

# HOW Summary

## The Future of Scientific Computing

Amber Boehnlein  
IT Division Director  
CIO

Jefferson Lab



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science



# There are a hundred ways to waste paint...

- And #2 is to write a 'A future of' Talk
  - I agreed to this?
- The following remarks are solely my opinions.
  - They may be uninformed
  - They are certainly biased
  - And unvarnished
  - And limited to topics where I had something to say in this moment
- Open to making this a conversation--as long as everyone can still get to Dulles.....



# Compliments and Kudos

---

- There is so much good to say
  - Congratulations to HEP Software Foundation on the road mapping process
  - Congratulations to OSG—forward looking vision and solid funding model
  - Congratulations on looking for ways to expand the WLCG model to facilitate other communities within HEP and beyond
  - Congratulations to the IRIS-HEP team for establishing a US based collaborative project
- This week was too information dense for anyone to take in everything
  - The presentations were well prepared
  - The LHC experiment presentations looked consistent and coordinated
  - Lots of humor

This is reflective of a true community that has emerged around HEP Computing and Software.

# A View of LHC Computing and Software

- Physics Analysis is a job that requires finesse
- Dealing the data volumes, processing time, production tasks is a job that requires a 'backhoe'
- And somehow these competing requirements have to be reconciled
  - On a very tight budget
- This contradiction leads to one of the most common framing devices

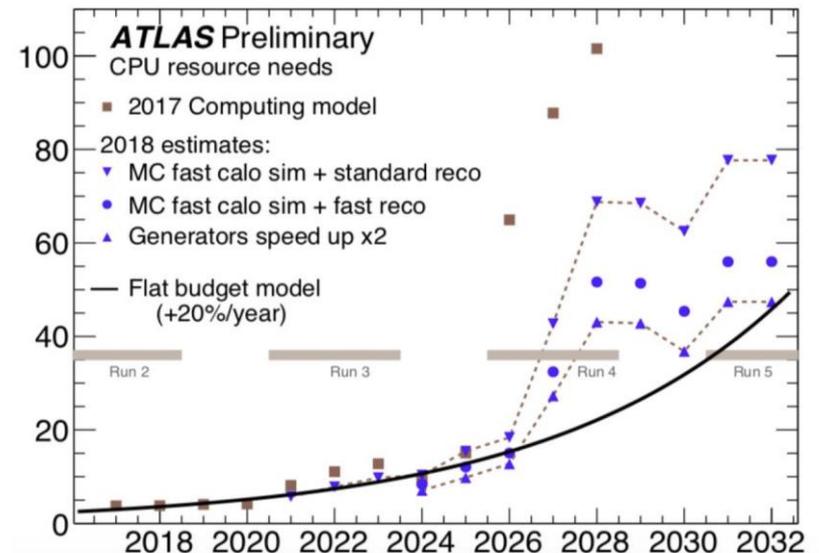


# A View of LHC Computing and Software

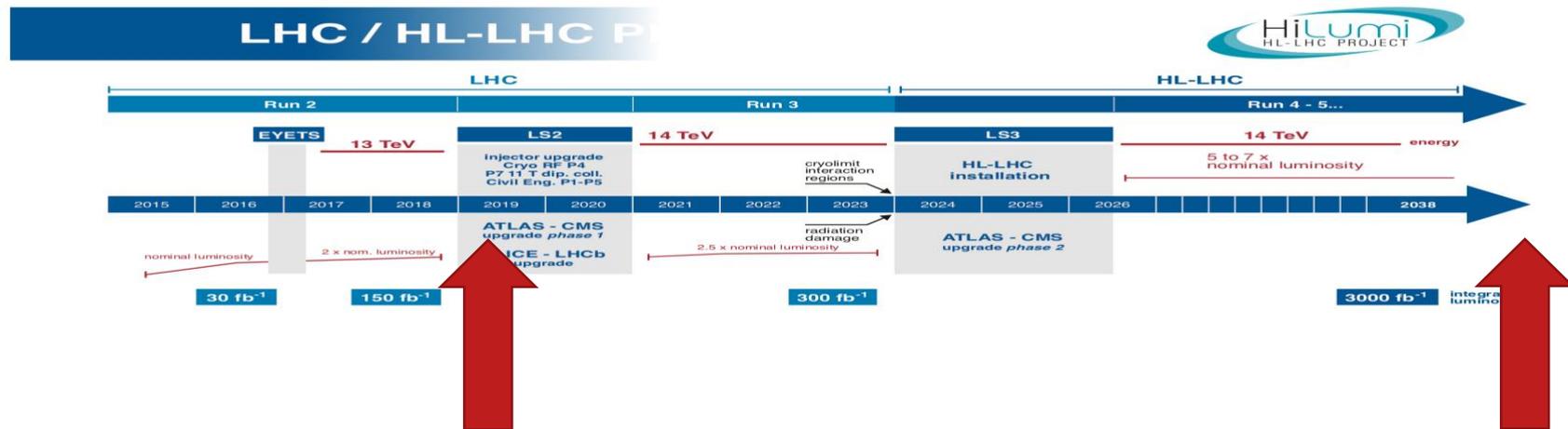
- Physics Analysis is a job that requires finesse
- Dealing the data volumes, processing time, production tasks is a job that requires a 'backhoe'
- And somehow these competing requirements have to be reconciled
  - On a very tight budget
- This contradiction leads to one of the most common framing devices
  - Illuminates the special challenges of LHC



Run 4: 75% of simulation with fast sim



# Sustainability and ability to evolve and to transform



- For Run 1, the sheer challenge of making working systems within all the constraints was overwhelming.
  - Succeeded beyond all hope!
- For HL-LHC
  - Additional constraint of sustained operations with
  - Nearly complete change out of the workforce with
  - Massive Changes in Technology while burdened with legacy systems
- Anything reasonable that can be done to automate or simplify operations should be done
  - View as a design constraint
- Building in the ability to evolve and innovate is hard, but is essential
  - Road mapping as a process is a key tool

# A Roadmap for HEP Software and Computing R&D for the 2020s

- The road mapping process represents a best practice.
  - It is clearly a consensus document
- Before I read the roadmap, I made a list
  - If I thought the topic was addressed well
    - I ticked it off and probably won't mention it
  - If I thought a topic was quite well addressed
    - it got a star and I might mention it
  - For the rest, I withheld judgement until after this meeting
    - And some of those now have stars.
    - The roadmap is software focused.
  - Generally everything is at least on a good track
  - And some things much better
- For some stuff, it might be time to....

# KonMari



# What to do about the hoarder?

- Assigning more value to what we own than to what we don't is Human Nature
  - Reject at the trigger level; lost forever
  - Write it to tape: carve it in stone
- It doesn't work for the closet
- It doesn't work for science
  - Makes C&S much harder
  - Which makes everyone's life harder
- Not Sustainable or Evolvable
- How can the collaborations engage to address?



- Compute everything I would possibly need
- Store everything I would possibly need
- Leave it there forever, even if I stop using it
- Never ask myself how much does it cost

Rizzi

# DOMA and DAWG

- Grappling with exactly the right long terms questions
- DOMA is and should be a focus at this time
  - Interesting R&D
- Giving analysts tools and leaving to understandable
  - And it created a divide
  - revisit a user ‘aware’ approach
  - ‘Hiding the how is good’
- Also managing expectations with trains and nano
- All of the experiments are also looking at HLT based analysis



- **Reducing I/O for heavy lifting:**
- **Trains** an accepted solution, can more workflows use this concept?
- **Common nano-AOD** centrally produced, less reinventing the wheel on format
  - Q. How much bandwidth can go this route?
- **Turbo stream** calibrate once
  - Q. How much bandwidth can go this route? How strong is the physics case to justify not doing that?
- Convergence on **Jupyter** notebooks as analysis platform, **hiding the how is good**
- Trend towards **declarative analysis**, especially for LHCb/Belle II
  - Does anything prevent other experiments?
- Addressing **systematics** is still a challenge, see **Andrea's** talk
  - Can we attack (some of) this as a community? What is best practice?
  - Not covered - **Monte Carlo**

Laycock

# Event Generation and MC

- It's a science and cost driver
  - The LHC physics program is changing—there will more emphasis on precision measurements and QCD in HL-LHC
    - Have we embraced this?
    - Engaged with the theory Community?
    - Good place to look for partners
  - Many productive lines of inquiry for detector simulation
    - I particularly like the GAN concept
      - But will be a limited solution
  - When will it be time come to completely rethink MC particle transport?
- we need
    - better algorithms Pokorski
    - better code
    - efficient use of current (and future) computing architectures
  - **we need to modernize our code!**

# Machine Learning/Deep Learning/AI

- ML has been an important tool for analysis for almost a generation
  - In a domain specific, labor intensive context.
- It's now a 'thing'-industry standard platforms, R&D work
- In 9 months since CHEP 2018, there has clearly been a lot of foundational work in this area for our field
  - Leaving the 'magic pony' phase
- Tend to use ML to *replace*
  - Do a defined task better or cheaper
  - 'small rock/medium rock problems'
- Are we thinking deeply enough about the ways in these approaches could be transformational?
  - What are the big rocks?



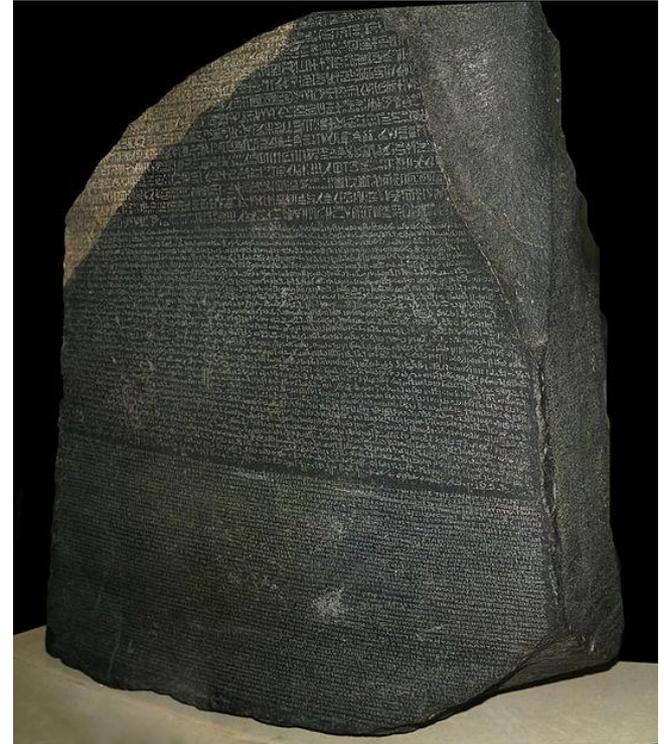
# Technology and HPC strategy

- This community clearly has an excellent grasp on hardware technology futures
  - There is some external pressure to use traditional HPC resources at major compute centers
  - There are also some nice work and prototypes for using them
    - Scavenger mode
    - Event generation
    - ML/Data Science
    - Compute intensive analysis tasks
  - It's also an opportunity to engage with Computer Scientists and Applied Mathematicians
    - They have a lot to offer
  - Pulling together the current landscape and then extending is probably sufficient 'Strategy'
- We want to have the ability to iteratively tune and explore our measurement spaces (i.e. explore systematics)
  - Current single pass workflows don't easily permit this type of feedback mechanism easily
  - Not enough grid computing resources to realistically do this
  - HPC environments can both accelerate these analysis workflows and enable iterative approaches

Norman-NOVA

# Data 'Preservation'

- Consideration of the time scales and of changes to the physics program intensifies the concept of data preservation and open data
  - Some things are nearly impossible to retro-fit that would have been relatively simple to build in from the start
  - Is this an area that deserves more attention?



- I hope you have enjoyed your time at Jefferson Lab
- We have enjoyed having you—we learned so much
- Safe Travels!

