

Cloud Computing in Ground Segments

Earth Observation processing campaigns

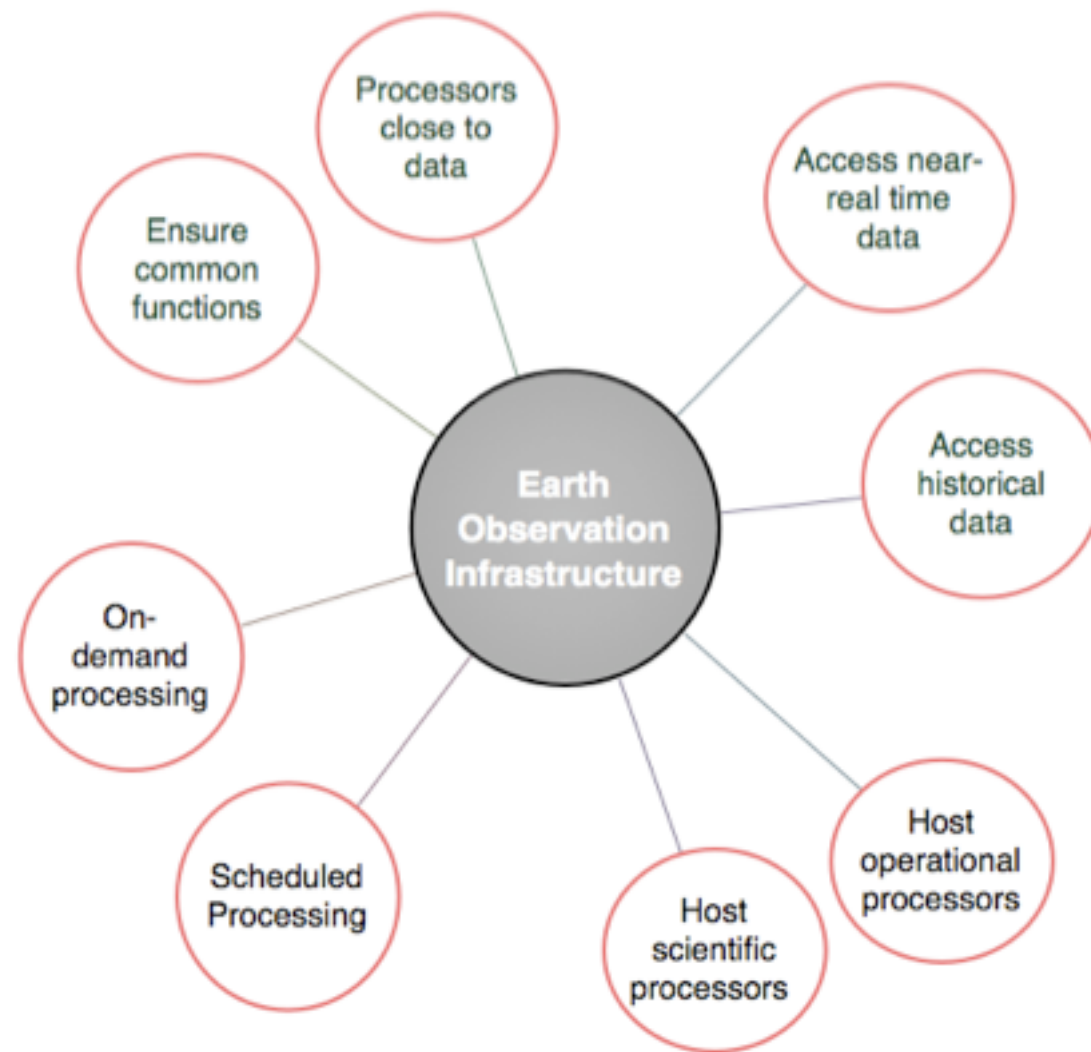
Fabrice Brito – Terradue Srl

GSAW – 2010 Geospatial Clouds



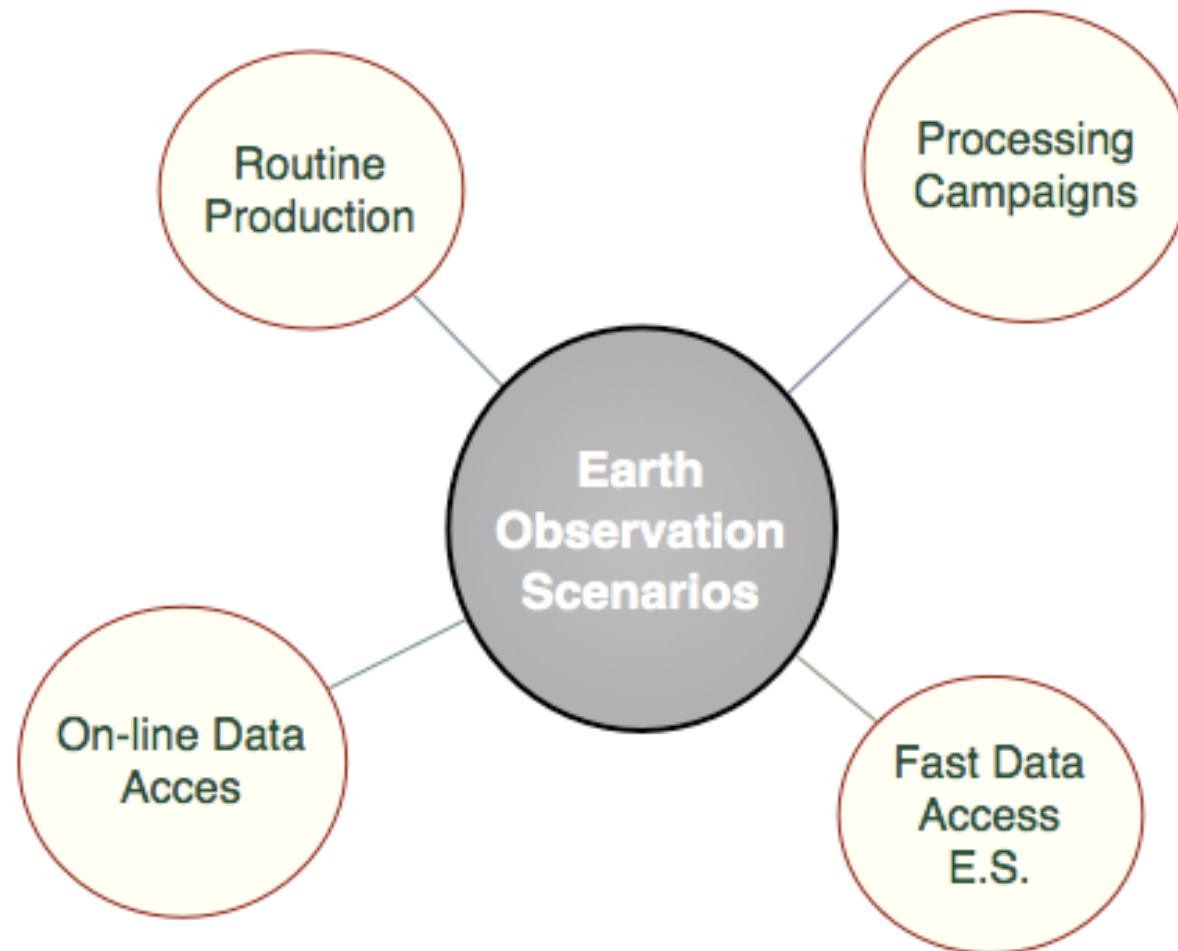
Earth Observation IT

EO computing infrastructures have particular requirements



Earth Observation Scenarios

EO computing infrastructures deal with several scenarios



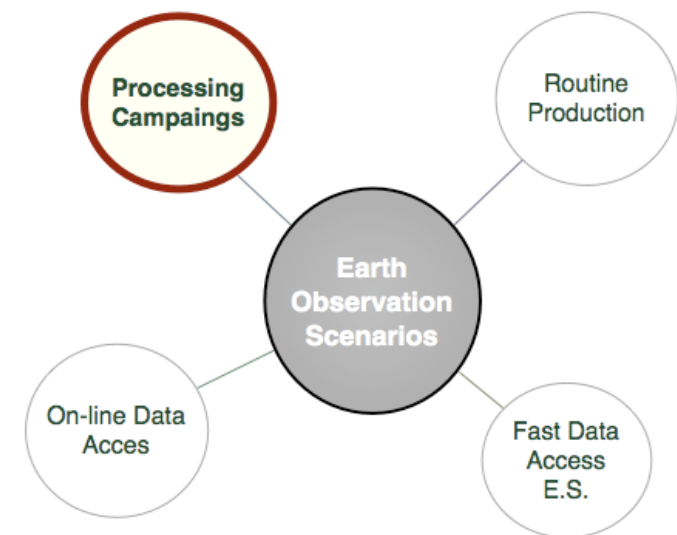
Processing Campaigns

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



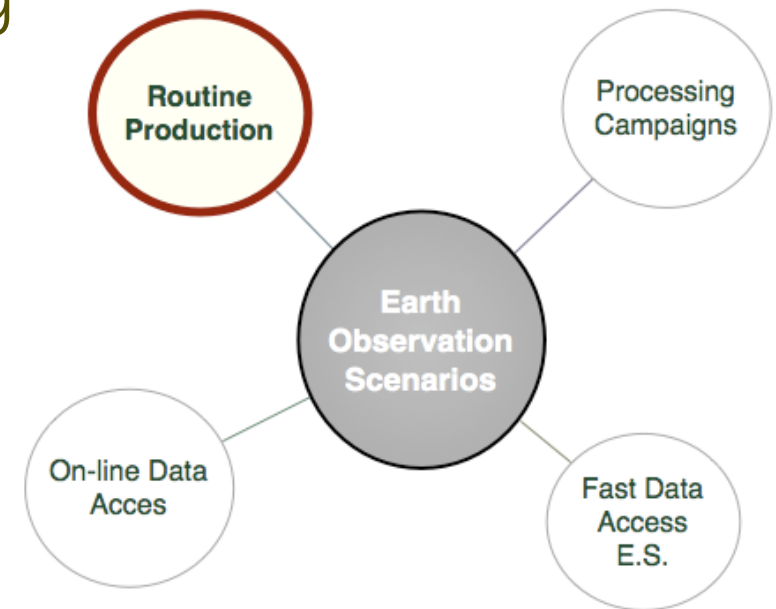
Routine Production

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



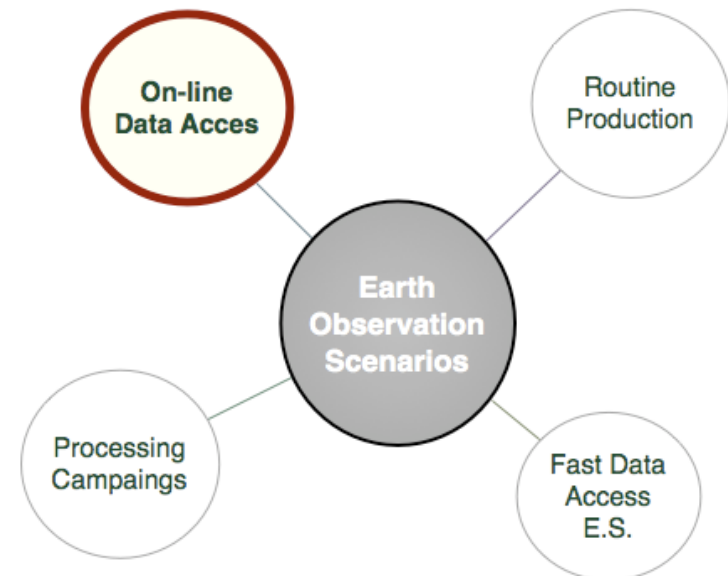
On-line data access

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

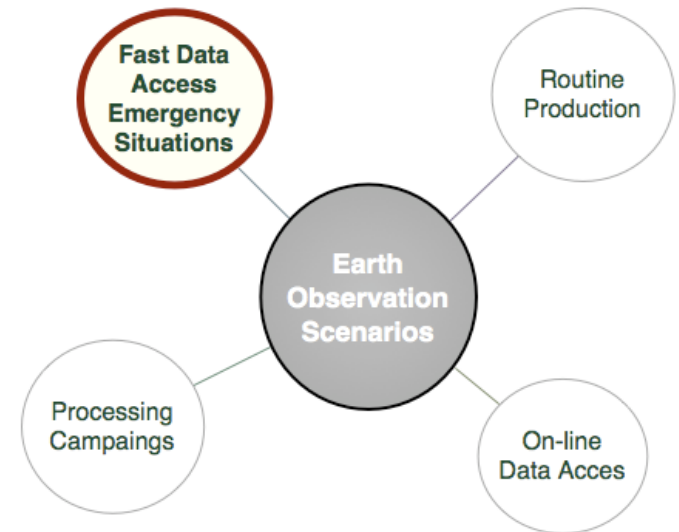


Fast data access for Emergencies

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?

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

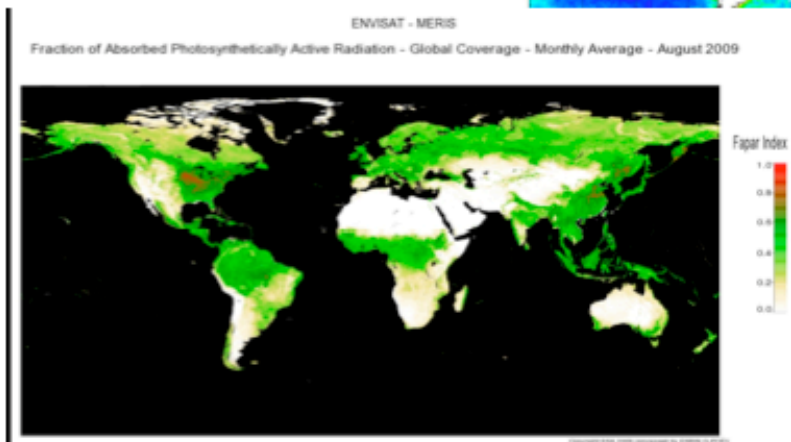
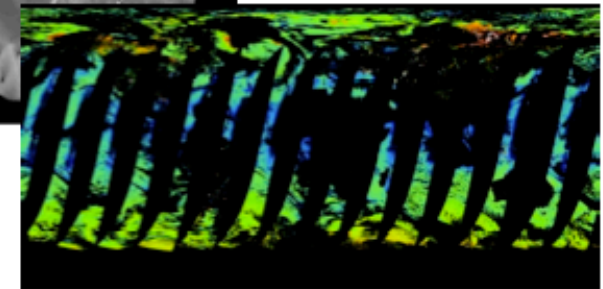
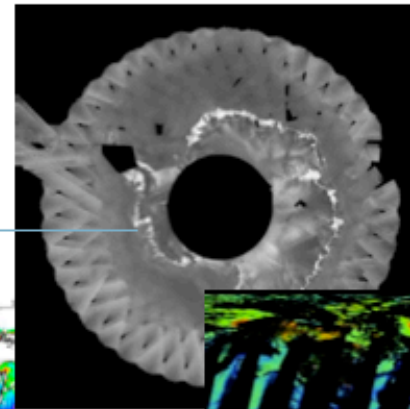
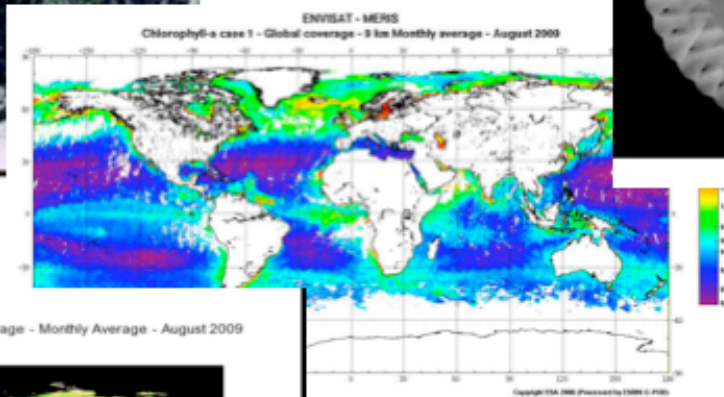
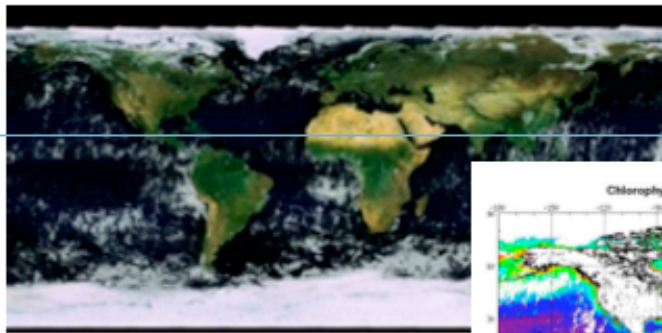
- Access and use of EO mission data available at ESA
- Provide computing infrastructure and tools to assist the generation of “scientific added value products”
- Supporting the development of earth science applications requiring significant data and processing resources
- Fostering of new partnerships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



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

G-POD Products

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



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

G-POD Cat-1 – Support to science



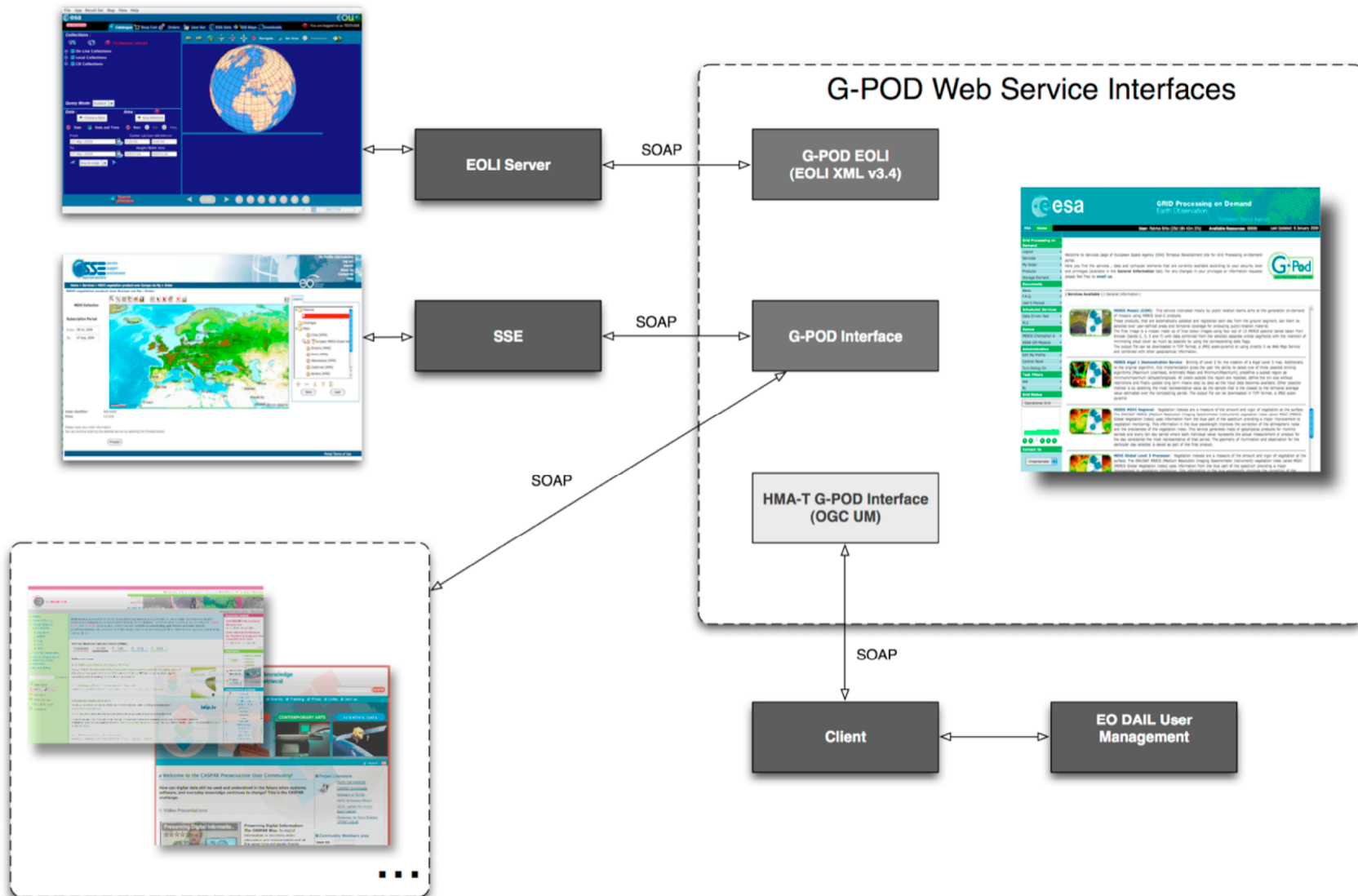
The ESA "G-POD CAT-1" opportunity creates a partnership opportunity for conducting Earth Science research activities through grid technology.

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



<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

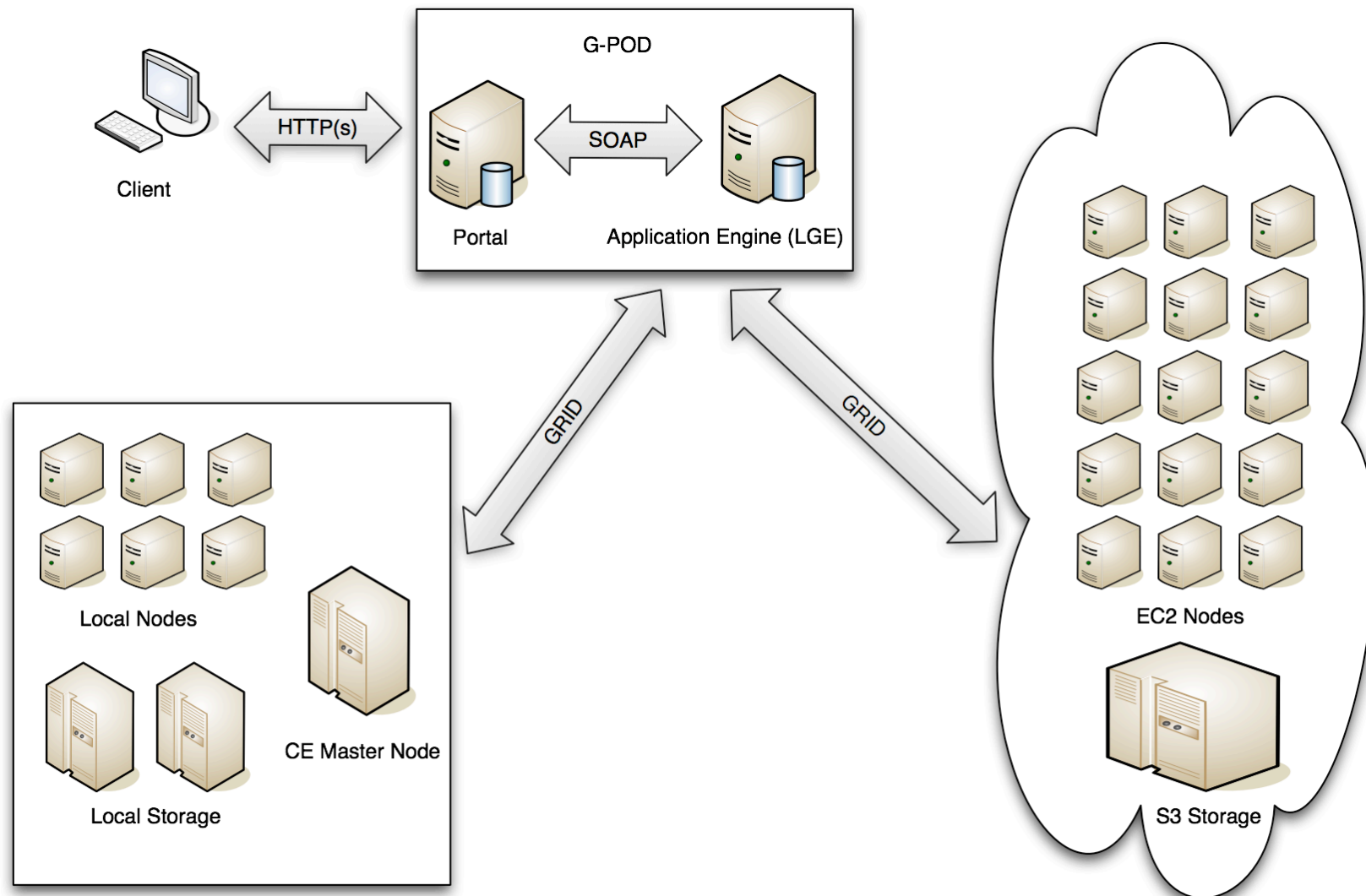
Amazon seems the most mature and flexible service today:

- Amazon use a lot of computer to sell books 
- Amazon AWS is a collection of remote-computing services
- With Amazon EC2 (Elastic Compute Cloud) provides computers for 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

```
<rdf:RDF>
- <rdf:Description rdf:about="http://engine.terra due.com:8900/catalogue/gpod/mer_rr_1p/xml/?ce=ify-ce03.terra due.com&start=2004-08-31&stop=2004-09-01T06:00:00">
  <os:totalResults>3</os:totalResults>
  <os:startIndex>0</os:startIndex>
  <os:itemsPerPage>20</os:itemsPerPage>
  <dc:date>2009-11-09T16:46:08</dc:date>
  <dc:SizeOrDuration>0.325 sec</dc:SizeOrDuration>
</rdf:Description>
- <dc:identifier>Series</dc:identifier>
  <dc:identifier>MER_RR_1P</dc:identifier>
  <dc:description rdf:resource="http://engine.terra due.com:8900/catalogue/gpod/mer_rr_1p/description"/>
  <dc:title>MER_RR_1P</dc:title>
  <dc:abstract>Level 1 Reduced Resolution</dc:abstract>
  <dc:subject>ENVISAT</dc:subject>
  <env:isat:sensor>MERIS</env:isat:sensor>
- <dc:spatial>
  POLYGON((-180 -90,-180 90,180 90,180 -90,-180 -90))
</dc:spatial>
</dc:Series>
- <dc:DataSet rdf:about="gridftp://storage.terra due.com:2811/EO_DATA/MER_RR_1P/2004/09/MER_RR_1PQACR20040901_011337_000026292030_00017_13097_0000.N1">
- <dc:onlineResource>
  <ws:GSIFTP rdf:about="gridftp://storage.terra due.com:2811/EO_DATA/MER_RR_1P/2004/09/MER_RR_1PQACR20040901_011337_000026292030_00017_13097_0000.N1"/>
</dc:onlineResource>
- <dc:onlineResource>
  <ws:S3 rdf:about="s3://eodata/MER_RR_1P/2004/09/MER_RR_1PQACR20040901_011337_000026292030_00017_13097_0000.N1"/>
</dc:onlineResource>
- <dc:identifier>
  MER_RR_1PQACR20040901_011337_000026292030_00017_13097_0000.N1
</dc:identifier>
```

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



The screenshot displays the terrardue 2.0 web application interface. At the top, the terrardue 2.0 logo is visible. Below it, a terminal window shows the execution of AWS CLI commands to create and manage EC2 instances. The output includes details about reservations, instances, and their states (pending, running, etc.). A table summarizes the instances, showing columns for instance ID, AMI ID, instance type, state, and creation time. At the bottom, there is a 'Refresh' button and a footer with copyright information and a note about being powered by gridify™.

```
Queue aws1254921577 initialised 20091007.1519
RESERVED ami-2842695c -n 5 --instance-type m1.small --kernel aki-540d2520 --debug-mode 0 --master ify-ce03.terrardue.com
Queue aws1254921577 started 20091007.1520
RESERVATION r-a2d0cbbd6 907518642958 terrardue pending lge-dev 0 m1.small 2009-10-07T13:20:05+0000
INSTANCE i-5942a32e ami-2842695c pending lge-dev 1 m1.small 2009-10-07T13:20:05+0000
INSTANCE i-4742a330 ami-2842695c pending lge-dev 2 m1.small 2009-10-07T13:20:05+0000
INSTANCE i-4542a332 ami-2842695c pending lge-dev 3 m1.small 2009-10-07T13:20:05+0000
INSTANCE i-4342a334 ami-2842695c pending lge-dev 4 m1.small 2009-10-07T13:20:05+0000
INSTANCE i-4142a336 ami-2842695c pending lge-dev 4 m1.small 2009-10-07T13:20:05+0000
Creating queue aws1254921577 at ify-ce03.terrardue.com, retrieving IPs at 20091007.1520
Polling for IPs of i-5942a32e i-4742a330 i-4542a332 i-4342a334 i-4142a336 at 20091007.1520
... pending at 2009-10-07T15:20:13+1000
... pending at 2009-10-07T15:20:23+1000
... pending at 2009-10-07T15:20:33+1000
... pending at 2009-10-07T15:20:43+1000
... pending at 2009-10-07T15:20:53+1000
Instances pending at 2009-10-07T15:21:32+1000

Queue aws1254921577
queue_type = Execution
total_jobs = 0
state_count = Transit:0 Queued:0 Held:0 Waiting:0 Running:0 Exiting:0
resources_max.walltime = 720:00:00
resources_default.needsnodes = aws1254921577
mtime = Wed Oct 7 15:20:06 2009
max_user_run = 100
enabled = True
started = True
```

Refresh

[Return to /cgi-bin/aws_nodes](#) [Return to homepage](#)

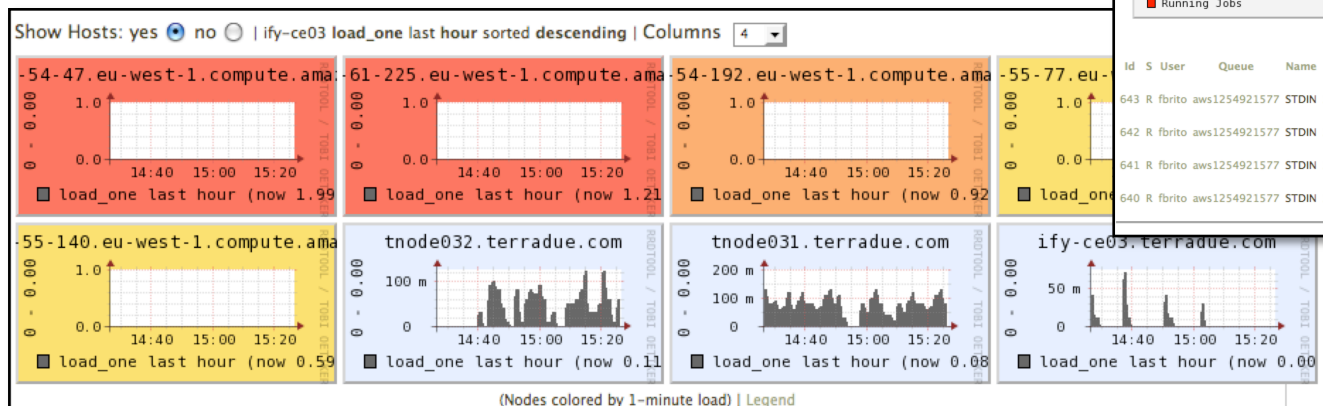
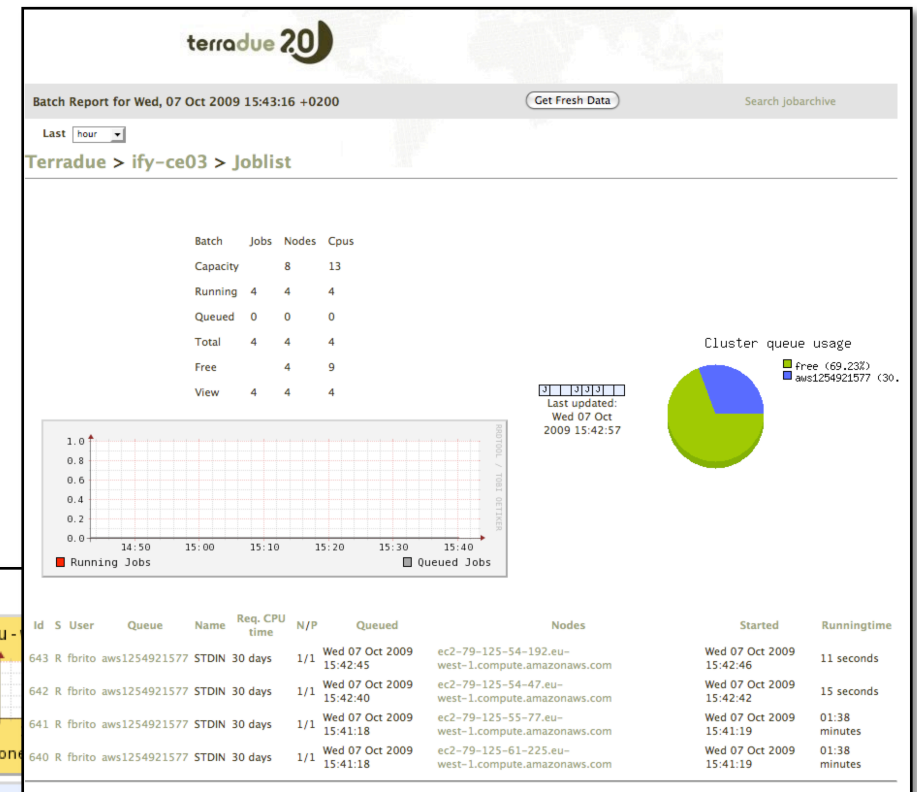
© terrardue srl 2006-2009 | powered by gridify™

Resource Monitoring

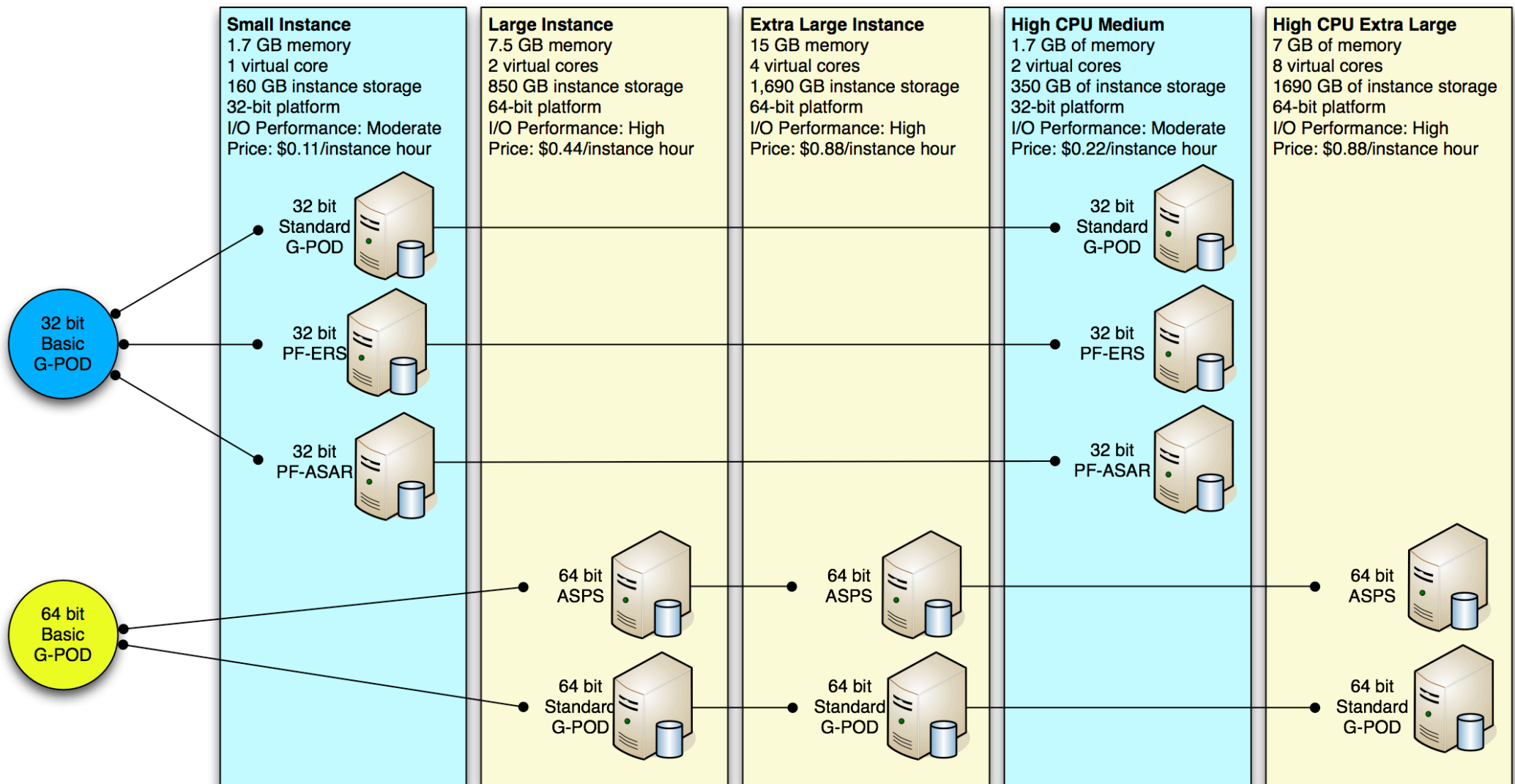


Cloud based Computing Element Monitoring

Secure Web Application allowing to monitor Computing Resources



AWS EC2 Machine types



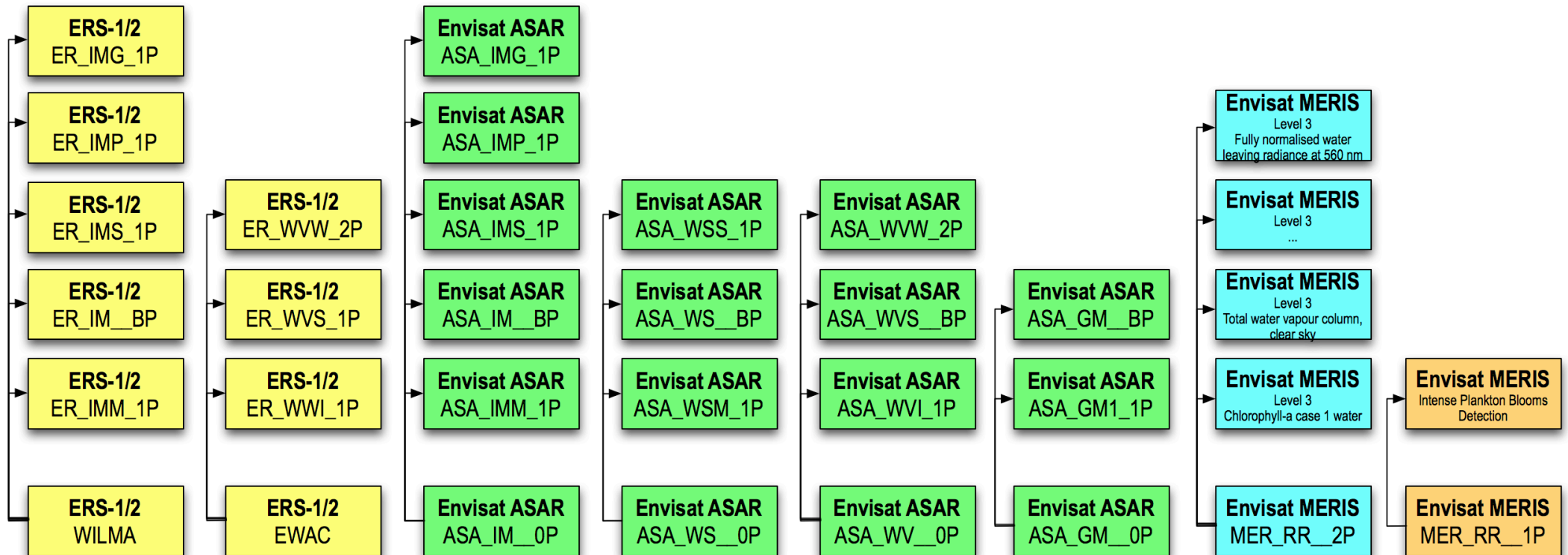
G-POD Services in the Cloud



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

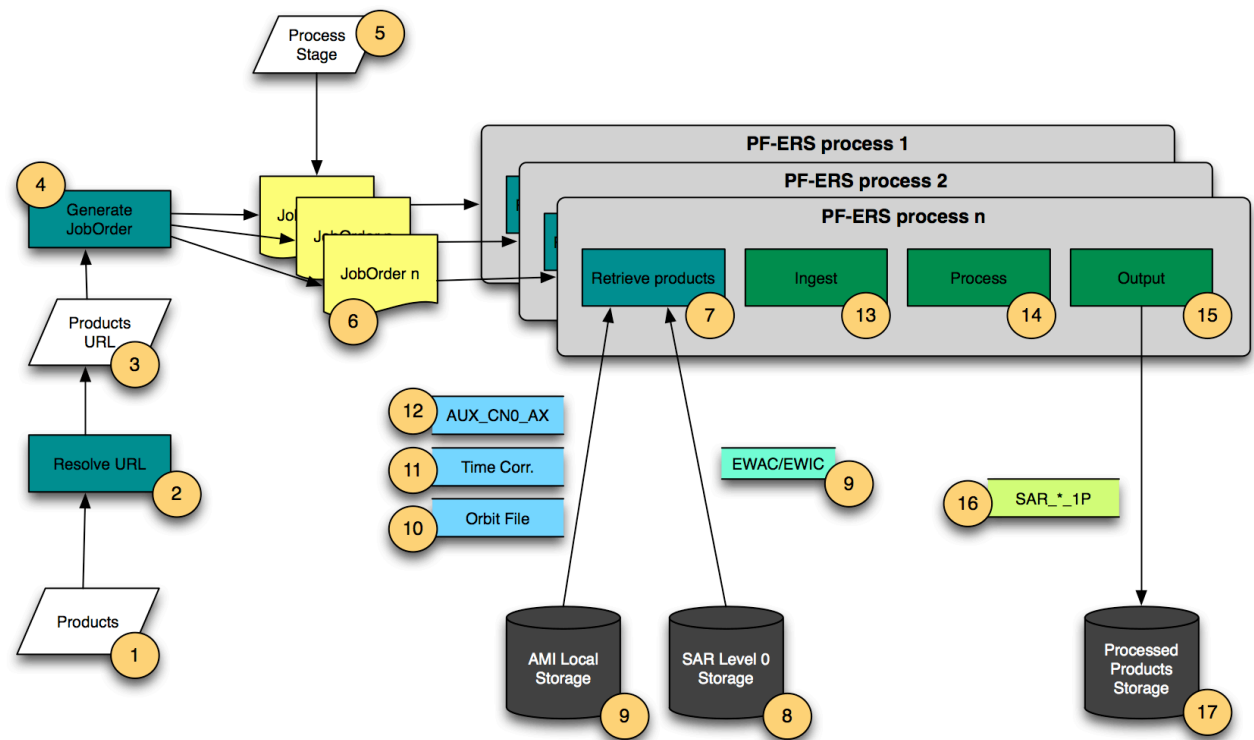


G-POD ERS SAR Processing



ERS-2 SAR Wave processing campaign

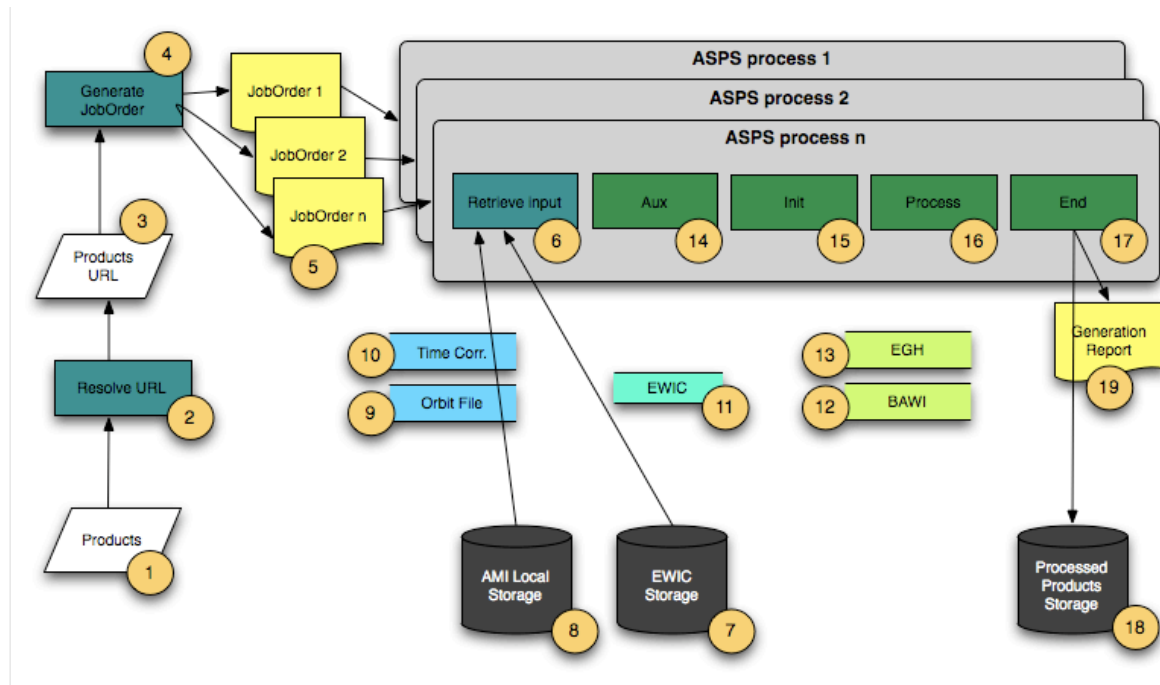
- 80 cycles of data
- Input 5 TB
- Output 20 TB



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

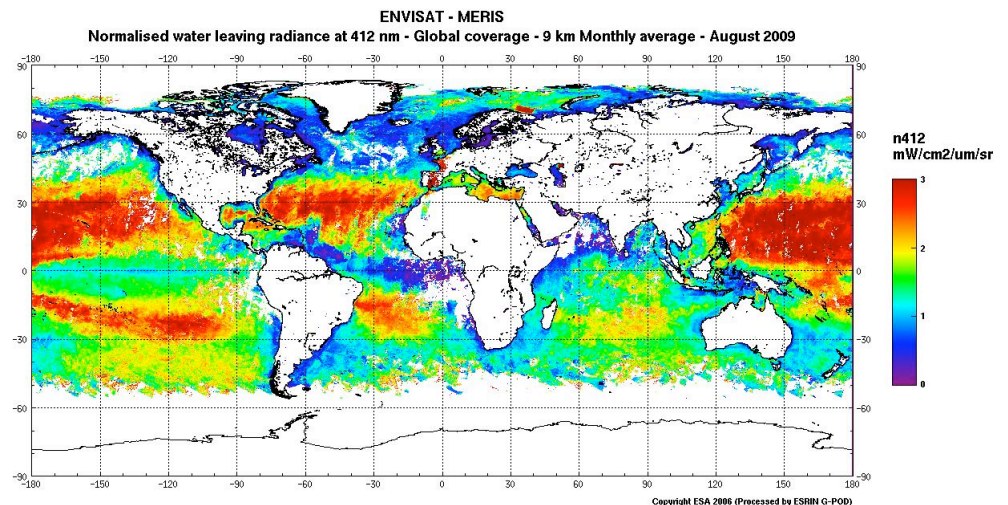
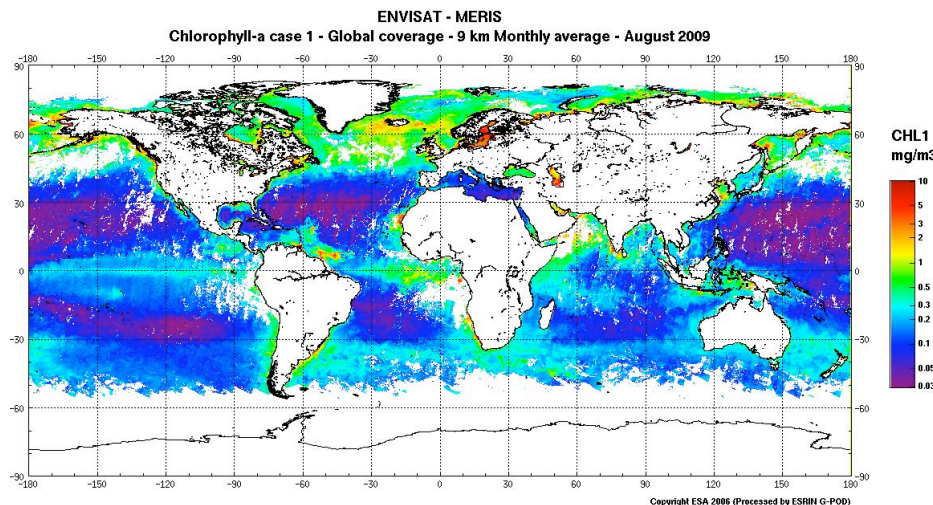
G-POD Daily Processing

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



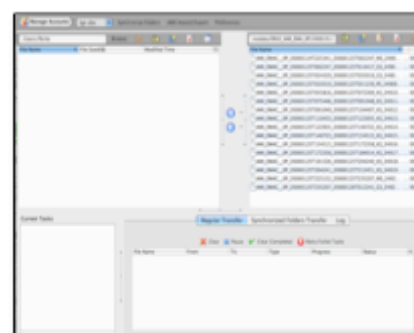
7 Submit the Task



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@terraviva.com