VISAGE – A VISUALIZATION AND EXPLORATION FRAMEWORK FOR ENVIRONMENTAL DATA

Helen Conover[1], Todd Berendes[1], Patrick Gatlin[2], Manil Maskey[2], Aaron Naeger[1], Stephanie Wingo[3], Ajinkya Kulkarni[1], Charles Collins[1], Abdelhak Marouane[1], Lihua Wang[1], Shivangi Gupta[1], Khomsun Singhirunnusorn[1]

[1]The University of Alabama in Huntsville, [2]National Aeronautics and Space Administration Marshall Space Flight Center, [3]Universities Space Research Association

ABSTRACT

Diverse airborne and ground-based environmental observations are important technologies for disaster assessment and response, as well as for the validation of environmental satellite observations and atmospheric models which can improve forecasts. The VISAGE (Visualization for Integrated Satellite, Airborne and Ground-based data Exploration) project is working to provide three-dimensional visualization and basic analytics capabilities for such datasets in an interactive user interface. The use of cloud-native, serverless technologies for analysis optimized data storage will position VISAGE for integration with other technologies into a Data Analytic Center Framework.

Index Terms **—** 3D visualization, data integration, GPM ground validation, analysis optimized data store

1. INTRODUCTION

Diverse airborne and ground-based environmental observations are important technologies for disaster assessment and response, as well as for the validation of environmental satellite observations and atmospheric models which can improve forecasts. These data typically have irregular coverage, widely varying temporal and spatial scales, and are often in unique or specialized data formats. All of these factors make such data difficult to integrate into a data analytics framework. The VISAGE (Visualization for Integrated Satellite, Airborne and Ground-based data Exploration) project is working to provide three-dimensional visualization and basic analytics capabilities for such datasets in an interactive user interface. The use of cloud-native, serverless technologies for analysis optimized data storage will position the VISAGE data framework and its visualization and analytics tools for integration with other technologies in a Data Analytic Center Framework.

This paper introduces the science context and technical challenges addressed in the VISAGE project, then discusses various aspects of the project's implementation approach including its use of cloud-native, serverless technologies; analysis optimized data storage; and interactive 3D visualization research. Finally, we discuss how the use of open technologies and cloud-native services positions the VISAGE project for integration into a common disaster and environmental data analytics center framework.

2. SCIENCE VALUE

In order to address a sufficient variety of diverse environmental data, VISAGE is focusing on several science use cases from the Global Precipitation Measurement (GPM) mission's Ground Validation (GV) program [1], which investigated different precipitation regimes and related hazards including snowstorms, severe weather and flooding. GPM GV data provides a wealth of intensive, concentrated observations of atmospheric phenomena from a wide variety of ground-based, airborne and satellite instruments. These data have diverse temporal and spatial scales, variables, and data formats and organization.

3. TECHNICAL CHALLENGES

Key technical challenges for the VISAGE project include:

- Incorporation of diverse data into a common analytics framework;
- Efficient rendering and visualization of multiple high-volume, diverse three-dimensional datasets on a web-based platform;
- Temporal alignment of data with diverse time scales and resolutions;
- 3D data interrogation via map user interface; and
- Computations on data fields across instruments and platforms.

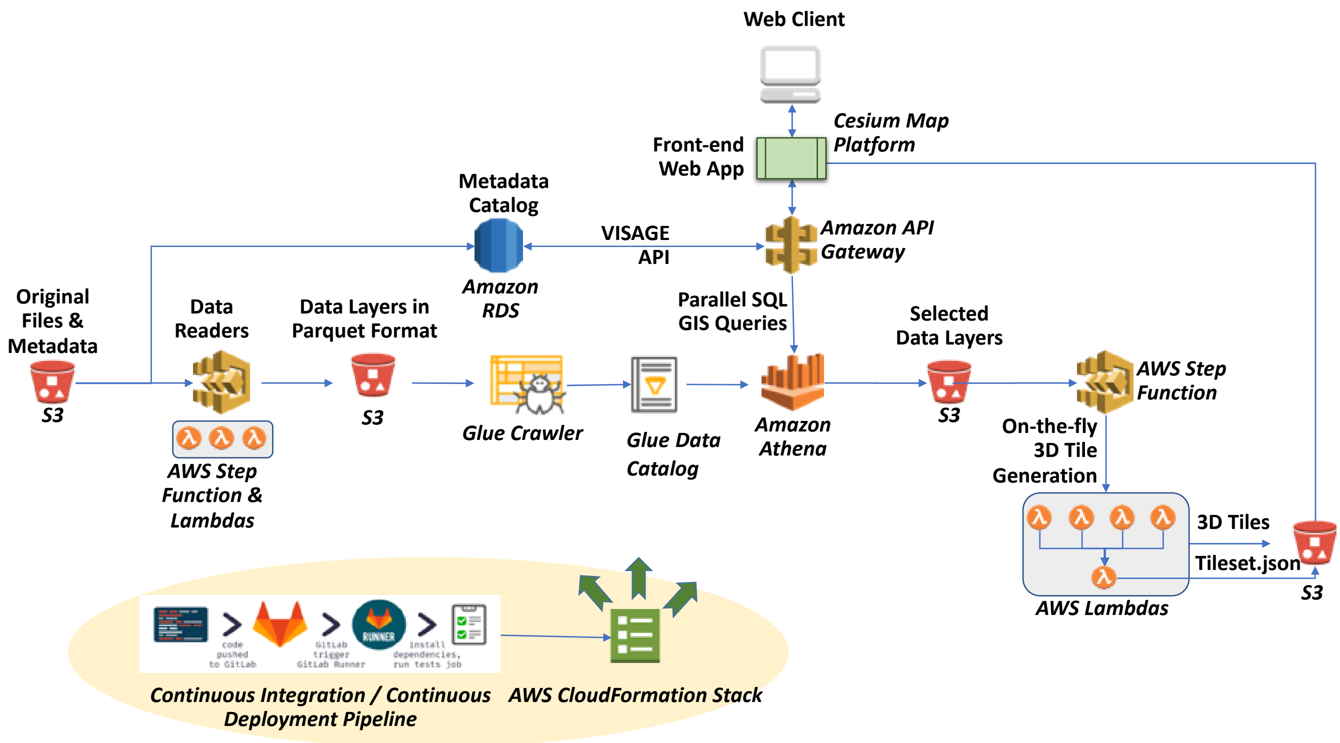

Figure 1. VISAGE cloud-native architecture with AWS Serverless technologies.

4. CLOUD-NATIVE TECHNOLOGIES

In order to better cope with large data volumes, and to align with current NASA data management advances, VISAGE is built on cloud-native technologies. Serverless computing tools from Amazon Web Services (AWS) allow for flexible scaling to support on-demand data rendering, interrogation and analytics at minimal cost – the project pays for cloud resources only when in use. Figure 1 shows the VISAGE architecture and identifies underlying AWS technologies.

5. ANALYSIS OPTIMIZED DATA STORE

Analysis optimized data, typically stored in the cloud for easy access by a broad user community, are formatted and organized for immediate use with a minimum of additional user effort. Analysis optimized data storage, is an active research area, with several implementation approaches including Zarr, a Python package providing chunked, compressed data arrays [2] and Apache Parquet columnar storage format [3].

VISAGE is using Amazon's Athena stateless query service as the interactive interface to VISAGE data. Our initial work with Athena focused on data in CSV (comma separated values), but we have moved to Apache Parquet for analysis optimized data storage. Parquet uses very efficient compression and encoding schemes compared to CSV, with an efficient nested data structure described in [4]. It allows for lower data storage costs and maximizes the effectiveness of querying the data with serverless technologies like Athena and RedShift Spectrum.

5.1. Data Ingest API

VISAGE uses a suite of configurable data readers to translate different types of data files from their native formats, (e.g., netCDF, Radar Universal Format) to JavaScript Object Notation (JSON) for posting to the data ingest application programming interface (API). The API invokes a Lambda function that automatically organizes the data into Parquet files using Hive-style partitioning [5]. Amazon Glue is used to crawl the partitioned Parquet files and construct a data catalog, which provides a single view into the data no matter where or how it is stored. Partitioning restricts the amount of data to be scanned by Athena in a given query, for lower cloud cost and greater application efficiency.

5.2. Data Access API

A common API is used to access any Parquet data file in the VISAGE repository, via the Athena query service. API functions include rendering data layers for visualization,

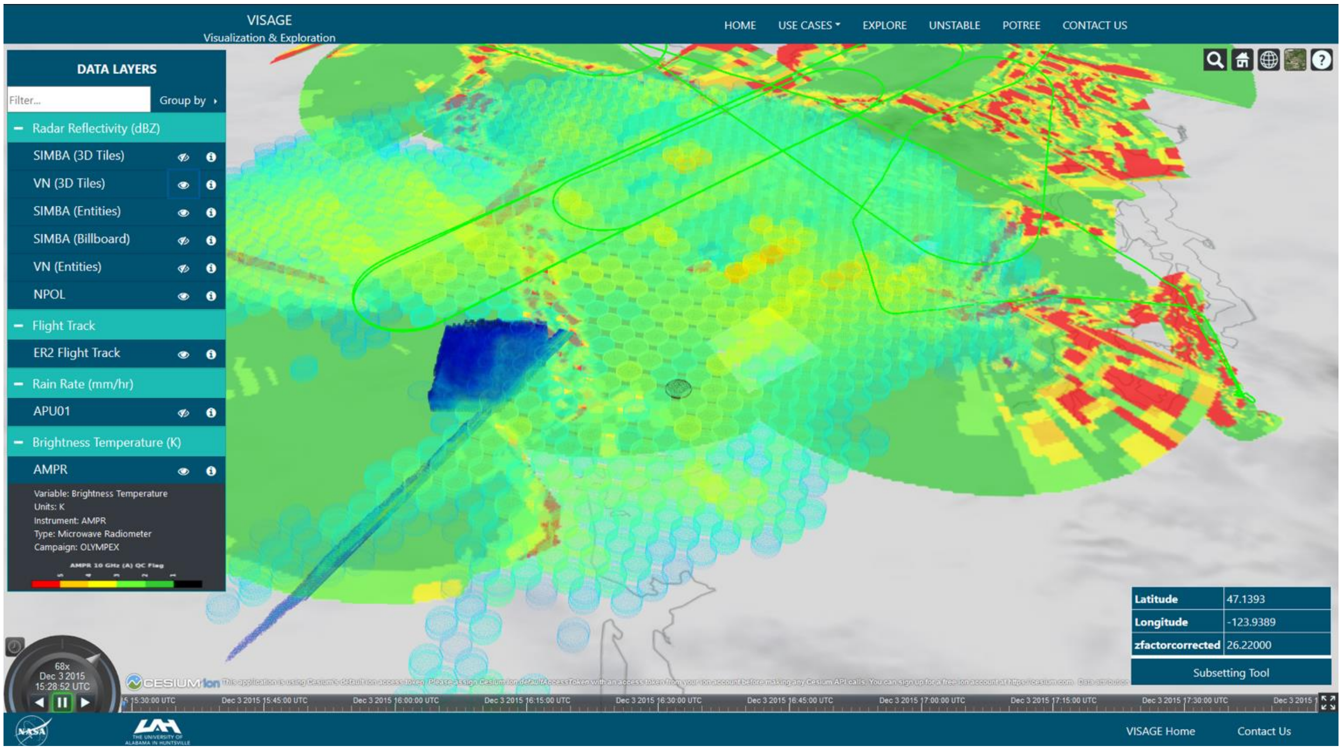

Figure 2. VISAGE user interface showing multiple datasets from the GPM GV Olympic Mountains Experiment in 2015.

providing temporal and spatial data subsets, and performing basic analytics such as statistics or histograms.

6. 3D VISUALIZATION

VISAGE is using the open source Cesium JavaScript library [6] for web-based globe and map data visualization. In particular, Cesium uses 3D Tiles [7], an open specification for streaming large, heterogeneous, geospatial datasets. This specification is being considered as a Community Standard by the Open Geospatial Consortium (OGC) and was part of the OGC Testbed-13 in 2017 [8] and Testbed-14 in 2018 [9]. The primary 3D visualization approach used in VISAGE is 3D Tile Point Clouds. We have made significant improvements to the appearance of these point clouds using stippling (offsetting each point a random distance and direction from the grid to reduce appearance of lines and other artifacts in the data) and variable transparency (lower data values more transparent so user can see through to less transparent higher data values) [10]. For more efficient handling of millions of data points, the VISAGE team is working with Entwine open source software, which organizes massive point cloud collections into the 3D Tiles Point Cloud format for native display in Cesium.

Potree [11] is an alternative approach for rendering large point clouds in web browsers. Potree point clouds use less memory than comparable 3D Tiles; their smaller binary files result in faster loading. The Potree renderer also provides more memory control than Cesium, in that it can free memory when a given point cloud is removed from view. In order to display Potree point clouds in the VISAGE Cesium-based user interface, the team integrated the two systems, using the Cesium viewer for 2D data and some 3D volumes, with Potree running in a transparent ThreeJS layer to render Potree point clouds. The team has addressed several integration challenges, include supporting user interrogation of data in both layers via "mouse click", synchronizing camera angle for both data viewers. This research is ongoing.

Figure 2 shows six datasets from the GPM GV Olympic Mountains Experiment in the VISAGE user interface. Types of data layers include vector (aircraft flight track), raster imagery (satellite and airborne microwave radiometer) and 3D data volume (satellite and ground-based radar).

7. TOWARD A COMMON FRAMEWORK

The use of both open technologies and cloud-native services positions the VISAGE project for integration into a common disaster and environmental data analytics center framework. VISAGE's core user interface technologies include the open source Cesium software library and the 3D Tiles open community standard for streaming geospatial data. The VISAGE metadata catalog is based on the Unified Metadata Model used for NASA's Common Metadata Repository. Note that UMM is compatible with the International

Organization for Standardization (ISO) 19115 family of geographic metadata standards. In addition, we are exploring cloud-based analysis optimized data storage, and following current research in this area within the Earth science analytics community. The data API can serve analysis optimized data and visualization results on demand for an analytics center.

8. ACKNOWLEDGEMENT

This work was funded by the NASA Earth Science Technology Office through their Advanced Information Systems Technology (AIST) program, grant number 80NSSC17K0414.

9. REFERENCES

[1] Schwaller, M. R., and K. R. Morris, "A Ground Validation Network for the Global Precipitation Measurement Missions," *J. Atmos. Ocean. Tech.*, 28, 301-319, 2011.

[2] Zarr Developers, "Zarr," Revision d94256d3, 2018, online at https://zarr.readthedocs.io/en/stable/, accessed 10 Jan 2019.

[3] Apache Foundation, "Apache Parquet," 2018, online at http://parquet.apache.org/documentation/latest/, accessed 10 Jan 2018.

[4] Melnik, S, Al Gubarev, J. Long, et al., "Dremel: Interactive Analysis of Web-Scale Datasets," *Proc. of the 36th Int'l Conf on Very Large Data Bases*, pp. 330-339, 2010, online at https://ai.google/research/pubs/pub36632, accessed 2 Jan 2019.

[5] Thusoo, A., J Sarma, N Jain, et al., "Hive – A Warehousing Solution over a Map-Reduce Framework," Proceedings of the VLDB Endowment 2(2):1626-1629, August 2009. DOI: 10.14778/1687553.1687609

[6] Cesium Consortium, "CesiumJS – Geospatial 3D Mapping and Virtual Globe Platform," online at https://cesiumjs.org/, accessed 10 Jan 2019.

[7] Cozzi, P., "Introducing 3D Tiles," *The Cesium Blog*, 10 August 2015, https://cesium.com/blog/2015/08/10/introducing-3d-tiles/, accessed 2 Jan 2019.

[8] Coors, V., ed., "OGC Testbed-13: 3D Tiles and I3S Interoperability and Performance Engineering Report," 5 March 2018, online at http://docs.opengeospatial.org/per/17-046.html, accessed 2 Jan 2019.

[9] Butler, H., ed., "OGC Testbed-14: Point Cloud Data Handling Engineering Report," 8 March 2019, online at http://docs.opengeospatial.org/per/18-048r1.html, accessed 21 May 2019.

[10] Conover, H., T. Berendes, A. Kulkarni, et al., "Interactive Visualization of Diverse Datasets in VISAGE", Abstract IN51D-0609, presented at 2018 *AGU Fall Meeting*, Washington, D.C., 10-14 Dec, 2018.

[11] Schuetz, M., Potree: *Rendering Large Point Clouds in Web Browsers*, Diploma Thesis submitted to the Faculty of Informatics at the Vienna University of Technology, 19 Sept 2016, online at https://www.cg.tuwien.ac.at/research/publications/2016/SCHUETZ-2016-POT/SCHUETZ-2016-POT-thesis.pdf, accessed 21 May 2019.