U.S. DEPARTMENT OF ENERGY | Office of Science

Future Computing in High Energy Physics

HEP Software Foundation Meeting

January 23, 2017

Abid Patwa
Program Manager, High Energy Physics
Office of Science, U.S. Department of Energy

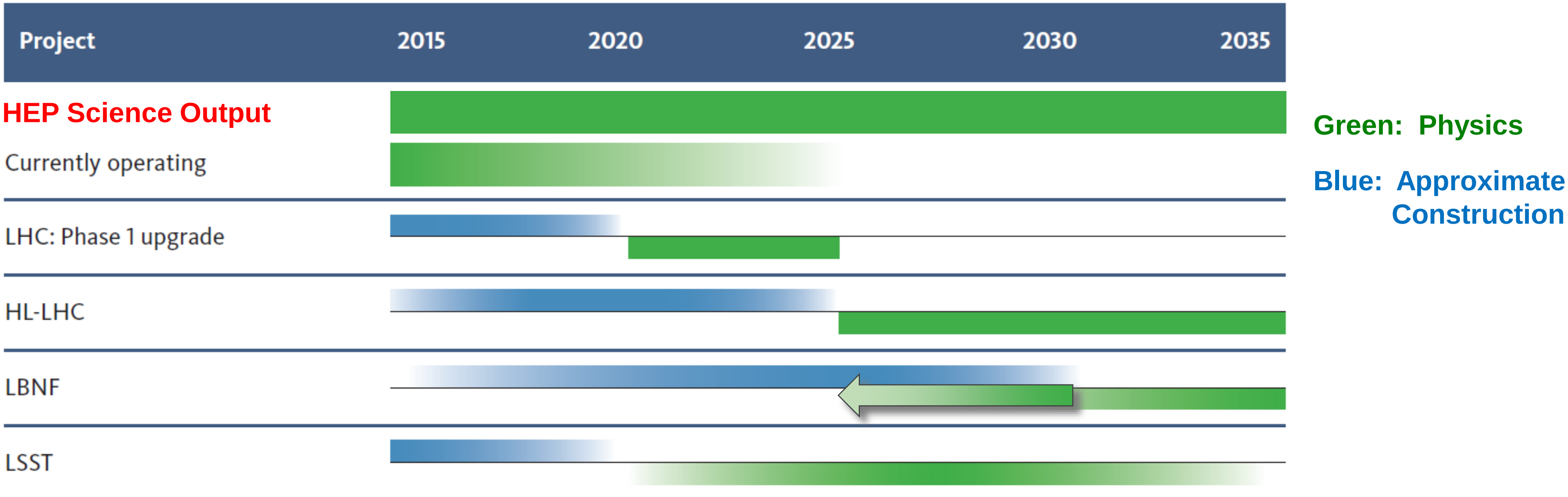
The Science Drivers of Particle Physics

The P5 report identified five intertwined science drivers, compelling lines of inquiry that show great promise for discovery:

- **Use the Higgs boson as a new tool for discovery** **2013*
- **Pursue the physics associated with neutrino mass** **2015*
- **Identify the new physics of dark matter**
- **Understand cosmic acceleration: dark energy and inflation** **2011*
- **Explore the unknown: new particles, interactions, and physical principles**

**** Since 2011, three of the five science drivers have been lines of inquiry recognized with Nobel Prizes***

P5 Science and Construction Timeline (abridged)



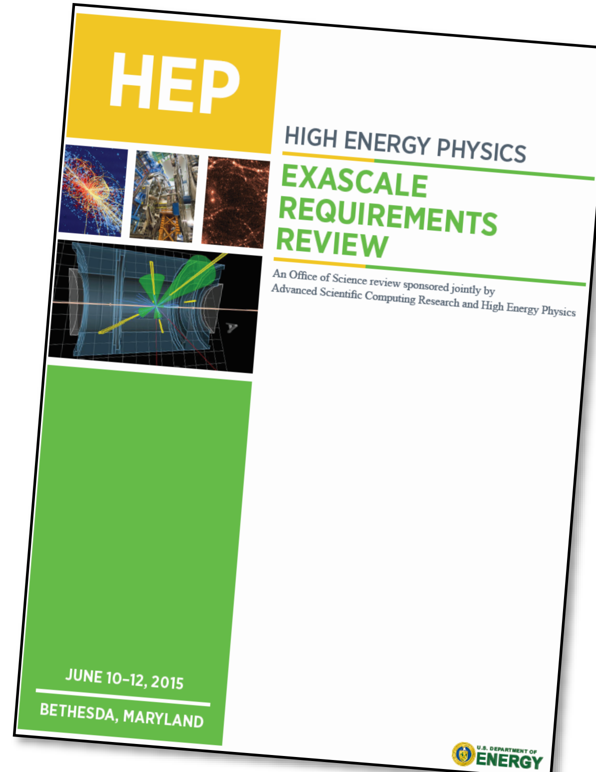
- The P5 report recommends a limited, prioritized and time-ordered list of experiments to optimally address the science drivers
  - Covers the small, medium and large investment scales
  - Will produce results continuously throughout a 20-year timeframe
- HEP is implementing the discovery-driven strategic plan set within a global vision for particle physics as presented in the P5 report
- Realizing this vision will require a shift in approaching the networking and computing challenges for the data that these future experiments will present!

P5 Vision for Computing

- **P5 report recognized the importance of computing:**
  - "Rapidly evolving computer architectures and increasing data volumes require effective crosscutting solutions"
  - "[Need] investments to exploit next-generation hardware and computing models"
  - "Close collaboration of national laboratories and universities across the research areas will be needed"
- **P5 Recommendation 29:**
  - Strengthen the global cooperation among laboratories and universities to address computing and scientific software needs, and provide efficient training in next-generation hardware and data-science software relevant to particle physics. Investigate models for the development and maintenance of major software within and across research areas, including long-term data and software preservation.
- **HEP Response to P5 Recommendation 29:**
  - Initiated HEP Center for Computational Excellence (CCE) http://hepfce.org/

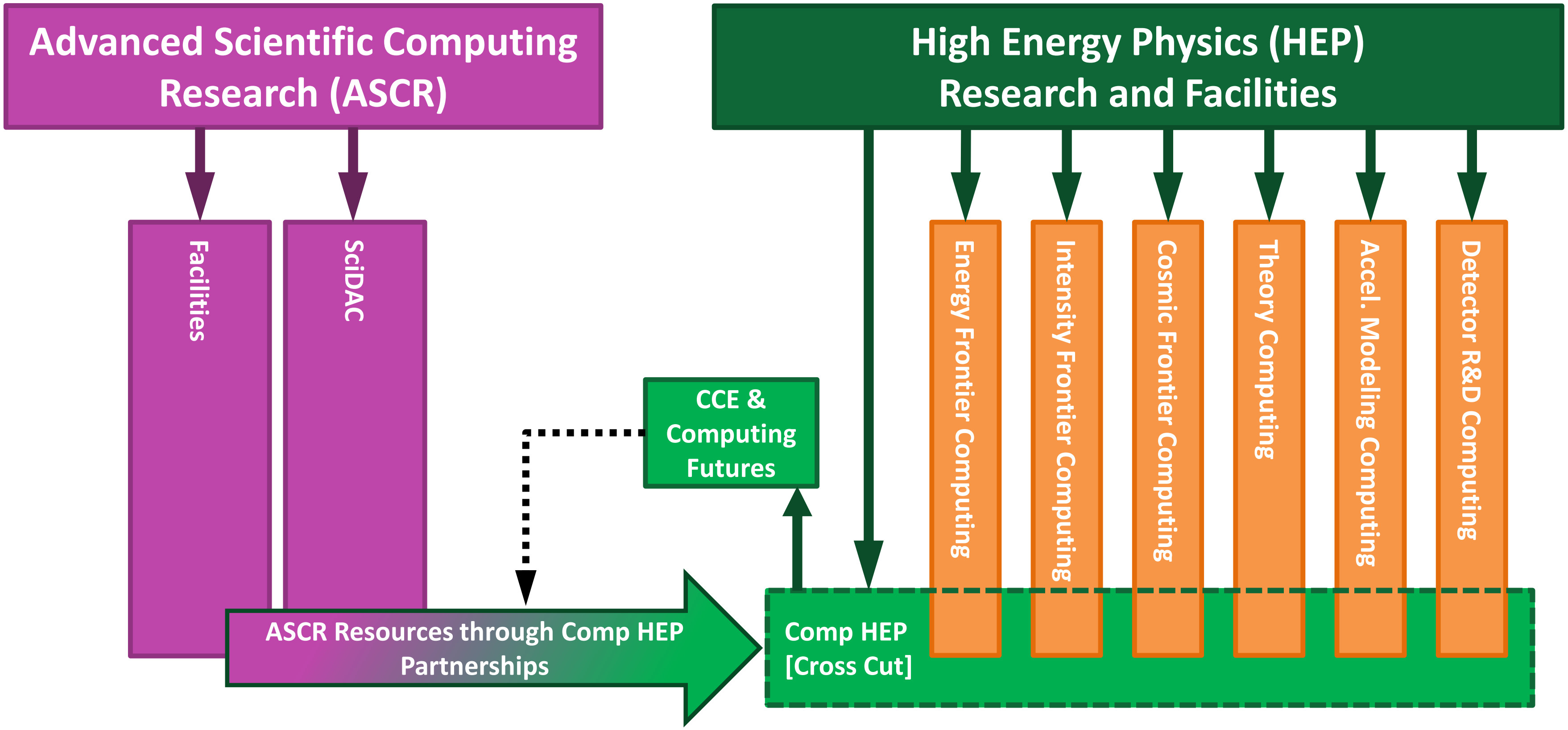
HEP Center for Computational Excellence (HEP-CCE)

- **Primary Mission**
  - Bring next-generation computational resources to bear on pressing HEP science problems
  - Develop cross-cutting solutions leveraging ASCR expertise and resources
- **Technical knowledge to the HEP Community**
  - Software management for HPC systems (containers)
  - Edge services for HPC systems
  - Petascale data transfer project with ESnet
  - Large-scale data analysis using commercial services or opportunistic resources
- **Cross-cutting CCE successes include:**
  - Range of South Pole Telescope projects, including implementing data analysis pipeline in HPC environment to prepare for SPT-3G deployment
  - Ported the Dark Energy Survey data reduction pipeline to a NERSC system by using Docker
    - Replicated and validated pipeline in 2 months running on 20 nodes, with data pushed/pulled via ESnet
  - Adapted serial event generator Alpgen to the parallel HPC environment of Mira for increased performance



Joint ASCR-HEP Exascale Requirements Review (co-organized by HEP-CCE)

Computational HEP and Overall HEP Computing



- **Within DOE HEP, computing is primarily funded within Frontiers and/or Thrusts for program needs**
  - ***e.g.*, LHC computing mainly supported through U.S. LHC Ops within Energy Frontier**
- **Computational HEP, with input from CCE, identifies where external partnerships & cross cuts are possible and fosters them**
  - ***e.g.*, Leverage ASCR partnerships to support LHC computing with ESnet, opportunistic HPC**

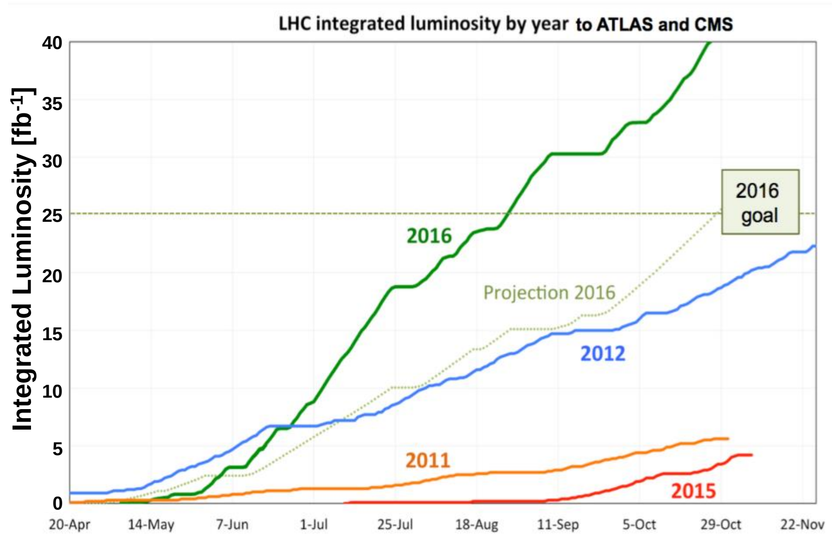
Committee of Visitors (COV)

2016 HEP Committee of Visitors report (covering FY 2013-2015) included a dedicated section on computing *[online at: https://science.energy.gov/hep/hepap/reports/]*

- **Highlights from COV Comments:**
  - **Demands for a healthy software and computing program are increasing**
  - **Need scientists to be engaged in R&D and to develop and maintain a computing infrastructure that is useful and adapted to the scientific data workflow needs of experiments**
    - **Requires labs and universities to strengthen partnerships on software engineering and support**
  - **Continue to encourage particle physics community to develop a clear technical vision of how to address technology issues**
    - **How to make effective use of new hardware, scale data management capabilities, …**
  - **Encourage community to engage more and across frontiers in taking next steps**
    - **HEP-CCE should bring the relevant parties together in more areas, work with ongoing complementary community efforts, and enable coordination and strong collaboration internationally and across agencies**
- **COV Recommendation 18: Include planning for computing and software development into the planning for new initiatives.**
  - **HEP is now addressing these comments and recommendations**

Excellent LHC Performance and Near-term Challenges

- Excellent LHC performance in 2016 (Run 2)
  - Unprecedented peak instantaneous luminosity >40% beyond LHC design!
  - Data accumulation ~60% beyond 25 fb^{-1} goal for 2016!



- Congratulations to the CERN accelerator team for the hard work in operating the LHC, and to the experiments for the high performance efficiency in acquiring data!

- Immediate challenges for additional Run 2 computing resources being coordinated with experiments and CERN via the U.S. LHC Ops program
  - Plan to address at the DOE-NSF U.S. LHC Operations Program Review, January 24-27, 2017 *(this week at UC Irvine…)*

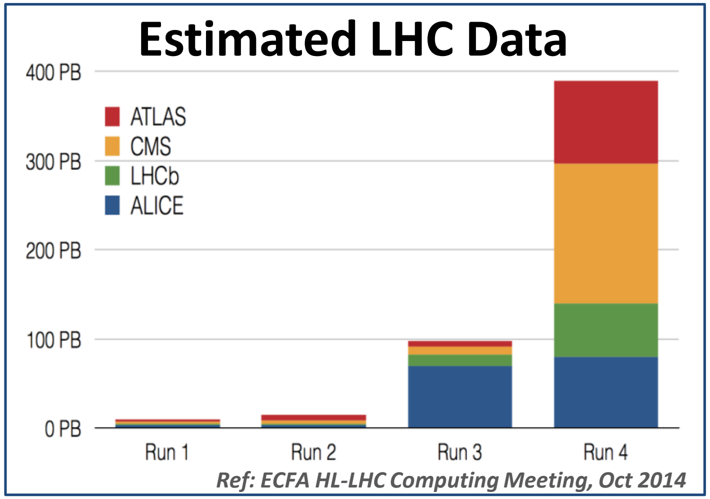
Projected Shortfalls in HEP Computing Resources

- **Expected to be 10x – 100x shortfall by 2025**
  - Shortfalls
    - Data movement – needing smart networks
    - Hardware for simulations, data analysis, and storage
    - Workforce, expertise & training
  - Computing Ecosystem critical to workflows and results
  - Challenging to handle entirely within HEP resources and HEP subprograms
  - Need a shift in strategy to best prepare for future while managing current operations and using resources external to HEP

- **Resources that could serve as a basis for future studies:**
  - SNOWMASS Computing Frontier Reports
  - 2013 Computing Planning Meeting
  - FCE (CCE) Working Group Reports
  - HEP-ASCR Exascale Requirements Report
  - European initiatives

Options for Future Computing

- **Example: LHC Run 4 (2026 and beyond) will start the exabyte era for HEP!**
  - How will the data be processed and analyzed?



☐ **Buy facilities**
- ✓ Pro: Own it! No impediment to running at full capacity when needed
- × Con: Must invest for peak utilization even if not used

☐ **Use services from other providers**
- ✓ Pro: Others make capital investments
- × Con: Will usage be available/affordable when needed?

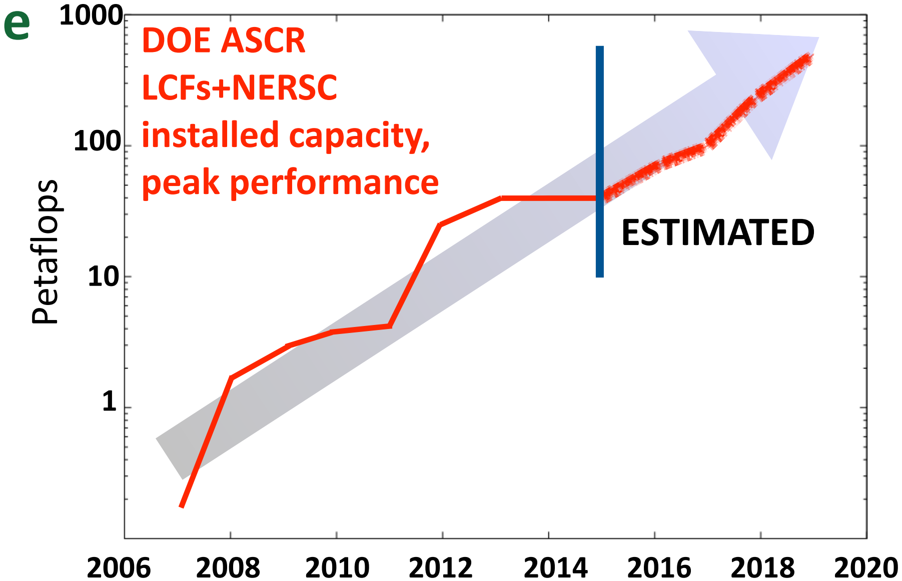
- **Current computing model provides examples of infrastructure not owned by HEP**
  - Like ESNet, NERSC, commercial clouds; not necessary for HEP to purchase all hardware

☑ **Hybrid model**
- ✓ Own baseline resources that will be used at full capacity
  - Reliable cycles available for reconstruction, MC generation, etc.
- ✓ Use service providers for peak cycles when needed
  - Conference analysis season, special collaboration needs, etc.
- ? Community and agencies exploring this approach, *but* future cost model uncertain

- **To achieve P5 global vision, all partners need to bring in their available resources!**

DOE Computing Resources Outside HEP

- DOE's major computing resource is the through the Advanced Scientific Computing Research (ASCR) program
- ESNET provides valuable network resources for science
  - Initiation of the ESNet TransAtlantic networking is a successful example of ASCR-HEP partnership



Peak performance as a function of time: approach to exascale by LCF/NERSC systems

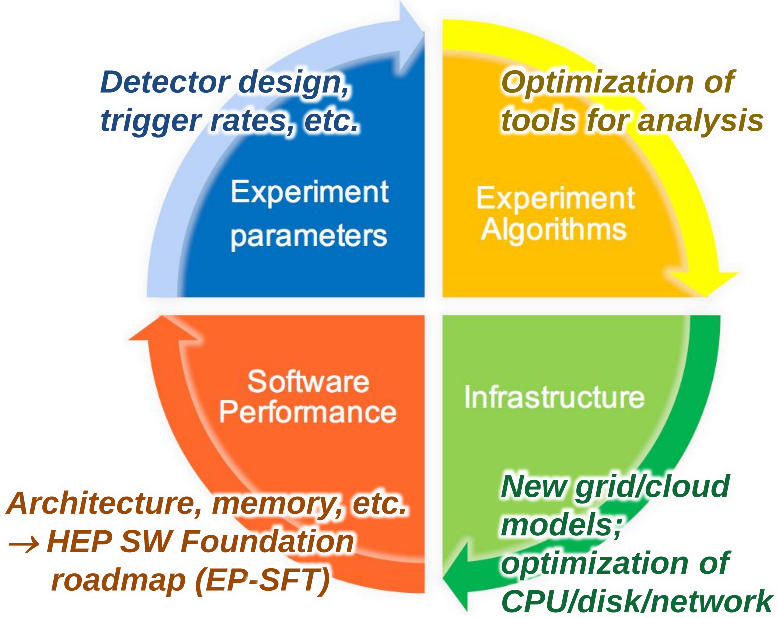
- Through National Energy Research Scientific Computing Center (NERSC), ASCR provides reliable HPC resources
  - ✓ Use of NERSC resources will exist for "free"
  - ✓ Any additional next-generation hardware will become available for HEP to buy
  - ✗ Con: HEP must port code to work in HPC environment!
- HEP is working collaboratively across the program to optimize use of DOE resources, including HEP cloud, through CCE and other efforts

HEP Computing Future Challenges

- **Future computing will intertwine different computing paradigms**
  - **High Throughput Computing (HTC)**
    - Increasing HEP experiment demand may outpace Grid computing resources
    - **Need**: New hardware/software exploits
  - **High Performance Computing (HPC)**
    - Classic use of HPC resources by theorists
    - **Need**: Event services for simulations and dedicated front-ends for job packaging
  - **Data-Intensive Scalable Computing (DISC)**
    - Analysis of simulations & comparison to observational data without HTC lead times
    - **Need**: True interactive largescale computing
- **Challenge will be to adapt the HEP computing model to optimize operations and analysis workflow to exploit all resources**
  - **New paradigms are on the horizon in computing that HEP should also pay close attention to:**
    - Deep learning
    - Smart networking
    - National Strategic Computing Initiative → Neuromorphic, Quantum computing
  - **Directed computing investments in partnership with ASCR are needed to meet future needs**
    - ASCR welcomes partnerships with HEP through CCE

Framework for Future HEP Computing

- **HEP aims to optimally leverage DOE resources in developing future computing solutions to meet our mission needs**
  - In particular, aim to leverage our National Laboratories and ASCR resources
- **White House decision for Exascale has led to ASCR's work on hardware design**
  - Community has opportunity to work with ASCR to ensure HW is useful for future needs
  - Optimized HW then becomes widely available and lower cost to HEP
  - Timescale for this change well-matched to HEP needs (early 2020s)
- **Framework of mission-driven model for future HEP computing:**
  - Hardware
    - Develop centralized resources that can be efficiently and effectively used by large portions of the community
  - Data storage
    - Develop options to consolidate, in partnership with Laboratories and ASCR
  - Software frameworks and analysis tools
    - Develop shared solutions within and across frontiers, including those coordinated through CCE and the labs
  - Computing for successful experiment operations
    - Assess computing needs early enough to help inform experiments' project and operations planning



Community Input Useful to DOE HEP

- The most useful community input for DOE HEP in developing plans for future computing, at this stage, is the anticipated requirements for future experimental operations
  - Resources necessary for operations and analysis of each experiment:
    - Data storage (disk, tape, …)
    - Networking
    - Computing cycles
    - Porting and software development
  - Models that link experiment performance to computing resources needed
    - Increased instantaneous luminosity and pile-up at the LHC
    - Extended data-taking
    - Alternative computational operations schemes

- Understanding such requirements for all experiments, across frontiers, is necessary to develop an optimal plan for future HEP computing resources for operations into 2026 and beyond

U.S. DEPARTMENT OF ENERGY

Office of Science

Demo on Google Cloud

- Generated > 1 Billion events in 48 hours during the 2016 International Conference for High Performance Computing, Networking, Storage and Analysis
- Doubled the size of global CMS computing resources during these two days

Relevant National Events External to HEP

- **National Strategic Computing Initiative (NSCI) Executive Order 13702, signed by U.S. President on July 29, 2015**

  - **DOE is one of the lead agencies executing the mission – with ASCR and NNSA primarily responsible – Exascale Computing Project**
  - **Three HEP groups have earned ASCR funded code development projects:**
    - **Cosmic Frontier codes**
    - **Plasma acceleration codes**
    - **Lattice QCD software**
- **Many of you participated in HEP-ASCR Exascale Requirements Review**
  - **Advantages to HEP:**
    - **Use of powerful computing capabilities**
    - **Exploring new hardware for HEP facilities**
    - **Develop advanced computing ecosystem:**
      - **Data – data movement – networks – hardware – software**
  - **Report available at: *https://arxiv.org/abs/1603.09303***

U.S. LHC Detector Operations

- Successful interagency partnership between DOE and NSF to support the U.S. ATLAS and U.S. CMS Detector Operations program
- Objectives of the Operations program:
  - Operate and maintain U.S. built detectors or detector components
  - Meet U.S. CMS and U.S. ATLAS M&O common fund costs
  - Tier-1s (DOE) and Tier-2s (NSF) computing facilities; U.S.-CERN LHC Trans-Atlantic Network (DOE)
  - Enable physics analysis by U.S. physicists on CMS and ATLAS by providing computing hardware and core software in direct support for all phases of analysis
  - Upon completion of fabrication of U.S.-built detector components, and delivery to CERN, of the initial [Phase-1] CMS and ATLAS detector upgrades, complete installation and commissioning activities for each international collaboration
  - For NSF: conduct project planning and R&D activities leading to a construction-ready proposal (MREFC) for the HL-LHC CMS and ATLAS detector upgrades