Aloha!

Tapis-CHORDS Integration: Time-Series Data Support in Science Gateway Infrastructure

Sean B. Cleveland, Anagha Jamthe, Smruti Padhy, Je'aime Powell, Joe Stubbs, Michael D. Daniels,

Suzanne A. Pierce and Gwen A. Jacobs

UNIVERSITY OF HAWAI'I

Overview

- Project Background
  - Motivation
    - Hawai ' i EPSCoR - 'Ike Wai Gateway
    - Planet Texas 2050 – (Je'aime Powell up NEXT!)
  - The Challenge
    - Tapis
    - CHORDS
- Implementation
- Performance testing
- Future Work

TACC
TEXAS ADVANCED COMPUTING CENTER

UNIVERSITY OF HAWAI'I
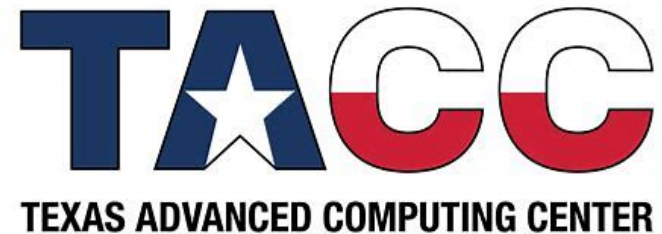
Motivation - 'Ike (knowledge) Wai (water) Gateway

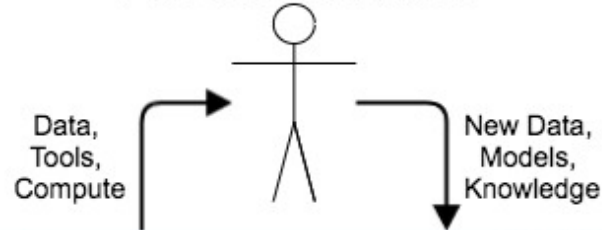Climate Change, Incidents, Growth & Sustainability

- **What is it?**

  A data and modeling platform for Hawai'i volcanic hydro-geology, economic modeling, and decision support. Bringing data, models, computation, analysis and visualization into one place

- **What purpose does it serve?**

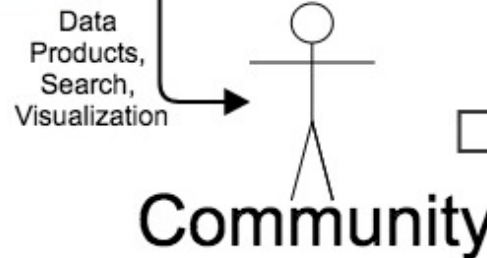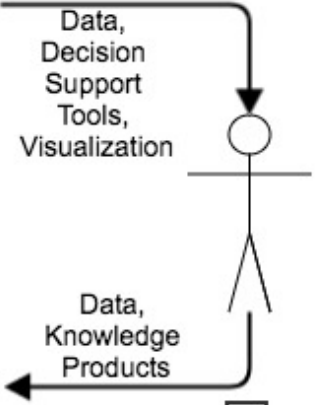 Enables decision and policy makers simpler access to better data and tools
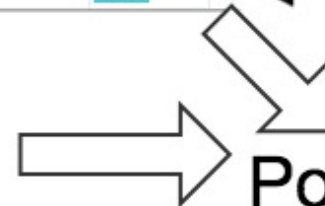
Sensor Data and time-series legacy data (SGD, Chemistry, Precipitation, Well)

The Challenge

- Support:
  - the storage of streaming data from sensors and legacy time-series data

  - advanced annotation of streaming data/products

  - spatial queries to discover this streaming data.

Intelligent Systems Research to Support Geosciences (IS-GEO)

- Hosted Workshop for sensors on the Big Island
- Used the Cloud-Hosted Realtime Data Services for the Geosciences (CHORDS) for hosting the streaming data
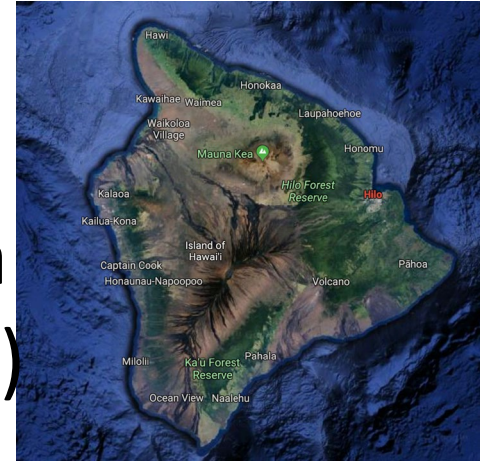
This seemed like something we should leverage…

Our Challenge With Technologies

- Tapis offers almost everything necessary for all the use cases. BUT
    - There was a gap for:
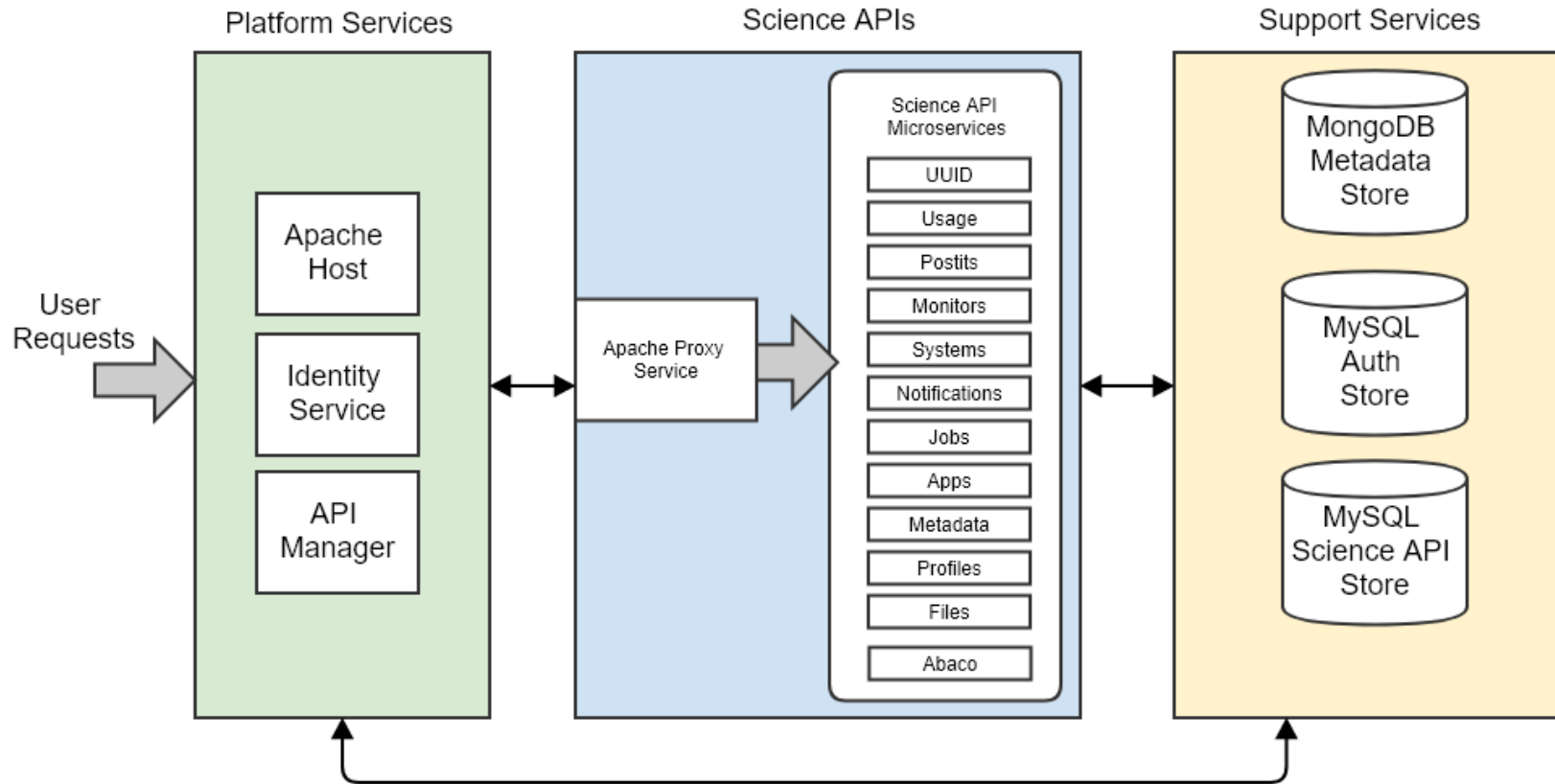   - Granular time-series data support.

- CHORDS support time-series data for instruments/sensors. BUT
    - There was a gap for:
        - Spatial data query support
        - Advanced/custom metadata annotation.
        - Multi-tenancy support.

UNIVERSITY OF HAWAI'I

Tapis Framework
Platform as a Service, REST APIs

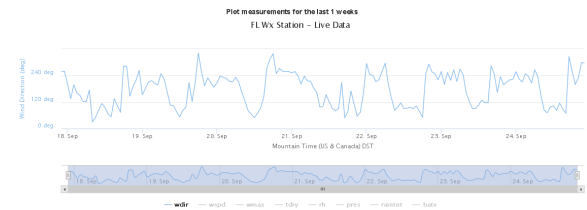

Auth, Data Management, Compute

UNIVERSITY OF HAWAI'I

CHORDS

- Ruby on Rails APIs and web UI
- InfluxDB for time series data
- MySQL for Auth & Metadata
- Optional Grafana visualizations

Our Solutions

Create a Proof of Concept **Streams API** that ties Tapis Metadata services together with CHORDS time-series data support.

Our initial implementation uses a NodeJS API

- Tapis Authentication and Authorization features to secure data and provide multi-tenancy support
- Tapis Metadata for Annotations
- CHORDS for measurement storage
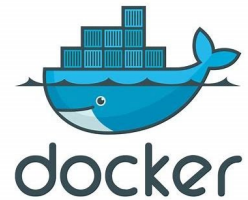
REST API Endpoints

- **Sites** – geographical location hosting 1-n instruments (WGS 84 GeoJSON)
  - **Instruments** – A source of measurements
    - **Variables** – A description of a measurement
      - **Measurements** – Recorded observation(s)
- **Spatial** - Query uses a GeoJSON polygon & returns matching Sites.

All requests require a Tapis API token with sufficient permissions in the inherited hierarchy(site/instr/var)

How does it work?
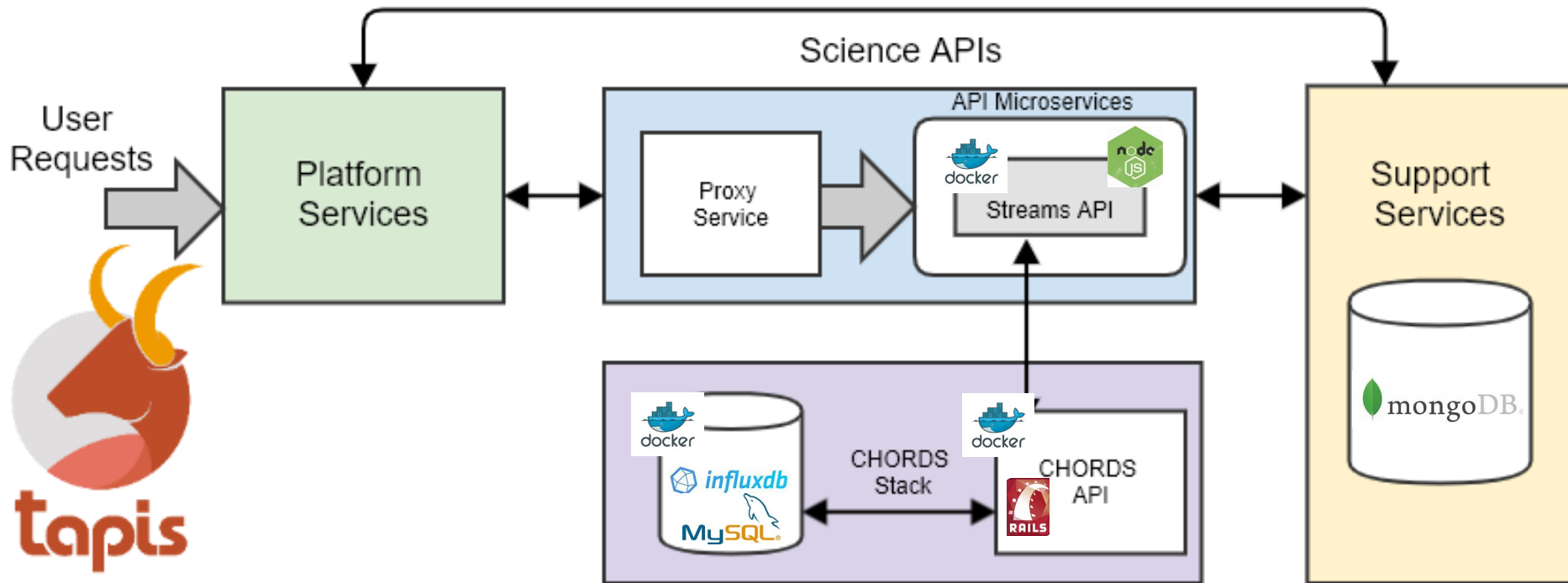
-All metadata(Site/Instrument/Variable) recorded in both Tapis metadata & CHORDS

-Utilizes a CHORDS service account for read/write

-CHORDS UI interface continues to work.

Performance Benchmarking

PERFORMANCE BENCHMARK: DATA INGESTION

Service	Number of POST requests per test run	Data Accuracy (%)	Avg. request response time (seconds)
CHORDS	100	100	0.11
CHORDS	200	100	0.11
CHORDS	500	100	0.12
Streams	100	100	0.65
Streams	200	100	0.6
Streams	500	100	0.6

The discrepancies in the times are related to the additional auth/authorization checks we have added on top of CHORDS (we plan to improve on these)

Performance Benchmarking

PERFORMANCE BENCHMARK:DATA FETCH

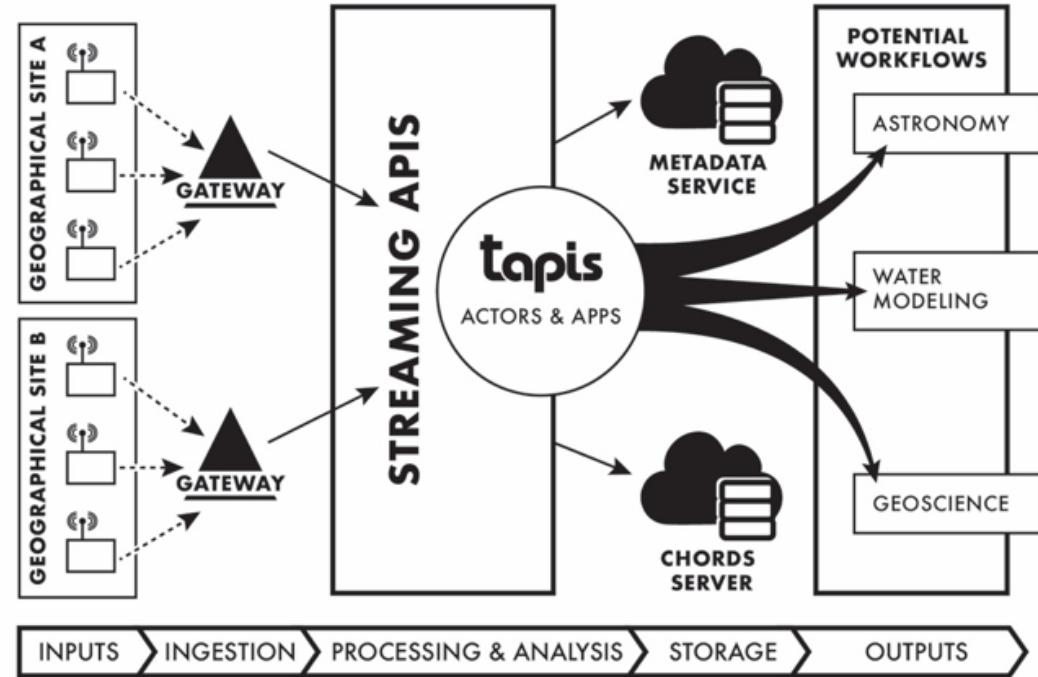| Number of GET requests per test run	Avg. request response time CHORDS (seconds)	Avg. request response time STREAMS(seconds)
100	0.61	0.98
200	0.67	1.0
500	0.64	0.91

Some recent benchmarks for fetching **50K+** measurements clock in around **2.5** second across 500 repeated requests.(We plan to improve on that – we may go directly to InfluxDB in some cases)

Future Work

- Enhance performance of the Streams API
- Production release of the Streams API in Tapis coming in 2020
- Integration of Streams API with data triggers for executing applications and actors for workflow support.

Funding Acknowledgement

Mahalo!

Questions?