Demythifying Fink services

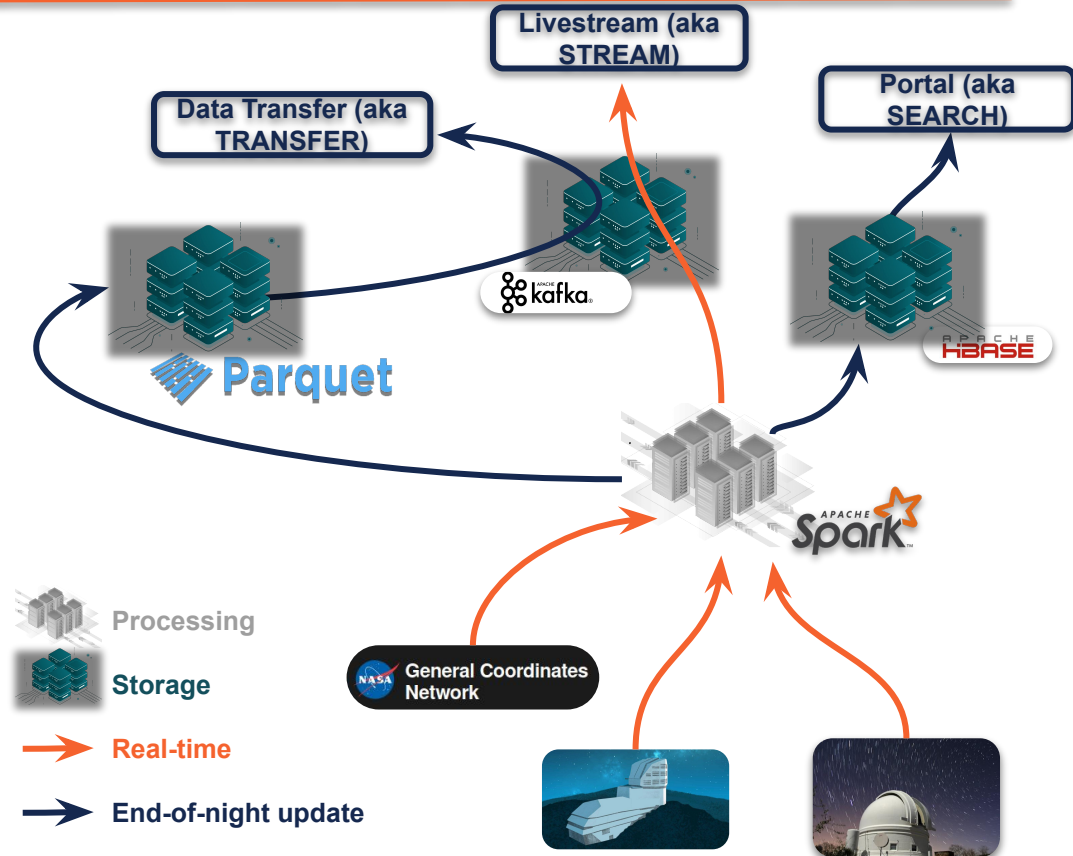## What's under the hood

Julien @ IJCLab
09/01/2024

What is what?

Data **provenance** is not the same for all services.

New data **availability** is not the same for all services.

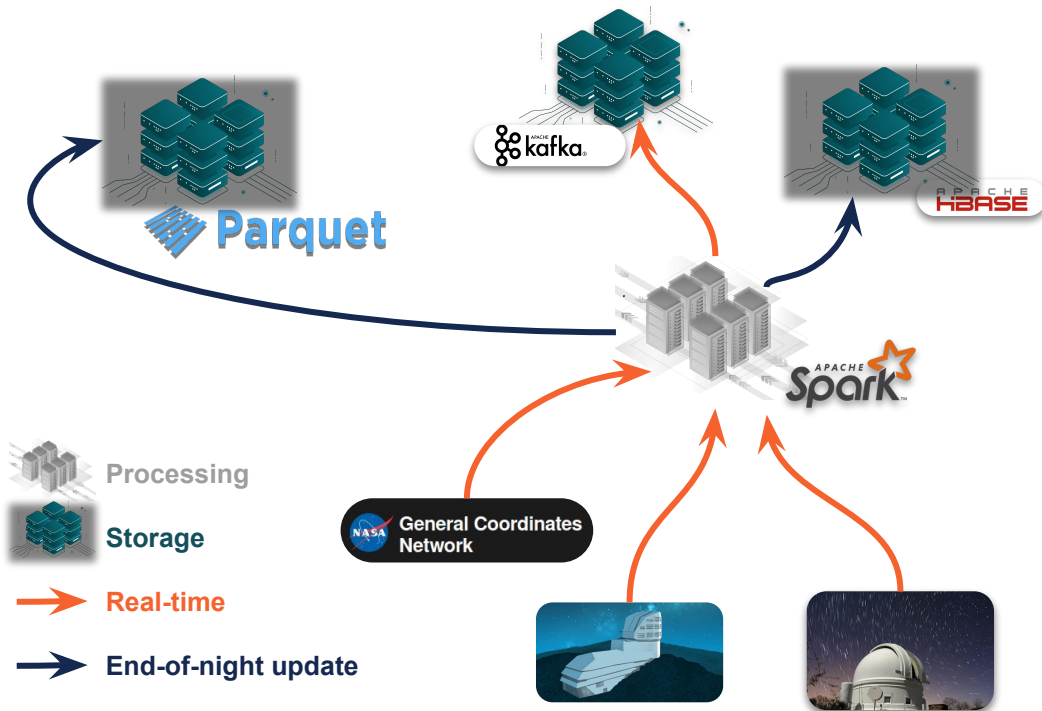Data **structure** is not the same for all services.

fink-broker

Data provenance: All external streams.

Data availability: "real-time"
- ZTF (20min)
- LSST (1min)

Data structure: original alerts (avro)
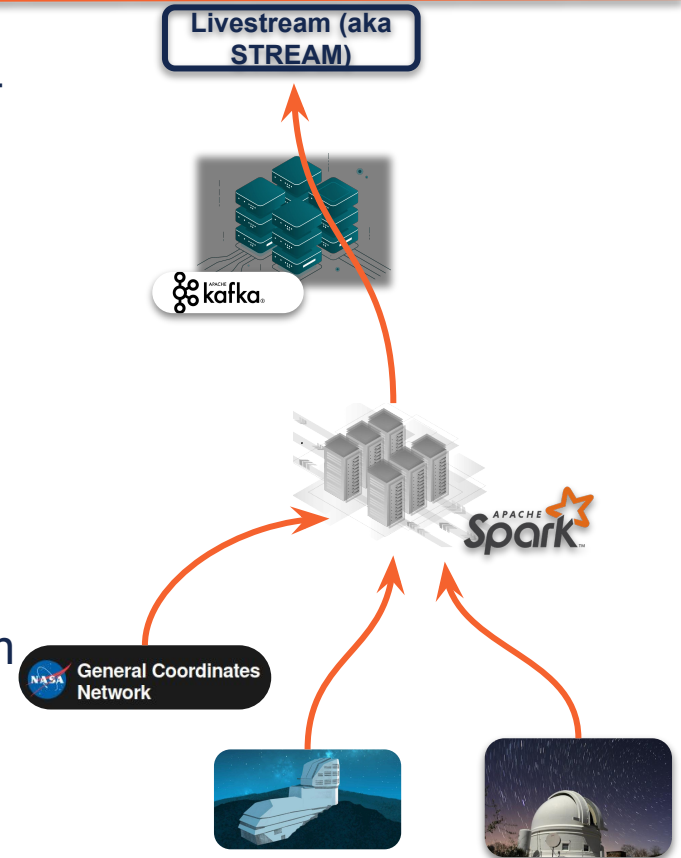- Science modules process the alert data.

Stream

Data provenance: All external streams. ZTF/LSST alone is handled by `fink-broker`, otherwise information from GCN is handled by `fink-mm`.

Data availability: "real-time"
- ZTF (20min) + Fink (~1min)
- LSST (1min) + Fink (30 seconds)

Data structure: processed alerts (avro)
- Original alert content + Fink added values from science modules
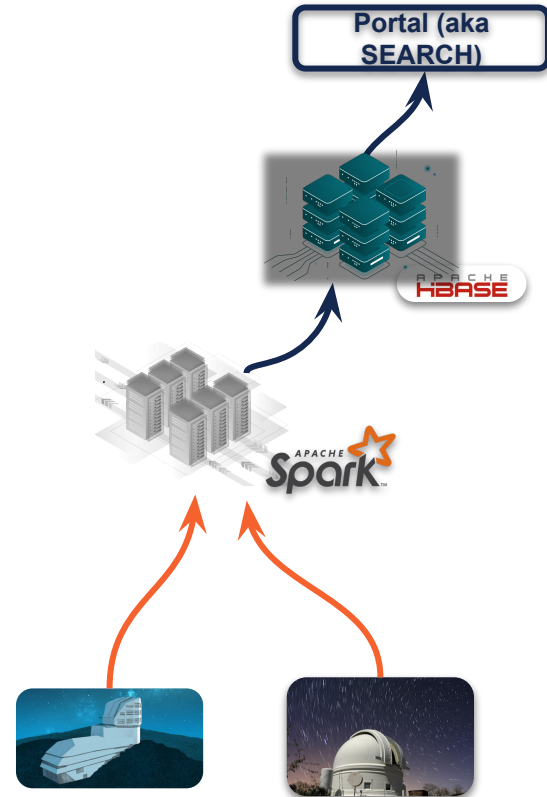- Filters select interesting alerts.

Livestream (aka STREAM)

General Coordinates Network

Search

Data provenance: Science Portal (aka SEARCH) gives access to ZTF alert data only.

Data availability: end-of-the night (heavy operations…)

Data structure: non-relational distributed database
- Tables with unique index
- Tables with column families (`i:`, `d:`, etc.) and columns (original ZTF or Fink name)
- Main table indexed by objectId (enable to reconstruct aggregated lightcurve efficiently)
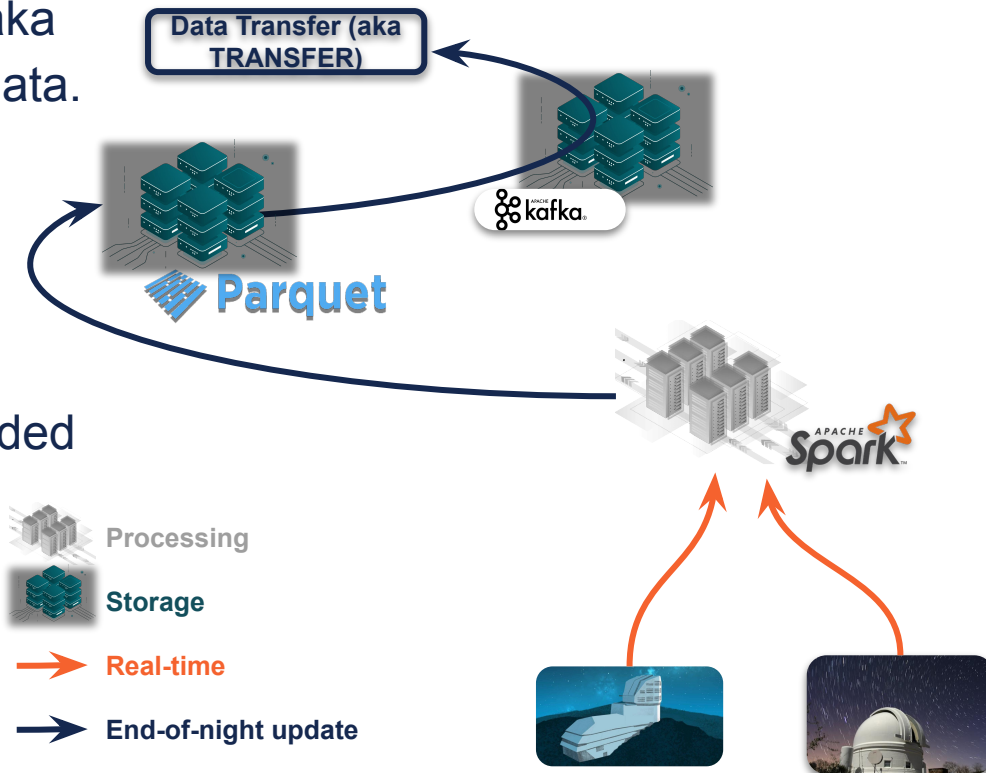- Pros/cons: past presentation

Transfer

Data provenance: Data Transfer (aka TRANSFER) gives access to ZTF data.

Data availability: end-of-the night.

Data structure: processed alerts (avro) like STREAM
- Original alert content + Fink added values from science modules

Data Transfer (aka TRANSFER)

Parquet

kafka

Spark

Processing

Storage

→ Real-time

→ End-of-night update

Recurring issues

Search is often frozen between 8 and 9pm Paris time
- We are having database issues T_T

I have a list of 1,000 objects, Search is slow to get me results

- Did you select some columns

I cannot install fink-client correctly
- We had a constraint on the environment which hopefully is resolved

Fink-GPT

Do you know why Search is often frozen between 8 and 9pm Paris time?

We are having database issues. The Hbase settings is not great, and we tried to improve many times. We think we understand the problem, but we are struggling to find a satisfying solution. It is solved at the other datacenter though.

Fink-GPT

I have a list of 10,000 objects, Search is slow to get me results. How can I improve the query time?

The database is not made for large queries. You could try to select only a subset of all columns. If this does not work, you should use the Data Transfer service instead.

Fink-GPT

What is the Search timeout?

The timeout in Search for a single query is 3 minutes.

Do you cache intermediate results?

Yes.

Fink-GPT

In Search, what is the meaning of columns?

You can get this information from this URL: https://fink-portal.org/api/v1/columns. Do not hesitate to bookmark it to access it quickly!

Fink-GPT

Can I get all Fink data?

Your question is odd, but if you really want so, you can use the Data Transfer service. Beware, you will need over 10TB of disk space and a good internet connection. You might warn Julien before.

Discussion

Case 1 : You are a user of Search. You enjoy it a lot (obviously!), but there are missing features you would like to see. Which ones?

Case 2 : You heard a lot about Fink (obviously!), and you are working in a scientific area where Fink could help you. How?

Case 3 : You thought adding a building brick in Fink (aka science module or a filter), but you do not know how to do it. Are you seeking help?

Other cases?