**Giovanni in the Cloud: Earth Science Data Exploration in Amazon Web Services**

**Maksym Petrenko**, Mahabal Hegde, Christine Smit, Hailing Zhang, Paul Pilone, Andrey Zasorin, Long Pham

NASA/Goddard EARTH SCIENCES DATA and INFORMATION SERVICES CENTER (GES DISC)

**What is Giovanni?**

Giovanni ([https://giovanni.gsfc.nasa.gov](https://giovanni.gsfc.nasa.gov)) is an online tool for exploration of geo-spatial data with:

- Twenty-two (22) analysis and visualization services at the click of a button
- Access to over 1600 data variables
- Persistent URLs for sharing data and visualizations

**Notable Features**

- Single page application for specifying service parameters, navigating and manipulating results
- Rapid exploration of geo-spatial data in time and space
- Serves broad spectrum of users from students to subject matter experts

**Pain Points**

- Emphasis on feature set rather than reliability and performance; the two key pillars of a well architected framework
- Victim of its own success; unable to meet spikes in demand during training, and "seasonal" events such as conferences and end of academic terms
- Increased demand on resources due to higher resolution data and user demand for data statistics

**Leveraging Cloud**

- Server-less architecture: AWS-managed solutions for services where possible.
  - AWS API Gateway for service endpoints
  - AWS Lambda, Simple Queueing Service (SQS), Simple Notification Service (SNS) for triggering request processing
  - AWS Simple Storage Service (S3) for webhosting and data storage
  - AWS Elastic MapReduce (EMR) for cluster computing
  - AWS Elastic Compute Cloud (EC2) for general computing (Example: Web Mapping Service)
- Microservices: each operation is an independent service, making chaining of services feasible
- OpenAPI based service specifications: enables language agnostic service definition
- Auto-scaling: to meet demand spikes and compute-intensive services
- Use of Apache Parquet, a columnar and open-source agnostic service
- Built-in synergistic evolution: as AWS evolves in hardware and software, Giovanni benefits

**Architecture**

**Prototyped Features**

- Services
  - Time-averaged map: most popular service in Giovanni
  - Area-averaged time series: most resource intensive service in Giovanni
- Plot-centric instead of variable-centric user interface: users add data to plots simplifying user experience

**Performance Analysis**

<table>
<thead>
<tr>
<th>Feature</th>
<th>On-premises Version</th>
<th>Cloud Version (estimated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Cache Size</td>
<td>30 TB</td>
<td>24.8 TB</td>
</tr>
<tr>
<td>(Approximately 1600 variables)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Cost Analysis**

<table>
<thead>
<tr>
<th>Feature</th>
<th>Quantity</th>
<th>Cost/Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage</td>
<td>25 TB</td>
<td>$600 per month</td>
</tr>
<tr>
<td>Compute</td>
<td>2500 requests/day</td>
<td>$1600 per month</td>
</tr>
</tbody>
</table>

**See also**

IN31A-0068: Use of Schema on Read in Earth Science Data Archives
IN41B-0039: The Value of Data and Metadata Standardization for Interoperability in Giovanni