

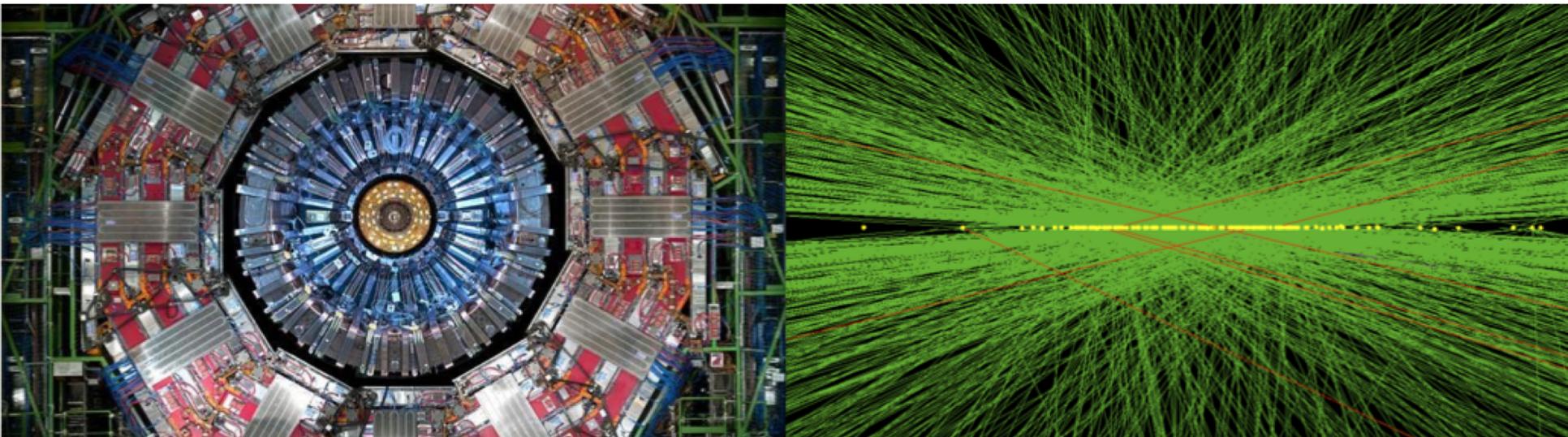


B04 - TDAQ DAQ upgrade project

Remi Mommsen (Fermilab)

HL LHC CMS CD-1 Review

October 23, 2019





Outline

- Technical Aspects of DAQ
 - Conceptual Design
 - Scope and U.S Deliverables
 - Progress since June 2018 IPR
 - QA/QC

- Managerial aspects of DAQ
 - Cost, Schedule, and Risks
 - Contributing Institutions
 - ES&H

- Summary



Who am I?

- Remi Mommsen, Computing Prof., Fermilab
 - PhD in Physics (Bern, Switzerland, 2001)
 - ATLAS (TDAQ), BaBar (central analysis) and DØ (track trigger)
 - Working on CMS DAQ for FNAL since 2008
 - Deputy project leader for CMS DAQ
 - US CMS level-2 DAQ operations manager
 - Project leader for new CMS Online Monitoring System (OMS)
 - Event-builder software and performance tuning
 - Run-1 storage-manager software



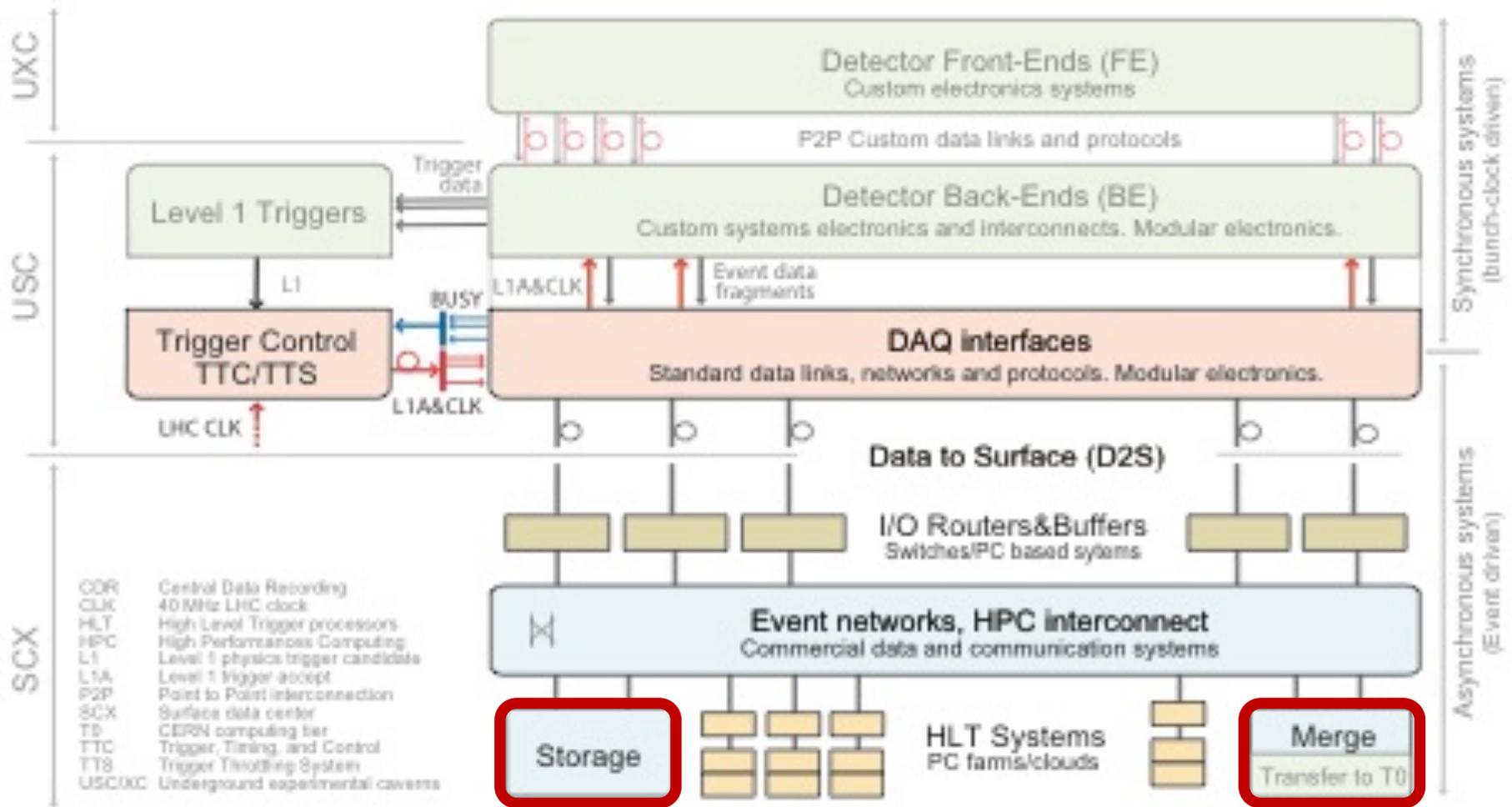
Conceptual Design

Design Considerations for 402.6.6

- Baseline HL-LHC DAQ/HLT architecture same is today
 - Events are read out and assembled at the L1 trigger rate
 - Complete events are delivered to the High-Level Trigger (HLT)
 - HLT runs on commercial processors and selects ~1% of events
 - Accepted events are collected from HLT nodes and temporarily stored before transferred to tier 0

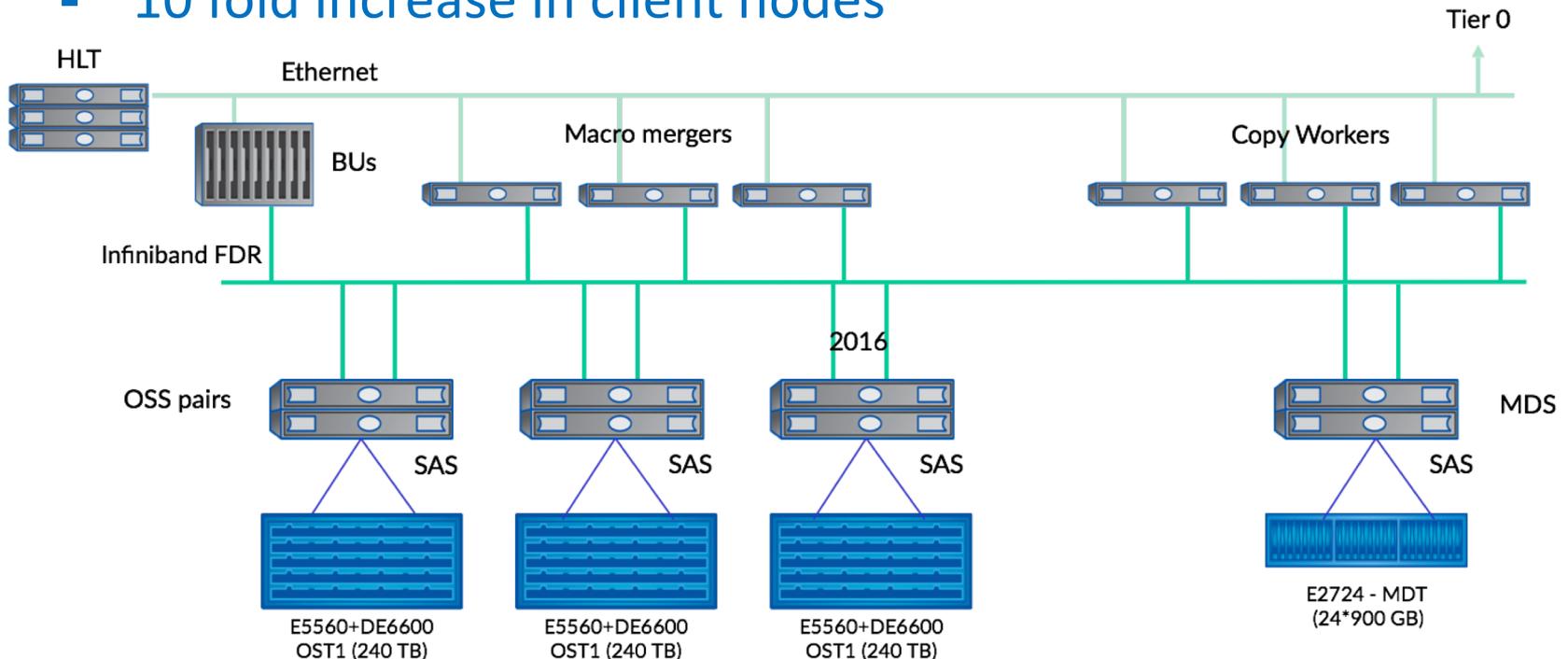
CMS detector Peak \langle PU \rangle	LHC Run-2	HL-LHC Phase-2	
	60	140	200
L1 accept rate (maximum)	100 kHz	500 kHz	750 kHz
Event Size	2.0 MB ^a	5.7 MB ^b	7.4 MB
Event Network throughput	1.6 Tb/s	23 Tb/s	44 Tb/s
Event Network buffer (60 seconds)	12 TB	171 TB	333 TB
HLT accept rate	1 kHz	5 kHz	7.5 kHz
HLT computing power ^c	0.5 MHS06	4.5 MHS06	9.2 MHS06
Storage throughput	2.5 GB/s	31 GB/s	61 GB/s
Storage capacity needed (1 day)	0.2 PB	2.7 PB	5.3 PB

Design Considerations for 402.6.6



Design Considerations for 402.6.6

- No fundamental change foreseen w.r.t. today's system
 - Cluster file system on commercial hardware (Lustre/NetApp)
 - 31 GB/s throughput with 2.7 PB storage for startup system
 - 12 times the performance of today's system
 - ~10 fold increase in client nodes



Deliverables for 402.6.6

- Storage Manager and Transfer System
 - Collect events accepted by HLT from the filter farm
 - Aggregate monitoring information from the HLT processors
 - Transfers the data to tier 0 for offline reconstruction or to consumers at the experimental site for online data-quality monitoring or fast calibration
- Based on a commercial storage system
 - Buy hardware as late as possible to profit from cost decrease
 - Small scale test system (3 GB/s) early 2025
 - Low-pileup startup system (31 GB/s) ordered by end of 2025

- Commercial h/w with required specs is available today
 - Continuous market watch for emerging technologies
 - Evaluation of system with best performance/cost will be done during 2024
- Current storage system to be replaced for run 3 (2021)
 - Run 3 system has similar requirements than run-2 system
 - Purchase of new hardware foreseen during 2020
 - Allows for a cost/performance re-assessment
 - Might reduce uncertainty on cost on timescale of CD-2

- Baseline DAQ architecture documented in iTDR
 - “The Phase-2 Upgrade of the CMS DAQ Interim Technical Design Report”, CERN-LHCC-2017-014, CMS-TDR-018, 12 Sep 2017, <https://cds.cern.ch/record/2283193>
- Storage system is a COTS system
 - Will be bought in 2025
 - System with the HL-LHC requirements can be bought today
 - Reasonable extrapolations of costs based on the cost of the current system and vendor quotes



Quality Assurance and Quality Control

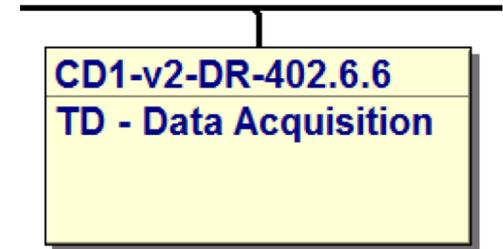
- QC for storage hardware will be done upon reception from vendor
 - Includes burn-in and performance tests of the installed system

Cost and Schedule

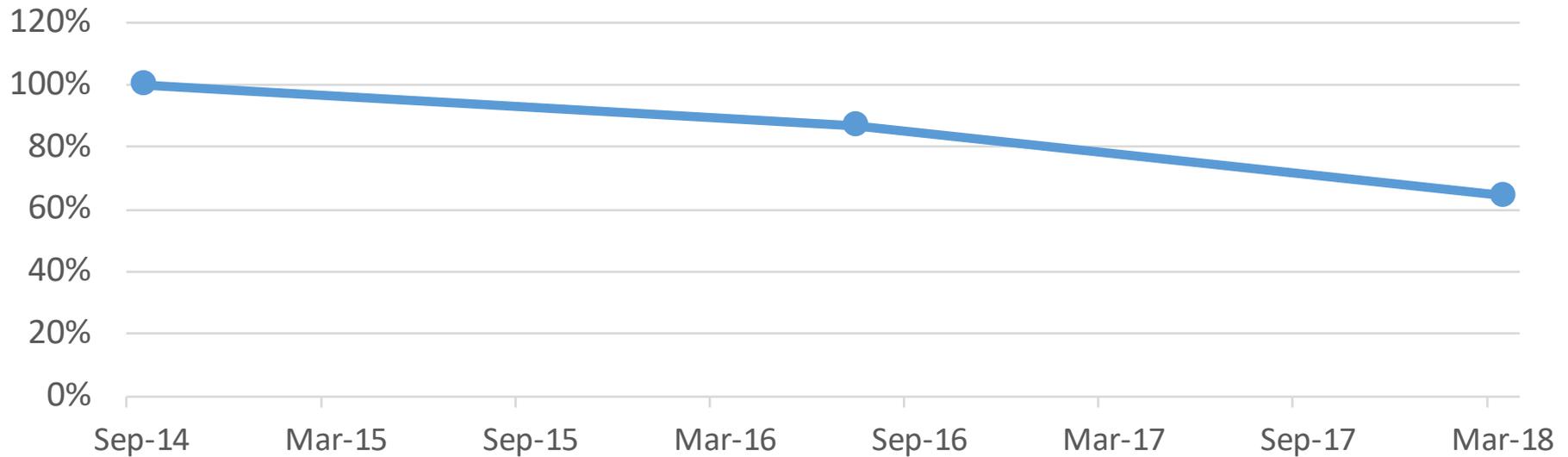


402.6.6 WBS Structure

- DAQ is not broken down into L4 areas



Cost Estimate Technique



- Based on cost payed for today's storage system
 - Initial purchase in Sep 2014 & addition of 3rd OSS in July 2016
 - Cost/performance improvement of 7.2% per year
 - Trend confirmed by a vendor quote in March 2018
- Estimated cost is consistent with a quote from another vendor for a system with HL-LHC specifications
- New storage system for run 3 will be bought during 2020
 - Might allow to reduce uncertainty on cost for HL-LHC storage system

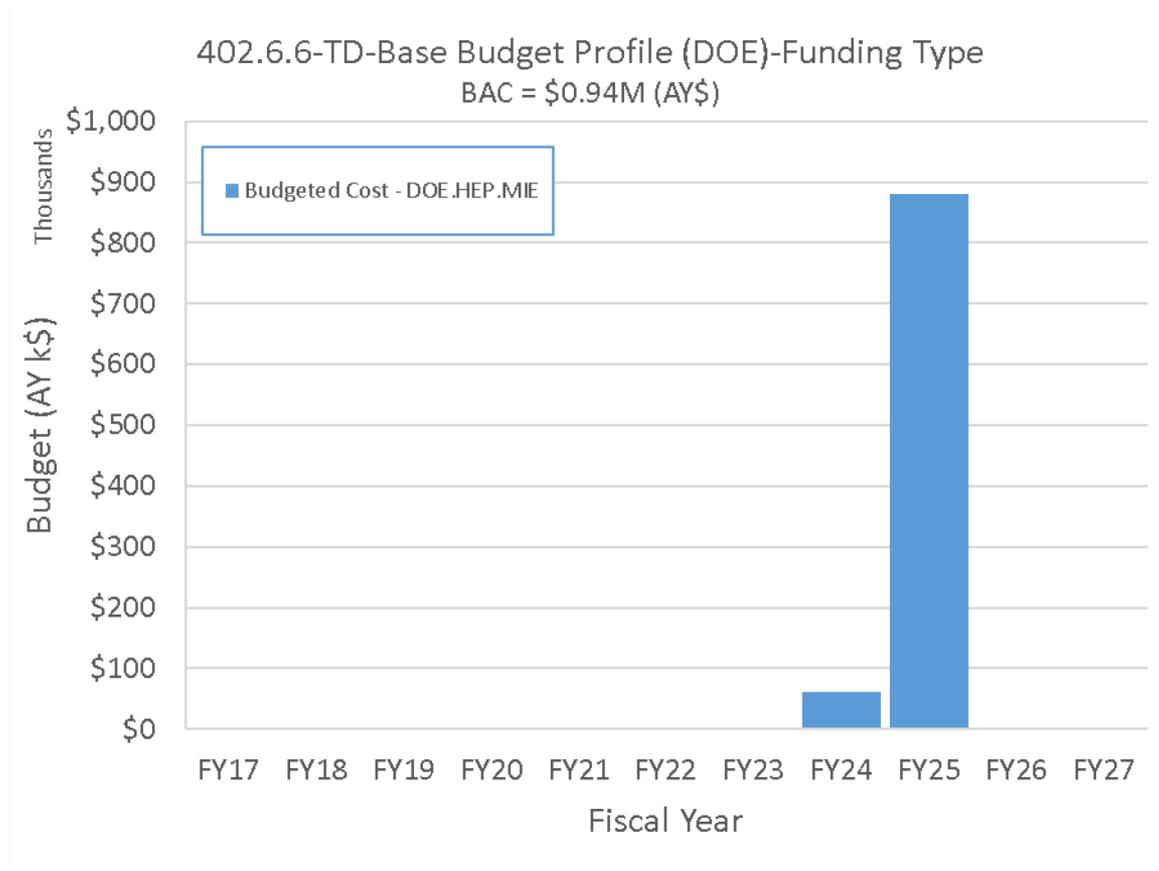
Cost Estimate

WBS	Direct M&S (\$)	Labor (Hours)	FTE	Direct + Indirect + Esc. (\$)	Estimate Uncertainty (\$)	Total Cost (\$)
DOE-CD1-402.6 402.6 TD - Trigger and DAQ (at DOE CD1)	4,004,125	107439	60.77	9,087,893	2,438,143	11,526,036
DOE-CD1-402.6.2 TD - Management (DOE)	171,594	25470	14.41	215,167	13,745	228,912
DOE-CD1-402.6.3 TD - Calorimeter Trigger	1,400,335	31043	17.56	3,266,174	808,451	4,074,625
DOE-CD1-402.6.5 TD - Correlator Trigger	1,664,196	50926	28.80	4,667,023	1,146,183	5,813,206
DOE-CD1-402.6.6 TD - Data Acquisition	768,000	0	0.00	939,529	469,764	1,409,293

- Only M&S for storage hardware
 - 80k for test system with ~10% capacity of production system
 - 138k for metadata servers and head nodes
 - 550k for storage system for low-pileup running (31 GB/s)
 - 50% “Conceptual” (M5) uncertainty



Fiscal Year Cost Profile



Contingency Breakdown

- Current estimate based on a scaled-up version of the current storage system
 - Preliminary estimate of event sizes and HLT output rate
 - Poorly known cost/performance improvement over 10 years
- Assign “Conceptual” (M5) contingency to cost estimate
 - Choose mid-point of 50% for uncertainty estimate
- Uncertainty of 470k USD

Risk – RU-402-6-07-D

- The I/O requirements of the DAQ STMS could change in response to changes in event size or HLT output event rate. This event, in turn, requires more (or less) STMS resources in response.
 - Purchase cost of DAQ STMS (768k\$) scaled down/up by -25%/50%/100% (-192/384/768 k\$)
 - 20% probability
 - Open until purchase of system (3-Jun-2025)
- Mitigations could include
 - Higher compression of event data
 - Less monitoring/calibration data
 - Reduced event content
 - Accept less events

Risk – RU-402-6-07-D

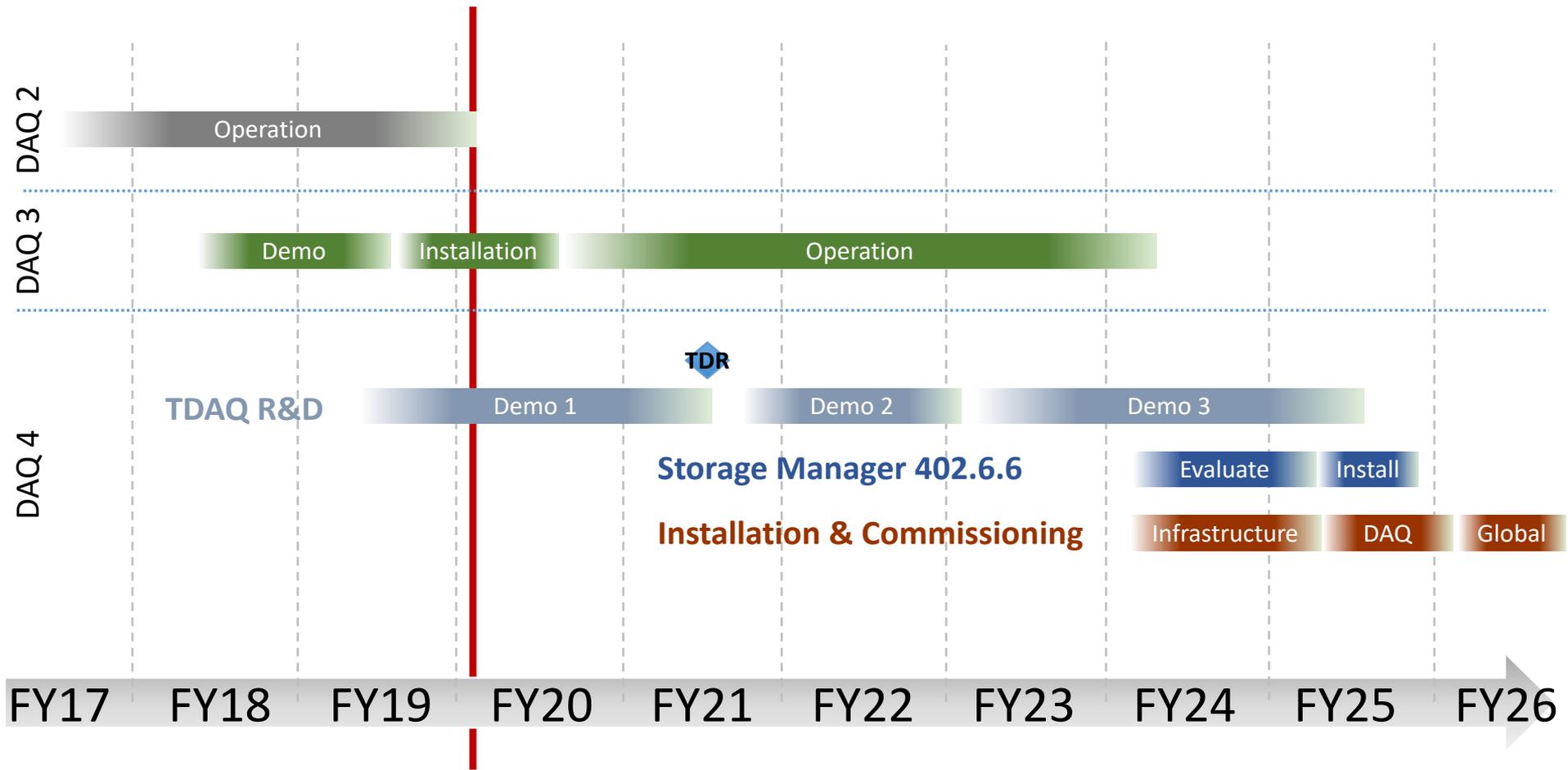
RU-402-6-07-D TD - DAQ STMS I/O performance does not meet requirements (DOE)

Risk Rank:	2 (Medium) Scores: Probability : 2 (L) ; Cost: 2 (M) Schedule: 0 (N))	Risk Status:	Open
Summary:	The I/O requirements of the DAQ STMS could change in response to changes in event size or HLT output event rate. This event, in turn, requires more (or less) STMS resources in response.		
Risk Type:	Uncertainty	Owner:	Jeffrey W Berryhill
WBS:	402.6 TD - Trigger and DAQ (DOE)	Risk Area:	Technical Risk / Requirements
Probability (P):	20%	Technical Impact:	2 (M) - significantly substandard
Cost Impact:	PDF = 3-point - triangular Minimum = -192 k\$ Most likely = 384 k\$ Maximum = 768 k\$ Mean = 320.0 k\$ P * <Impact> = 64.0 k\$	Schedule Impact:	PDF = 1-point - single value Minimum = N/A Most likely = 0.0 months Maximum = N/A Mean = 0 months P * <Impact> = 0 months
Basis of Estimate:	Purchase cost of DAQ STMS (768k\$) scaled down/up by -25%/50%/100%		
Cause or Trigger:	Specifying HLT output event rate or event size different from TDR values.	Impacted Activities:	Procurement cost of DAQ STMS
Start date:	1-Oct-2018	End date:	3-Jun-2025
Risk Mitigations:	Ongoing evaluation of DAQ configuration in response to HLT design and T0 computing.		
Risk Responses:	Negotiation with CMS DAQ on re-configuration of project to meet startup goals.		
More details:			

- Added risk for change in I/O requirements (+\$130k impact)



Schedule Overview

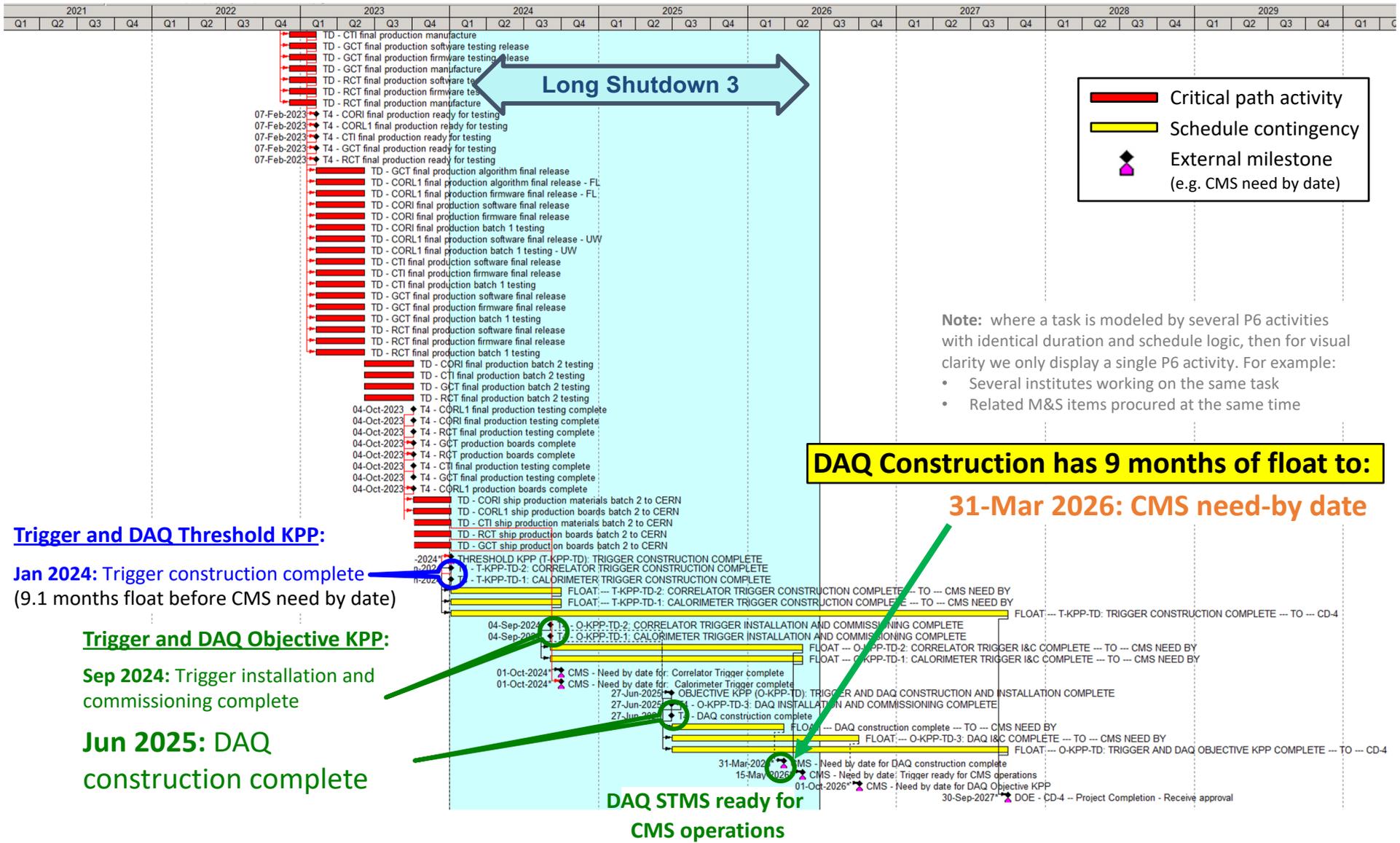


Schedule Overview

- Detailed documentation of DAQ system with better estimates in DAQ TDR which is due mid 2021
 - DAQ installation completed mid 2025
 - Ready for detector commissioning by end 2025
 - Full performance required by start of run 4 (fall 2026)
- Evaluation of storage hardware 2024/25
 - Select storage hardware and cluster filesystem in fall 2024
 - Procure test system end of 2024
 - Evaluate system capabilities and adapt software
- Storage hardware procurement in FY25
 - Commissioning and performance tuning
 - Ready for data taking in fall 2026



Critical Path and Float



Contributing Institutions and Resource Optimization



Contributing Institutions

■ MIT

- Responsible for storage and transfer system since run 1
- Storage hardware evaluation, commissioning and monitoring
- Development of the merger and transfer software
- Darlea is our key expert for storage h/w and Lustre file system

■ UCSD

- Network configuration, monitoring and optimization

■ FNAL

- Definition of interfaces
- Coordination with CMS DAQ and stakeholders
- Mommsen has a long experience from run 1 and 2 with integration of the storage manager system into the online system

■ Rice

- Experience in providing a monitoring system for online users

- Storage system installation is done by the vendor
 - Hardware deployment at the experiments side (pt.5)
 - Configuration to reach agreed level of raw performance
 - Hardware and configuration support
- Integration into the DAQ and adaption of software done by MIT/UCSD/FNAL/Rice
 - Long experience with similar systems
 - Custom s/w tailored to needs of CMS
 - Coordination with the rest of the DAQ system and tier 0

ES&H

- As with entire project, we follow the Integrated Safety Management Plan ([cms-doc-13395](#)) and have documented our hazards in the preliminary Hazard Awareness Report ([cms-doc-13394](#))
- There are no specific Hazards for 402.6.6.



Summary

- Storage Manager and Transfer System is an US responsibility since the dawn of CMS
 - Long experience with similar systems within the US DAQ groups
- The storage hardware is a COTS
 - Systems fulfilling the HL-LHC requirements can be bought today
 - System will be purchased as late as possible to profit from cost/performance improvements
 - Cost estimate base on recent purchases of a similar system and vendor quotes
 - 50% contingency on estimate