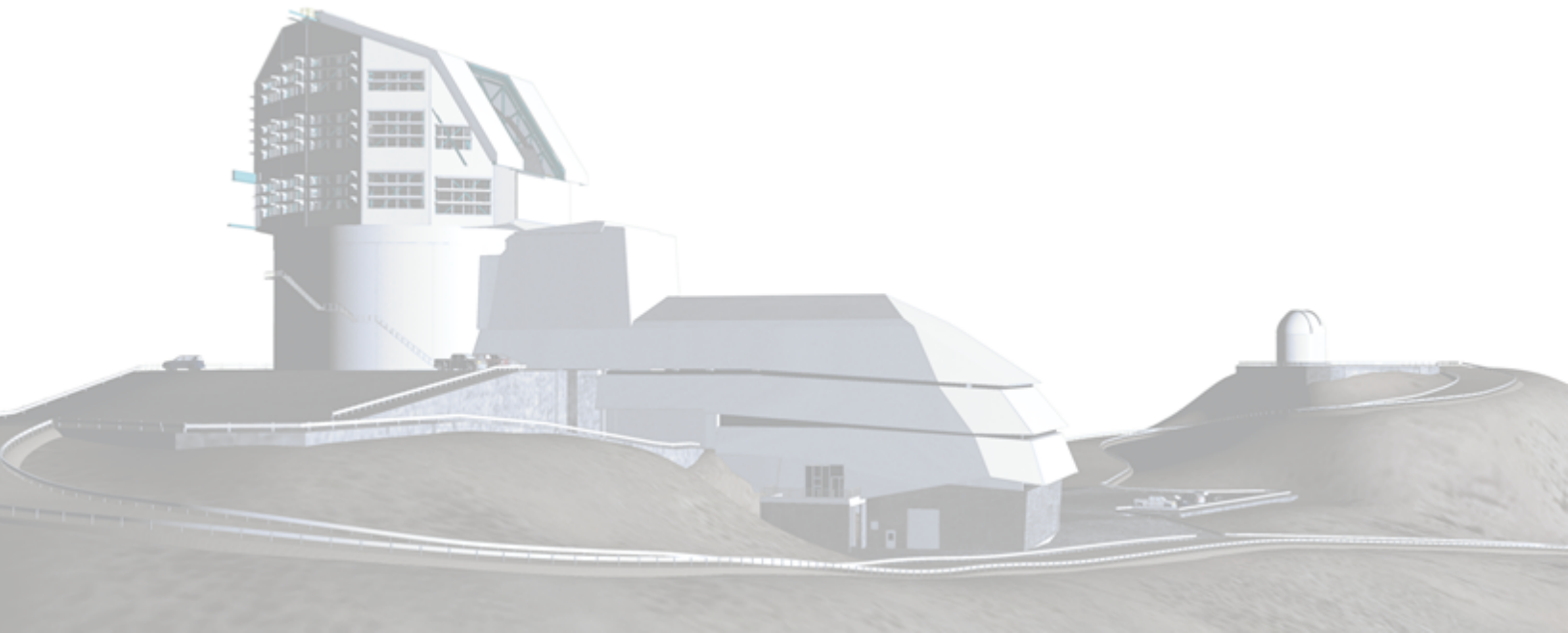
Vera C. Rubin Observatory
Data Management

LSST Alerts: Key Numbers

M. L. Graham, E. Bellm, L. Guy, C. T. Slater, G. Dubois-Felsmann,
M. Jurič and the Data Management System Science Team

DMTN-102

Latest Revision: 2024-04-26



Abstract

A quantitative review of the key numbers associated with the Large Synoptic Survey Telescope (LSST) Alert Stream.

Change Record

Version	Date	Description	Owner name
1	2019-02-19	Released.	Melissa Graham
1.1	2019-01-06	Minor updates, added glossary.	Melissa Graham
1.2	2024-04-09	Update estimates for moving objects.	Melissa Graham
1.3	2024-04-24	Release via Zenodo.	Melissa Graham

Document source location: `https://github.com/lsst-dm/dmtn-102`

Cite as: M. L. Graham, E. Bellm, L. Guy, C. T. Slater, G. Dubois-Felsmann, and M. Jurič. (2024). LSST Alerts: Key Numbers. Zenodo `https://zenodo.org/doi/10.5281/zenodo.11062349`.

Contents

LSST Alerts: Key Numbers

1 Introduction

The LSST Data Management System's (Data Management Subsystem (DMS)) Alert Production (Alert Production (AP)) pipeline will process new data as it is obtained by the telescope. Difference Imaging Analysis (Difference Image Analysis (DIA)) will be performed, and all sources with a signal-to-noise ratio >5 (in positive or negative flux) will be considered "detected", a record will be instantiated in the source catalogs, and an alert generated. Each alert is a packet containing LSST data about the source, such as coordinates, photometry, and image cutouts. For a full description of detected sources and alert packet contents, see LSE-163. The LSST alert stream will be delivered to several community-developed brokers, and also accessible to users[1] via an alert filtering service through the LSST Data Access Centers (DACs). Plans and policies for alert distribution are provided in LDM-612.

transSNR
LSR-REQ-0101
DMS-REQ-0269
DMS-REQ-0274

The purpose of this document is to quantitatively inform broker developers, and the broader scientific community planning to use alerts, on the key numbers regarding alert generation, distribution, and access via the LSST alert filtering service. The goals of this document are threefold: (1) to provide all of the key numbers regarding alert generation in one place; (2) to include any and all basis information, assumptions, and derivations that contributed to the key number; and (3) to be clear about whether each key number represents an estimate, a requirement, or a boundary. Wherever possible, the reference to a specific LSST requirement and any relevant requirement parameters are provided in the right-hand column. In this document we use 8 bits per byte (Byte (8 bit) (B)), and 1024 B per KiloByte (KB), 1024 KB per MegaByte (MB), and so forth.

2 Alert Stream

The concept and existence of the LSST alert stream was first introduced by the highest-level document, the LSST Science Requirements Document [LPM-17], which specifies that information about the detections of transient, variable, and moving objects be released promptly as a data stream.

[1] In this case, "users" is restricted to individuals with LSST data rights and access privileges.

2.1 Alert Release Timescale

It is a requirement that the DMS be capable of supporting the distribution of at least 98% of alerts for each visit within 60 seconds of the end of image readout[2].

OTT1
OTR1
LSR-REQ-0101
LSR-REQ-0025
OSS-REQ-0127
DMS-REQ-0004

This requirement applies to visits resulting in fewer than 40,000 alerts, and the term "distribution" includes all steps up to and including the transmission of the alert packet out of the LSST Data Facility (i.e., it does not include the time it takes for a broker to receive or ingest the alert). It is furthermore specified that all delayed alerts be made available at the next opportunity (LPM-17; see also the discussion regarding delayed/failed alert distribution in § 2.7).

2.2 Number of Alerts per Visit (and per Night)

It is a requirement that the DMS support the distribution of at least 10,000 alerts per standard visit[3] on average during a given night, and at least 40,000 alerts per single standard visit.

transN
LSR-REQ-0101
nAlertVisitAvg
OSS-REQ-0193
nAlertVisitPeak
DMS-REQ-0393

An extension of the above is that the DMS will support a long-term average number of 10^7 alerts distributed per night (assuming an average of 1,000 visits per night). It is furthermore specified that the performance of alert distribution shall degrade gracefully beyond these values, meaning that visits resulting in an excess of alerts should not cause any DMS downtime.

The value of 10,000 alerts per visit is a requirement on the DMS and not a scientific estimate of the intrinsic rate of transients and variables in the universe. However, estimates for the most common transients and variables can be derived from the Science Book (LSST Science Collaboration, 2009) by making some significant assumptions, as follows:

- Variable Stars: LSST is predicted to observe a total of ~135 million variable stars. Making the simple assumption that 20% (80%) of the stars are in extragalactic (Galactic) fields, and that of the ~18,000 degree; unit of angle (deg)^2 surveyed by LSST, 80% (20%) of the fields are extragalactic (Galactic), and that 10% of all variable stars are detectably variable at any given time, then a typical extragalactic (Galactic) field would yield ~1,800 (~28,800) alerts per visit [4]. Averaged over all fields, and weighted by 80% and 20% of the

[2]The design, minimum, and stretch values for the alert release timescale are 1, 2, and 0.5 minutes [LPM-17].
[3]The design, minimum, and stretch values for the number of alerts per visit are 10^4, 10^3, and 10^5 [LPM-17].
[4]Since "detectably variable" means "significantly different from the template", the value of 10% does depend

fields being extragalactic and Galactic, respectively, this is an average of 7,200 alerts per visit.

- Supernovae: LSST is predicted to observe a total of ∼10 million supernovae in 10 years, or ∼1 million per year. Since SNe are typically only visible for a few months, there might be ∼0.3 million detectable at any given time. Over 15,000 deg^2 of extragalactic survey area, that's ∼20 SNe deg^{-2} or ∼200 alerts for SNe per visit.
- Active Galactic Nuclei: LSST is predicted to observe millions of AGN. If ∼10% of them are detectably variable at any given time, then the estimate is that ∼0.1 million alerts over 15,000 deg^2 would generate ∼7 alerts deg^{-2}, or ∼70 alerts per visit for AGN.
- Moving Objects: The number of Solar System objects that LSST is predicted to observe is dominated by the 5.5 million main-belt asteroids. Due to their concentration along the ecliptic, estimates for the number of moving objects range from ∼400 alerts per visit on average but up to ∼5000 alerts per visit in the densest areas of the ecliptic.

Therefore, astrophysical estimates for the occurrence rates of alerts caused by the most common types of transients and variables yield ∼5,100 (∼32,000) alerts per visit in extragalactic (Galactic) fields, with an average of ∼10,500 alerts per visit.

2.3 Alert Packet Size

The size of an individual alert packet is estimated to be $\lesssim$82 KB.

There are no requirements regarding the alert packet size. The statement above is an estimate based on the planned content of the alerts as described in Section 3.5 of LSE-163. Simulated alert packets based on the Apache Avro format are at most ∼82 KB. This volume represents an alert packet for a variable star with a full 12 month history of detections, and the history alone accounts for ∼27KB of the alert packet (∼33%). Cutout stamps included in the alert will be at least 30×30 pixels and contain flux (32 bit/pix), variance (32 bit/pix), and mask (16 bit/pix) extensions for both the template and difference image, plus a header of metadata [LSE-163]. The stamps alone will contribute $\gtrsim$18 KB to the total size of the uncompressed alert packet (i.e., ∼20%). The application of gzip compression can further reduce the size of an alert to ∼65 KB (JIRA ticket Data Management (DM)-16280).

on how the template is generated.

"Lite" Packet Options – Brokers that plan to do their own source association, compile source catalogs based on alerts, or not use the image stamps might prefer a stream of packets with appropriately reduced information. The LSST DM team is currently open to exploring options for supporting "Lite" versions of alert packets. Individual broker teams may indicate which information they require (or would like removed from the packets in their stream) during the broker proposal process [LDM-723]. As previously mentioned, removing the image stamps would reduce packet size by $\gtrsim$18 KB, and removing the historical records of past detections could reduce packet size by up to $\sim$27 KB. A few of these options might also be available to users of the LSST alert filtering service (§ 3).

2.4 Alert Stream Data Rate

The time-averaged data rate of the alert stream is estimated to be $\sim$0.2 Gbps, potentially with bursts of up to 5.4 Gbps.

There are no requirements regarding the alert stream data rate. The values quoted in the statement above are estimates based on the expected size of an alert packet ($\sim$82 KB, Section 2.3), the number of alerts per visit, and the alert distribution mechanism. Using an average of 10,000 alerts released per standard visit, this leads to a *time-averaged* alert stream data rate of $\sim$0.2 Gbps. As discussed in § 2.2, the number of alerts per field will vary in extragalactic and Galactic fields from $\sim$2,000 to $\lesssim$40,000, which would produce *time-averaged* alert streams of $\sim$0.04 to $\lesssim$0.8 Gbps. However, in order to release alerts within 60 seconds of image readout (§ 2.1), the stream will not be continuous in time, but periodic, with potential bursts: if all 10,000 alerts are issued within the last 5 seconds of that window the data rate would be 1.3 Gbps. In galactic fields with $\sim$40,000 alerts per visit this could be as high as 5.4 Gbps.

2.5 Number of Selected Brokers

It is a requirement that the DMS be capable of supporting the transmission of at least 5 full alert streams within 60 seconds of image readout.

numStreams
DMS-REQ-0391

This requirement is based in part on what estimates of the alert stream data rate and the bandwidth allocated to alert distribution have shown will be possible to support.

2.6 Alert Database Volume

The estimated size of the alerts database after 10 years is ∼2.2 PB.

There are no requirements on the alerts database volume. The statement above is an estimate based on the alert packet contents, the number of alerts per night, and the expected number of observing nights per year. As described in § 2.2, the DMS system will support an average of ∼10 million alerts per night (which approximately matches the expected scientific yields). Assuming the upper estimate of ∼82 KB per alert (§ 2.3), that leads to a total of ∼782 GB per night. After accounting for downtime and weather the total number of observing nights is 300 per year [LSE-30], which leads to an estimate of ∼2.2 PB after 10 years.

OSS-REQ-0080
OSS-REQ-0081
OSS-REQ-0082

2.7 Delayed/Failed Alert Distribution

It is a requirement that no more than 1% of all standard visits fail to have at least 98% of its alerts distributed within 60 seconds of image readout, and that no more than 0.1% of all standard visits fail to distribute alerts.

sciVisitAlertDelay
sciVisitAlertFailure
OSS-REQ-0112
DMS-REQ-0392

These requirements apply to standard visits which should have produced ≤40,000 alerts. For example, a visit would be considered ”delayed” and count towards that 1% limit if >2% of its alerts were distributed with a latency of >60 seconds. The requirement that no more than 0.1% of all science visits fail to generate and/or distribute alerts is integrated over all stages of data handling, not just alert distribution, and includes failures at any stage of prompt processing.

For an average of 10,000 alerts per visit and 1,000 visits per night, this requirement allows the DMS to distribute up to 2% (200,000 alerts per night) with a latency >60 seconds after image readout. The worst-case scenario for a night of alert distribution which still meets these requirements is if 989 visits all have just under 2% of their alerts delayed by >60 seconds and distributed within 24 hours (197,800 alerts delayed), *and* 10 visits (1%) have all of their alerts distributed with >60 seconds and <24 hours (100,000 alerts delayed), *and* 1 visit (0.1%) completely fails to generate and/or distribute any alerts (10,000 alerts).

2.8 Alert Stream Completeness and Purity

It is a requirement that DM derive and supply threshold values for a spuriousness[5] parameter, which can be used to filter alerts into a subsample of transient and variable objects with a given completeness and purity.

transSampleSNR
transCompleteness-Min
transPurityMin
OSS-REQ-0353

The requirement is that DM calculate a spuriousness parameter for all alerts, and derive and supply a spuriousness threshold value that filters the full stream into a subsample of alerts for transient and variable objects[6] that is 90% complete and 95% pure for all sources with a signal-to-noise ratio >6. While the requirements on purity and completeness are specified as point thresholds, DM expects to provide information to enable users to choose spuriousness threshold values that can be used to filter the stream to a desired level of completeness and purity, thereby reducing the fraction of false positives (sources detected that are not astrophysical in origin) to a level that is appropriate for their science goals. Brokers could request a pre-filtered stream that includes a restriction on spuriousness.

3 The LSST Alert Filtering Service

It is a requirement that the LSST provide an alerts filtering service for users.

DMS-REQ-0342
DMS-REQ-0348

The LSST alerts filtering service is a mechanism by which users — individuals with LSST data rights and access — can receive alerts via pre-defined filters that have been optimized for established transient classifications such as supernovae and/or create and apply their own filters to the stream [LPM-17; LSE-61].

3.1 Number of Simultaneous Users

It is a requirement that the LSST alert filtering service be able to support at least 100 simultaneous users.

numBrokerUsers
DMS-REQ-0343

This requirement is driven by outbound bandwidth limitations from the Data Access Center at the National Center for Supercomputing Applications (National Center for Supercomputing

[5] In this context, spuriousness is like a real/bogus score.

[6] See also Observatory System Specifications; LSE-30 (OSS)-REQ-0354 for the required parameters for a subsample of transient and variable objects [LSE-30].

Applications (NCSA)); the DM team is currently investigating approaches that would support larger numbers of active users [LDM-612]. During LSST Operations, if the total number of simultaneous users is oversubscribed then a proposal process may be instituted [LSE-163].

3.2 Number of Alerts per Visit Returned

It is a requirement that the LSST alert filtering service be able to return 20 full-sized alerts per visit per user.

numBrokerAlerts
DMS-REQ-0343

Assuming 1,000 visits per night (§ 2.2), each user's filter will be capable of returning 20,000 alerts per night, which would amount to ∼1.6 GB (§ 2.3).

4 Alerts Archive

It is a requirement that all alerts be stored in an archival database and be available for retrieval.

DMS-REQ-0094

The term "available for retrieval" applies to users with data rights and access to the LSST Science Platform. Like all other Prompt data products, the alerts archive will be updated within 24 hours [LSE-29]. The alerts archive is not a part of the LSST alert filtering service, but is included in this section to raise awareness of its existence.

L1PublicT
LSR-REQ-0104

The LSST DM team anticipates that the alerts archive will support queries by their unique alert identification numbers, but might not support searches by coordinate, time, magnitude, or other alert attributes. For this reason, the alerts archive should *not* be considered a viable alternative for users who *do* wish to study transient, variable and moving objects with the LSST, but who *do not* require immediate (i.e., same-night) access to sources detected via difference image analysis. In other words, queries to the alerts archival database should not be construed as a viable alternative to community brokers or the LSST alert filtering service. Instead, users with science goals that are achievable with a latency of ≥24 hours should plan to use the Prompt data products described in Section 3 of LSE-163. Furthermore, users with science goals that are achievable with latencies of a year or more (i.e., archival time-domain studies) should plan to use the Data Release data products described in Section 4 of LSE-163.

References

[LDM-612], Bellm, E., Blum, R., Graham, M., et al., 2020, Plans and Policies for LSST Alert Distribution, URL https://ldm-612.lsst.io/,
Vera C. Rubin Observatory Data Management Controlled Document LDM-612

[LDM-723], Bellm, E., Blum, R., Graham, M., et al., 2020, Call for Proposals for Community Alert Brokers, URL https://ldm-723.lsst.io/,
Vera C. Rubin Observatory Data Management Controlled Document LDM-723

[LSE-29], Claver, C.F., The LSST Systems Engineering Integrated Project Team, 2017, LSST System Requirements (LSR), URL https://ls.st/LSE-29,
Vera C. Rubin Observatory LSE-29

[LSE-30], Claver, C.F., The LSST Systems Engineering Integrated Project Team, 2018, Observatory System Specifications (OSS), URL https://ls.st/LSE-30,
Vera C. Rubin Observatory LSE-30

[LSE-61], Dubois-Felsmann, G., Jenness, T., 2019, Data Management System (DMS) Requirements, URL https://lse-61.lsst.io/,
Vera C. Rubin Observatory LSE-61

[LPM-17], Ivezić, Ž., The LSST Science Collaboration, 2018, LSST Science Requirements Document, URL https://ls.st/LPM-17,
Vera C. Rubin Observatory LPM-17

[LSE-163], Jurić, M., Axelrod, T., Becker, A., et al., 2023, Data Products Definition Document, URL https://lse-163.lsst.io/,
Vera C. Rubin Observatory LSE-163

LSST Science Collaboration, 2009, ArXiv e-prints (arXiv:0912.0201), doi:10.48550/arXiv.0912.0201, ADS Link

A Glossary

Alert A packet of information for each source detected with signal-to-noise ratio > 5 in a difference image during Prompt Processing, containing measurement and characterization parameters based on the past 12 months of LSST observations plus small cutouts of the single-visit, template, and difference images, distributed via the internet.

Alert Production The principal component of Prompt Processing that processes and calibrates incoming images, performs Difference Image Analysis to identify DIASources and DIAObjects, packages and distributes the resulting Alerts, and runs Solar System Processing..

AP Alert Production.

Archive The repository for documents required by the NSF to be kept. These include documents related to design and development, construction, integration, test, and operations of the LSST observatory system. The archive is maintained using the enterprise content management system DocuShare, which is accessible through a link on the project website www.project.lsst.org.

B Byte (8 bit).

Center An entity managed by AURA that is responsible for execution of a federally funded project.

Data Management The LSST Subsystem responsible for the Data Management System (DMS), which will capture, store, catalog, and serve the LSST dataset to the scientific community and public. The DM team is responsible for the DMS architecture, applications, middleware, infrastructure, algorithms, and Observatory Network Design. DM is a distributed team working at LSST and partner institutions, with the DM Subsystem Manager located at LSST headquarters in Tucson.

Data Management System The computing infrastructure, middleware, and applications that process, store, and enable information extraction from the LSST dataset; the DMS will process peta-scale data volume, convert raw images into a faithful representation of the universe, and archive the results in a useful form. The infrastructure layer consists of the computing, storage, networking hardware, and system software. The middleware layer handles distributed processing, data access, user interface, and system operations services. The applications layer includes the data pipelines and the science data archives' products and services.

deg degree; unit of angle.

DIA Difference Image Analysis.

DM Data Management.

DMS Data Management Subsystem.

Document Any object (in any application supported by DocuShare or design archives such as PDMWorks or GIT) that supports project management or records milestones and deliverables of the LSST Project.

flux Shorthand for radiative flux, it is a measure of the transport of radiant energy per unit area per unit time. In astronomy this is usually expressed in cgs units: erg/cm2/s.

JIRA issue tracking product (not an acronym but a truncation of Gojira the Japanese name for Godzilla).

KB KiloByte.

LSST Large Synoptic Survey Telescope.

MB MegaByte.

metadata General term for data about data, e.g., attributes of astronomical objects (e.g. images, sources, astroObjects, etc.) that are characteristics of the objects themselves, and facilitate the organization, preservation, and query of data sets. (E.g., a FITS header contains metadata).

NCSA National Center for Supercomputing Applications.

Operations The 10-year period following construction and commissioning during which the LSST Observatory conducts its survey.

OSS Observatory System Specifications; LSE-30.

pipeline A configured sequence of software tasks (Stages) to process data and generate data products. Example: Association Pipeline.

Release Publication of a new version of a document, software, or data product. Depending on context, releases may require approval from Project- or DM-level change control boards, and then form part of the formal project baseline.

Science Platform A set of integrated web applications and services deployed at the LSST Data Access Centers (DACs) through which the scientific community will access, visualize, and perform next-to-the-data analysis of the LSST data products.

transient A transient source is one that has been detected on a difference image, but has not been associated with either an astronomical object or a solar system body.

Visit A sequence of one or more consecutive exposures at a given position, orientation, and filter within the LSST cadence. See Standard Visit, Alternative Standard Visit, and Non-Standard Visit.