

LHCb Stripping Project: Continuing to Fully and Efficiently Utilize Legacy Data

Nate Grieser, on behalf of the Collaboration

University of Cincinnati

CHEP 2024 Krakow

21-10-2024

Overview and Motivation

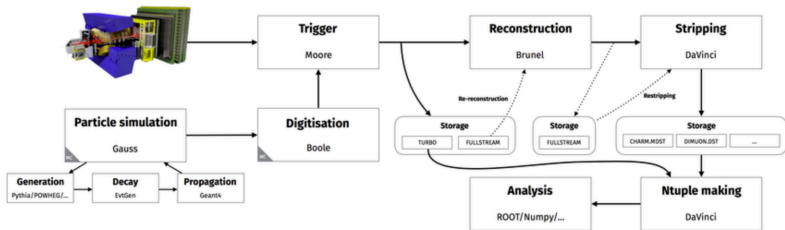
Evolving experiments have evolving dataflows and data models

→ **How do we maintain software to ensure continued data utilization?**

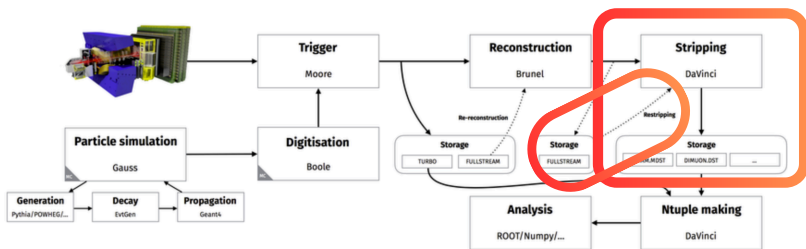
- Simple structure that can be learned and adapted easily
- Regular software testing to follow impact of sporadic changes
- Efficient workflows to adapt to rapidly changing operational situations
- Successful knowledge transfer, adapting new procedures when needed

→ **Provide an easy-to-use, sustainable legacy workflow!**

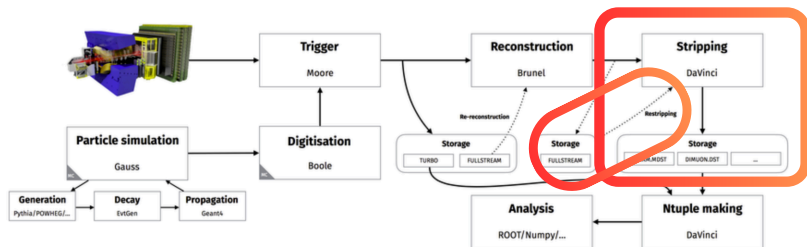
LHCb Legacy DataFlow



LHCb Legacy DataFlow

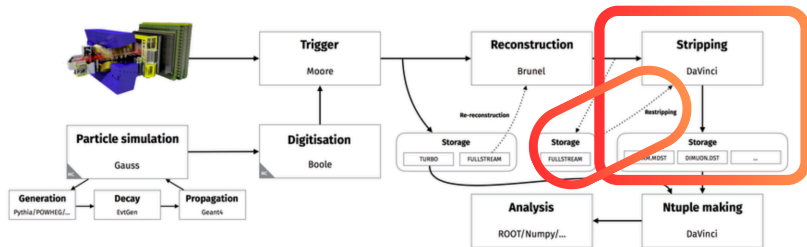


LHCb Legacy DataFlow



- 2-3x reduction in events, with 2x reduction in average event size

LHCb Legacy DataFlow

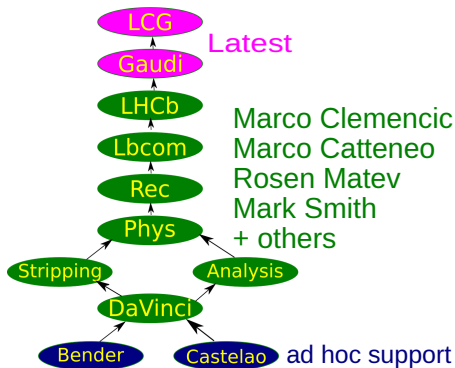


- 2-3x reduction in events, with 2x reduction in average event size
- Consistency of selections between years when possible

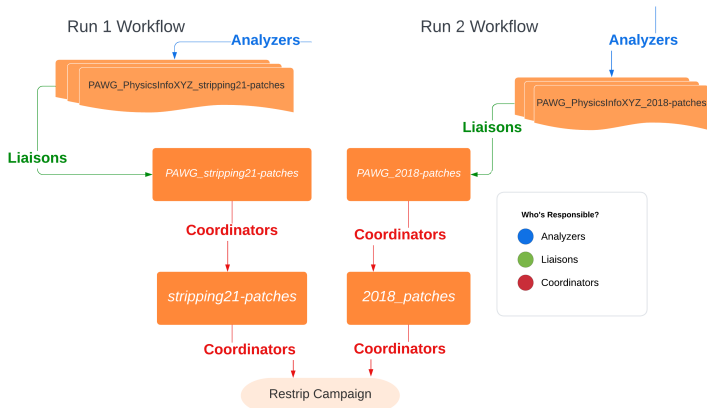
Supporting the Processing – Run 1/2 Legacy Stack

For analyzing Run 1/2 data there is a legacy stack maintained

- Everything from LHCb upwards is updated
 - Builds on latest LCG & Gaudi
 - Maintenance for obsolete projects are dropped
- Collaboration with core computing team to ensure stability and performance
- **New tools can be added to process the legacy data after the productions!**
 - Only release when necessary

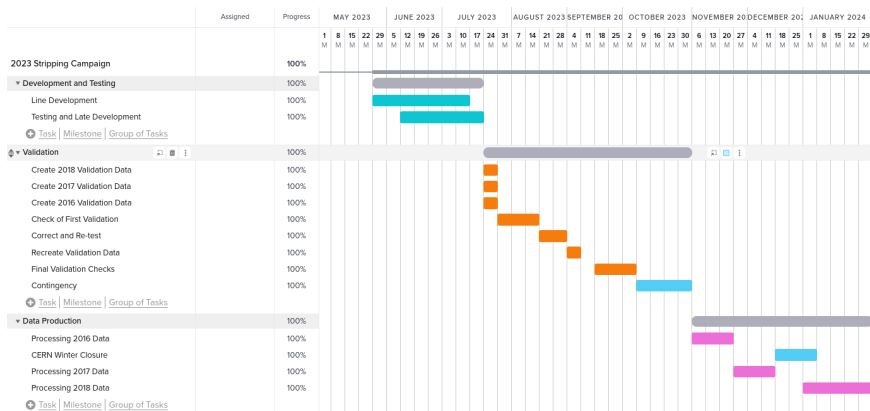


Getting Started – Fresh Faces, Fresh Ideas



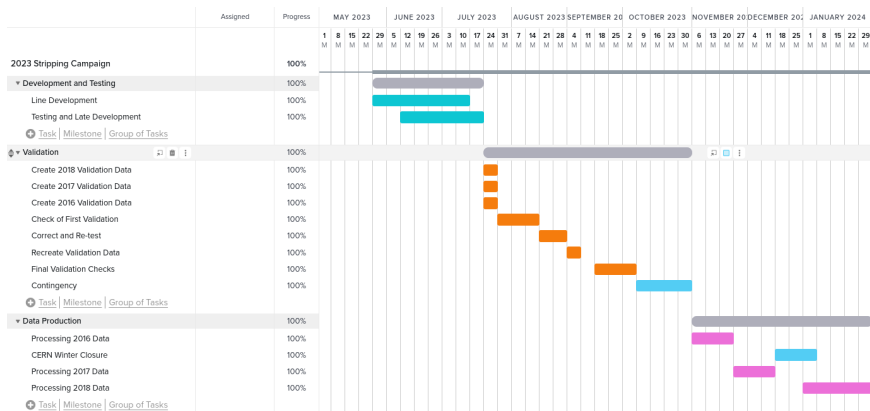
- Liaisons from Physics WG to support the campaign
- In depth training to assign roles, technical crash course, and update workflows

Advertise the Plan... Then Deliver!



- Limited training and development time → Need to be efficient!

Advertise the Plan... Then Deliver!



- Limited training and development time → Need to be efficient!
- Tight production windows → No contingencies, have to get it right first!

GitLab Milestones – Modern Bookkeeping

The screenshot shows the GitLab Merge requests interface. At the top, there are filters for 'Issues' (0), 'Merge requests' (125), 'Participants' (0), and 'Labels' (0). Below these are four main categories:

- Work in progress (open and unassigned):** 0 items.
- Waiting for merge (open and assigned):** 0 items.
- Rejected (closed):** 18 items. Examples include:
 - Draft: New LLP->eta pi pi line (ID: 11847, assigned to GEE)
 - adding D0->K3pi mode for DICARM (ID: 11755, assigned to B&G)
 - Draft: add D0->K3pi mode for DICARM
- Merged:** 107 items. Examples include:
 - QEE LineConfigDictionaries and Init file (ID: 11868, assigned to GEE)
 - Reinstate pi0/eta -> e+ e- gamma lines, preserving brems photons (ID: 11857, assigned to GEE)

- Analysts required to add the bookkeeping, alleviating overhead of coordinators
- Liaisons and coordinators can follow things simply
- Neatly available statistics at end of campaigns to feedback to management

The screenshot shows the 'Assign milestone' dialog box for user Ifan Williams. It includes an 'Edit' link and a search bar. The list of milestones is as follows:

- ✓ No milestone
- 2023 Re-Strip Campaign -- Development and Testing
- 2023 Re-Strip Campaign -- Validation
- 2023 Re-Strip Campaign
- HLT2 Tracking speedup (expired)

The 'Milestone' dropdown at the bottom is currently set to 'Milestone'.

Development – Concise and Complete MRs

- Source branches use naming requirements to run specific WG tests
 - Efficient testing!
- All updated/new lines by name and their test output information
 - Rate information**
 - Timing information**
- WG labels and milestones used to track development process

QEE dmcontrollines 2018 patches

Code ▾

Merged Saul Lopez Solino requested to merge QEE_dmcontrollines_2018-pa... into QEE_2018-patches 1 year ago

Overview 17 Commits 3 Pipelines 7 Changes 1

All threads resolved! ⌵

Hi! Me, @jcidvida and @Cvazquez are adding some new control lines and changing the cuts for some of the already existing LambdaDecaysDM lines. The report would be this one.

StrippingReport	INFO Event 60000, Good event 60000
=Decision name	=Rate, % =Accepted = *Mult =ms/evt =
StrippingGlobal	0.4017 241 10.775
StrippingSequenceStreamBhadronCompleteEvent	0.4017 241 10.751
! StrippingLambdaDecaysDMLambdaToPiPiLine	0.0333 20 1.250 6.216
! StrippingLambdaDecaysDMLambdaToPiPiLine	0.0567 34 2.059 0.697
! StrippingLambdaDecaysDMLambdaToPiLine	0.0367 22 1.045 0.170
! StrippingLambdaDecaysDMLambdaToKLine	0.0383 23 1.217 0.249
! StrippingLambdaDecaysDMLambdaToKPiLine	0.0450 27 1.000 0.283
! StrippingLambdaDecaysDMLambda1S2BControlline	0.0200 12 1.250 0.056
! StrippingLambdaDecaysDMLambda2S95Line	0.0233 14 1.286 0.275
! StrippingLambdaDecaysDMLambda2S95Controlline	0.0350 21 1.333 0.163
! StrippingLambdaDecaysDMLambdaToDKLine	0.0683 42 1.146 0.114
! StrippingLambdaDecaysDMLambdaToDKControlline	0.0450 27 1.111 0.138
! StrippingLambdaDecaysDMLambda1S2BLine_TIMING	0.0333 20 1.250 0.235
! StrippingLambdaDecaysDMLambdaToPiPiLine_TIMING	0.0567 34 2.059 0.183
! StrippingLambdaDecaysDMLambdaToPiLine_TIMING	0.0367 22 1.045 0.044
! StrippingLambdaDecaysDMLambdaToKLine_TIMING	0.0383 23 1.217 0.043
! StrippingLambdaDecaysDMLambdaToKPiLine_TIMING	0.0450 27 1.000 0.248
! StrippingLambdaDecaysDMLambda1S2BControlline_TIMING	0.0250 15 1.333 0.050
! StrippingLambdaDecaysDMLambda2S95Line_TIMING	0.0233 14 1.286 0.048
! StrippingLambdaDecaysDMLambda2S95Controlline_TIMING	0.0383 23 1.565 0.100
! StrippingLambdaDecaysDMLambdaToPiControlline_TIMING	0.0550 33 1.212 0.062
! StrippingLambdaDecaysDMLambdaToDKControlline_TIMING	0.0517 31 1.323 0.085

Assignee

 Saul Lopez Solino

Reviewer

 Xiaolin Wang

Labels

QEE

Milestone

2023 Re-Strip Campaign --
Development and Testing (expired)

Time tracking

No estimate or time spent

6 Participants



Handshakes With Computing Team

Vital to communicate regularly between production and analysis teams

2023 Re-Stripping Campaign -- 2016 (S28r2p2)

Stripping production request -- WIP

Summary

Currently prepared is a YAML for 2016 MD.

For testing locally, we need staged some samples:

Validation runs: 184604, 184642 (MU), 175835 (MD)

YAML(s)

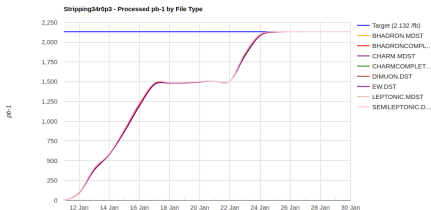
```

- author: ngrieser
  name: Stripping28r2p2 # E.g. "Passthrough / Exclusive Sprucing (2023 MagDown)"
  comment: 2023 incremental restripping of 2016 data # E.g. "Test for first time"
  type: Stripping

  wg: DPA
  inform: # Feel free to add any other relevant person
  - ngrieser
  - avenkate
  - fredri
  priority: 1a

  input_dataset:
    # 94000000 / 90000000 for Exclusive / Passthrough Sprucing
    event_type: 90000000
    # Is there a check on consistency e.g. of magnet polarity and name?
    conditions_description: Beam6500GeV-VeloClosed-MagDown
    conditions_dict:
      configName: LHCB
      configVersion: Collision16 # E.g. "Collision23"
      inFileType: RAW
      inProPass: Real Data
      inDataQualityFlag: OK # To be changed as soon as DQ is fully in place

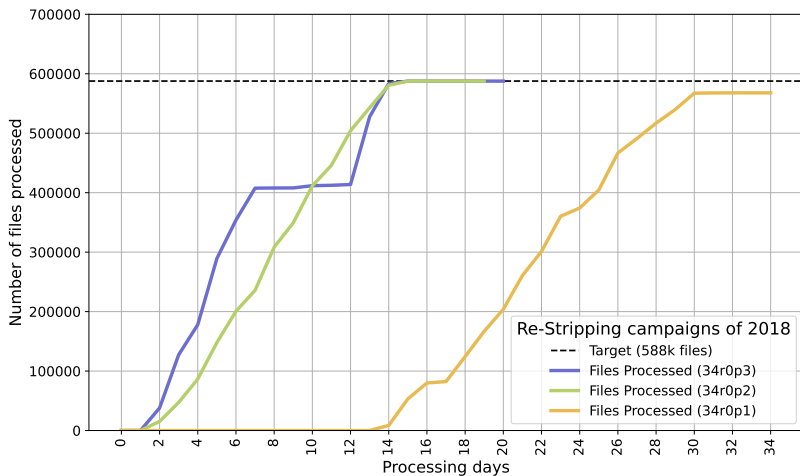
```



↑ Live feedback of sample processing allows to catch any serious oversights in development

← Production requests are steered using GitLab issues and yaml files that production team can apply directly

Improving the Production Times

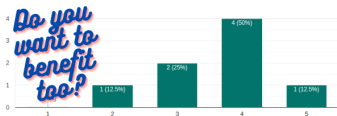


Learning From Each Other

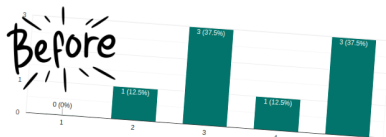
Legacy productions will not happen regularly: **Lots of turnover to handle**
 → New minds can be a good thing, too!

- Training and procedures progress over years → need to make sure it's actually improving!
- Impact of training, comfortableness with the role, and impact of the role on their physics training were all considered!

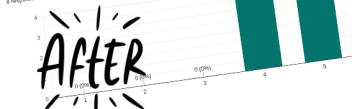
How do you feel your preparedness to further your physics career benefitted from performing this liaison role?
 8 responses



How comfortable did you feel with your responsibilities as a liaison prior to the liaison training this summer?
 8 responses



How qualified did you feel with your responsibilities as a liaison AFTER the liaison training this summer?
 8 responses



Conclusions

→ LHCb continues to have a thriving legacy physics program

Conclusions

→ LHCb continues to have a thriving legacy physics program

- Software and development is maintained to allow for large data reprocessing campaigns

Conclusions

→ LHCb continues to have a thriving legacy physics program

- Software and development is maintained to allow for large data reprocessing campaigns
- Continued collaboration with computing and operations makes success!

Conclusions

→ LHCb continues to have a thriving legacy physics program

- Software and development is maintained to allow for large data reprocessing campaigns
- Continued collaboration with computing and operations makes success!
- Learn and grow with modern workflows

Conclusions

→ LHCb continues to have a thriving legacy physics program

- Software and development is maintained to allow for large data reprocessing campaigns
- Continued collaboration with computing and operations makes success!
- Learn and grow with modern workflows



BACKUP

Using PyConf to Make Selections

```
def makeB2Charged2Body( name,
                       trChi2, trGhostProb, minPT, minIPChi2,
                       maxPT, maxIPChi2, combMassLow, combMassHigh, doca,
                       bPT, bIPChi2, bTAU, massLow, massHigh ) :
    _daughters_cuts = "(TRGHOSTPROB < %(trGhostProb)s) & (TRCHI200F < %(trChi2)s) & (PT > %(minPT)s * MeV) & ( MIPCHI2DV(PRDIMARY) > %(minIPChi2)s )" % locals()
    _combination_cuts = "(AMAXCHILD(MAXTREE('pi'+='ABSD,PT)) > %(maxPT)s) & ( AMAXCHILD(MAXTREE('pi'+='ABSD,MIPCHI2DV(PRDIMARY))) > %(maxIPChi2)s) & (AMAXDOCA(''') < %(doca)s) & (M > %(combMassLow)s * MeV) & (M < %(combMassHigh)s * MeV)" % locals()
    _mother_cuts = "(PT > %(bPT)s * MeV) & (M > %(massLow)s * MeV) & (M < %(massHigh)s * MeV) & ( BPVPOCHI2() < %(bIPChi2)s) & (BPVLTIME() > %(bTAU)s )" % locals()
    CombineB2Charged2Body = CombineParticles( DecayDescriptor = 'BB -> pi+ pi-',
                                             DaughtersCuts = ( "pi" : _daughters_cuts,
                                                             CombinationCut = _combination_cuts,
                                                             MotherCut = _mother_cuts )
    return Selection( name,
                    Algorithm = CombineB2Charged2Body,
                    RequiredSelections = [ StoNePIDsPions ] )
```

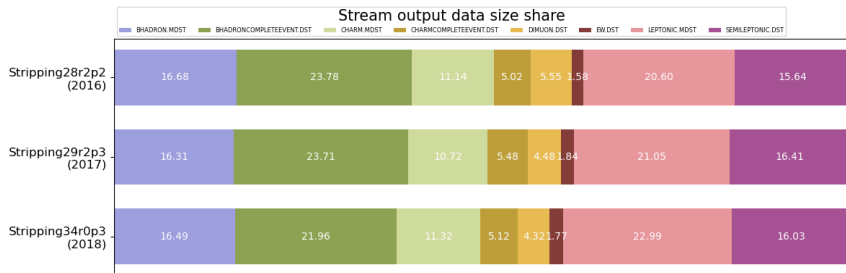
↑ Define decay selections with
builders in Python

Steer selections with config libraries
→ Harmonize code while specific
analysis selections available →

```
default_config = {
    'NAME' : 'B2HHBDT',
    'WGS' : ['BnoC'],
    'BUILDERTYPE' : 'B2HHBDTLines',
    'CONFIG' : { 'PrescaleB2HHBDT' : 1.,
                'TrChi2' : 4,
                'TrGhostProb' : 3,
                'PionPT' : 1000,
                'SumPT' : 4500,
                'DOCACHI2' : 9,
                'BIPCHI2' : 9,
                'BDIRA' : 0.99,
                'BPT' : 0,
                'BMassWinLow' : 4700,
                'BMassWinHigh' : 6200,
                'BMassLow' : 4800,
                'BMassHigh' : 6200,
                'PionIPCHI2' : 16,
                'BFDCHI2' : 100,
                'BDTCut' : -1,
                'BDTWeightsFile' : "$THVAWEIGHTSROOT/data/B2HH_BDT_v1r5.xml"
            },
    'STREAMS' : ['BhadronCompleteEvent']
```

Giving the User a Manageable Dataset

→ Strength of Stripping project on display: Huge compression for analysts



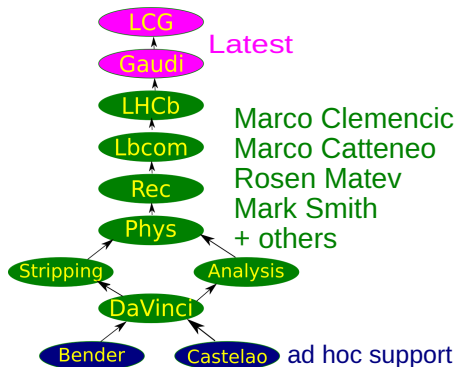
config version	processing pass	reco	magnet polarity	# events (in)	Total size (in)	Avg kB/event (in)	# events (out)	Total size (out)	Avg kB/event (out)	# events (RAW)	Total size (RAW)	Avg kB/event (RAW)	# streams (out)	# evts out/in	Size reduction
Collision16	Stripping28r2p2	Reco16	Down	21'942'558'323	1088.0	49.6	7'759'577'117	250.9	32.3	21'948'653'743	1395.0	63.6	8	0.4	4.3
			Up	20'170'963'459	970.7	48.1	6'899'525'674	225.8	32.7	20'195'067'584	1262.0	62.5	8	0.3	4.3
Collision17	Stripping28r2p3	Reco17	Down	17'662'321'998	951.5	53.9	7'323'509'172	228.2	31.2	17'662'323'449	1161.0	65.7	8	0.4	4.2
			Up	16'870'981'928	947.5	56.2	6'943'947'373	216.3	31.2	16'871'076'018	1115.0	66.1	8	0.4	4.4
Collision18	Stripping34r0p3	Reco18	Down	19'236'992'943	1073.0	55.8	8'745'097'218	272.9	31.2	19'237'119'448	1331.0	69.2	8	0.5	3.9
			Up	20'596'358'630	1139.0	55.3	9'188'827'963	284.8	31.0	20'596'464'277	1423.0	69.1	8	0.4	4.0

- Consistency of streams between different years
- 2-3x reduction in events in vs. out, with 2x reduction in average event size

Supporting the Processing – Run 1/2 DaVinci Stack

For analysing Run 1/2 data there is a legacy stack maintained

- Everything from LHCb upwards is updated
 - Builds on latest LCG & Gaudi
 - Maintenance for obsolete projects are dropped
- Latest LCG, ROOT etc as well as recent Python & C++ & compilers/platforms
 - Collaboration with core computing team to ensure stability and performance
- **New tools can be added to process the legacy data after the productions!**
 - Users can open MR towards the legacy branches, and steer the bookkeeping with GitLab labels
 - Full software stack releases occur as needed to ease burden of maintainers



Getting Started – Fresh Faces, Fresh Ideas

FRIDAY, 2 JUNE

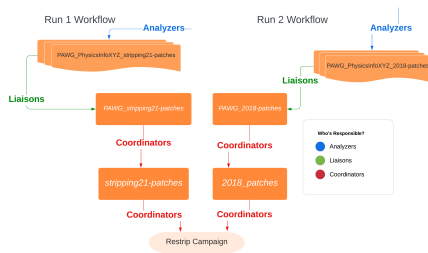
09:30 – 16:00 Hackathon 🕒 6h 30m

10:00 – 12:00 Stripping liaisons training

Conveners: Aravindhan Venkateswaran (EPFL - Ecole Polytechnique Federale Lausanne (CH)), Federico Leo Redi (CERN), Dr Nathan Grieser (University of Cincinnati (US))

📎 GMT20230602-080... 📎 How to write a Strip... 📎 StrippingLiaisons_G... 📎 StrippingLiaisons_O...

- Liaisons offered from Physics WG to support the campaigns
 - New liaisons providing support → Lots to learn!
- Assign roles and technical crash-course
- In-depth training on GitLab use for the campaign → Apply new workflows!



- Workflow applied similarly to run 3 project workflows
- Continuous integration tests used to allow for a rolling testing of developments

Improving the Production Times

