Time-domain astronomy at scale: lessons from a broker

Anais Möller on behalf of the Fink team
ARC DECRA Fellow at Swinburne

Rubin data products

Rubin Observatory (2025+)

- **20TB of images / night**
- **1TB of alerts / night**: x100-x1000 above current streams

Now
Raw Data
Sequential 30s image, 20TB/night

60s
Prompt Data Product
Difference Image Analysis
Alerts: up to 10 million per night

24h
Prompt Products DataBase
Images, Object and Source catalogs from DIA
Orbit catalog for ~6 million Solar System bodies

Year
Annual Data Release
Accessible via the LSST Science Platform &
LSST Data Access Centers.

End
Final 10yr Data Release
Images: 5.5 million x 3.2 Gpx
Catalog: 15PB, 37 billion objects

Rubin brokers

Rubin will send the full alert stream to seven brokers

- ALERCE, AMPEL, ANTARES, Babamul, Fink, Lasair, Pitt-Google

Serve a large scientific community by ingesting, **classifying**, filtering, and redistributing alerts. Classification is a community-driven effort.

Prototyping on ZTF (300k alerts/night), and testing deployment of the Rubin Alert Distribution system.
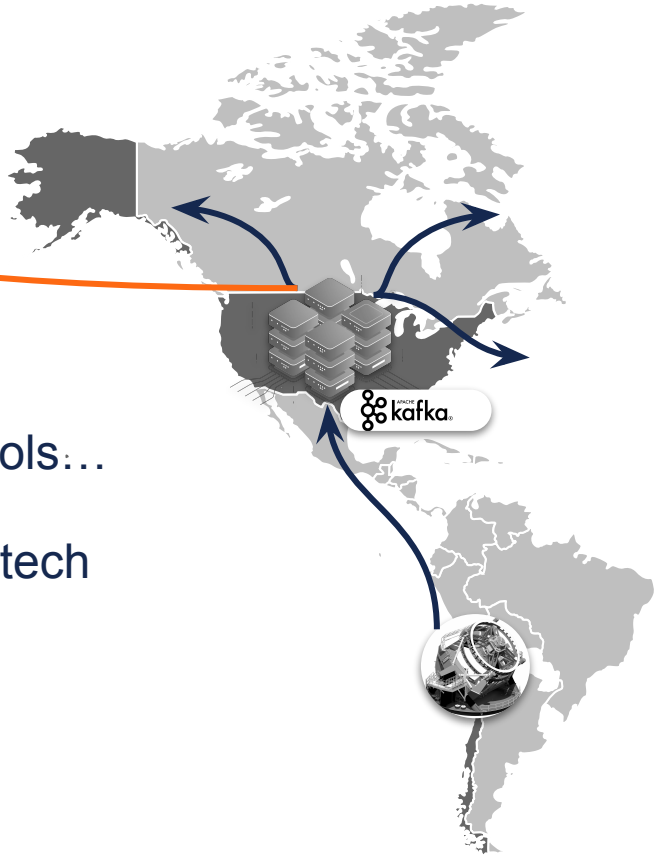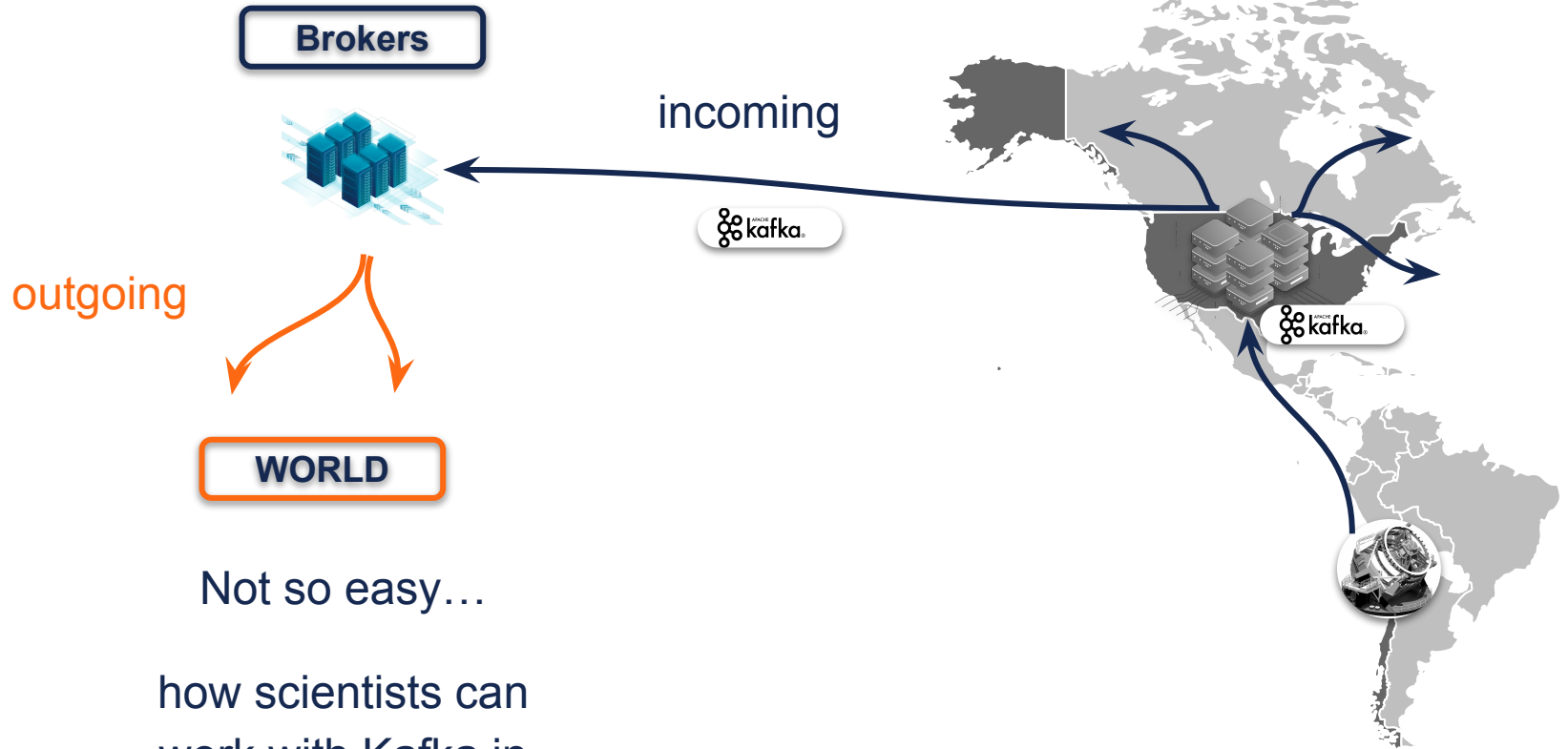
Brokers

WORLD

Brokers

WORLD

APACHE kafka
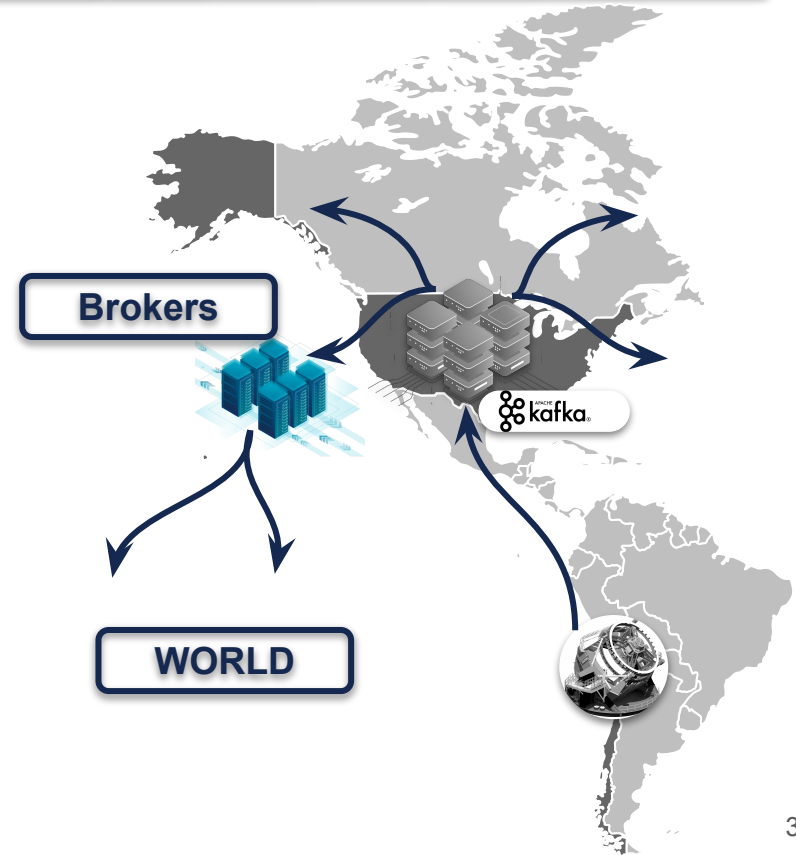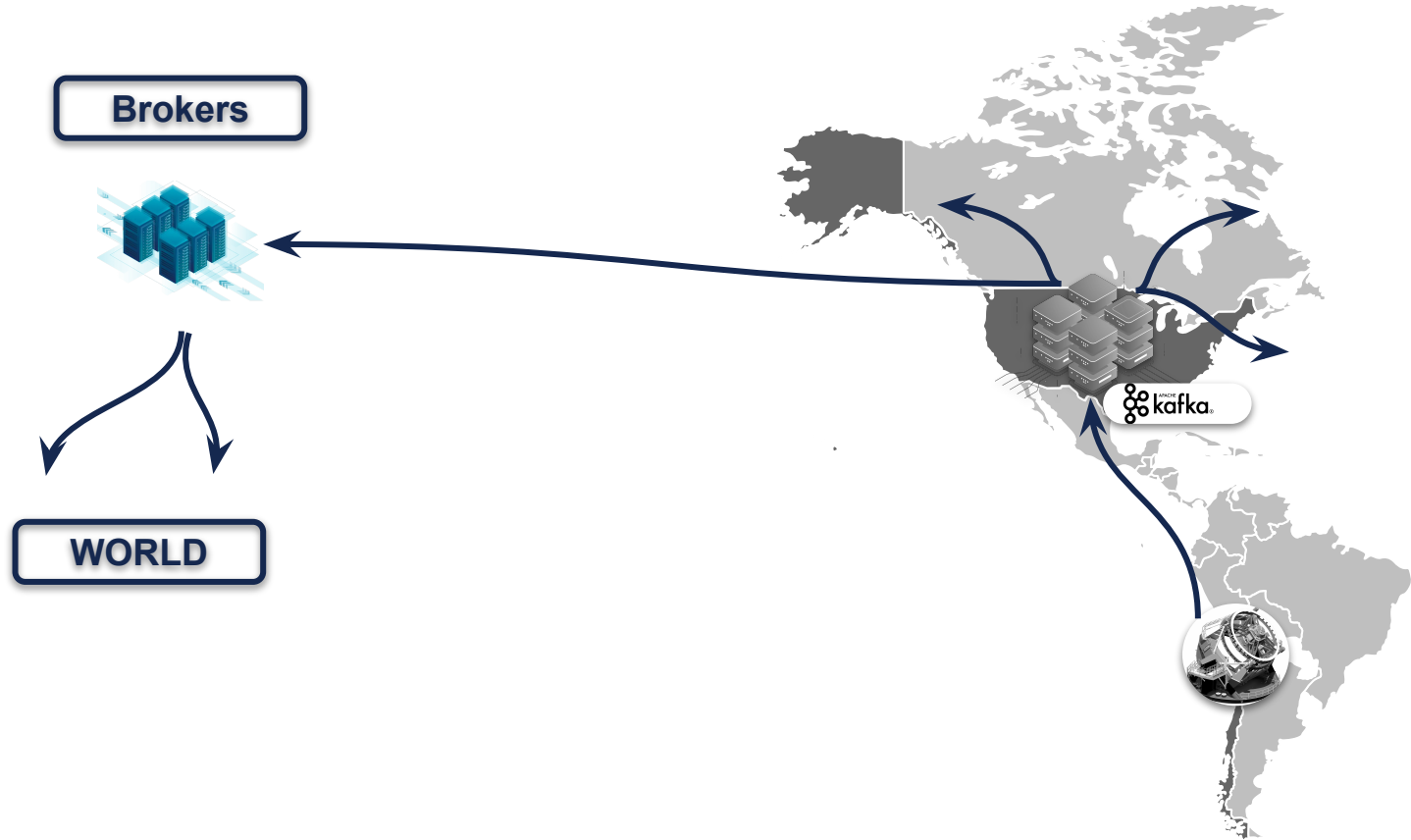
Brokers

incoming

Thanks to protocols…

"easy" because tech
talk to tech

Brokers

incoming

outgoing

WORLD

Not so easy…

how scientists can
work with Kafka in
practice?

Fink: cloud-based broker

60+ members:

- engineers and scientists
- >10 countries
- Engineer lead Julien Peloton

15+ scientific topics covered

Current: ZTF stream

Upcoming: Rubin

+ Connections to GCN and other astro streams
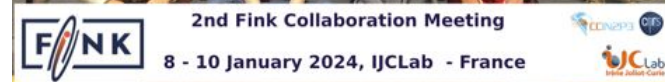

Photo: Carl Knox - OzGrav, Centre of Excellence for Gravitational Wave Discovery
3-5 May 2023, Melbourne - Australia

2022 Fink Hackathon
19-26 November
Grimentz, Swizertland
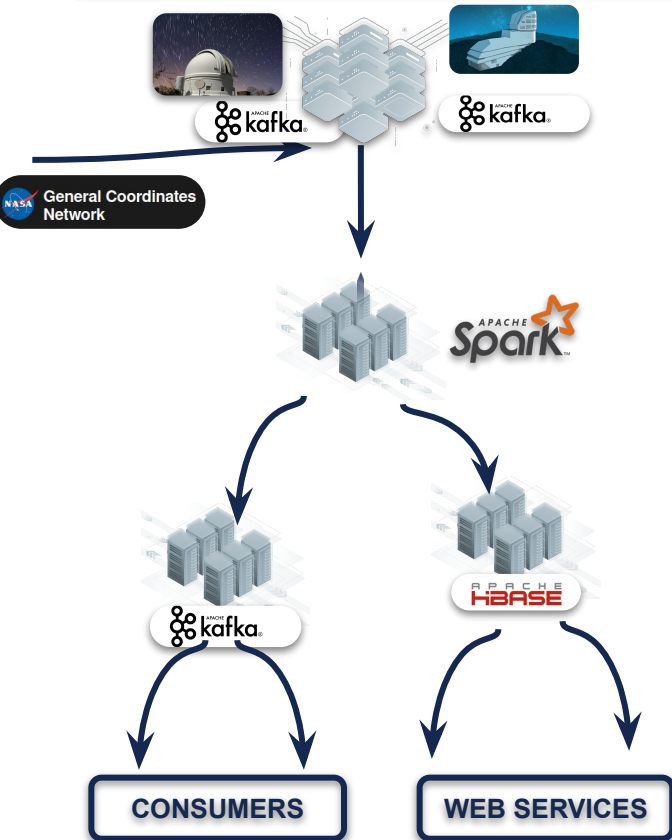

2nd Fink Collaboration Meeting
8 - 10 January 2024, IJCLab - France

Fink: cloud-based broker

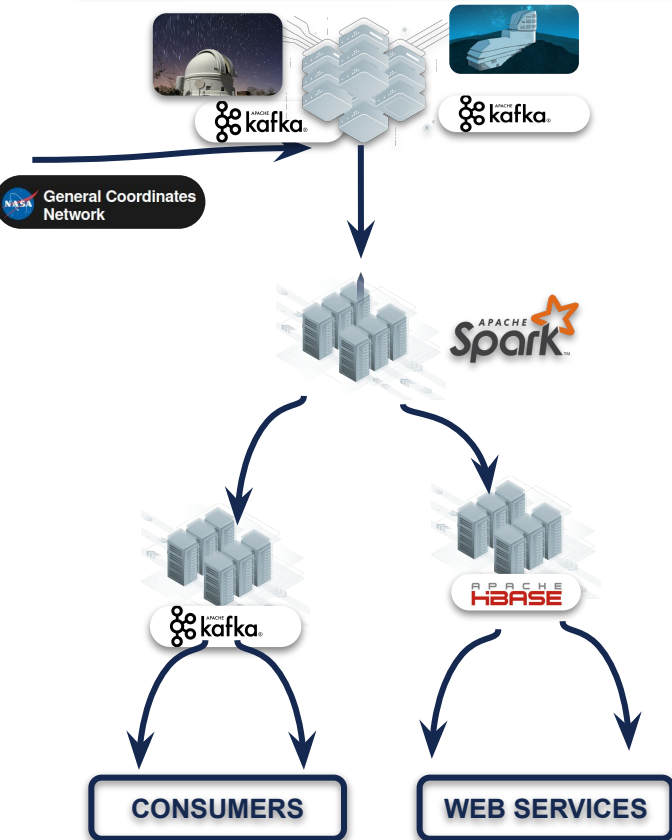Services deployed on large OpenStack clouds (UPSaclay 15k vCPUs & CC-IN2P3)

- Computing (Spark), database (HBase), streaming (Kafka), storage (Ceph & HDFS)
- Orchestration: Mesos & kubernetes
- Autoscaling based on the load

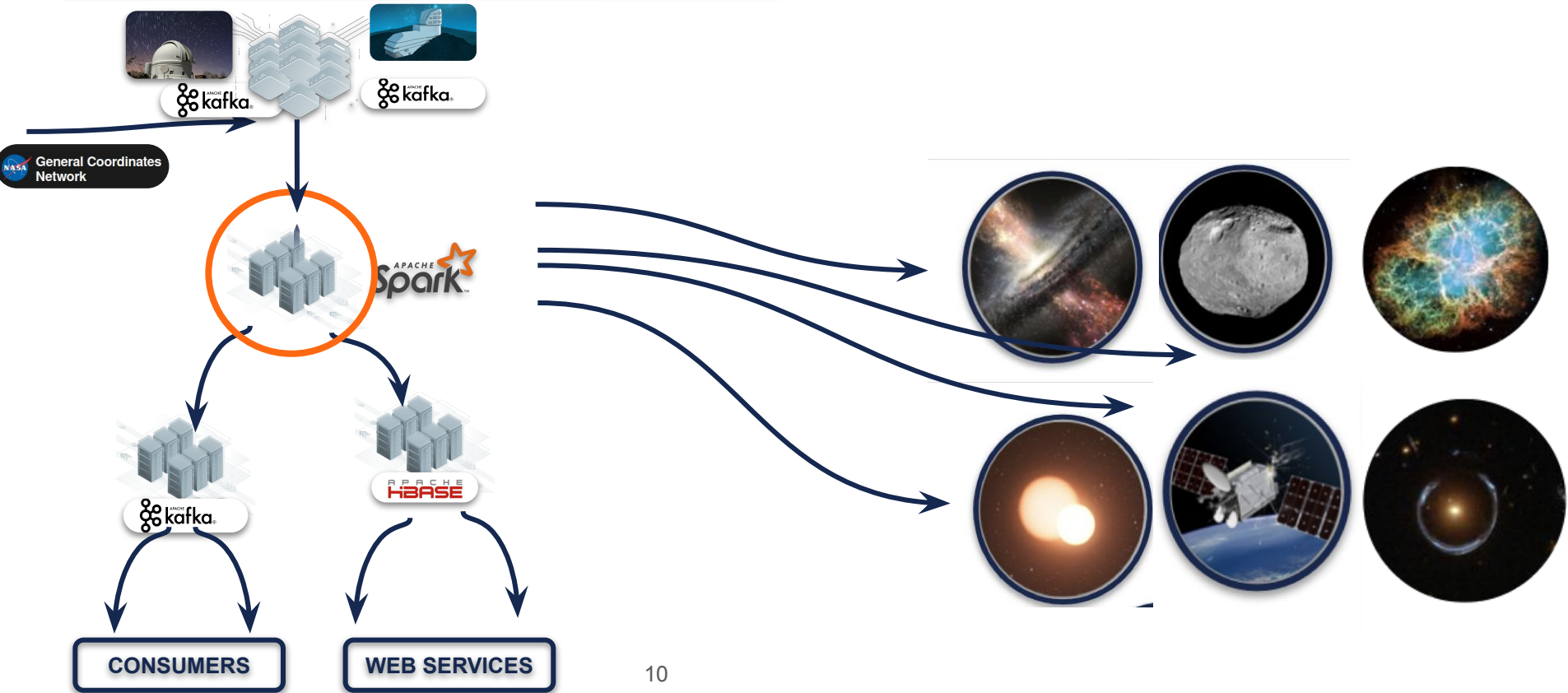Operating 24/7 since 2019, serving 100 unique users per day (scientists & follow-up facilities).

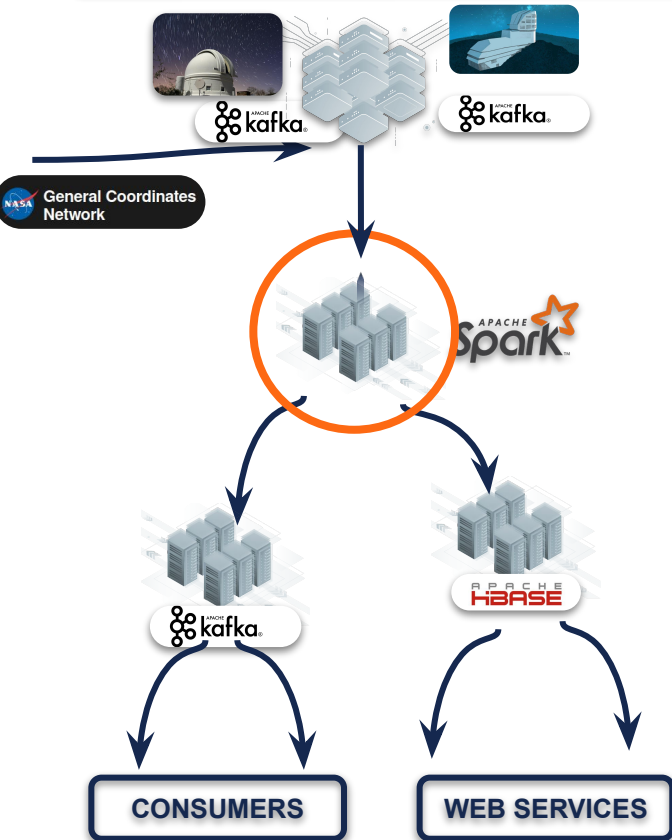Current rate @ 300k alerts/night (tested up to 50M/night). Science database of 8TB (250M events).

8

Fink

CONSUMERS

WEB SERVICES

9

Fink: computing challenge

General Coordinates Network

CONSUMERS   WEB SERVICES

10

Fink: computing challenge

Domain experts are the crucial agent for scientific discoveries

- Huge legacy of codes…
- … but they rarely meet computing requirements

Stronger interplay between the computing model & user software

- Software engineering role is increasing
  - Tailored service to integrate codes developed by the community
  - Infrastructure should be created to adapt to specific user needs
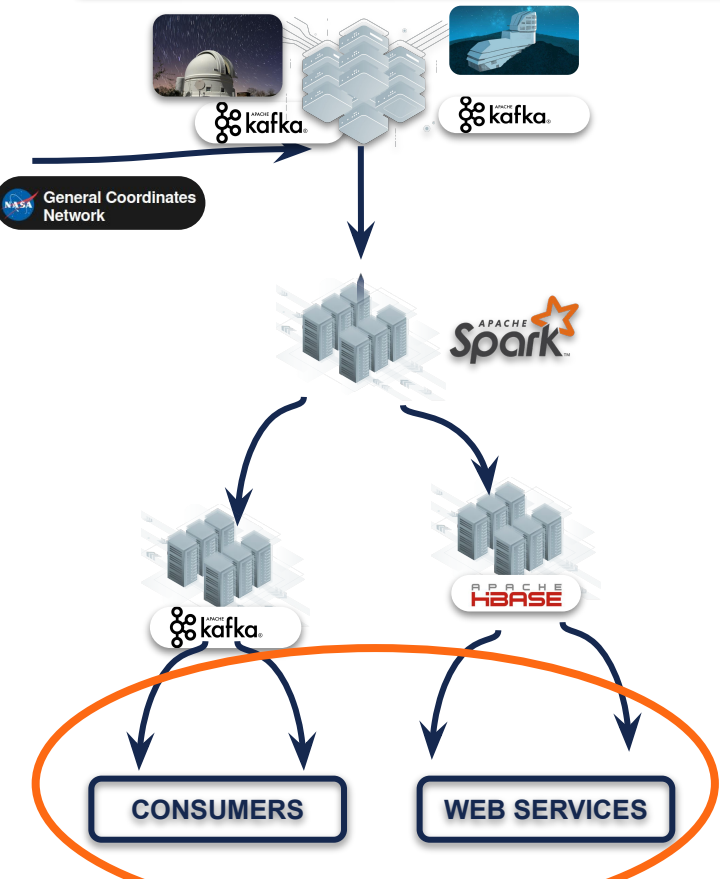
11

Challenge: open & big data
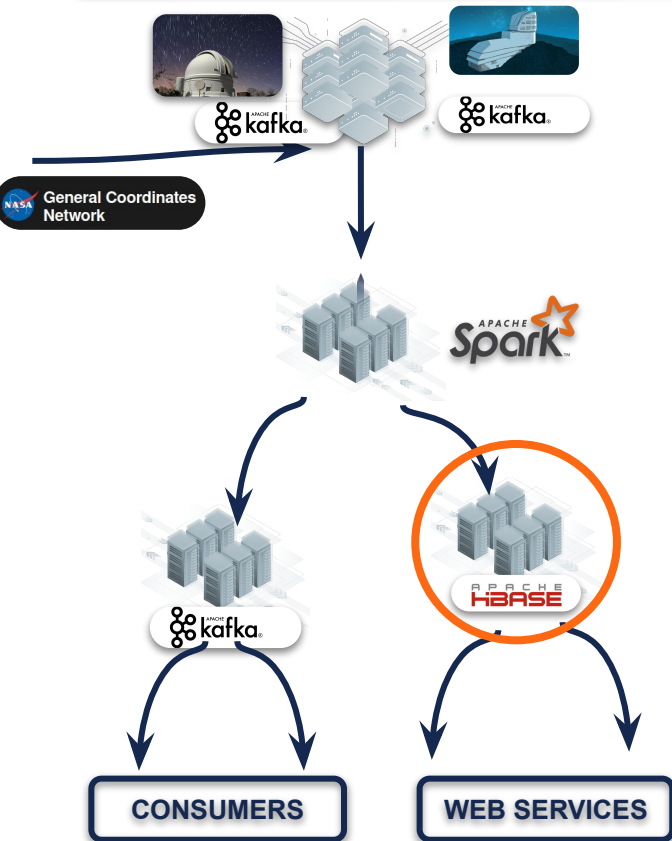
- How to expose a 3PB database to the world efficiently?
  - Multi-indexing is not trivial
  - Random access is cheap, but exotic processing requires computing resources
  - Exposing is not enough
- Non relational and graph oriented databases are explored
  - Multidatabase (Hrivnac, CHEP 2023), inspired from ATLAS

Challenge: users
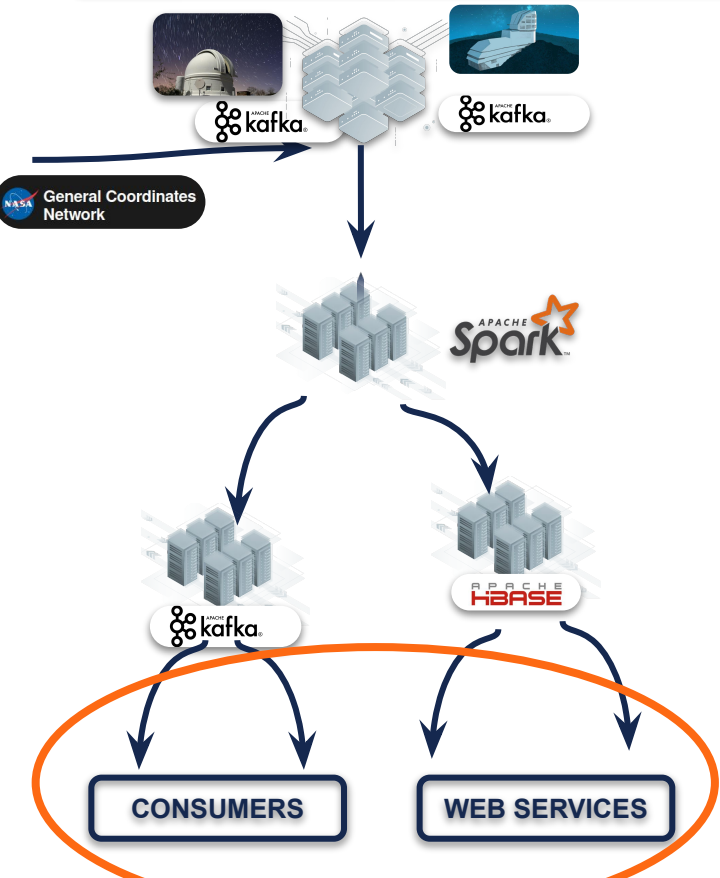
Challenge: users

Kafka, Spark are great but they are obscure to users!

How scientists can work with Kafka in practice?

1. Bring interfaces to existing tools (e.g. TOM Toolkit, SkyPortal, ...)
2. Fink data service: Kafka to stream data. Processing on the fly (to not download huge amount of useless data → Apache Spark)

14

Challenge: interoperability

- ○ Are current standards suitable at PB scale?
- ○ How to support science platforms with analysis close to data?
- ○ How to support new data-types driven by growth in size and complexity of data sets?

[Astro-COLIBRI 2](#)
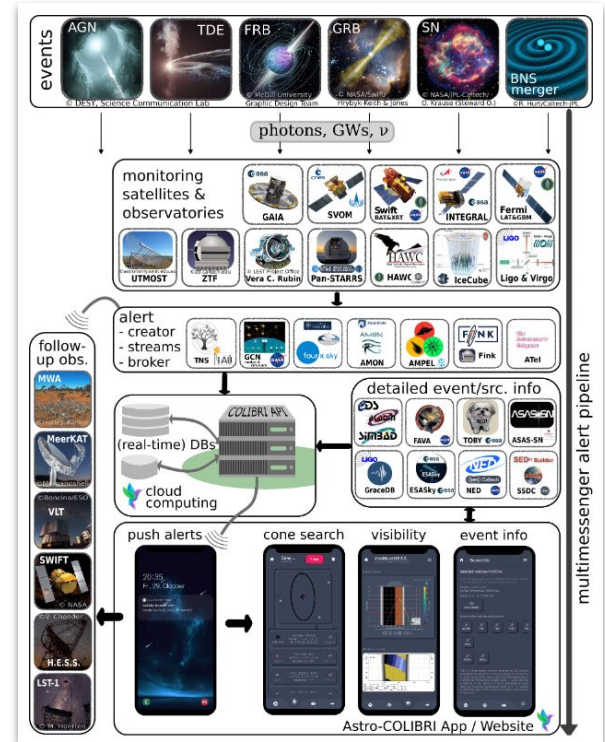
Conclusion

Challenge for alerts is dominated by the computing (networks are fast)

- survey → brokers → scientific community (defines the science roadmap)

Fink: processing is centralised, science is decentralised

- Cloud computing, allows to scale out resources
- Brokers provide data, computing, storage & web services for the community
- Open source backbone of the structure

Various challenges remain:

- user-driven & evolving analysis, open & big data, interoperability for multi-messenger & multi-wavelength analyses…
- How can broker and IVOA roadmaps can be synergetic?

- contact@fink-broker.org