



The FIFE project: supporting non-LHC Experiments at Fermilab

Ken Herner HOW 2019 18 March 2019

Fabric for Frontier Experiments (FIFE)

- Standardize computing framework for non-LHC **FNAL** experiments
- Offer a robust, shared, and modular set of tools for experiments, including:
 - Job submission and monitoring
 - Workflow management software
 - Data management and transfer tools
 - **Continuous Integration service**
- Work closely with experiment contacts during all phases: from experiment workflow development and testing to production campaigns. Standing meetings; regular workshops and tutorials
 - Provide an on-boarding process for an experiment's production team









NOvA



18 Mar 2019

FIFE Project Benefits



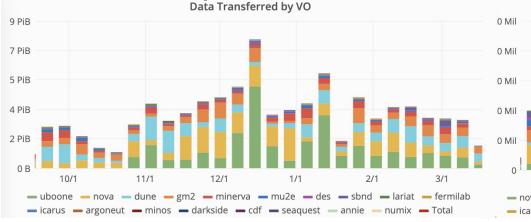
- Can leverage each other's expertise and lessons learned
- Allow to identify and address common problems and **provide common solutions**
 - Greatly simplifies life for those on multiple experiments (same toolset)
- Shift burden from experiments to gain access to new resources
 - OSG
 - European Grid
 - HPC
- Simplify transitions from one common service to another in transparent ways:
 - From FermiGrid to HEP Cloud
 - From GUMS & VOMS-Admin to a new authorization registry service (FERRY)
- FIFE Toolset includes
 - Job submission (jobsub), Workflow management (POMS)
 - Data handling and distribution
 - Continuous integration
 - Integration with new technologies and resources: GPUs and HPC
- Many experiments use ART, module software framework forked from CMSSW
- Also of note: LArSoft, common framework for LAr TPC experiments

FIFE Vision

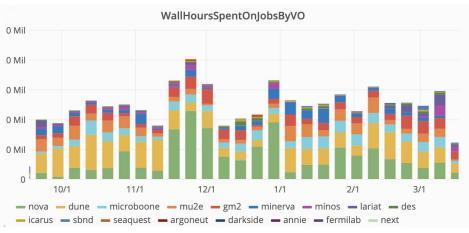
- Mature set of tools used by more that 600 users from different experiments and projects
- All tools are still "alive": new features/improvements are being implemented.
- Layered architecture: could easily replace an obsolete tool with a better one while keeping UI the same
- Monitoring that encompass all the aspects: from individual file transfer to campaign execution
- Well defined on-boarding process for all new experiment.
- Unified users' support

Supported Experiment Data and Job Volumes

- 35 experiments; more than 600 unique users
- 29 sites in 9 countries
- About 160K jobs per day; 170M hours per year
- Combined numbers approaching scale of LHC (6-7x smaller wrt ATLAS+CMS)



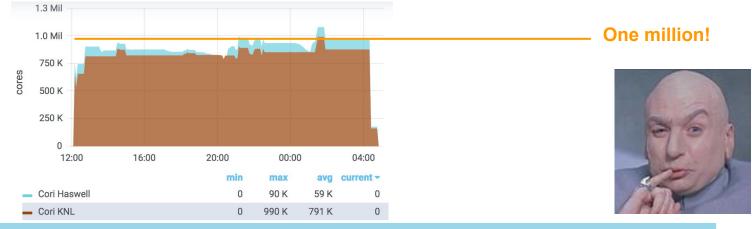
Total weekly data transferred by experiment, last 6 months



Total weekly wall time by experiment, last 6 months

HEPCloud Demonstration Success Story for long-baseline: NOvA @ NERSC

- NOvA experiment uses Feldman-Cousins procedure as part of final analysis.
- Using standard HTC resources, analysis of full dataset would have taken several weeks; impossible to complete in time without significantly expanding resources.
- NOvA was able to use 2M hardware threads simultaneously at NERSC via HEPCloud, achieving an overall speedup factor of about 150.
- Results presented at Neutrino 2018.



Short-Baseline Neutrino Program

Important precursor to DUNE

Consists of **MicroBooNE** (running), **ICARUS** (commissioning this fall), **SBND** (2020); all LArTPC

Similar near-far detector concepts as NOvA/DUNE

ICARUS very interested in adopting Rucio ASAP

Common challenges:

Memory footprint

Learning how to effectively use multiple cores =>

Effectively using HPC resources

Storage/data model (all are multi-PB experiments)



Muon Experiments

g-2: Probe hints of anomalous magnetic moment of the muon

Currently taking data

Mu2e: Look for muon -> electron decay (lepton flavor violation)

Design and construction continuing One of the earlier experiments to get onto OSG (50 M+ hours in 2015-2016)

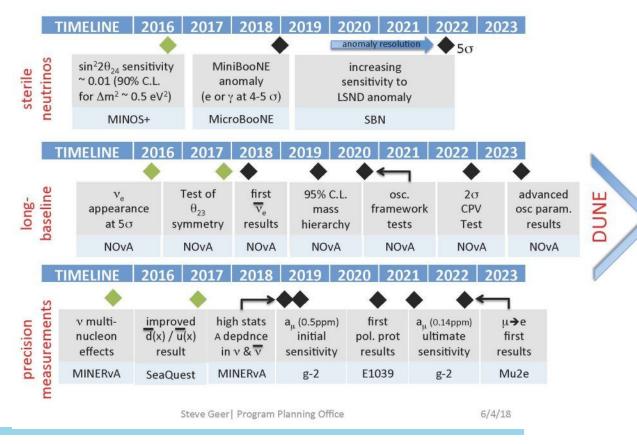


Experiment schedules

Fermilab Accelerator Hosted Science Plan

Significant overlap between running experiments in 2019-21

Additional challenges include DUNE Near Detector design/construction



Tools to ease the crunch

Auxiliary File Delivery via OSG's StashCache



Example of the problem: several neutrino experiments use O(100 MB)"flux files" for simulating neutrinos in rock, etc.

- Several files are used per simulation job
- Total input is few GBs
- Not well-suited to CVMFS for several reasons (each job uses random subset; can thrash worker node cache)
- Transferring all the time sub-optimal (plus no caching at all that way)

StashCache service developed and widely used by OSG to solve just such a problem:

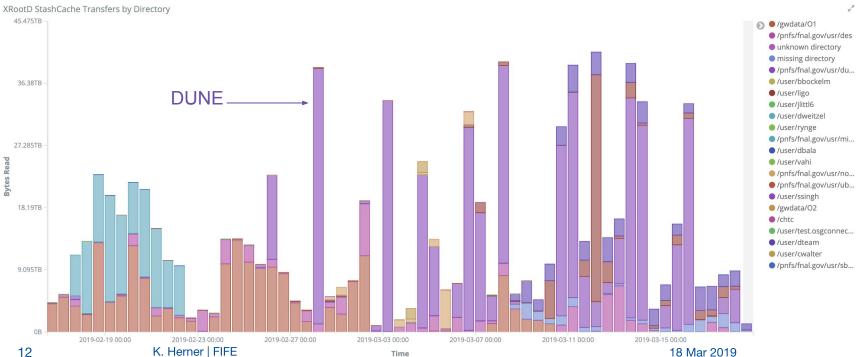
FNAL experiments using the service so far: DES, DUNE, Microboone, Minerva, Mu2e, Nova, SBND (all are Stash + CVMFS, access via /cvmfs/expt.osgstorage.org)

FIFE assists experiments with setting up and populating area.

One issue we've seen: *inconsistent ways of configuring such repositories between US and EGI sites.* Makes it take longer to get this working. Closer communication as far as Stash and CVMFS configs would help here

StashCache Usage

DUNE: largest StashCache user in past 30 days 3 of top 5 are FNAL expts



Time

XRootD StashCache By Logical Directory

Directory ≑	Bytes ≑
/pnfs/fnal.gov/usr/dune	390.114TB
/gwdata/O1	180.212TB
/pnfs/fnal.gov/usr/minerva	99.352TB
/pnfs/fnal.gov/usr/des	63.204TB
/user/ligo	58.211TB

Workflow Management with POMS

- Started working on in-house Production Operation Management Service (POMS) suitable for production workflows for FIFE experiments (now opening to analyzers as well- same interface!)
- Tracks what processing needs to be done ("Campaigns") and job submissions from a campaign
- Make automatic job submissions and set automatic recovery options
 - Allows dependent workflow stages
- Interface with SAM to track files processed by submissions and Campaigns
- Provides "Triage" interface for examining individual jobs/logs and debugging failures.

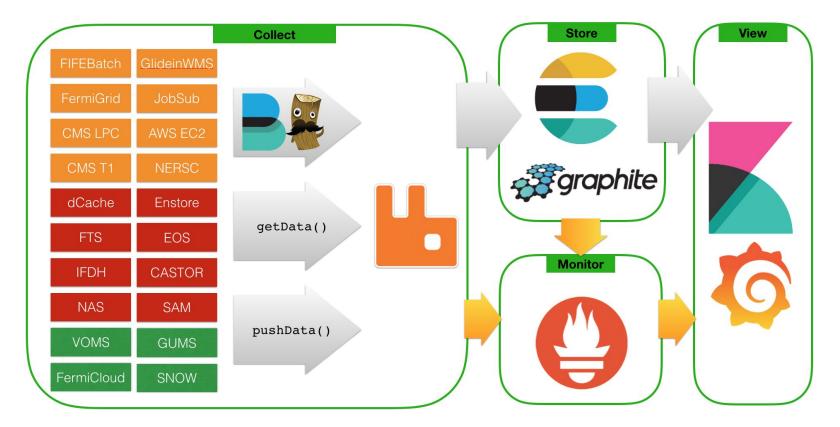
Workflow Management with POMS (2)

					🛔 Kenneth Herner		
Campaign Search Q	Operations Management Campaign Editor for			dune		\$	
External Links	System	mcc11_DP_RITM	0761645 🔞		production		\$
Dune Logbook							
OMS ServiceNow Page			campaign_stage				
wntime Calendar			name	mcc11_DP_RITM07616	45_reco		
periment			vo_role	Production			
mbership (POMS)	Campaign mcc11_DP_RITM0761645 Save		software_version dataset_or_split_data	v08_03_00 POMS_DATASET			
			cs_split_type	None		Edit	
npaign Data	Add Stage 🖉 Add Dependency 🛞 Dele	ete selected	completion_type	complete (default)	\$		
paigns paign Stages			completion_pct param overrides	100	/dunepro/poms_MCC11/DF	Edit	
ple Campaigns	GUI editor for setting up d	ependent	test_param_overrides		duriepro/poins_MCC11/DF	Edit	
		opondon	login_setup	mcc11_gen_g4_launch			
nfigure Work mpose Login/Setup	stages		job_type Reset	mcc11_gen_g4 (default	:) \$ OK	Canaal	
npose Job Type		mcc11_DP_RITM0761645	Reset		UK	Cancer	
mpose Campaign Stages							
				C			· · · · · ·
onfig File Templates	mcc11_DP_RITM076	61645_gen_g4	645_detsim	mcc11_DP_RITM076			
	mcc11_DP_RITM076	61645_gen_g4	645_detsim	mcc11_DP_RITM076		ustor	nizabl
s	mcc11_DP_RITM076	61645_gen_g4	645_detsim	mcc11_DP_RITM076		ustor	nizabl
)s (Last Daγ)	(mcc11_DP_RITM076	61645_gen_g4	645_detsim	mcc11_DP_RITM076		ustor	nizabl
s Last Day) ed By Exit Code and Campaign Stage ne (Last Day)	(mcc11_DP_RITM076	a1645_gen_g4	345_detsim	mcc11_DP_RITM076		ustor	nizabl
hfg File Templates S (Last Day) ed. By Exit Code and Campaign Stage ne (Last Day) ed. By Exit Code and Campaign Stage ne (Last Week) ed. By Exit Code and Node Name ti Day)	mcc11_DP_RITM076	61645_gen_g4	345_detsim	mcc11_DP_RITM076		ustor	nizabl

Workflow Management with POMS (3)

A POMS				*	Kenneth Herner	
Campaign Search Q	Operation Operations Management	une	\$			
External Links	System	mcc11_DP_RITM	10761645 🔞	р	roduction	\$
Dune Logbook			-			7
POMS ServiceNow Page			campaign stage			
Downtime Calendar			Param Editor			
Experiment			Key 😧	Space 🕜	Value 😧	
Membership (POMS)	Campaign mcc11_DP_RITM0761645 Sa	ive	-c		/dune/app/home/dune	
Communitient Darba			-Oglobal.utilquals=		e17: prof	
Campaign Data	Add Stage 🥢 Add Dependency 👔	3) Delete selected	-Oglobal.quals=		e17: prof	
Campaigns Campaign Stages			-Oglobal.utilversion=		v08_02_00	
Sample Campaigns			-Oglobal.version=		%(version)s	
Sample Sampagns			-Oglobal.basename=		%(dataset)s	
Configure Work			-Oglobal.recofclfile=		rawhitfinding_dune10	
Compose Login/Setup		mcc11 DP RITM0761645	-Oglobal.nevents=		100	
Compose Job Type			-Oenv_pass.SAM_STA		dune	
Compose Campaign Stages			-Osubmit.n_files_per_j		1	
Config File Templates	mcc11_DP_RI	ITM0761645_gen_g4 mcc11_DP_RITM076	-Osubmit.expected-life		24h	
Jobs			-Osubmit.memory=		3000MB	
All (Last Day)			Oglobal.tarfile=		dunepro_mcc11_2019	
Failed_By Exit Code and Campaign Stage Name (Last Day)	Workflow-specific	overrides possible	-Oglobal.tardir=		/pnfs/dune/resilient/u:	
Name (Last Day) Failed By Exit Code and Campaign Stage	(can reuse basic s	stage template for	-Osubmit.dataset=		%(dataset)s	
Name (Last Week)		•	-Osubmit.email-to=		mgreenwo@fnal.gov	
Failed_By Exit Code and Node Name (Last Day)	Campaign Login/Setups and Jobtypes	S)	-Ojob_output.dest=		/pnfs/dune/scratch/du	
	Campaign Login/Setups and Jobtypes		-Oglobal.logdir=		/pnfs/dune/scratch/du	0044
		mcc11 gen g4 launch	Cancel Accept			

Monitoring (Landscape Program)

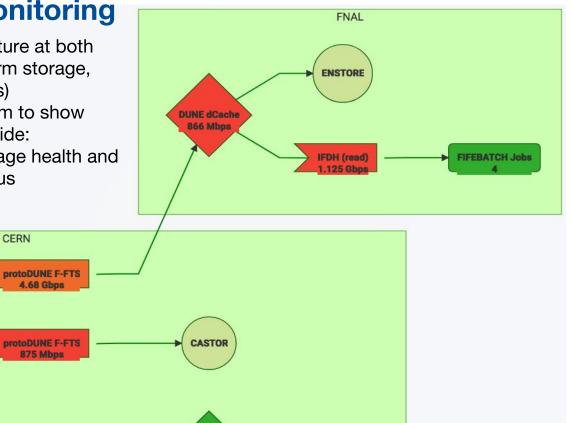


protoDUNE Cross Lab Monitoring

- protoDUNE has computing infrastructure at both CERN and FNAL (buffer disk, long-term storage, reconstruction and analysis resources)
- Developed a unified monitoring system to show data from both laboratories side-by-side:
 - Covers data transfer rates, storage health and usage, job submission and status

Public EOS (write)

16.7 Gbps



LXBATCH Jobs

18 Mar 2019

Public EOS (read)

30.3 Gbps

protoDUNE F-FTS

948 Mbp

DAQ

FTS-Light

0.07 wps

Short and Intermediate Term plans

- FABRIC FOR FRONTIER EXPERIMENTS
- **Containers** (Docker, Singularity, etc.) are becoming more important in increasingly heterogeneous environments (including GPU machines). Help shepherd users through this process and create some common containers for them
- Seamlessly integrate dedicated, opportunistic, HPC, and commercial computing resources
- Lower barriers to accessing computing elements around the world in multiple architectures
 - Help to connect experimenters and computing professionals to drive experiment SW to increased multithreading and smaller memory per core footprints
 - Federated identity management (reduced access barriers for international partners)
- Integrate new technologies (e.g. Rucio)

Summary



FIFE project provides valuable service for FNAL non-LHC experiments. Makes life much easier for those on multiple experiments

Plans include increasing containerization and transition to other

Following existing approaches saves time, effort, enables easy international cooperation

Eagerly seeking common approaches as widely as possible in future projects (Rucio, HPC, IRIS-HEP, etc.)

BACKUP



FABRIC FOR FRONTIER EXPERIMENTS

Efficient User Code Distribution

- Until January, 2018 analysers were able to access their Bluearc code dev area
 - Was NFS mounted on all FermiGrid nodes.
- After Bluearc has been unmounted analyzers have started to use dCache scratch area to upload tarballs and stage them from dCache to worker nodes. It turned out to be a bad idea:
 - too much hammering on single pools when many jobs of the same type start
 - can completely block access to files owned by other users on a given pool
- Interim solution: stage tarballs in "resilient" dCache area (files replicated 20x on different pools):
 - Advantages: keeps the system running, doesn't block users, more efficient.
 - Disadvantages: tarballs use 20x space; need careful cleanup policies
- Longer-term solution: looking into automated distribution via CVMFS.
 - User tarball created, hashed, stored on job submission server at submit time.
 Avoids repeats copying if contents unchanged

Off-line Production Shifter Dashboards

Production Shifter -					111 to 12	🖺 🏟 < Q > 🕐 Last 24 hours 📿
vo ubcone +						
r	Completed Production Job	Status	Active Production J	ob Status		FTS Status
		total percentage — Cancelled 50 1% — Fall 436 11% — Held 50 1%	istie jobs - ranning jobs miss istie time -	69		ending files - 0 alting on tape - 219 r rate - 18.73 MB/s
		🕳 Success 3.57 K 87%	Production Batch Jold	1.45%	Persiste	torage Usage ent dCache - 77.4% IuaeArc - 97%
			Transfer Star 90th percentile - 1	.54 min.		
			Active/Recent POMS			
	Campaign Name		Active/Recent POMS	Batch Jobs Completed	Average CPU Eff	Average Mern Eff
604	MCC8_BNB_Cosmic_T		Active/Recent POMS	Batch Jobe Completed 137	89%	413
604 598	MCC8_BNB_Cosmic_T dqmCalibrationTrees_e	ext_v7	Active/Recent POMS	Batch Jobs Completed 137 1 K	89% 58%	41% 58%
604 598 597	MCC8_BNB_Cosmic_T dqmCalibrationTrees_t dqmCalibrationTrees_t	ext_v7	Active/Recent POMS	Ratch Jobe Completed 137 1 K 1 K	89% 58% 55%	41% 58% 57%
1598 1597 1396	MCC8_BNB_Cosmic_T dqmCalibrationTrees_t dqmCalibrationTrees_t crt_swizzle4_v2	ext_v7	Active/Recent POMS	Batch Jobs Completed 137 1 K 1 K 8	99% 58% 55% 70%	41% 68% 57% 50%
ID - 1604 1598 1597 1396 1395 1394	MCC8_BNB_Cosmic_T dqmCalibrationTrees_t dqmCalibrationTrees_t	ext_v7	Active/Recent POMS	Ratch Jobe Completed 137 1 K 1 K	89% 58% 55%	41% 58% 57%

- Customizable dashboards for each experiment
- Display information experiments deem relevant to shifters in a single place
- Could drill down to more details

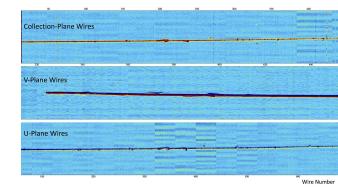
campaign 1604 -						## POMS Campaign Efficiency	## POMS Campaign Jobs	# POMS Campaign Stats	## POMS Tasks
C ²	Batch Job Status			Active Job Status			Batch Job Efficiency		
	- Car - ## - ## - #00	celled	total percentage 41 19% 8 4% 41 19% 129 59%	ensa tide finar-i 5 min. Lide piza + 0 noming jutar-0					
				Active/Recent POMS Tasks 👻					
ID +	POMS_LAUNCHE	R		Batch Jobs Completed	Average CPU Eff		Average Mem Ef	7	
219031	renlu23			137	89%		41%		

The FIFE tools in action: ProtoDUNE

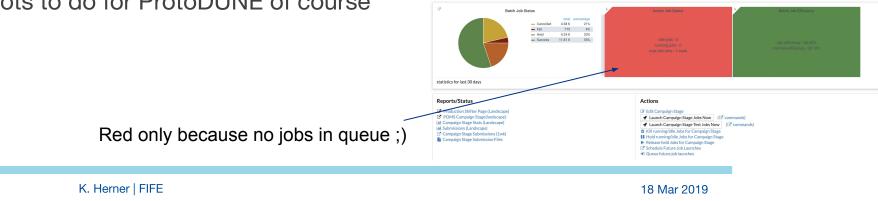
ProtoDUNE using SAM, F-FTS, jobsub, GlideinWMS, POMS (workflow management), CI, Landscape

Using common tools makes it easy for people to contribute quickly

Still lots to do for ProtoDUNE of course



Campaign Stage protoDUNE_SP_cosmics_7.5ms_v07_08_00_05 in Campaign protodunesp_keepup_cosmics_v07_08_00_05_7.5ms__3261



Fermilab Experiment Schedule

Fermilab Program Planning 5-April-18

LONG-RANGE PLAN

		FY18	FY19	FY20	FY21	FY22	FY23	FY24	FY25	FY26	FY27	FY28	FY29	FY30	
LBNF /	SANFORD		1		DUNE	DUNE	DUNE	DUNE	DUNE	DUNE	DUNE	DUNE	DUNE	DUNE	
PIP II	FNAL					LBNF	LBNF	LBNF	LBNF	LBN F	LBNF	LBNF	LBNF	LBNF	
NuMI	MI	IINER v.	IINERv.	OPEN	OPEN	OPEN	OPEN	DPEN							
NUM	IVII	NOvA	NOvA	NOvA	NOvA	NOvA	NOvA	NOv/							h
() ()		BooNI	BooNI	BooNE	OPEN	OPEN	OPEN	DPEN			OPEN	OPEN	OPEN	OPEN	
BNB	В	ICARUS	CARUS	CARUS	ICARUS	ICARUS	ICARUS	DPEN			OPEN	OPEN	OPEN	OPEN	
	364254	SBND	SBND	SBND	SBND	SBND	SBND	DPEN			OPEN	OPEN	OPEN	OPEN	
D.G	Comulau	g-2	g-2	g-2				100	C CHI MIN					OPEN	
wuon	Complex	Mu2e	Mu2e	Mu2e	Mu2e	Mu2e	Mu2e	Mu2	G SHUTD	JWN	Mu2e	Mu2e	Mu2e	OPEN	μ
	MT	FTBF	FTBF	FTBF	FTBF	FTBF	FTBF	FTBF			FTBF	FTBF	FTBF	FTBF	
SY 120	MC	FTBF	FTBF	FTBF	FTBF	FTBF	FTBF	FTBF			FTBF	FTBF	FTBF	FTBF	р
	NM4	OPEN	E1039	E1039	E1039	E1039	OPEN	DPEN			OPEN	OPEN	OPEN	OPEN	1
		FY18	FY19	FY20	FY21	FY22	FY23	FY24	FY25	FY26	FY27	FY28	FY29	FY30	Г
	Construction / commissioning Run Subject to PAC review Shutdown														

- NOTES: 1. Mu2e estimates 4 year running starts mid-FY22 after 18 months commissioning. Assume, with contingency, 5.5 years data taking.
 - 2. DUNE: 1st 10kT detector module commissioned in FY24. Runs without beam FY25 to mid-FY26.
 - 3. NOvA runs as long as possible [in the spirit of PAC Nov 2017].
 - 4. Assume NuMI in nubar mode through FY19 facilitates 12E20 POT for MINERvA [PAC Nov 2017]. Assumption may need revision.
 - 5. Assume g-2 completed before Mu2e commissioning start mid-FY20. Very tight. Needs scrutiny.
 - 6. Assume E1039 fully approved & commissioned by mid-FY19.
 - Experiment estimates 2 yrs run. Add 1 yr contingency. [Stage 1 approval PAC June 2013, update July 2017]
 - 7. FY19 and FY20 MicroBooNE running subject to future PAC review [PAC July 2017].

CI Example

Multiplatform Continuous Integration for LarCI

	Build 💡	Start Time 💡	Build Type 💡	setup_environment ?	checkout 🕜	build 🕜	unit_test 💡	install 💡	ci_tests 💡
Se	uboone_ci/498 (LArSoft uBooNE)	2018-07-05 12:11:37	d16 c2:prof	\checkmark	~	\bigotimes	?	?	?
	Build 🕜	Start Time 🕜	Build Type 🕜	checkout 🕜	build 🕜	unit_test 🕜	install 🕜	ci_tests 💡	ci_validation 🕜
ł b	lar_ci/2775 (uBooNE)	2018-07-05 11:48:16	slf6 e15:prof	\checkmark	~	\otimes	~	×	0
bι	lar_ci/2774 (uBooNE)	2018-07-05 10:18:34	slf6 e15:prof	\checkmark	~		~	×	0
	Build 🕜	Start Time 🕜	Build Type 🕜	setup_environment 💡	checkout 🕜	build 🕜	unit_test 🕜	install 🕜	ci_tests 🕜
Se Fi	uboone_ci/497 (LArSoft uBooNE)	2018-07-05 09:43:09	slf7 e15:prof	\checkmark	~	\checkmark	8	~	
	uboone_ci/496 (LArSoft uBooNE)	2018-07-05 09:36:04	slf7 e15:prof	~	\checkmark	\checkmark	8	~	
	uboone_ci/495 (LArSoft uBooNE)	2018-07-05 09:17:23	slf6 e15:prof	~	\checkmark	\checkmark	8	~	\checkmark
Þli	uboone_ci/494 (LArSoft uBooNE)	2018-07-05 09:06:57	slf6 e15:prof	~	~	\checkmark	8	~	\checkmark
Βu	Build 🕜	Start Time 🕜	Build Type 🕜	checkout 🕜	build 🕜	unit_test 🕜	install 🕜	ci_tests 🕜	ci_validation 💡
Te	lar_ci/2773 (uBooNE)	2018-07-05 08:57:44	slf6 e15:prof	~	~	\otimes	~	×	0
∀a	lar_ci/2770 (uBooNE)	2018-07-05 07:49:52	slf6 e15:prof	~	\checkmark	\otimes	~	×	0
	Build 💡	Start Time 💡	Build Type 💡	setup_environment 💡	checkout 💡	build 💡	unit_test 💡	install 💡	ci_tests 💡
	uboone_ci/493 (LArSoft uBooNE)	2018-07-04 21:40:36	slf6 e15:debug	~	~	\checkmark	0	~	\otimes