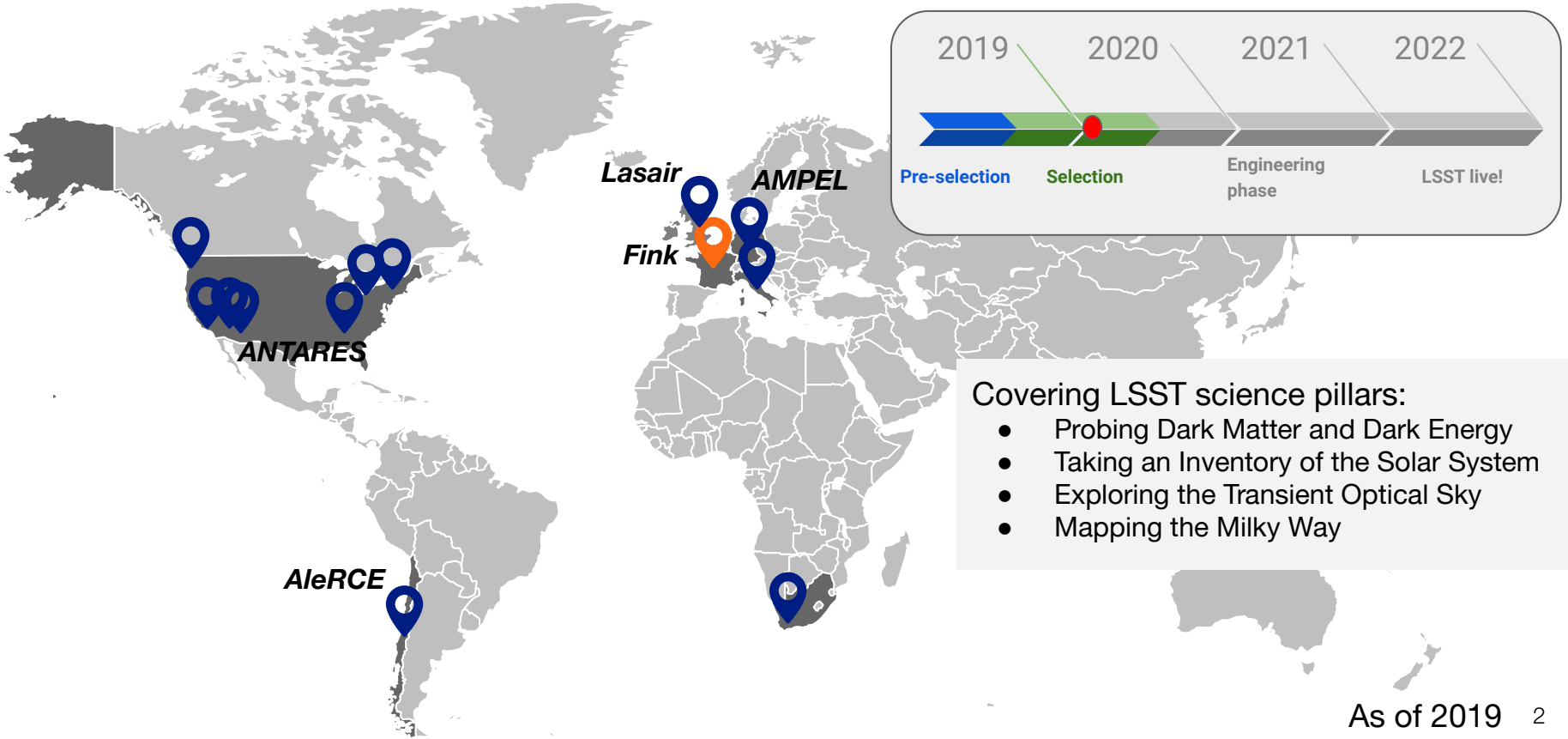
Fink: status and roadmap

Anais Möller and Emille E. O. Ishida
(LPC-Clermont)

Julien Peloton
(CNRS/IJCLab)

On behalf of the Fink initiative

LSST Broker landscape



Covering LSST science pillars:

- Probing Dark Matter and Dark Energy
- Taking an Inventory of the Solar System
- Exploring the Transient Optical Sky
- Mapping the Milky Way

As of 2019

Fink *in the broker landscape*

Lasair

- Main added value is content + cross-match with static data base, ML under development
- Built to fulfill the needs of the British transient community
- *Main users focused on UK telescopes*

Alerce

- High emphasis in hierarchical classification, ML, interdisciplinarity
- Aim to explore the potential of follow-up facilities in Chile
- *Hosts data challenges and hackathons*

ANTARES

- High emphasis in the front-end and api development
- Important ML component being adapted, focus on early classification
- *No specific events for community engagement*

Fink *in the broker landscape*

Lasair

- Main added value is content + cross-match with static data base, ML under development
- Built to fulfill the needs of the British transient community
- *Main users focused on UK telescopes*

Alerce

- High emphasis in hierarchical classification, ML, interdisciplinarity
- Aim to explore the potential of follow-up facilities in Chile
- *Hosts data challenges and hackathons*

ANTARES

- High emphasis in the front-end and api development
- Important ML component being adapted, focus on early classification
- *No specific events for community engagement*

Fink

- State of the art ML techniques: adaptive ML and Bayesian NN
- Aim to fulfil the needs of the French+ community and explore the potential in the LSST data base hosted at CC
- *Emphasis on community-driven science*

Fink *in the broker landscape*

Lasair

- Main added value is content + cross-match with static data base, ML under development
- Built to fulfill the needs of the British transient community
- *Main users focused on UK telescopes*

Alerce

- High emphasis in hierarchical classification, ML, interdisciplinarity
- Aim to explore the potential of follow-up facilities in Chile
- *Hosts data challenges and hackathons*

ANTARES

- High emphasis in the front-end and api development
- Important ML component being adapted, focus on early classification
- *No specific events for community engagement*

Fink

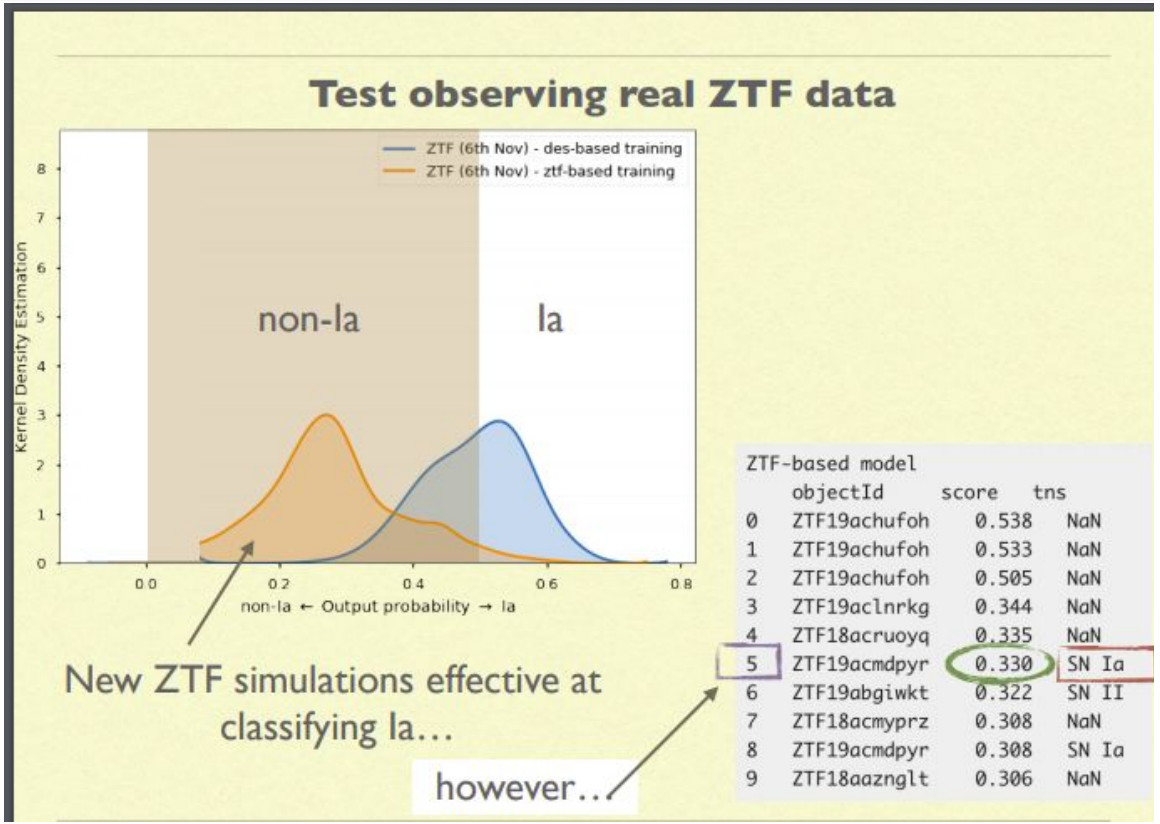
- State of the art ML techniques: adaptive ML and Bayesian NN
- Aim to fulfil the needs of the French+ community and explore the potential in the LSST data base hosted at CC
- *Emphasis on community-driven science*

Fink *important remark*

Fink *example modules*

Supernova Classification

by Marco Leoni (IJCLab)



Fink *example modules*

Connection with other facilities: SVOM, GRANDMA

By Nicolas Leroy (IJCLab)

Gravitational waves, Kilonova, GRBs

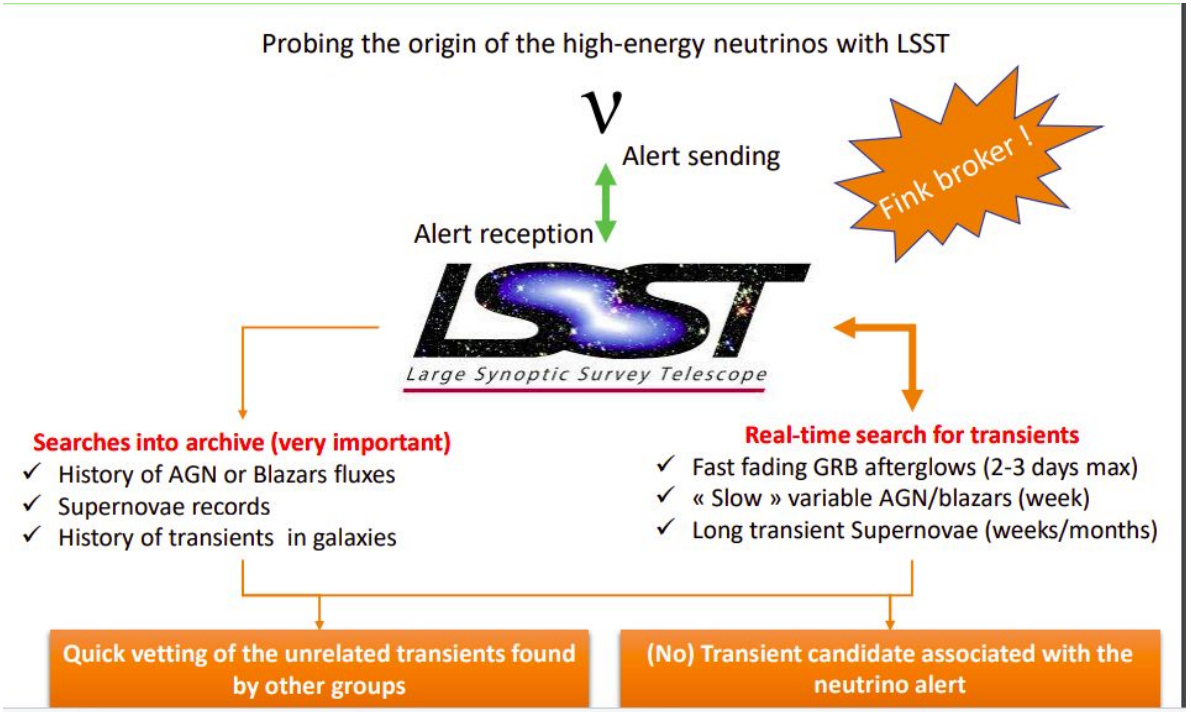
FINK and GRANDMA

FINK and SVOM

Fink *example modules*

High Energy Neutrinos source identification

by Damien Turpin (APC)



Fink *modules under discussion*

- First tests with microlensing
  - Tristan Blaineau
- Connections with other brokers
  - Alexis Coleiro, Andrii Neronov, Volodymyr Savchenko
- Deblending
  - Alexandre Boucaud

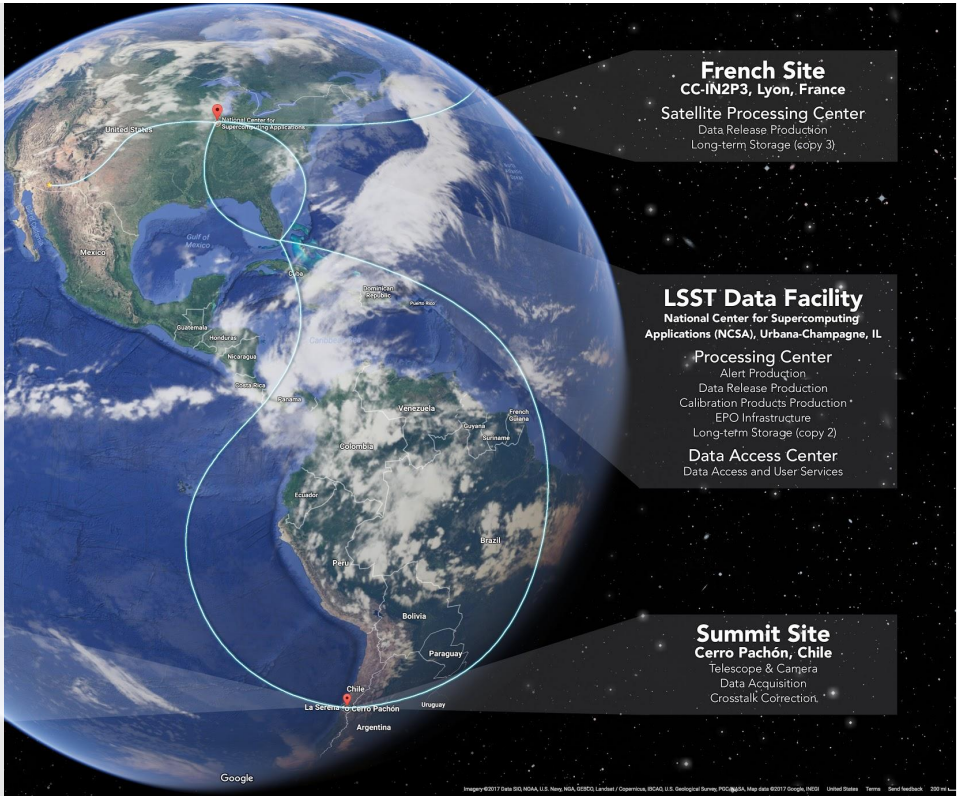
<u>Next in line:</u>

- Anomaly detection

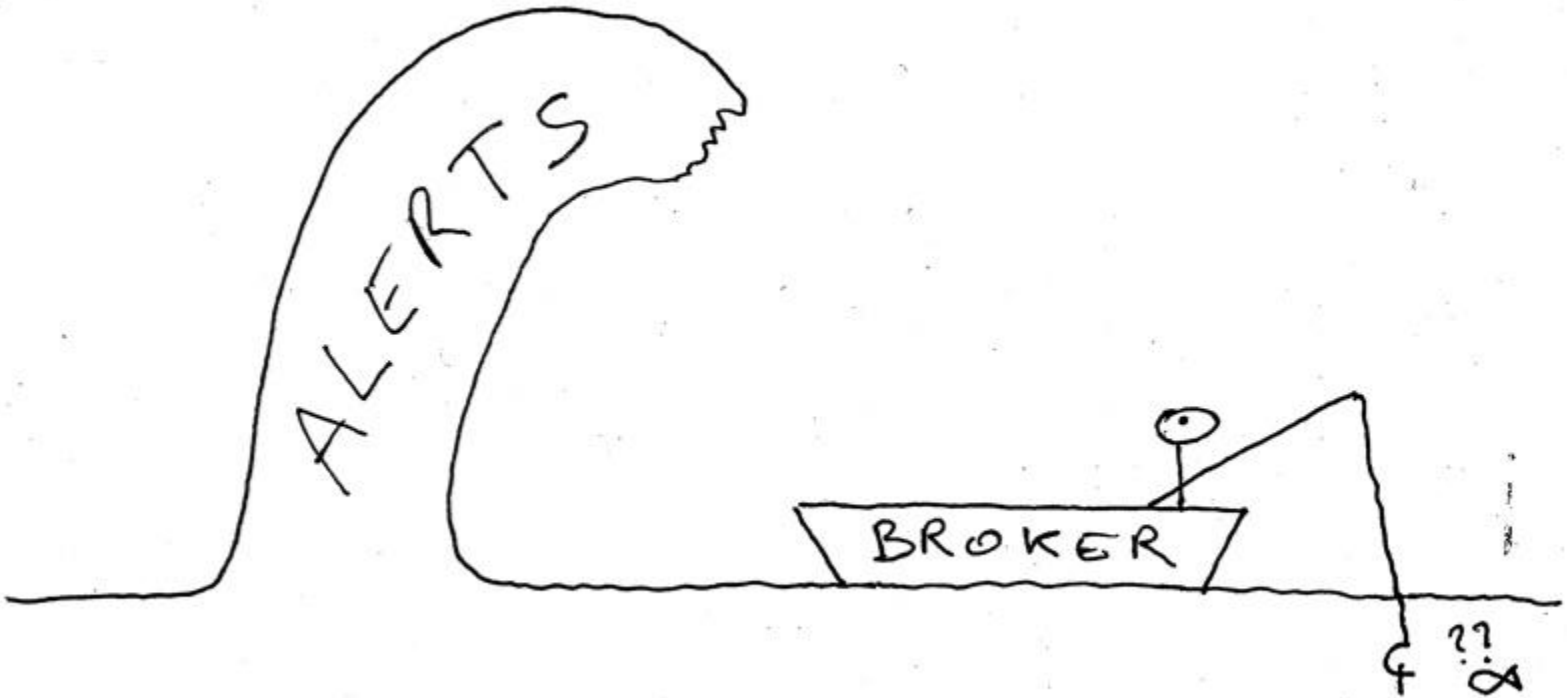
Fink *next steps*

Some Data Challenges...

- Forecasted: 10 million alerts per night…
  - Current serialisation implies ~82KB/alert, 800 GB/night, 3PB in 2030.
- 98% of alerts must be transmitted with 60 seconds of readout...
  - … and processed before the next night!
- Wires to send alerts worldwide are not infinitely big...



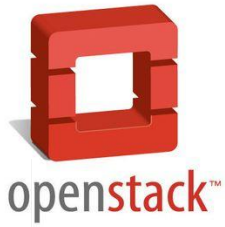
Concretely...



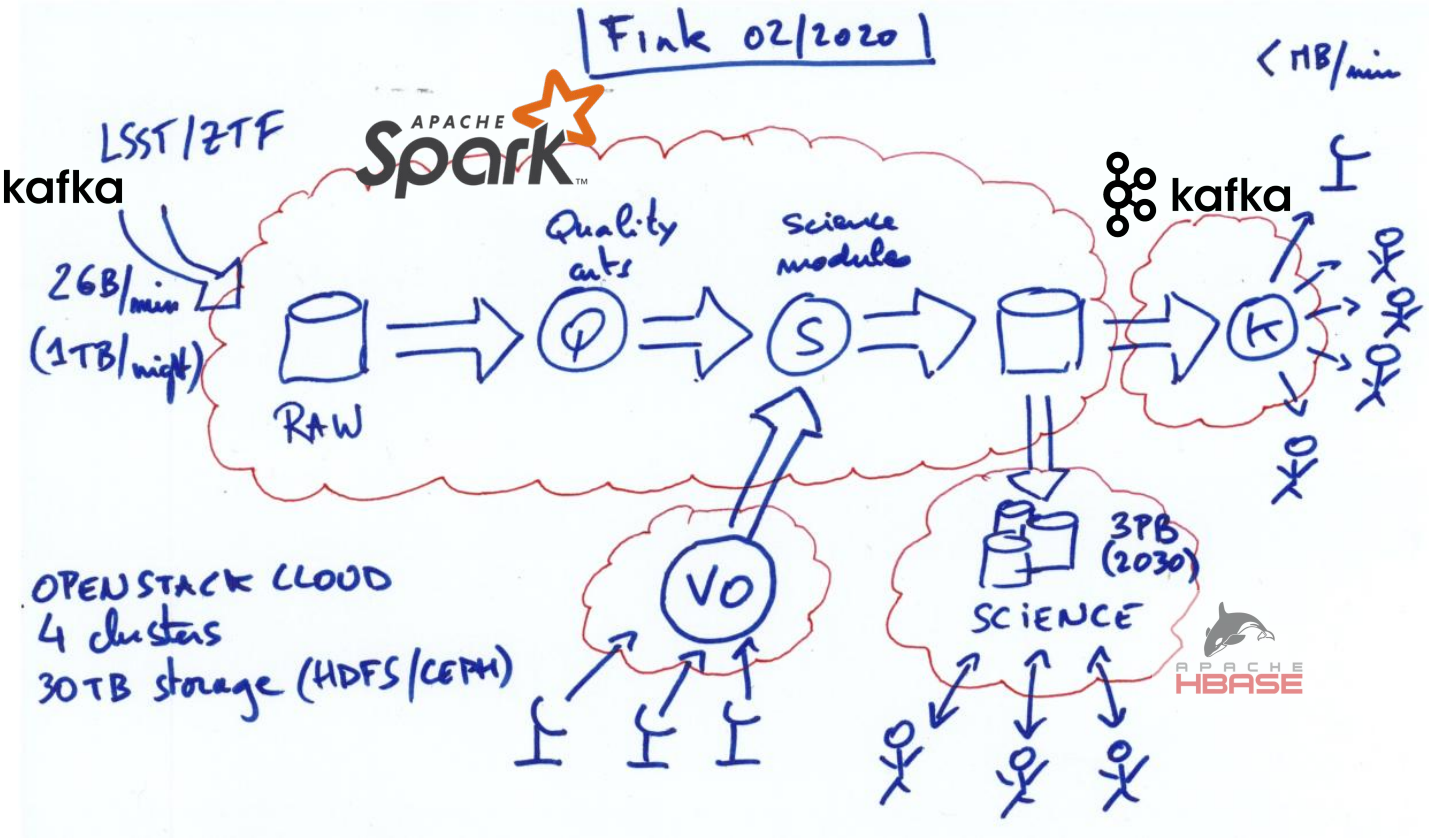
Fink challenges & design

Fink's design is driven by:

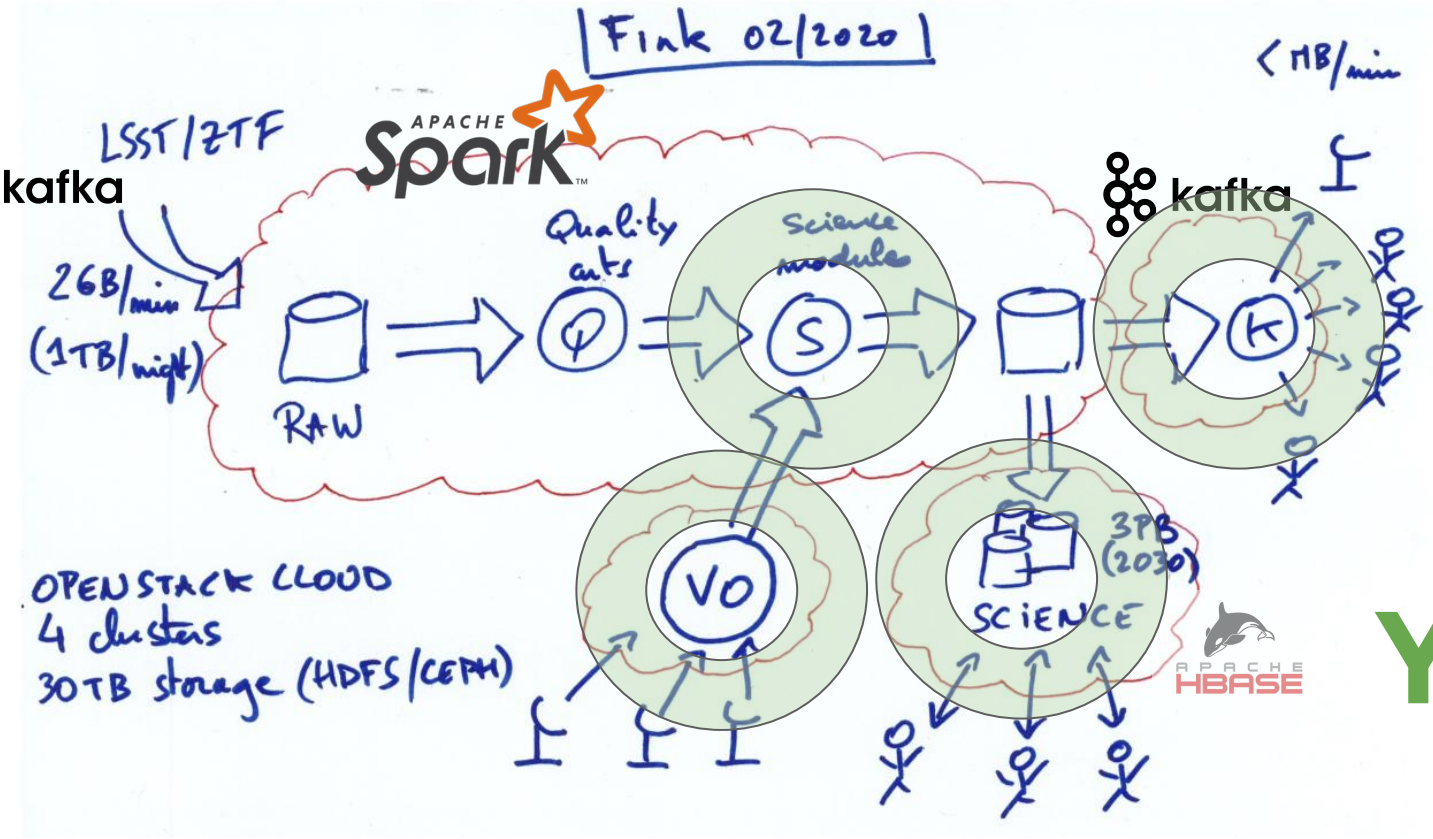
- Maximizing the scientific return on LSST and related experiments over the next decade: SVOM, CTA, Integral, KM3NET, ...
- Working efficiently at scales: real time and post-processing.
- Having a good integration with the current ecosystem: we are not alone!



Alert processing in Fink



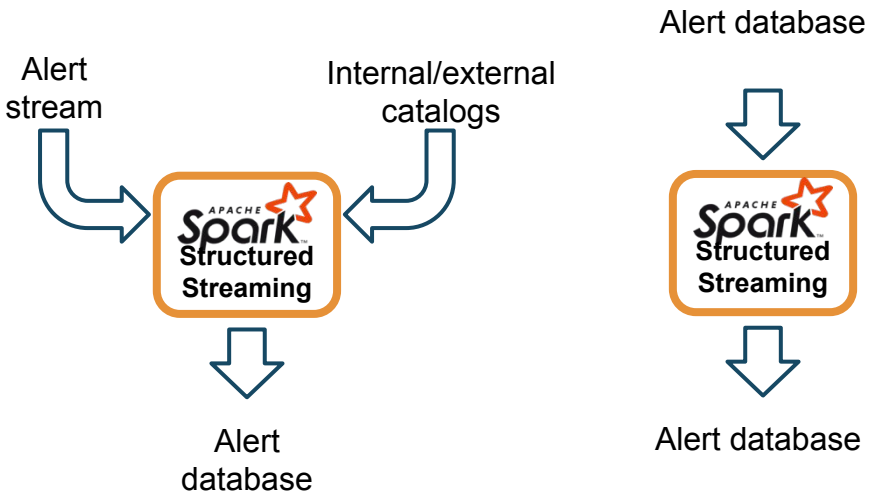
Fink community

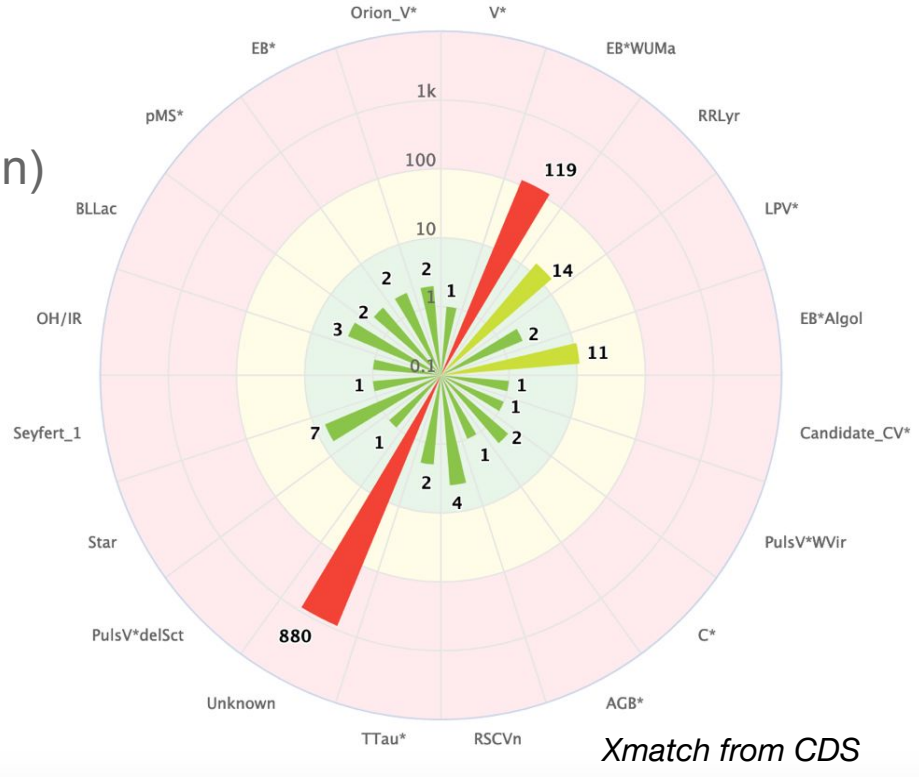


Science modules

Add values to the raw alerts

- Broker services (e.g. cross-match)
- User-defined modules (e.g. classification)

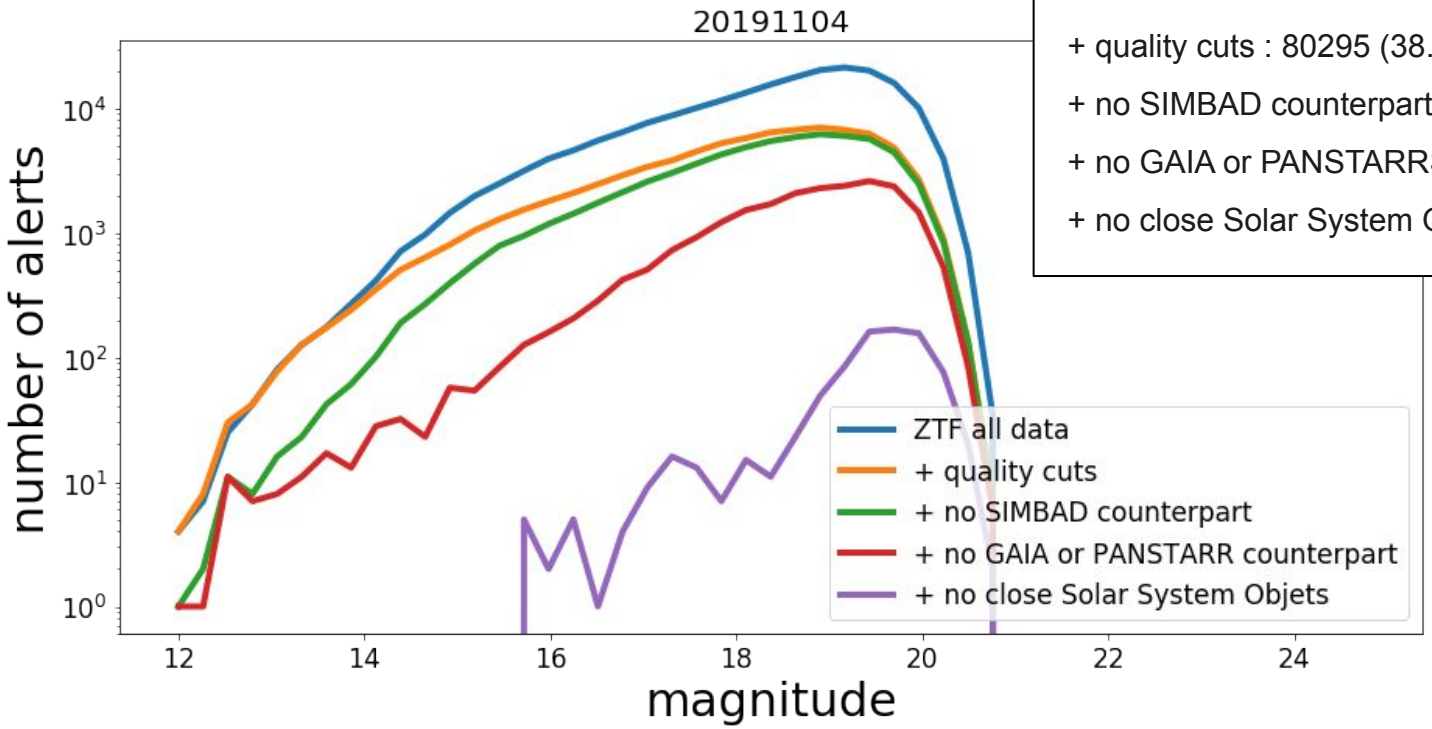




Xmatch from CDS

C. Arnault, E. Ishida, M. Leoni, A. Möller + Fink community

Filters and data reduction



Entries for ZTF all data : 208920 (100.0%)

+ quality cuts : 80295 (38.4%)

+ no SIMBAD counterpart : 65141 (31.2%)

+ no GAIA or PANSTARRS counterpart : 21895 (10.5%)

+ no close Solar System Objects : 827 (**0.4%**)

Filters are crucial to not be spammed!

Communication protocols

Apache Avro:

- Data serialisation format for alerts
- Extremely efficient (header/data, compression, ...)
- Need a schema to decode the information

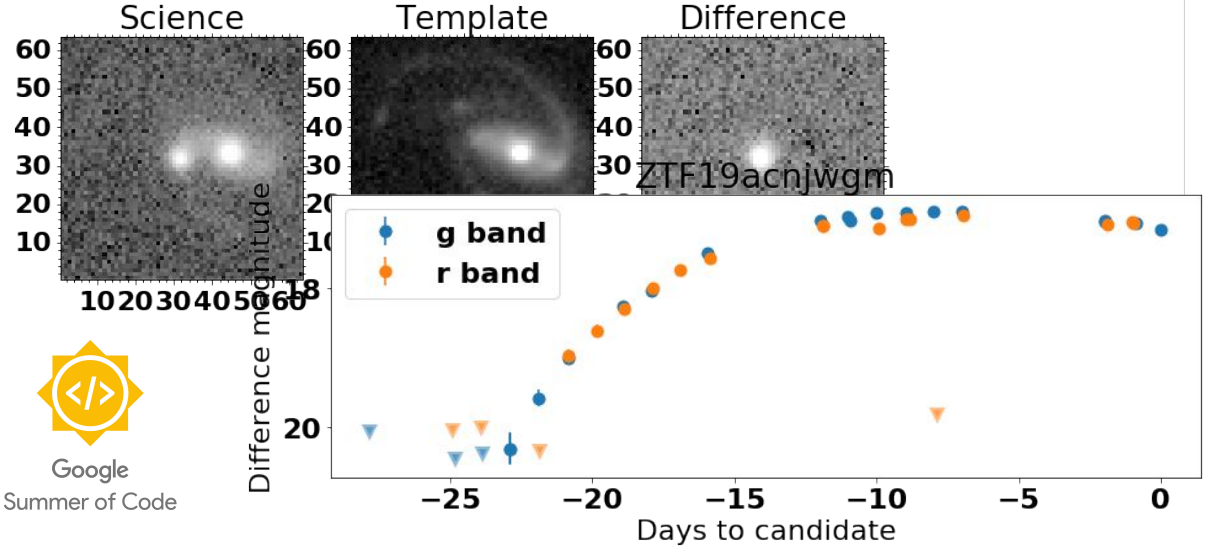
Apache Kafka:

- Stream-processing software platform (pub/sub)
- Distributed management of alerts (with his friend Zookeeper)
- Kafka can handle thousands of users with thousands of topics

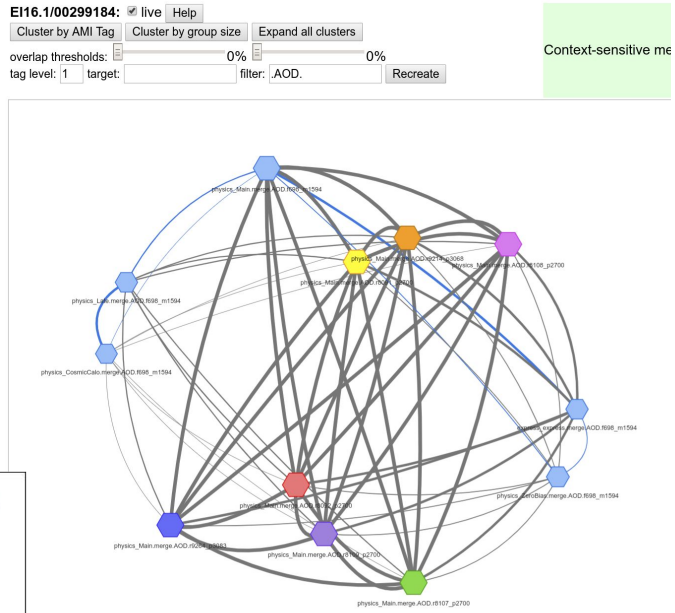
User interface(s)

Two entry points for users:

- Fink streams: This workshop!
- Science DB: Graph oriented DB (J. Hrivnac).



Google Summer of Code



J. Hrivnac, ATLAS

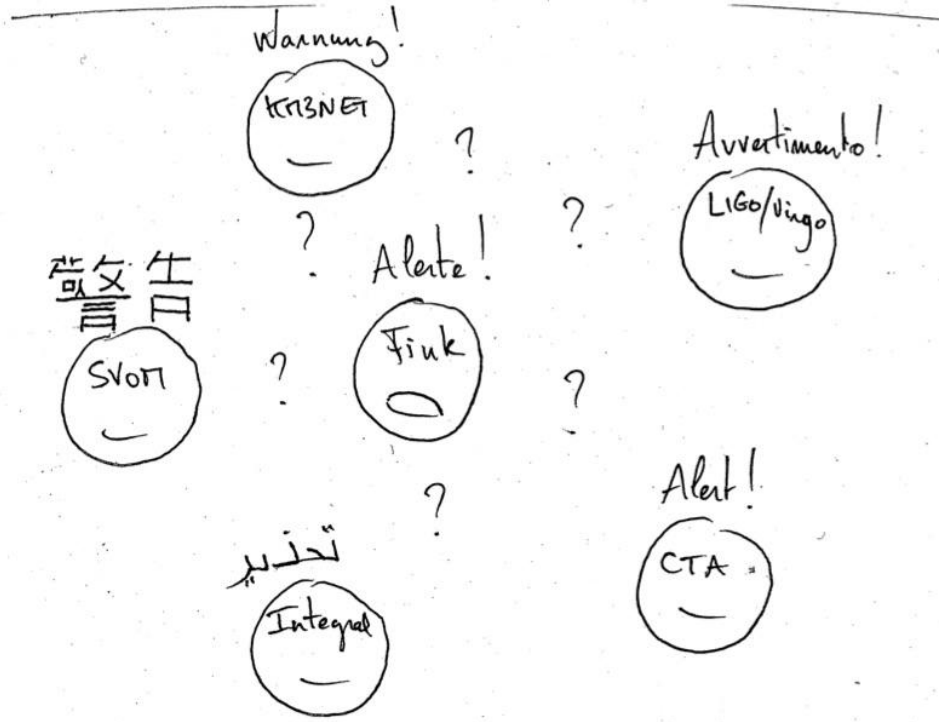
Joining information

Challenge: different data formats, different communication protocols.

Current solution:

- Use Comet to receive VOEvents
- Convert on-the-fly into a Fink-friendly stream
- Perform coincidence using a temporal window of few minutes

Status: (largely) experimental.



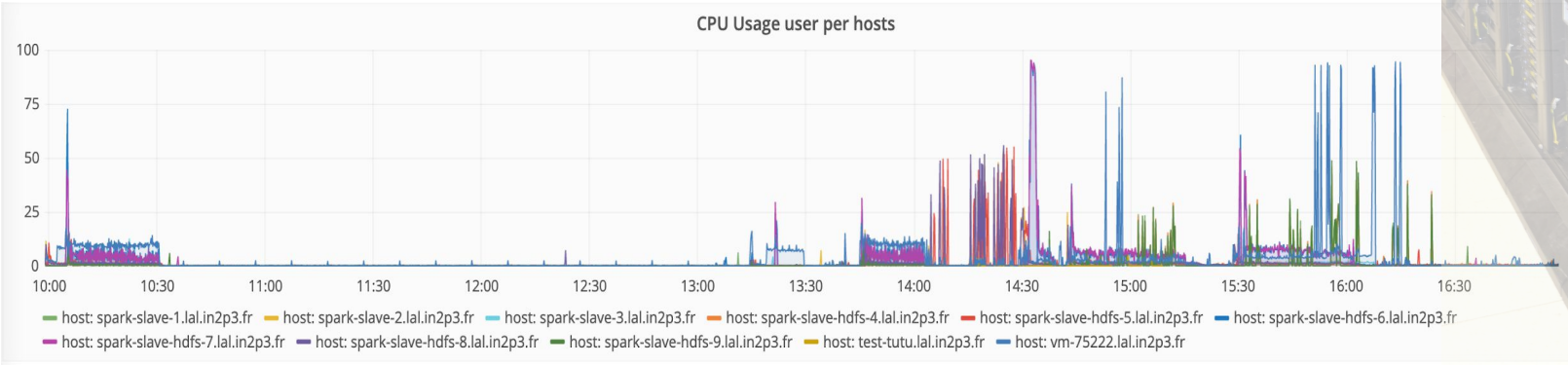
Contact us if you are using VOEvents.

Prototype status

Deployed broker instance for R&D in the VirtualData Cloud (UPSaclay)

- **Communication**: Apache Kafka cluster (5 machines, 20 cores)
- **Processing**: Apache Spark cluster (11 machines, ~200 cores)
- **Science DB**: Apache HBase (1 machine, 6 cores).

Science storage: O(10)TB distributed storage (HDFS, Ceph + s3)



Workshop goals

Tutorials to explore alert data and the Fink Kafka client:

- Exploring ZTF Alerts
- Connect to Fink alert streams
- Fink filters: how they work?
- Fink science modules & broker added values
- Fink and external alert streams

Throughout all tutorials, we will use **ZTF alert data** that are currently available and have similar structure to the expected one for LSST alerts.

Happy coding!

To do before starting...

Apply (small) patches ~~done at midnight last night…~~

1. cd fink-client; git pull
2. cd fink-tutorials; git pull

Connection ID: on the whiteboard.

The full proposal for LSST

- Considered as private communication
- Max 10 pages
- Due June 15, 2020
- Expected content:
  - Scientific goals: specific science modules
  - Stream access: do you want the full stream? Can you forward smaller streams to secondary brokers?
  - Data products and services: user access, added values, follow-up strategies (if any)
  - Technical implementation
- **Previous results**
  - Example data products
  - Plans to engage the community to use your system
- Management Plan
  - Number of FTE during X years
  - Potential funding resources
- Proposing team (and their experiences)
- Result by August 2020