SSC19-WKVI-02

LASP SmallSat Science Data Services

Alexandria Ware, James Craft, Thomas Baltzer, Christopher Pankratz
Laboratory for Atmospheric and Space Physics
1234 Innovation Drive, Boulder, Colorado; 303-492-6835
alex.ware@lasp.colorado.edu

ABSTRACT

We are developing of a set of turn-key science data services for smallsat data management, processing, and hosting. Using cloud computing resources and existing infrastructure, we can rapidly deploy a modular data system for a mission or project. A basic system includes reliable, secure data storage, an API for fast data access worldwide, and a lightweight website with information about the mission and data API documentation. Optional add-ons include the ability to deploy science processing software using Docker containers, interactive web-based data displays, and archive deliveries to NASA or other archive facilities. The use of AWS CloudFormation templates to build new systems makes deployment and support straightforward and cost-efficient, and provides a consistent interface for both mission teams and science data users.

INTRODUCTION

As smallsats grow more numerous and complex, there is an increasing need to manage and process the data they generate. Missions have varying needs for data processing, security, and distribution, which up until now have often been handled in an ad-hoc fashion. For smallsat data to be useful to the wider community, data providers need to follow best practices such as using standard formats, providing metadata, and making data easily accessible to users. We plan to apply our extensive experience in data management systems for large missions to establish a framework through which smallsat data providers can quickly and easily make data available to their users and the general public.

IMPLEMENTATION

Cloud Computing Architecture

Our science data services will reside entirely in the cloud, using Amazon Web Services (AWS) technology. Cloud hosting allows us to set up a system which is simultaneously secure, yet accessible by authorized users anywhere in the world, with low latency and high system stability. We have designed a system architecture which takes advantage of AWS's offerings to provide the most functionality for the lowest cost.

We have based the design of the Smallsat Data Services on existing LASP-operated science data centers, while adapting the architecture to be more modular to meet the needs of smaller missions and projects. The basic architecture is captured in an AWS CloudFormation template. This allows us to provision a new system within minutes, while ensuring that each system is built to the same easy-to-maintain standard.

Each separate project is assigned its own storage area consisting of an AWS Simple Storage Service (S3) "bucket," where telemetry data can be delivered from operations centers anywhere in the world. Processing systems are set up inside virtual private clouds (VPC), which is secure and completely isolated from other projects. Access can then be restricted to only those individuals or institutions authorized by the project. Within that VPC, an Elastic Cloud Compute (EC2) instance offers on-demand, scalable computing power to carry out data processing as needed. Serverless computing is also available via AWS Lambda to run code that carries out short-term applications.

Data Storage and Access

Storage volume is instantly scalable, providing small missions with only what they need, when they need it. AWS hosting includes built-in storage redundancy such that data durability is guaranteed to 99.999999999%.

In general, the best practice for cloud computing is to "work where the data is," using computing resources in the cloud rather than downloading data to local storage. To this end, we provide the option to run automated processing in the cloud, to use Docker to deploy code into the processing pipeline, or to set up a Jupyter hub for interactive data analysis.

For those projects or users wishing to download the data, there are built-in APIs available for most operating systems and programming languages, including Python, Java, Ruby, C++, and more. We also offer the option of a lightweight Wordpress website for publishing data and

information about the mission for public release. Data can then be accessed via https, with additional options for searching and filtering available data before downloading. Identity and access management is available such that project administrators can easily control access for team users while making data available to the public.

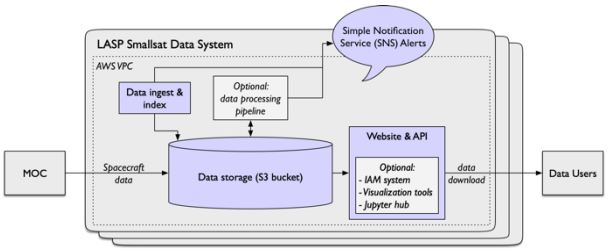

Figure 1: Smallsat Data System data flow

Data Processing

Nearly all mission data requires some form of automated data processing. Depending on the volume of data and complexity of code, this can be carried out either by serverless Lambda computing, or EC2 instances which effectively act as Linux servers. Code can be deployed directly or via Docker containers, which allow developers to work on code locally before pushing it to an identical environment in the cloud.

A Simple Notification Service (SNS) can be set up to deliver notifications via text or email directly to users, or via an API to trigger other events. The SNS triggers can be customized for notifications such as new data ingest, processing completion, processing errors and so on.

Data Archiving

Until now, archiving smallsat data has not been a high priority for most smallsat missions. NASA has expressed interest in capturing smallsat datasets in its science data archives, such as the Planetary Data System (PDS), the Space Science Data Coordinated Archive (NSSDCA), and the Earth Observing System Data and Information System (EOSDIS). We have extensive experience delivering data to all of these archives, and can assist in designing and constructing archive data structures, metadata, and documentation in accordance with archive standards. Delivery to archives can then be established on an automated cadence, or carried out at end of mission.

CONCLUSION

Smallsat data management is becoming increasingly essential as the number, volume, and scientific value of datasets grows. Data needs to be easily and rapidly accessible to users. We are using AWS to architect a standardized data system which can be quickly and easily replicated for multiple small missions at a low cost. We intend to make this service available to all interested projects, offering a basic system at a low cost such that small missions and projects can easily get their data online quickly and easily and make smallsat data available to the world.