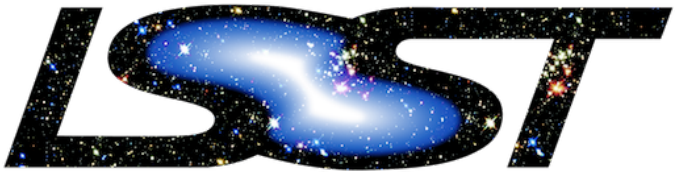

LARGE SYNOPTIC SURVEY TELESCOPE

Large Synoptic Survey Telescope (LSST)
Data Management

LSST Alerts: Key Numbers

M. L. Graham, E. Bellm, L. Guy, C. T. Slater, et al.

DMTN-102

Latest Revision: 2019-01-15

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 $\texttt{transSNR} > 5$ in positive or negative flux will be considered "detected", instantiate a record in the source catalogs, and generate an alert (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. In this work we use 8 bits per byte (B), and 1024 B per KB, 1024 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.

Regarding the DMS's ability to generate alerts, the SRD states that *"data on likely optical transients ... will be released with a latency of at most* OTT1 *minutes"*, where the design, minimum, and stretch values for OTT1 are 1, 2, and 0.5 minutes, respectively [LPM-17]. The SRD's requirement flows down to the LSR, which also defines the fraction of alerts per visit that must be made available within OTT1 as OTR1= 98% (LSR-REQ-0101,-0025 LSE-29). The LSR's requirements on OTT1 and OTR1 flow down to the OSS (OSS-REQ-0127, LSE-30) and the DMSR (DMS-REQ-0004, LSE-61). It is the intent of these requirements that OTT1 covers only the time that the DMS has control over: from the end of camera readout to the time the alert packet "crosses the border" of NCSA and becomes publicly available for the brokers (i.e., OTT1 does not include the time it takes for a broker to receive or ingest the alert). It is sometimes colloquially said that OTT1 starts *at the time of shutter close*, but it actually starts *at the end of readout* (i.e., 2 seconds later, or longer if slower reads are adopted). The LSR also specifies that all delayed alerts be made available at the next opportunity (see also the discussion regarding formal requirements on 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 SRD states that the DMS *"should be capable of reporting such data for at least* `transN` *candidate transients per field of view and visit,"* where the design, minimum, and stretch values for `transN` are 10^4, 10^3, and 10^5 respectively [LPM-17]. The SRD's requirement flows down to the LSR, in which it is unclear whether `transN` refers to an average or an instantaneous limit (LSR-REQ-0101, LSE-29). The OSS clarifies this with a formal requirement that the DMS *"shall be sized to accommodate an average value of at least* `nAlertVisitAvg` *alerts generated per standard visit"*, where `nAlertVisitAvg` is 10^4 (OSS-REQ-0193 in LSE-30). Pending updates to the OSS and the DMSR will clarify that the DMS must support a long-term average number of alerts per night of `nAlertNightAvg` $= 10,000,000$ (i.e., $10,000$ alerts per visit for 1000 visits per night), and instantaneous peaks of `nAlertVisitMax`$= 40,000$ alerts per visit.

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 extra/galactic fields, and that of the $\sim 18000\ \mathrm{deg}^2$ 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 $\sim 1800/28800$ alerts per visit. Averaged over all fields, and weighted by $80/20\%$ of the fields being extra/galactic, this is 7200 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 $15000\ \mathrm{deg}^2$ 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 $15000\ \mathrm{deg}^2$ 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 ~ 18000 deg^2 survey area (even though they're not, as they're found primarily along the ecliptic) leads to ~ 3000 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 $\sim 5100/32000$ alerts per visit in extra/galactic fields, with an average of ~ 10500 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 formal requirements regarding the alert packet size; this 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 $\sim 82/126$ 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 ~ 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 $\gtrsim 18$ 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 some options will be possible, and brokers may propose an option that works for them during the selection process [LDM-612]. As previously mentioned, removing the image stamps would reduce packet size by $\gtrsim 18$ KB. Removing the historical records of past detections would reduce all alert packets to be equivalent in size to a new unassociated source. 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/sec, poten-

tially with bursts of up to $640\ \mathrm{MB/sec}$**.**

There are no formal requirements regarding the alert stream data rate, these 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 $\sim\ 10000$ alerts released per $\sim\ 30$ 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 ~ 2000 to $\lesssim 40000$, which would produce *time-averaged* alert streams of ~ 5 to $\lesssim 100$ $\mathrm{MB\ s^{-1}}$. However, in order to release alerts within $\mathtt{OTT1} = 60$ seconds of image readout (§ 2.1), the stream will not be continuous in time, but periodic, with potential bursts. For example, if all 10000 alerts are issued within the last 5 seconds of $\mathtt{OTT1}$, this would produce a data rate of $160\ \mathrm{MB\ s^{-1}}$; in galactic fields with $\lesssim 40000$ alerts this could be as high as $640\ \mathrm{MB\ s^{-1}}$.

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; this is an estimate based on the alert stream data rate and the planned alert distribution mechanism. As described in Section 2.2.3 of LDM-612, *"An allocation of* $10\ \mathrm{Gbps}$ *is baselined for alert stream transfer from the [LSST Data Facility], with an estimated packet size of* $82\ \mathrm{KB}$ *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 $\lesssim 3\ \mathrm{PB}$**.**

There are no formal requirements on the alerts database volume; this 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 $\sim 82\ \mathrm{KB}$ per alert (§ 2.3), that leads to a total of $\sim 782\ \mathrm{GB}$ 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 could be reformatted (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 $< 1\%$ ($< 0.1\%$) of all science visits experience delayed (failed) alert generation and distribution.

The OSS specifies that *"no more than* `sciVisitAlertFailure` *% of science visits ... shall fail to be subjected to alert generation and distribution, integrated over all stages of data handling"*, where `sciVisitAlertFailure` $= 0.1\%$, and that *"no more than* `sciVisitAlertDelay` *% of science visits ... shall have their alert generation and distribution completed later than [*`OTT1`*]"*, where `sciVisitAlertDelay` $= 1\%$ (OSS-REQ-0112; LSE-30). Assuming an estimated $10,000,000$ alerts per night (§ 2.2), brokers can rely on the DMS to not exceed an upper limit of $\lesssim 200,000$ alerts ($\lesssim 2\%$) made available with a latency $>$`OTT1`$= 60$ seconds, and an upper limit of $\lesssim 10,000$ alerts (or 1 visit) that fail to be 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 SRD makes no statements about alert stream purity or completeness, but does state that the *"minimum signal-to-noise ratio in difference image for reporting detection of a transient object"* has a design specification of `transSNR` $= 5$ [LPM-17]. The LSR contains essentially the same definition for `transSNR`: *"the signal-to-noise ratio in single-visit difference images above which all optical transients are to be reported"* (LSR-REQ-0101; LSE-29). There is no minimum specification or stretch goal associated with `transSNR`. However, as described in the DPDD, detections with `transSNR` > 5 but *"with high probability of being instrumental non-astrophysical artifacts may be excluded"* (Section 3.1, LSE-163), and a limited number of `transSNR` < 5 sources might be

allowed to generate alerts (e.g., sources near a gravitational keyhole which could make an asteroid potentially hazardous; Section 3.2.1, LSE-163).

Regarding the alert stream completeness and purity, the OSS requires that *"there shall exist a spuriousness threshold* `T`*, for which the completeness and purity of selected difference sources are higher than* `transCompletenessMin` *and* `transPurityMin`*, respectively, at the SNR detection threshold* `transSampleSNR`*. This requirement is to be interpreted as an average over the entire survey"* (OSS-REQ-0353; LSE-30). The values of these requirement parameters for transients are `transSampleSNR` $= 6$, `transCompletenessMin` $= 90\%$, and `transPurityMin` $= 95\%$. (For moving objects there is a separate specification for a threshold pertaining to `transSampleSNR` $= 5$, `mopsCompletenessMin` $= 99\%$, and `mopsPurityMin` $= 50\%$; OSS-REQ-0354, LSE-30). This spuriousness threshold `T` will allow users to filter their stream to a fiducial completeness and purity, thereby reducing the fraction of false positives (sources detected that are not astrophysical in origin). Brokers may 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 a simple alerts filtering service for users (individuals with LSST data rights and access to the Science Platform), which is hereafter referred to as the LSST alert filtering service (LAFS).

The SRD specifies that *"users will have an option of a query-like pre-filtering of the [alert] data stream in order to select likely candidates for specific transient type"* and that *"several pre-defined filters optimized for traditionally popular transients, such as supernovae and microlensed sources, will also be available"* [LPM-17]. This flows down directly to the DMSR, which has a formal requirement that *"a basic, limited capacity, alert filtering service shall be provided that can be given user defined filters to reduce the alert stream to manageable levels"*, and that this service include *"a predefined set of simple filters"* (DMS-REQ-0342, -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.

The DMSR specifies that the LSST *"alert filtering service shall support* `numBrokerUsers` *simultane-*

ous users", where `numBrokerUsers` = 100 (DMS-REQ-0343; LSE-61).

3.2 Number of Alerts per Visit Returned by LAFS

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

The DMSR specifies that within the LSST alert filtering service *"each user [shall be] allocated a bandwidth capable of receiving the equivalent of* `numBrokerAlerts` *alerts per visit"*, where `numBrokerAlerts` = 20 (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.

The OSS states that *"All published transient alerts ... shall be available for query"*, and this applies to users with data rights and access to the LSST science platform (OSS-REQ-0185; LSE-30). Like all other Prompt data products, the alerts database will be updated within `L1PublicT` = 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 $< 24h$ should be querying 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