

Modernisation of the LHCb continuous integration build system

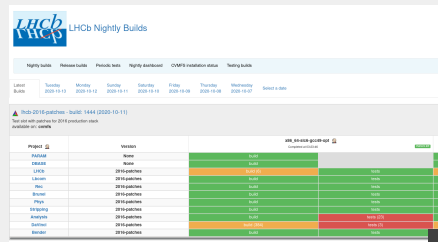
Maciej Szymański on behalf of the LHCb collaboration

CERN

ICHEP 2022 Bologna, 9th of July 2022

Nightly Builds Pipelines

- Critical service for the software development in LHCb
 - ▶ centralised monitoring
 - ▶ validation and testing
- ~ 50 software stacks (aka slots) composed of up to ~ 30 C++ interdependent projects
 - ▶ checkout, build and (unit) test
 - ▶ every night
 - ▶ on demand for most important use cases
 - ▶ several platforms (architecture, OS, compiler, build type)
- Essential to provide **fast turnaround** of produced builds
 - ▶ summaries in the dashboard
 - ▶ artifacts (e.g. binaries deployed to shared file system)



The screenshot shows the LHCb Nightly Builds dashboard. At the top, there's a navigation bar with links: Nightly builds, Release builds, Periodic tests, Nightly dashboard, CVMFS installation status, and Testing builds. Below this is a calendar view for October 2020, with the 11th highlighted. A warning message indicates a test set with patches for the 2014 production stack is available on the example. The main table displays build results for various projects across different versions and platforms.

Project	Version	Build	Status
ROOT	6.08.04	Build 101	Success
CRUISE	None	Build 101	Success
LHCb	2016-patches	Build 101	Success
Lumi	2016-patches	Build 101	Success
Reco	2016-patches	Build 101	Success
Brunei	2016-patches	Build 101	Success
Flays	2016-patches	Build 101	Success
Standalone	2016-patches	Build 101	Success
Analysis	2016-patches	Build 101	Success
Guinea	2016-patches	Build 101	Success
Backend	2016-patches	Build 101	Success

Motivation for modernisation

Old design has reached its scalability limit

- increasing number of *builds on demand* makes the system a continuous integration framework rather than just nightly
- monolithic tasks giving no flexibility
- redundant job executions

Jenkins, our automation server, causes quite frequent issues

- does not prove stable enough with our workload
- used only to schedule tasks on remote machines
- problematic upgrades



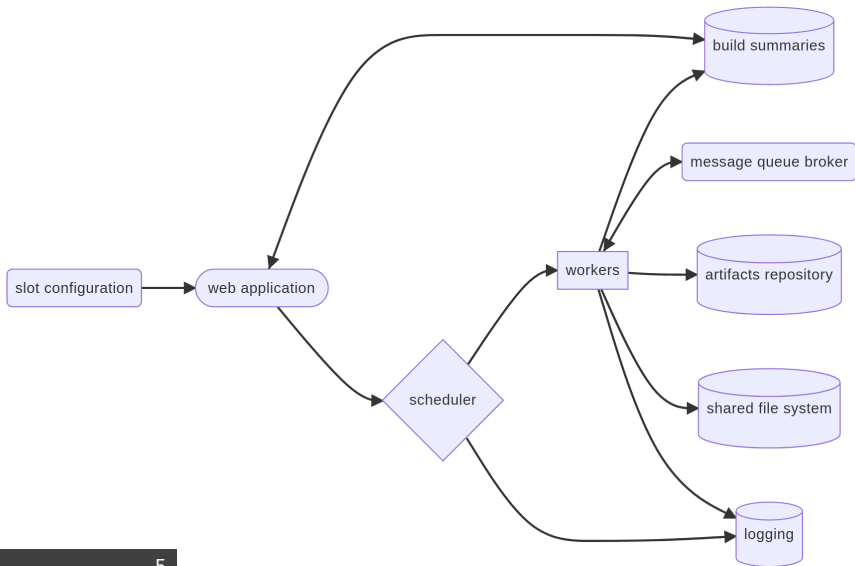
Design of the new continuous integration build system

- Checkout, build and test **split** per project/platform
 - ▶ instead of per slot
 - ▶ parallel jobs increase the overall throughput
- Tasks organised in a **directed acyclic graph**
 - ▶ be faster by doing less!
 - ▶ profit from reusing the **cached artifacts**, whenever possible
 - ▶ e.g. no need for running checkout if sources unchanged
- **Remote execution**
 - ▶ distributing the actual CPU intensive workload to the build farm
- **Family of Python** packages with focused responsibilities
- Aim for simpler and cleaner solution compared with the legacy one
- Easy deployment for development and production environments

Optimisation using deployed artifacts

- In the legacy system, tasks spend significant resources downloading the artifacts from the repository and extracting files from archives
- Recently, LHCb software publication rate to CVMFS has been greatly improved ([EPJ Web Conf. 251 \(2021\) 02034](#))
 - ▶ deployment of the binaries takes typically less than 5 minutes
- **Saving IO** by using artifacts deployed to CVMFS!
 - ▶ trigger the installation of sources and binaries asap
 - ▶ use directly deployed dependencies for subsequent builds and tests

High-level architecture overview

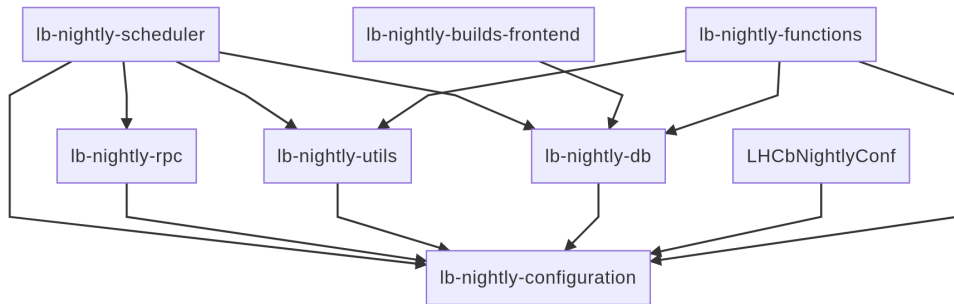


Underlying services

- Flask web application for user interactions
 - CouchDB instance to store the result summaries
 - RPC services to schedule and distribute the workload
 - RabbitMQ instance as a RPC communication protocol
 - MySQL as a RPC tasks backend
 - S3 repository for the artifacts
 - Nexus frontend for S3
 - OpenSearch instance to collect the logs
-
- Deployed in OpenShift, systemd + Puppet, CERN Database On Demand service, CERN infrastructure, or self-hosted
 - Started investigating deployment of a dedicated Kubernetes cluster
 - Development infrastructure based on docker-compose



lb-nightly packages



Releases uploaded to PyPI and conda-forge

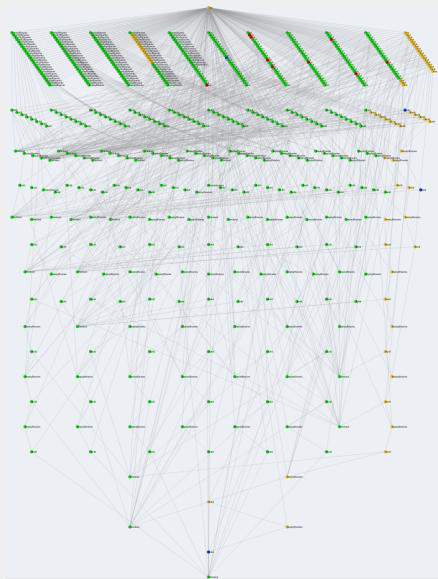
Scheduling the tasks

- Using [Luigi](#), a Python task manager
- Similar to make in the way it handles tasks, dependencies and artifacts
- Takes care of (dynamic) dependency resolution, workflow management, handling failures
- Code in [lb-nightly-scheduler](#)



Luigi visualiser: DAG of lhcb-head slot

- Nodes denote checkout, build, test, deployment tasks for all the projects in a slot
- Edges show dependencies between
 - ▶ types of tasks (e.g. *deploy sources* depends on *checkout*)
 - ▶ projects (e.g. *LHCb* depends on *Gaudi*), **resolved dynamically**



Remote execution

- Based on [Celery](#), a distributed task queue system
- Application delegating tasks to workers
 - ▶ using message queues (RabbitMQ)
 - ▶ responsible for routing the task depending on the architecture
 - ▶ enables introducing job priorities
 - ▶ defining retry policy
 - ▶ setting up the workers (e.g. build worker should not run concurrent tasks)
- Code in [lb-nightly-rpc](#)



Remainder of the family of `lb-nightly` packages

`lb-nightly-functions`

- actual functions used to checkout, build and test projects
 - ▶ provide wrappers to run within Singularity containers
 - ▶ log collection using AsyncIO and UNIX sockets

`lb-nightly-configuration`

- definition of basic abstractions (e.g. `Project`, `Slot`)

`lb-nightly-db`

- functions to communicate with the database

`lb-nightly-utils`

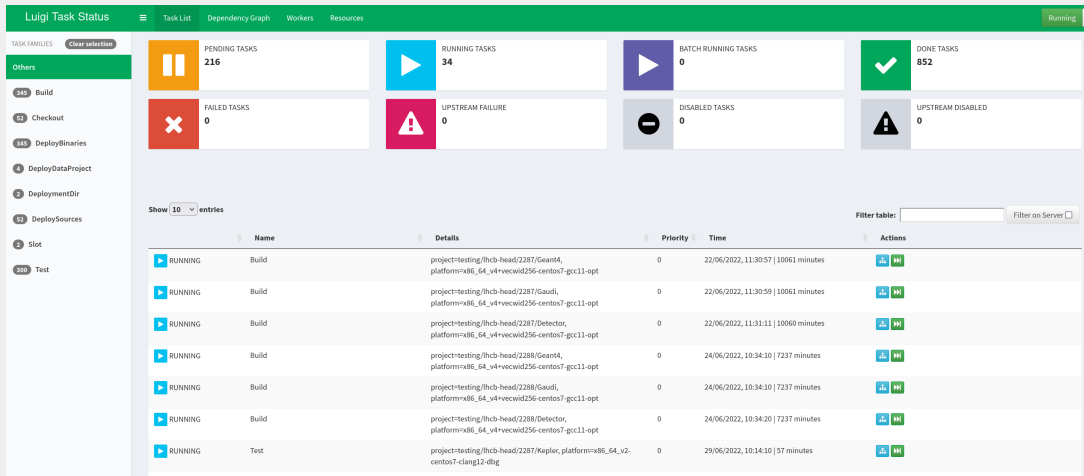
- e.g. artifacts repository class abstracting the location of the artifacts

Environments

- Essential to keep versions of **dependencies under control**
- Separate environments for
 - ▶ workers (lb-nightly-rpc + dependencies)
 - ▶ scheduler (lb-nightly-scheduler + dependencies)
 - ▶ functions (lb-nightly-functions + singularity + cmake + ninja etc.)
- Making use of conda package manager
- Environments defined by hash of the content
- **Automatically deployed** to CVMFS through GitLab CI



Monitoring the scheduler



Monitoring the workers and tasks in Flower

Flower

Dashboard

Tasks

Broker

Docs

Co

Active: 19

Processed: 22021

Failed: 10251

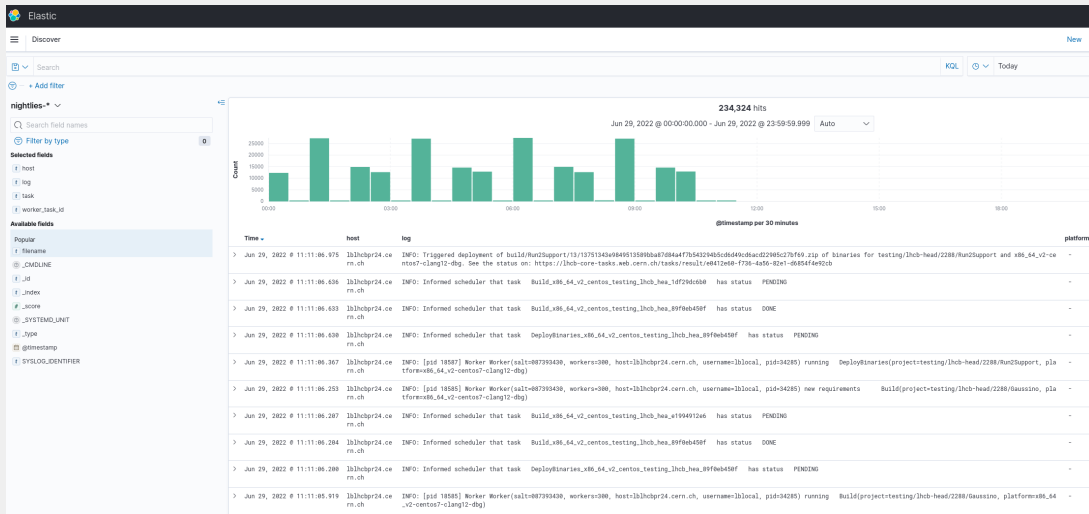
Succeeded: 11447

Retried: 0

Search:

Worker Name	Status	Active	Processed	Failed	Succeeded	Retried	Load Average
celery@lblhcbpr22.cern.ch	Offline	0	0	0	0	0	70.77, 56.22, 61.21
build@lblhcbpr24.cern.ch	Online	0	2430	36	2387	0	9.7, 12.63, 20.78
scheduler@lblhcbpr24.cern.ch	Online	2	52	0	9	0	9.7, 12.63, 20.78
checkout@lblhcbpr22.cern.ch	Online	0	653	128	525	0	29.97, 29.56, 29.16
build@lblhcbpr22.cern.ch	Online	0	2249	23	2220	0	29.97, 29.56, 29.16
checkout@lblhcbpr24.cern.ch	Online	0	653	117	536	0	9.7, 12.63, 20.78
test@lblhcbpr24.cern.ch	Online	8	8019	4970	2899	0	9.7, 12.63, 20.78
test@lblhcbpr22.cern.ch	Online	9	7965	4977	2871	0	29.97, 29.56, 29.16

Monitoring the raw logs in OpenSearch Dashboards


































Nightlies dashboard

lhcb-head/2285 (2022-06-16) [prev](#) [next](#)

head of everything against Gaudi/master and LCG_101


build 78% failing 14 errors tests 79% ✓ 5720 | ✗ 619

Project	Version	x86_64_v2-centos7-gcc11-opt		x86_64_v2-centos7-gcc11-dbg		x86_64_v2-centos7-clang12-opt		x86_64_v2-centos7-clang12-dbg		x86_64_v3-centos7-gcc11-opt+g		
		build	tests	build	tests	build	tests	build	tests	build	tests	
LCG	101											
Gaudi	 master	0 / 0	284 / 0	0 / 0	284 / 0	0 / 0	283 / 1	0 / 0	284 / 0	0 / 0	280 / 3	
Detector	 HEAD	0 / 0	32 / 0	0 / 0	32 / 0	0 / 0	32 / 0	0 / 0	32 / 0	0 / 0	32 / 0	
LHCb	 HEAD	0 / 0	243 / 0	0 / 0	243 / 0	0 / 0	243 / 0	0 / 0	243 / 0	0 / 0	243 / 0	
Online	 HEAD	0 / 0	58 / 1	0 / 0	59 / 0	0 / 0	58 / 1	0 / 0	57 / 2	0 / 0	57 / 2	
Lbcom	 HEAD	0 / 0		0 / 0		0 / 0		0 / 0		0 / 0		
Boole	 HEAD	0 / 0	17 / 0	0 / 0	17 / 0	0 / 0	17 / 0	0 / 0	17 / 0	0 / 0	17 / 0	
Rec	 HEAD	0 / 0	57 / 0	0 / 0	57 / 0	0 / 0	57 / 0	0 / 0	57 / 0	0 / 0	57 / 0	
Allen	 HEAD	0 / 0	4 / 2	0 / 0	4 / 2	0 / 0	4 / 2	0 / 0	4 / 2	0 / 0	4 / 2	
Moore	 HEAD	0 / 0	152 / 14	0 / 0	151 / 13	0 / 0	151 / 13	0 / 0	 error	0 / 0	152 / 14	
Analysis	 HEAD	0 / 0	3 / 0	0 / 0	3 / 0	0 / 0	3 / 0	0 / 0	3 / 0	0 / 0	3 / 0	
DaVinci	 HEAD	0 / 0	47 / 1	0 / 0	48 / 0	0 / 0	48 / 0	0 / 0	48 / 0	0 / 0	48 / 0	
Alignment	 HEAD	0 / 0	14 / 0	0 / 0	14 / 0	0 / 0	13 / 1	0 / 0	4 / 0	0 / 0	14 / 0	
MooreOnline	 HEAD	0 / 0	1 / 1	0 / 0	1 / 1	0 / 0	1 / 1	0 / 0	1 / 1	0 / 0	1 / 1	
Panoptes	 HEAD	0 / 0	5 / 0	0 / 0	5 / 0	0 / 0	5 / 0	0 / 0	5 / 0	0 / 0	5 / 0	
Kepler	 HEAD	2 / 0	2 / 0	2 / 0	2 / 0	2 / 0	2 / 0	2 / 0	2 / 0	2 / 0	2 / 0	
AlignmentOnline	 HEAD	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	0 / 0	
LHCbIntegrationTests	 HEAD	0 / 1		0 / 1		0 / 1		0 / 1		0 / 1		
MooreAnalysis	  HEAD	0 / 0	15 / 6	0 / 0	15 / 6	0 / 0	15 / 6	0 / 0	15 / 6	0 / 0	15 / 6	
Run2Support	  HEAD	0 / 0	6 / 0	0 / 0	6 / 0	0 / 0	6 / 0	0 / 0	6 / 0	0 / 0	6 / 0	
Panoramix	 HEAD	1 / 0		1 / 0		1 / 0		1 / 0		1 / 0		
Geant4	  HEAD	0 / 0		33 / 0		268 / 0		268 / 0		0 / 0		
Gaussino	  HEAD	25 / 0	7 / 0	25 / 0	7 / 0	23 / 5	0 / 7	23 / 5	0 / 7	25 / 0	7 / 0	
Gauss	  HEAD	53 / 0	7 / 65	52 / 0	7 / 65	0 / 0		0 / 0		53 / 0	7 / 65	
DBASE	 None											
PARAM	 None											

Monitoring the build summaries

CERN Accelerating science

DirectorySigned in as: macosmanLogout

Build log of Gaussino for x86_64_v2-centos7-gcc11-opt in [testing/lhcb-head/2285](#)

Nightly buildsRelease buildsPeriodic tests

Report a bugOther links

Links

- checkout report
- raw build report

Summary

6 unique warnings (25 total occurrences)

```
/cvmfs/lhcbdev.cern.ch/nightlies/testing/lhcb-head/2285/Geant4/InstallArea/x86_64_v2-centos7-gcc11-opt/lib/Geant4-10.6.2.5/Geant4PackageCache.cmake:9: warning: Value of &#x27;CLHEP_DIR&#x27; is already set and does not match value set at Geant4 build-time
6 occurrences: configure #1, configure #2, configure #3, configure #4, configure #5, configure #6

/cvmfs/lhcbdev.cern.ch/nightlies/testing/lhcb-head/2285/Geant4/InstallArea/x86_64_v2-centos7-gcc11-opt/lib/Geant4-10.6.2.5/Geant4PackageCache.cmake:9: warning: Value of &#x27;XercesC_INCLUDE_DIR&#x27; is already set and does not match value set at Geant4 build-time
6 occurrences: configure #1, configure #2, configure #3, configure #4, configure #5, configure #6

/cvmfs/lhcbdev.cern.ch/nightlies/testing/lhcb-head/2285/Geant4/InstallArea/x86_64_v2-centos7-gcc11-opt/lib/Geant4-10.6.2.5/Geant4PackageCache.cmake:9: warning: Value of &#x27;XercesC_LIBRARY_RELEASE&#x27; is already set and does not match value set at Geant4 build-time
6 occurrences: configure #1, configure #2, configure #3, configure #4, configure #5, configure #6

/home/lblocal/override_envs/functions/lib/python3.9/site-packages/lb/nightly/functions/data/cmake/GaudiProjectConfig.cmake:2911: warning: failure computing dependencies of QMTest tests
5 occurrences: configure #1, configure #2, configure #3, configure #4, configure #5

make[4]: warning: jobserver unavailable: using -j1. Add "+&#x27; to parent make rule.

/workspace/Gaussino/Sim/GiGaMTCORE/src/truth/TruthStoringTrackAction.cpp:177:11: warning: variable &#x27;type&#x27; set but not used [-Wunused-but-set-variable]
```

Environment

Environment variables

CPU info

Full log

hide/show clean sections

configure (23 warnings)

HepMC3 (1 warnings)

Sim/GiGaMTCORE (1 warnings)

Summary

- Crucial to provide a **robust continuous integration system** for building LHCb software stacks
- Newly designed system is much **more efficient** and cleaner than the legacy one
 - ▶ increased overall throughput and performance
 - ▶ **splitting and parallelising** the tasks
 - ▶ **caching** and reusing the artifacts to save the usage of resources
 - ▶ relevant **monitoring**
 - ▶ **better control** after dropping dependency on Jenkins
 - ▶ usage of **Open Source** tools
- Working towards deployment of the new system in production