Software and Computing Highlights at BNL

Torre Wenaus (BNL NPPS), Ofer Rind (BNL SDCC), and contributions from many

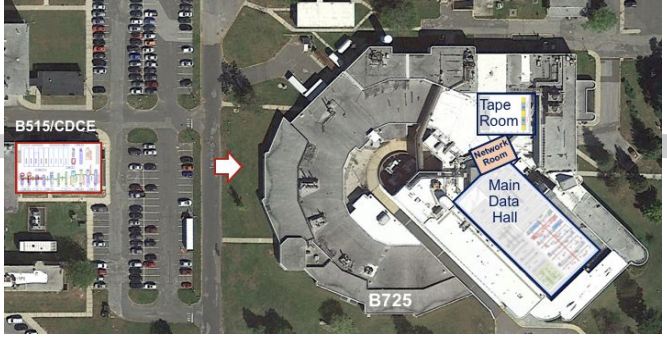
Software and Computing Round Table
December 7 2021

@BrookhavenLab

This talk

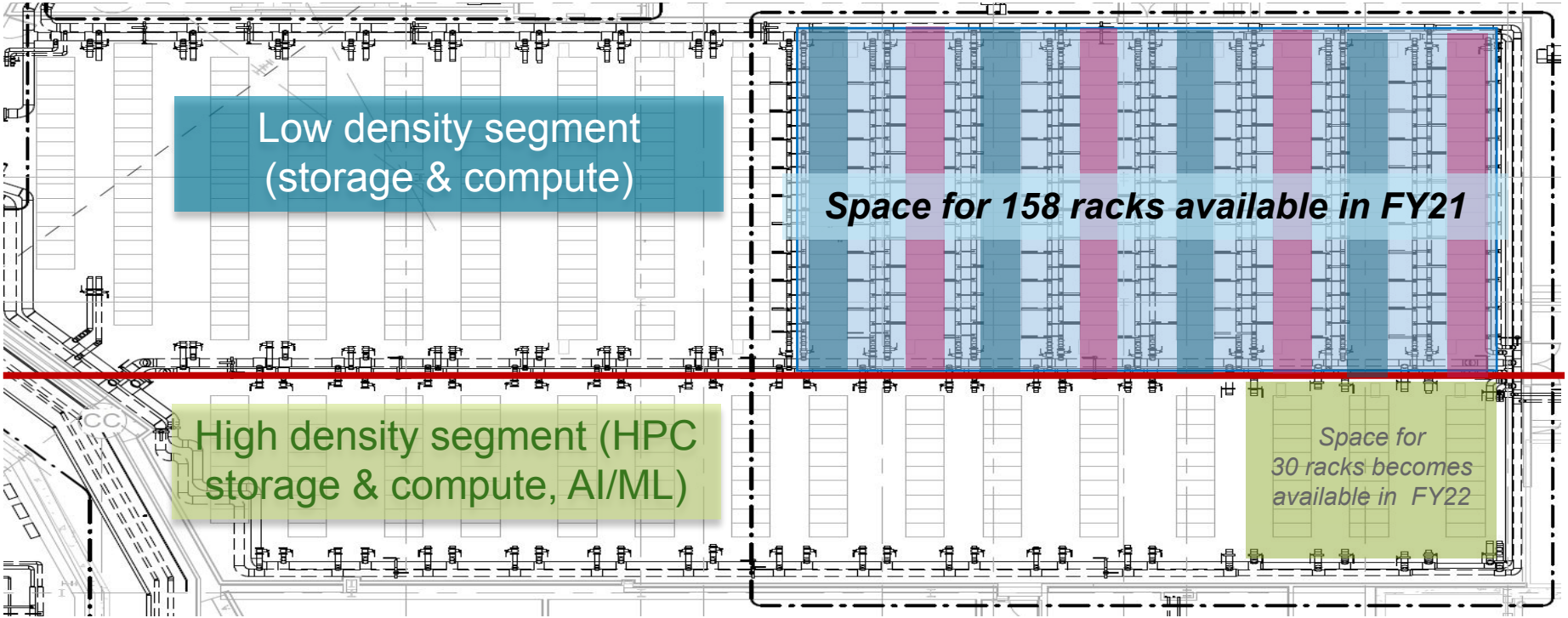
- Scope is experimental nuclear and particle physics in the BNL physics department
- Caveat:
  - Selective and incomplete
  - With more of an emphasis on software
- Nonetheless, still too many slides :-) (but about 20 fewer than last year! ;-)
  - Will talk over them rather than deeply into them
  - With apologies, some in supplementary, especially more steady state activities
- No time for background on the experiments, will assume they are familiar
- Outline…
  - Computing facilities: SDCC
  - Software: NPPS
  - NP experiments
  - HEP experiments
  - AI/ML
  - Plans and priorities

Scientific Data and Computing Center



- New B725 data center became operational in 2021
  - Part of BNL core revitalization project
- Main Data Hall for storage and compute resources
  - Can support 478 racks with up to 9.6 MW of IT load
- Initially supports 188 rack positions in the Main Data Hall
  - 158 rack positions with 2.4 MW of power/cooling are available since late FY21, expected to get 35 racks populated by the end of 2021
  - 30 high density rack positions with 900 kW of power/cooling are to become available later in FY22
  - 1.2 MW of diesel generator backup power, to be increased to 2.4 MW before the end of CY21
- Dedicated Network Room hosting all the central network equipment fully deployed
- Dedicated Tape Room capable of supporting up to six 8-frame IBM TS4500 libraries
  - 2x 8-frame IBM TS 4500 libraries installed

SDCC's new Main Data Hall



Distributed and Central Disk Storage

More than **75 PB** data served on distributed disk storage managed by **dCache** and **XRootD**

Major 2021 updates

- sPHENIX dCache now in production, as for ATLAS, PHENIX, Belle II, and DUNE
- ATLAS and DUNE dCache now running latest golden release with auth token capabilities (ready for SciToken testing, etc.)
- dCache v7.3 testbed deployed for data Quality of Service (QoS) development/testing
- STAR XRootD write capability enabled on central storage portion

16 PB, > **2 billion files** now accessible on centralized **GPFS** and **Lustre** file systems

Major 2021 updates

- Deployed major GPFS update motivated by security vulnerabilities
- Lustre footprint expansion to 7 PB (v2.12.7)
- Deployed new MinIO S3 endpoints for EIC and Center for Functional Nanomaterials

dCache.org

XRootD

lustre

MINIO

Tape Storage (HPSS)

More than **180 PB** now stored on tape at SDCC and managed by HPSS
Over **20 PB** written, **175 PB** read in 2021

In 2021:

- Major HPSS upgrade (v7.4 to 8.3) with new file grouping capabilities
- Installed two 10K-slot IBM TS4500 tape libraries, for ATLAS production in 2021
- New LTO8 tape technology deployed for Lattice QCD group
    - STAR and ATLAS moving to LTO8 late 2021
    - PHENIX and Archive will remain on LTO6
- Soon TBD: final library configuration and tape technology for sPHENIX

Data Carousel and WLCG Tape Challenge

- SDCC continues to play a leading role in the ATLAS Data Carousel project, as well as the 2021 WLCG Tape Challenge
  - Close collaboration between SDCC and NPPS on data carousel
- Almost all ATLAS workflows (MC simulation, data reprocessing, derivation) run in Data Carousel mode now. Effort continues to optimize utilization of tape resources.
  - New HPSS features have been deployed at SDCC that will better group ATLAS files on tape (by dataset) for better staging efficiency.
- WLCG Tape Challenge was conducted this October, with all four LHC experiments and WLCG Tier1 sites participating. First of its kind.
  - Both ATLAS and BNL had a good run, reaching the target ATLAS Run3 tape read and write rates.

SDCC tape traffic during Tape Challenge

WLCG Network Data Challenge status

2021 target: 10% of HL-LHC in 2027 and the highest rate for special periods of Run-3

Achieved in the last period:

In: **50 Gb/s**

With **25 Gb/s from CERN**

Out: **60 Gb/s**

with **20 Gb/s to Tier1s**

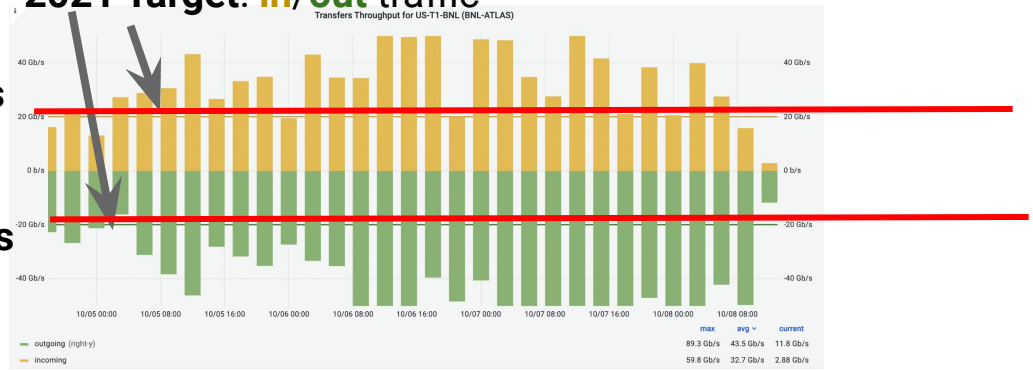
40 Gb/s to Tier2s

Transfer throughput for BNL (all activities)

2021 Target: **in**/**out** traffic

20 Gb/s

20 Gb/s



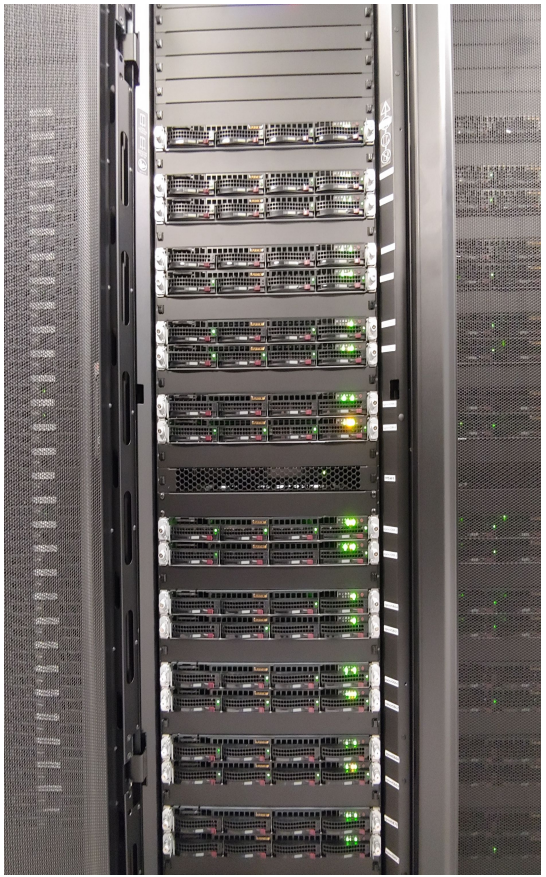
The target has been reached with no special storage and network issues

High Throughput Computing

SDCC operates **~2,000** compute nodes providing **~80,000** logical cores

In 2021:

- Taking delivery of another seven racks with **210** Supermicro Xeon Cascade Lake servers (**20K+** cores)
- Update of HTCondor batch to v9 with new support for token authentication
- Actively evaluating RHEL8 and RHEL8-based distributions including Rocky Linux 8, due to discontinuation of Scientific Linux and early end of life for CentOS 8 at end of 2021



High Performance Computing

Currently supporting 5 HPC clusters. All allocations subject to approval.

Institutional Cluster (IC): 216 HP XL190r Gen9 nodes with EDR IB
Skylake Cluster: 64 Dell PowerEdge R640 nodes with EDR IB
KNL Cluster: 142 KOI S7200AP nodes with dual rail Omnipath Gen.1 interconnect
ML Cluster: 5 HP XL270d Gen10 nodes with EDR IB
NSLS2 Cluster: 30 Supermicro nodes with EDR IB

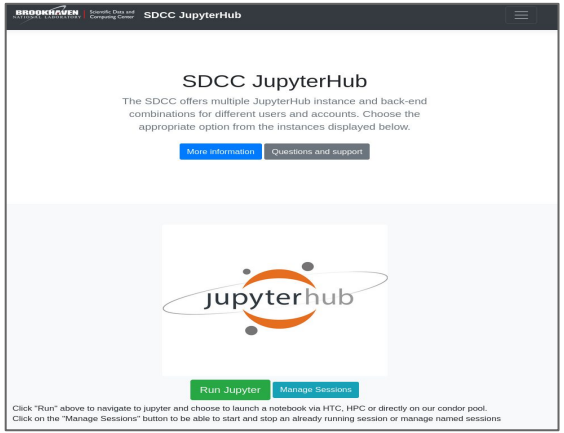
Evaluating test units for **next generation institutional cluster** with the following specs:

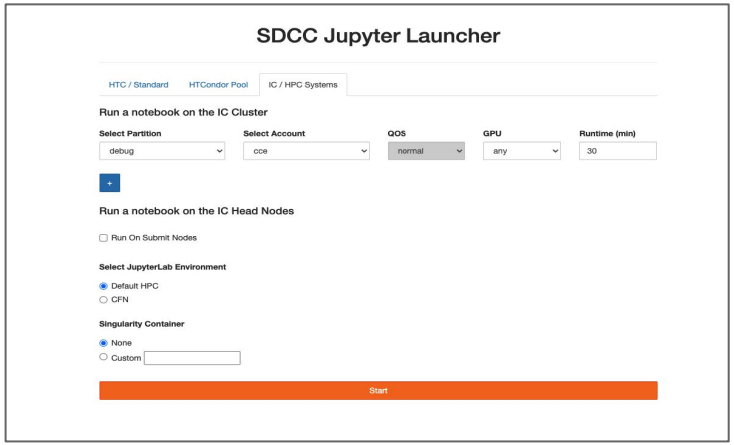
- 2x Intel Xeon (Ice Lake)
- 512GB DDR4-3200
- HDR InfiniBand interconnect (200GbE)
- 2x Nvidia A100 (40GB and/or 80GB)

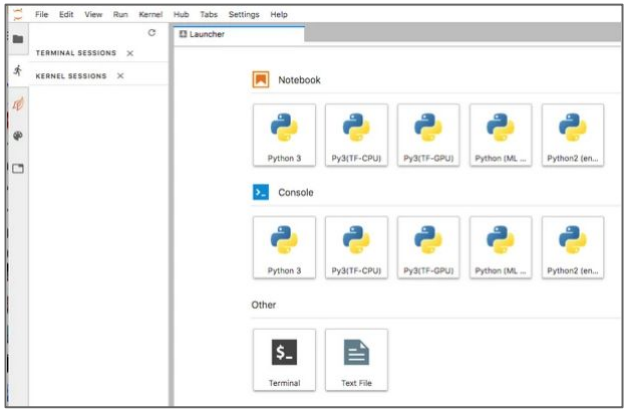
NSLS2 Cluster

Jupyter

- **Jupyter service available at SDCC**
  https://jupyter.sdcc.bnl.gov
  - Web browser-based interactive analysis
  - Interface for using our HPC and HTC resources
    - Can run notebooks on dedicated VMs, and via HTCondor or SLURM batch
    - Dask extension enabled
  - Externally available to all users with SDCC accounts
  - Updated JupyterHub to v1.4
  - Experiments can maintain their own kernels

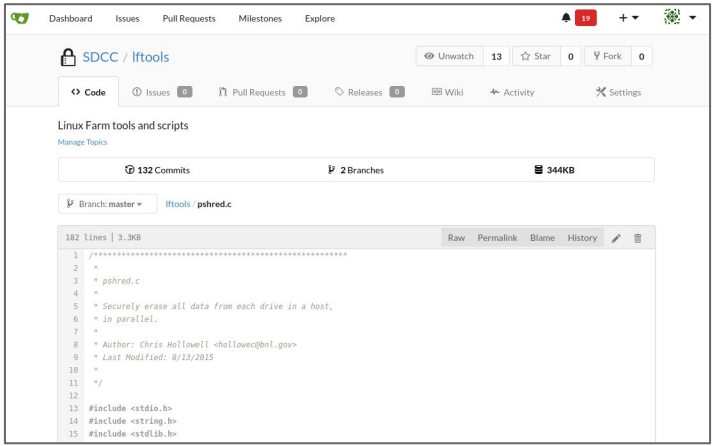






Cloud-based tools & services

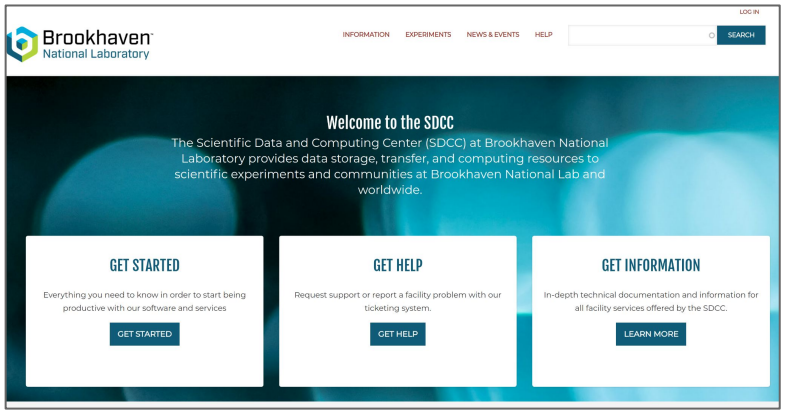
- Overleaf LaTeX editor service added this year
  - Available to collaborative activities with a BNL research program
  - A BNL computer account or appointment is not required to obtain a license
  - More info at https://www.overleaf.com/org/bnl
- Jira
  - SDCC holds licenses for NPP users for Atlassian Jira Work Management (formerly Jira Core) project & task management service
  - Currently hosting 55 projects with 77 users
  - To request an account: RT-RACF-UserAccounts@bnl.gov
- Gitea
  - Self-service git service with shared/private repo support
  - Wikis, issue tracking, continuous integration (CI) hooks



Website Developments

- **SDCC website improvements include**
  - Improved search results based on user requests and analytics
  - Remodelled site layout and navigation to improve user experience
  - Content improvements with better guidance on SDCC services

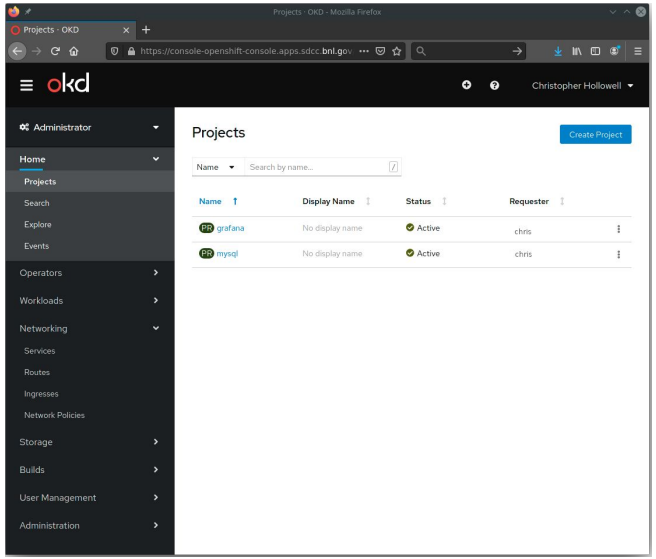
- **Sites currently being maintained and developed by the SDCC**
  - **Quantum Astronomy** - quantastro.bnl.gov
  - **Cosmology & Astrophysics** - in development
  - **sPHENIX** - sphenix.bnl.gov
  - **US ATLAS** - in development
  - **LuSEE at Night** - in development



OKD (Kubernetes) at SDCC

- **OKD provides a platform for container orchestration, similar to Kubernetes (k8s)**
  - Community-supported release of Openshift
  - Allows for simplified deployment of services via helm charts and Openshift templates
  - Contains numerous security enhancements out of the box vs k8s
    - Users are never root by default

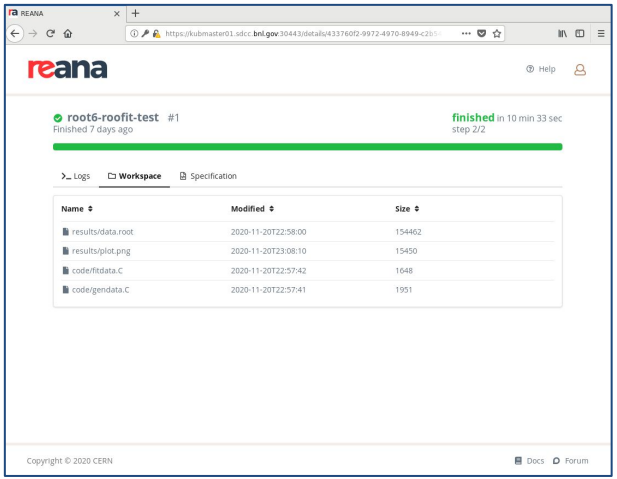
- **Test/development OKD 4.7 cluster online at SDCC**
  - A number of VMs and two repurposed latest-generation farm nodes
  - Currently only accessible from inside BNL
  - Test users manage projects and service deployments from our interactive compute nodes



- **Production cluster hardware purchased for sPHENIX**
  - Expect delivery soon

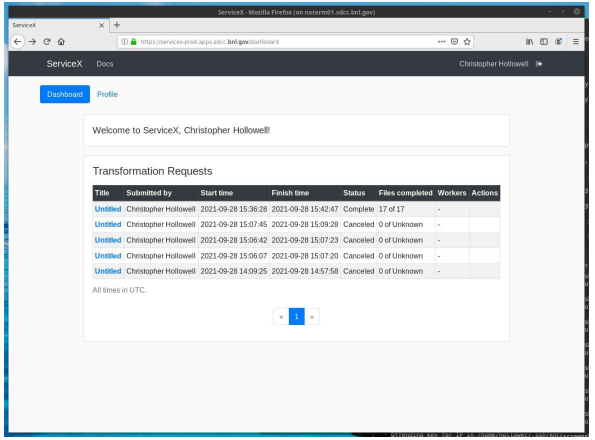
REANA at SDCC

- REANA testbed deployed at SDCC
- A platform for reproducible scientific data analysis
  - https://www.reanahub.io/
  - Primarily being developed by CERN, currently V0.8.0
  - May be an important part of future HEP analysis facilities
- Being used by a number of test users
  - Working with NPPS for PHENIX and PanDA use
  - Modified to support PHENIX CVMFS access
  - Used by PHENIX workshop attendees in June
- Web interface currently accessible via SSH tunnel/SOCKS proxy
- Working on tying accounts/auth to our LDAP/K5/IDP
- Can interface and submit container jobs to SLURM on the institutional cluster
- We can provide instructions to anyone interested in trying it



ServiceX at SDCC

- ServiceX test deployment at SDCC
- Columnar data delivery and pre-processing service being developed by IRIS-HEP, ATLAS, CMS
- May also be an important part of future HEP analysis facilities
- Deployed in our test OKD cluster
  - Currently only available from within the SDCC network
  - Contact us if you are interested in testing
- Working with developers to support OKD upstream




```
#!/usr/bin/env python3
from func_adl_servicex import ServiceXSourceXAOD

dataset_name =
"mc15_13TeV:mc15_13TeV.361106.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Zee\
.merge.DAOD_STDM3.e3601_s2576_s2132_r6630_r6264_p2363_tid05630052_00"
src = ServiceXSourceXAOD(dataset_name)
df = src \
    .SelectMany('lambda e: e.Jets("AntiKt4EMTopoJets")') \
    .Select('lambda j: j.pt()/1000.0') \
    .AsPandasDF('JetPt') \
    .value()
printf(df)
```

Read/filter ATLAS xAOD into columnar pandas format

```
          JetPt
0         36.319766
1         34.331914
2         16.590844
3         11.389335
4          9.441805
...             ...
857133     6.211655
857134    47.653145
857135    32.738951
857136     6.260789
857137     5.394783

[11355980 rows x 1 columns]
```

Nuclear and Particle Physics Software (NPPS) Group

The Nuclear and Particle Physics Software (NPPS) Group consolidates much (not all) of the NPP software development in the Physics Department. See https://npps.bnl.gov

- Currently 23 NP and HEP members working on
  - NP: EIC, PHENIX, sPHENIX, STAR
  - HEP: ATLAS/HL-LHC, Belle II, DESC, DUNE, Rubin Observatory, soon LuSEE at Night
- Group mandate and emphasis: cross-experiment common efforts across HEP and NP
  - Shared personnel, expertise, software
  - 15 group members are working on more than one experiment
- Many collaborations
  - SDCC/CSI at BNL, WLCG (LHC), HSF, IRIS-HEP (NSF), HEP-CCE (DOE)
- Many BNL NP and HEP software activities beyond NPPS
  - In the STAR, (s)PHENIX, EDG (neutrino physics), Omega (ATLAS), Cosmology & astrophysics, theory and other groups

Collaborations

Stony Brook University

Argonne NATIONAL LABORATORY

HSF
HEP Software Foundation

Jefferson Lab

aws

Google Cloud

iris hep
Institute for Research and Innovation in Software for High Energy Physics

BERKELEY LAB

THE UNIVERSITY OF TEXAS AT ARLINGTON

Fermilab

SCIENTIFIC DATA MANAGEMENT
RUCIO

OAK RIDGE National Laboratory

ROOT
Data Analysis Framework

XRootD

BROOKHAVEN NATIONAL LABORATORY *Computational Science Initiative*

CMS

DIRAC
THE INTERWARE

CERN

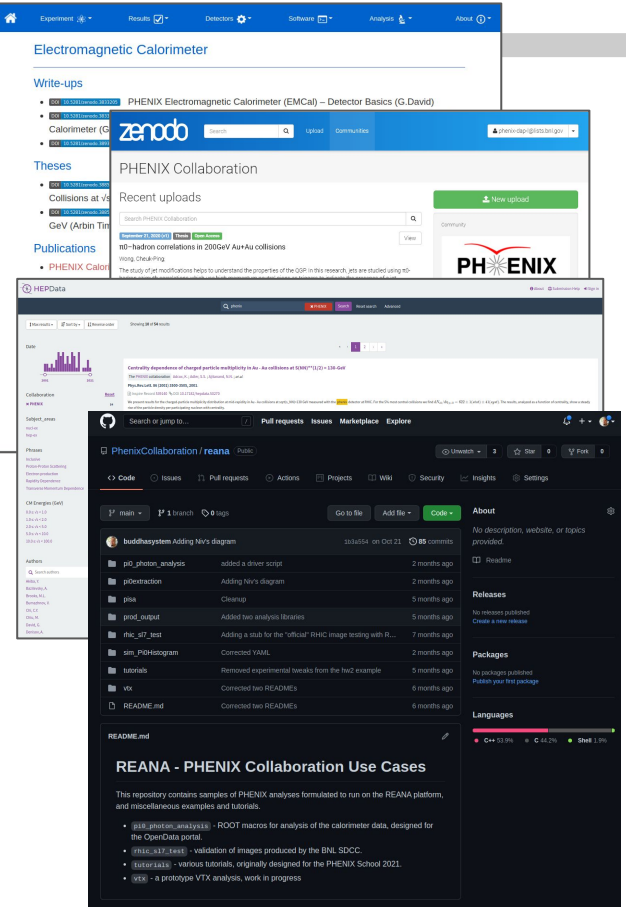
CERN openlab

WLCG
Worldwide LHC Computing Grid

SLAC

Brookhaven National Laboratory

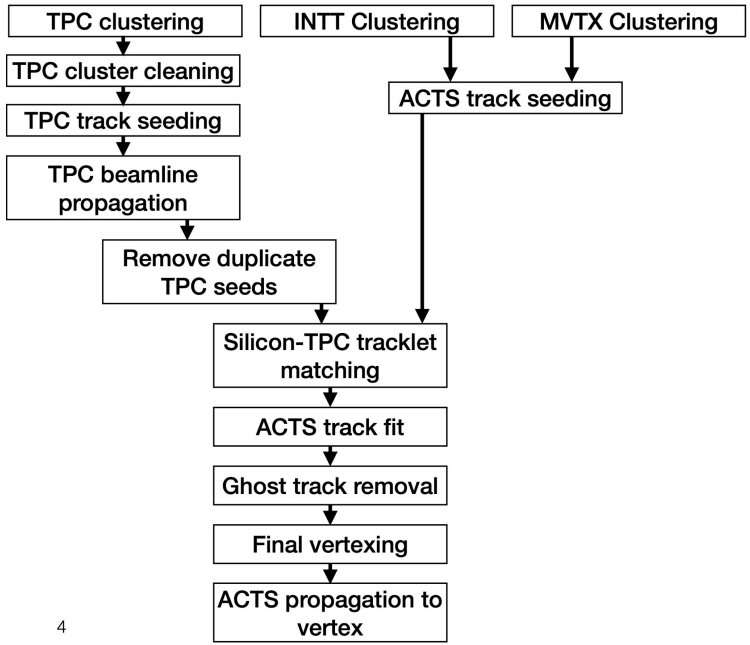
PHENIX: Data and analysis preservation (DAP)



- Leveraging community (especially CERN) practices and tools
- **GitHub** - Web development, code preservation, documents etc
  - New GitHub/Jekyll site aggregates info from many sources
- **Zenodo** - Document preservation service
  - About 500 (so far) theses, publications, tutorials, … collected, curated, tagged with metadata/keywords
- **HEPData, OpenData** - Research data preservation
  - >50 HEPData entries now.
- **Docker** - Capture and preserve the PHENIX software environment
  - PHENIX environment preserved as Docker images and packages deployed on CVMFS
- **REANA** - Platform to preserve reproducible analysis
  - Realistic elements of analyses have been run on the REANA platform
  - Working on porting a complete analysis
  - In the future, benefit from ATLAS work on PanDA/REANA integration
- HEPData, REANA tutorials held during the PHENIX School in 2021

sPHENIX: ACTS tracking

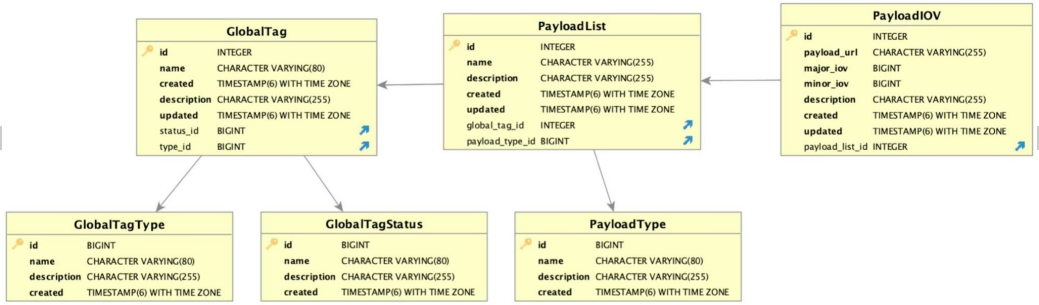
- **Great progress** in developing ACTS based tracking, now the official baseline
- **Meeting targets in resolution, timing, memory**
  - still with room for improvement
- Improved reco for realistically smeared event vertex
- Improved capability for secondary track reco
- **Second mock data challenge** over next few months
  - aim for realistic reco for analyzers
- Much remaining to do
  - **integrate realistic TPC space charge distortions** in standard simu/reco
  - implement iterative reco
  - much calibration, alignment work to do
  - ...



4

sPHENIX: Conditions DB

- **Evolved the Belle II CDB we're responsible for into experiment agnostic CDB**
- Updated backend: Java Tomcat => Django
- Streamlined schema based on global tags aggregating payloads with intervals of validity
- REST-like web service API, json format
- Like the Belle II CDB, architecture follows (more closely) that of the HSF CDB white paper
- Implementing a C++ CDB library wrapping the CDB interactions for use in the sPHENIX framework
- Like the Belle II CDB, collaborative with SDCC
  - NPPS software deployed on SDCC OKD
- **For the future: bring the generic CDB to HSF as a reference implementation?**



Get example

```
GET /api/cdb/payloadiovs/?gtName=ARICHdata&majorIOV=3&minorIOV=3

HTTP 200 OK
Allow: GET, HEAD, OPTIONS
Content-Type: application/json
Vary: Accept

[
    {
        "id": 8246,
        "name": "ARICHModuleTest",
        "global_tag": {
            "id": 106,
            "name": "ARICHdata",
            "description": "",
            "created": "2021-08-27T10:39:03.282104",
            "updated": "2021-08-27T10:39:03.282180",
            "status": 3,
            "type": 1
        },
        "payload_type": {
            "id": 1,
            "name": "Type1",
            "description": "",
            "created": "2021-08-04T13:54:47.835407"
        },
        "payload_iov": [
            {
                "id": 792659,
                "payload_url": "dbstore/ARICHModuleTest/dbstore_ARICHModuleTest_rev_3.root",
                "major_iov": 0,
                "minor_iov": 2,
                "payload_list": 8246,
                "created": "2021-08-27T10:39:04.370853"
            }
        ],
        "created": "2021-08-27T10:39:03.741195"
    },
```

Payload URL, major and minor IOVs

Dubbed the No Payload DB (NPDB), IOV based requests return payload locations (URLs), not payloads themselves

Enables effective payload caching

Intelligent use of "pages" to use cacheable queries

9

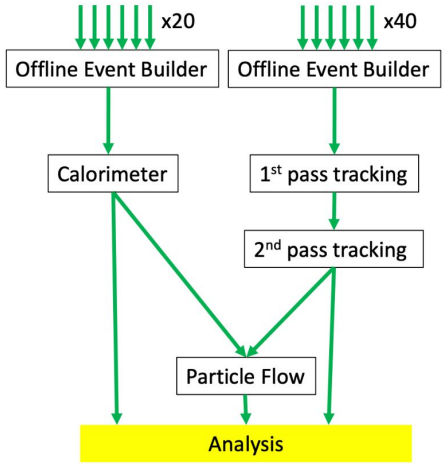
sPHENIX: Distributed computing

- sPHENIX has adopted **PanDA** & its ecosystem for workflow/workload management, and **Rucio** for data management, both services to be BNL SDCC based
- **PanDA:** new BNL service being implemented now with SDCC
  - Replacing Oracle in the back end with PostgreSQL
- **Rucio:** SDCC is familiar with running Rucio services from Belle II and will run one for sPHENIX
  - Leverage Rucio expertise in both NPPS and SDCC
- sPHENIX will make use of PanDA's new support for workflows directed at efficient resource utilization and flexible DAG-described workflows
  - **data carousel** to efficiently process primarily tape-resident data
  - new 'pchain' command line tool to run **arbitrarily complex workflows** can be an easy mechanism for setting up sPHENIX workflows

Production Workflow

0. Event building + Calibrations (partly run in sPHENIX counting house)
1. Event building (20/40 input streams, 1 output streams [each])
   - Tracking Detectors
   - Calorimeters
2. Calorimeter
3. 1st pass tracking
4. 2nd pass tracking (time lag between 1st and 2nd pass)
5. Particle Flow (combines tracks & calorimeter data)

6. Simulations on the Grid

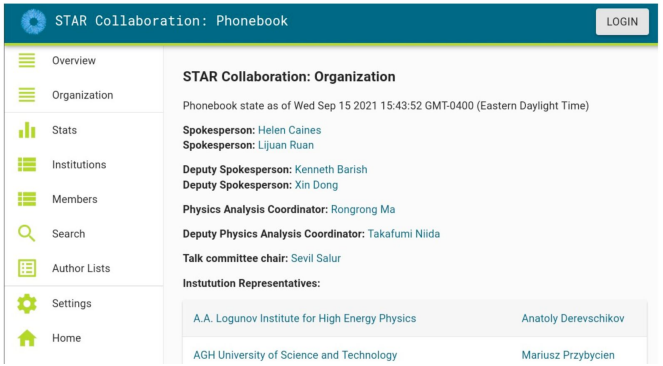


T. Maeno, S. Padolski, A. Jaikar et al

STAR: Databases & collaborative tools

- NPPS develops and maintains for STAR and RHIC:
  - **Databases & online tools:** Online conditions, Offline calibrations, FileCatalog, Online meta-data collector (MIRA), Document System, ShiftSignup, RunLog, Event Display, Experiment Status Plots
  - **Collaboration Record Keeping** (aka PhoneBook): In-house tools developed because of a lack of off-the-shelf software suitable to scientific collaborations
    - An integrated suite with uniform accounts, access, data
    - The past year: **reengineering** to address wider interest, e.g. sPHENIX
      - Updated technology stack: React, GraphQL, Material-UI

https://phonebook.sdcc.bnl.gov/demo/



Material-UI

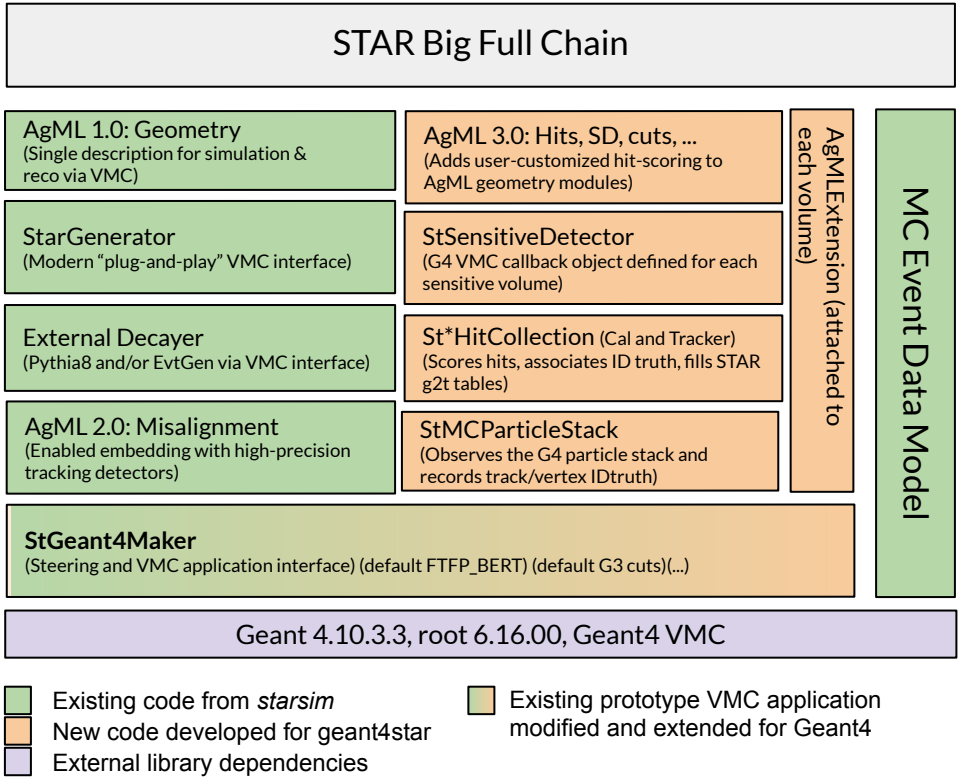
React

Live Data

Dmitry Arkhipkin

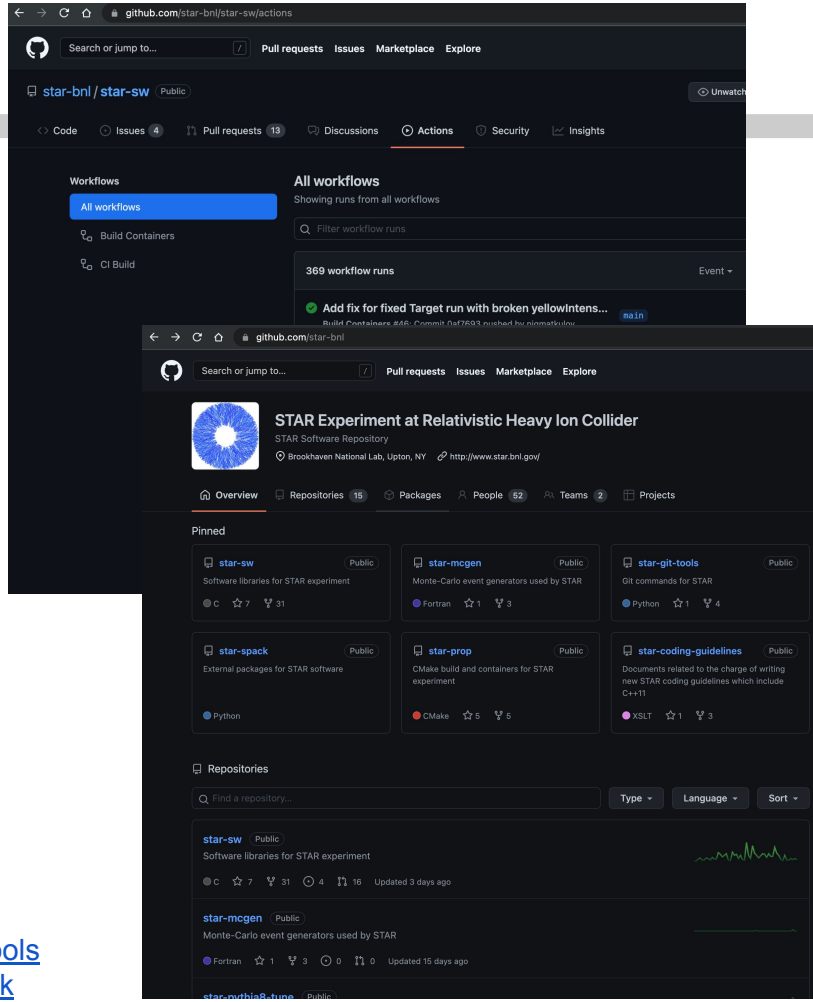
STAR: geant4star

- geant4star was developed to support the **forward upgrade physics program** of the STAR experiment
  - leveraging **Geant4's improved hadronic physics** models over GEANT3
- Comprehensive CI test suite to compare/validate against production simu
- First full release Oct 2021
  - Geant4.10.3.3, ROOT 6.16.00
- Focus now is on physics validation, starting with calorimetry



STAR: Modernized infrastructure

- Migrated from CVS to git and **GitHub**
- Designated code maintainers, code review/status checks prior to merge
- CI for STAR repo set up using **GitHub Actions**
  - code is compiled on every pull request
- CI build performed in a **Docker** container with self-contained STAR environment & dependencies
- Dependencies installed via **Spack**

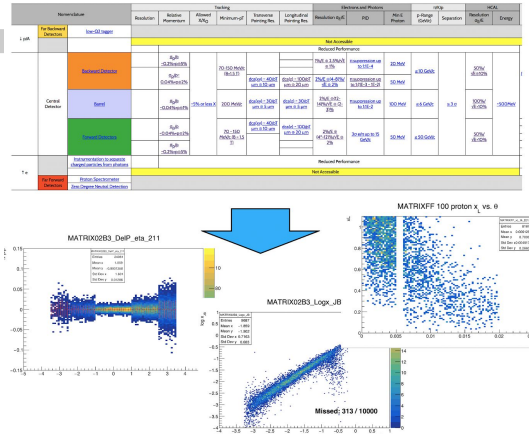


https://github.com/star-bnl/star-sw
https://github.com/star-bnl/star-git-tools
https://github.com/star-bnl/star-spack

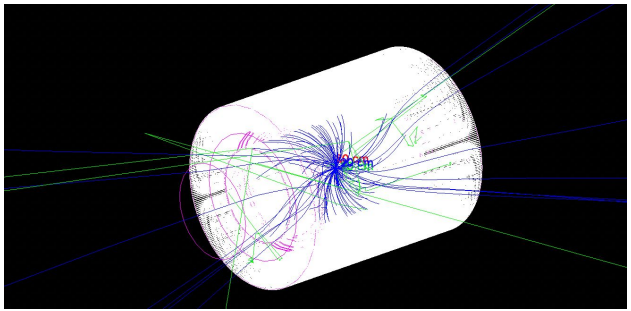
EIC: Simulation

- NPPS responsible for **eic-smear**, the workhorse **fast simulation** for EIC
- With EIC now in the (proto)collaboration phase, developing detailed detector designs, **full simulation becoming more important**
- EIC software working group is developing a common Geant4 simulation toolkit for EIC
  - the **eAST** (eA Simulation Toolkit) project
  - EIC-directed, detector-agnostic simu toolkit using latest Geant4 (soon V11), led by one of the fathers of Geant4, Makoto Asai
  - **Aim: common denominator simu toolkit 2022+**
  - Proto-collaborations well engaged in the project
  - NPPS contributions:
    - adapting proto-collaboration simulations to eAST
    - prototyping integrated fast/full simu in eAST
      - an area where eAST can add unique value
    - software/documentation infrastructure

eic-smear implementing detector capability matrix

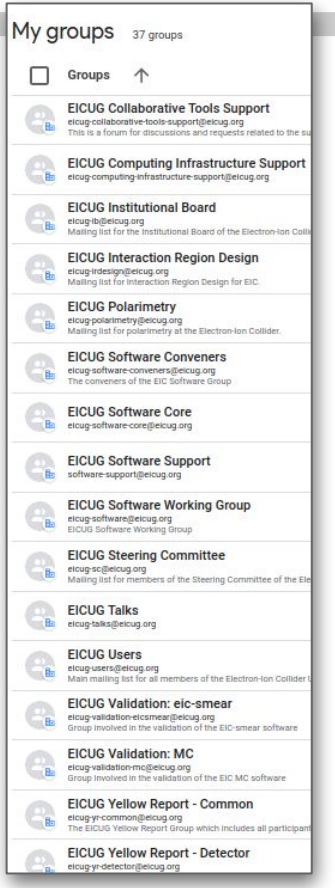


ECCE in eAST

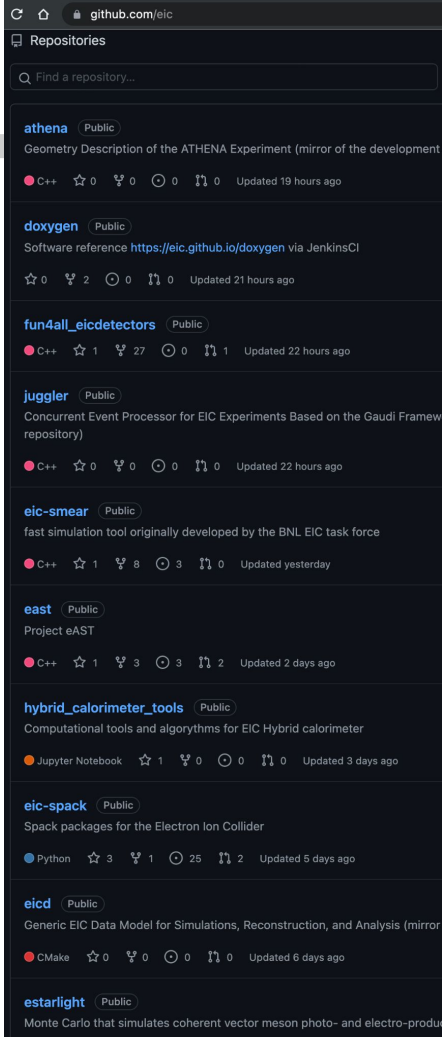


Kolja Kauder, Maxim Potekhin, Chris Pinkenburg

EIC: Infrastructure, collaboration



- EIC-wide collaborative tools developed/maintained mainly by NPPS
  - The main EICUG website
  - Google services: 37 groups for distribution of e-mail, group calendars
  - BNL-hosted: Indico, Wiki, the EICUG Phone Book
- https://github.com/eic software repositories
  - Efficient collaboration in core software, documentation, web development
  - Tools supporting developers: dOxygen, LXR code browser
- The EICUG Software Group Website https://eic.github.io/
  - Hosted on GitHub Pages, secure, fast and efficient
  - A hub for all EIC-related software, fostering teamwork and collaboration



ATLAS: Offline software

- Towards Run-3
  - Co-coordinating ATLAS software as **multi-threaded Athena** (Release 22) has entered production for Run-3, now doing a full reprocessing of Run-2 data
  - We've led the design and much of the implementation of the new **Run-3 analysis model**, including new **compact analysis formats**
  - Leading implementation of **ARM** in current releases
- Towards Run-4 HL-LHC
  - Leading a proof of concept study of **RNTuple for xAOD**
  - Co-leading the US HL-LHC S&C R&D effort directed at
    - resolving how we can effectively use **large heterogeneous HPCs**, in particular their accelerators, following DOE's mandate to do so
    - developing **AI/ML capability within Athena** (inference), and beginning to explore how event batching might boost accelerator utilization to reach significant gains
    - completing and deploying **FastChain fast simu/reco** capable of meeting 90% of Run-4 simulation needs

- Less disk and CPU usage for same data content
  - 25% smaller files, x2-5 better single-core performance
  - 10GB/s per box and 1GB/s per core sustained end-to-end throughput (compressed data to histograms)
- Native support for object stores (targeting HPC)
- Lossy compression
- Systematic use of exceptions to prevent silent I/O errors

RNTuple

Seamless transition from TTree to RNTuple

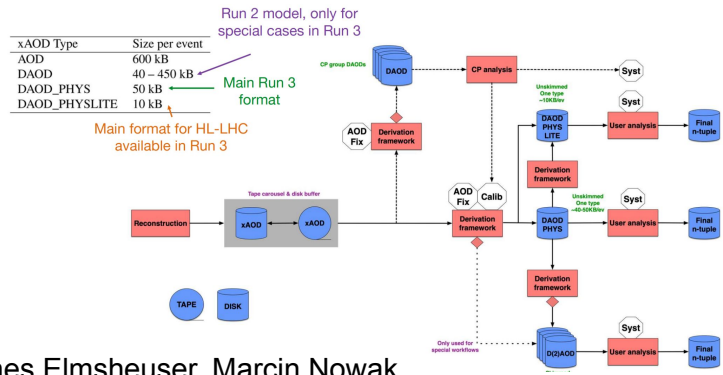
Event iteration
Reading and writing in event loops and through RDataFrame
`RNTupleDataSource, RNTupleView, RNTupleReader/Writer`

Logical layer / C++ objects
Mapping of C++ types onto columns
e.g. `std::vector<float>` ↦ index column and a value column
`RField, RNTupleModel, REntry`

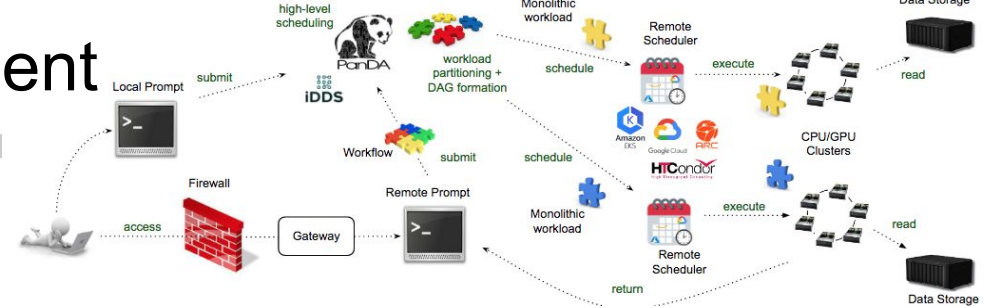
Primitives layer / simple types
"Columns" containing elements of fundamental types (`float, int, ...`) grouped into (compressed) pages and clusters
`RColumn, RColumnElement, RPage`

Storage layer / byte ranges
`RPageStorage, RCluster, RNTupleDescriptor`

Run-3 analysis model



ATLAS: Workflow management

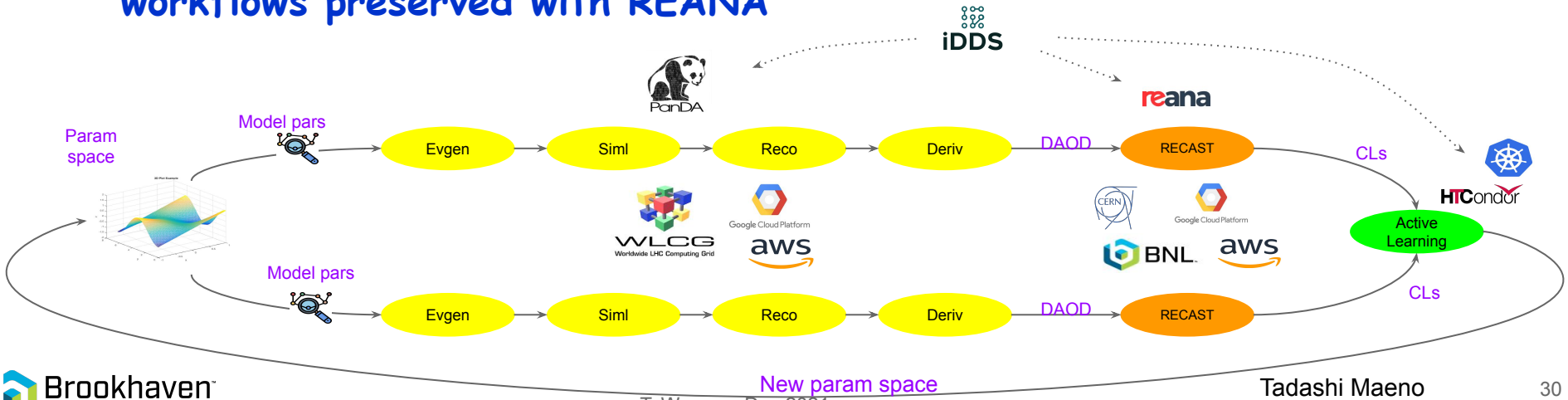


Steadily growing number of use cases and features supported in the PanDA ecosystem

- Recent focus on **analysis-directed** use cases such as **ML** (hyperparameter optimization, training), **active learning**, **pseudo-interactive analysis** (integrations with Jupyter, Dask), **analysis clouds** (Google cloud, AWS, K8s clusters), **REANA** (analysis preservation), **pchain CLI** (complex workflows at the command line)
  - iDDS (intelligent data delivery system) a major element of this, able to orchestrate workflows of arbitrary DAG-described complexity. Collaboration with IRIS-HEP
- **PanDA/iDDS based ML services** supports scaling up of AI/ML applications in ATLAS and beyond
- **BNL-led ATLAS-Google R&D collaboration** has captured the interest and imagination of key ATLAS analysis experts, working with us on innovative cloud analysis applications
- **Support for (friendly) HPCs** has had huge payoff: a huge new HPC in Europe delivered throughput on the order of the entire ATLAS grid for months
- Extensive work from the core PanDA team supporting **wider PanDA usage** (Rubin, sPHENIX, ...)
  - Major documentation upgrade, iDDS scaling, Postgres back end, modern authentication mechanisms, quasi real time log access, containerized services, ...

PanDA/REANA: One example of a novel new integration

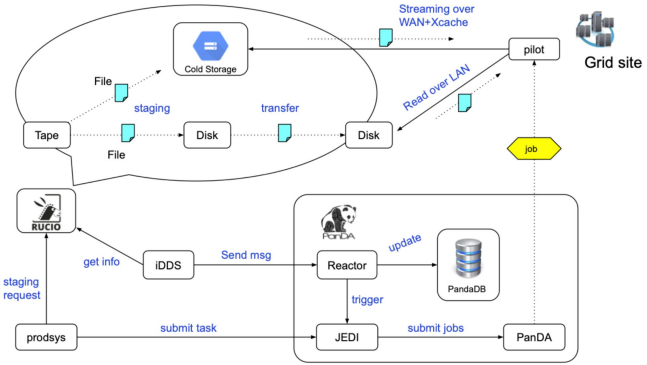
- ATLAS workflow automation via PanDA + iDDS (intelligent data delivery service) being enhanced to send workflows to other automation systems, including REANA
- Active learning supporting ATLAS analyses is the first target
  - Model parameter points in the original parameter space → Full production chain for each point → RECAST for each point to calculate CLs → Iterate with active learning to intelligently and efficiently explore the parameter space
- Aiming to demonstrate the full machinery by the end of the year
- **Will bring a new DAP capability: PanDA as the engine for DAP workflows preserved with REANA**

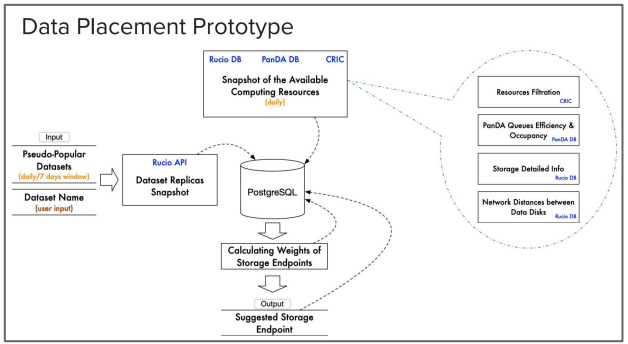


ATLAS: Data management

- **Storage is the tightest resource currently and at HL-LHC**
- **Data carousel** tightly integrating workflow and data management for efficient processing of tape-resident data in full production
  - BNL NPPS and SDCC have led its development
  - *See supplementary slide*
- Leadership in managing ATLAS resource usage
- Rucio core development
  - Reengineered and new components/features
    - ‘conveyer’ transfer daemon
    - ‘reaper’ deletion daemon
    - multi-hop
  - Shared Rucio expertise with Belle II
- ATLAS data management readiness for Run-3
  - Updated replication policies, managing large data export streams with higher Run-3 data volumes, improved automated cleanup and copy reduction
- Applying data popularity knowledge to intelligent data placement

Data carousel





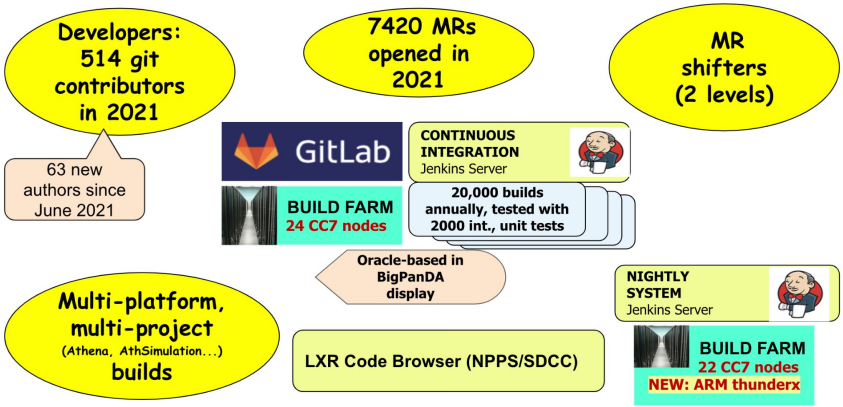
C. Serfon, A. Klimentov, T. Maeno et al

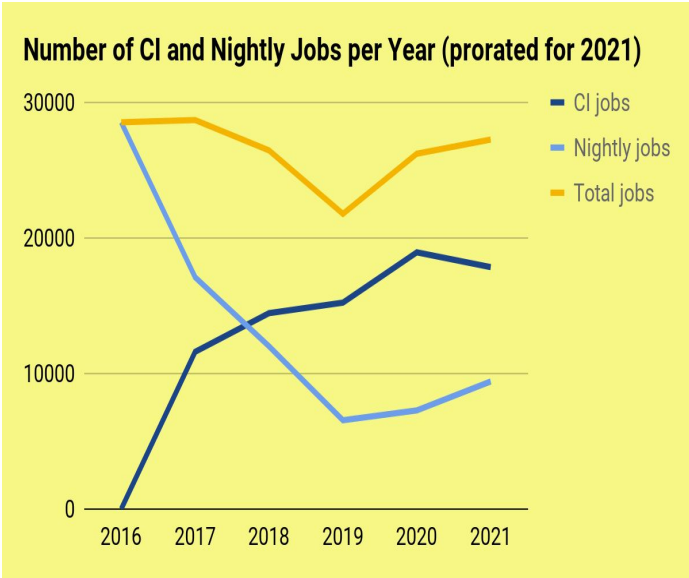
ATLAS: Software infrastructure

Short CI and Nightlies status summary

- **Stability:** <1% of jobs with technical problems
- **Regular software updates**, security patches applied
- **Best possible Linux VMs** (no CPU steal, local SSD)
- **ARM/aarch64** machine recently added
- **Sufficient reserve:** can run 20% more jobs
- **Excellent BigPanDA-based monitoring**
- **Good progress** of OpInt CI project

➢ READY FOR RUN3

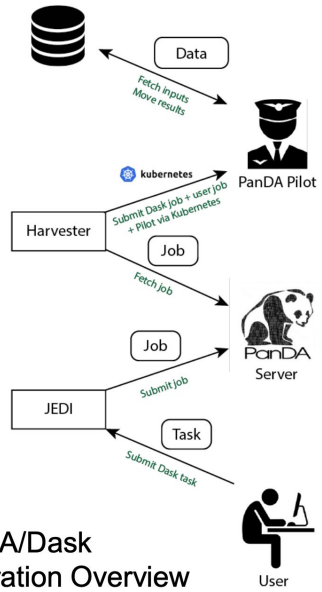




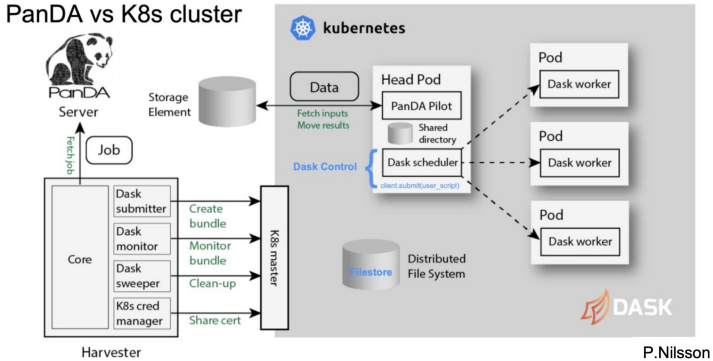
Alex Undrus, Shuwei Ye

ATLAS: R&D project with Google

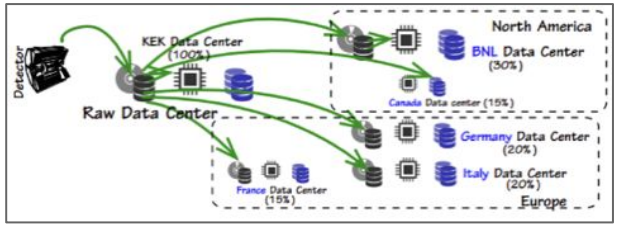
- Now in Year 2 of the BNL-led ATLAS-Google R&D collaboration
- Has become an R&D sandbox and technology evaluation platform
  - pythonic ecosystems
    - PanDA+Jupyter+Dask for pseudo-interactive PHYSLITE analysis
  - columnar analysis
    - BigQuery database for column-wise analysis
  - data delivery/transformation/storage technologies
    - Parquet translation of PHYSLITE analysis format
  - Kubernetes ecosystem
    - PanDA now able to scale up workers via Kubernetes in GKE, run large scale tasks on demand
- Cost evaluation for possible research use
  - NB the Google cloud is the computing facility for Rubin at present
- Progressing well on the Year 2 goal: demonstrate usefulness of Google cloud for analysis
  - and discussing Year 3 plan



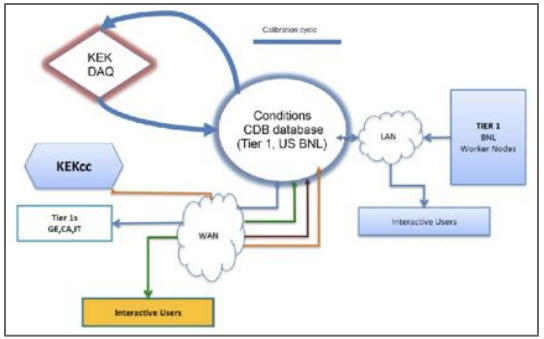
PanDA/Dask
Integration Overview



Belle II: Data, Calibration, Conditions

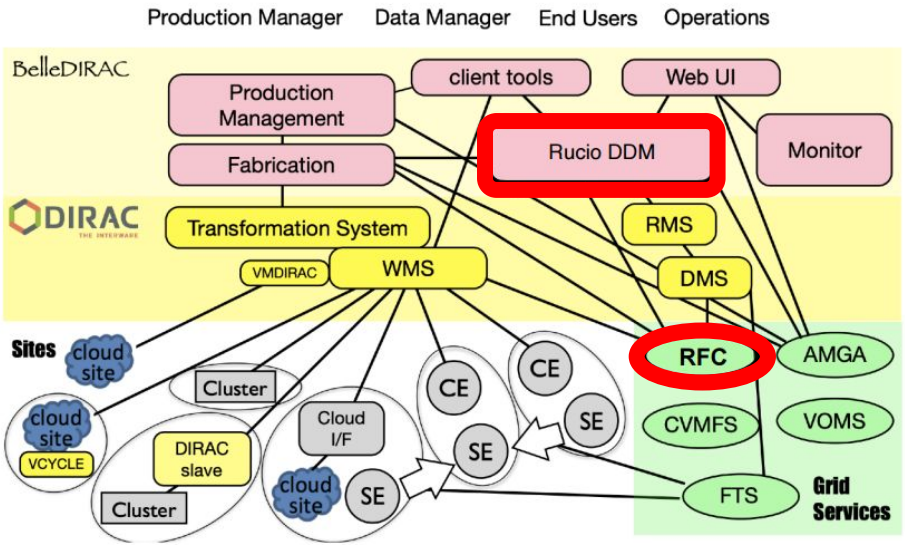


- BNL is Belle II US Tier 1
- The biggest raw data center outside of KEK
- Prompt calibration moved from KEK to SDCC in 2020
- Belle II S&C at BNL a collaboration between NPPS, SDCC, and the neutrino group (EDG)



- Belle II conditions database operated from SDCC
- Software development in NPPS
- Software being reengineered/generalized for sPHENIX and experiment-agnostic generic product
- NPPS provides database group co-convenor and Global Tag manager

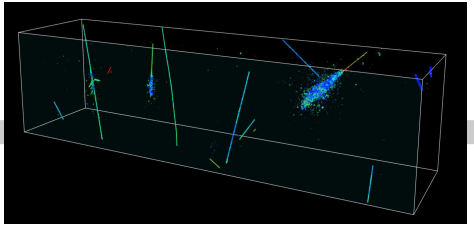
Belle II: Distributed data management



- Successful upgrade of Belle II data management to Rucio this year
  - New DIRAC plugin for Rucio
  - Updated Belle II Dirac APIs
- Performed transparently with minimal disruption to the running experiment
  - Positive feedback
  - new user functionality: async deletion, collections
- ~100M replicas, 18PB registered data, request rate 80-100Hz
- Integrating new features since the migration
  - ATLAS commonality: WebUI, popularity
  - Belle II developments fed back to the community, e.g. lightweight monitoring

Cedric Serfon, Ruslan Mashinistov, Paul Laycock

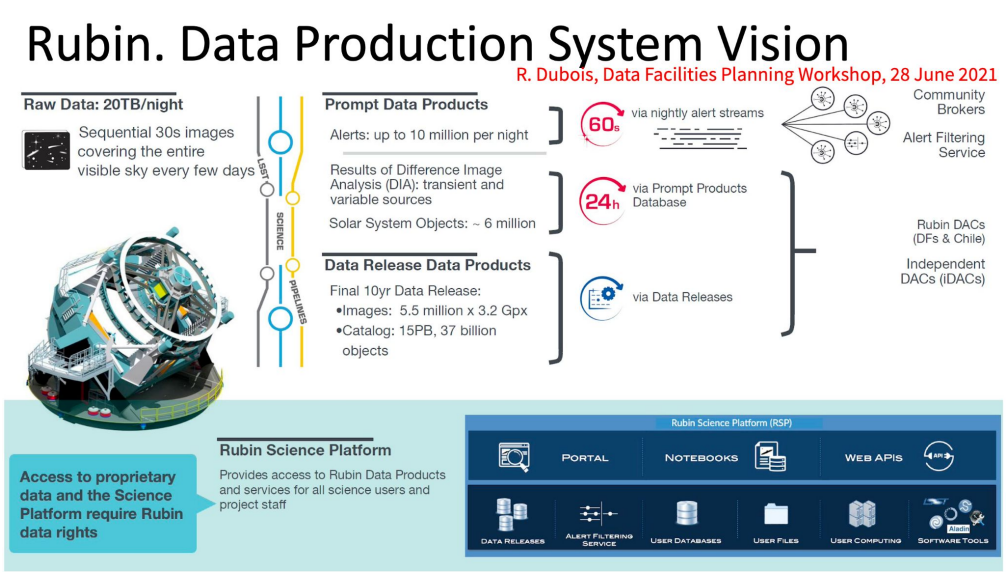
DUNE: Software and databases



- Software
  - BNL neutrino group leads development of Wire-Cell LAr TPC simu/reco/visualization package
    - Collaboration with CSI via HEP-CCE on adapting Wire-Cell for GPUs
  - NPPS has leading role in software framework requirements/design
    - Recently reviewed by HSF, moving towards an implementation plan
    - Addressing HPC and GPUs, data formats, potential use of HDF5
  - AI/ML R&D underway with CSI on precision GAN-based sim/data systematics
- Databases
  - NPPS co-coordinates DUNE databases
  - New NPPS PostDoc starting on conditions DB
    - Offline conditions workflows using ProtoDUNE data
  - Participating also in other DB development: Parts Identifier, TimeScaleDB (slow controls)
- Participating in DUNE S&C CDR planned for 2022

Rubin Observatory: Workflow management

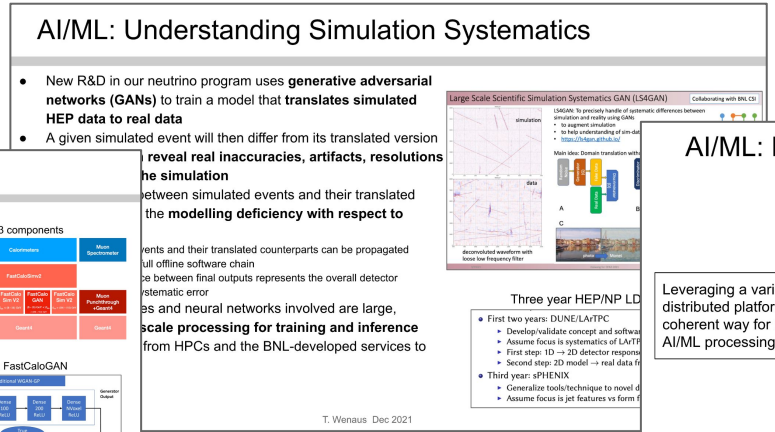
- After 1+ years of an evaluation involving substantial NPPS effort investment, Rubin decided in August to adopt PanDA for workflow management
- A challenging environment, approaching ATLAS in scale: ~200k concurrent jobs, complex workflows
- PanDA's new DAG support via iDDS is key to flexibly and fully supporting workflows
- NPPS now hiring a new PanDA core team member to contribute to Rubin and other PanDA projects. BNL staff position open!
- PanDA instance at SLAC will be established (PostgreSQL based, like the sPHENIX BNL one)

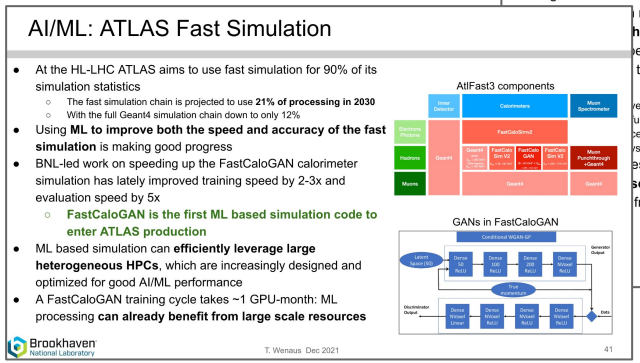


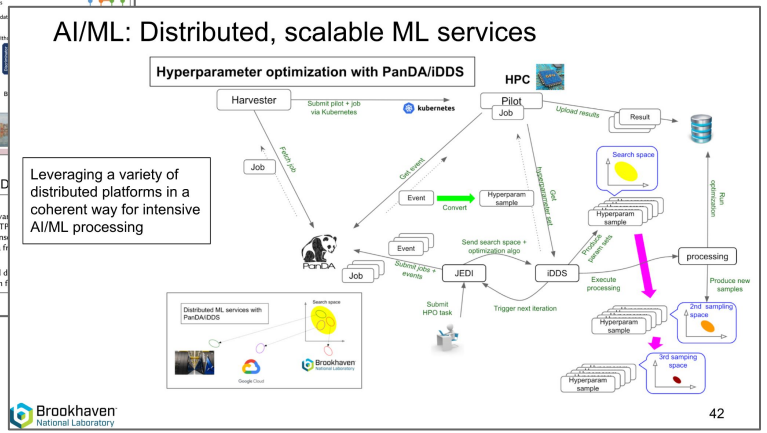
AI/ML

BNL HEP AI/ML Vision

- Growing rapidly at BNL, as everywhere
- Our approach: apply AI/ML after exhausting well understood "conventional" ways to extract information from data
  - Use accrued domain knowledge to wisely guide the application of AI/ML
  - Produce results that can be understood and trusted, including their error properties
- Examples of AI/ML projects underway are in the supplemental slides, no time!







Plans and priorities

- Sustaining and improving S&C in our running and soon to be **running experiments**: ATLAS, Belle II, STAR, sPHENIX, Rubin
- Continuing to **grow cross-experiment** contributions, common activities, collaborations
- **sPHENIX readiness** as BNL's next major experiment, to start up in 2023
- The intensive development program remaining for **HL-LHC readiness**
- **Supporting the EIC** as it moves to detector collaborations and the detailed design phase
- Last year: "**Growing new opportunities**: Rubin Observatory, DUNE, …" … check, check! :-)
- Building more **AI/ML** experience and activity
- Building **data & analysis preservation** as an integral part of the experiment life cycle, from PHENIX to others

Many thanks

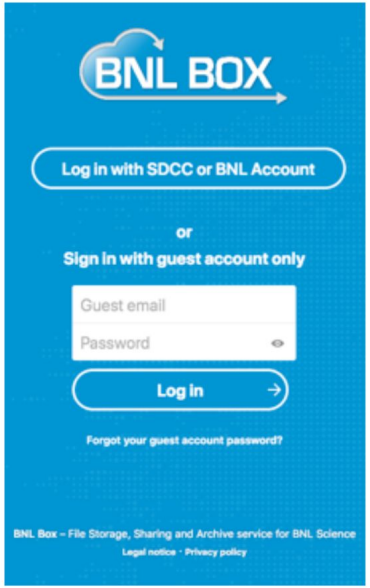
- Many thanks to those in the BNL Physics Department and elsewhere who contributed to this talk, and many more who have contributed to the work described
- An inevitably partial list of contributors:
  - David Adams, Dmitry Arkhipkin, Fernando Barreiro Megino, Costin Caramarcu, Tim Chou, Kaushik De, John De Stefano, Johannes Elmsheuser, Yuri Fisyak, Carlos Gamboa, Vincent Garonne, Robert Hancock, Chris Hollowell, Jin Huang, Amol Jaikar, Kolja Kauder, Hongwei Ke, Alexander Kiselev, Alexei Klimentov, Eric Lancon, Paul Laycock, Christian Lepore, Jane Liu, Meifeng Lin, Tadashi Maeno, Ruslan Mashinistov, Paul Nilsson, Marcin Nowak, Joe Osborn, Sergey Padolski, Louis Pelosi, Victor Perevoztchikov, Chris Pinkenburg, Maxim Potekhin, Tejas Rao, Ofer Rind, Cedric Serfon, Dmitri Smirnov, Jason Smith, Will Strecker-Kellogg, Alex Undrus, Gene Van Buren, Brett Viren, Jason Webb, Tony Wong, Iris Wu, Shuwei Ye, Alex Zaytsev, Xin Zhao

Supplementary

BNL Box Cloud Storage

BNL Box - Enterprise File Sync-and-Share system based on Nextcloud and integrated into the SDCC to provide flexible, easy-to-use, unified cloud storage for all BNL scientific users

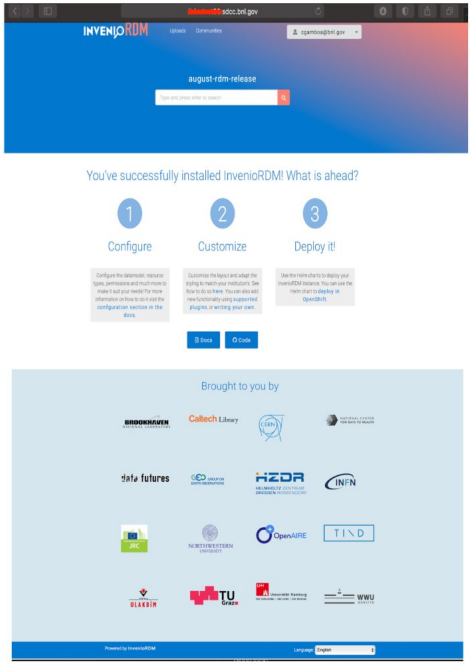
- Account with 50GB (default) available to all BNL users with SDCC or AD accounts — currently 293 users, ~4TB data
- High performance Lustre storage (1 PB usable) with 40 Gbps WAN-to-storage network capacity
- Data access via command line, browser, desktop & mobile clients, including automated synchronization
- Flexible file-sharing methodologies for sharing data with external collaborators (e.g. CFN scans, Belle-II logs,...)
- Custom tape archiving capability (via Lustre & HPSS) for large data sets



Brookhaven National Laboratory

Invenio-based Digital Repositories at SDCC

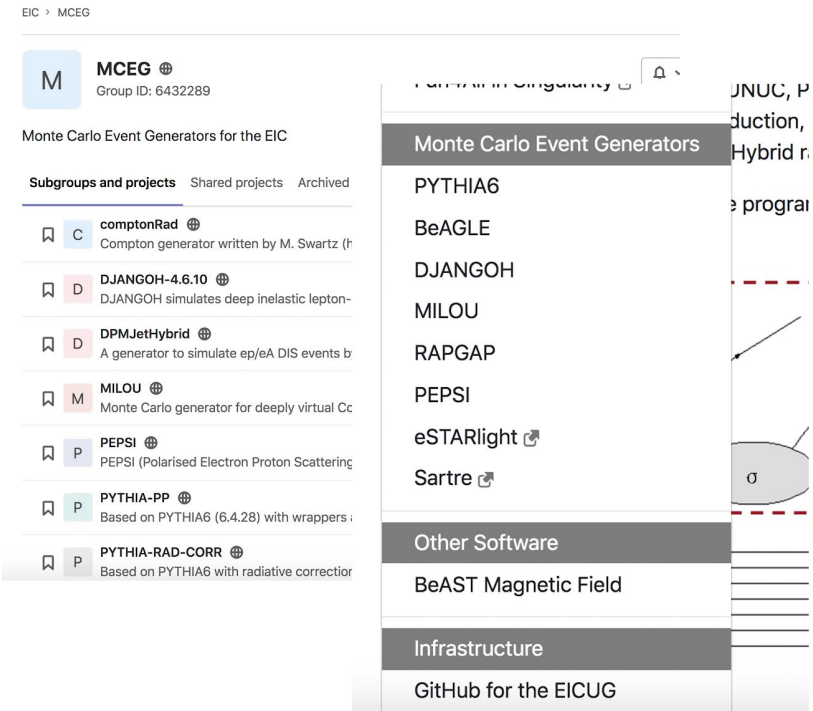
- Invenio is a open source framework to build scalable digital repositories e.g CERN Videos and Zenodo
- SDCC also supports computing infrastructure for Invenio based applications, along with customized network, storage and Authentication infrastructure enabled to host services (production, testing and developing).
- Invenio is used to build community driven digital repositories for example the **SET**: National Nuclear Security Administration repository
- SDCC is part of the invenioRDM project, a multi-disciplinary effort to build a *community supported* Research Digital Repository based on invenio. InvenioRDM will replace Zenodo
- Different BNL communities will migrate their Zenodo based repositories to invenioRDM (EIC-Zenodo)
  - InvenioRDM v6.0 (with long-term support) released in August 2021

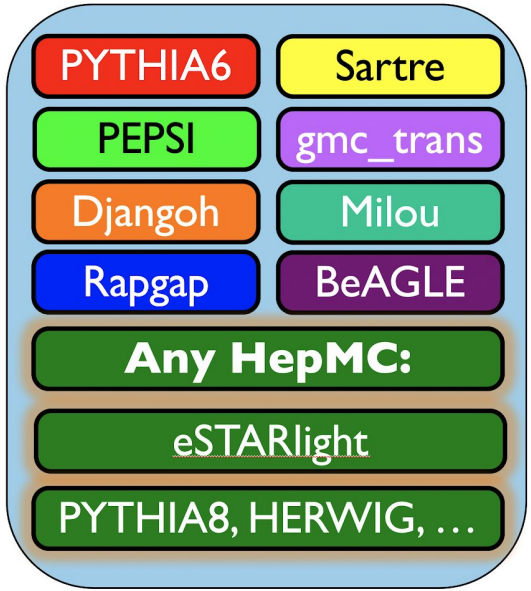


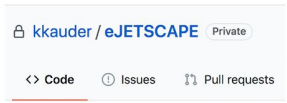
Brookhaven National Laboratory

EIC: Generators

- EIC MC collection is consolidated





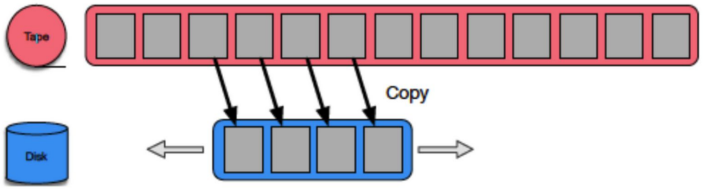


- … and growing!

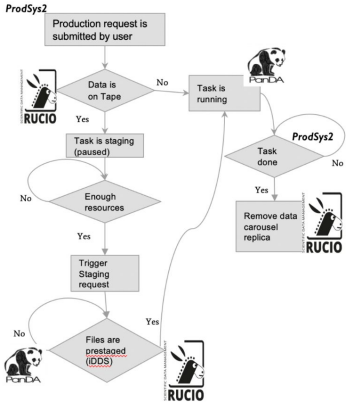
Data Carousel

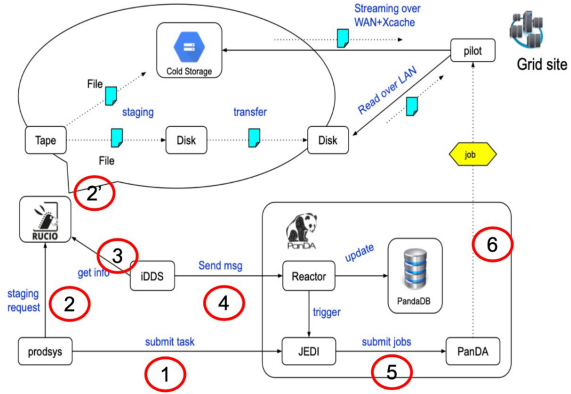
Recent talk

- Data Carousel: on-demand reading from tape without pre-staging
- Uses a rolling disk buffer whose size can be tuned to suit available resources and production requirements
- Key to success: rate at which data can be staged to disk at the Tier-0 and Grid sites and processed
- Technique can eventually be used for any experiment
  - Two tape challenges during 2021 to address I/O tape performance at WLCG Tier-1s (4 LHC VOs)
- In ATLAS production today for data reprocessing, derivation and Monte-Carlo production



Data Carousel workflow and New distributed software component : intelligent Data Delivery Service (iDDS)





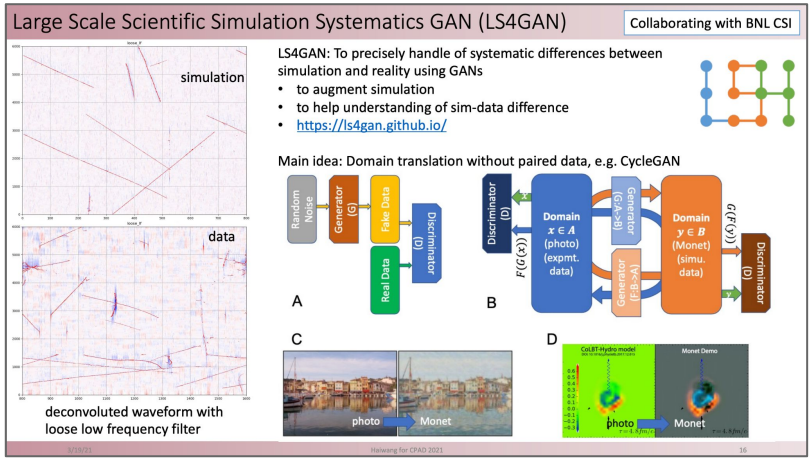
AI/ML: BNL HEP AI/ML Strategy

BNL HEP AI/ML Vision

- ML is accepted as an important tool capable of improving our physics reach
  - As it has been for over 30 years
- The ‘deep learning revolution’ over the last decade has made AI/ML an ever more powerful tool in our analytical toolkit
  - Leveraging rapid growth in AI/ML and related analytic techniques/tools (e.g. back propagation), and accelerated architectures tailored to AI/ML
- A pillar of BNL’s scientific computing strategy: *Apply AI/ML in creating solutions to scientific computing challenges in experiment and theory*
- **We see AI/ML as a tool and approach to apply after exhausting well understood "conventional" ways to extract information from data**
  - Use accrued domain knowledge to wisely guide the application of AI/ML
  - Produce results that can be understood and trusted, including their error properties
- A few example activities...

AI/ML: Understanding Simulation Systematics

- New R&D in our neutrino program uses **generative adversarial networks (GANs)** to train a model that **translates simulated HEP data to real data**
- A given simulated event will then differ from its translated version in ways that can **reveal real inaccuracies, artifacts, resolutions and biases in the simulation**
- The difference between simulated events and their translated form represents the **modelling deficiency with respect to actual data**
  - Simulated events and their translated counterparts can be propagated through the full offline software chain
  - The difference between final outputs represents the overall detector simulation systematic error
- The data volumes and neural networks involved are large, requiring **large scale processing for training and inference** that can benefit from HPCs and the BNL-developed services to use them easily



Three year HEP/NP LDRD program

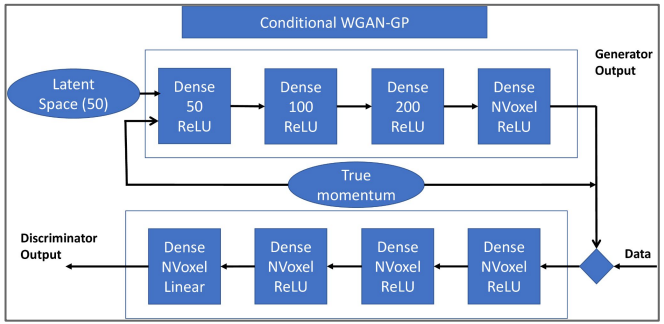
- First two years: DUNE/LArTPC
  - Develop/validate concept and software tools.
  - Assume focus is systematics of LArTPC detector response.
  - First step: 1D → 2D detector response model as "fake" → "real"
  - Second step: 2D model → real data from single-phase ProtoDUNE
- Third year: sPHENIX
  - Generalize tools/technique to novel domain.
  - Assume focus is jet features vs form factors and detector response.

AI/ML: ATLAS Fast Simulation

- At the HL-LHC ATLAS aims to use fast simulation for 90% of its simulation statistics
  - The fast simulation chain is projected to use **21% of processing in 2030**
  - With the full Geant4 simulation chain down to only 12%
- Using **ML to improve both the speed and accuracy of the fast simulation** is making good progress
- BNL-led work on speeding up the FastCaloGAN calorimeter simulation has lately improved training speed by 2-3x and evaluation speed by 5x
  - **FastCaloGAN is the first ML based simulation code to enter ATLAS production**
- ML based simulation can **efficiently leverage large heterogeneous HPCs**, which are increasingly designed and optimized for good AI/ML performance
- A FastCaloGAN training cycle takes ~1 GPU-month: ML processing **can already benefit from large scale resources**

AtlFast3 components

GANs in FastCaloGAN



AI/ML: Distributed, scalable ML services

Leveraging a variety of distributed platforms in a coherent way for intensive AI/ML processing