# DOMA: Preparing for Year 2 and Year 4



Institute for Research & Innovation in Software for High Energy Physics









/rə'trēt/

verb

"the French retreated in disarray" synonyms: withdraw, retire, draw back, pull back, pull out, fall back, give way, give ground, recoil, flee, take flight, beat a retreat, beat a hasty retreat, run away, run off, make a run for it, run for it, make off, take off, take to one's heels, make a break for it, **bolt**, make a quick exit, clear out, make one's getaway, escape, head for the hills; More

### noun



an act of moving back or withdrawing. "a speedy retreat"

synonyms: withdrawal, pulling back, flight; rare katabasis "a counteroffensive caused the retreat of the imperial army"

2. a signal for a military force to withdraw. "the bugle sounded a retreat"

Translations, word origin, and more definitions





### (of an army) withdraw from enemy forces as a result of their superior power or after a defeat.



## **Unifying DOMA Vision**





The DOMA team is working across the following areas:

- **Modeling**: understanding how we utilize the data management systems today and potential impact of changes.
- **Organization**: What data is written to disk and how it is serialized.
- Management: Bulk data movement between storage facilities.
- Access: how "data sinks" compute clusters or analysis systems accesses the data.

The strength of the team is we can have an end-toend system view!

FEARLESS SCIENCE

3









We have a number of small-to-medium projects ongoing:

- Modeling of cache usage.
- **XCache**: Xrootd-based proxy file cache.
- Data movement with **HTTP-TPC**.
- **IDDS**: Delivering events, esp. to production systems.
- ServiceX: Delivering columns, esp. to analysis facilities.
- **SkyHook DM**: Managing data (particularly, HEP-like data) with Ceph.

What's needed for Year 2?





## Year 1 - Commonalities

## Year 1:

- Brought together several teams. •
- Continued a number independent efforts.
- Started a few small projects to address specific needs.
- broad number of areas – but many of these were *independent*!





Had several reporting milestones to capture knowledge or community progress.

This allowed us to bring the DOMA team together and make significant progress on a

#### MORGRIDGE INSTITUTE FOR RESEARCH FEARLESS SCIENCE 5



## **Milestones and Deliverables**

	Data Organization, Management and Access	Done	Late	In-Progress		Not-Started / Due soon										
Label	Description	Type (M/D)	Y1Q1	Y1Q2	March 2019	April 2019	May 2019	June 2019	July 2019	Aug. 2019	Sept. 2019	Oct. 2019	Nov. 2019	Y2Q2	Y2Q3	Y2Q4
G3.1	Start a process for regular exchanges between ATLAS and CMS on understanding use data from both experiments	D	DUE													
G3.2	Begin technical work on caches and interoperable data transfer mechanisms	м	DUE													
G3.3	Organize topical meeting on data delivery through CDNs within LHC and comparisons with other users	D		DUE												
G3.4	Initial web presence explaining the DOMA area activities	D		DUE												
G3.5	Organize discussions for "data lake" and "IDDS" use cases, stakeholders, and development plans	D		DUE												
G3.6	First draft of IDDS design	М		DUE												
G3.7	Prepare internal document outlining opportunities for minimizing event data storage sizes	D		DUE							Projec	ted Dor	ie			
G3.8	Deliver report on LHC data access patterns, usage and intelligent caching approaches for the HL-LHC	D								DUE	Projec	ted Dor	e			
G3.9	First technical demo of IDDS	М								DUE	Projec	ted Dor	e			
G3.10	Report on cache usage on the WLCG and potential use cases & deployment scenarios for the US LHC facilities	D								DUE	Projec	ted Dor	ie			
G3.11	Prototype of HEP event data on object storage using alternate organization approaches	D											DUE			
G3.12	With OSG, help transition 30% of data transfers at one USLHC site to use a non-Globus Toolkit implementation	м												DUE		
G3.13	Report on feasibility of hardware accelerators for event data	D												DUE		]
G3.14	IDDS prototype that can automatically transform / optimize data formats.	М												DUE		
G3.15	With SSL, scale test a data lake prototype across multiple sites.	м														DUE
G3.16	First version of IDDS at production site	D														DUE



6







## **The Big Milestones – Year 2**

The most complex milestones for Year 2 are:

- Prototype HEP data on an object store (1 December). lacksquare
- With OSG, transition 30% of data transfers at one site to a non-GridFTP protocol (1 March). IDDS prototype for transforming data (1 March) ... and at a production site (1 Sept). With SSL, scale test a "data lake prototype" (1 Sept).
- What is shared amongst these? Integration!
- For example, the object store milestone needs to receive data (think: IDDS or ServiceX) and • then needs to serve the data to a consumer (think: Coffea, func-adl?).
- There is need for integration within DOMA (IDDS, ServiceX, SkyHook), within IRIS-HEP (DOMA, SSL, AS), and with external entities (Coffea project).

## How to best achieve this?







## **Data Lake Prototype**

I'd rather have them become "data lake challenges". Something along the lines of:

- **Challenge #1 (3mos)**: Use ServiceX to deliver 1M events to SkyHook.
- Challenge #2 (6mos): Via Rucio & HTTP, orchestrate movement of 200TB of CMS  $\bullet$ NanoAOD through a ServiceX instance on the SSL and process for a "real analysis" with Coffea.
- **Challenge #3 (12mos)**: Demonstrate processing of 1PB of events via IDDS (maybe on an lacksquareHPC?), transform them using ServiceX, cache in SkyHook, and deliver to Coffea for analysis.

discuss the contents at today's DOMA breakout post-lunch.

Organizing along these lines helps illustrate the impact of the DOMA area. 





We have distinct deadlines in 3, 6, and 12 months. Rather than treat the milestones independently,

Don't mind the contents of the above three – the important aspect is increasing scale. We will

### MORGRIDGE INSTITUTE FOR RESEARCH FEARLESS SCIENCE 8



## Year 4 - Vision

- As we plan out the tactics to finish Year 2, we should think of the vision for Year 4 (2021). • Year 4 will be mostly through the execution phase; we will potentially be starting to write a new proposal. Computing TDRs will (hopefully) be mostly done will be in Run3. What accomplishments will we want under our belt?
- lacksquareHere's my view for DOMA:
- "Over the hump" of the transition away from GridFTP. Majority of traffic at most sites will use something else.
- Regularly delivering 100's of TBs of ntuples to object-store-based analysis facilities.
- Caches are widely deployed via k8s as buffers and for heavily-used data. What are the missing pieces we should add to this?



MORGRIDGE INSTITUTE FOR RESEARCH FEARLESS SCIENCE 9





## **RETREAT!**

re-treat

/rəˈtrēt/

### verb

verb: retreat; 3rd person present: retreats; past tense: retreated; past participle: retreated; gerund or present participle: retreating

### (of an army) withdraw from enemy forces as a result of their superior power or after a defeat.

"the French retreated in disarray" synonyms: withdraw, retire, draw back, pull back, pull out, fall back, give way, give ground, recoil, flee, take flight, beat a retreat, beat a hasty retreat, run away, run off, make a run for it, run for it, make off, take off, take to one's heels, make a break for it, bolt, make a quick exit, clear out, make one's getaway, escape, head for the hills; More antonyms: advance, dig in

- · move back or withdraw. "it becomes so hot that the lizards retreat into the shade" synonyms: go out, ebb, recede, flow out, fall, go down "the tide was retreating" antonyms: come in
- withdraw to a quiet or secluded place. "after the funeral he retreated to the shore"
- change one's decisions, plans, or attitude, as a result of criticism from others. plans; More
- (of shares of stock) decline in value. "shares retreated 32 points to 653 points"
- CHESS move (a piece) back from a forward or threatened position on the board.





"his proposals were clearly unreasonable and he was soon forced to retreat" synonyms: change one's decision, change one's mind, change one's attitude, change one's









This material is based upon work supported by the National Science Foundation under Grant No. <u>1836650</u>. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## FEARLESS SCIENCE

## morgridge.org

