Testing Cloud Computing for Massive Space Data Processing, Storage and Distribution with Open-Source Geo-Software

Rubén Pérez
Elecnor Deimos
Francia 9, Polígono Ind. La Nava
13500, Puertollano, Spain
+ 34 926 44 35 78
ruben.perez@elecnor-deimos.es

Gerardo González
Elecnor Deimos
Francia 9, Polígono Ind. La Nava
13500, Puertollano, Spain
+ 34 926 44 35 78
gerardo.gonzalez@elecnor-deimos.es

Jonathan Becedas
Elecnor Deimos
Francia 9, Polígono Ind. La Nava
13500, Puertollano, Spain
+ 34 926 44 35 78
jonathan.becedas@elecnor-deimos.es

Félix Pedrera
Elecnor Deimos
Ronda de Poniente 19,
28760 Tres Cantos, Spain
+ 34 91 806 34 50
felix.pedrera@deimos-space.com

Manuel J. Latorre
Elecnor Deimos
Ronda de Poniente 19,
28760 Tres Cantos, Spain
+ 34 91 806 34 50
manuel-jose.latorre@deimos-space.com

ABSTRACT
Processing and distribution of big space data still present a critical challenge: the treatment of massive and large-sized data obtained from Earth Observation (EO) satellite recordings. Remote sensing industries implement on-site conventional infrastructures to acquire, store, process and distribute the geo-information generated. However these solutions do not cover sudden changes in the demand of services and the access to the information presents large latencies.

In this work we present the detailed design, architecture and implementation of the GEO-Cloud experiment to value if future Internet technologies can be used to overcome the previously defined limitations.

The whole system implements a complete highly demanding EO system making use of future internet technology and cloud computing to define a framework to make EO industry more competitive and adaptable to the new requirements of massive data processing and instant access to satellite information.

Categories and Subject Descriptors
C.0 [**General**]: *System architectures.*
C.2.3 [**Network operations**]: *Network monitoring*.
C.2.4 [**Distributed Systems**]: *Distributed applications*.
C.4 [**Performance of systems**]: *Performance attributes*.
H.3.4 [**Systems and software**]: *Distributed systems, performance-evaluation, user profiles and alert services*.
H.3.5 [**Online information services**]: *Web-based services*.
J.2 [**Physical sciences and engineering**]: *Aerospace*.

General Terms
Measurement, Performance, Design, Economics, Reliability, Experimentation, Verification.

Keywords
Cloud Computing, Software Architecture, Earth Observation, Remote Sensing, Distributed Systems.

1. INTRODUCTION
EO commercial data sales have increased a 550% in the last decade [1]. The field is considered a key element in the space industry and an opportunity market for the next years. Nevertheless, EO still presents critical challenges. The main one is the treatment of massive and large-sized data obtained from satellite recordings. EO industries implement on-site conventional infrastructures to acquire, store, process and distribute the geo-information generated. However these solutions have the risks of over/under size the infrastructure, they are not flexible to cover sudden changes in the demand of services and the access to the information presents large latencies.

In addition there is an increasing demand of EO services from new sectors and user typologies, some of which are newcomers to the EO industry. These new users need more flexible, easy and instant access to EO products and services through the Web. This demand has traditionally been driven through Space Data Infrastructures and heavy standards (ISO TC/211 and OGC) which are focused on interoperability rather than the real demand from the end-users.

The use of cloud computing technology can overcome the previously defined limitations because of its elasticity, scalability and on-demand use characteristics [2].

We designed and implemented the GEO-Cloud experiment in the European federated framework Fed4FIRE [3], [4] to test cloud computing for processing, storing and distributing geo-information on demand, through web services, by using open source geo-software.

The GEO-Cloud experiment utilizes the free platforms PlanetLab Europe [5], [6], Virtual Wall [7], [8] and BonFIRE [9], [10] future internet resources to go beyond conventional services and infrastructures in the EO sector and implement and test in cloud a complete EO system.

2. EXPERIMENT DESCRIPTION
The GEO-Cloud experiment consists of the emulation of a realistic and complete EO System that provides services using cloud technology. For that purpose software that simulates a constellation of 17 satellites communicating with a network of 12 ground stations that ingest the data in a cloud computing infrastructure was developed. The data is processed and distributed to end users around the world. To study the potential of cloud computing in the EO field, a novel cloud architecture

that incorporates open source software was also designed and implemented.

The GEO-Cloud experiment is divided into two sub-experiments:

- One experiment in a system integrated in Virtual Wall and BonFIRE cloud emulating the whole EO system. In Virtual Wall, the constellation of satellites and the ground stations are simulated so as to end users requesting geo-information to a web service computed in cloud. In the BonFIRE cloud, we implemented a novel cloud architecture, the processing chain of the data, its archive, catalogue and its distribution through an Image Distribution and Visualization (IDV) module. Both Virtual Wall and BonFIRE are interconnected to transfer information from one testbed to the other and viceversa.
- One experiment based in PlanetLab. This consists of real networks interconnected around the world, which we used to emulate the real behavior of the links between the ground stations and the cloud and from the cloud to the end users. The network performance, bandwidth, loss-rate and jitter were measured. Those parameters once measured were used to update the parameters of the models implemented in Virtual Wall.

The main purpose of the GEO-Cloud experiment is to validate if future internet technology provides a viable and cost-efficient solution to provide high added value services in the Earth Observation/Remote Sensing market. The experiment is also used to validate the EO models implemented, i.e. the models of the satellite constellation, the ground stations network, the cloud architecture and the end users accessing to the web services. Additionally, GEO-Cloud will additionally validate the Fed4FIRE infrastructure as a free tool to carry out large scale experiments. To achieve the previous objectives, the following metrics are analyzed:

- Processing times, varying the user models for different scenarios.
- A Benchmark comparing traditional EO infrastructures and cloud based EO systems. It includes the study of the limitations and costs of both options.
- Performance of the whole system under different highly demanding testing scenarios.

A wider description of the GEO-Cloud experiment can be found in [11].

2.1 Scenarios definition

The GEO-Cloud experiment simulates different scenarios which were defined to widely cover several applications that fit the resolution, revisit time and spectral band, among other characteristics of the satellite system. In [12] a review of different applications using optical satellites in function of the previous parameters can be found.

Trying to cover a wide variety of applications, the following scenarios were designed:

- **Scenario 1: Emergencies – Lorca's Earthquake (Spain):** Government and humanitarian organizations need the images of the affected region before and after the earthquake that took place in 2011 in order to manage the rescue and perform the damage assessment.
- **Scenario 2: Infrastructure monitoring - Affection in railway infrastructures by sand movement in desert areas (United Arab Emirates):** Providing information about dunes movements and speed during the creation of a railway infrastructure.
- **Scenario 3: Land Management – South West of England:** Land cover and land cover change products will be generated from mosaics with a weekly frequency.
- **Scenario 4: Precision Agriculture – Argentina:** A mosaic image of Argentina is provided in a daily basis. Processing offers information about irrigation planning, improved management of fertilizer usage, meteorological data affecting crops and fruit maturity.
- **Scenario 5: Basemaps – Worldwide:** With the global daily coverage of the GEO-Cloud constellation of satellites, a monthly true-color basemap is built with very high detail and coverage percentage.
- **Scenario 6: Online Catalogue / Ordering Worldwide:** An online catalogue that manages queries to the catalogue by using CSW standard.

3. EXPERIMENT DESIGN AND IMPLEMENTATION

The EO system tested in the GEO-Cloud experiment is constituted of i) a Space System Simulator including models of a satellite constellation and a ground stations network; ii) a cloud computing system that incorporates a novel architecture that integrates an orchestrator, an archive and catalogue module, an IaaS system with image processors to process the raw data acquired by the satellites and transform it into ortho-rectified images, and finally the IDV module that provides high added value web services; iii) a model of end users making requests and receiving information from the cloud web service in each of the previously described scenarios; iv) and a PlanetLab network to measure real impairments and update the system implemented in Virtual Wall and BonFIRE. Figure 1 represents the interconnectivity between all the components: the Satellite simulators, Ground Station simulators and end users were executed in individual Virtual Wall nodes with corresponding network impairments, while BonFIRE provided a cloud platform for the implementation of the cloud architecture.

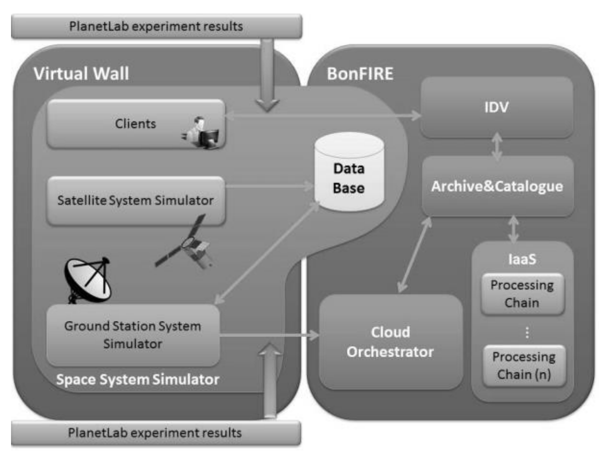

Figure 1. Architecture

3.1 Space System Simulator

The Space System Simulator is constituted of two simulators: one simulates the constellation of satellites and the other one simulates the network of ground stations. The required data to simulate the acquisition and downloading of the raw data has to be ingested from a distributed database.

3.1.1 Database

The distributed database includes the parameters representing each scenario. It is implemented in the BonFIRE cloud. From this, all the components of the spatial system simulator can retrieve the necessary information to execute every scenario. Depending on the scenario, the Space System Simulator renders the scenario, identifies satellite acquiring images and simulates the ground stations receiving the raw data.

3.1.2 Satellite System Simulator

This constellation is composed by 17 satellites in a Sun Synchronous orbit at 646 km. Images are acquired with 6.7 m Ground Sample Distance (GSD) in scenes of 160 ×160 km^2. The constellation was designed to provide daily coverage of the Earth landmass.

Once the Satellite Simulator is initialized by consulting the parameters of the distributed database, the different orbits and interactions with the Ground Stations are scheduled. Figure 2 shows the architecture of the Satellite System Simulator.

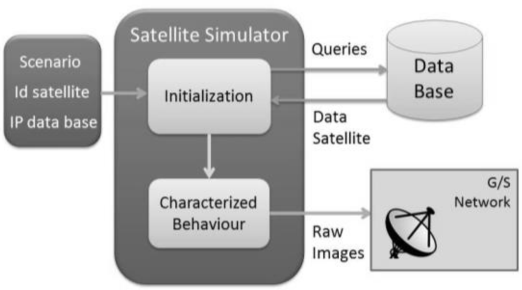

Figure 2. Satellite Simulator's Architecture

3.1.3 Ground Station System Simulator

The Ground Station System Simulator reproduces the behavior of a network of 12 ground stations distributed around the world. All the acquired data is daily downloaded to them. The amount of data generated is over 20 TBytes per day. Figure 3 shows the distribution and coverage of the ground stations.

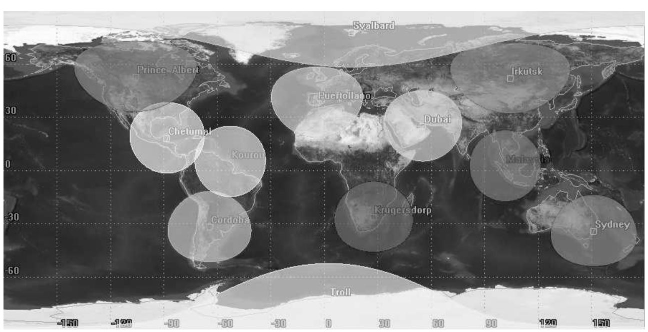

Figure 3. Footprints of the Ground Stations.

Once the Ground Station Simulator is initialized by consulting the parameters of the database, it opens a socket to listen the incoming connections from the satellites. When the satellites arrive to an area with ground station coverage, they download the raw data. When the satellite finishes the pass over a ground station the Orchestrator is able to ingest the raw data in the cloud from the ground stations. In Figure 4 the architecture of the Ground Stations Simulator is depicted.

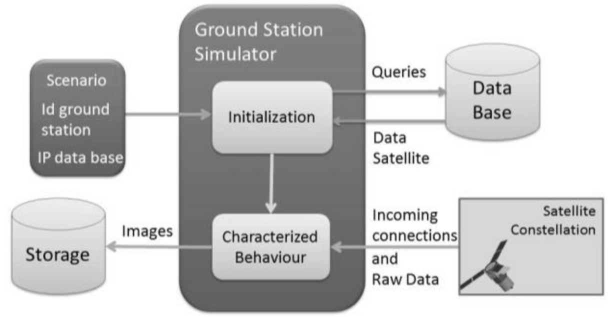

Figure 4. Ground Station Simulator's Architecture

3.2 Cloud Architecture

The designed cloud architecture facilitates the acquisition of satellite data, the on demand processing, archive and catalogue, and the distribution of geo-information to the end users Web Services. The cloud architecture is constituted of two main layers:

i) Layer 1 involves the basic satellite imagery services. It acquires satellite raw data; processes it to obtain calibrated, geo-referenced and ortho-rectified images; and stores the processed data. It is constituted of the following components:
   - Cloud Orchestrator: manages the ingestion of satellite data in the cloud, the Infrastructure as a Service (IaaS) and the Archive and Catalogue module.
   - Infrastructure as a Service (IaaS) module: creates elastic processing chains on demand to take advantage of all resources that the cloud platform provides.
   - Archive and Catalogue module: stores the images and its required metadata. It also provides a standardized CSW Service for layer 2. It is based on the open source GeoServer stack and its CSW extension.

ii) Layer 2 offers high added value services (including post-processing and analysis of imagery data) to end users through web services. The main part is the Image Distribution and Visualization module (IDV), which is based on current open source geospatial stack and infrastructures (GDAL, GeoServer, GeoWebCache). It is constituted of the following components:
   - Interface with the Archive and Catalogue through a CSW Service.
   - Datastore module, whose function is to additionally process and optimize the images stored in the Archive and Catalogue.
   - EO server, which implements OGC Web Services.
   - Tiles cache, which cache map tiles coming from a variety of sources such as OGC Web Map Service (WMS).

3.3 Modelling end users

End users are designed to test how massive requests of geo-information affect the cloud architecture. The demand is modeled according to the scenario with the following features:

- Service Type: It can be Basic (direct download of ortho-rectified images), of High Added Value (post-processed images, analysis and online visualization) or Hosting.
- Loads in Cloud Technology: Processing (Low, Medium or High), Storage (Low, Medium or High) and Communications (Urgent or Not urgent).
- Demand Variability: Constant, Variable or Highly Variable, depending on the scenario.

3.4 PlanetLab Experiment

The parameters that characterize the networks between the ground stations implemented in Virtual Wall and the BonFIRE cloud, and between the cloud and the end users models also implemented in Virtual Wall are in reality measured in PlanetLab Europe and its extension to PlanetLab Central. Those parameters are the following: bandwidth, loss-rate and jitter. By measuring those parameters we obtained realistic temporal evolution models for the networks.

We communicated nodes from PlanetLab Europe and PlanetLab Central located in the nearest location to the real ground stations with one representing the cloud.

The experiment consisted of transferring UDP (to obtain the loss-rate and jitter) and TCP packets (to obtain the maximum bandwidth the network is able to proportionate) between nodes.

3.5 Used Tools

The tools used in the GEO-Cloud experiment are the following:

- Python (https://www.python.org/).
- NEPI (http://nepi.inria.fr/).
- GeoServer (http://geoserver.org/display/GEOS/Welcome).
- MySQL (http://www.mysql.com/).
- Apache Tomcat (http://tomcat.apache.org).
- GeoWebCache (http://geowebcache.org).
- MarlinRenderer (https://github.com/bourgesl/marlin-renderer).
- Java Advanced Imaging API (http://www.oracle.com/technetwork/java/)
- PostgreSQL (http://www.postgresql.org).
- PostGIS (http://postgis.net).

4. STATE PROGRESS AND PRELIMINARY RESULTS

Currently, the Space System Simulator has been implemented and tested for scenarios 1 to 4. It was proven the correct execution of the simulator. The cloud architecture was implemented and successfully tested with the Archive and Catalogue, the shared database and the IDV. However the Orchestrator and the Processing Chain have not been tested in the cloud yet, but in local they worked properly. Moreover, the IDV horizontal scalability is still in implementation and tests have been performed only on a client-server architecture. Finally, the PlanetLab experiment was successfully executed. An evolution model of the network was concluded from the results to reproduce the Virtual Wall network impairments. For example in the Australian node, the obtained results are summarized as follows: the average bandwidth is 1.381413 *Mbytes/s*, the standard deviation is 0.303152, the jitter value is 0.619 *ms* and the loss-rate is 0.0085%.

5. CONCLUSIONS AND FUTURE WORK

The experiment is almost completely implemented and we are starting the experimentation stage. All the modules of the system were tested and work as expected. From them we obtained preliminary satisfactory results.

These results and the project evolution suggest the feasibility of the idea of using cloud computing for massive space data management as well as the efficiency of using open-source geo software in the image distribution and visualization.

The tasks to be accomplished in future work are the following:

- Implementation in cloud of the Orchestrator and the Processing Chain in the IaaS.
- Communication between the models implemented in Virtual Wall and BonFIRE testbeds to carry out the execution of the whole experiment.

6. ACKNOWLEDGMENTS

This work was carried out with the support of the FP7 Fed4FIRE-project (”Federation for FIRE”), an Integrated project funded by the European Union’s Seventh Programme for research, technological development and demonstration under grant agreement No [318389]. It does not necessarily reflect the views of the European Commission. The European Commission is not liable for any use that may be made of the information contained herein.

7. REFERENCES

[1] Euroconsult. 2010. Satellite-based Earth Observation. *Market Prospects to 2019*. Euroconsult.

[2] Ambrust M. 2010. A view of cloud computing. *Communications of the ACM*. 53, 4, 50-58.

[3] Fed4FIRE. 2014. FED4FIRE Project. https://portal.fed4fire.eu/

[4] Vanderberghe, W. 2013, July. Architecture for the heterogeneous federation of Future Internet experimentation facilities. *Future Network and Mobile Summit (FutureNetworkSummit),* 2013, 1-11.

[5] Europe, P. 2014. PlanetLab Europe Project. http://www.planet-lab.eu/.

[6] Fdida, S., Friedman, T., & Parmentelat, T. 2010. OneLab: An open federated facility for experimentally driven future internet research. in new network architectures. Springer Berlin Heidelberg. 141-152.

[7] Wall, V. 2014. Virtual Wall Facility. *Virtual Wall Facility*.

[8] Bouckaert, S., Becue, S., Vermeulen, B., Jooris, B., Moerman, I., & Demeester, P. 2012. Federating wired and wireless test facilities through Emulab and OMF: the iLab.t use case. In Testbeds and Research Infrastructure. Development of Networks and Communities. Springer Berlin Heidelberg. 305-320.

[9] BonFIRE. 2014. BonFIRE Project. http://www.bonfire-project.eu/

[10] Hume, A. C., Al-Hazmi, Y., Belter, B., Campowsky, K., Carril, L. M., Carrozzo, G., et al. 2012. BonFIRE: A Multi-cloud Test Facility for Internet of Services Experimentation. In Testbeds and Research Infrastructure. Development of Networks and Communities. Springer Berlin Heidelberg. 81-96.

[11] Becedas, J. 2014. The GEO-Cloud Experiment: Global Earth Observation System Computed in Cloud, The Eighth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, (Birmingham, United Kingdom, July 02-04, 2014).

[12] Sandau, R. 2009. Satellite Earth Observation and Surveillance Payloads. In *Systems Concepts and Integration Panel*.