Guidance for Community Broker Development

Beth Willman, LSST Deputy Director
Eric Bellm, Alert Production Science Lead
LSST SAC Meeting

August 14, 2017
Talk outline

High-level policy and technical considerations - Beth

Alert format and testing - Eric
LSST’s Alert Stream

LSST will generate 10 million alerts per night. LSST will offer a limited filtering service for these alerts to our data rights community.

It is expected that Community developed and operated Brokers will also be utilized by the community for specialized analyses of the alert stream. Alerts will be world-public, so Community Brokers can deliver LSST alerts to anyone.

The SAC has requested the Project to develop guidelines for Community Brokers. In this talk – we provide an overview of policy and technical factors, and propose a timeline and process for publishing guidelines.
Why will there be requirements on brokers?

Networking resources - Each full alert stream requires 1 gbs bandwidth

Staffing resources - It is anticipated that community brokers will require some level of support from the LSST operations team

Science integrity – We would like reliability for the community

World-public – We would like to prioritize resource investment into brokers that will serve a broad community
Networking Capacity – How many brokers?

Initial sizing estimates supported only four full streams due to network bandwidth limitations.

Current capacity projections indicate a larger number may be possible, but with some uncertainty.

- Depends on actual network capacity and usage patterns
- Depends on details of the alert format
- Depends on the achieved purity of the alert stream

Don’t think we can make a final decision on the number of brokers until we have at least processed data from ComCam.
Possible Requirements/Factors Impacting Prioritization

- Demonstration of availability of appropriate technical and personnel resources
- Potential to add scientific that serves a large community, enables high-profile science, and/or provides unique capabilities
- Willingness to make products world public (preferred)
- Integration with follow-up resources and/or the broader time-domain ecosystem
- Demonstrated performance and/or community adoption on precursor streams
- Demonstration of scientific validity on precursor data or by data challenge
- Applicability of international agreements, if relevant
Community brokers will need...

Large inbound and outbound network bandwidth (full alert stream is a few TB/night)

Petabytes of disk capacity

Databases handling of billions of sources

Compute resources to handle sophisticated classification and filtering tasks in real time at scale

Personnel to develop and maintain the service

What about Intermediate Brokers? – One can imagine brokers that only pull in a fraction of the full stream and are built to satisfy a specific community. Will this be the dominant mode?
What can the LSST “mini-broker” do?

LSST will host a basic filtering service at NCSA (ls.st/dpdd)

Enables user-uploaded simple filters that act on information within a single alert packet (no crossmatches to external catalogs)

Anticipate allowed capacity up to 20 alerts passed per visit per user

Limits on filter runtime and total alert throughput are expected
How will alert formats be developed?

Present VOEvent 2.0 standard and VOEvent Transport Protocol currently insufficient for LSST’s needs.

- No ability to embed cutouts in the packet
- XML format is space-inefficient
- No ability to validate packet contents
- Lifetime VOEvents handled by SkyAlert is 1/10th of one night of LSST

Now prototyping potential new alert format based on Apache Avro and new transport mechanism based on Apache Kafka.

- Zwicky Transient Facility (ZTF) survey plans to deploy a fork publically in 2018 (see session Thursday)

Expect to evaluate community experience with ZTF stream and work with VO to develop ”VOEvent 3.0” in ~2019.
What resources for testing are expected?

Next year, plan to make available a testbed alert distribution system with sample canned datasets (e.g., DECam) using the prototype formats.

(ZTF alert stream is also expected to be available in that timeframe)

During commissioning, alerts from ComCam and LSSTCam will be made available with substantial latency

driven by structure of commissioning observations
and the need to commission relevant pipelines
What is the proposed timeline?

March 2018: present draft policy document to SAC for feedback
August 2018: obtain Project approval for policy document
Mid 2018: Set up test alert stream with sample precursor data
2019: VO interactions to develop format
2020: Call for proposals for community brokers
Sample commissioning alerts with ComCam
2021: Finalize number of full streams
Sample commissioning alerts with LSSTCam
Selection of LSST brokers for early operations
2022: Begin operations

It is reasonable to expect periodic review of broker performance and usage during LSST Operations.