LARGE SYNOPTIC SURVEY TELESCOPE

**Large Synoptic Survey Telescope (LSST)
Data Management**

LSST Alerts: Key Numbers

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

DMTN-102

Latest Revision: 2019-02-19

Abstract

A quantitative review of the key numbers associated with the LSST Alert Stream.

Change Record

Version	Date	Description	Owner name
1	2019-02-19	Released.	Melissa Graham

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

Contents

LSST Alerts: Key Numbers

1   Introduction

The LSST Data Management System's (DMS) Alert Production (AP) pipeline will process new data as it is obtained by the telescope. Difference Imaging 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.

<div style="text-align: right">transSNR
LSR-REQ-0101
DMS-REQ-0269
DMS-REQ-0274</div>

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 (B), and 1024 B per KB, 1024 KB per 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 formal requirement that the DMS make available 98% of alerts for each visit within 60 seconds of the end of image readout[2].

<div align="right">

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

</div>

The term "make available" refers to the point at which an LSST alert packet becomes available for broker retrieval (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 formal requirement that the DMS support the generation of at least 10,000 alerts per visit[3] on average, and an instantaneous maximum of 40,000 alerts per visit.

<div align="right">

transN
LSR-REQ-0101
nAlertVisitAvg
OSS-REQ-0193

</div>

An approximate upper limit on the number of standard visits (the equivalent of 30 seconds of exposure time, plus ~4 seconds of overhead) per night is 1,000 visits, which corresponds to ≲9.5 hours of observing time. Pending updates to LSST documentation (LSE-30 and LSE-61) will clarify that the DMS must support a long-term average number of alerts *per night* of 10^7 (i.e., 10,000 alerts per visit for 1,000 standard visits per night) and instantaneous peaks of 40,000 alerts *per visit*. At the time of writing, the latter was only defined in the DM sizing model [LSE-81].

The value of 10,000 alerts per visit is a formal 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 deg² 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,

---

[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].

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 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. Assuming that they are spread evenly over the ~18,000 deg^2 survey area (even though they're not, as they're found primarily along the ecliptic) leads to ~3,000 alerts per visit due to moving objects.

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 ⩽82 KB.

There are no formal 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. The application of gzip compression can further reduce the size of an alert to ~65 KB (JIRA ticket DM-16280). 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 ⪆18 KB to the total size of the uncompressed alert packet (i.e., ~20%).

---

[4]Since "detectably variable" means "significantly different from the template", the value of 10% does depend 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 brokers may indicate which information they require (or would like removed from the packets in their stream) during the broker proposal process [**?**]. As previously mentioned, removing the image stamps would reduce packet size by $\gtrsim$18 KB. Removing the historical records of past detections would also reduce the size of all alert packets. 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 ~25 MB s^{-1}, potentially with bursts of up to 640 MB s^{-1}.

There are no formal 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 (~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 ~25 MB s^{-1}. As discussed in § 2.2, the number of alerts per field will vary in extragalactic and Galactic fields from ~2,000 to $\lesssim$40,000, which would produce *time-averaged* alert streams of ~5 to $\lesssim$100 MB s^{-1}. 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 160 MB s^{-1}. In galactic fields with ~40,000 alerts per visit this could be as high as 640 MB s^{-1}.

2.5 Number of Selected Brokers

The DM team estimates that resources will allow for the delivery of the alert stream to 4 to 7 community brokers.

There are no formal requirements on the number of brokers. The statement above is an estimate based on the alert stream data rate and the bandwidth allocated to alert distribution. As described in Section 2.2.3 of LDM-612: *"An allocation of 10 Gbps is baselined for alert stream*

transfer from the [LSST Data Facility], with an estimated packet size of 82 KB and up to 10,000 alerts per visit. For illustration, based on these numbers up to 7 brokers could receive the full stream if 5 seconds is budgeted for outbound data transfer."

2.6   Alert Database Volume

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

There are no formal 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. Compression could drastically lower this, as could reformatting: every alert contains a ~12 month historical record and links to the most recent DIAObject and DR Object catalogs. The set of alerts for the same transient/variable would contain a significant amount of redundant information, which is provided to enable rapid filtering but could be reformatted for long-term storage (i.e., removed from the individual alerts and compiled elsewhere in an alerts database).
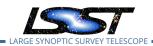
OSS-REQ-0080
OSS-REQ-0081
OSS-REQ-0082

2.7   Delayed/Failed Alert Distribution

It is a formal requirement that no more than 1% (0.1%) of all science visits experience delayed (failed) alert generation and distribution.

sciVisitAlertDelay
sciVisitAlertFailure
OSS-REQ-0112

Given the requirement in § 2.1 that 98% of all alerts be made available within 60 seconds, a visit would be considered "delayed" and count towards that 1% limit if, and only if, >2% of its alerts were made available 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 (i.e., includes failures at any stage of prompt processing).

Assuming ~10 million alerts per night (§ 2.2), brokers can rely on the DMS to not exceed an upper limit of 200,000 alerts (2%) being made available with a latency of >60 seconds, and

an upper limit of 10,000 alerts (or 1 visit's worth, i.e., 0.1% of 1,000 visits per night) that fail to be generated and/or made available. In other words, the absolute worst-case scenario for a night — even though we might not expect it to happen — : still meets these specifications is if 989 visits all have just under 2% of their alerts (i.e., up to 197800 alerts) delayed by >60 seconds and made available within 24 hours, *and* 10 visits (1%) have all of their alerts made available with >60 seconds and <24 hours (i.e., are "delayed" but not "failed"), *and* 1 visit (0.1%) completely fails to generate and/or make available any alerts.

2.8 Alert Stream Completeness and Purity

It is a formal 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 formal 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 formal 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].

---

[5]In this context, spuriousness is like a real/bogus score.
[6]See also OSS-REQ-0354 for the required parameters for a subsample of transient and variable objects [LSE-30].

3.1　Number of Simultaneous Users

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

OSS-REQ-0185

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., co authors, 2018, *Plans and Policies for LSST Alert Distribution*, LDM-612, URL `https://ls.st/LDM-612`

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

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

[LSE-81], Dubois-Felsmann, G., 2013, *LSST Science and Project Sizing Inputs*, LSE-81, URL `https://ls.st/LSE-81`

[LSE-61], Dubois-Felsmann, G., Jenness, T., 2018, *LSST Data Management Subsystem Requirements*, LSE-61, URL `https://ls.st/LSE-61`

[LPM-17], Ivezić, Ž., The LSST Science Collaboration, 2018, *LSST Science Requirements Document*, LPM-17, URL `https://ls.st/LPM-17`

Ivezić, Z., et al., 2008, ArXiv e-prints (`arXiv:0805.2366`), ADS Link

[LSE-163], Jurić, M., et al., 2017, *LSST Data Products Definition Document*, LSE-163, URL `https://ls.st/LSE-163`

LSST Science Collaboration, 2009, ArXiv e-prints (`arXiv:0912.0201`), ADS Link

[LDM-151], Swinbank, J.D., et al., 2017, *Data Management Science Pipelines Design*, LDM-151, URL `https://ls.st/LDM-151`