Fink: status and roadmap

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

Julien Peloton
(CNRS/IJCLab)

On behalf of the Fink initiative

Fink *How did we get here?*

- Jan/2019: First conversations about Spark and alert streams @ Clermont
- Feb/2019: DESC Broker workshop in Berkeley, USA
- May/2019: Fink is born: letter of intent (LoI) was submitted with 31 signatures
- Jun/2019: Broker workshop in Seattle with those who submitted LoI
- Aug/2019: All LoIs are accepted, full proposal due June/2020
- Sep/2019: Call for proposals for science modules

Fink *How did we get here?*

- Jan/2019: First conversations about Spark and alert streams @ Clermont
- Feb/2019: DESC Broker workshop in Berkeley, USA
- May/2019: Fink is born: letter of intent (LoI) was submitted with 31 signatures
- Jun/2019: Broker workshop in Seattle with those who submitted LoI
- Aug/2019: All LoIs are accepted, full proposal due June/2020
- Sep/2019: Call for proposals for science modules

Since then … a lot has happened ...

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 (as of April/2019)*

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*

Focusing in **specific** science cases is vital!

Fink *important remark*

Focusing in **specific** science cases is vital!

How?

Early development of tailored science modules.

Tools at our disposal:

- Increasingly accurate classification and anomaly scores (adaptive learning)
- Calibrated probabilities (Bayesian

Fink *modules under development*

- Supernova photometric classification in ZTF alert stream
  - see Marco's talk

- Enabling kilonova discovery with ML in Fink (also GW)
  - Biswajit Biswas - M1 student working now remotely, at LPC-Clermont from April/2020

- Looking for remote calcium reach transients in ZTF data
  - Christopher Frohmaier (ICC - Portsmouth, UK)

- Connections with SVOM
  - see Nicolas' and Damien's talks

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*

- Major advantage of working with Fink is the opportunity to develop a tailored science module with the help of the development team.

- If you have one in mind, the time to talk is now!

Fink *next steps*

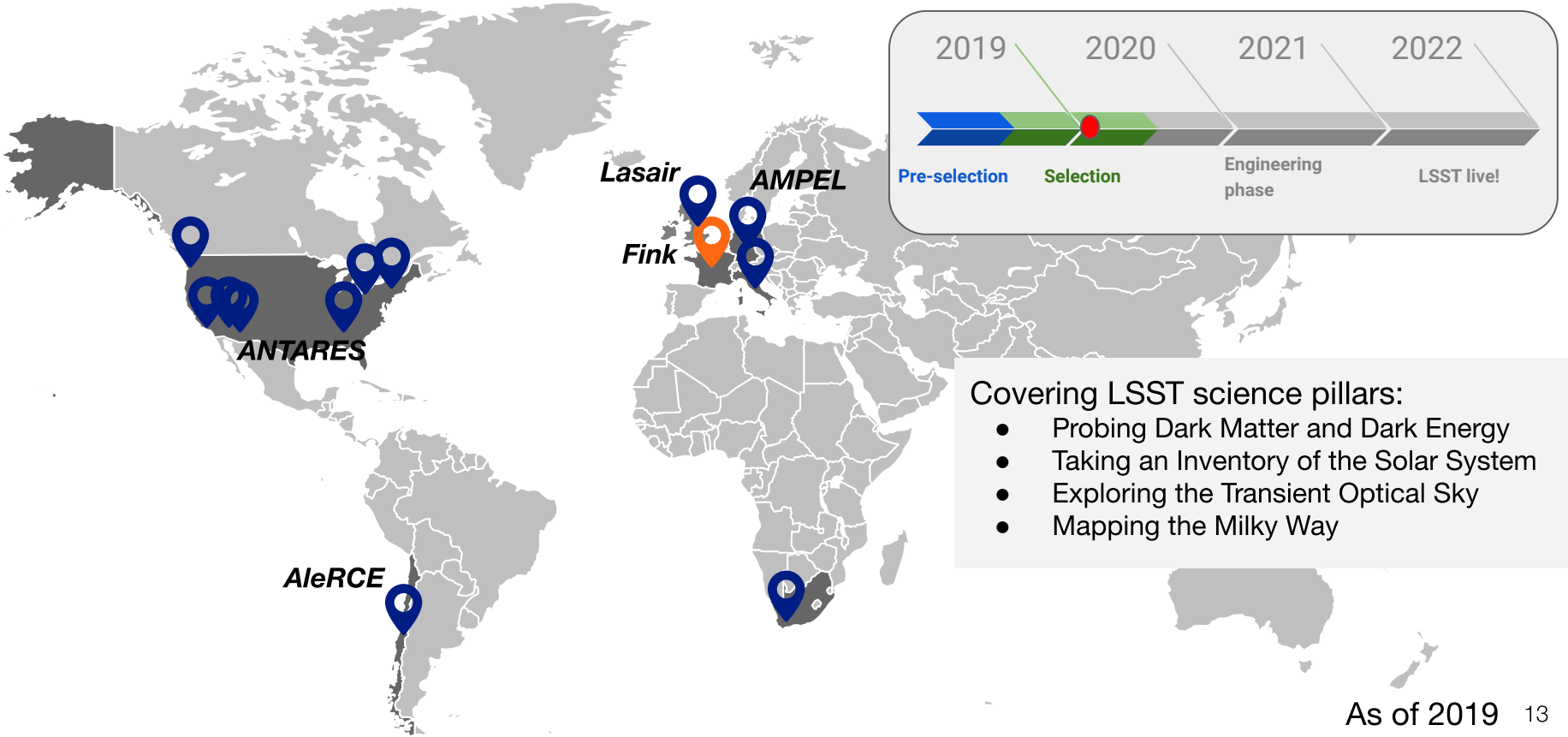
- Major advantage of working with Fink is the opportunity to develop a tailored science module with the help of the development team.

- If you have one in mind, the time to talk is now!

Regular monthly telecons to start in March/2020

Opportunity to discuss science modules more broadly

LSST Broker landscape



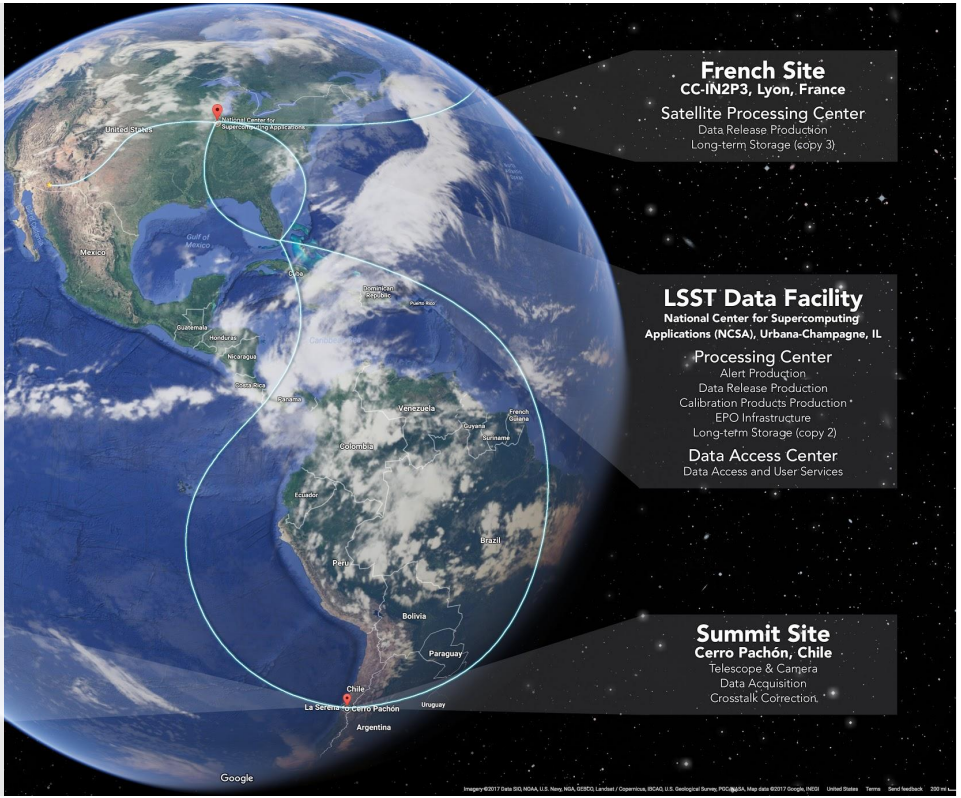
Covering LSST science pillars:

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

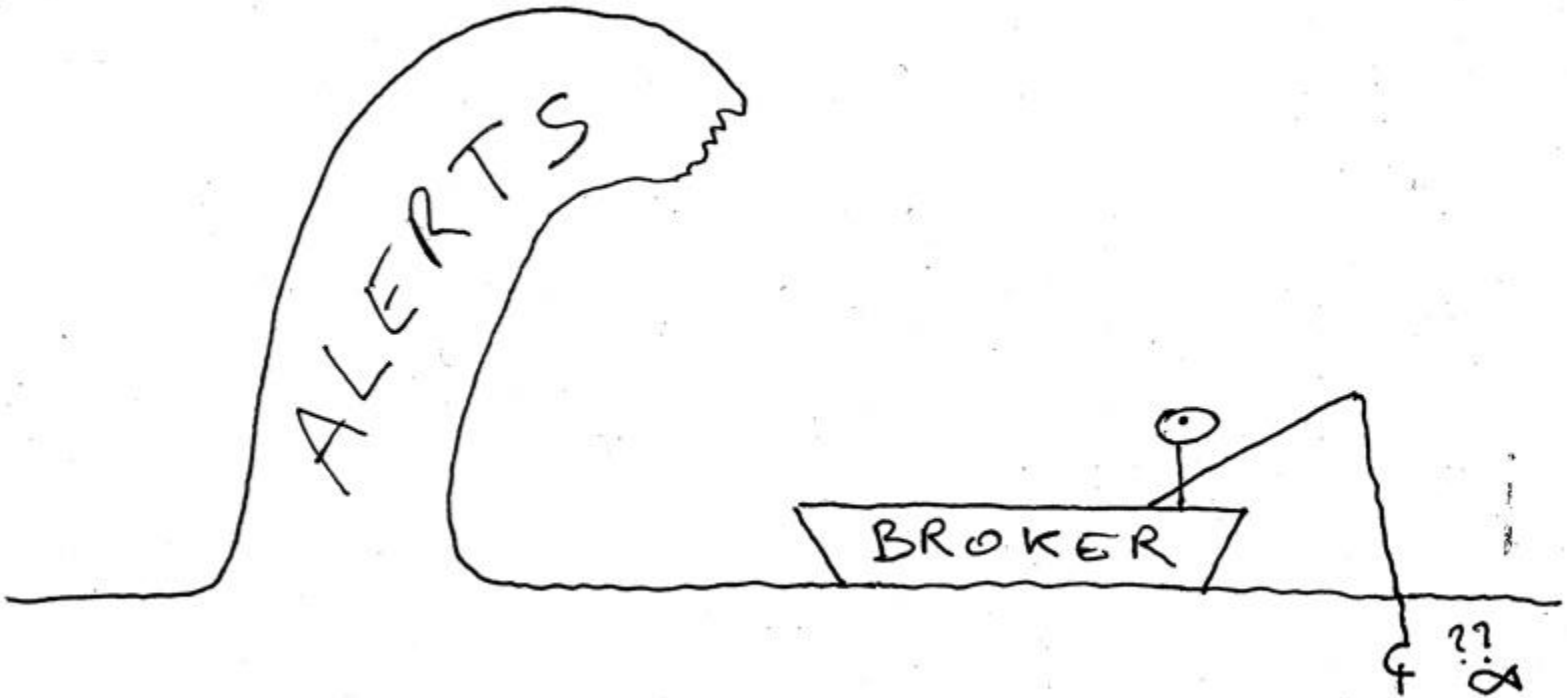
As of 2019

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...



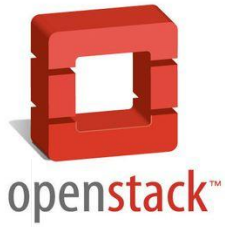
Concretely...



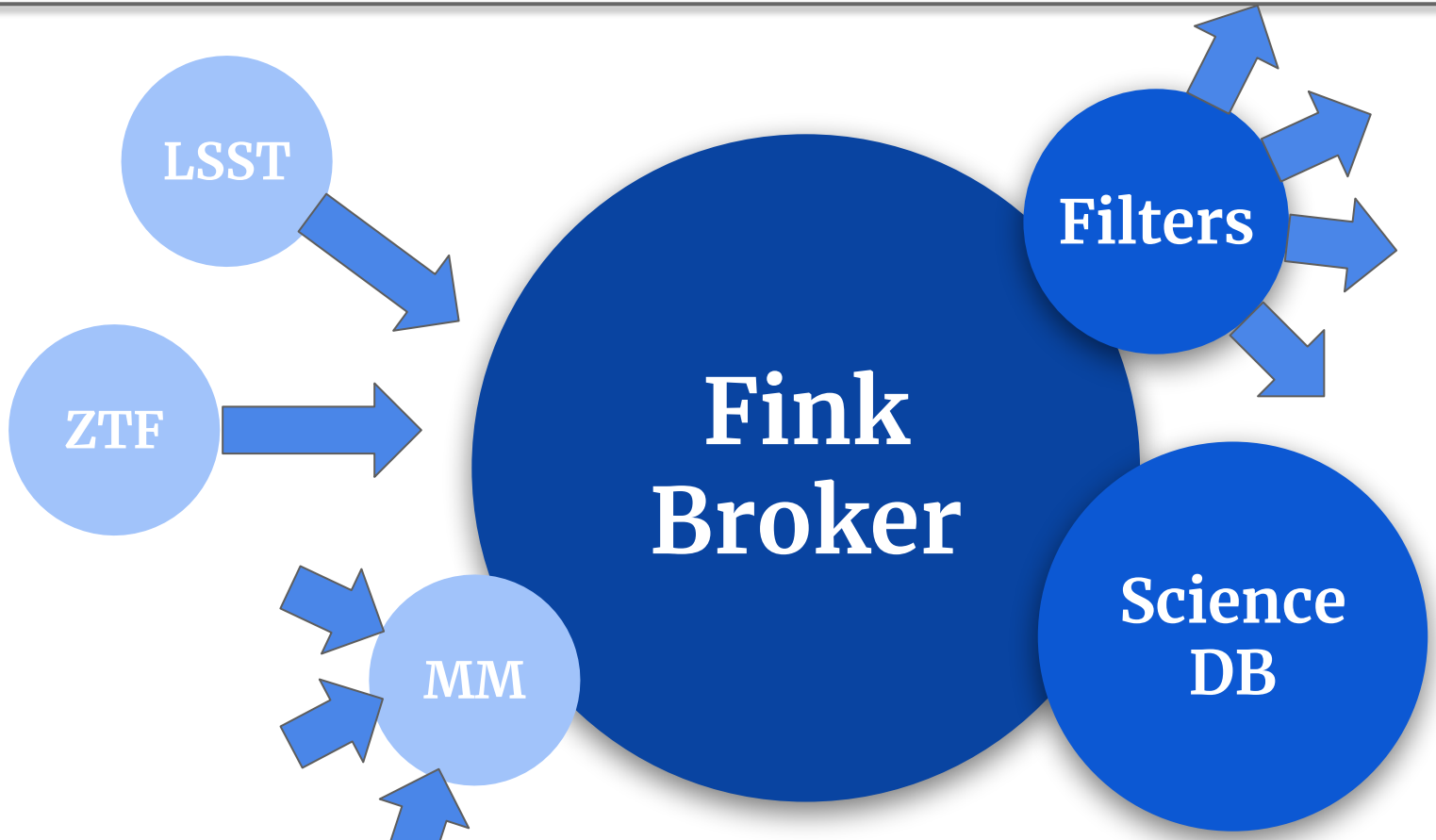
Fink challenges & design

Fink's design is driven by:

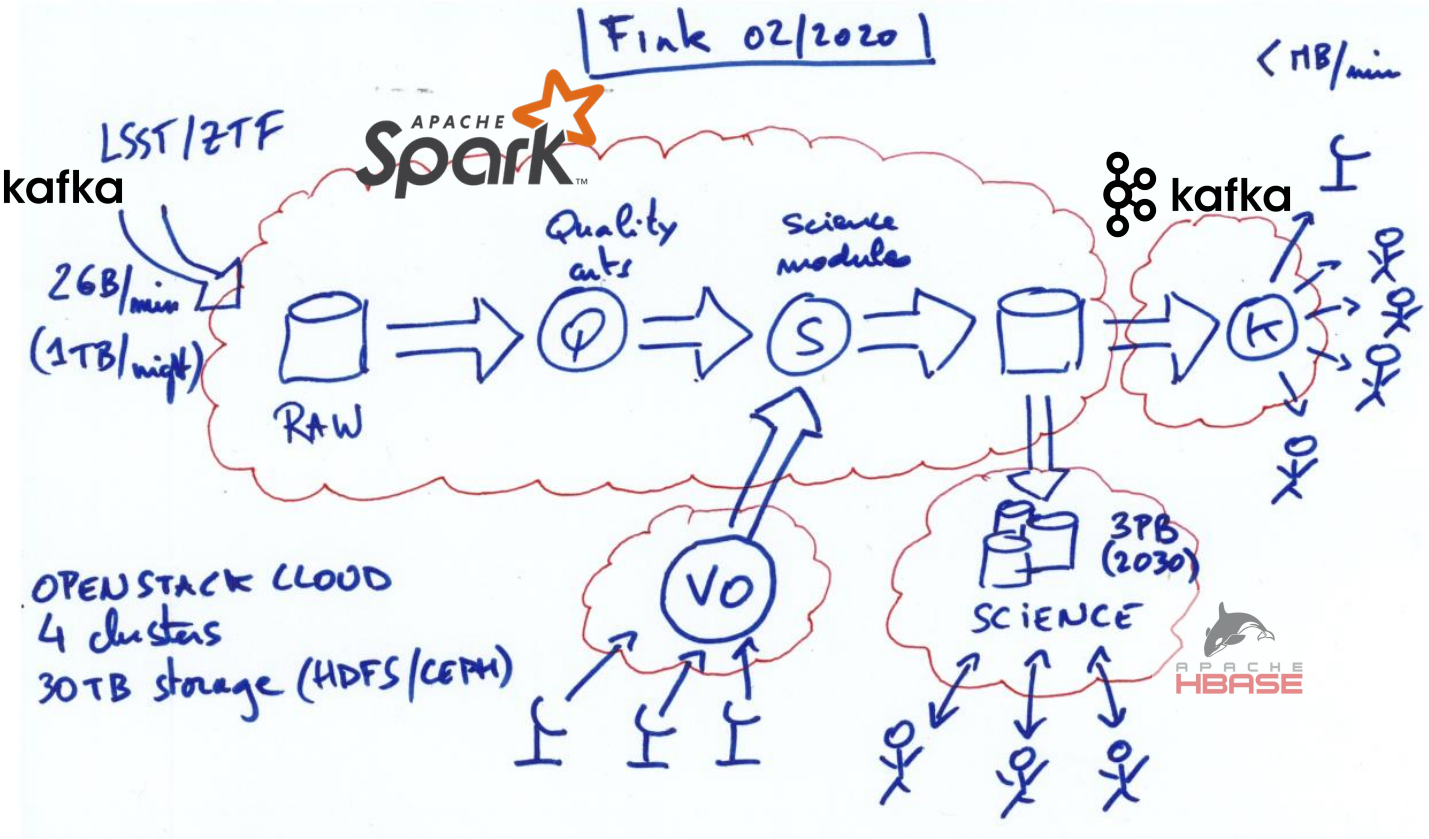
- Maximizing the scientific return on LSST and related experiments over the next decade: SVOM, CTA, Integral, KM3NET, ...
- Working efficiently at scales: real time and post-processing.
- Having a good integration with the current ecosystem: we are not alone!



Alert processing in Fink



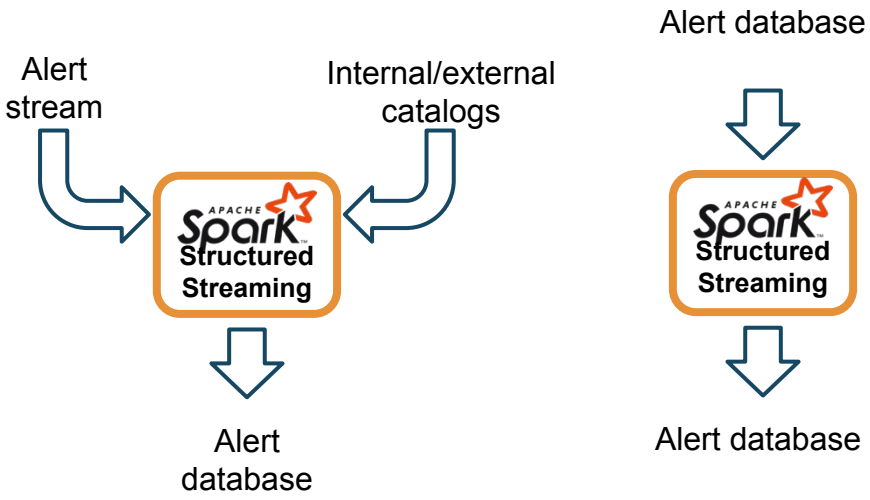
Alert processing in Fink

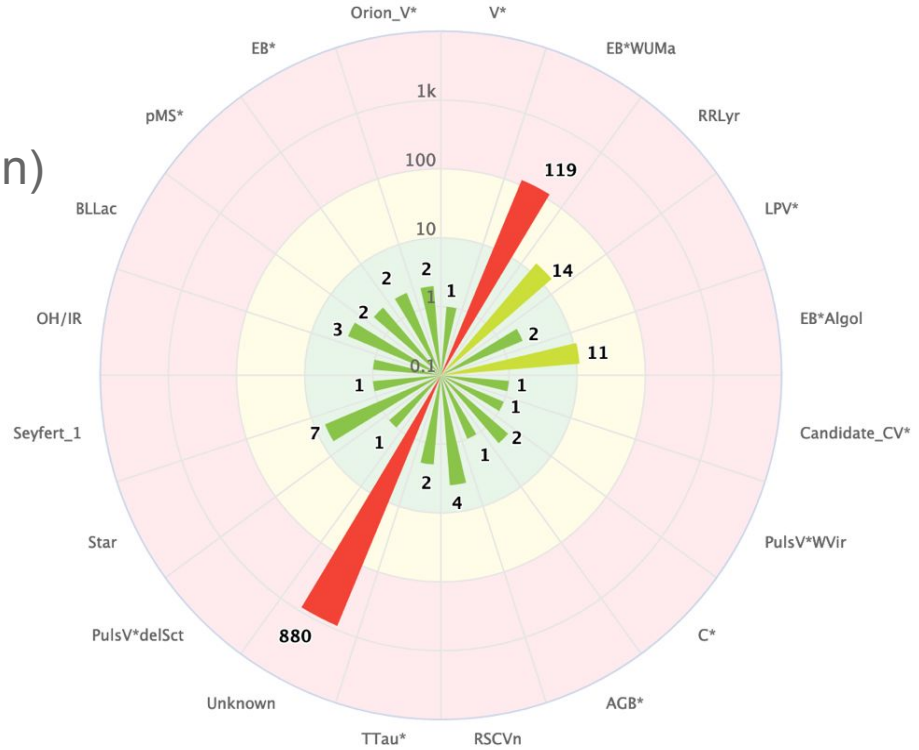


Science modules

Add values to the raw alerts

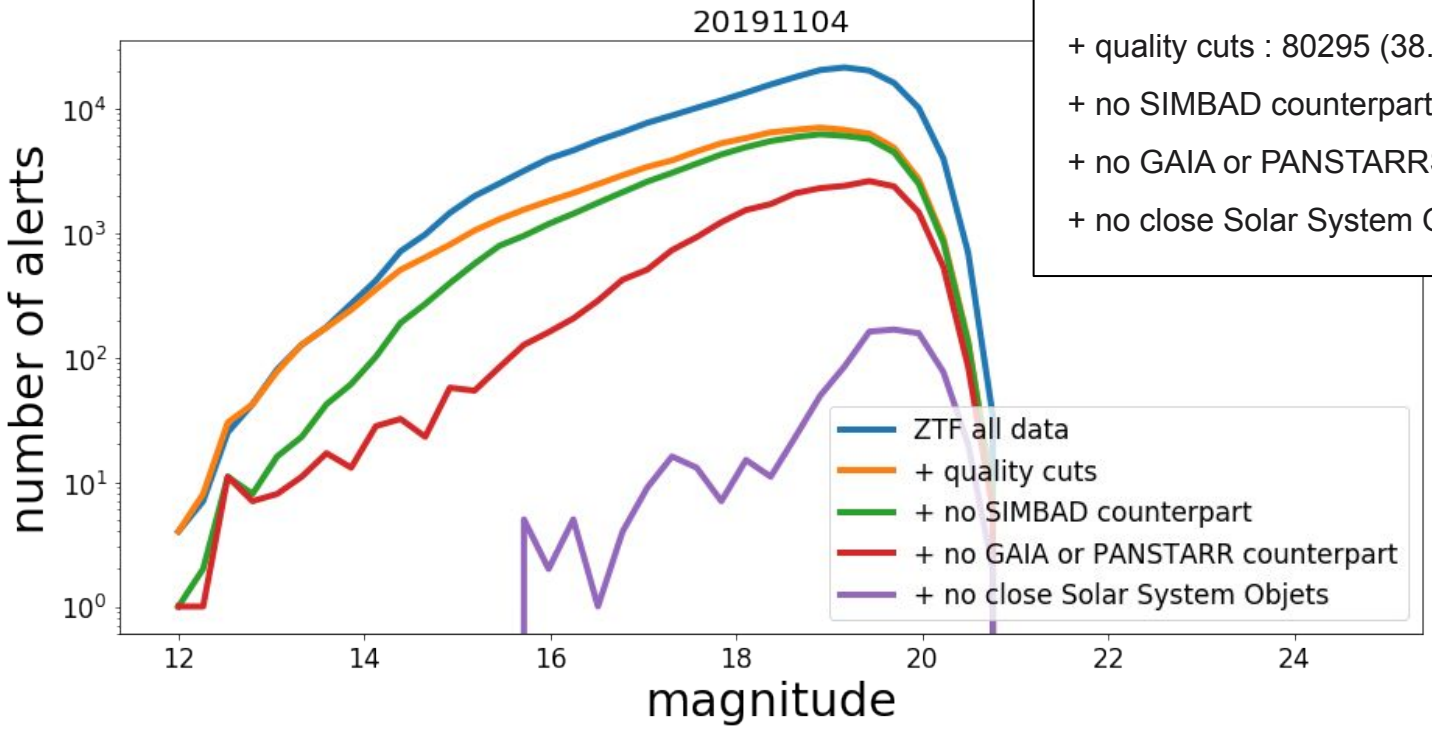
- Broker services (e.g. cross-match)
- User-defined modules (e.g. classification)





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%**)

More at the Fink Workshop on Thursday

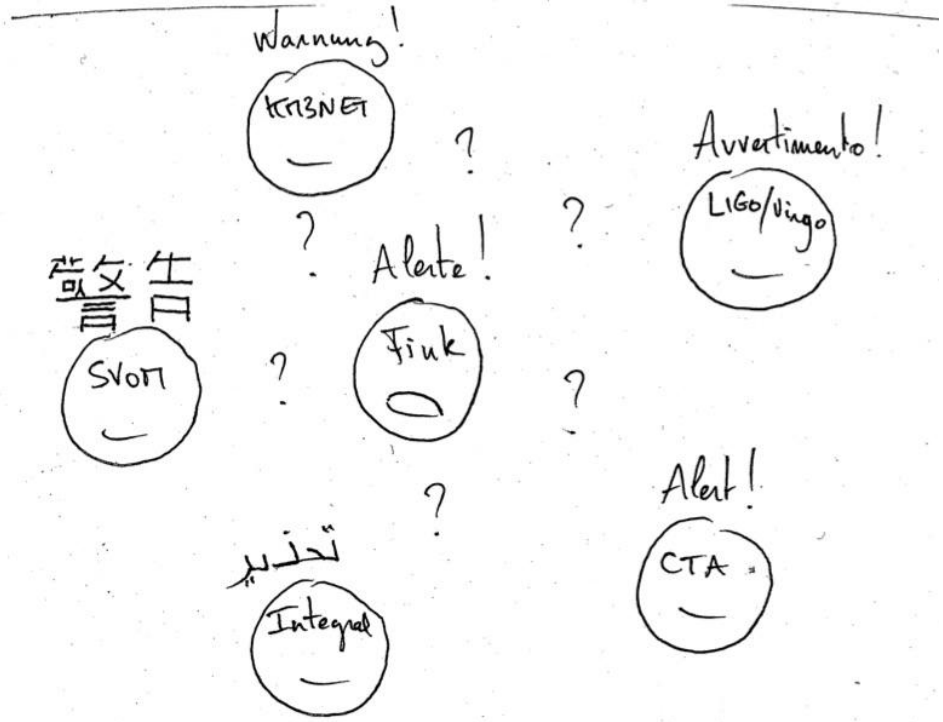
Joining information

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

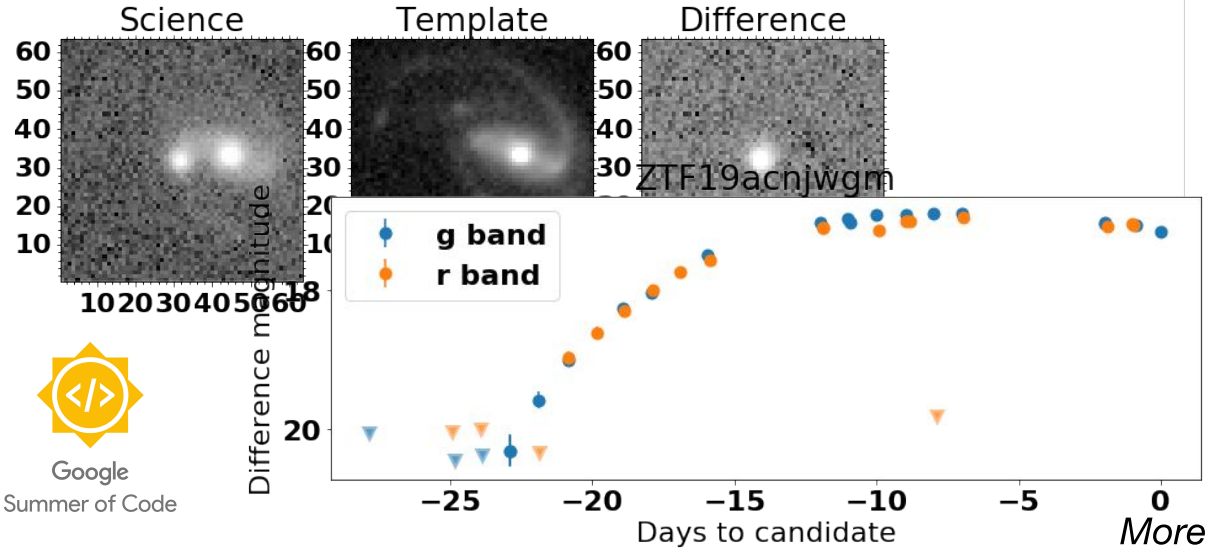
Status: (largely) experimental.



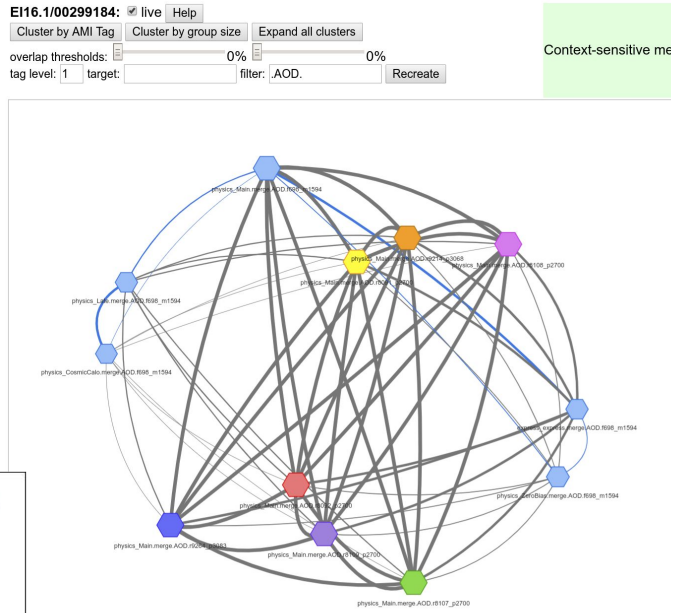
User interface(s)

Two entry points for users:

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



Google Summer of Code



J. Hrivnac, ATLAS

More at the Fink Workshop on Thursday

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)

Tests in near-real condition with good scalability:

- **LSST rate:** 10,000 alerts / 30 seconds (tested up to 10x this with margins).
- **LSST science content:** Alerts coming from ZTF (LSST pathfinder).
- **Science filters:** CDS xmatch, SN Ia like, ...

More information on the Calcul session.

Monitoring performances

Fink R&D projects

Continuous R&D projects to explore the outside world, e.g.

- Improving storage layer to enforce data integrity (C. Arnault)
- Introducing Graph DB for visualising data at Petascale (J. Hrivnac)
- Distributed Machine Learning to classify objects faster than light
- ...

Towards the full proposal...

Recent milestones: MoU with ZTF

- Fink joins AMPEL, ANTARES, ALeRCE, Lasair, …

Keep developing the client and the science DB access

- Improve your experience
- How to explore and visualise efficiently PB of data?

Keep integrating science modules

- Ongoing work with SN, GRB, ...

Improve Fink integration within the current networks and communities

Deploy the production version at CC-IN2P3