Rubin Observatory

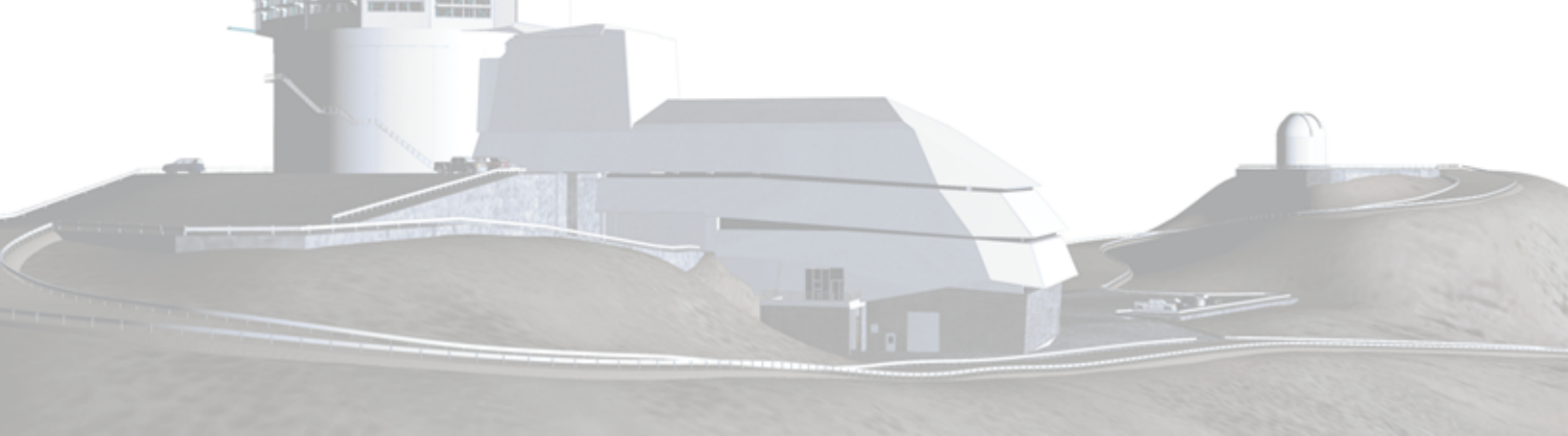
Vera C. Rubin Observatory
Systems Engineering

Batch Production Service Requirements

Mikolaj Kowalik, Michelle Gower, Rob Kooper

LDM-636

Latest Revision: 2019-06-11



Rubin Observatory

Change Record

Version	Date	Description	Owner name
	2019-03-07	Initial version	M. Gower
1.0	2019-07-16	First release, approved in RFC-602	M. Gower

Document source location: MagicDraw SysML
Version from source repository: 239

Rubin Observatory

Contents

Batch Production Service Requirements

This document describes the requirements relating to the Batch Processing Services. These services should not be confused with IT-level batch services like PBS, Slurm, and HTCondor. The LSST Batch Production Services are a layer that sits above the IT-level batch services that executes and manages science payloads as ”campaigns” consisting of a defined pipeline, a defined configuration, and defined inputs and outputs.

This document does not cover Operation’s requirements that are outside of the scope of the Batch Production Services, e.g., having change board approval (or pre-approved rules) prior to changing a Production configuration. Where Use Cases are mentioned, they are defined in LDM-633 which also includes a glossary.

1 Processing Reliability

ID: DMS-BPS-REQ-0001

Specification: Except in cases of major disaster, the BPS shall have no unscheduled outages of the DMS pipelines extending over a period greater than productionMaxDowntime. A major disaster is defined as a natural disaster or act of war (e.g. flood, fire, hostile acts) that compromises or threatens to compromise the health and integrity of the DMS physical facility computing equipment, or operational personnel.

2 BPS Unscheduled Downtime

ID: DMS-BPS-REQ-0002

Specification: The BPS shall be designed to facilitate unplanned repair activities expected not to exceed DMDowntime days per year.

3 Processing Throughput

ID: DMS-BPS-REQ-0003

Specification: The BPS shall dispatch and manage pipelines at scale needed to meet LSST objectives within their time constraints. Each of the pipelines have their unique requirements as specified in LSE-81 (rows 215 through 223).

Discussion: The BPS shall be designed to facilitate unplanned repair activities expected not to exceed DMDowntime days per year.

4 Nightly Data Accessible Within Specified Time

ID: DMS-BPS-REQ-0004

Specification: The BPS shall be capable of executing the offline Prompt Production pipelines in a time no greater than **L1PublicT**, without impacting observatory operations. It includes catch-up of missed nightly processing as well as daytime processing such as the Moving Object Processing System.

Discussion: This will put a requirement on the available resources at any point in the future. **L1PublicT** time also includes DBB replication time.

5 Calibration Images Available Within Specified Time

ID: DMS-BPS-REQ-0005

Specification: The BPS shall be capable of executing the Daily Calibration Products Update payload producing Calibration products from a group of up to **nCalExpProc** related exposures that should be processed together making the outputs available from the DMS image archive within **calProcTime** of the end of the acquisition of images/data for that group.

Discussion: This will put a requirement on the available resources at any point in the future. **calProcTime** includes data transfer times external to the BPS. It also includes DBB replication

time.

6 Pre Runtime

6.1 Configuration Validation

ID: DMS-BPS-REQ-0006

Specification: The BPS shall verify validity of the BPS configuration of a (sub)campaign to the extent possible at submit time.

Discussion: This will not guarantee the science validity of the actual campaign execution or that it will successfully execute on particular platforms or input datasets, but should show that the campaign is generically capable of being executed, and there are no missing configurations.

6.2 Configuration Overrides

ID: DMS-BPS-REQ-0007

Specification: The BPS shall allow the Operator to override the configuration options at any level, i.e., global, site, campaign, payload, pipeline step definition.

Discussion: These overrides include but are not limited to: execution configuration (e.g. memory needed, computational platforms to use or avoid), science configuration (e.g. updating particular threshold), excluding certain data from a campaign. Differentiating between changes requiring change board approval vs. pre-approved is beyond the scope of this document and should be addressed in higher level document concerned with Operation requirements. (Use Case: PRE1)

6.3 Pipeline Modification

ID: DMS-BPS-REQ-0008

Specification: The BPS shall allow the Operator to modify sequence of pipeline step definitions. Modification includes addition of new pipeline step definitions, deleting, and reordering existing ones for science reasons.

Discussion: Operations may require to run either only few first steps in a larger pipeline to produce outputs for debugging or few of its last steps starting from previous executions. (Use Case: PRE2)

6.4 Software selection

ID: DMS-BPS-REQ-0009

Specification: The BPS shall allow the Operator to submit (sub)campaigns using specific versions of packages from the LSST Software Stack.

Discussion: This does not require the BPS to accept lists of individual packages. For example, BPS could track container versions or personal EUPS tables both of which appear to BPS as a single entity yet allowing mix-and-match of individual software packages. The goal is to have regimented tracking for production, yet allow flexibility for development. (Use Case: PRE3)

6.5 Execution order

ID: DMS-BPS-REQ-0010

Specification: The BPS shall allow the Operator to set the order of a workflow execution.

Discussion: For example, choose between breadth-first search and depth-first search order. The default shall be breadth-first search order. (Use Case: PRE4)

6.6 Dynamic Workflow Generation

ID: DMS-BPS-REQ-0011

Specification: The BPS shall create workflows dynamically in locations designated at submit time.

Discussion: It is not required that the BPS should be able to do this efficiently for every job in the workflow. It is expected that there will be only a few if any such locations needed by pipelines. The dynamic generation is related to data discovery only. It does not include a support for conditional execution or rerunning jobs or parts of the workflow based on some quality metrics. May not need if Pipelines guarantee to always generate output dataset and succeed in cases of non-fatal error. (Use Case: PRE5)

6.7 Reducing Number of Jobs

ID: DMS-BPS-REQ-0012

Specification: The BPS shall allow the Operator to group portions of the workflow into smaller number of compute jobs.

Discussion: (Use Case: PRE6)

6.8 Conditional Workflow Aborts

ID: DMS-BPS-REQ-0013

Specification: The BPS shall allow the operator to specify on a per pipeline step definition basis whether to stop or continue workflow execution on failures.

Discussion: May not need if Pipelines guarantee to always generate output dataset and succeed in cases of non-fatal error. (Use Case: PRE7)

6.9 Selective Dataset Persistence

ID: DMS-BPS-REQ-0014

Specification: The BPS shall allow the Operator to specify which datasets to persist to Data Backbone

Discussion: For debugging purposes, the Operator may want to inspect files which are not collected during regular payload executions. (Use Case: PRE8)

6.10 Unexpected Data Handling

ID: DMS-BPS-REQ-0015

Specification: The BPS shall allow the Operator to specify whether to save any unexpected (or faulty) files.

Discussion: A files is considered to be unexpected (or faulty) when they are in a wrong place, have a wrong name, or are missing required metadata. (Use Case: PRE9)

6.11 Campaign Prioritization

ID: DMS-BPS-REQ-0016

Specification: The BPS shall allow the Operator to prioritize (sub)campaigns at the time of submission.

Discussion: (Sub)campaigns with higher priorities will be dispatched for execution before those with lower ones. (Use Case: PRE10)

6.12 Processing of Data from Special Programs

ID: DMS-BPS-REQ-0017

Specification: The BPS shall be able to process special programs data with the offline production pipelines alongside data from the main survey.

Discussion: See DMS-REQ-0320's discussion for limitations

6.13 Campaign Submission

ID: DMS-BPS-REQ-0018

Specification: The BPS shall allow the Operator to submit many payloads that can be managed as a single group (a campaign).

6.14 Programatic Submission

ID: DMS-BPS-REQ-0019

Specification: The BPS shall provide an API allowing Operators to submit campaigns/pipelines programmatically.

Discussion: (Use Case: PRE11)

6.15 Simultaneous Processing

ID: DMS-BPS-REQ-0020

Specification: The BPS shall be able to run multiple workflows originating from different campaigns/payloads simultaneously without impacting each other and observatory operations.

Discussion: Not having enough resources available can affect offline efficiency.

7 Runtime

7.1 Immediate Termination

ID: DMS-BPS-REQ-0056

Specification: The BPS shall allow the Operator to immediately terminate processing of a (sub)campaign.

Discussion: (Use Case: RUN1)

7.2 Pausing Campaign

ID: DMS-BPS-REQ-0021

Specification: The BPS shall allow the Operator to pause dispatching new workflows/jobs

associated with a selected running (sub)campaign.

Discussion: The freeing of resources can be done to start campaigns of a higher priority or for maintenance. This does not include the ability to change configuration, etc. for the (sub)campaign (see restarts). Either killing the job resulting in a controlled failure or waiting for the job to finish, but not starting any new jobs is sufficient to meet this requirement. (Use Case: RUN2)

7.3 Alter Processing Resource

ID: DMS-BPS-REQ-0022

Specification: The BPS shall allow the Operator to assign new compute resources to a (sub)campaign put on hold.

Discussion: This does not include the ability to change configuration, etc. for the (sub)campaign (see restarts). (Use Case: RUN3)

7.4 Alter Processing Priority

ID: DMS-BPS-REQ-0023

Specification: The BPS shall allow the Operator to alter the priority for a pending (sub)campaign.

Discussion: (Use Case: RUN5)

7.5 Processing Resource Selection

ID: DMS-BPS-REQ-0024

Specification: If explicitly configured at submit time to run on specific computational resources (platforms or machines), the BPS shall dispatch workflows/jobs only to those resources.

Discussion: (Use Case: RUN4)

7.6 Processing Resource Avoidance

ID: DMS-BPS-REQ-0025

Specification: If explicitly configured at submit time to avoid running on specific computational resources (platforms or machines), the BPS shall not dispatch workflows/jobs to those resources.

Discussion: (Use Case: RUN4)

7.7 Processing Platform Diversity

ID: DMS-BPS-REQ-0026

Specification: The BPS shall support execution of workflows/jobs on computational platforms with varying architectures.

Discussion: The BPS will support computational platforms with or without the shared filesystem, e.g., the LSST verification cluster as well as clusters in CC-IN2P3. (Use Case: RUN4)

7.8 Automatic Avoidance of Faulty Resources Using BPS Data

ID: DMS-BPS-REQ-0027

Specification: The BPS shall use failure information from workflow executions to determine if a resource is suspicious and automatically be avoided for future work.

Discussion: Must be configurable (on/off, threshold level). Failing resources (e.g. nodes for which number of failed jobs exceeds a given threshold) will be marked and excluded from the pool of available resources. (Use Case: RUN10)

7.9 Reporting Faulty Resources

ID: DMS-BPS-REQ-0028

Specification: The BPS should provide data to LSST Monitoring Services regarding suspected faulty resources in cases where not reportable by other means.

7.10 Automatic Avoidance of Faulty Resources Using LSST Monitoring Services

ID: DMS-BPS-REQ-0029

Specification: The BPS should use information from LSST Monitoring Services to determine if a resource is suspicious and automatically be avoided for future work.

Discussion: Some monitoring information may be complicated to interpret to mean that the resource is unusable for BPS. (Use Case: RUN10)

7.11 Pipeline Middleware

ID: DMS-BPS-REQ-0030

Specification: The BPS shall be able to execute (sub)campaigns which use the current stable version of Pipeline Middleware.

Discussion: The initial generation for BPS development will be Gen3. (Use Case: RUN6)

7.12 Optional Datasets

ID: DMS-BPS-REQ-0031

Specification: The BPS shall support execution of workflows where pipeline step instances can have optional inputs and outputs.

Discussion: Certain steps in a pipeline can proceed even when not all declared inputs are present. The BPS should continue pipeline execution in such cases providing the minimal requirements are met. May not need if Pipelines guarantee to always generate output dataset and succeed in cases of non-fatal error. (Use Case: RUN7)

7.13 Automatic Retries

ID: DMS-BPS-REQ-0032

Specification: The BPS shall automatically retry failing jobs/pipeline step instances providing certain criteria are met.

Discussion: The preference is to rerun the minimal portion of the workflow necessary, but it may involve rescheduling a job to a different node or platform. (Use Case: RUN8)

7.14 Persisting Datasets When Retries

ID: DMS-BPS-REQ-0033

Specification: The BPS shall upload only the files from the final retry to the Data Backbone.

Discussion: (Use Case: RUN8, RUN9)

7.15 Persisting Logs and Standard Streams

ID: DMS-BPS-REQ-0034

Specification: Depending on the configuration, the BPS shall upload files with logging information, stdout, and stderr from each retry to the Data Backbone.

Discussion: (Use Case: RUN9)

7.16 Staging Files

ID: DMS-BPS-REQ-0035

Specification: For computing platforms that have staging areas for pre/post job file transfers, the BPS shall transfer datasets between these areas and the Data Backbone.

Discussion: (Use Case: RUN9)

7.17 Managing Staging Areas

ID: DMS-BPS-REQ-0036

Specification: The BPS shall manage its file staging areas.

Discussion: It will remove files that are no longer needed to make room for new files.

7.18 Managing Input Datasets

ID: DMS-BPS-REQ-0037

Specification: Inside a compute job, the BPS shall transfer input data required by the pipeline step instances to the computing node on which they are being executed.

Discussion: The transfers are from either a staging area or in the case of no staging area the Data Backbone directly. (Use Case: RUN9)

7.19 Managing Output Datasets

ID: DMS-BPS-REQ-0038

Specification: Inside a compute job, the BPS shall transfer output data produced by the pipeline setup instances from the computing node on which they were executed.

Discussion: The transfers are to either a computing platform staging area or in the case of no staging area the Data Backbone directly. (Use Case: RUN9)

7.20 Output Verification

ID: DMS-BPS-REQ-0039

Specification: Based on payload configuration, the BPS shall verify if the required outputs were generated and properly persisted for each successful workflow.

Discussion: (Use Case: RUN12)

7.21 Multinode Support

ID: DMS-BPS-REQ-0040

Specification: The BPS shall be able to spread the execution of independent pipeline steps across multiple nodes based upon operational configuration.

Discussion: It does not imply that the BPS will support execution of tasks using MPI (i.e., co-scheduling tasks on different machines).

7.22 Collecting Provenance

ID: DMS-BPS-REQ-0041

Specification: The BPS shall record workflow provenance information for each pipeline step instance including any retries.

Discussion: The provenance information will be detailed enough to allow for later re-runs with the same data dependencies, but not necessarily in the same exact order of execution. The BPS system can execute a task multiple times, for debugging and provenance we should track the multiple executions. (Use Case: RUN8)

7.23 Runtime Monitoring

ID: DMS-BPS-REQ-0042

Specification: The BPS shall record and monitor runtime metrics for workflows/jobs/pipeline step instances including:

- number of pending, running, finished, and failed job/pipeline step instances;
- amount of computer resources (e.g., CPUs, memory, disk space) in use vs idle;

- job runtime information (e.g., host name, memory, wall/CPU time, data input and output volume);
- what job, pipeline step instance, or framework step is currently running (if possible seeing stdout/stderr from the step) for when pipelines seem to be taking too long.

Discussion: Some of these metrics will be stored permanently. Existing monitoring tools will be leveraged when possible. (Use Case: RUN11)

7.24 Tracking Individual Tasks

ID: DMS-BPS-REQ-0043

Specification: The BPS shall track execution of individual pipeline step instances which were grouped into a single compute job.

Discussion: Tracking information may or may not be available during runtime. (Use Case: RUN11)

7.25 Monitoring API

ID: DMS-BPS-REQ-0044

Specification: The BPS shall provide an API that allows the Operator to programmatically monitor the progress of workflows/jobs.

Discussion: (Use Case: RUN11)

7.26 Event Notifications

ID: DMS-BPS-REQ-0045

Specification: The BPS shall be capable of sending notifications about (sub)campaign level events.

Discussion: The events include (sub)campaign failure, completion, jobs taking longer than some Operator-specified threshold. (Use Case: RUN13)

8 Post Runtime

8.1 Provenance-Based Reruns

ID: DMS-BPS-REQ-0046

Specification: The BPS shall have the ability to re-run a (sub)campaign based on provenance information.

Discussion: This means repeating the same steps on same inputs and configurations in the same data dependency order, but it does not imply identical order of executions.

8.2 Automatic Restart Designation

ID: DMS-BPS-REQ-0047

Specification: The BPS shall automatically designate points of failures as restart points.

8.3 Manual Restart Point Designation

ID: DMS-BPS-REQ-0048

Specification: The BPS shall allow the Operator to designate the point of restart.

Discussion: Sometimes a pipeline step needs to be rerun despite not having explicitly failed during previous processing attempt. (Use Case: POST2)

8.4 Campaign Restarts

ID: DMS-BPS-REQ-0049

Specification: The BPS shall allow the Operator to restart a (sub)campaign execution from the designated restart point.

Discussion: Restarting a pipeline may include one or more of the following: changing software stack to be used, changing science configuration, changing execution configuration, etc. (Use Case: POST1, POST2)

8.5 Campaign Comparison

ID: DMS-BPS-REQ-0050

Specification: The BPS shall persist the data from both successful and failed (sub)campaigns such that the Operator can use similar procedures and tools for both.

Discussion: (Use Case: POST3, POST6)

8.6 Programatic Access to Runtime Metrics and Provenance

ID: DMS-BPS-REQ-0051

Specification: The BPS shall provide an API that allows the Operator to programmatically view any runtime metrics and provenance saved.

Discussion: (Use Case: POST4)

8.7 Runtime Monitoring

ID: DMS-BPS-REQ-0052

Specification: For each execution step (which includes pipeline step instances, scheduling, data transfer, data loading, workflow generation, etc), the BPS shall provide access to information including at least:

- machine it was executed on;
- when the step was running and time it took to complete;
- amount of memory a process used during its execution;
- version of software used;
- job's environment;
- what input datasets were used (if applicable);
- what output datasets were produced (if applicable);
- whether the step finished successfully or failed (from the BPS perspective).

Discussion: (Use Case: POST4)

8.8 Persisting Logs and Standard Streams

ID: DMS-BPS-REQ-0053

Specification: The BPS shall persist stderr/stdout/log files of finished jobs.

Discussion: The minimum to fulfill this requirement is access to files easily discoverable for a specific pipeline step instance. Ideally, the data would be stored in a manner to make analyzing patterns easy. (Use Case: POST5)

8.9 Campaign Summary

ID: DMS-BPS-REQ-0054

Specification: The BPS shall provide a summary of a (sub)campaign containing execution status and runtime metrics.

Discussion: This summary will not include scientific metrics. If possible, pipelines deviating from a norm will be highlighted. (Use Case: POST7)

8.10 Campaign Dataset Coverage

ID: DMS-BPS-REQ-0055

Specification: It shall be possible to ask the BPS what images have been processed in specific campaigns.

Discussion: Saving provenance and (sub)campaign execution status minimally satisfies this requirement.

References

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

[2] **[LDM-633]**, Kowalik, M., Gower, M., Kooper, R., 2019, *Offline Batch Production Services Use Cases*, LDM-633, URL https://ls.st/LDM-633