USING REMOTE SENSING IMAGES AND CLOUD SERVICES ON AWS TO IMPROVE LAND USE AND COVER MONITORING

K. R. Ferreira[a,*], G. R. Queiroz[a], G. Camara[a], R. C. M. Souza[a], L. Vinhas[a], R. F. B. Marujo[a], R. E. O. Simoes[a], C. A. F. Noronha[a], R. W. Costa[a], J. S. Arcanjo[a], V. C. F. Gomes[a] and M. C. Zaglia[a]

[a] INPE, National Institute for Space Research, São José dos Campos, Brazil - (karine.ferreira, gilberto.queiroz, gilberto.camara, ricardo.cartaxo, lubia.vinhas, rennan.marujo, rolf.simoes, carlos.noronha, raphael.costa, jeferson.arcanjo, vitor.gomes, matheus.zaglia)@inpe.br

KEY WORDS: Cloud Computing, Remote sensing images, land use and cover monitoring, Amazon Web Service (AWS)

ABSTRACT:

The National Institute for Space Research (INPE) is one of the main institutes in Brazil that produce official information about deforestation as well as land use and cover in the country, based on remote sensing images. Recently, the open data policy adopted by many space agencies and governments worldwide has made accessible petabytes of remote sensing images. To properly deal with this vast amount of images, novel technologies have been proposed and developed based on cloud computing and big data systems. This paper describes the INPE's initiatives and experiences in using remote sensing images and cloud services on the Amazon Web Services (AWS) in order to improve land use and cover monitoring.

1. INTRODUCTION

The National Institute for Space Research (INPE) is the main federal research institutes in Brazil in Earth Observation and Space Science areas. Since 1988, INPE has led projects that produce official information about deforestation and Land Use and Cover Change (LUCC), such as PRODES (INPE, 2019), DETER (Diniz et al., 2015) and TerraClass (Almeida et al., 2016). This information is created from remote sensing images and is crucial for Brazilian government to formulate public policies in the environmental area.

Since 2004, the DETER project aims at producing deforestation alerts for the Brazilian Amazon and other biomes in Brazil (Diniz et al., 2015). The PRODES project monitors shallow-cut deforestation in the Brazilian Amazon since 1988 and is responsible for providing accurate official deforestation rates for all biomes in Brazil (INPE, 2019). The data sets generated by DETER and PRODES are disseminated in a web platform called TerraBrasilis (Assis et al., 2019). The TerraClass project investigates what the deforested areas detected by PRODES have become, understanding and explaining LUCC processes in the Brazilian Amazon and Cerrado biome. Based on remote sensing data analysis and geoinformation techniques, interpreters classify deforested areas into different land use and cover classes and evaluate the spatiotemporal and semantic dynamics of these areas (Almeida et al., 2016).

Recently, the open data policy adopted by many space agencies and governments worldwide has made accessible petabytes of remote sensing images of different spatial, spectral and temporal resolutions. For effectively managing and analyzing these vast amounts of images, novel technologies have been proposed based on cloud computing and big data systems. Given these challenges, cloud computing can be seen as a viable option to deal with the data volume and processing needs (Wang et al., 2013).

Nowadays, several cloud computing environments exist, being Microsoft Azure Cloud Services and the Amazon Web Services (AWS) two of the most commonly. Microsoft Azure is a cloud service platform from Microsoft. It was launched in 2010 and offers services that can be categorized as Platform as a Service (PaaS), software as a service (SaaS) and Infrastructure as a service (IaaS). AWS is a cloud service platform by Amazon. It was launched in 2006 and also can be categorized as PaaS, SaaS and IaaS.

Both platforms are highly scalable and contains developer tools, storage, database and networking services. Considering the volume of remote sensing images, when compared to Azure, AWS has advantages because it contains several of the most used remote sensing collections, which includes Landsat-8, Sentinel-2 and CBERS-4 images.

In the current era of big Earth observation data and cloud computing environment, it is crucial to improve the land use and cover monitoring projects of INPE in order to actually take advantage of the big amount of remote sensing imagery freely available. In this context, this paper describes the INPE's initiatives and experiences in using remote sensing images and cloud services on the AWS in order to improve its projects for land use and cover monitoring. To execute this work, the INPE's team is using the credits earned under the GEO (Group on Earth Observations) and AWS Earth Observation Cloud Credits Programme.

The general objective of this work is to use and evaluate the AWS cloud computing environment, developing applications and services to improve the land use and cover monitoring projects of INPE. The current development on AWS focuses on three fronts: (1) Forest Monitor Application, described in Section 2.; (2) Brazil Data Cube generation, described in Section 3.; and (3) LUCC classification, described in Section 4.. Some final remarks are presented in Section 5..

2. FOREST MONITOR APPLICATION

Forest Monitor is a web-based platform to support the detection of deforestation alerts by accessing and visualizing the remote sensing images stored in the AWS buckets. The system allows the visualization of Sentinel-2A/MSI, Sentinel-2B/MSI Landsat-8/OLI and CBERS-4/AWFI image collections from the moment they are published on AWS. Besides that, it provides functionalities for image contrast adjustment, spatiotemporal visualization

*Corresponding author.

Figure 1: BDC Forest Monitor Application.

using a temporal slider and creation of vector layers. A picture of the prototype platform can be seen in Figure 1.

The Forest Monitor platform integrates all remote sensing images of medium spatial resolution available in the AS with the deforestation polygons detected by PRODE and DETER projects. The main idea is to provide this platform for image-interpreters in order to improve the analysis and evaluation of deforested areas using different kinds of remote sensing images. Using this platform, the interpreters can compare the detected deforestation areas with the most recently satellite image available in the AWS. When interpreters observe a deforestation, they can create new deforested areas in the platform by drawing polygons.

Figure 2 presents the Forest Monitor platform architecture. Based on services provided by RemotePixel.ca, we created Tile Map Services (TMS) using AWS Lambda service to access satellite images of the Sentinel, Landsat and CBERS AWS Buckets. For image collection metadata access, we used the STAC (SpatioTemporal Asset Catalog) API provided by the Development Seed (https://sa api.developmentseed.org/search/stac) for the Sentinel and Landsat image collections and by Kepler (https://cbers.stac.cloud/) for CBERS images.

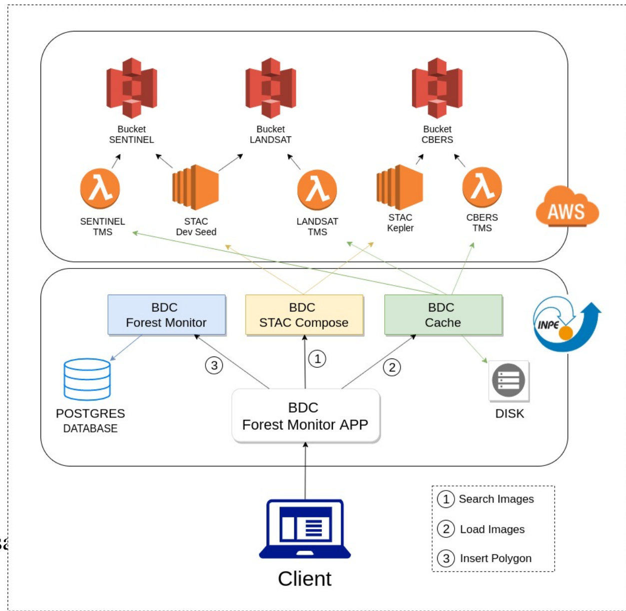

Figure 2: BDC Forest Monitor platform architecture.

3. BRAZIL DATA CUBE GENERATION

Since the beginning of 2019, the INPE's team is working in a project called Brazil Data Cube (BDC) (http://brazildatacube.org/). This project has four main objectives: (1) create analysis-ready data (ARD) from remote sensing images of medium spatial resolution (10 to 30 meters) for all Brazilian territory, including images from the Earth observation satellites Landsat, CBERS and Sentinel; (2) model these ARD data sets as multidimensional cubes with three or more dimensions that include space, time and properties; (3) propose and develop novel methods and techniques to store, process and analyze these big Earth observation data sets using satellite image time series analysis, image processing procedures and machine learning methods; and (4) create LUCC information for Brazil using the data cubes and methods developed in this project. The Brazil Data Cube project follows a worldwide trend in creating multidimensional data cubes from remote sensing images for a specific country, such as the Australian Data Cube (Lewis et al., 2017) and the Swiss Data Cube (Giuliani et al., 2017) projects.

The scripts developed in the Brazil Data Cube project are free, open source and are available on the github `https://github.com/brazil-data-cube`. The image acquisition and preprocessing scripts obtain Landsat-8 OLI and Sentinel (2A and 2B) MSI sensor images, store the images metadata in an internal database catalog and process these images to generate the surface reflectance products using LaSRC (Vermote et al., 2016) and Sen2cor (Louis et al., 2016) atmospheric correction, respectively. Images from CBERS-4 AWFI are acquired already processed as surface reflectance product using MS3 system. Image collections from

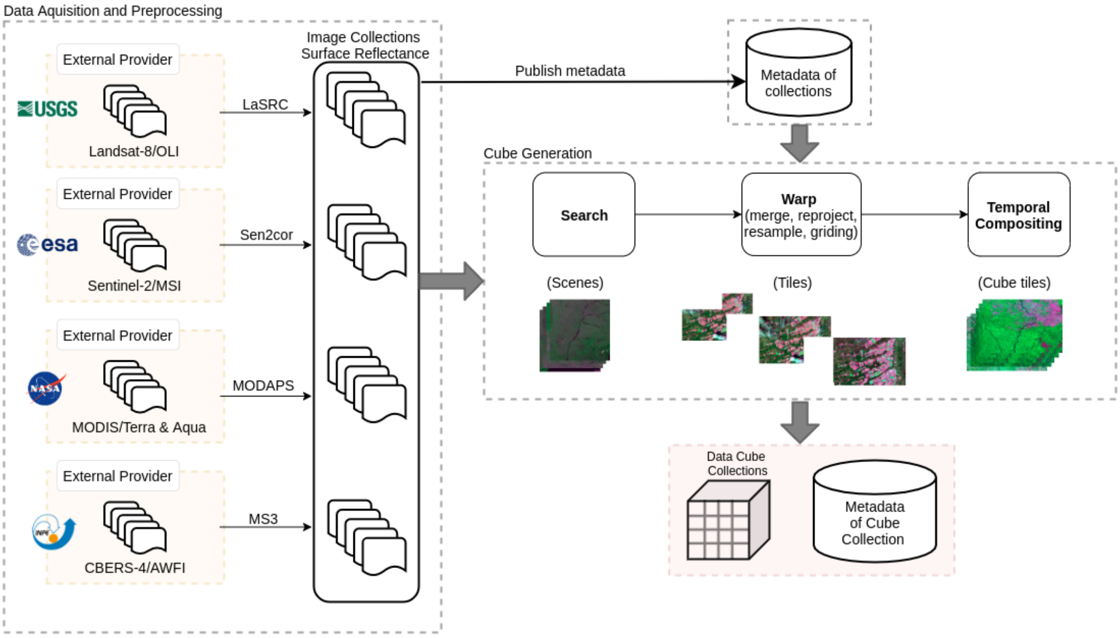

Figure 3: Brazil Data Cube generation.

CBERS-4, MODIS, Landsat-8 and Sentinel-2 are used to generate the multidimensional data cubes.

Figure 3 illustrates the data acquisition, preprocessing and multidimensional data cubes generation. To generate a data cube, a grid is used to search all available images of a given image collection, these images are merged, reprojected, resampled and gridded, hereafter, a temporal compositing function is used to build regular intervals (16 days or monthly) and reduce the data dimensionality using functions as median and best pixel approaches.

The data cube generation scripts on AWS are being developed adopting the same strategies as BDC local scripts, thus enabling interoperability between the local and cloud computing environments. The image data cubes creation is based on the orchestration of several services. Once processed locally, the surface reflectance products are being sent for storage in the Simple Storage Service (S3), which is a storage service that delivers scalability, availability, security and performance and are cataloged through a STAC service.

Figure 4 illustrates the data cube generation procedure on AWS. The AWS Lambda service allows code to run without the need to provision or manage servers, and can run up to 1000 lambdas in parallel. In the context of cube generation lambdas perform image manipulation. These lambdas are scaled by the AWS Simple Queue Service (SQS) service. The AWS Dynamo DB service is a key-value database that offers high performance. In the context of data cube generation it is used in the activities orchestration. AWS Relational Database Service (RDS) enables the configuration, operation, and scalability of relational databases in the cloud, using an instance of MariaDB/MySQL that stores the product metadata. The Kinesis service serializes messages from lambdas and enables other lambdas that can process these messages in an orderly manner, for example access to DynamoDB and RDS. At the end of the cube generation process, images are stored in S3 and their metadata are organized in RDS, so that users can accesss the repository.

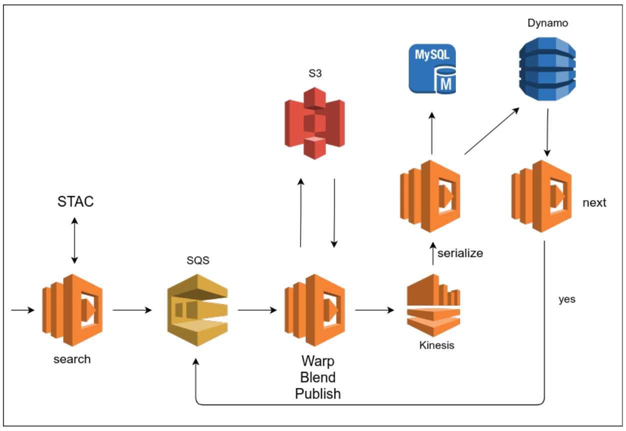

Figure 4: Brazil Data Cube generation on AWS.

4. LUCC CLASSIFICATION

The INPE's team is researching new ways to extract LUCC information from big Earth observation data sets, using satellite image time series analysis, machine learning algorithms and image processing procedures (Picoli et al., 2018) (Santos et al., 2019). These methods are being developed in an open source R package called Satellite Image Time Series (SITS) (https://github.com/e-sensing/sits). The SITS package provides a set of tools for working with analyses, clustering and classification of satellite image time series.

Figure 5 illustrates the LUCC classification on AWS, using the SITS package running in a R Studio Server in the Amazon Elastic Compute Cloud (EC2) service and the MOD13Q1 product files stored in AWS S3 buckets. To process the classification in a EC2 service, we choose a machine type called r4.8xlarge with 32 CPUs and 240 GB of RAM memory. Using this AWS EC2 service machine running the SITS R package, we took 1,5 days to classify 18 years of image time series from MOD13Q1 product for the Mato Grosso state, Brazil. The LUCC map resulting from this classification is shown in Figure 5.

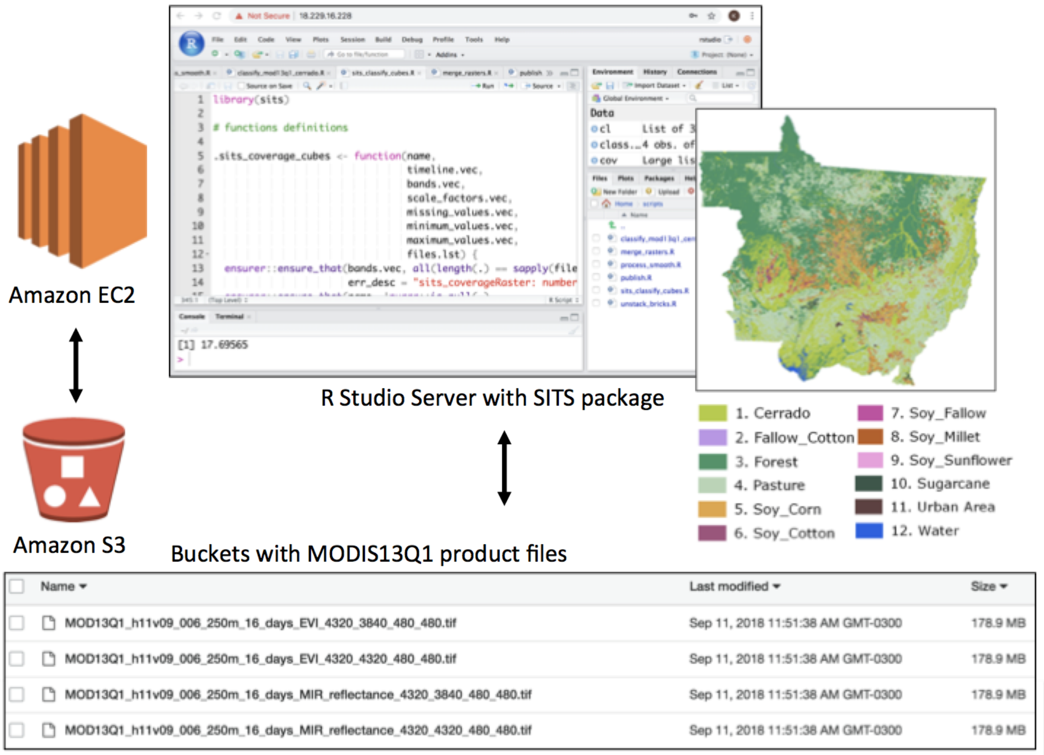

Figure 5: LUCC classification on AWS.

The Mato Grosso classification processed around 660 thousands of raster files of 480x480 pixels stored in AWS S3 buckets. Together, these files composed all time series used in our classification method. To process all these volume of files, we made a catalog metadata indexing them. This catalog, that is being developed as a web service, is a central component of the data cube technology. It enables the search and retrieval of data in raster stored in buckets and facilitates the interoperability between services and clients software.

5. FINAL REMARKS

This paper presents three INPE' initiates in using and evaluating the remote sensing images and cloud services on AWS: (1) Forest Monitor application; (2) Brazil Data Cube generation; and (3) LUCC classification. In this work, we describe the software architecture of each initiative and the AWS services used.

The AWS services tested in this work are very fast and useful, specially the AWS Lambda service that allows users to run code without provisioning or managing servers. Table 1 shows the cost to generate a data cube of Sentinel 2 images with spatial resolution of 10 meters and temporal resolution of 16 days (called Sen10m16d), comparing the use of a premise hardware with AWS lambdas in this generation. In this table, we can see the gain in using AWS Lambdas to generate multidimensional data cubes from remote sensing images.

The main drawback faced in this work is that Earth on AWS is missing surface reflectance products (ARD data) of Sentinel 2, Landsat 8 and CBERS 4. To generate the data cubes described in Section 3., we need to download the original images from USGS and ESA at INPE, run the software tools Sen2Cor and LaSRC at INPE to generate the surface reflectance images and upload them to AWS. Thus, we need to pay AWS S3 buckets to store these surface reflectance images for all Brazilian territory.

As future work, we intend to use and evaluate the AWS SageMaker service to build, train, and deploy machine learning models.

ACKNOWLEDGEMENTS

The authors would like to thank the Fundo Amazônia, Amazon AWS and GEO (Group on Earth Observations) for all support in this project.

Data Cube Name	Temporal extent	Spatial extent	Premise hardware (1 machine with 32 32 CPUs / 128 GB Ram)	AWS using 260 Lambdas	AWS cost using 260 Lambdas
Sen10m16d	1 year	1 tile of the Brazil data cube grid	1242 minutes = 20 hours	5 minutes	US$ 3,00 (without S3 cost)
Sen10m16d	1 year	All Brazilian territory (560 tiles of the Brazil data cube grid)	11592 hours = 483 days	2800 minutes = 46 hours	US$ 840,00 (without S3 cost)

Table 1: Cost to generate a Sentinel-2 data cube on AWS.

REFERENCES

Almeida, C. A. d., Coutinho, A. C., Esquerdo, J. C. D. M., Adami, M., venturieri, A., Diniz, C. G., Dessay, N., Durieux, L. and Gomes, A. R., 2016. High spatial resolution land use and land cover mapping of the Brazilian Legal Amazon in 2008 using Landsat-5/TM and MODIS data. Acta Amazonica 46, pp. 291 – 302.

Assis, L. F. F. G., Ferreira, K. R., Vinhas, L., Maurano, L., Carvalho, A., Rodrigues, J., Maciel, A. and Camargo, C., 2019. Terrabrasilis: A hybrid spatial data analytics infrastructure using micro-services for large-scale thematic mapping projects in brazil. pp. 1–26.

Diniz, C. G., d. A. Souza, A. A., Santos, D. C., Dias, M. C., d. Luz, N. C., d. Moraes, D. R. V., Maia, J. S., Gomes, A. R., d. S. Narvaes, I., Valeriano, D. M., Maurano, L. E. P. and Adami, M., 2015. Deter-b: The new amazon near real-time deforestation detection system. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 8(7), pp. 3619–3628.

Giuliani, G., Chatenoux, B., Bono, A. D., Rodila, D., Richard, J.-P., Allenbach, K., Dao, H. and Peduzzi, P., 2017. Building an earth observations data cube: lessons learned from the swiss data cube (sdc) on generating analysis ready data (ard). Big Earth Data 1(1-2), pp. 100–117.

INPE, 2019. Prodes project: Brazilian amazon forest monitoring by satellite. `http://www.obt.inpe.br/OBT/assuntos/programas/amazonia/prodes`. (28 Dec. 2019).

Lewis, A., Oliver, S., Lymburner, L., Evans, B., Wyborn, L., Mueller, N., Raevksi, G., Hooke, J., Woodcock, R., Sixsmith, J., Wu, W., Tan, P., Li, F., Killough, B., Minchin, S., Roberts, D., Ayers, D., Bala, B., Dwyer, J., Dekker, A., Dhu, T., Hicks, A., Ip, A., Purss, M., Richards, C., Sagar, S., Trenham, C., Wang, P. and Wang, L. W., 2017. The Australian Geoscience Data Cube — Foundations and lessons learned. Remote Sensing of Environment 202, pp. 276–292.

Louis, J., Debaecker, V., Pflug, B., Main-Knorn, M., Bieniarz, J., Mueller-Wilm, U., Cadau, E. and Gascon, F., 2016. Sentinel-2 sen2cor: L2a processor for users. European Space Agency, (Special Publication) ESA SP SP-740(August), pp. 9–13.

Picoli, M. C. A., Camara, G., Sanches, I., Simões, R., Carvalho, A., Maciel, A., Coutinho, A., Esquerdo, J., Antunes, J., Begotti, R. A., Arvor, D. and Almeida, C., 2018. Big earth observation time series analysis for monitoring Brazilian agriculture. ISPRS Journal of Photogrammetry and Remote Sensing 145(August), pp. 328–339.

Santos, L., Ferreira, K. R., Picoli, M. and Camara, G., 2019. Self-organizing maps in earth observation data cubes analysis. In: International Workshop on Self-Organizing Maps and Learning Vector Quantization, Clustering and Data Visualization (WSOM+), Vol. 13, Barcelona, Spain, pp. 70–79.

Vermote, E., Justice, C., Claverie, M. and Franch, B., 2016. Preliminary analysis of the performance of the landsat 8/oli land surface reflectance product. Remote Sensing of Environment 185, pp. 46–56.

Wang, P., Wang, J., Chen, Y. and Ni, G., 2013. Rapid processing of remote sensing images based on cloud computing. Future Generation Computer Systems 29(8), pp. 1963–1968.