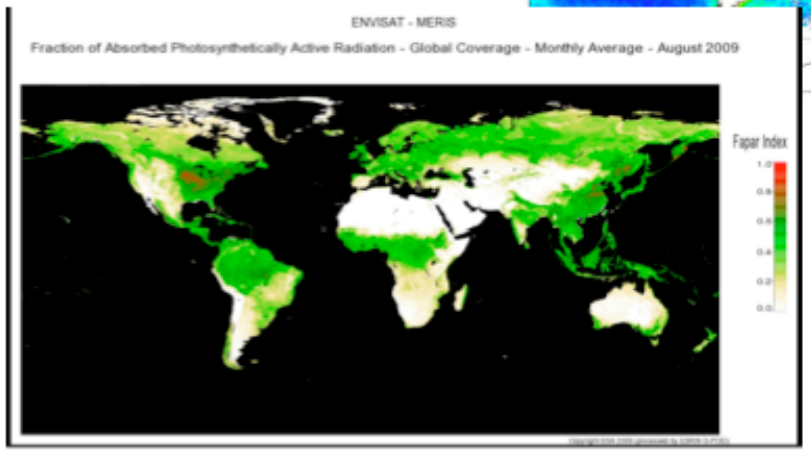
terradue 2.0

Several ESA and third-party applications and services are currently integrated in G-POD








GRID PROCESSING on DEMAND

http://gpod.eo.esa.int/

# G-POD Cat-1 – Support to science

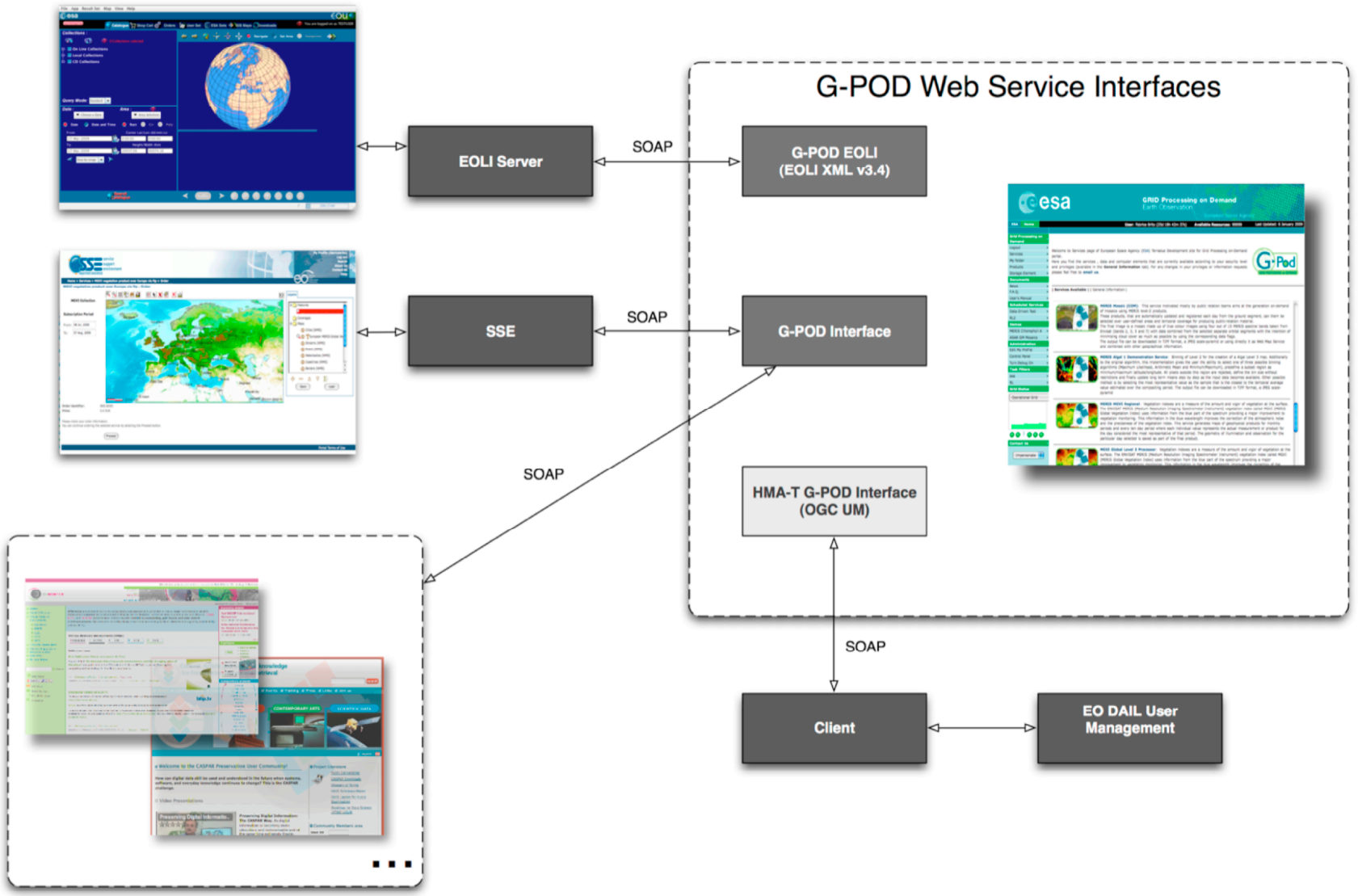
The ESA "G-POD CAT-1" opportunity creates a partnership opportunity for conducting Earth Science research activities through grid technology.

ESA offers on-line access to data, together with the G-POD attached computing infrastructure to host and run the partner's applications.


GRID PROCESSING on DEMAND

http://gpod.eo.esa.int/

# G-POD Interoperability



# Cloud IaaS in G-POD

Cloud IaaS strongly enhances G-POD by

- ✓ Providing virtually infinite storage space
- ✓ Providing large and scalable computing power
- ✓ Allowing running “any” applications
- ✓ Being secure and reliable

# Cloud IaaS in G-POD

G-POD today's computing power infrastructure:

- More than 200 working nodes
- Around 180 TB of data

It would take ***a few minutes*** to add another 200 nodes to G-POD's computing power

… it would also take a ***few minutes*** to shut-down such an infrastructure…

…maintaining G-POD's enhanced features…

# Cloud IaaS Amazon AWS

terradue 2.0

Amazon seems the most mature and flexible service today:

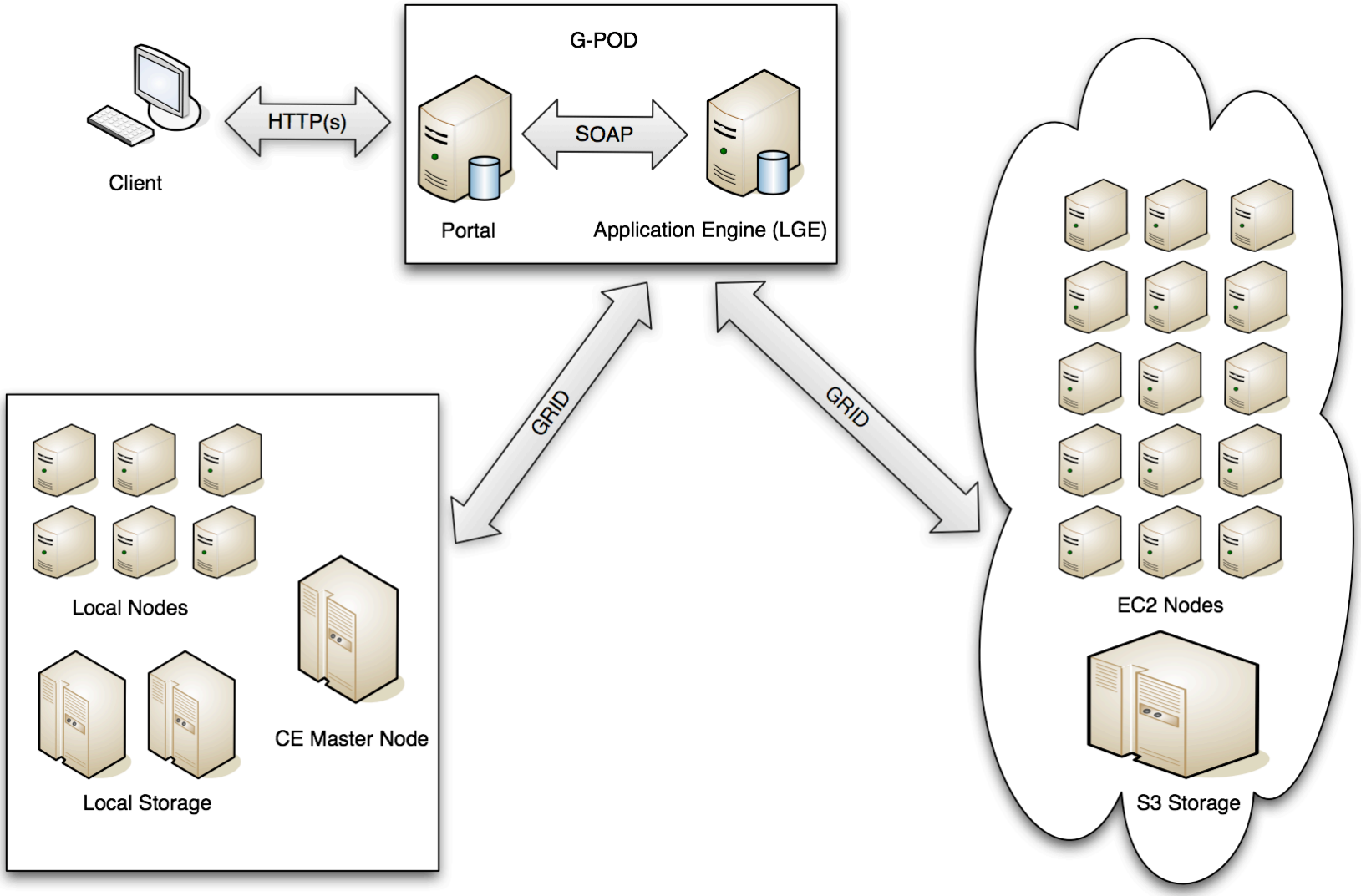
- Amazon use a lot of computer to sell books

amazon.com

- Amazon AWS is a collection of remote-computing services
- With Amazon EC2 (Elastic Compute Cloud) provides computers fro rent by the hour
- Amazon S3 (Simple Storage Service) provides virtually unlimited storage
- Simple cost model: pay as you go

# G-POD + Cloud Architecture

G-POD fruits the computing power Local and Cloud cluster resources via Grid M/W

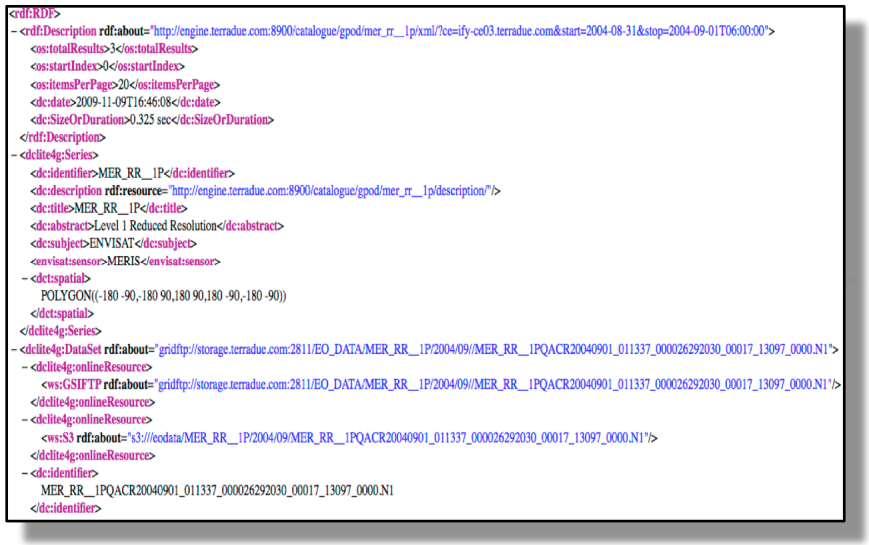


# G-POD data discovery & access

G-POD Catalogue contains:

- metadata for each Earth Observation product
- information to retrieve the product for processing from a given computing resource (Cloud based or not)

Data access: tools to execute PUT, GET, LIST, DELETE



# Resource Management

## Cloud based Computing Element Management

Secure Web Application allowing to define create/destroy resources based on:

- AWS EC2 Instance type, AMI Id
- Number of Working Nodes and Master CE

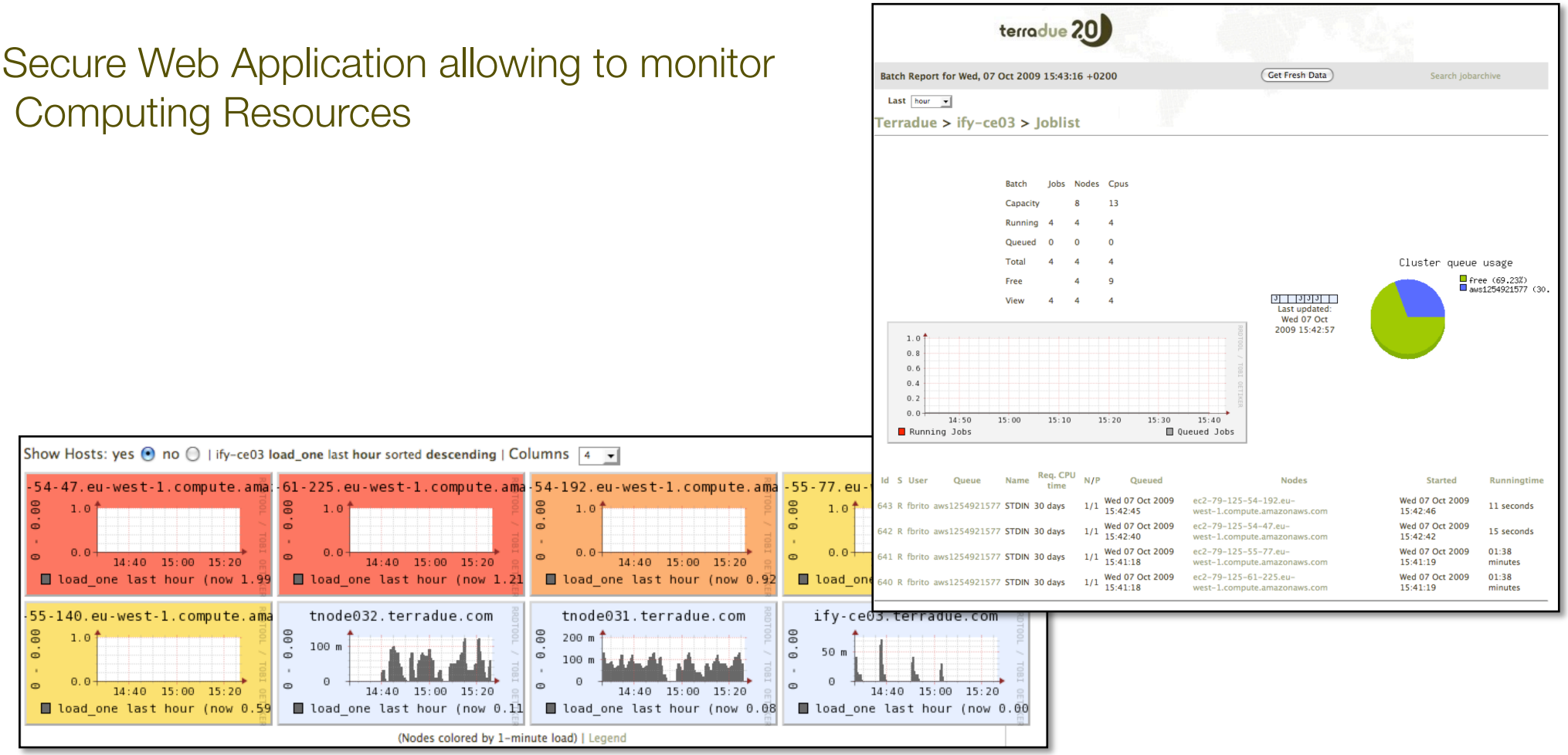
It dynamically allows to:

- Add, change and delete queues in the Computing Element scheduler
- Add/remove Cloud Computing resources to already existing queues

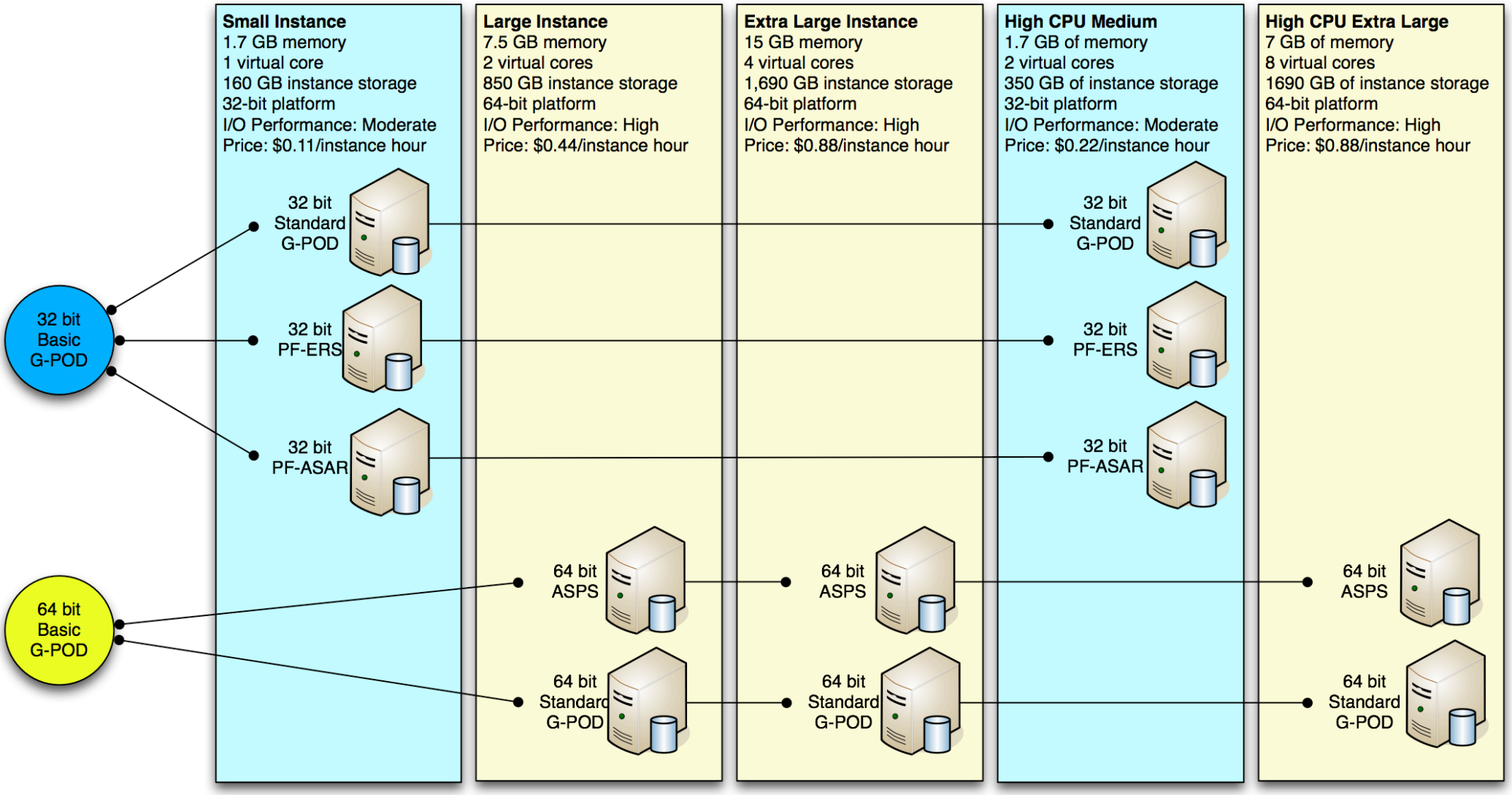
# Resource Monitoring

## Cloud based Computing Element Monitoring

Secure Web Application allowing to monitor Computing Resources



# AWS EC2 Machine types



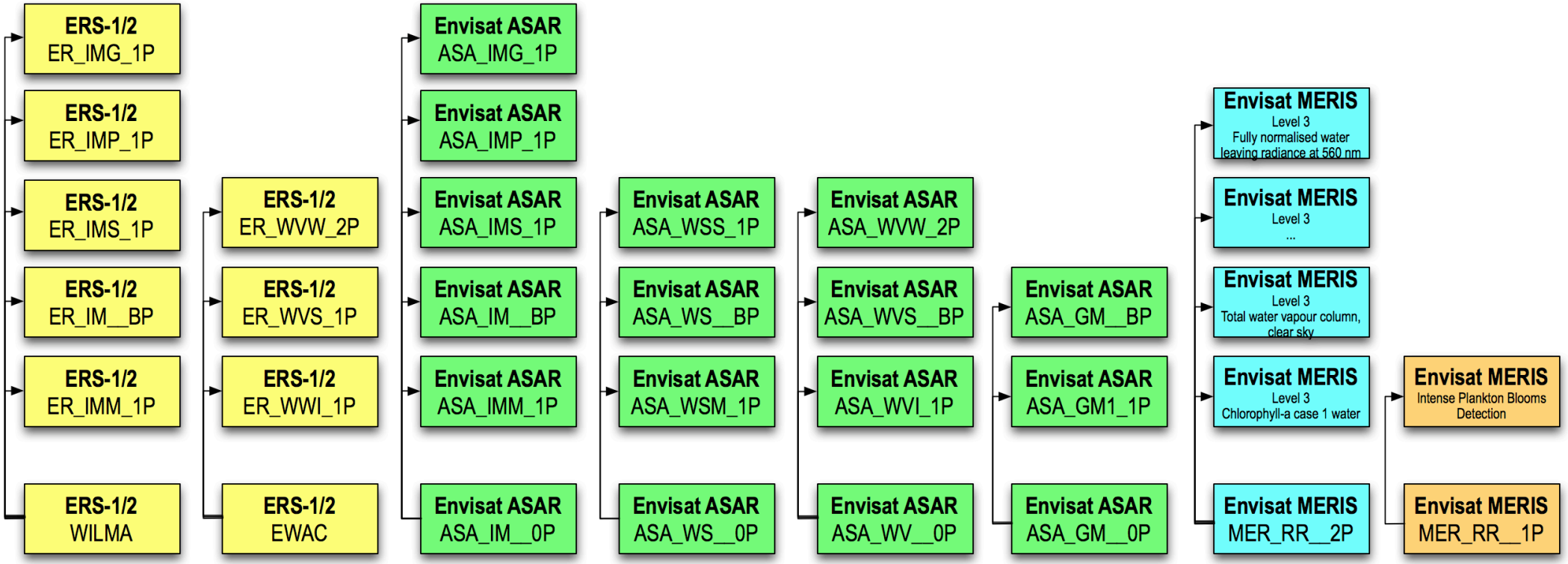
# G-POD Services in the Cloud

terradue 2.0

Services supported in G-POD Cloud today:

- PF-ERS – ERS-1/2 EWAC and WILMA L0 to L1 (and L2) processing
- PF-ASAR – Envisat ASAR L0 to L1
- MKL3 Toolbox – Envisat MERIS L2 to L3
- BEAM Band Arithmetic – Envisat MERIS L1 to L2 and L3

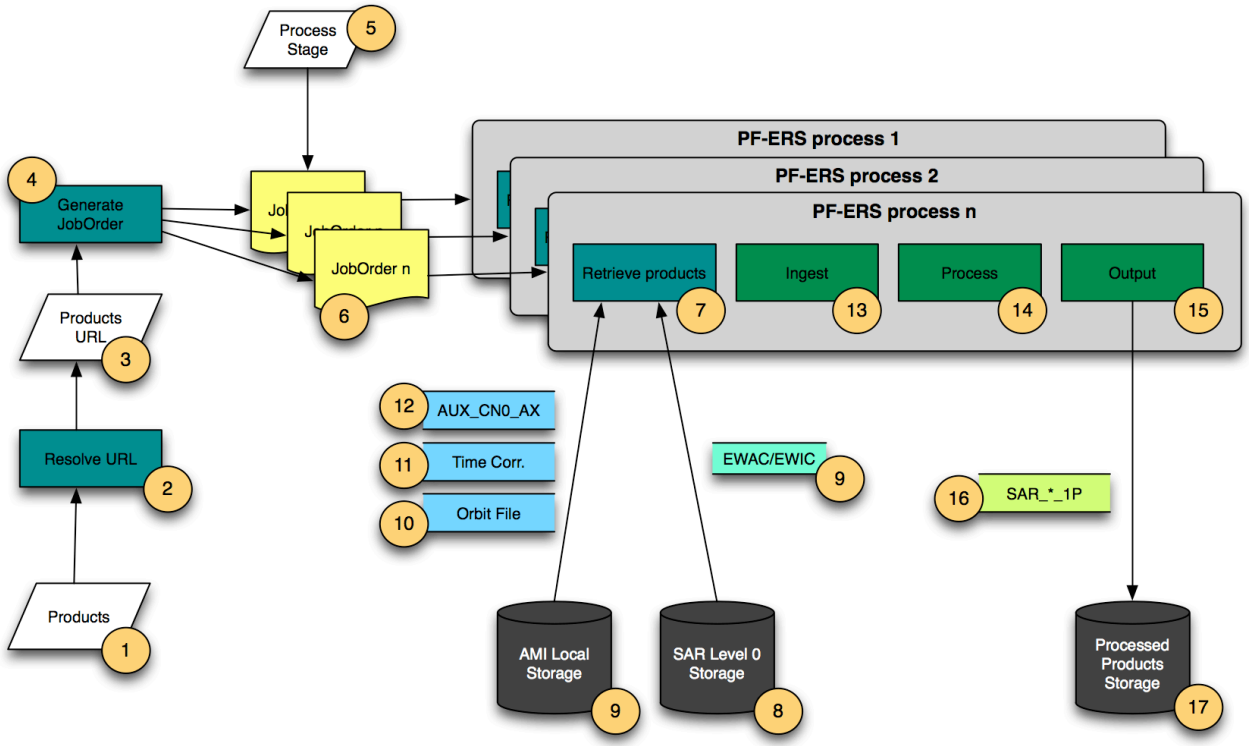
# G-POD Cloud Product Tree



# G-POD ERS SAR Processing

## ERS-2 SAR Wave processing campaign

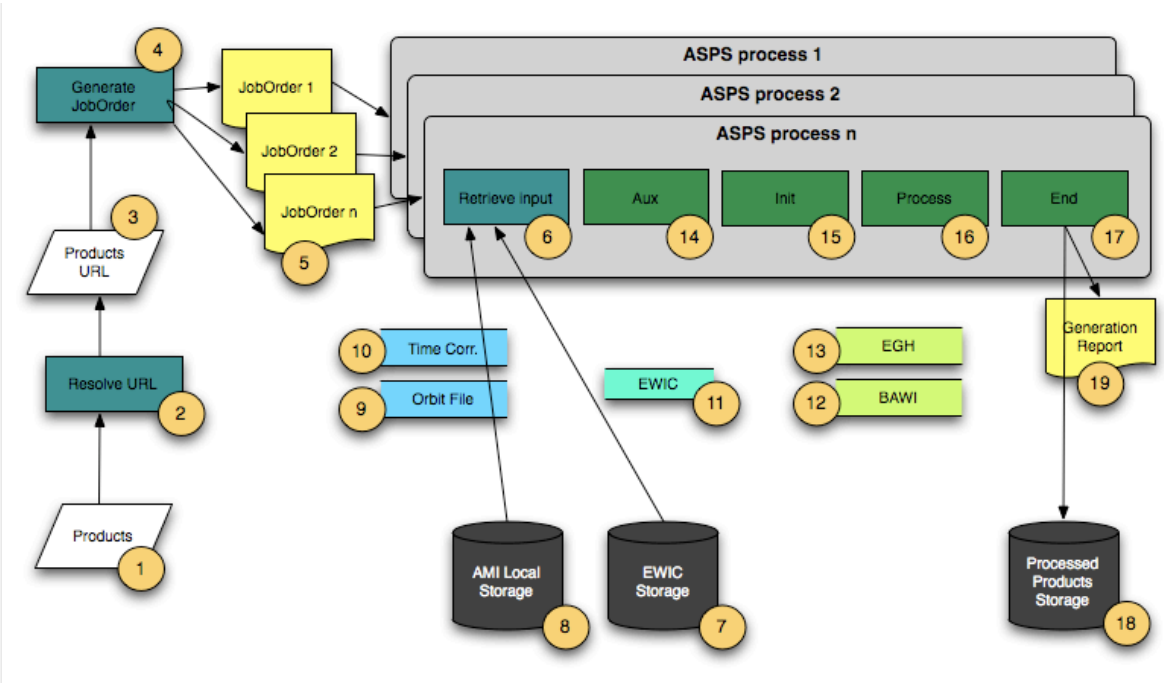
- 80 cycles of data
- Input 5 TB
- Output 20 TB




All products processed and stored using Cloud IaaS

# G-POD ERS Wind Processing

## ERS-2 Wind Scatterometer processing campaign




All products processed and stored using Cloud IaaS

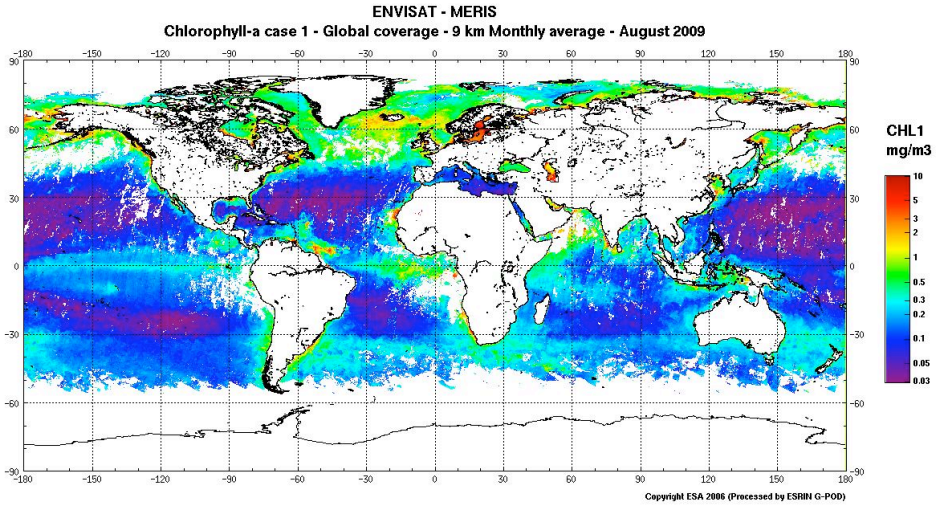
# G-POD Daily Processing

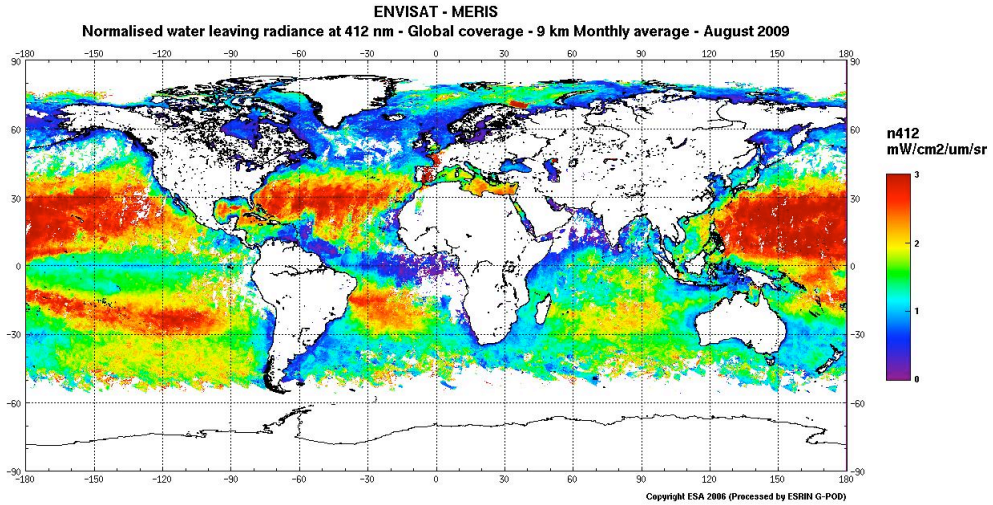
Processor used: MKL3 Toolbox

Input data: MERIS RR Level 2 ~600 MB/product

Output data: 120 Mb (HDF, netCDF, geoTiff, …)

Time to produce Level 3 product: ~1,5 hour



# G-POD Cloud Script

1 Create a CE

2 Update the CE

3 Select the Service

4 Define the Task

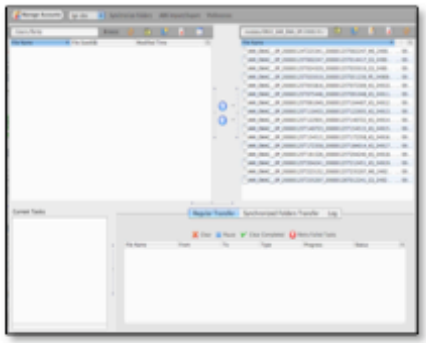
5 Save the Task

6 Check CE

7 Submit the Task

8 Monitor Resources

8 Wait for the conclusion



9 Browse the results

Repeat with other Resources and Services

# Towards standardization

- Catalogues using standards in approval phase by OGC - OpenSearch (CSW3.0)


- Processing Services walking towards OGC WPS
- Resource Management - current implementation uses EC2 API

  We are looking into OCCI for Cloud resource discovery and access

# The end

Thank you!

fabrice.brito@terradue.com