Contrib. Astron. Obs. Skalnaté Pleso **55/2**, 95 – 105, (2025)
https://doi.org/10.31577/caosp.2025.55.2.95

An overview of astronomical transient brokers in Rubin era

V. Vujčić[1], V.A. Srećković[2], S. Babarogić[3] and J. Aleksić[1]

[1] *Astronomical Observatory, Volgina 7, 11060 Belgrade, Serbia (E-mail: veljko@aob.rs)*

[2] *University of Belgrade, Institute of Physics Belgrade, PO Box 57, 11001 Belgrade, Serbia*

[3] *University of Belgrade, Faculty of Organizational Sciences, Jove Ilića, 11000 Belgrade, Serbia*

Received: December 12, 2024; Accepted: December 12, 2024

Abstract. Vera C. Rubin Observatory, formerly known as the Large Synoptic Survey Telescope (LSST) is approaching its operational phase, and all actors in the scientific/software ecosystem surrounding it are in the last phases of preparations. One of the types of the data that Rubin will publicly deliver are real-time alerts - a data stream of transient astronomical events. These alerts will be disseminated to the public through seven project-approved 'community brokers' - applications used to ingest data stream and provide various additional functionalities. In this review paper we compare software designs and analyze features of seven Rubin community brokers - ALeRCE, AMPEL, ANTARES, Babamul, Fink, Lasair, Pitt-Google.

Key words: astronomical transients – large astronomical survey – lsst – real time event processing – stream processing – data

1. Introduction

Both data suppliers, intermediary software and tool providers, and end users are getting ready for the start of the data flow as Vera C. Rubin Observatory - Legacy Survey of Space and Time, formerly known as the Large Synoptic Survey Telescope (LSST, Ivezić et al. (2019)), nears its operational phase. Real-time alerts, or data stream of transient astronomical occurrences, are one of the two main types of data products that LSST will deliver. Alert stream will be publicly available, but because of the infrastructural and bandwidth limits LSST approved seven 'community brokers' - software platforms that ingest, transform, add value and redistribute the stream - to act as intermediary layer between the data stream and the scientific community. Most broadly defined, the task of the 'broker' is not to act as a definitive scientific tool but to be an intermediary agent, reducing the stream to smaller number of events with most potential for particular science case and offering automatization capabilities to programmatically connect to the next tool in the process. That being said, brokers differ in their functionalities and architecture, and the aim of this paper is to analyze and explain these differences.

Nightly pipeline - or Alert Production pipeline, one of several LSST data pipelines (Jurić et al., 2015), will process difference image data after it is obtained by the telescope. Each detection of an astronomical object which surpasses threshold of SNR>5 after performing subtraction from the template image will be treated as a potentially significant event and included in the alert stream. Every alert includes a package with LSST information about the source, including photometry, coordinates, and picture cutouts (Graham et al., 2019a). The stream will be broadcast in near-real time with a 60-second latency to account for processing time and source association.

LSST plans to ship alerts in Apache Avro serialization format with 12-month history, with approx 82KB size per alert packet - additionally LSST may ship 'lite' packets, drastically smaller without history and cutout. Broadband rate of the stream is estimated to be between 0.2 Gbps, and 5 Gbps, and this was the main reason LSST decided to limit the number of direct consumers by soliciting external(community) 'brokers'. After nine teams submitted full proposals (in 2020), seven were selected to receive direct full stream and two a downstream through an intermediary. Majority of selected brokers are developed and tested using data stream from Zwicky Transient Facility (ZTF Graham et al. 2019b) - a precursor to LSST at 10% data volume, which shares similar alert schema, serializaton format and streaming platform (Apache Kafka). LSST makes a distinction between a diaobject (dia as in difference image analysis), an astronomical object, and a diasource - a single measurement of the object, with one-to-many relationship. This terminology may differ in individual broker's documentation (object− >alert, locus− >alert, etc.) but the cardinality stays the same.

In this review we compare software designs and analyze features of seven Rubin community brokers - ALeRCE, AMPEL, ANTARES, Babamul, Fink, Lasair, Pitt-Google.

2. Overview of LSST community brokers

Brokers (a term long ago accepted in the astronomical community (Borne, 2008), more precisely stream/event processing engines) are intended to deal with real time, high throughput (big data is term more commonly used in the community) input streams of astronomical events.

2.1. Alerce

ALeRCE (`https://alerce.science/`, `https://github.com/alercebroker`) - Automatic Learning for the Rapid Classification of Events - is a Chilean event processing platform, rich in features, with online services and Python API. It includes the real-time ingestion, aggregation, cross-matching, machine learning (ML) classification, and visualization (see Förster et al., 2021). ALeRCE performs two ML classification methods - a Balanced Hierarchical Random Forest (BHRF) on lightcurve features Sánchez-Sáez et al. (2021) and convolutional neural network classifier of cutout images - stamps Carrasco-Davis et al. (2021). More recently, ALeRCE tested novel approaches including Deep Learning Transformer Cabrera-Vives et al. (2024) which outperformed ALeRCE's own BHRF at the ELAsTiCC campaign[1]. For the outliers that don't belong to the proposed taxonomy, ALeRCE introduces anomaly detector based on deep support vector data description / autoencoder neural networks (Perez-Carrasco et al., 2023). ALeRCE offers a variety of services: a web portal wi for retrieval of astronomical objects via simplified graphical query with depiction of objects within a portfolio consisting of graphical lightcurves, classification assessments, cross-matching records, cutout stamps etc; a web portal used for real-time supernova (SN) candidates discovery updates, including visual spatial depiction; a Python client for interaction with Alerce databases and services (https://github.com/alercebroker/alerce_client); Target and Observation Manager - a tool for authorized users who can send follow-up requests; direct database access, Jupyter notebooks etc.

2.2. AMPEL

AMPEL (https://github.com/ampelproject, https://github.com/AmpelAstro) Alice in Modular Provenance-Enabled Land - alert management, photometry, and evaluation of light curves - is a scalable Python framework for general processing of large datasets, with specific application for astronomic alert streams (Nordin et al., 2019) developed at DESY, Hamburg. AMPEL is flexible - it can process real time or batch (existing) datasets, and extendable - it offers users to write additional code via python interface, or to easily plug an existing classification or analytical tool. AMPEL defines 'channels' as a subset of processes which belong to an individual user and is designed to avoid redundant processing, i.e resources will be shared between channels. Use of AMPEL is available through a fork or online live query API which requires pre-authorization.

2.3. ANTARES

ANTARES (https://antares.noirlab.edu/, https://gitlab.com/nsf-noirlab/csdc/antares/) - Arizona-NOIRLab Temporal Analysis and Response to Events System - is a broker which was conceived and funded long before the others, as early as 2014 (see e.g. Matheson et al., 2014). Its aim is to provide a web platform for filtering, flagging, categorizing and classifying relevant candidates and enable users to do science in various use cases. It has access through web portal which offers graphical querying, object portfolio, and defining custom science-enable filters written in Python. ANTARES can also be

[1] https://portal.nersc.gov/cfs/lsst/DESC_TD_PUBLIC/ELASTICC/

accessed via HTTP API, python client and webkit (Matheson et al., 2021). In the ANTARES system, 'locus' is a synonym for for an astrophysical object.

2.4. Babamul

Babamul (https://github.com/babamul/babamul), is conceived at Caltech as a lightweight and fully featured LSST broker but it seems like it was never fully developed or documented. A Babamul-related presentation was offered at the 2024 ESO - LSST workshop[2] where it was said that there is still a need for a low latency broker that is not heavy on hardware resources, without being specific about current status or plans for development. It was also mentioned that Fritz (https://www.ztf.caltech.edu/ztf-fritz.html), data management system within ZTF which was also developed at Caltech, would meet the broker function and handle LSST data rates. For the lack of more precise information, we will not further analyze Babamul in this work.

2.5. Fink

FINK Broker (https://fink-broker.org/, https://github.com/astrolabsoftware/fink-broker), an international project centered around French National Centre for Scientific Research, is a robust alert processor based on Apache Spark engine for main streams (ZTF, LSST) and Apache Kafka for transformation of secondary streams. It is designed to efficiently handle high throughput and different timeframes of diverse astronomical phenomena (Möller et al., 2021). Fink defines science modules as additional functionalities written in Python - such as cross-matching, machine and deep learning classifiers, aggregators and additional community modules. The modules operate in an adjustable system, where input of one module can be used as the output of other modules. Fink distinguishes alert classification (which it does) from object classification. Fink is available through science (web) portal - which offers querying and rich objects portfolios, Python API, as well as redistributed Kafka filtered stream.

2.6. Lasair

Lasair (https://lasair-ztf.lsst.ac.uk/, https://github.com/lsst-uk/lasair-lsst), an Edinburgh/Oxford/Belfast collaboration, is a well-tested platform for filtering, adding value and storing alerts (Williams et al., 2024). As formulated in the documentation, 'Lasair is a platform for scientists to make science; it does not try to make the science itself'. Lasair web portal introduces concepts of filters, watchlists and annotators which are shareable among users. Filters are written as SQL queries to be applied on the running stream. Watchlists are personalized lists of points of interest in the sky from which alerts are

[2] https://www.eso.org/sci/meetings/2024/lsst.html

collected. Annotators include third party added-value tools, such as ML classifiers. Lasair also uses an internal classifier - Sherlock (Young, 2023), which uses spatial context implemented through search of archival catalogues with a boosted decision tree algorithm to give a primary classification assessment and spatial association (Smith, 2019).

2.7. Pitt-Google

Pitt-Google Alert Broker (https://github.com/mwvgroup/Pitt-Google-Broker, https://github.com/mwvgroup/pittgoogle-user) is an University of Pittsburgh project running on Google Cloud (GC) and provides alert stream filtering, distribution, processing, analysis, value adding with the focus on providing broad public access and flexibility. Users can use GC services or move data out of the cloud /citepwood2024pitt. Also, users can define their preferred level of pre-processing by using Pub/Sub services. Pitt-Google is highly flexible and customizable and can serve as a stream replicator (as LSST restricts direct shipping of alerts) with or without basic filtering, as well as full scale analytical/classification tool, where users access data through GC platform. Pitt-Google will offer subscription models, mainly to cover the costs of the GC services.

3. Comparison

We compare conceptual design of brokers to the concepts of stream processing as defined in relevant literature, cross-examine their functional and technical features, and offer a glimpse into the performance.

3.1. Design concepts

While most brokers were conceived and developed in the last five years, the idea and surrounding concepts are around two decades old (e.g Borne (2008)). A dedicated conference 'Hot-wiring the Transient Universe', which started in 2007, includes discussions on numerous aspects of automatization of processes involving transient event data. In comparison to the design based on relevant event processing literature (Fig.1), where blue rectangles present internal event processing agents and yellow hexagons interoperable communication channels, most of the brokers are built on similar premises - applying simple filtering for narrowing selection, enrichment by adding historical data (such as spectrography or additional photometry) from internal or external DB's, applying machine learning tools and/or predefined patterns, selecting most interesting candidates for followup[3] and finally redistribute the altered stream in the same interoperable data format[4]. Some of them keep modular flexibility to skip or combine any of these steps. Somewhat in contrast to Fig.1, we can note prevalence of machine learning instead of pattern matching inference mechanism which can be explained with general tendency of ML/DL methods research within the astro community. We may remark here that while some methods offer significant accuracy, no ML classifier for significant phenomena such as Supernovae is valid to the point of conclusiveness (i.e it has to be verified further by a follow-up), and they are sometimes resource-heavy within the broker workflow. This trend shifted broker design perspective from high performance stream engines (designed to transform, reduce and apply inference mechanisms), to centralized cluster-hosted feature-rich scientific tools. It will take several years from now to see which paradigm is optimal.

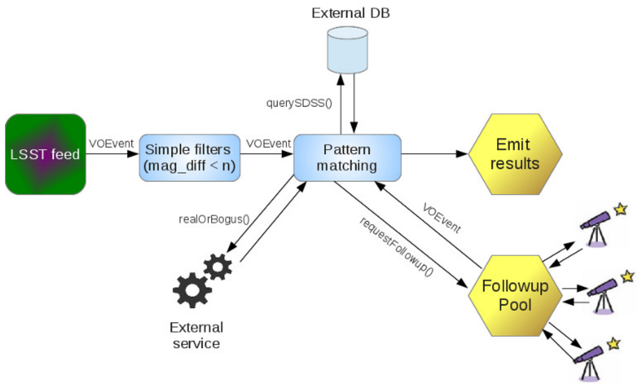

Figure 1. A proposed event processing network for real-time detection and classification of astronomical alerts - Figure from Vujčić (2014); Vujčić & Jevremović (2020), based on stream processing building blocks as defined in Etzion & Niblett (2010).

3.2. Functional and technical features

As we saw in the Sec. 2, functionalities of different brokers can vary or overlap and it's common to use different terms for the same concept. For example, AMPEL uses 4-tier event processing architecture, with following concepts - *add*,

[3]Software tools that manage 'Followup Pool', i.e that coordinate requests between brokers and telescopes/facilities for additional observation, are now called TOMs - from Target and Observation Manager. For more info, see https://lco.global/tomtoolkit/ and https://tom-toolkit.readthedocs.io/en/stable/

[4]In the late 2010's LSST switched from the International Virtual Observatory Alliance (https://ivoa.net) VOEvent XML/json standard for exchange of astronomical events to a more resource-efficient avro format (https://github.com/lsst/alert_packet).

combine, *complement*, *react* which correspond to design patterns of filter, correlate, enrich and pattern match as defined in literature (Tsimelzon, 2006). Core concepts of event processing (and synonyms used in brokers' documentation, in brackets), as defined in the seminal book on event processing, Etzion & Niblett (2010), include: ingestion, filtering, correlation (joining streams, multimessaging), enrichment (complement, annotate, query, cross-match, xmatch), matching (machine learning, classify, characterize, react) and distribution. Brokers often add some of the features taken over from other types of software, such as rich UX, object portfolios, programmable/extensible APIs etc. In this sense, functionalities of some brokers aim to push more ambitiously into scientific exploration and overlap with other multipurpose, integrated astronomical platforms such as SkyPortal (Coughlin et al., 2023).

For concrete cross-comparison of functional and technical aspects, we show Table 1. We compare brokers across twelve operational and infrastructure categories. Where functionalities of all brokers overlap are most fundamental concepts - basic stream filtering capabilities, cross-matching (with internal storage/external access differences) and API's that offer programmatic extension to external tools (like custom analytic tools or TOMs/Marshalls). Also, all of the brokers are open-source and are published at online git repositories. Most brokers are free - Pitt-Google will work on a subscription model, charging for cloud services and some (like AMPEL) offer limited access to the 'live' version running on their own resources. Most use machine (deep) learning algorithms for classification estimate - here we make a distinction between the teams who made efforts to develop a variety of detailed methods ('rich'), a single/provisional method(s) ('basic') and an option to include other's methods ('external'). Only Fink and Google-Pitt base their back-end on a proven industry stream processor platform, while all of the brokers apply cluster storage solutions, whether relational, document or a combination of both. ALeRCE, ANTARES, Fink and Lasair each developed a convenient web-portal with options for online querying and portfolios of astronomical objects, among other project-specific features. ANTARES, Fink, Lasair and Google-Pitt can redistribute alerts in the original avro format. Although built on different premises, it could be said that AMPEL, Fink, Lasair and Google-Pitt share concept of modularity, offering users/developers to choose which aspects or functions of the system will they include in their own workflow.

3.3. Performance

There were two LSST broker challeges in 2022 and 2023 called ELAsTiCC[5], with taxonomy tree defined by the ELAsTiCC team (Malz & Knop, 2022). where "all brokers have demonstrated the ability to classify objects in less than one day, and often in less than 3 hours" (Knop & Team, 2023). During June

[5]Extended LSST Astronomical Time-series Classification Challenge

2024, Operations Rehearsal 4 Rubin Observatory processed simulated ComCam images sent from the summit and sent live alerts to brokers where all brokers operated at the same time. However, the purpose of the test was not to evaluate performance metrics in any specific sense, but rather to demonstrate that the series of processes operates successfully.

According to the private communication with some of the authors of the brokers, including Julien Peloton of Fink and Roy Williams of Lasair, attempting direct comparisons between brokers in order to compare some measurable value is a common misconception. The brokers associated with the Rubin Observatory are not alternative implementations designed for the same scientific objective but distinct implementations tailored to diverse scientific goals that often do not overlap. According to members involved, the questions that brokers are designed to address are shaped by the specific needs and priorities of their respective user communities, each of which defines "performance" differently.

4. Conclusion

We gave the overview of the current pallette of brokers for transient astronomical events - software tools designed to act as an intermediary layer between high throughput astronomical streams and more specific scientific tools. Different brokers match in some aspects and diverge in other aspects of their functionalities and conceptual design. Our broad conclusion points would be: Lasair and ALeRCE offer richest web portal features and may be appealing to widest scientific community; ALeRCE team has put most effort into development of ML classifying methods, but that might prove not to be of primary importance as some brokers allow external classifiers; Fink and Pitt-Google have robust architecture which may be critical for latency/volume ratio of some scientific use cases; Fink, AMPEL, Pitt and Lasair show various ranges of systemic modularity, where AMPEL might prove most flexible in terms of low coupling of modules and workflow extensibility. Overall, Fink fulfills all functional, design and infrastructure requests that we analyzed - not necessarily being dominant in each one of them. Choice of the broker will depend on specific scientific use cases, flexibility for further refinement as the LSST survey proceeds and assignment/availability of resources.

Table 1. Comparison of community brokers' features

	ALeRCE	AMPEL	ANTARES	FINK	Lasair	Pitt-Google
Filtering	x	x	x	x	x	x
ML/classifiers integration	rich	basic		rich+external	basic+external	basic
Cross-match external dbs	x	x	stored internally	x	stored internally for classification, portal only for annotation	x
Web portal with object portfolio	x		x	x	x	via api only
Client API	x	x	x	x	x	x
HTTP/REST API			x	x		
Based on industry stream processor				apache spark		google cloud platform
Cluster storage	s3 + postgresql cluster	mongodb	cassandra + mysql cluster	hbase	cassandra + galera mysql	bigquery
Open source / online repo	x	x	x	x	x	x
Redistribution in the same format			x	x	x	x
Association with other real time streams				x		
Modularity		x		x	x	x

Acknowledgements. This research was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia (MSTDIRS) through contract no. 451-03-66/2024-03/200002 made with Astronomical Observatory (Belgrade), 451-03-47/2023-01/200024 made with Institute of physics Belgrade. The authors acknowledge the networking opportunities from COST Action CA22133 - The birth of solar systems (PLANETS) supported by COST (European Cooperation in Science and Technology).

References

Borne, K., A machine learning classification broker for the LSST transient database. 2008, *Astronomische Nachrichten: Astronomical Notes*, **329**, 255, DOI:10.1002/asna.200710946

Cabrera-Vives, G., Moreno-Cartagena, D., Astorga, N., et al., ATAT: Astronomical Transformer for time series and Tabular data. 2024, *Astronomy and Astrophysics*, **689**, A289, DOI:10.1051/0004-6361/202449475

Carrasco-Davis, R., Reyes, E., Valenzuela, C., et al., Alert Classification for the ALeRCE Broker System: The Real-time Stamp Classifier. 2021, *Astronomical Journal*, **162**, 231, DOI:10.3847/1538-3881/ac0ef1

Coughlin, M. W., Bloom, J. S., Nir, G., et al., A data science platform to enable time-domain astronomy. 2023, *The Astrophysical Journal Supplement Series*, **267**, 31, DOI:10.3847/1538-4365/acdee1

Etzion, O. & Niblett, P. 2010, *Event processing in action* (Manning Publications Co., United States)

Förster, F., Cabrera-Vives, G., Castillo-Navarrete, E., et al., The Automatic Learning for the Rapid Classification of Events (ALeRCE) Alert Broker. 2021, *Astronomical Journal*, **161**, 242, DOI:10.3847/1538-3881/abe9bc

Graham, M., Bellm, E., Guy, L., Slater, C., & Dubois-Felsmann, G. 2019a, *LSST Alerts: Key Numbers* (DMTN-102, URL https://dmtn-102. lsst. io, LSST Data Management Technical Note)

Graham, M., Kulkarni, S., Bellm, E., et al., The Zwicky Transient Facility: Science Objectives. 2019b, *Publications of the Astronomical Society of the Pacific*, **131**, 078001, DOI:10.1088/1538-3873/ab006c

Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al., LSST: From Science Drivers to Reference Design and Anticipated Data Products. 2019, *Astrophysical Journal*, **873**, 111, DOI:10.3847/1538-4357/ab042c

Jurić, M., Kantor, J., Lim, K., et al., The LSST data management system. 2015, *arXiv preprint arXiv:1512.07914*

Knop, R. & Team, E., Simulated Transient Alerts and Broker Classifications for ELAsTiCC. 2023, in *American Astronomical Society Meeting Abstracts*, 117–02

Malz, A. & Knop, R. 2022, ELAsTiCC taxonomy, https://github.com/LSSTDESC/elasticc/blob/elasticc2/taxonomy/taxonomy.ipynb, [Online; accessed 12-2024]

Matheson, T., Saha, A., Snodgrass, R., & Kececioglu, J., ANTARES: A Prototype Transient Broker System. 2014, in American Astronomical Society Meeting Abstracts, Vol. **223**, *American Astronomical Society Meeting Abstracts #223*, 343.02

Matheson, T., Stubens, C., Wolf, N., et al., The ANTARES astronomical time-domain event broker. 2021, *The Astronomical Journal*, **161**, 107, DOI:10.3847/1538-3881/abd703

Möller, A., Peloton, J., Ishida, E. E., et al., FINK, a new generation of broker for the LSST community. 2021, *Monthly Notices of the Royal Astronomical Society*, **501**, 3272, DOI:10.1093/mnras/staa3602

Nordin, J., Brinnel, V., Van Santen, J., et al., Transient processing and analysis using AMPEL: alert management, photometry, and evaluation of light curves. 2019, *Astronomy & Astrophysics*, **631**, A147, DOI:10.1051/0004-6361/201935634

Perez-Carrasco, M., Cabrera-Vives, G., Hernandez-García, L., et al., Alert classification for the alerce broker system: The anomaly detector. 2023, *The Astronomical Journal*, **166**, 151, DOI:10.3847/1538-3881/ace0c1

Sánchez-Sáez, P., Reyes, I., Valenzuela, C., et al., Alert Classification for the ALeRCE Broker System: The Light Curve Classifier. 2021, *Astronomical Journal*, **161**, 141, DOI:10.3847/1538-3881/abd5c1

Smith, K., Lasair: the transient alert broker for LSST: UK. 2019, *The Extragalactic Explosive Universe: the New Era of Transient Surveys and Data-Driven Discovery*, 51

Tsimelzon, M. 2006, Complex Event Processing: Ten Design Patterns, [Online; accessed 12-2024]

Vujčić, V., Use of complex event processing engines in astronomy. 2014, in *Big Data Conference proceedings, Szombately Hungary*

Vujčić, V. & Jevremović, D., Real-time stream processing in astronomy. 2020, in *Knowledge Discovery in Big Data from Astronomy and Earth Observation* (Elsevier), 173–182

Williams, R. D., Francis, G. P., Lawrence, A., et al., Enabling science from the Rubin alert stream with Lasair. 2024, *RAS Techniques and Instruments*, **3**, 362

Young, D. R. 2023, *Sherlock. Contextual classification of astronomical transient sources*, DOI:10.5281/zenodo.8038057