REACHING FOR THE CLOUD: ADVANCING THE NATIONAL DMAC ARCHITECTURE

Steering Committee Meeting Kickoff

June 10, 2022

IQOS Association

rps MAKING COMPLEX EASY

AGENDA

1. Background
2. Project Goals and Overview
3. Overview and Progress
4. Roles and Responsibilities
5. Prototypes

Project Background

- **Funding:** FY21 Implementation of the U.S. Integrated Ocean Observing System (IOOS) – Topic Area 2, *Advancing the National Data Management and Cyberinfrastructure System Architecture.*
- **Topic Area 2 Theme:** incorporate advances in HPC, data science, and data modeling into DMAC
- **RPS focus:** opportunities provided by the cloud to improve DMAC and meet a variety of needs from across the IOOS community

Project Goals

- Goal: increase the use of IOOS data and promote connections to other disciplines by lowering the barriers for entry
- Objectives:
  - understanding the current state, challenges, and opportunities
  - Identify state-of-the-art technology and standards that build toward vision
  - developing a technical roadmap for what DMAC should look like in 5-10 years
  - prototypes and demonstration datasets

Year 1: Gather requirements, develop prototypes, draft roadmap

Year 2: Demonstrate Value

Year 3: Final Roadmap and Recommended Architecture

Workplan

- **Task 1**: Establish Steering Committee
- **Task 2**: Conduct RA Assessment
- **Task 3**: Roadmap/Architecture
- **Task 4**: Design and implementation of prototypes
- **Task 5**: Host community webinars
- **Task 6**: Final report and recommendations

Task 1: Establish Steering Committee

- 7-12 members with representation of the RA management, RA DMAC, IOOC agencies, and other community members
- Provide input and guidance on technical implementation, facilitate understanding of the IOOS vision for the cloud and the development of recommendations and best practices.
- Meet quarterly to review the results of the assessment to provide input on the prototypes, and serve as a sounding board and to develop recommendations for next steps

Task 2: RA Assessment

- Virtual assessment of RA interest, potential needs and opportunities for using the cloud;
- Follow up through virtual interviews with RA directors, DMAC coordinators and other representatives
- Summarized in a short document describing use and desired use of the cloud by IOOS regions.
- Resulting in additional guidance for the direction of the project, subsequent prototypes developed, and final recommendations

Task 3: RoadMap/Architecture:

- Y1: Draft roadmap: key concepts, elements, and challenges
  - evaluate current/emerging technologies
  - assessment of cost, performance and scalability, ease of use, maintenance, and adoptability
- Y2: Refined roadmap and proposed architecture
- Y3: Finalized roadmap and proposed architecture for critical components
- Y3: Guidelines for RAs and others to extend DMAC components in line with recommended architecture and directions, while also satisfying certification requirements

Task 4: Prototypes

- Defined using input from the project team and the steering committee
- Engage with broader user groups and software communities > informed by parallel activities
- Focus on both storage of datasets in the cloud and developing serverless functions and library interfaces to act against them
- Demonstrate connections to existing national elements such as the NOAA Open Data Dissemination
- Successful prototypes will be made available to the wider IOOS DMAC community for testing and demonstration.
- Gather cost and usage metrics;

Task 5: Webinars

- Foster awareness and dialog about how the cloud can be used to meet IOOS objectives.
- Topics will be developed to address issues including those identified in the regional assessment, NOAA and IOOS Federal agencies cloud strategies, lessons learned related to the cloud and to showcase and gather feedback on prototype development.

Task 6: Final Report

- Summary of the RA assessment, the outcomes of the prototypes, information on cost, performance and scalability, ease of use, maintenance, adoptability
- A list of opportunities and guidelines for how RAs can utilize the cloud, recommendations for governance and operations, and recommendations for next steps.
- Technology gap analysis highlighting which areas can move forward now with existing technology and which require additional investment and development to proceed.

Project Progress

Outreach and Communication

- Presented project to RA Directors
- IOOS Code Sprint
- Talk at *Cloud-Native Geospatial* (April 2022)
- Regional Association Assessment

Technical Implementation

- Numerous technical brainstorming sessions
- Identified major technical themes and technologies of interest
  - Early vetting of: Zarr, Xarray, Xpublish, STAC, Xcube, FastAPI Kubernetes, TileDB
- Developed high-level, idealized system architecture diagrams
- Discussed software development processes and core values (e.g. modular, language-agnostic, etc.)
- Development of prototype ideas for discussion with the steering committee
- Progressed 4 prototype ideas

Roles and Responsibilities

RPS Team

- Strategy and Vision
- Tools, Interim Solutions, Prototypes, POCs
- Status Reporting and Finances
- Final Report and Recommendation

IOOS Program Office

- Contract management
- Lead IOOS and DMAC strategy and vision
- Communicate NOAA cloud strategy
- Represent IOOS on Steering Committee

IOOS Association

- Conducting Assessments
- Organizing Steering Committee
- Organizing Webinars and Workshops
- Assisting with Final Report

Steering Committee (RPS, RA Management, RA Data Managers, IOOS Program Office, 3rd parties)

- Provide input on project direction / Prioritize prototypes
- Communicate external considerations and influences
- Guide IOOS/RA cloud strategy

Regional Associations/Stakeholders

- Participate in assessment of requirements & needs
- Share cloud aspirations and challenges
- Attend workshops and webinars
- Test prototypes and provide feedback

User Groups / Software Communities

- Test and use mature prototypes
- Participate in open-source collaboration

Questions?

PROTOTYPE DISCCUSSION

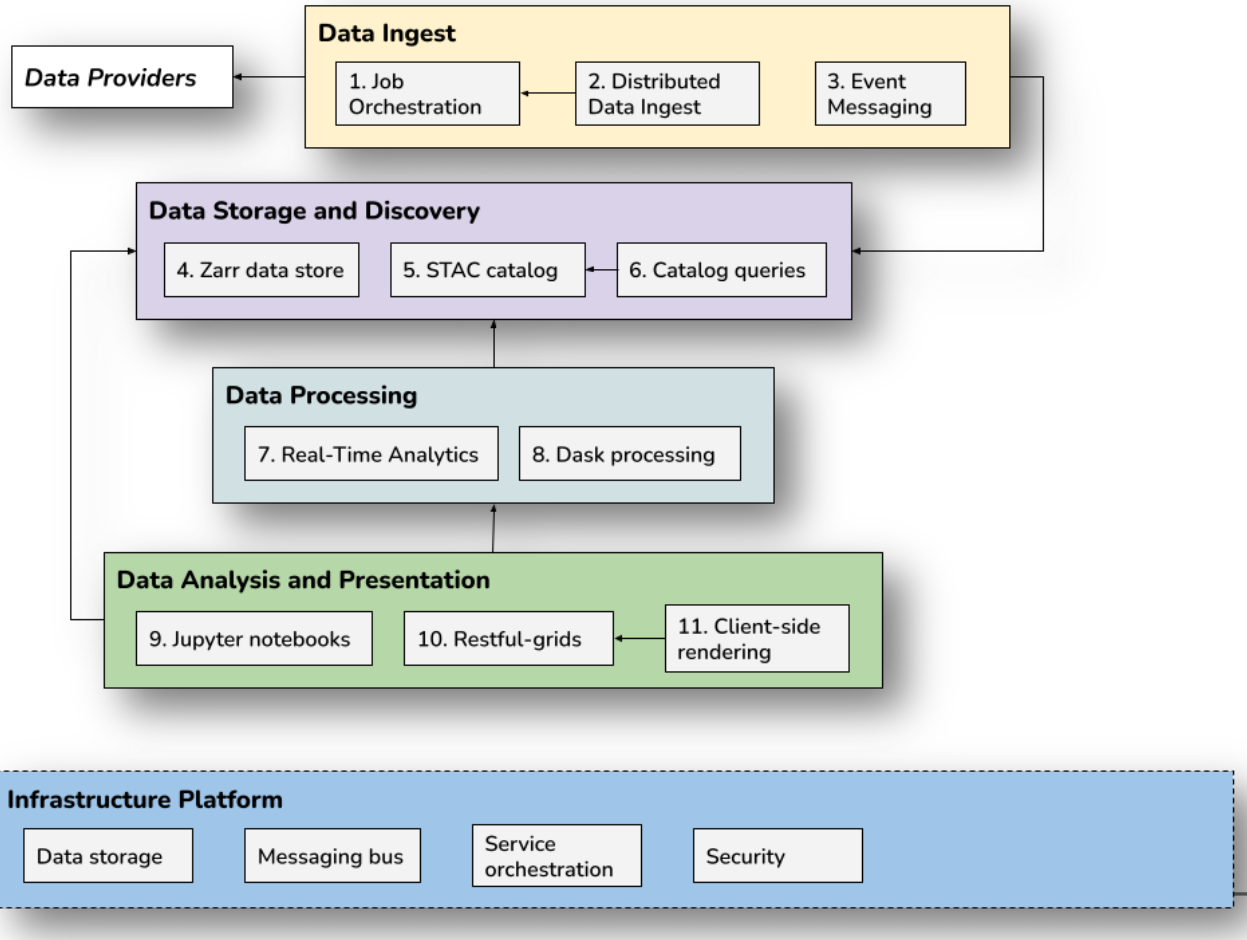
Prototype List

1. Job Orchestration
   - Visibility into data flow status, high reliability, one standard platform for managing data flows
2. Distributed Data Ingest
   - Faster system-wide data availability, independent data feeds, contained failures, easier to add more datsets, support for many data types
3. Event Messaging
   - Faster data availability for downstream products, easier to add on new services without system impacts
4. Zarr Data Store
   - Faster, scalable data access
5. STAC Catalog
   - Interoperability with other STAC Catalogs. External users can easily discover and develop against NOAA data on the cloud, and we can also point to other external STAC catalogs for easier data fusion.

Prototype List

6. Catalog Queries
   - Ability to find datasets by searching and filtering metadata using queries
7. Real-Time Analytics
   - Faster availability of critical information, run analytics quickly without waiting for stored & catalogued data
8. Dask Processing
   - Capability to run large computations against the data store on the cloud
9. Jupyter Notebooks
   - Provide users a cloud-based interface for data analysis, bring processing to the data (no data downloads)
10. Restful-grids
    - Simplified data dissemination through open standards, removes need for specialized tooling to interact with model data, easier for downstream developers to create products using model data
11. Client-side data rendering
    - Better user experience, faster data delivery

Prototype Relationship Diagram



Prototype Highlights

- Zarr Data Store
  - Converted several sample datasets to zarr and hosted on Amazon S3 for testing
- Spatio-Temporal Asset Catalog (STAC)
  - Created a sample STAC catalog for zarr datasets
  - Developed architecture diagram for STAC service implementation
  - Still need work in STAC Extensions to capture important NOAA metadata
- Jupyter Notebooks
  - Began installation of JupyterHub and tested various user access methods
- Restful-grids
  - Started development at the 2022 IOOS Code Sprint
  - Created data retrieval and Web Map Tiling Service demonstration
- Client-side Data Rendering
  - Demonstrated rendering of zarr data in web browser using WebGL

Demonstration

- Tiled data is loaded using web services reading data in zarr format
- Data is rendered in web browser, allowing for immediate customizations
- Requires further investigation to assess scalability and interoperability with other use-cases
- Additional prototypes will support backend services necessary to extend these capabilities to more data products