

Giovanni



NASA GES DISC Giovanni Current and Future

Jennifer Wei, A. Li, M. Hegde, H. Zhang, C. Smit, L. Pham, D. Meyer

NASA Goddard Earth Sciences Data and Information Services Center (GES DISC)

<https://disc.gsfc.nasa.gov/>

<https://giovanni.gsfc.nasa.gov/>



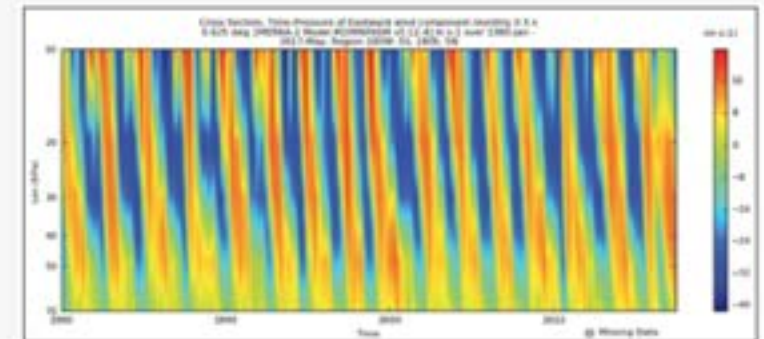
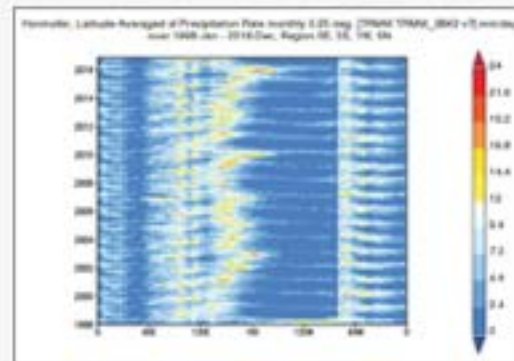
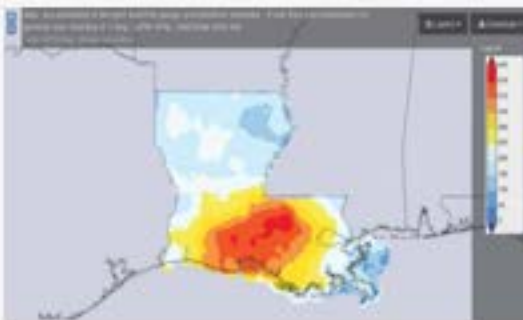
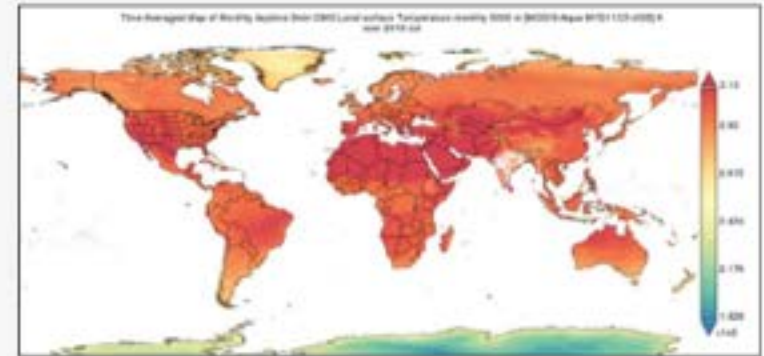
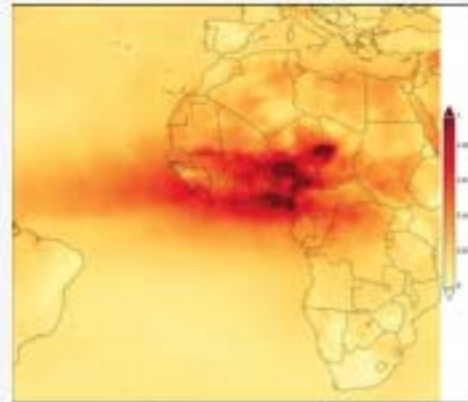
GES DISC logo

2019 AOGS IG20-A011



Giovanni the Bridge Between Data and Science

<https://giovanni.gsfc.nasa.gov/>





Outline



Introduction

**Big data
archive -
NASA GES
DISC
Big data
handler -
Giovanni**



Giovanni Communities

**Registered
Users
Peer-
Reviewed
Publication
User
inquiries**



Giovanni Improvements

**Scientific
Challenges
& Solutions
Engineering
Challenges
& Solutions**

Current

Future



The GES DISC - Big Data Holding



GES DISC
Archives, Composites, Tools & Data Codes, and Data Services

Welcome to GES DISC

If you are new to Goddard Earth Sciences Data and Information Services Center, consider taking a brief tour to get an overview of the site features that will allow you to browse, search, access, and learn about our data. Thank!

Don't show again

<https://disc.gsfc.nasa.gov>

Projects & Milestones

Cloud Absorption Pathfinder (CAP) - The Cloud Absorption Pathfinder (CAP) is an archive multi-wavelength, scanning radiometer that can perform several functions including a...
MIPAS-Plus: Mapping Earth System Data Records for Use in Research Environments, is a NASA project, funded through Research Opportunities in...
MIPAS - The Shuttle Solar Backscatter Ultraviolet (SSBUV) needs dataset to Nimbus II (SSBUV) and NOAA SSBUV instruments from eight space...
View All Projects & Milestones

Tools

Search
MIPAS Software
Data Tools for Hydrology
MIPAS
MIPAS Data
MIPAS Data Services
MIPAS and CAP

Resources

Help
Library
FAQ
News
Contact Us

About Us

Who We Are
Using Our Data
Contact Us
User History Group

- ❑ Archives total volume > **2.3 Petabytes** consisting of >**110 million data files** covering >**2500 public and restricted collections, distributes >23 Petabytes**
- ❑ **Multi-disciplinary** data holdings include observations and model data of **atmospheric composition, water/energy cycles, climate variability**
- ❑ These include for example Aqua AIRS, Aura HIRDLS/MLS/OMI-TROPOMI, SORCE, TOMS, TOVS, TRMM-GPM, UARS, LDAS and MERRA/MERRA-2.
- ❑ Through various available tools and services, the GES DISC provides users with **multi-sensor and model visual comparisons** and data access for a number of **projects spanning several disciplines.**



Big Data Handler - Giovanni



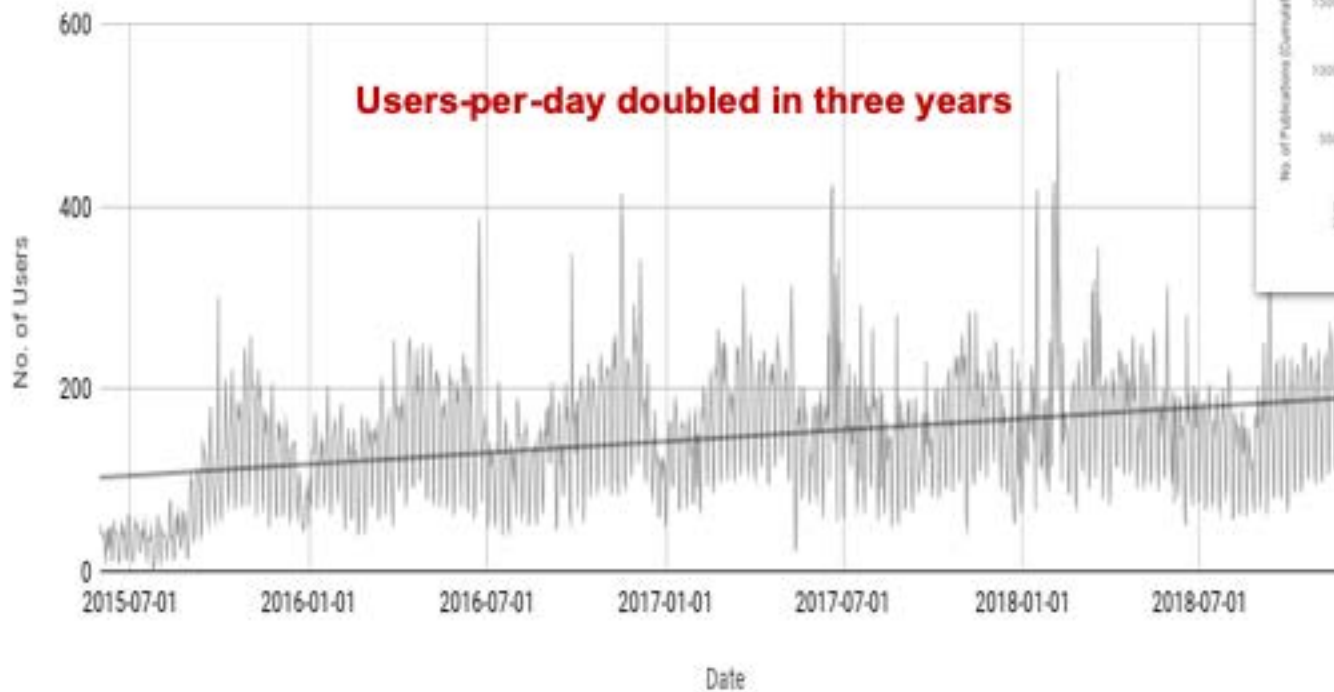
- Provide **quick** online access to data through visualizations **without learning esoteric data formats and metadata representations.**
- Access to about **2000 NASA Earth Science data variables**
- **Twenty-two (22)** analysis and visualization services
- Data downloads in **netCDF, CSV, and GeoTIFF**
- Image downloads in **PNG and KMZ**
- **Shareable URLs**



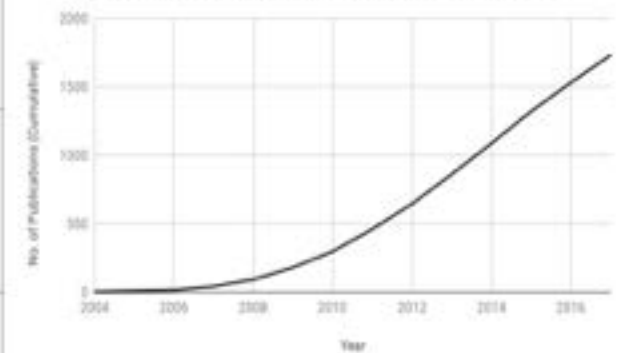
How Community Use Giovanni?



Giovanni Usage



Peer-reviewed publications acknowledging Giovanni



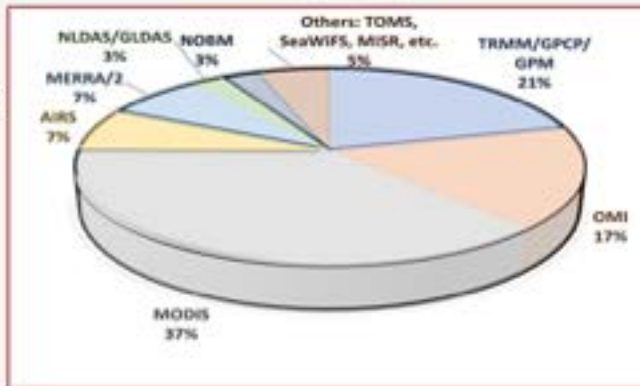
Peer-reviewed publications increased over the years (~200/year for recent years)



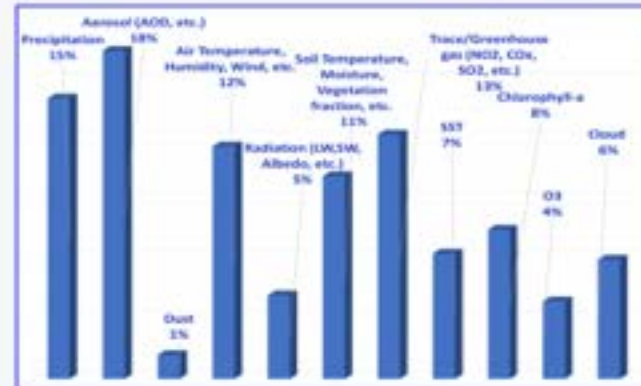
Giovanni Usage – Peer Reviewed Publication (2017)



Studied Missions



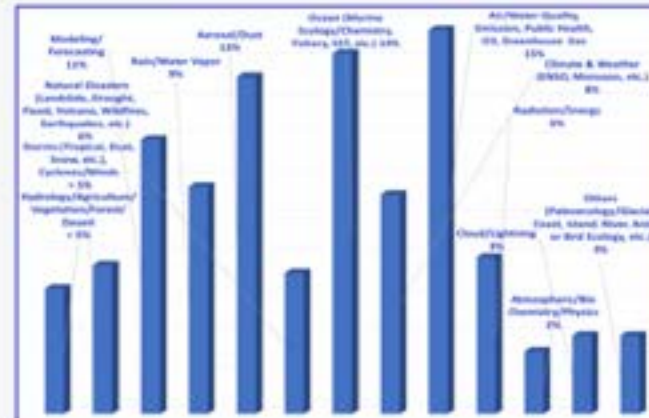
Studied Variables



Studied Regions



Studied Subjects



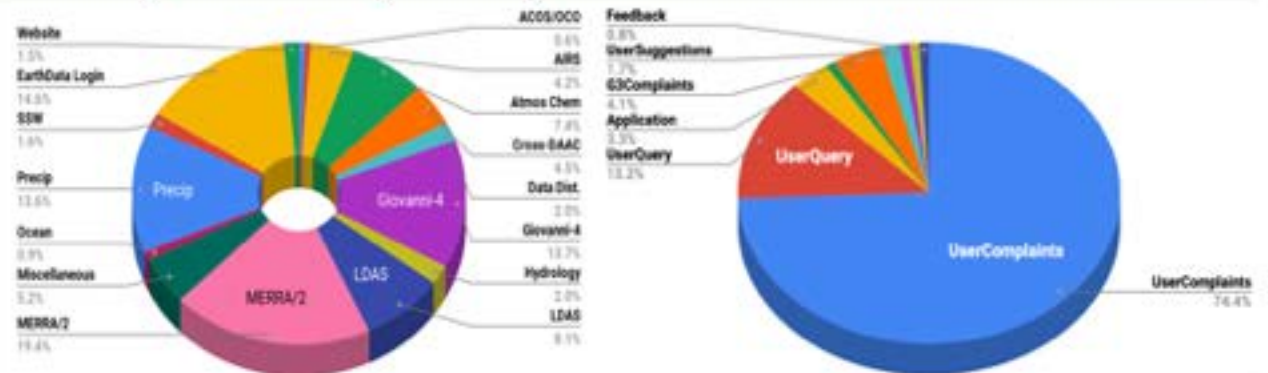
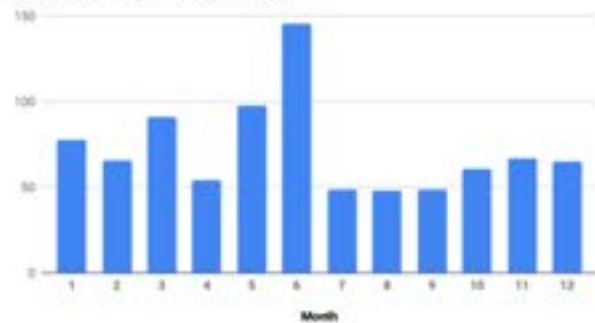


Giovanni Usage – GES DISC User Inquiries (2017)

gsfc-help-disc@lists.nasa.gov



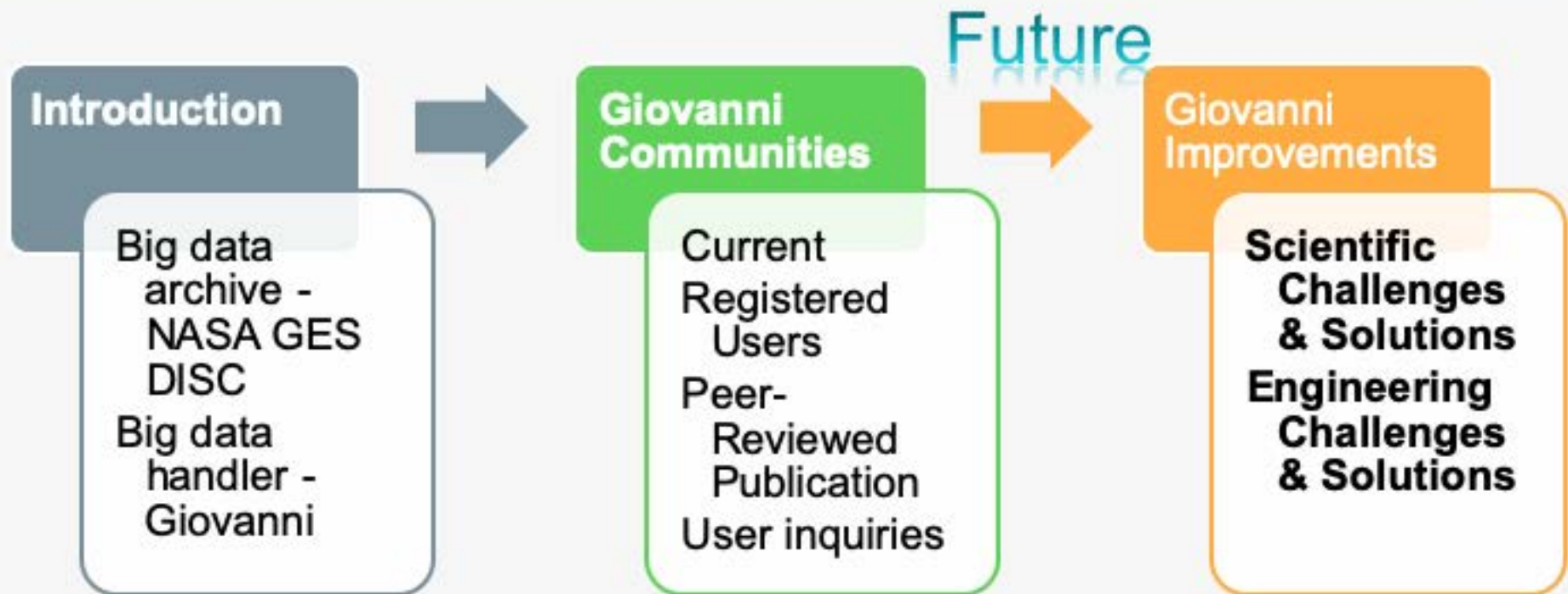
2017 User Inquiry (Total: 873)



- Total over 40K registered users through NASA Earthdata Login
- Over 900 YouTube NASAGESDISC subscribers
- Over 10K Twitter followers on @nasa_gesdisc, over 3K followers on @nasa_giovanni
- Around 800 user inquiries annually, mostly informing some glitches in our services



Outline (Part III)

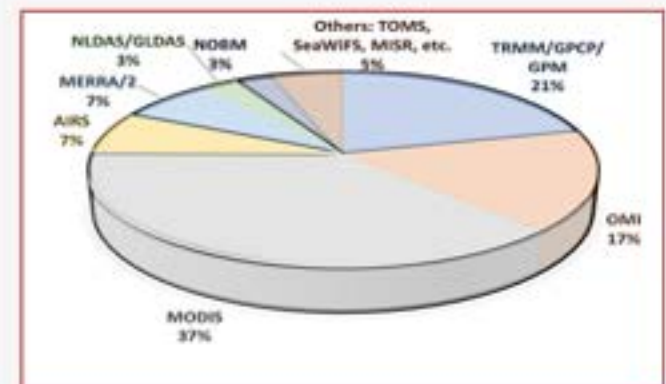
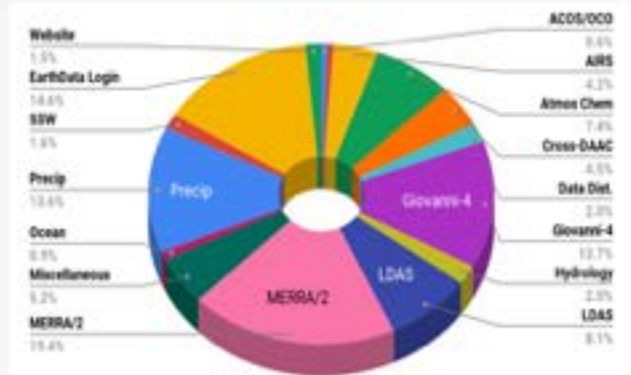




Giovanni Challenges - Scientific Perspective



- Data Content/Discovery
 - How can I find your data from
 - Google/NASA/GESDISC
 - Data reside in multiple DAACs
 - Where can I find related documentation?
- Data Usability/Diverse users
 - Is this dataset best suitable for my research?
 - Which variable should I use in my application?





Possible Solutions for Science Content



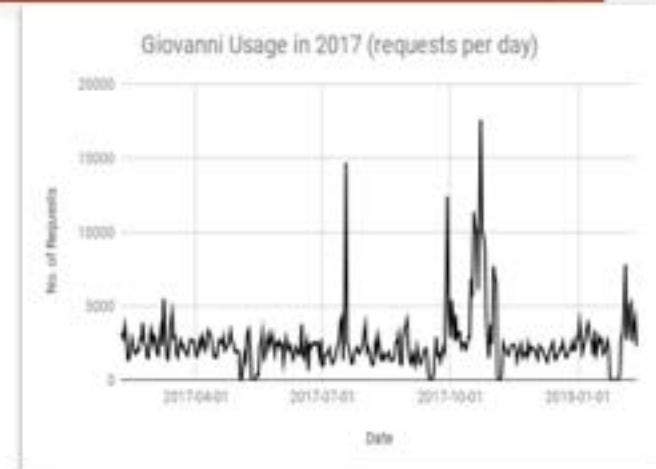
- Enrich the metadata with science content
- Curate data with user community in mind
- Provide tutorials/guidance
- Use machine learning (ML) to mine peer reviewed publication



Giovanni Challenges - Engineering Perspective



- **Data Interoperability/Structure**
 - User request for supporting non GES DISC data (Giovanni without border)
- **Service Performance**
 - Timed out visualization
 - Timed out workflow
 - Usage spikes (during training, conference, etc)
 - User request for resource intensive services
 - Area statistics in area averaged time series
 - Increasing or removing spatio-temporal limits for services





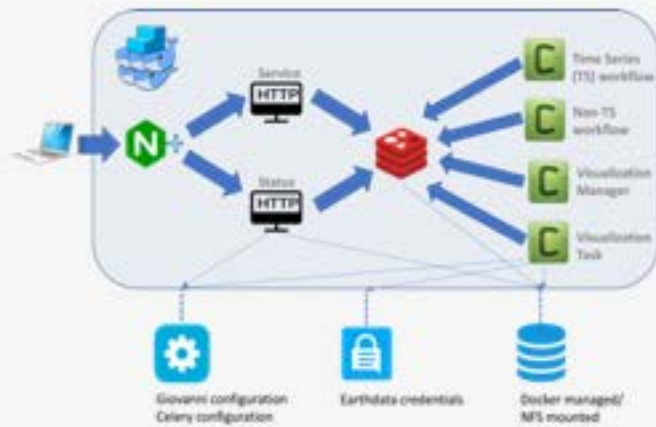
Possible Engineering Solutions



- Queue based Visualization & Workflow
- Cluster Giovanni
- Cloud Giovanni



Clustered Giovanni and Giovanni-in-the-Cloud

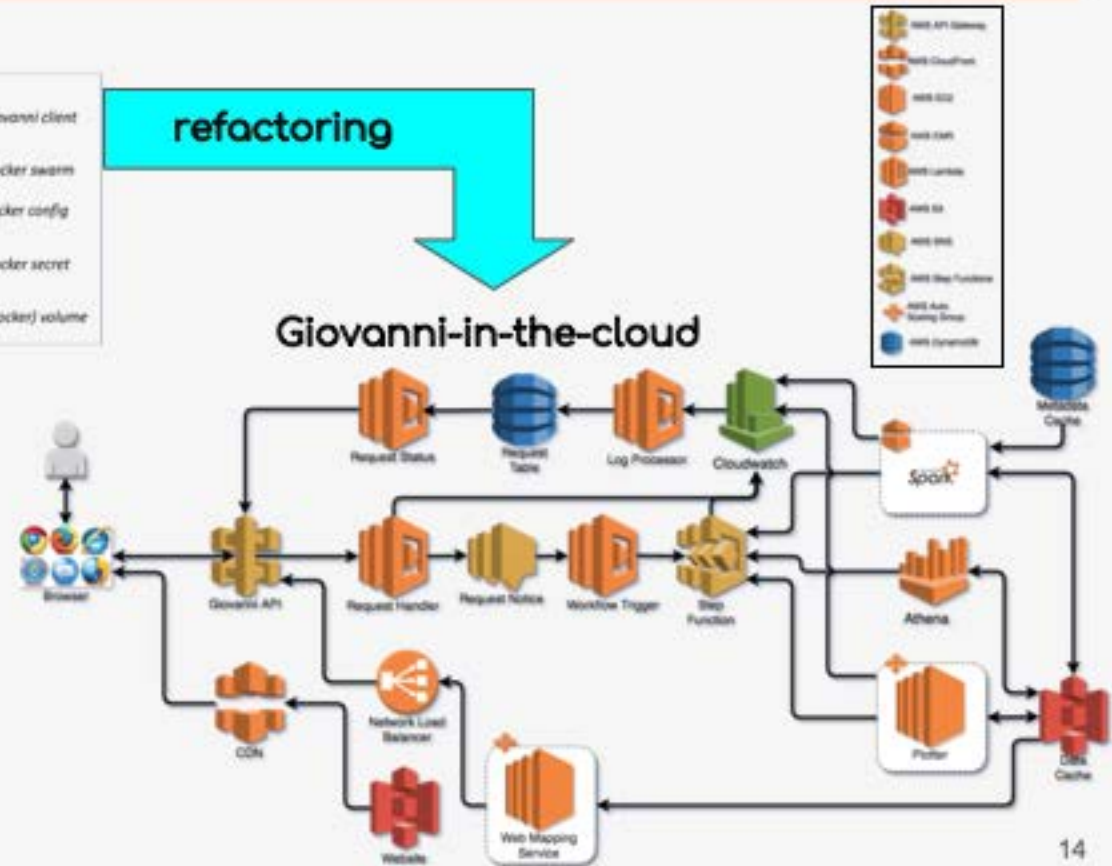


Clustered Giovanni (On Premises)

- Giovanni client
- Docker swarm
- Docker config
- Docker secret
- (Docker) volume



Giovanni-in-the-cloud

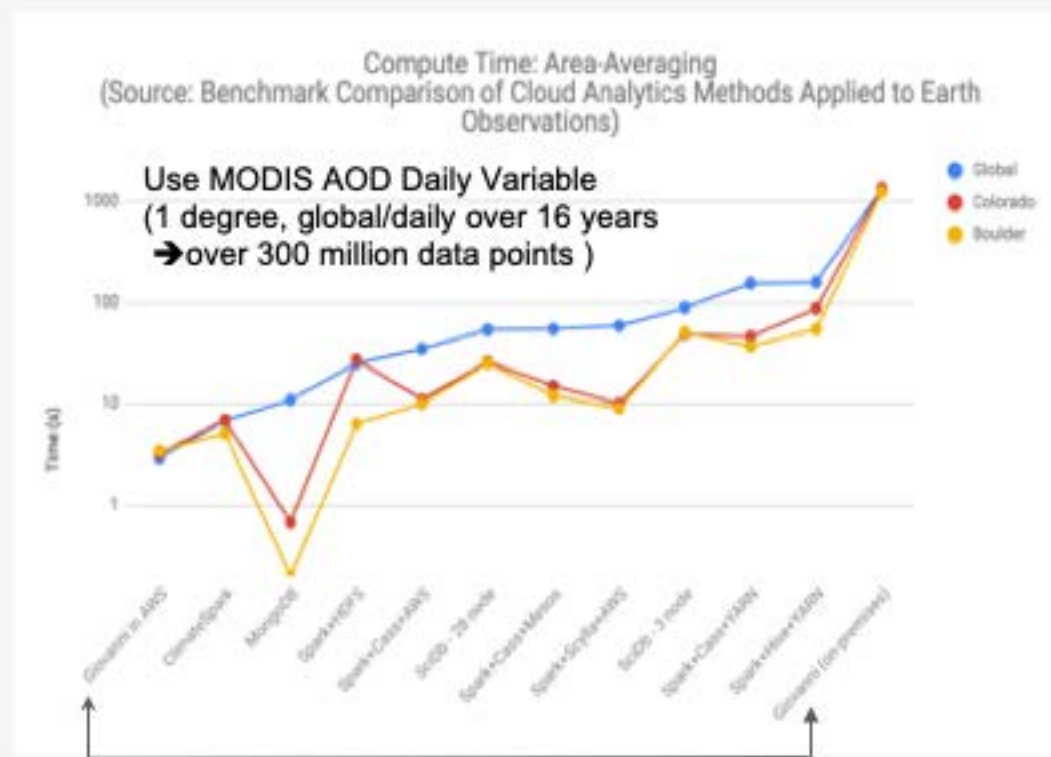


- AWS API Gateway
- AWS CloudFront
- AWS S3
- AWS IAM
- AWS Lambda
- AWS CloudWatch
- AWS Step Functions
- AWS Amazon ElastiCache
- AWS Amazon Athena
- AWS Amazon EC2
- AWS Amazon DynamoDB

Powered by Giovanni team



Giovanni in the Cloud Performance Benchmark



Improvement of cloud version over current Giovanni


Highlights

- ❑ Achieved close to fixed-cost (per request) both in clock time and dollar amount
- ❑ **Extreme reduction** in compute cost: < \$5 per month for a service which is a cause of frequent server overloading.
- ❑ **Portable** Cloud-based data store
- ❑ Certain computationally intensive services saw a **500x** increase in performance (from 1.5 hours to 11 seconds)



Questions?



- Videos on **You Tube**: subscribe “NASAGESDISC”
- Twitter : @NASA_GESDISC, @NASA_Giovanni
- How-To's Data Cookbook:

Download

Case Study

Visualization

- Feedback: 
- Help Desk: gsfc-help-disc@lists.nasa.gov
- Visit us at:

<https://giovanni.gsfc.nasa.gov/>

<https://disc.gsfc.nasa.gov>

