**Earth Science Technology Forum**  
2018-06-12

**PI:** Helen Conover, UAH  
**Co-Is:** Todd Berendes, Aaron Naeger, UAH  
Manil Maskey, Patrick Gatlin, MSFC  
Stephanie Wingo, USRA/NPP

**Technical Team:** Ajinkya Kulkarni, Charles Collins, UAH  
**Students:** Sriraksha Nagaraj, Shivangi Gupta, UAH  
**Collaborators:** David Wolff, GSFC/WFF  
Walt Petersen, MSFC

AIST-16-0094
Introduction

• VISAGE is a data exploration tool that will facilitate more efficient Earth Science investigations by providing visualization and analytic capabilities for diverse coincident datasets, with a focus on airborne field campaigns.

• Expected outcomes:
  • Interactive user interface for visualization and analytics
  • VISAGE repository containing data specific to the selected use cases
  • Support for target user community – NASA Precipitation Measurement Mission Science Team
  • Long term vision – a robust multi-sensor, multi-format integration system suitable for a wide array of applications
Science Value

- Proof-of-concept to be centered around the **GPM Ground Validation program**
  - Valuable source of intensive, coincident observations of atmospheric phenomena
  - Data from a wide variety of ground-based, airborne and satellite instruments
  - Diversity in spatial and temporal scales, variables, formats, etc., makes these data difficult to use together

- VISAGE can bring together these diverse measurements into a common framework to
  - facilitate selection of weather events or features for study
  - improve the data discovery process
  - assist with both qualitative and quantitative analysis of the measurements
  - facilitate more efficient research and analysis

- Focus on “golden cases” where most ground instruments were in operation and multiple research aircraft sampled a significant weather event, ideally while the GPM Core Observatory passed overhead
Science Use Cases from the GPM GV Program


- **Use Case:** complex baroclinic system with orographic enhancement on 3 Dec 2015; excellent sampling coordination with simultaneous satellite, airborne, & ground-based observations.

**Integrated Precipitation and Hydrology Experiment (IPHEX)** - Warm season precipitation and hydrologic processes in complex terrain (W North Carolina, April – June 2014).

- **Use Case:** warm-season convective storm with severe hail on 23-24 May 2014; observations from ground-based radars, two aircraft, and GPM Core satellite with very good GMI and DPR coverage.

**GPM Cold season Precipitation Experiment (GCPEX)** - Microphysical properties of precipitating snow (Ontario, Canada, Jan – Feb 2012)

- **Use Case:** Microphysical observation & simulation of the entire life cycle of a significant precipitation band along a warm front; multiple airborne and ground observations on 18 Feb 2012 (before GPM Core satellite launch in Feb 2014)
Technical Approach

• Key Technical Challenges
  • **3D data rendering and visualization** of multiple diverse datasets on a web-based platform
  • Data interrogation via map user interface, especially 3D data
  • Temporal alignment of data with diverse time scales and resolutions
  • Computations on data fields across instruments and platforms

• Current Research Areas
  • **Serverless Cloud-Native Technologies**
    • Amazon Web Service (AWS) Athena stateless query service for searching data stored in S3 buckets
    • AWS Step Functions and Lambdas to orchestrate and run data processing and rendering code without provisioning or managing servers, automatically scaling resources as needed
  • **3D tiles for data visualization and exploration**
    • Cesium open source geospatial 3D mapping platform
    • Point clouds, a proposed OGC community standard, to render millions of data points.
    • Experimentation with numbers of points, 3D tiles, and tileset hierarchy for efficient generation and rendering of tiles for visualization
Planned Architecture

Original Files & Metadata
S3

AWS Step Function & Lambdas

Data Readers
S3

Formatted Data Files, Layers

Amazon RDS

Parallel SQL GIS Queries

AWS Step Function

On-the-fly 3D Tile Generation

AWS Lambda

3D Tiles
S3

Tileset.json

Continuous Integration / Continuous Deployment Pipeline

AWS CloudFormation Stack

Web Client

Cesium Map Platform

Front-end Web App

VISAGE API Gateway

Amazon API Gateway

Amazon Athena

AWS Lambdas

Continuous Integration / Continuous Deployment Pipeline

12 June 2018
Earth Science Technology Forum
Cloud Native Technologies: AWS (mostly) Serverless Platform

S3 – Simple Storage Service - within S3, data objects are stored in buckets

Lambda – provides capability to run code without provisioning or managing servers, with automatic scaling

Step Functions – used to coordinate components and step through the functions of an application, e.g., to orchestrate Lambda functions

Athena – serverless, interactive query service analyze data in S3 using standard SQL

Amazon API Gateway – service to create, publish, maintain, monitor, and secure APIs

AWS CloudFormation – tools to describe and provision all the infrastructure resources in the cloud environment

AWS RDS – Relational Database Service supporting different database engines

Earth Science Technology Forum
12 June 2018
Current Implementation

Web Client

Cesium Map Platform

Front-end Web App

Amazon API Gateway

Metadata Catalog

Amazon RDS

S3

Original Files & Metadata

Data Readers

AWS Lambdas

Formatted Data Files, Metadata Catalog, Layers

S3

SQL GIS Queries

Amazon Athena

3D Tile Generation

AWS Lambdas

S3

3D Tiles

AWS Lambdas

Formatted Data Files, Metadata Catalog, Layers

S3

SQL GIS Queries

Amazon Athena

3D Tile Generation

AWS Lambdas

S3

3D Tiles

AWS Lambdas

S3

3D Tiles
Metadata Catalog (RDS)

- Initial basic metadata
  - VISAGE use cases
  - Data collections and granules
  - Information about variables, mapped to
    - GCMD science keywords
    - Instruments and platforms
    - Color tables
  - Index of 3D tiles

- Relationship to CMR/UMM*
  - Subset of CMR metadata for collections and granules
  - Metadata for variables as needed for VISAGE, but should be compatible; will evaluate against UMM-Var and UMM-Vis as these models evolve
  - Metadata for use cases specific to VISAGE, comparable to events

*ESDIS Common Metadata Repository and Unified Metadata Model
Interactive Data Repository (Athena Query Service)

- Stateless query service provides cost efficiency for VISAGE database (i.e., charged only for queries, not to maintain database uptime)

- Data files stored as comma separated value (CSV) files in S3 buckets. Athena “tables” support SQL queries to select data by time, location or value.
  - Common structure for all variables (time, location, value) in Athena data repository
  - Variable metadata (name, units, scale, range) in RDS metadata catalog

- Data is partitioned based on most commonly used query fields for better cost efficiency and response time

```
visage-db
  vn
    campaign/year/month/day/hours/minutes/seconds
  simba
    campaign/year/month/day/hours/minutes/seconds
  rain-gauge
    campaign
```
Data Transformations

Original data files to Athena database
  - Native data files are converted to CSV (and compressed)
    - Dataset specific code to translate to CSV
    - Possibility of using OPeNDAP server for translation

Athena to 3D tiles
  - Common, reusable Lambda functions
  - Working toward on-the-fly generation using parallel Lambda functions

Diagram:

1. Original Files & Metadata (S3)
2. Data Readers
3. AWS Lambdas
4. Formatted Data Files, Metadata Catalog, Layers (S3)
5. SQL GIS Queries
6. 3D Tile Generation
7. AWS Lambdas
8. 3D Tiles (S3)
3D Data Rendering and Visualization

• Why 3D tiles?
  • Designed for streaming massive heterogeneous 3D geospatial datasets
  • Natively supported by Cesium map visualization user interface
  • Supports several data models
    • Batched 3D Model - Textured terrain and surfaces, 3D building exteriors and interiors, massive models
    • Instanced 3D Model - Trees, windmills, bolts
    ➢ Point Cloud - Massive amount of points
      • Point cloud density can be adjusted to render solid or semi-transparent appearance
      • Proposed OGC Community Standard
  • Vector Data - Polygons, polylines, and placemarks (draft spec only)
  • Composite - Combine heterogeneous tile formats
  • Declarative Styling - Style features using per-feature metadata
Experiments in Making 3D Tiles

Goal: on-the-fly generation and display of data as 3D tiles

Sample dataset: GPM Validation Network of coincident satellite and ground radar reflectivity averages within a cylindrical GPM view volume were rendered as cylindrical volumes using an adjustable spatial density of points in Cesium 3D tile point cloud files.

Amazon API Gateway

SQL GIS Queries

3D Tile Generation

AWS Lambdas

3D Tiles

Web Client

Cesium Map Platform

2500 cylinders
3.9 million points

Amazon Athena

Front-end Web App
Evolution of 3D Tile Software

- **Dec 2017**: py3dtiles – public domain Python module to create point cloud 3dTiles
  
  - [https://github.com/Oslandia/py3dtiles](https://github.com/Oslandia/py3dtiles)

- **Mar 2018**: CesiumTiles – locally developed Java function implements additional features of the point cloud 3dTile

- **June 2018**: VisageTiles – lambda function to generate 3D tiles on-the-fly and render them in Cesium. Multiple VisageTiles lambdas can be invoked in parallel

<table>
<thead>
<tr>
<th></th>
<th>py3dtiles</th>
<th>CesiumTiles</th>
<th>VisageTiles</th>
</tr>
</thead>
<tbody>
<tr>
<td># points</td>
<td>3.9 million</td>
<td>3.9 million</td>
<td>3.9 million</td>
</tr>
<tr>
<td># tiles</td>
<td>2500</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Time to create</td>
<td>13+ min</td>
<td>11 sec</td>
<td>6 sec total elapsed time</td>
</tr>
<tr>
<td>Time to display</td>
<td>30+ min</td>
<td>&lt;1 sec</td>
<td></td>
</tr>
</tbody>
</table>

**Note**: time to create 3D tiles is related to number of points, while time to display depends more on number of tiles
Experiments in Point Cloud Density

1. Denser point clouds can appear opaque
2. Fewer points can be rendered more quickly, and provide for better user interface performance
3. Sparser point clouds provide some transparency, allowing users to see into the data.

Cylinders with interior points
Cylinders with exterior points only
Sparser point cloud at high zoom level
VISAGE is a serverless web application hosted on an AWS S3 bucket. The VISAGE API is developed in .NET Core using AWS API Gateway and AWS Lambda functions.

Cesium JavaScript library is used for data and map visualization

A layer manager, to allow users to manipulate datasets loaded in Cesium, is being developed using the Angular framework. The web application also allows for mouse-over data interrogation.

Most data is in point cloud format from the 3D tiles specification

Simpler datasets, such as navigation data and flight paths, are visualized using CZML, a JSON-based format designed specifically for use with Cesium.
User Interface – Merged GPM and Ground Radar

Data Interrogation
User Interface – Merged GPM and Ground Radar
User Interface – NPOL radar
User Interface – ER-2 Flight Track
VISAGE demo: https://youtu.be/B60b_IbAE4A
Issues and Planned Mitigations

• Data processing and volume
  • Balancing processing and performance requirements for on-demand 3D tile
generation against cost of pre-generating and storing 3D tiles for all datasets
    ➢ Continued experimentation with cloud native approaches, such as using
      multiple Lambdas orchestrated with Step Functions to generate many 3D tiles
      in parallel

• Need for custom software to handle diversity of data types
  ➢ Store all data as CSV files in Athena database
    • Common software to generate data visualizations
    • Common tools for basic analytics such as differences and ratios across
data products fields
  ➢ Use OPeNDAP services for data in standard formats and write custom
    handlers for others
Plan Forward – Data Services

• Import additional field campaign data into Athena database
  • Experiment with OPeNDAP as common service to convert many target data types to CSV
  • Experiment with Lambda functions to convert data files to CSV files on demand

• Continue experiments with configuration of parallel Lambda functions for 3D tile generation to increase performance and avoid memory and/or timeout problems

• Continue experiments with composition of 3D tiles to increase performance
  • Increase size of individual points to improve display quality
  • Experiment with styling to resize tiles as user zooms in and out
  • Find optimal combination of tile resolutions for best display quality and interactivity

• Experiment with additional 3D visualization approaches such as volume rendering with ray casting
Plan Forward – UI Features

• Provide better support for interactive data interrogation
  • Help user select specific points to interrogate in the 3D point cloud
  • Determine whether to store data point location and values in 3D tiles, or query source data file

• Investigate methods for temporal alignment of data with diverse time scales and resolutions

• Allow user to access metadata and download data associated with on-screen visualizations

• Prototype cloud-native data services for subsetting and computations on data fields across instruments and platforms
Back-up Slides
Initial Focus Use Case: 3 Dec 2015 from the OLYMPEX Field Campaign

Complex baroclinic system with orographic enhancement; excellent sampling coordination with simultaneous satellite, airborne, & ground-based observations

- Evolving shortwave trough with shallow, developing frontal boundary
- Southerly flow & widespread stratiform precipitation with embedded variability
- Stacked aircraft observations within GPM satellite coverage
- Post-frontal behind wind shift later after GPM overpass

Data Available:
- Most ground instruments, including:
  - Radars: NPOL, KLGX, D3R, DOW
  - Disdrometers, gauges, profilers, soundings
- ER-2, DC-8, and Citation aircraft:
  - AMPR, CRS, APR-3, CoSMIR, microphysics probes, etc
- GPM overpass at 15:22 UTC
- Both GMI & DPR swaths
- Select SIMBA columns
- DPR and ground radar match-ups
- Select WRF model subsets

3 Dec NPOL reflectivity, middle panel shows GPM DPR and GMI overpass swaths (bold line is nadir track) and select SIMBA column locations

Locations of stacked ER-2, DC-8, and Citation flights within GPM DPR Ku-band swath for ~1 h centered on the GPM overpass time, DPR Ku reflectivity shown

Relative nominal aircraft altitudes shown on cross section of DPR Ku reflectivity

Houze et al. (2017)
Targeted Data Products: GPM Ground Validation Archive

GPM GV data are archived at the GHRC DAAC. Dataset publication is ongoing. This collection includes:

- Series of field campaigns collecting detailed measurements of precipitation and related physical processes in a variety of diverse metrological regimes
- Ground and airborne precipitation datasets supporting validation of satellite-based precipitation retrieval algorithms
- Related extended observations from additional sites
Targeted Data Products: GPM Validation Network (VN)

- Compares data from satellite radars (GPM DPR, TRMM PR) and microwave imagers (GMI, TMI and others) to ground-based scanning weather radar observations
- Subsets satellite and ground radar (GR) data for coincident observations of precipitation
- Generates vertical profiles with matching coincident DPR and GR data for precipitation events

Example of DPR gate averaging at GR sweep intersections. Shaded areas show individual DPR gates intersecting the vertical extent of two GR sweeps (dashed) at different elevation angles. The reflectivity values of the individual DPR gates are averaged over the vertical extent of the GR sweeps, resulting in two matching volumes for the single DPR ray shown in this case.
Targeted Data Products: SIMBA
System for Integrating Multi-platform data to Build the Atmospheric column

- Higher-level data product, fusing GPM satellite and ground-based observations into a gridded atmospheric column data file
- Interpolates and/or resamples observations from various scales to set data into a common, user-specified 3D grid
- Encodes observations from diverse data formats into unified netCDF file
- Attributes preserve key operation parameters for each sensor
  - Location, operation mode, timestamps, algorithms, product versions, etc.

Wingo et al. – in review (JTECH)

SIMBA fuses targeted satellite- and ground-based observations collected in various formats & coordinate systems to a single, used-defined 3D column grid

5 April 2018
WRF Model Simulations

- Triple-nested 9, 3, and 1 km grid setup with high-resolution innermost nest over OLYMPEX field site
- Cloud microphysical schemes in WRF model can lead to large uncertainties in the precipitation forecasts
- Field campaign data can help intensively validate cloud microphysical schemes, but collecting/analyzing the large amount of field and model data can be tedious
- VISAGE will help promote more efficient model validation work, which can ultimately help improve precipitation forecasts

(left) WRF model grid setup, (right) model precipitation versus ground-based observations

(left) WRF model reflectivity at lowest level over the Olympic peninsula, (right) WRF model reflectivity (shaded) and rime mass (solid black) along NPOL RHI scan (black dashed)
Entry and Exit TRL

- **Entry TRL:** This new capability is still in the concept phase, **TRL 2**
- **Current TRL:** We are currently researching technical feasibility of 3D Tiles / Point Clouds and Amazon Athena database with representative data, **TRL 3.** However, proposed analytics functions remain at concept phase, **TRL 2,** until Year 2.
- **Exit TRL:** During the two-year VISAGE project, we expect to advance FCX to **TRL 4,** a standalone implementation using full-scale data and providing the functionality needed to address the selected use cases

<table>
<thead>
<tr>
<th>VISAGE component</th>
<th>TRL</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon Cloud services</td>
<td>8</td>
<td>Fully operational with most, but not all, documentation</td>
</tr>
<tr>
<td>Cesium platform</td>
<td>7</td>
<td>Operational but with limited documentation</td>
</tr>
<tr>
<td>Data readers</td>
<td>3</td>
<td>Native data files to CSV, for representative data</td>
</tr>
<tr>
<td>3D tile generators</td>
<td>3</td>
<td>Athena to point cloud, for representative data</td>
</tr>
<tr>
<td>3D vis web app</td>
<td>3</td>
<td>Proof of concept demonstration</td>
</tr>
<tr>
<td>Data interrogation function</td>
<td>3</td>
<td>Proof of concept demonstration</td>
</tr>
<tr>
<td>Basic analytics functions</td>
<td>2</td>
<td>Concept only</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>AIST</td>
<td>Advanced Information Systems Technology</td>
<td></td>
</tr>
<tr>
<td>AMCE</td>
<td>AIST-managed Cloud Environment</td>
<td></td>
</tr>
<tr>
<td>AMS</td>
<td>American Meteorological Society</td>
<td></td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
<td></td>
</tr>
<tr>
<td>AWS</td>
<td>Amazon Web Services</td>
<td></td>
</tr>
<tr>
<td>CMR</td>
<td>Common Metadata Repository</td>
<td></td>
</tr>
<tr>
<td>CSV</td>
<td>Comma Separated Value</td>
<td></td>
</tr>
<tr>
<td>DAAC</td>
<td>Distributed Active Archive Center</td>
<td></td>
</tr>
<tr>
<td>DPR</td>
<td>Dual-frequency Precipitation Radar (on GPM)</td>
<td></td>
</tr>
<tr>
<td>ESDIS</td>
<td>Earth Science Data and Information System</td>
<td></td>
</tr>
<tr>
<td>FCX</td>
<td>Field Campaign Explorer</td>
<td></td>
</tr>
<tr>
<td>GCPEX</td>
<td>GPM Cold-season Precipitation Experiment</td>
<td></td>
</tr>
<tr>
<td>GHRC</td>
<td>Global Hydrology Resource Center</td>
<td></td>
</tr>
<tr>
<td>GMI</td>
<td>GPM Microwave Imager</td>
<td></td>
</tr>
<tr>
<td>GPM</td>
<td>Global Precipitation Monitoring</td>
<td></td>
</tr>
<tr>
<td>GPM GV</td>
<td>GPM Ground Validation</td>
<td></td>
</tr>
<tr>
<td>GR</td>
<td>Ground Radar</td>
<td></td>
</tr>
<tr>
<td>GRA</td>
<td>Graduate Research Assistant</td>
<td></td>
</tr>
<tr>
<td>GSFC</td>
<td>Goddard Space Flight Center</td>
<td></td>
</tr>
<tr>
<td>IPHEX</td>
<td>Integrated Precipitation and Hydrology Experiment</td>
<td></td>
</tr>
<tr>
<td>HAMSR</td>
<td>High Altitude Monolithic microwave integrated circuit (MMIC) Sounding Radiometer</td>
<td></td>
</tr>
<tr>
<td>HS3</td>
<td>Hurricane and Severe Storm Sentinel</td>
<td></td>
</tr>
<tr>
<td>MSFC</td>
<td>Marshall Space Flight Center</td>
<td></td>
</tr>
<tr>
<td>netCDF</td>
<td>network Common Data Form</td>
<td></td>
</tr>
<tr>
<td>NPP</td>
<td>NASA Postdoctoral Program</td>
<td></td>
</tr>
<tr>
<td>OLYMPEX</td>
<td>Olympic Mountain Experiment</td>
<td></td>
</tr>
<tr>
<td>OPeNDAP</td>
<td>Open-source Project for a Network Data Access Protocol</td>
<td></td>
</tr>
<tr>
<td>PPS</td>
<td>Precipitation Processing System</td>
<td></td>
</tr>
<tr>
<td>PR</td>
<td>Precipitation Radar (on TRMM)</td>
<td></td>
</tr>
<tr>
<td>S3</td>
<td>Simple Storage Service</td>
<td></td>
</tr>
<tr>
<td>SIMBA</td>
<td>System for Integrating Multi-platform data to Build the Atmospheric column</td>
<td></td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
<td></td>
</tr>
<tr>
<td>TMI</td>
<td>TRMM Microwave Imager</td>
<td></td>
</tr>
<tr>
<td>TRL</td>
<td>Technology Readiness Level</td>
<td></td>
</tr>
<tr>
<td>TRMM</td>
<td>Tropical Rainfall Measurement Mission</td>
<td></td>
</tr>
<tr>
<td>UAH</td>
<td>The University of Alabama in Huntsville</td>
<td></td>
</tr>
<tr>
<td>UI</td>
<td>User Interface</td>
<td></td>
</tr>
<tr>
<td>UMM</td>
<td>Unified Metadata Model</td>
<td></td>
</tr>
<tr>
<td>USRA</td>
<td>Universities Space Research Association</td>
<td></td>
</tr>
<tr>
<td>VISAGE</td>
<td>Visualization for Integrated Satellite, Airborne and Ground-based data Exploration</td>
<td></td>
</tr>
<tr>
<td>VN</td>
<td>Validation Network (for GPM)</td>
<td></td>
</tr>
<tr>
<td>WFF</td>
<td>Wallops Flight Facility</td>
<td></td>
</tr>
<tr>
<td>WRF</td>
<td>Weather Research and Forecasting</td>
<td></td>
</tr>
</tbody>
</table>