



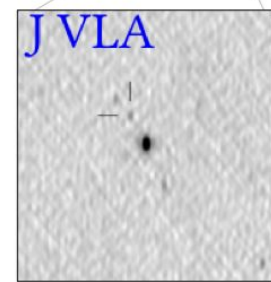
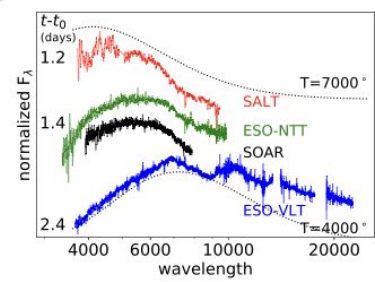
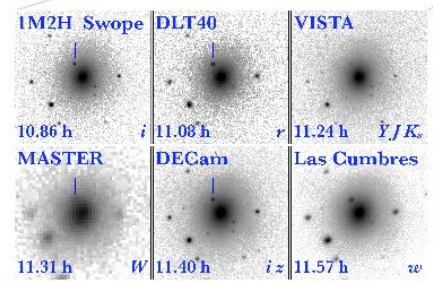
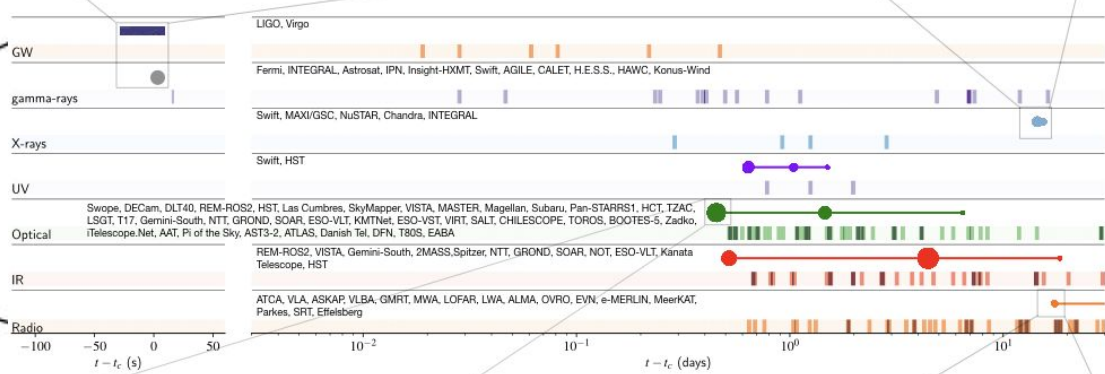
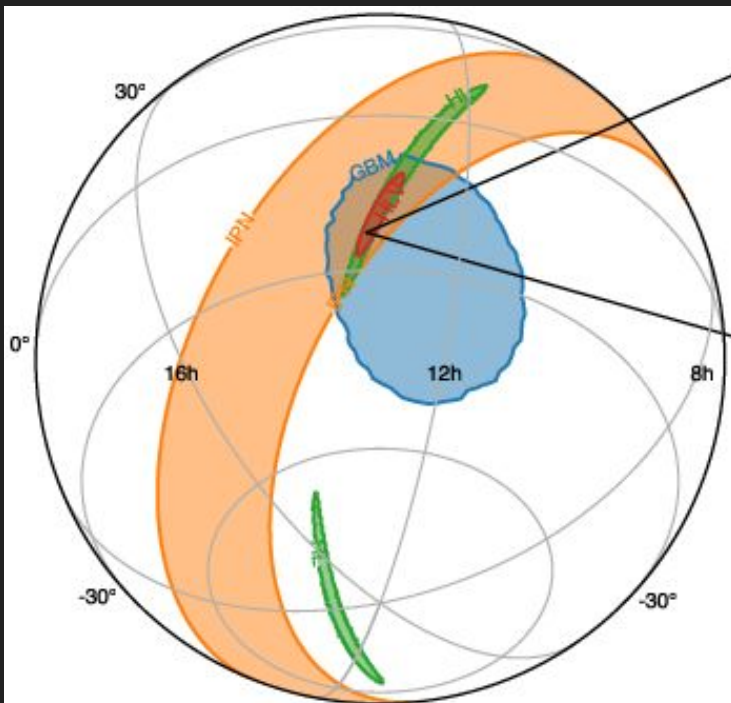
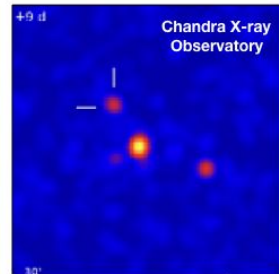
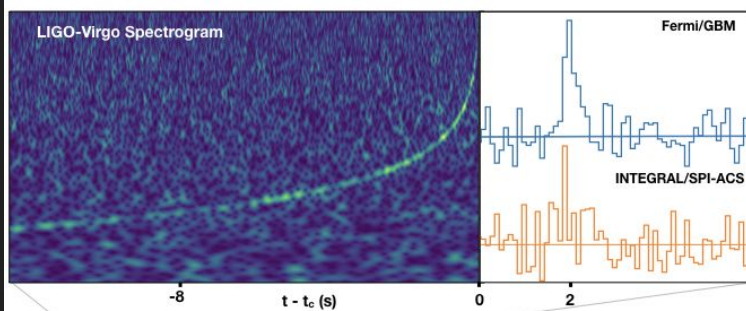
# Low latency data and the DATA MESH

IC4MF  
January 2024

Jameson Rollins  
LIGO Caltech

# LIGO GW170817

First binary neutron star merger multi-messenger observation

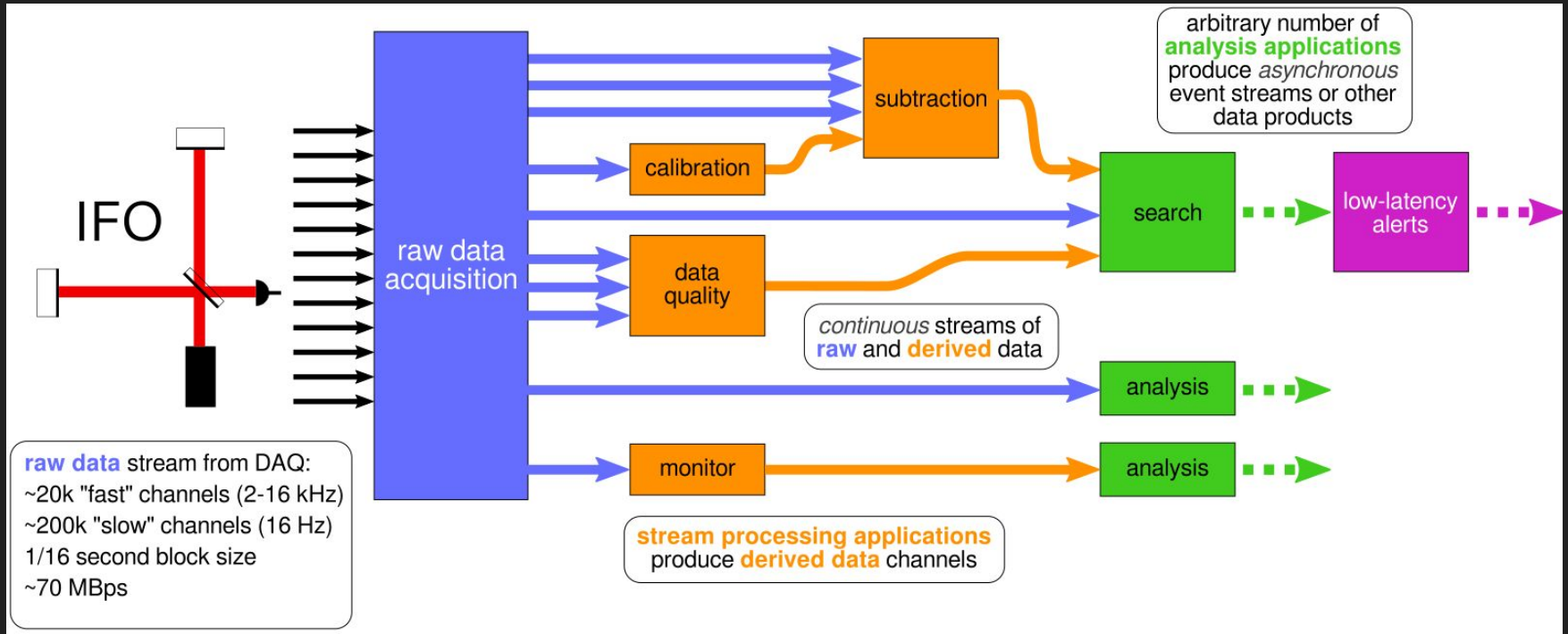


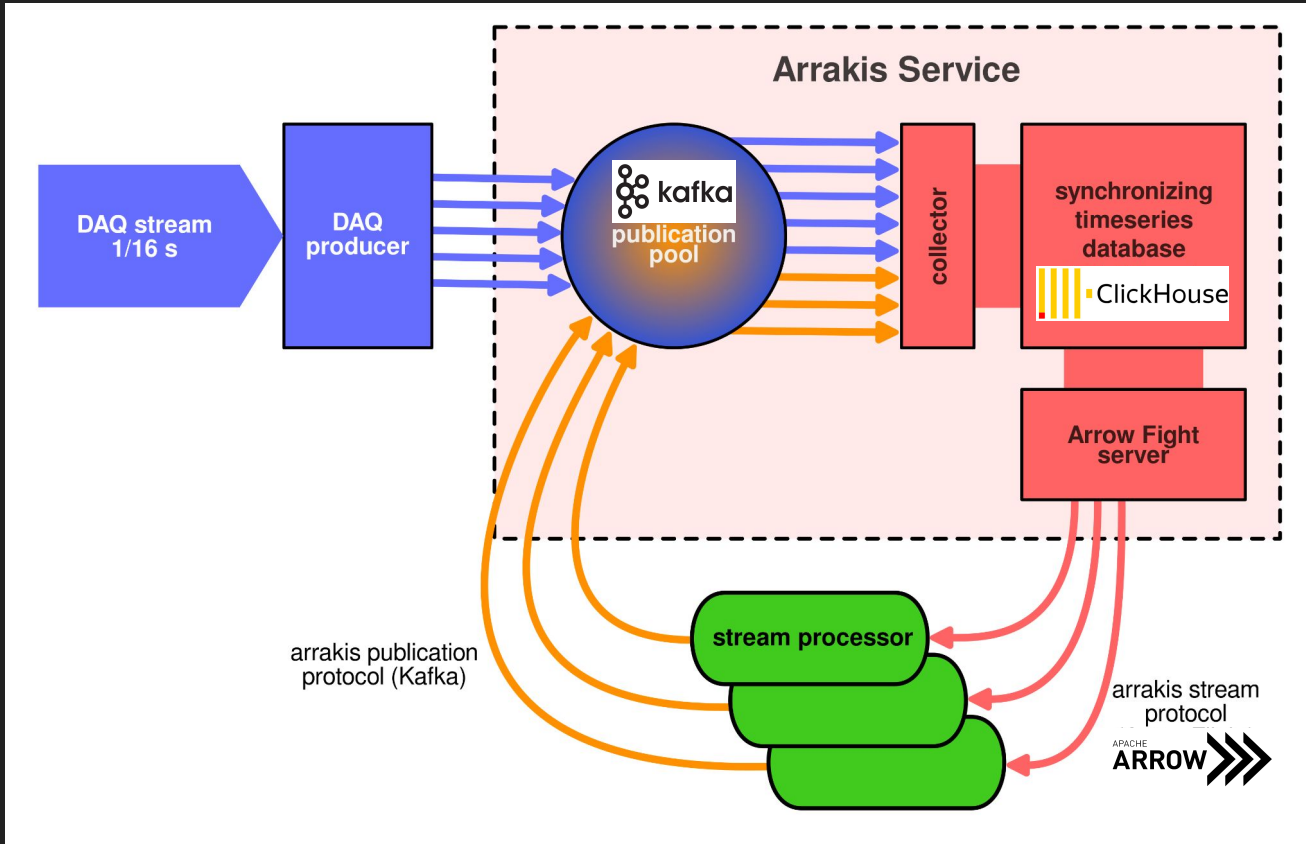


## Next generation data delivery for LIGO

LIGO project undertaking a complete overhaul of its stream processing and data distribution infrastructure. Goals:

- Online/offline data access with the same interface
- Ability publish “derived” data streams that are immediately discoverable and accessible.
- Automatic data discovery







## Data Mesh

Data Mesh is a data management and distribution concept that seems to be gaining some traction in the tech world. The basic tenants are:

- *Distributed domain-driven architecture*
  - No centralized data store, different domain manage their own data products.
- *Data as a product*
  - Data is *discoverable, addressable, trustworthy, self-describing*, etc.
- *Self-serve data infrastructure*
  - Individual domains serve their own data.
- *Federated data governance*
  - Global standards for protocols and APIs, centralized registry of data products, etc.

How could we apply these principles to our needs?

- Each “domain” (site, pipeline, working group, etc.) defines and manages their own data products.
- Common shared protocol, API, and interfaces for access
- Standardized data types
- Centralized registry of domains
- Data discoverable across domains (detectors, pipelines, etc)

