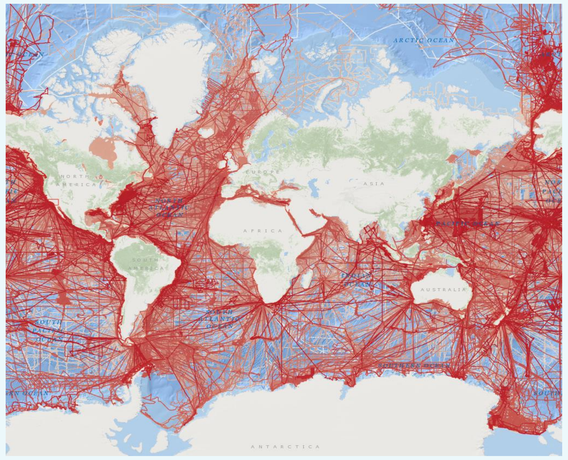
Trackline Geophysics in the Cloud

Jordan Schweizer (CIRES/NOAA NCEI), Brian Meyer (NOAA NCEI), Chris Slater (CIRES/NOAA NCEI), Peemin Chen(CIRES/NOAA NCEI)

Introduction

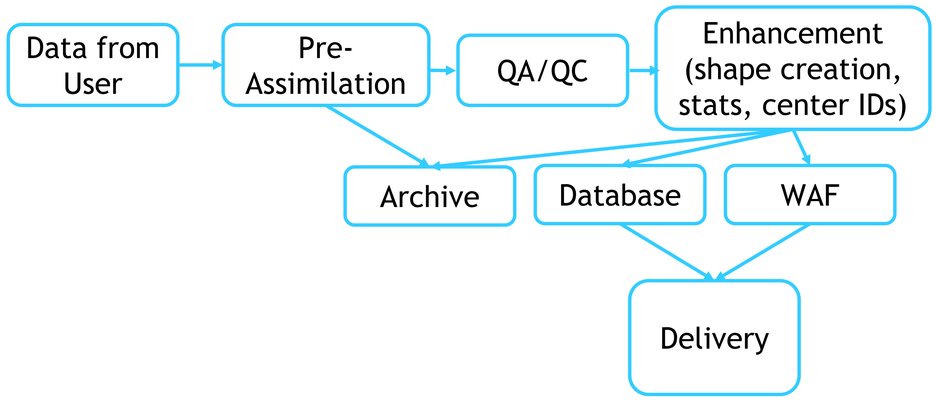
The Trackline Geophysics data set consists of single beam bathymetry, gravity, magnetic, and sub-bottom seismic data. This also includes historical seismic data and scanned microfilm, and will soon include side scan sonar. These data had previously been managed using the Geophysical Data System (GEODAS), which has since become outdated and inefficient. Following the deprecation of GEODAS, a new system of data assimilation, metadata processing, and QA/QC tools was required. This lead to the development of a new hybrid on-prem/cloud based data ingest pipeline, known as **TrackPipe**.

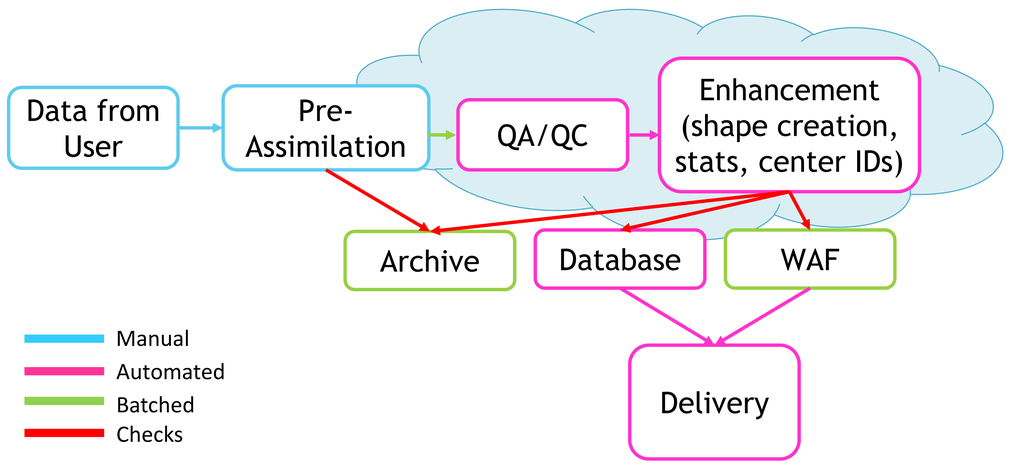


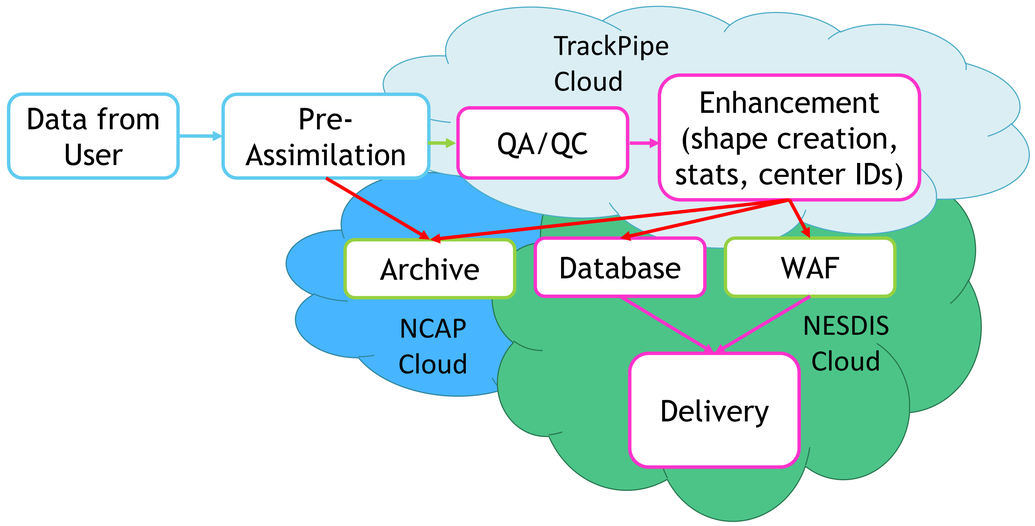
Planning

While planning for a cloud compatible data pipeline, the first objective was break down workflow

- What needs to be manual?
- What can be automated?
- What can be batched?
- Where to put checks?







Transitioning to the Cloud

- Identify the core functions
- In/Out of the cloud is expensive
- Think about error handling
  - How to report errors
- Encapsulate

The Future of the Cloud

- Connect the Clouds
- Coordinate with similar datasets
- Leverage open source and shared libraries
- Be proactive
  - Reach out to NCAP and NESDIS Cloud to stay compatible

Conclusion

- **Maintainable**
  - Emphasize shared and open source tools
- **Modular**
  - Break down components to fundamental units
  - Moved independently
- **Coordinate**
  - Reach out to NCAP and NESDIS Cloud teams
  - Work with related data managers
- **Refactor**
  - Take advantage of cloud strengths
    - Scalable
    - Powerful
    - Automated Workflows