DM sizing model and purchase plan for the remainder of construction.

**Michelle Butler, Kian Tat Lim, William O'Mullane**

2019-11-26

# 1 Introduction

# 2 Proposed budget

Based on the needs in Table 6 and the costs in Table 7 and Table 8 the following budgets can be considered.

A high level bottom line is given in Table 1. Since Xeon is more expensive that is the number used for the budget calculation, should we get Rome and it really performs we may save a little. The remainder of the document is all the details that went into that.

Table 1: This table pulls together all the information in a high level summary - in this table Xeon pricing is used since that is the more expensive but better known option. Price factors, defined in Table 6 are applied post 2020.

| Year | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|
| Compute (2019 pricing) | $300,102 | $0 | $3,286,841 | $7,976,643 |
| Applying price factor (CPU) | | $0 | $2,629,473 | $5,583,650 |
| IN2P3 (50% of compute) | | | | -$2,791,825 |
| Storage (2019 pricing) | $164,213 | $245,799 | $1,760,037 | $9,281,372 |
| Applying price factor (Storage) | $164,213 | $233,509 | $1,584,034 | $7,889,166 |
| **Total budget (using price factors)** | **$164,213** | **$233,509** | **$4,213,506** | **$10,680,991** |

In Table 1 we should note that IN2P3 do 50% of processing so we reduce the processing cost by half. This does not reduce the storage cost. We have applied a modes cost reduction assuming that processors and disks get a little cheaper - that percentage is given in Table 6.

# 3 Potential scope option

In the 2019 JSR we discussed the possibility of delaying purchasing DR1 hardware in DM. Table 2 defines what this would be worth using the cost/sizing model in this document.

Table 2: Considering a scope option of delaying the purchase of LOY1 processing hardware and only purchasing what is needed for commissioning we would only purchase up to and including 2022 hardware of Table 1. If we consider that amount and the current remaining construction budget for hardware the potential worth of such a scope option is given in this table.

| | |
|---|---|
| Budget for commissioning (to 2022) | $4,611,228 |
| DM construciton budet remaining | $14,000,000 |
| **Total potential to delay to ops** | **$9,388,772** |

## 3.1 Buy Xeon for compute

Table 3 gives the price of compute based on Xeons.

Table 3: Implementation with Intel Xeon

| Year | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|
| Number of Xeon | 30.01 | 0.00 | 328.68 | 797.66 |
| Approximate cost | $300,102.25 | $0.00 | $3,286,840.87 | $7,976,642.82 |

## 3.2 Buy Rome for compute

Table 4 gives the price of compute based on Rome -small and large.

Table 4: Implementation with AMD Rome (we have no good proce for these reallly)

| Year | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|
| number of small rome | 21.74 | 0.00 | 222.12 | 555.31 |
| Approximate cost of small rome | $282,662.34 | $0.00 | $2,887,620.17 | $7,219,050.42 |
| number of large rome | 7.00 | 0.00 | 71.53 | 178.84 |
| Approximate cost of large rome | $165,956.54 | $0.00 | $1,695,377.79 | $4,238,444.48 |

## 3.3 Storage

Table 5 gives the price of storage using all 3 types that we need. i This would be needed regardless of the compute chosen.

Table 5: Total storage cost estimate

| Year | 2020 | 2021 | 2022 | 2023 |
|---|---|---|---|---|
| Fast Storage | $11,842.11 | $11,842.11 | $8,433.33 | $50,600.00 |
| Normal Storage | $0.00 | $39,141.45 | $878,147.54 | $4,789,001.84 |
| Latent Storage | $31,852.03 | $55,741.05 | $318,074.72 | $1,616,046.66 |
| High Latency Storage | $120,518.73 | $139,074.68 | $555,381.66 | $2,825,723.27 |
| **Total** | **$164,212.87** | **$245,799.29** | **$1,760,037.25** | **$9,281,371.77** |

# 4 Models

## 4.1 Sizing model

An exhaustive and detailed mode is provided in [LDM-138; LDM-144] - here we concentrate on the needs for the final years of construction. We explore the compute and storage needed to get us through commissioning and suggest a 2023 purchase for DR1,2 processing which could be pushed to operations.

Table 6 gives the annual requirements for the next few years.

Table 6: Various inputs for deriving costs - 2019 represents currentl holdings.

| **Year** | **2019** | **2020** | **2021** | **2022** | **2023** |
|---|---|---|---|---|---|
| FLOPs Needed Total (DRP) | | 2.37E+20 | 2.37E+20 | 2.66E+21 | 1.48E+22 |
| Annual Increase | | 2.37E+20 | 0.00E+00 | 2.42E+21 | 1.21E+22 |
| Time to Process days | | 100.0 | 100.0 | 100.0 | 200 |
| Time to Process seconds | | 8,640,000 | 8,640,000 | 8,640,000 | 17,280,000 |
| Instantaneous GFLOP/ s (DRP) Annual increase | 39168 | 2.74E+04 | 0.00E+00 | 2.80E+05 | 7.00E+05 |
| Instantaneous GFLOP/ s (Alerts) | | 0 | 0 | 20203 | 0 |
| GFLOP/ s (Alerts) Annual increase | | 0 | 0 | 20203 | 0 |
| Instantaneous GFLOP/ s (SciPlat) | 18,278 | 18,278 | 18,278 | 18,278 | 46,810 |
| GFLOP/ s (SciPlat) Annual increase | | 0 | 0 | 0 | 28531 |
| **Total Annual Increase** | | **27,427** | **0** | **300,391** | **729,001** |
| Fast Storage (TB) | | 12 | 24 | 32 | 83 |
| Annual Increase (Fast) | | 12 | 12 | 8 | 51 |
| Normal Storage (TB) | 3000 | 1883 | 2173 | 8678 | 44152 |
| Annual Increase (Normal) | | 0 | 290 | 6505 | 35474 |
| Latent Storage (TB) | | 319 | 876 | 4057 | 20217 |
| Annual Increase (Latent) | | 319 | 557 | 3181 | 16160 |
| High Latency (TB) | | 1883 | 4056 | 12734 | 56886 |
| Annual Increase (High Latency) | | 1883 | 2173 | 8678 | 44152 |
| Annual price decrease CPU | | 10% | | | |
| Annual price decrease Storage | | 5% | | | |

## 4.2 Compute and storage

We which to base our budget on reasonable well know machines for which we have well know prices. Table 8 gives an outline of a few standard machines we use and a price. This table also gives a FLOP estimate for those machines. Table 7 gives costs for different types of storage - we will require various latency for different tasks and those have varying costs. These tables are used as look ups for the cost models in Section 2

Table 7: Storage types and costs used as inputs used for calculations

| Storage type | cost |
|---|---|
| fast – NVME (50GB/ s each) / TB | $1,000.00 |
| normal - SATA GPFS file systems/ TB | $135.00 |
| latency – slower but on disk | $100.00 |
| high latency – very slow – on tape | $64.00 |
| | |

In Table 7 we should consider for NVME for each TB with file system servers two DDN NVME box with GPFS servers. The price is based on the TOP performer with best price . The Normal price is for each TB with file system disks and servers locally attached to production resources.

In the latency and high latency prices are only at NCSA: for each TB with file systems and all people/services. The complete service not usually attached. S3 bucket type. Can be mounted if needed but not for production worthy speeds. The complete service with data flowing to tape using policies.

Table 8: Machine types and costs used as inputs used for calculations

| Type of machine | Cores | Memory(GB) | GFLOP/ s | Cost | purpose/ use |
|---|---|---|---|---|---|
| xeon | 32 | 192 | 913.92 | $10,000.00 | current K8 node |
| qserv | 12 | 128 | 408 | $20,000.00 | current qserv node |
| small rome | 64 | 256 | 1261.4 | $13,000.00 | https:/ / www.microway.com/ product/ navion-1u-amd-epyc-gpu-server/ |
| large rome | 128 | 512 | 3916.8 | $23,700.00 | |
| current compute node | 24 | 128 | 816 | $9,000.00 | current compute node |
| | | | | | |

There is also an associated running cost for machines included in the total cost of ownership. These overheads are listed in Table 9.

Table 9: Overhead costs per rack

| Item | Number/ Cost |
|---|---|
| Compute nodes in a rack | 36 |
| Rack initial cost has power, networking switches, networking cables, ready for machine installation | $24,000.00 |
| recurring after 1st year costs (power, cooling, licenses, floor space) | |
| Cost for each machine | $666.67 |

# 5 Sizing inputs

The following simplified sizing was used to give the input sizes for the cost model in Section 2. The storage sizes are given in Table 12 while the compute is given in Table 13.

## 5.1 Storage Model

Table 10: Inputs used to calculate storage needs

| Parameters | unit | FY2020 | FY2021 | FY2022 | FY2023/ LOY1 | Notes |
|---|---|---|---|---|---|---|
| Objects | number | | | 4.58E+09 | 2.75E+10 | from LSE-81, scaled to 2 months for 2022, ComCam ignored |
| Sources | number | | | 1.50E+11 | 9.01E+11 | from LSE-81, scaled to 2 months for 2022, ComCam ignored |
| ForcedSources | number | | | 4.85E+11 | 2.91E+12 | from LSE-81, scaled to 2 months for 2022, ComCam ignored |
| Science users | users | 50 | 100 | 5000 | 5000 | “Stack Club” to 2021, DP users thereafter |

| | | | | | | |
|---|---|---|---|---|---|---|
| Storage per science user | TB | 0.005 | 0.01 | 0.05 | 0.1 | ramp; includes oversubscription |
| LSSTCam image size | TB | 0.0152 | | | | uncompressed, 32 bit, with overscan and corner rafts |
| Raw image compression | factor | 0.42 | | | | lossless-compressed divided by uncompressed for raws |
| Lossy image compression | factor | 0.250 | | | | lossy-compressed divided by lossless-compressed for PVIs |
| Observing nights per year | nights | 300 | | | | maximum |
| Visits per night | visits | 1000 | | | | maximum |
| Images per visit | images | 2 | | | | |
| Calibration images per day | images | 500 | | | | |
| LSSTCam Science images | images | | | 100000 | 600000 | test images until 2 months of science in 2022 |
| LSSTCam Test images | images | 25000 | 50000 | 50000 | | ramp to science images |
| LSSTCam Engineering images | images | 12500 | 12500 | 15000 | 6000 | decreasing ramp |
| LSSTCam Calibration images | images | 12500 | 25000 | 37500 | 150000 | estimates based on science and test images; actual for 2023 |
| Object table row size | bytes | 1840 | | | | from LDM-141 |
| Object_Extra tables row size | bytes | 20393 | | | | from LDM-141 |
| Source table row size | bytes | 453 | | | | from LDM-141 |
| ForcedSource table row size | bytes | 41 | | | | from LDM-141 |
| Qserv replication factor | factor | 3.0 | | | | |

### 5.1.1 Overview

This simplified storage model eliminates many details in the previous storage model [LDM-141] that end up being insignificant. There are relatively few data products that require significant amounts of fast SSD or slower disk or tape storage; the others complicate the model without giving much insight. In addition, it is assumed that bandwidth is not a significant constraint, other than the distinction between SSD and spinning disk. With the advent of highly-parallel shared and object storage, having large numbers of spindles solely to achieve high bandwidth for certain operations is not thought to be necessary.

Values are computed for the amount of storage expected to be ”on the floor” at the beginning of each fiscal year from FY2020 through FY2023 (which is LSST Operations Year 1). Not included is any storage already present at the end of FY2019 holding past data.

This version of the model assumes that all raw science images, all Commissioning processed visit images, and the first year’s processed visit images are kept on spinning disk. Initially, all raw images and image output data products are placed on ”normal” filesystem disk; after 1 year, they are assumed to move to object storage.

All data is backed up to tape permanently, including annual snapshots of filesystems. Any incremental backups are assumed to be reusable or otherwise purged and hence not significant.

### 5.1.2 Parameters

The numbers of Objects, Sources, and ForcedSources are taken from LSE-82, with the FY2022 numbers reduced by a factor of 2/12 to account for the anticipated 2 months of on-sky science validation time for LSSTCam before the survey begins. These numbers are ultimately based on models for stars in the galaxy and galaxies in the universe that are dependent on the limiting magnitude achieved in each year and are listed in Table 10

The numbers of science users are estimates, using ”Stack Club” users and Commissioning users for FY2020 and 2021, followed by US science users in FY2022 and FY2023 for Data Preview data. The bulk of US science users are not expected to arrive until after Data Release 1 at the beginning of FY2024.

Storage per science user is estimated based on today’s usage at NCSA, scaled up as users become more active.

The LSSTCam image size is uncompressed and includes overscan, 4 bytes of raw data per pixel, and both science and corner rafts.

The raw image compression factor was measured on simulated LSST images. The lossy image compression factor for processed visit images is the ratio between the lossy-compressed file size (estimated at 1/6 of uncompressed) and the lossless-compressed file size (estimated at 66% of uncompressed).

The number of observing nights per year and the number of visits per night are maximal estimates. 2 images per visit is still the baseline and a possibility that must be accounted for. The number of calibration images per day was derived from the calibration plan.

As stated above, the number of LSSTCam science images is scaled by 2/12 for FY2022 given the length of science validation time. The number of test images is estimated as a ramp up to the full science cadence. The numbers of engineering and calibration images are estimated as ramping-down fractions of the number of science and test images, with calibration images ending at the number per day given previously.

Sizes of rows in various data product tables is taken from LDM-141, which was in turn derived from the DPDD.

Qserv replicates its data for fault tolerance; a typical replication factor is selected here.

#### 5.1.3 Data Product Sizing

Images and the results of processing them are the dominant factor controlling storage sizing which is outlined in Table 11. Precursor survey and LSSTCam images are the largest; ComCam, at less than 5% of the size of LSSTCam and with little on-sky science time is negligible, as is LATISS, which is less than 1% of the size of LSSTCam, though it has considerable on-sky time.

The sizing of the Alert Production Database (APDB) is based on experiments in Salnikov (DMTN-113) which found that 57,000 visits took 4.5 TB including indexes. A simple linear scaling to a full year's visits was performed, with half that purchased in 2020 for large (but not full) scale testing.

HyperSuprime-Cam (HSC) RC2 is a relatively small dataset used for monthly processing tests. The size of the input images was taken from Wood-Vasey et al. (DMTN-091); the size of the outputs (image and Parquet/other non-image files) was measured from the latest execution. A similar size dataset based on DESC DC2 is assumed to be being used for an additional monthly processing test. Note that this is a very small subset of the full DESC DC2, which is expected to cover 300 square degrees to 10-year LSST depth (approximately 1000 epochs per point on the sky). The full DESC DC2 is not currently scheduled to be reprocessed by the construction team. Instead, twice-a-year processings of the full HSC SSP PDR2 dataset are assumed to occur. The size of this dataset was also taken from Wood-Vasey et al. (DMTN-091); it is 5654 visits of 104 CCDs, each of which occupies 18.2 MB.

Output sizes are assumed to scale linearly with input size, and by the same factor for each instrument.

Scratch space is set at 10% of the output image storage for LSSTCam processing; it is assumed to be already present for precursor processing.

Qserv Czar fast (SSD) storage is assumed to be used for the primary Object table; additional space for the so-called "secondary index" mapping object identifiers to spatial chunks is negligible in comparison.

The main Qserv database storage is based on the Parquet file sizing for precursor data and on the estimated numbers of Objects, Sources, and ForcedSources for LSSTCam data.

Note that no space is explicitly reserved for Qserv query result storage.

An additional 20% disk and tape storage is added to account for all other needs.

Table 11: Inputs on dataset sizes used to calculate storage needs

| Dataset Sizing | unit | FY2020 | FY2021 | FY2022 | FY2023/ LOY1 | Notes |
|---|---|---|---|---|---|---|
| HSC RC2 Area | deg2□ | 3.0 | | | | |
| HSC SSP PDR2 Area | deg2□ | | | | | |
| DESC DC2 Area | deg2□ | 300 | | | | |
| LSSTCam Area | deg2□ | | | 2000 | 17000 | |
| APDB | TB | 12 | 24 | 24 | 24 | 4.5/ 57K TB per visit; 1 year retention; 6 months in 2020 |
| HSC RC2 Input Images | TB | 0.8 | 0.8 | 0.8 | 0.8 | 428 visits * 104 CCDs * 18.2 MB uncompressed |
| HSC RC2 Output Images | TB | 2.4 | 2.4 | 2.4 | 2.4 | lossless-compressed, not including warps |
| HSC RC2 Output Coadd Images | TB | 0.7 | 0.7 | 0.7 | 0.7 | lossless-compressed |
| HSC RC2 Output Catalogs | TB | 1.4 | 1.4 | 1.4 | 1.4 | |
| HSC SSP PDR2 Input Images | TB | 27.4 | 27.4 | 27.4 | 27.4 | 14476 visits * 104 CCDs * 18.2 MB uncompressed (2 * PDR1) |
| DESC DC2 Input Images | TB | 455 | 455 | 455 | 455 | 300 sq deg, 10 year depth |
| LSSTCam Raw Images | TB | 319 | 557 | 1290 | 4816 | compressed, moves to object store |
| Precursor Output Images | TB | 236 | 236 | 236 | 236 | monthly RC2 and DC2 subset plus biannual PDR |
| Precursor Output Parquet | TB | 130 | 130 | 130 | 130 | |
| LSSTCam Output Images | TB | | | 2248 | 13485 | lossless-compressed, moves to object store |
| LSSTCam Output Coadd Images | TB | | | 455 | 3864 | |
| LSSTCam Output Parquet | TB | | | 1329 | 7973 | |
| Scratch | TB | | | 225 | 1349 | 10% of output images |
| Qserv Czar/ Object | TB | | | 8 | 51 | based on row sizes and counts |
| Qserv Database | TB | 394 | 394 | 569 | 3417 | based on Parquet for preliminary; based on row sizes and counts |
| Science User Home | TB | 0 | 1 | 250 | 500 | |
| Other/ Misc | TB | 316 | 366 | 1450 | 7266 | 20% of total |

### 5.1.4 Storage Sizing

Finally, storage is allocated to specific types as shown in Table 12. Fast storage (SSD) is used for the APDB and Qserv Czar, which accumulates data from year to year until Data Releases are retired. Normal storage is used for inputs, scratch, and output images (initially). It is also used for Qserv database storage, which accumulates from year to year. Object storage is used for output tables each year and output images after one year. Lossy compression is applied at this time. Since only one year of operational processing is in the model, nothing is removed from the object store; it accumulates from year to year. Tape is used for long-term archiving and filesystem backup. Again, this accumulates from year to year.

Note that no replication is assumed in the object store.

Table 12: On floor storage estimates based on Table 11 and Table 10

| Storage Sizing (on the floor) | unit | FY2020 | FY2021 | FY2022 | FY2023/ LOY1 | Notes |
|---|---|---|---|---|---|---|
| Fast | TB | 12 | 24 | 32 | 83 | SSD |
| Normal | TB | 1883 | 2173 | 8678 | 44152 | Enterprise SATA |

| | | | | | | |
|---|---|---|---|---|---|---|
| Object Store | TB | 319 | 876 | 4057 | 20217 | |
| Tape | TB | 1883 | 4056 | 12734 | 56886 | |

## 5.2 Compute Model

Table 13: Inputs used to calculate compute needs

| Parameters | units | | | | | Notes |
|---|---|---|---|---|---|---|
| Max FLOP/ sec per core | FLOP/ core/ sec | 4.0E+10 | | | | E5-2680 v3 @ 2.50GHz, 16 FLOPs/ cycle |
| Sustained efficiency | FLOP/ cycle | 13.60 | | | | |
| Sustained FLOP/ sec per core | FLOP/ core/ sec | 3.4E+10 | | | | |
| HSC PDR1 Input Images | TB | 13.7 | | | | 7238 visits of 104 CCDs |
| HSC PDR1 small-memory compute | core-hours | 64392 | | | | measured |
| HSC PDR1 high-memory compute | core-hours | 78523 | | | | measured |
| Additional DRP steps | factor | 1.5 | | | | image differencing, stackfit, etc. |
| DRP FLOPs per TB of input visits | FLOP/ TB | 3.2E+18 | | | | based on sustained FLOPS |
| ap_pipe sec/ CCD | core-sec/ CCD | 83 | | | | measured |
| Additional AP steps | factor | 0.25 | | | | DCR, real_bogus, etc. |
| AP FLOPs per visit | FLOP | 6.7E+14 | | | | based on sustained FLOPS |

### 5.2.1 Overview

This simplified computing model divides computation into three classes: Data Release Production (DRP), Alert Production, and LSST Science Platform. Calibration Products Production is assumed to be negligible.

The pipelines have advanced considerably in terms of fidelity and science performance since the previous computing model [LDM-138] was developed. Scaling compute needs based on an execution of the nascent DRP pipeline on HSC PDR1 data and nightly executions of the nascent ap_pipe pipeline on HiTS2015 data is thus appropriate, but the fact that several steps are still missing from these pipelines must be taken into account.

Times are measured on existing hardware. Given an assumed efficiency ratio specifying the number of floating point operations (FLOPs with lowercase ”s”) per clock cycle, the number of sustained FLOPs/sec (also written FLOPS with uppercase ”s”) can be computed. This number is then multiplied by the wall-clock time and number of cores to determine the total FLOPs for a pipeline executing on a dataset. This estimation methodology incorporates all I/O, memory bandwidth, cache miss, and other overheads into the single efficiency ratio, simplifying calculations.

### 5.2.2 Parameters

DRP executes on the verification cluster, which uses Intel Xeon E5-2690v3 CPUs at 2.6 GHz. The Alert Production executes on Kubernetes nodes, which are a bit slower; to be conservative, this is neglected.

The most recent run of DRP on HSC PDR1 data is described at `https://confluence.lsstcorp.org/x/WpBiB`. The input data is the same size as PDR2 from the storage sheet. Most jobs (but not most of the time) could run on relatively small-memory machines with 24 cores and 5 GB RAM per core. The largest and longest-running jobs, however, required up to 4 times as much memory, using half or a quarter of the cores. To be conservative, we assume that half the cores were used for the large-memory jobs. Since the HSC PDR1 processing did not include several steps from the Science Pipelines Design document [LDM-151] such as image differencing and full multi-epoch characterization, the time and FLOPs used are scaled up to the expected pipeline consumption.

The SQuaSH system reports the execution time of `ap_pipe` in seconds per CCD. A mean was taken over all processed CCDs, and it was assumed that each CCD is processed on a single core. A factor is added to account for additional steps like differential chromatic refraction compensation and false positive detection that are not well-represented in the current pipeline. Multiplying by the number of LSSTCam science CCDs and the sustained FLOPS per core gives the total number of floating point operations used per LSSTCam visit.

### 5.2.3 Data Release Production

The number of floating point operations per TB of input data is multiplied by the precursor (HSC RC2 and DESC DC2 subset for 12 months and HSC PDR2 twice a year) and LSSTCam input data sizes to determine the total number of FLOPs needed in each year, this is shown in Table 14. Approximately half of these FLOPs need to be provided by small-memory (4-5 GB/core) machines; the other half needs to come from large-memory (20 GB/core) machines.

Table 14: Inputs used to calculate compute needs for DRP

| Data Release Production | units | FY2020 | FY2021 | FY2022 | FY2023/ LOY1 | Notes |
|---|---|---|---|---|---|---|
| Precursor Input Size | TB | 74 | 74 | 74 | 74 | |
| LSSTCam Visit Input Size | TB | | | 758 | 4550 | raw images / images/ visit |
| Precursor FLOPs | FLOP | 2E+20 | 2E+20 | 2E+20 | 2E+20 | |
| LSSTCam FLOPs | FLOP | | | 2E+21 | 1E+22 | |
| **Total FLOPs** | **FLOP** | **2E+20** | **2E+20** | **3E+21** | **1E+22** | |

### 5.2.4 Alert Production

The floating point operations per visit are divided by the minimum visit length (30 sec plus 1 sec shutter motion plus 2 sec readout) to give the minimum FLOP/sec rate needed to keep up with image taking this is shown in Table 15. This could be provided over multiple ”strings” of nodes, at increased latency to delivery of alerts, if single cores are not fast enough.

Table 15: Inputs used to calculate compute needs for Alert production

| Alert Production | units | FY2020 | FY2021 | FY2022 | FY2023/ LOY1 | Notes |
|---|---|---|---|---|---|---|
| AP FLOPs | FLOP | | | 6.7E+14 | 6.7E+14 | |
| AP FLOP/ sec | FLOP/ sec | | | 2.0E+13 | 2.0E+13 | minimum necessary to keep up |
| AP FLOP/ core/ sec | FLOP/ core/ sec | | | 1.1E+11 | 1.1E+11 | assuming 1 core/ CCD |

### 5.2.5 LSST Science Platform

LSST Science Platform needs for external science users are derived as 10% of the DRP FLOP requirement and is shown in Table 16. The LSP floating point operations are assumed to be spread over a year, giving a mean FLOP/sec rate. As a reasonableness check, the number of FLOP/sec per science user is computed, but it must be noted that an oversubscription factor needs to be taken into account, since not all users are expected to be simultaneously active.

Table 16: Inputs used to calculate compute needs for the Science Platform

| LSST Science Platform | units | FY2020 | FY2021 | FY2022 | FY2023/ LOY1 | Notes |
|---|---|---|---|---|---|---|
| LSP FLOP/ sec | FLOP/ sec | | | 8.4E+12 | 4.7E+13 | 10% of DRP, over a year |
| LSP FLOP/ sec/ science user | FLOP/ sec/ user | | | 1.7E+09 | 9.4E+09 | includes oversubscription |

# A References

## References

**[LDM-141]**, Becla, J., Lim, K.T., 2013, *Data Management Storage Sizing and I/O Model*, LDM-141, URL https://ls.st/LDM-141

**[LSE-82]**, Dubois-Felsmann, G., Lim, K.T., 2013, *Science and Project Sizing Inputs Explanation*, LSE-82, URL https://ls.st/LSE-82

**[LDM-144]**, Freemon, M., Pietrowicz, S., Alt, J., 2016, *Site Specific Infrastructure Estimation Model*, LDM-144, URL https://ls.st/LDM-144

**[LDM-138]**, Kantor, J., Axelrod, T., Lim, K.T., 2013, *Data Management Compute Sizing Model*, LDM-138, URL https://ls.st/LDM-138

**[DMTN-113]**, Salnikov, A., 2019, *Performance of RDBMS-based PPDB implementation*, DMTN-113, URL https://dmtn-113.lsst.io, LSST Data Management Technical Note

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

**[DMTN-091]**, Wood-Vasey, M., Bellm, E., Bosch, J., et al., 2019, *Test Datasets for Scientific Performance Monitoring*, DMTN-091, URL https://dmtn-007.lsst.io/v/DM-15448/index.html, LSST Data Management Technical Note

## B Acronyms

| Acronym | Description |
|---|---|
| DB | DataBase |
| DBB | Data Back Bone |
| DDN | Data Delivery Network |
| DM | Data Management |
| DMTN | DM Technical Note |
| FLOP | FLoating point Operation |
| FLOPS | FLoating point Operation per Second |
| GFLOP | Giga FLOP |
| GPFS | General Parallel File System (now IBM Spectrum Scale) |
| LDF | LSST Data Facility |
| LDM | LSST Data Management (Document Handle) |
| LSP | LSST Science Platform |
| NCSA | National Center for Supercomputing Applications |
| NVME | Non Volatile Memory Express."DM IT" |
| Qserv | LSST's distributed parallel database. This database system is used for collecting, storing, and serving LSST Data Release Catalogs and Project metadata, and is part of the Software Stack |
| SATA | Serial Advanced Technology Attachment |
| TB | TeraByte |