



---

# A Look Under the Hood

How the JPL Tropical Cyclone Information System  
Uses Database Technologies to Present Big Data  
to Users

Brian Knosp, Michael Gangl, Svetla Hristova-Veleva,  
Richard Kim, Peggy Li, Joe Turk, and Quoc Vu

Jet Propulsion Laboratory, California Institute of Technology



# The JPL Tropical Cyclone Information System

---

- The TCIS brings together different types of data related to hurricanes:
  - Satellite Observation Data
  - Airborne Campaign Observation Data
  - Model Forecast Data
  - Simulated Instrument Data from Model Forecasts
- Our goal is to collect data to give users a full picture of what happens before, during, and after a tropical cyclone
  - Ultimately, to aid in future model improvements



# TCIS Data Portals

---

**The TCIS supports different types of data portals:**

1. Global data archive of storm-centric observations *for the creation of robust statistics*
2. Mission/Campaign specific interactive portals *for near real-time conditions and for post-campaign analyses*

# TCIS Near Real Time Simulation Data Selection Capabilities

The screenshot displays the Tropical Cyclone Information System (TCIS) interface, featuring a central 3D globe visualization of a tropical cyclone simulation. The interface includes several key components:

- Header:** Jet Propulsion Laboratory California Institute of Technology logo and navigation links (JPL HOME, EARTH, SOLAR SYSTEM, STARS & GALAXIES, SCIENCE & TECHNOLOGY).
- Page Title:** NORTH ATLANTIC HURRICANE WATCH [NAHW].
- Left Panel (Satellite & Aircraft Data):** A list of data sources with checkboxes, including:
  - Storm Track: BEST TRACK, POUCH TRACK
  - Satellite Data: AIRS, AOT (MODIS), GOES-East, GOES-West, IR, IR COLOR, WVPOR, VIS, Microwave Rain Signatures (99H Gz, 10V Gz, 13H Gz, 19V Gz, 22V Gz, 37CCZ Gz, 37H Gz, 37V Gz, 85V Gz, 85V Gz), Rain Indicator, WRS, SST, TFW, WIND, ASCAT SPEED, ASCAT VECTOR, Reanalysis SPEED, Reanalysis VECTOR, CloudSat
  - Airborne Data: Flight Tracks, PAMSR, Dropsonde
- Right Panel (Model & Simulation):** A control panel for model and simulation data, including a calendar for selecting a date (August 2015) and a table for selecting simulation data:

Model	S	M	T	W	T	F	S
ECMWF	02						
CRF5	02						
19A02-CR18-01	02						
19A02-CR18-02	02						
- Central Visualization:** A 3D globe showing a tropical cyclone simulation with various data overlays (clouds, rain, wind vectors, etc.). A yellow line indicates the storm track. A "MAPSUIT" button is visible at the bottom center.
- Bottom Panel:** Includes a "Save snapshot", "Load snapshot", and "Download snapshot" button, a "Status Bar", and "Grid" options.
- Footer:** Cesium logo and site information: Site Manager: Svetlana M. Ibratova-Volkva, Privacy, Webmaster: Quoc Vu (JPL Clearance: CL#08-2420).



# Data File Retrieval & Database Ingestion

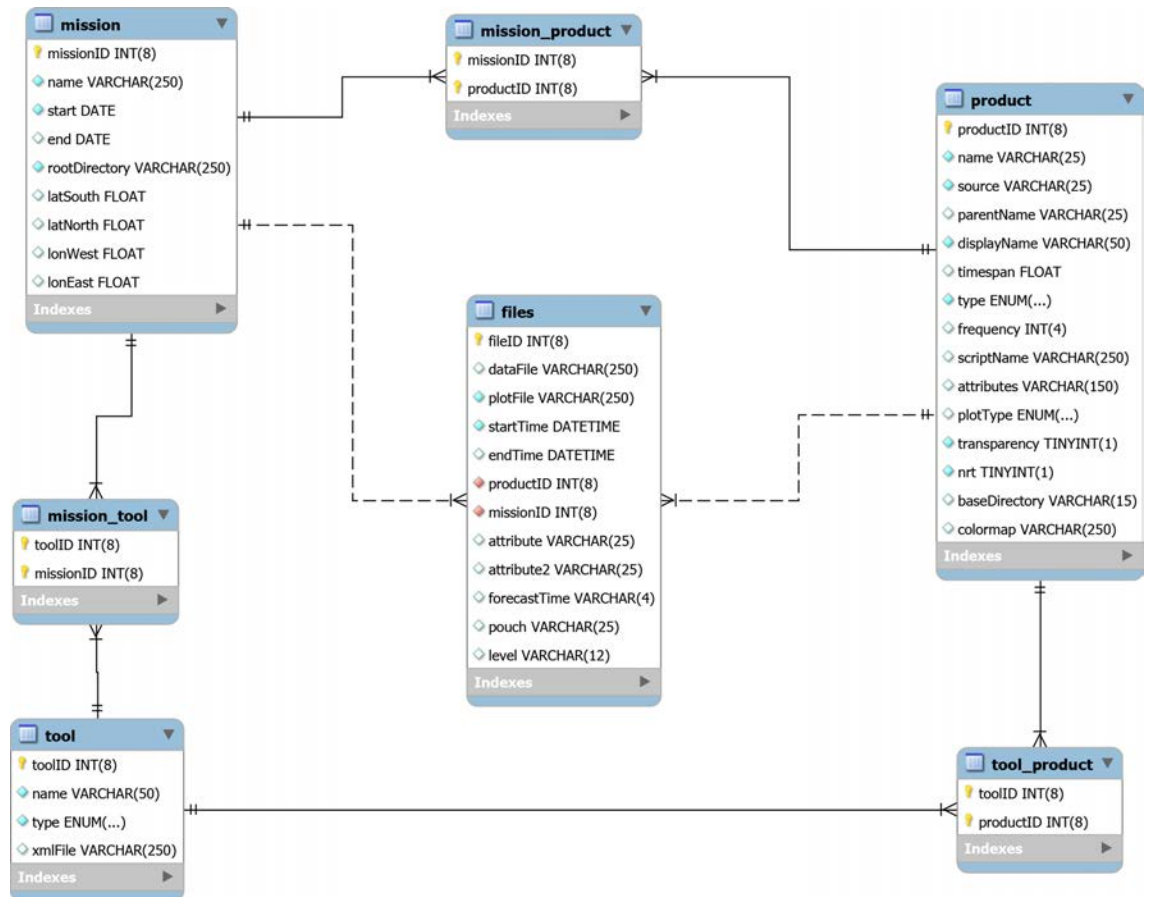
---

- Data is retrieved from data providers (e.g. a NASA DAAC) in hourly, daily, etc. intervals
- Data is subsetted and plotted to the geographical and temporal domains configured by the portal
  - Only the subsetted data and plots are stored (saves on storage)
  - Data does not have to be plotted on-the-fly (saves on user access time)
- Subsetted files and plots are ingested into the database and associated with each other
  - One plot matches with one data file
  - The MySQL database stores the metadata about the files



# NRT Portal Database Configuration

- TCIS NRT portals rely on the `campaigns` MySQL database
- The front end queries the database for all the information required to build the portal and display the plots





# Data Analysis Tools

---

- Some data products can be used for on-the-fly analyses
- The TCIS offers the following analysis tools:
  - Joint Probability Density Function (Joint PDF)
  - Wave Number Analysis
  - Automated Rotational Center Hurricane Eye Retrieval (ARCHER)
  - Data Cross-Section Slicer
  - Subsetting functions

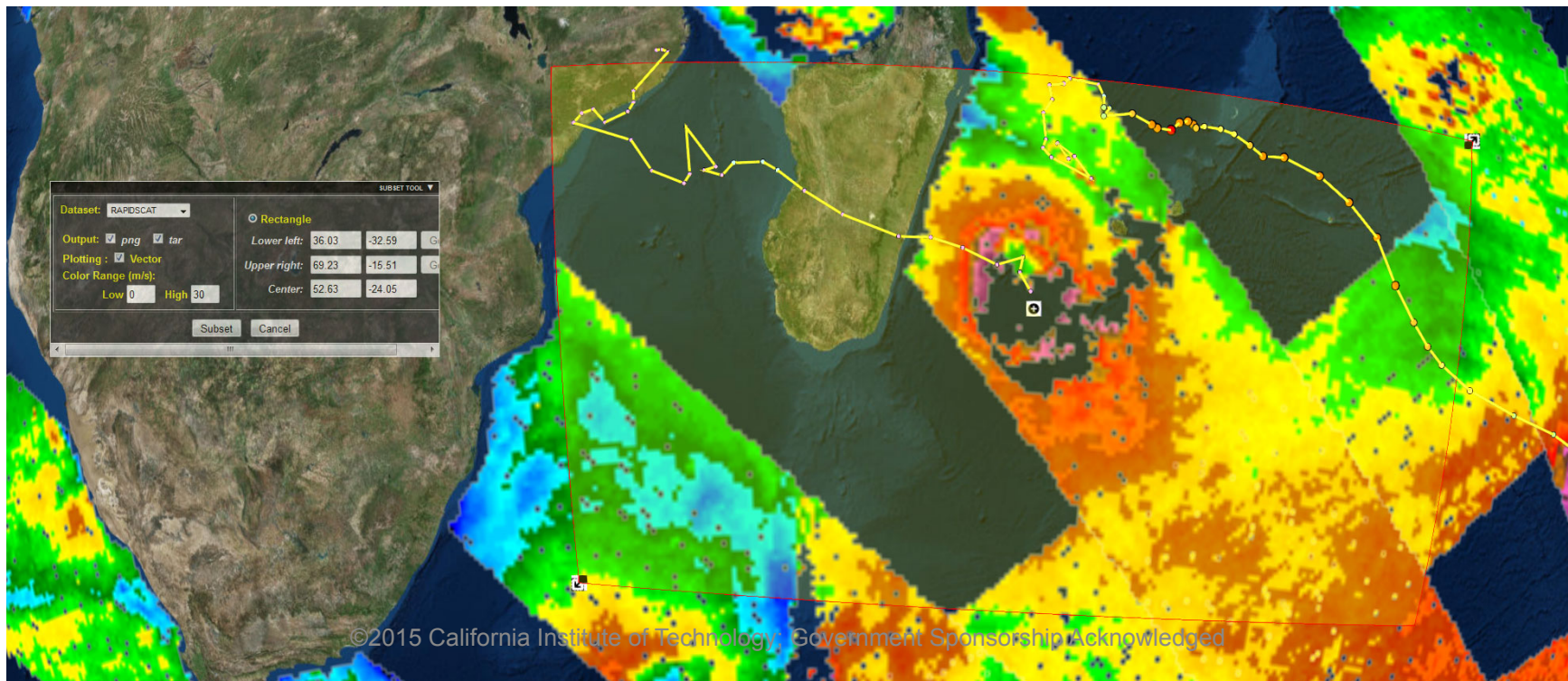


# Selecting Data for Analysis

Users select portions of the plots they see on their screens

**-then-**

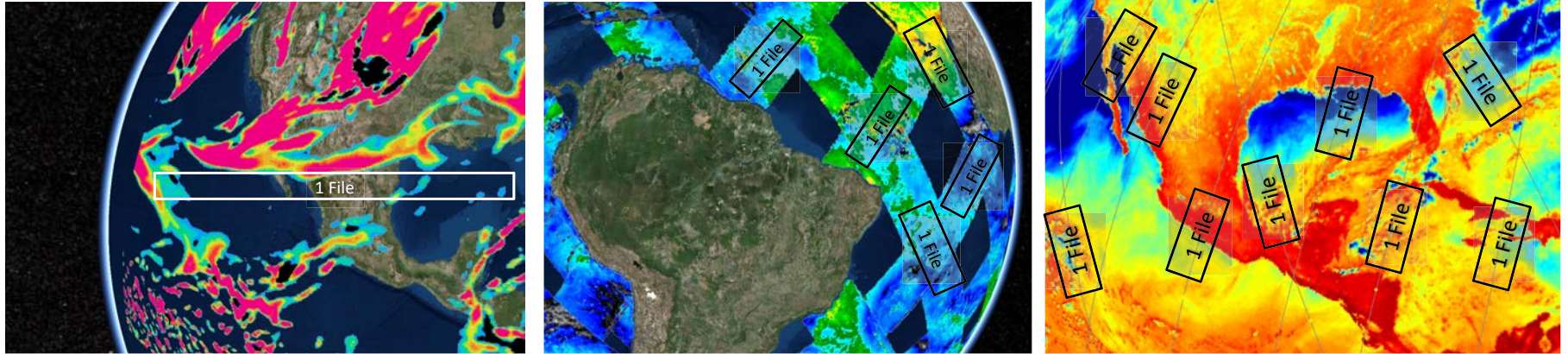
The analysis tool queries the database for the data files that created the selected area, finds them, runs the analysis, and return results







# Swath Data Challenge



Some data arrives to us as one daily file  
***1 plot = 1 data file***

Other data comes to us in multiple (swath) files throughout the day and are made into composite images  
***1 plot = X data files***



# Geospatial Search Implementation

---

- Decided to store swath footprints in the database to make them searchable
  - Goal was to have a system that is both FAST and ACCURATE
- Developed a custom Python script to extract footprints from satellite swaths, and ingested them into the database as polygon objects
- Created a table in MySQL with geospatial keys on the footprint field so that we could have an index of them
- We noticed some issues with query inaccuracies:
  - Python footprint extractor did not always include enough points
  - MySQL v5.5 used the Minimum Bounding Rectangle (MBR) method for searches

***Needed to update our database to improve query speed and accuracy***



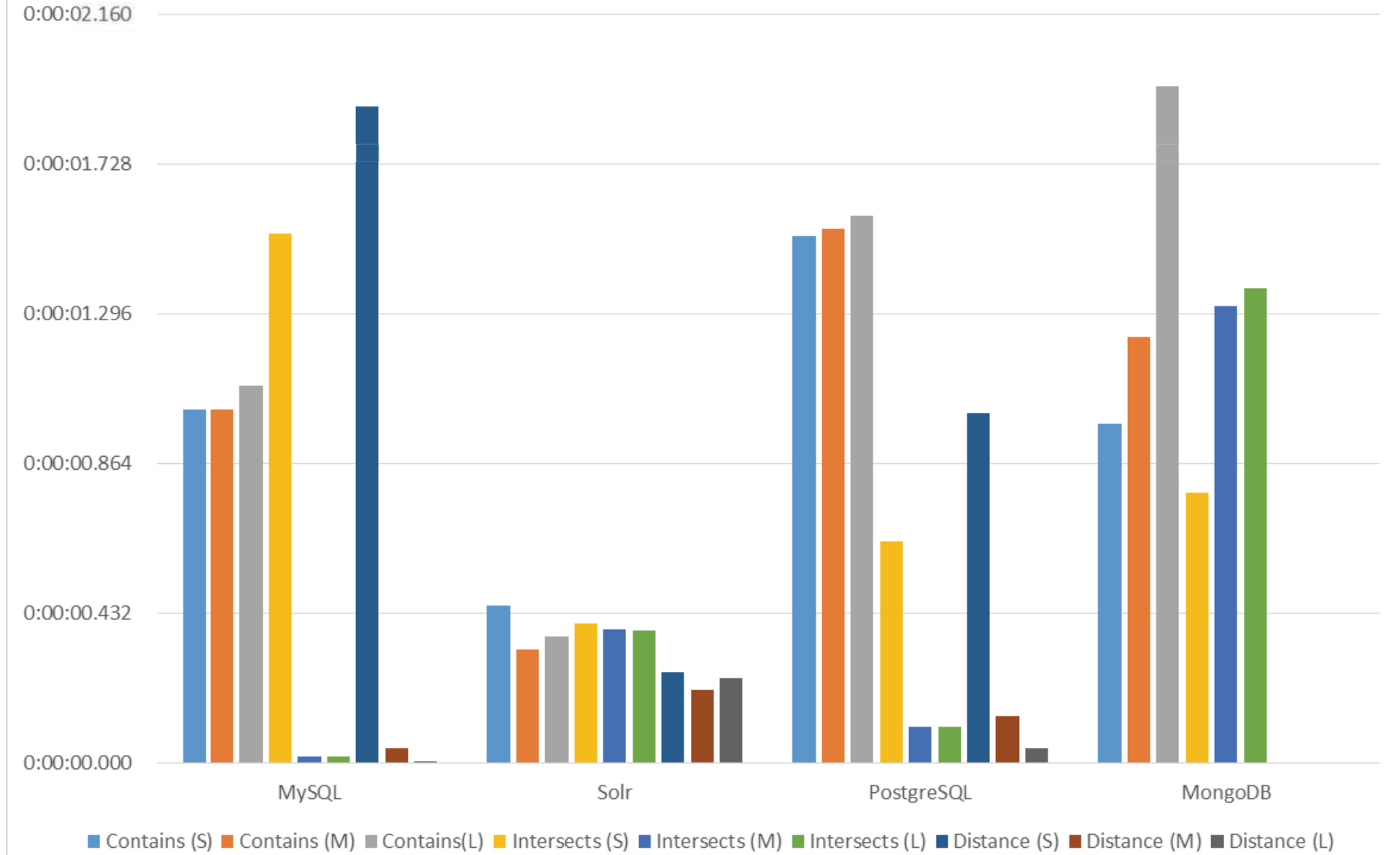
# Database Benchmarking

- Decided to test the geospatial functions of four different DBMSes in the TCIS environment
- Worked with the NASA PO.DAAC team to integrate their L2 Footprint Generator tool to improve footprint accuracy
  - Uses WKT format
  - Easy to configure for different types of data
  - Easy to adjust number of points you need in the footprint

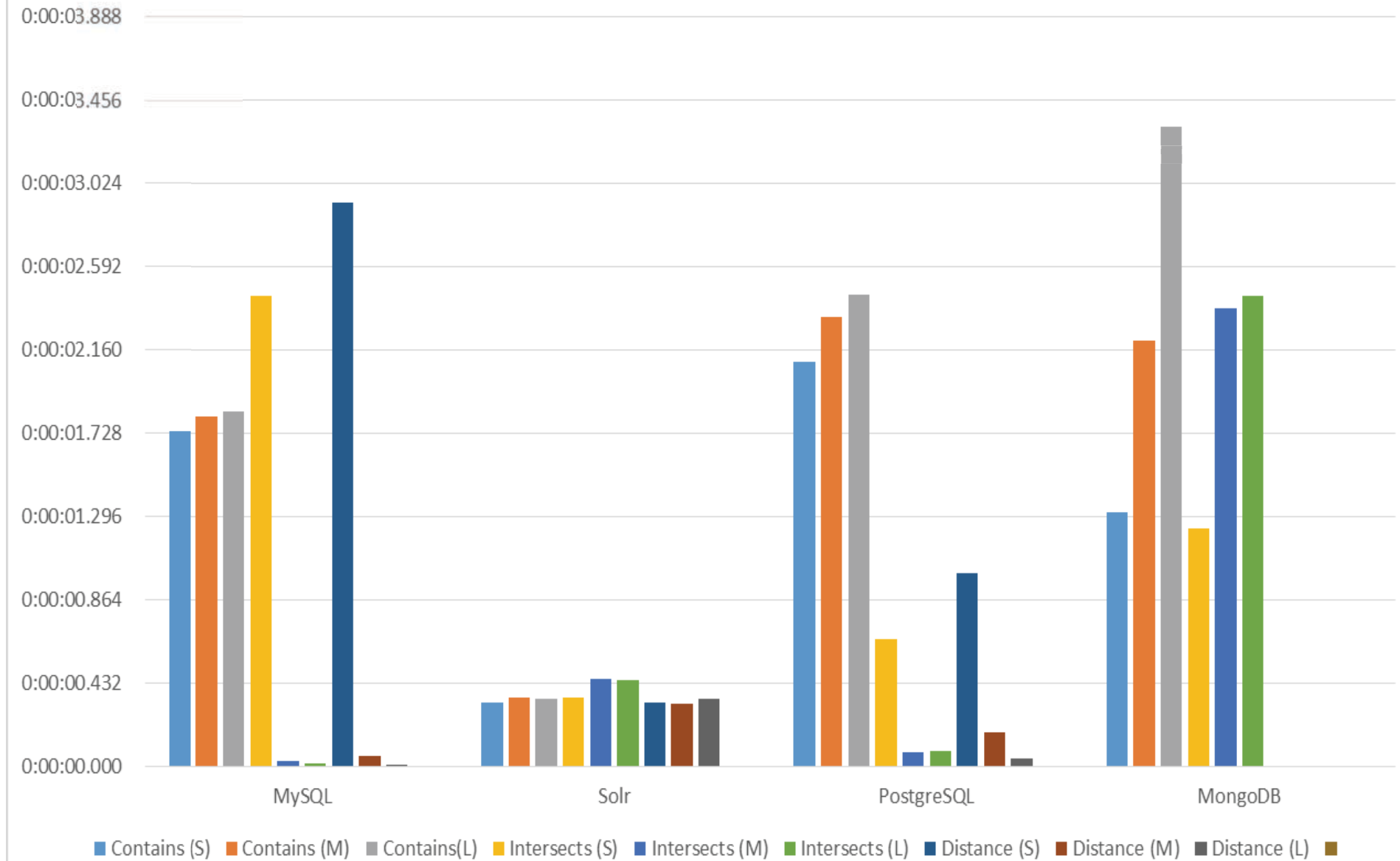
- Created a test data set of 2.1 million polygon object entries to run queries against in each DBMS

	Entries	Disk Space (GB)	Memory (GB)	Data Length (GB)	Index Length (GB)
MySQL	2108505	1.40	0.00	1.11	0.29
Solr	2108505	2.10	1.20	2.1	N/A
PostgreSQL	2108505	2.50	0.00	1.27	1.10
MongoDB	2106886	78.00	78.00	2.93	77.59

## Operation Overview (1 User AVG)



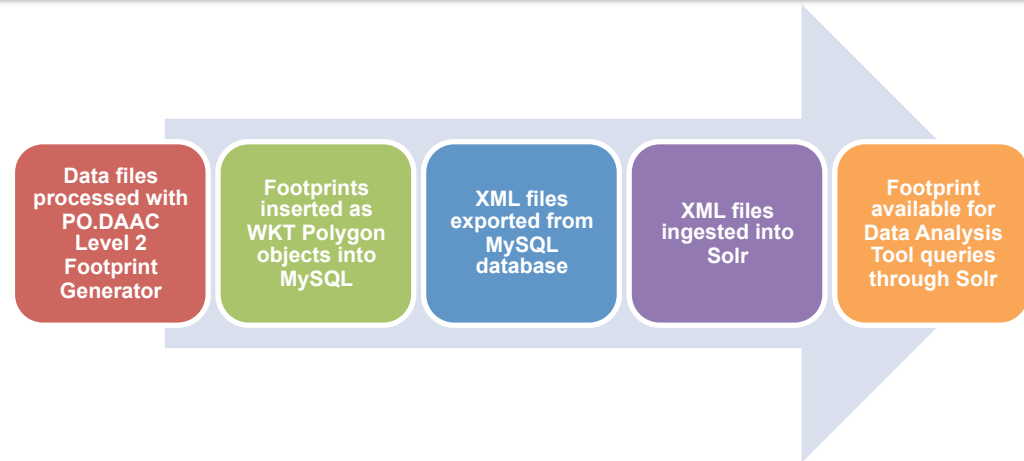
## Operation Overview (5 User AVG)





# Database Architecture Update

- Benchmarking tests showed that Solr was consistently the fastest tool for the data in our environment



- Moved to implement a MySQL + Solr architecture
- Needed fine tuning to make sure we were accurately storing and indexing footprints. This included:
  - Finding an optimal number of footprints points for each data set
  - Finding the optimal distance error percentage in the Solr RPT field that stored the footprints



# Future Portal Expansions

---

- Currently, we have three NRT portals:
  - **GRIP** Field Campaign Portal
  - **HS3** Field Campaign Portal
  - **RapidScat** Mission Portal
- Could be used for upcoming field campaigns
- Could expand portals for other campaigns and missions outside of “hurricane science”



---

## Website

<http://tropicalcyclone.jpl.nasa.gov>

## Contact

E-Mail: [Brian.Knosp@jpl.nasa.gov](mailto:Brian.Knosp@jpl.nasa.gov)

Twitter: [@that\\_knosp](https://twitter.com/that_knosp)