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 Subsystem Science Team

DMTN-102

Latest Revision: 2019-01-20

Abstract

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

Change Record

Version	Date	Description	Owner name
1	2018-12-06	Unreleased.	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", instantiate a record in the source catalogs, and generate an alert (LPM-17; LSR-REQ-0101, LSE-29; DMS-REQ-0269, -0274, LSE-61). 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 via the LSST Science Platform's Alert Filtering Service (LAFS). Plans and policies for alert distribution are provided in LDM-612.

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 is provided (in the format [DOC]-REQ-####). In this work we use 8 bits per byte (B), and $1024\ \mathrm{B}$ per KB, $1024\ \mathrm{KB}$ per MB, and so forth. The resources used in the preparation of this document are as follows:

- *LSST: From Science Drivers to Reference Design and Anticipated Data Products*, Ivezić et al. (2008)
- LSST Science Requirements Document (SRD), LPM-17.
- LSST System Requirements (LSR), LSE-29.
- Observatory System Specifications (OSS) document, LSE-30.
- Data Management System Requirements (DMSR) document, LSE-61.

- Science Requirements and System Specifications Spreadsheet (SR&SSS), LSE-81.
- Data Products Definitions Document (DPDD), LSE-163
- Plans and Policies for LSST Alert Distribution, LDM-612
- Data Management Science Pipelines Design, LDM-151

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 (SRD), which specifies that information about the detections of transient, variable, and moving objects be released promptly as a data stream.

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.

The formal requirements on the DMS's alert release timescale are defined by the SRD[1] [LPM-17], LSR-REQ-0101,-0025 [LSE-29], OSS-REQ-0127 [LSE-30], and DMS-REQ-0004 [LSE-61]. The term "make available" refers to the point at which an LSST alert packet becomes publicly available for broker retrieval (i.e., it does not include the time it takes for a broker to receive or ingest the alert). The LSR also specifies that all delayed alerts be made available at the next opportunity (see the discussion regarding delayed/failed alert distribution in § 2.7).

2.2 Number of Alerts per Visit

It is a formal requirement that the DMS sustain the generation of an average of at least $10,000$ alerts per visit, and an instantaneous maximum of at least $40,000$ alerts per visit.

The formal requirements on the DMS's ability to generate a minimum number of alerts per

[1] In the SRD, the design, minimum, and stretch values for the alert release timescale are 1, 2, and 0.5 minutes, respectively.

visit are defined by the SRD[2] [LPM-17], LSR-REQ-0101 [LSE-29], and OSS-REQ-0193 [LSE-30]. Pending updates to the OSS and the DMSR, these documents will clarify that the DMS must support a long-term average number of alerts per night of 10^7 (i.e., 10^4 alerts per visit for 10^3 visits per night) and instantaneous peaks of 4×10^4 alerts per visit.

The value of 10^4 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 extra/galactic fields, and that of the $\sim 18,000\ \mathrm{deg}^2$ surveyed by LSST, 80/20% of the fields are extra/galactic, and that 10% of all variable stars are detectably variable at any given time, then a typical extra/galactic field would yield $\sim 1,800$/$28,800$ alerts per visit. Averaged over all fields, and weighted by 80/20% of the fields being extra/galactic, this is $7,200$ 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 $15,000\ \mathrm{deg}^2$ of extragalactic survey area, that's $\sim 20\ \mathrm{SNe\ deg}^{-2}$ or ~ 200 alerts for SNe per visit.
- Active Galactic Nuclei: LSST is predicted to observe millions of AGN. If $\sim 10\%$ of them are detectably variable at any given time, then the estimate is ~ 0.1 million alerts over $15,000\ \mathrm{deg}^2$ would generate $\sim 7\ \mathrm{alerts\ deg}^{-2}$ 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 $\sim 18,000\ \mathrm{deg}^2$ survey area (even though they're not, as they're found primarily along the ecliptic) leads to $\sim 3,000$ 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 $\sim 5,100$/$32,000$ alerts per visit in extra/galactic fields, with an average of $\sim 10,500$ alerts per visit.

[2] In the SRD, the design, minimum, and stretch values for the number of alerts per visit are 10^4, 10^3, and 10^5, respectively.

2.3 Alert Packet Size

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

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 $\sim 82/126\ \mathrm{KB}$, without/with the schema, respectively. 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 $\sim 65\ \mathrm{KB}$ (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 $\gtrsim 18\ \mathrm{KB}$ to the total size of the uncompressed alert packet (i.e., $\sim 20\%$).

"Lite" Packet Options – Brokers which 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 currently expects that providing a subset of alert packet contents will be feasible, and brokers may indicate which information they require during the broker proposal process [LDM-612]. As previously mentioned, removing the image stamps would reduce packet size by $\gtrsim 18\ \mathrm{KB}$. 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 $\sim 25\ \mathrm{MB/sec}$**, potentially with bursts of up to** $640\ \mathrm{MB/sec}$**.**

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, the number of alerts per visit, and the alert distribution mechanism. The size of a single LSST alert will be $\sim 82\ \mathrm{KB}$ (including image stamps but not schema nor compression). Using an average of $\sim 10^4$ alerts released per ~ 30 second image $+2$ second readout, this leads to a *time-averaged* alert stream data rate of $\sim 25\ \mathrm{MB\ s^{-1}}$. As discussed in § 2.2, the number of alerts per field will vary in extra/galactic fields from $\sim 2,000$ 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 10^4 alerts are issued within the last 5 seconds of that window the data rate would be 160 MB s^{-1}. In galactic fields with $\sim 40,000$ 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 maximum upper limit for the full, 10-year alerts database is $\lesssim 3$ PB.

There are no formal requirements on the alerts database volume. The statement above is an estimate based on the alert packet contents and the number of alerts per night. 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. An extreme upper limit is 365 nights per year for 10 years, which would amount to ~ 2.7 PB *at the very most*. Therefore we quote an extreme upper limit on the alerts database as $\lesssim 3$ PB. 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).

2.7 Delayed/Failed Alert Distribution

It is a formal requirement that $\leq 1\%$ ($\leq 0.1\%$) of all science visits experience delayed (failed) alert generation and distribution.

The formal requirements on the fraction of visits that experience delayed or failed alerts are defined by OSS-REQ-0112 [LSE-30]. 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 $< 0.1\%$ 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 $\sim 10,000,000$ alerts per night (§ 2.2), brokers can rely on the DMS to not exceed an upper limit of $\lesssim 200,000$ alerts ($\leq 2\%$) being made available with a latency of > 60 seconds, and an upper limit of $\lesssim 10,000$ alerts or 1 visit's worth ($\leq 0.1\%$ of $\sim 1,000$ visits per night) that fail to be generated and/or made available.

2.8 Alert Stream Completeness and Purity

It is a formal requirement that DM derive and supply threshold values for a spuriousness parameter which can be used to filter alerts into a subsample with a given completeness and purity.

The formal requirement is that DM calculate a spuriousness[3] parameter for all alerts, and derive and supply a spuriousness threshold value that filters the full stream into a subsample of transient alerts which is 90% complete and 95% pure for all sources with a signal-to-noise ratio > 6 (OSS-REQ-0353, but see also -0354 for values related to moving objects LSE-30). While the requirement on purity and completeness is specified as a point threshold, DM expects to provide information to enable users to choose spuriousness threshold values which will 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] In this context, spuriousness is like a real/bogus score.

3 The LSST Alert Filtering Service

It is a formal requirement that the LSST provide an alerts filtering service for users, which is hereafter referred to as the LSST Alert Filtering Service (LAFS).

The formal requirement that LSST provide a mechanism by which users – individuals with LSST data rights and access – can receive alerts via pre-defined filters optimized for established transient classifications such as supernovae, and/or create and apply their own filters to the stream, is defined by the SRD [LPM-17] and by DMS-REQ-0342, and -0348 [LSE-61].

3.1 Number of Simultaneous LAFS Users

It is a formal requirement that the LAFS be able to support a minimum of 100 simultaneous users.

This formal requirement is defined by DMS-REQ-0343 [LSE-61]. If, during LSST Operations, the total number of simultaneous LAFS users is oversubscribed, a proposal process may be instituted [LSE-163].

3.2 Number of Alerts per Visit Returned by LAFS

It is a formal requirement that the LAFS be able to return 20 full-sized alerts per visit per user.

This formal requirement is defined by DMS-REQ-0343 [LSE-61]. However, note that in a footnote of LDM-612, it says that the *"requirement on the number of simultaneously connected users and number of passed alerts is largely driven by outbound bandwidth limitations from the DAC at NCSA. We are investigating approaches that would support larger numbers of active filters"* (page 12; LDM-612).

3.3 Alerts Database

It is a formal requirement that all alerts be stored in a database and available for query.

This formal requirement is defined by OSS-REQ-0185 [LSE-30], and the term "available for query" applies to users with data rights and access to the LSST Science Platform. Like all other Prompt data products, the alerts database will be updated within 24 hours (LSR-REQ-0104, LSE-29). The alerts database is not a part of LAFS, but is included in this section to emphasize that science users who do not need a latency of < 24 hours may wish to query the alerts database *instead of* creating LAFS filters.

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`