The HSF-India Project

David Lange
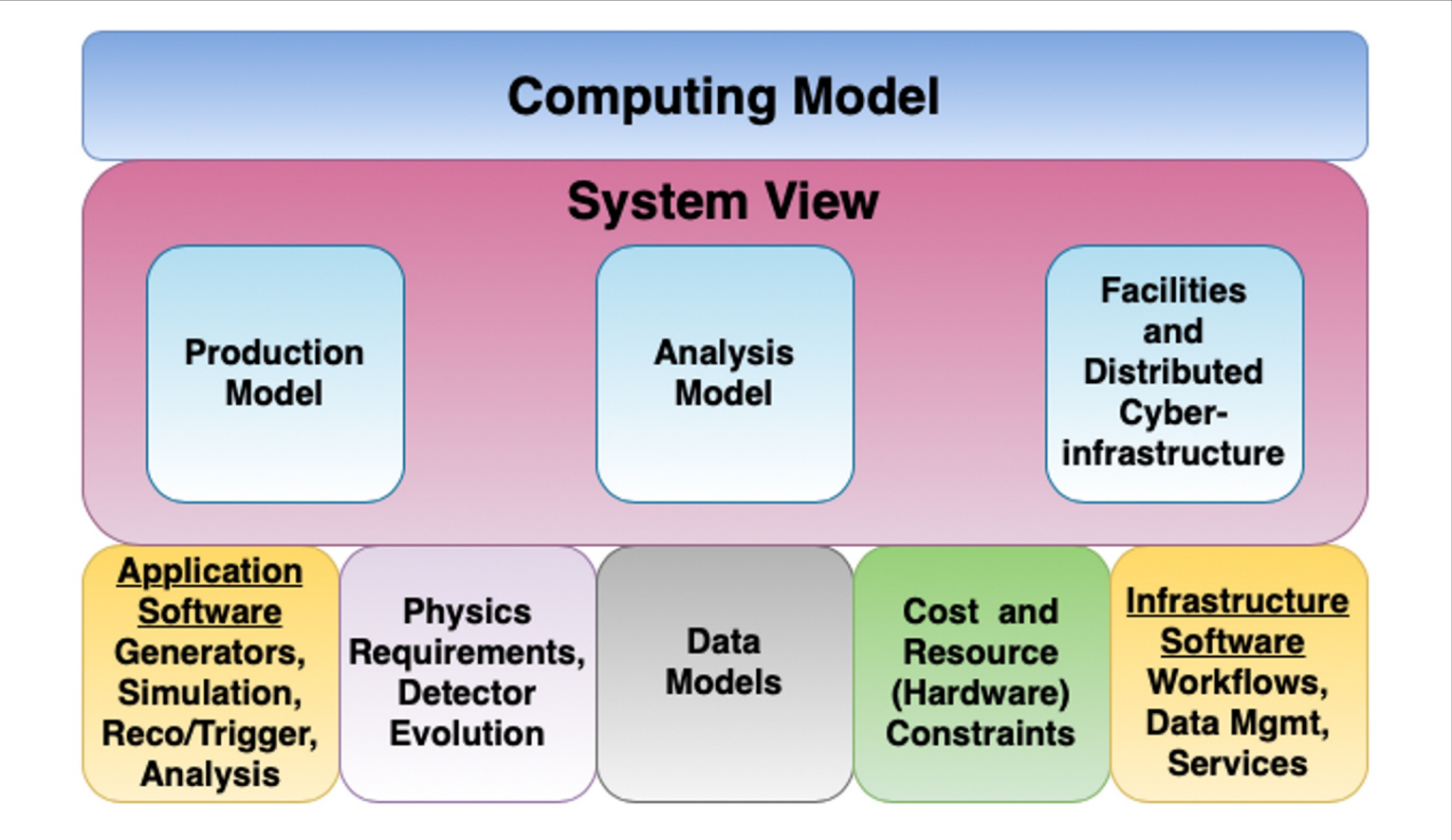
Princeton University

Over the last decade, it became clear that a new approach to large-scale research software is needed

Big (Team) Science projects need "Computing Models".

More and more software is considered as infrastructure

# HSF Community White Paper

January 2017
UCSD

June 2017
Annecy

"*The result: a Programme of Work for the field as a whole, a multifaceted approach to addressing growing computing needs on the basis of existing or emerging hardware.*"

Eckhard Elsen (CERN Director of Research and Computing), editorial published with CWP/Roadmap

Many workshops, involving a diverse group
- International participants
- Computing Management from the Experiments and Labs
- Individuals interested in the problems
- Members of other compute intensive scientific endeavors
- Members of Industry
- http://s2i2-hep.org/
- https://hepsoftwarefoundation.org/

Individual Papers on the arXiv:
Careers & Training, Conditions Data, DOMA, Data Analysis & Interpretation, Data and Software Preservation, Detector Simulation, Event/Data Processing Frameworks, Facilities and Distributed Computing, Machine Learning, Physics Generators, Security, Software Development, Deployment, Validation, Software Trigger and Event Reconstruction, Visualization

Integrated Community White Paper
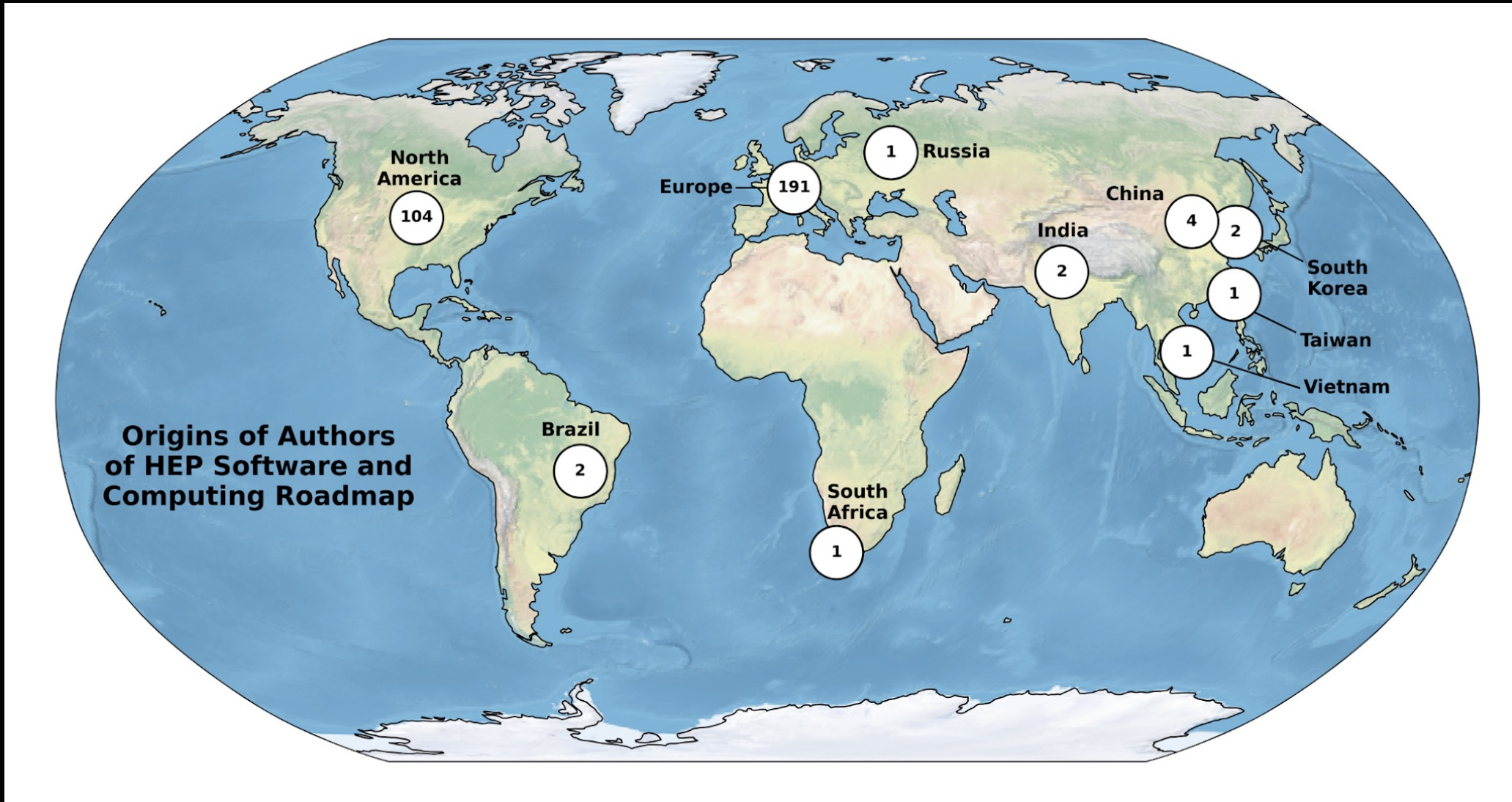
arXiv 1712.06982

IRIS-HEP

Community Consensus enables regional/national engagement

- The Institute for Research and Innovation in Software for High Energy Physics (IRIS-HEP) was established to meet the software and computing challenges of the HL-LHC, through R&D for the software for acquiring, managing, processing and analyzing HL-LHC data.

- IRIS-HEP is only one example of successful, collaborative, regional/national proposals that grew out of the CWP process.

IRIS-HEP: A distributed, virtual, Software Institute
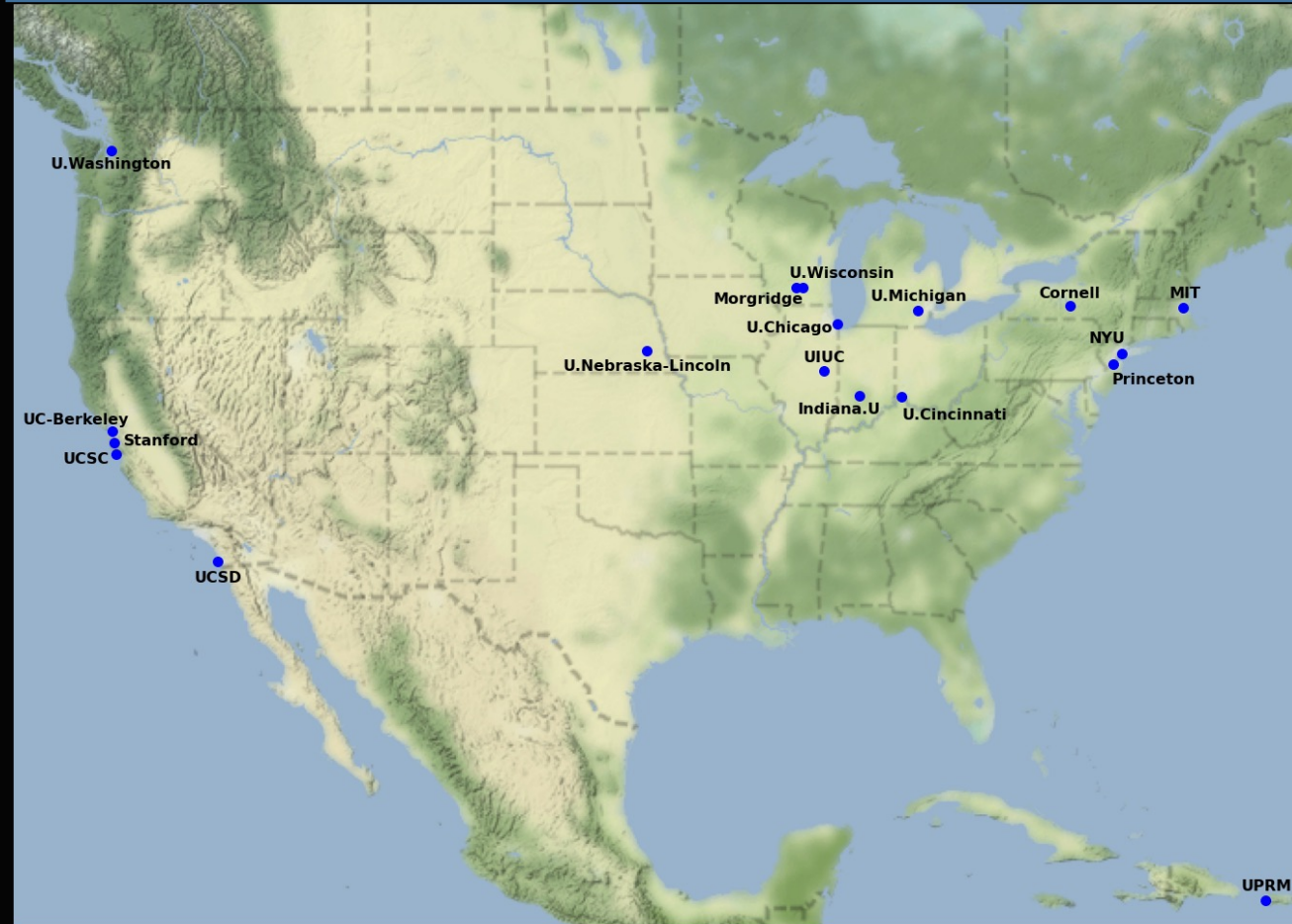
However, nearly all authors of the HSF Community Roadmap were from institutions in Europe and the US

Origins of Authors of HEP Software and Computing Roadmap

North America: 104
Europe: 191
Russia: 1
China: 4
South Korea: 2
India: 2
Taiwan: 1
Vietnam: 1
Brazil: 2
South Africa: 1

This is rather different from the representation in our experiments [3% vs. O(10%)]

	China	India	Japan	South Korea	Taiwan	Total authors
Atlas	164	0	90	0	9	2900
CMS	47	80	0	69	25	2200
Alice	46	84	14	33	0	1100
Dune	2	31	4	5	0	1300

Approximate accounting, all mistakes are mine…

- We looked for mechanisms to engage researchers who are already actively engaged in software to join collaborative research software projects (via mechanisms fostered by the HSF or otherwise)

Towards making this a global collaboration:
The HSF-India project

- HSF-India is a 5 year project funded by the US National Science Foundation (Princeton, UMass-Amherst, Oregon State) that aims to build international research software collaborations between <u>US, European, and India based researchers</u> to reach the science goals of experimental particle, nuclear and astroparticle research.

- Given the growing complexity of our scientific data and collaborations, these collaborations are increasingly important to raise the collective productivity of our research community.

Intended as a long-term investment in international team science.

# HSF-India is different from a "typical" research project

Much of our funding is to facilitate research collaborations, rather than directly fund a specific research activity

Around half of the funding is for supporting participant activities

- Training in research software skills
- Bidirectional research exchanges
- Summer or semester student programs

Bootstrap collaboration through software training
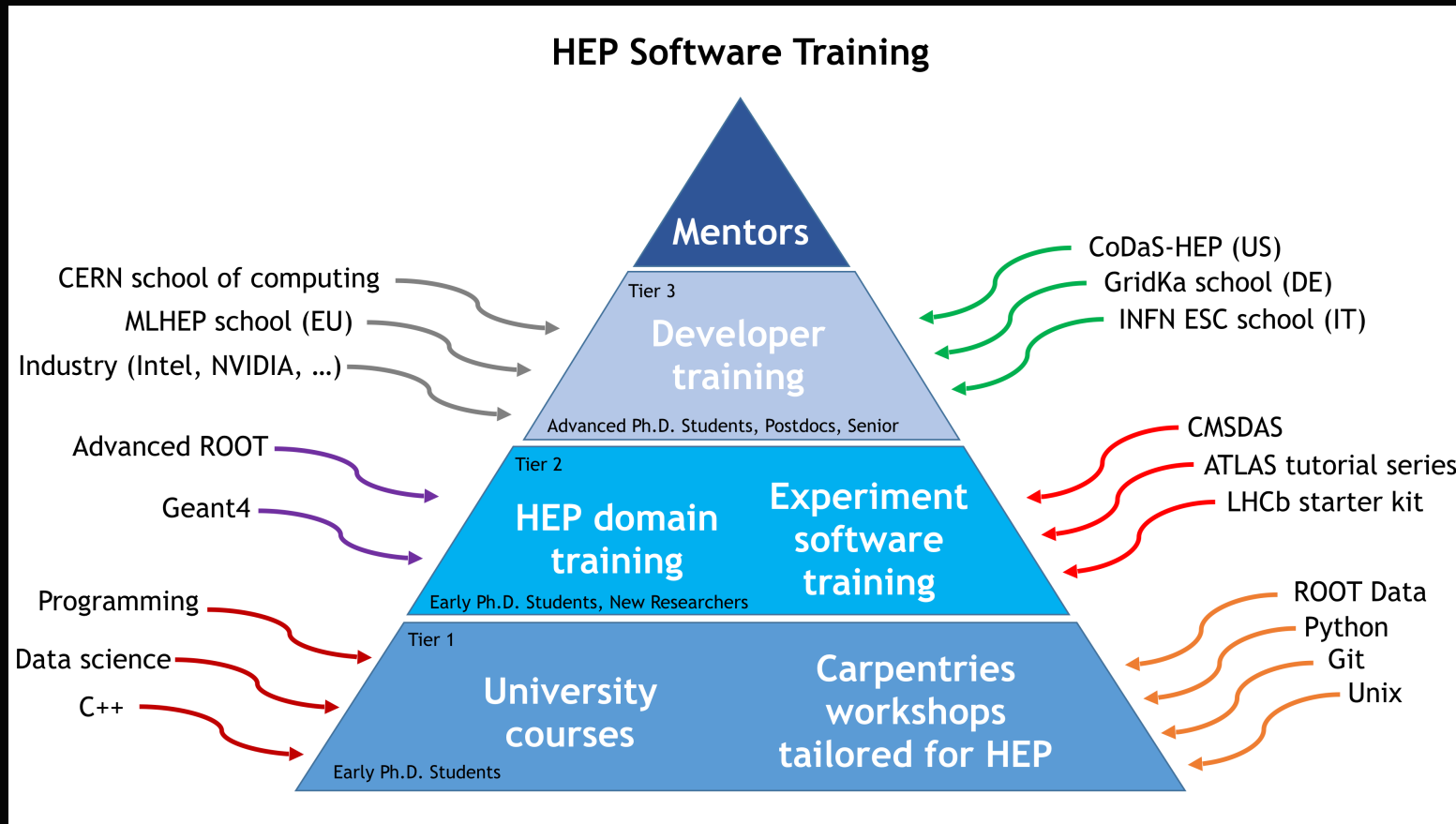
- A vision for training in HEP: researchers progress (vertically) from basic skills training, through user training in existing software to training in skills needed to develop new research software.

Our first software workshop at TIFR in Mumbai (April 2023)

- ~50 students registered ahead, growing registration during the week. Most students came from one of the universities in the Mumbai region.

- Mix of local and US instructors

- Materials derived from/patterned after HSF training courses

Scientific Python

Marshmallow Challenge

# NISER software workshop in December 2023(Bhubaneswar)

- Curriculum built around high-level (but hands-on) overviews of scientific python, machine learning, modern C++, GPU programming and topics of local interest (eg, simulation techniques)

Infrastructure – remote vs local resources

- Using remote resources at UChicago (SSL-River) gives us a working environment that we can reuse from workshop to workshop.
  - This has been great. A laptop+browser gives everyone the same environment
  - Available thanks to our connection with IRIS-HEP
  - Reproducibility, reproducibility, reproducibility….
- Things to consider
  - Images with CUDA and/or ML frameworks are large. They take time to setup and use.
  - Infrastructure allocates a full GPU (or slice of one) per binder instance. This means you need more GPUs than students
    - No one likes to look over the shoulder of their neighbor
  - Everything you knew about building consistent environments with GPUs and ML frameworks from one year ago is obsolete..
  - Timezones do not always facilitate real-time support…

Start of the day search for binderhub instances…
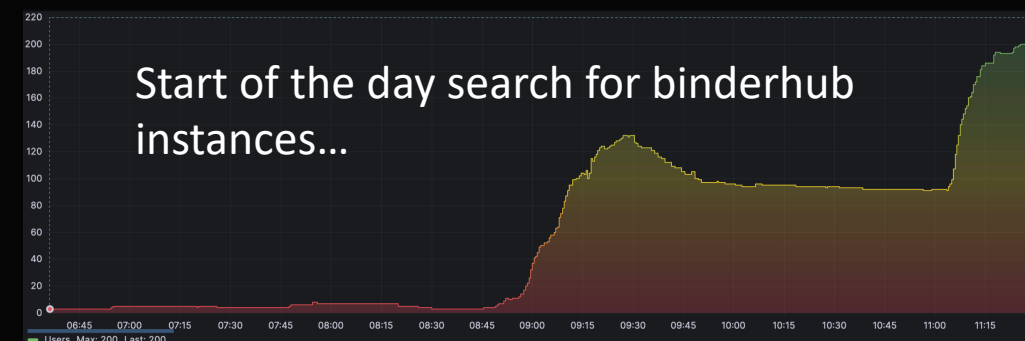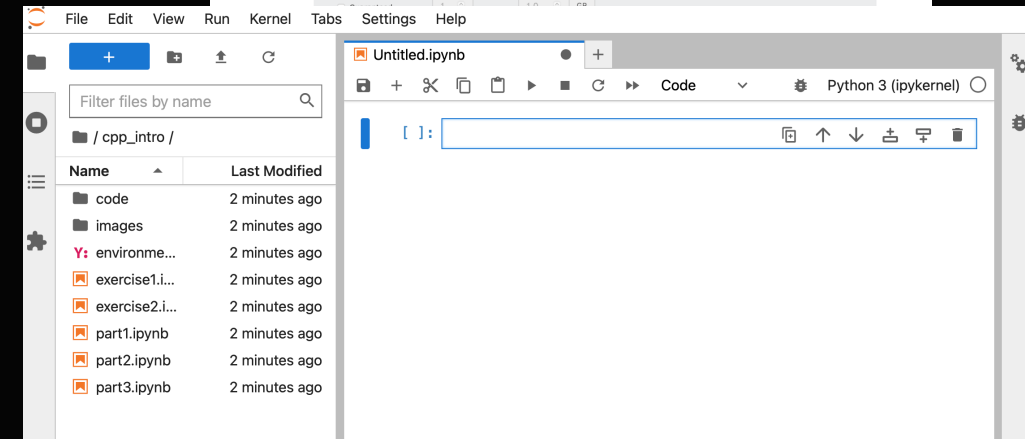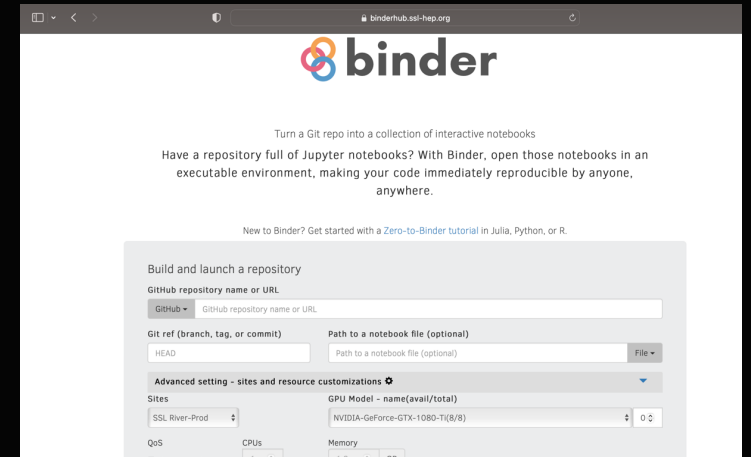
Upcoming HSF-India workshops

- Regional events help reduce travel burden (within India)
  - University of Delhi

  - [Next week](#)
    - 40-50 students expected
  - SINP / VECC – Kolkata
    - Planned for December
  - University of Hyderabad
    - Planned for January

- Always interested in engaging new researchers in these events

Summer project Fellows Program

- **Project focused** aiming to bring students into contact with "mentors" to work on a specific, pre-defined project, allowing them to grow their software skills and experience working in large projects

- These short term projects that build **longer-term collaborations** in research software and foster **scientific career progression**

Bidirectional Research Exchange Opportunity for WLCG/HSF

We also have funding for "research exchanges" that support travel costs for 1-3 months to work directly with another research group

Who can we support

- Researchers affiliated to a US university/lab doing an exchange based in India

- Researchers affiliated to a university/lab in India doing an exchange based in US [or at CERN to work with a US affiliated group]

Interested to talk with anyone about either project/host offers or interest in doing an exchange.

Lessons learned (so far)

- There is a lot of interest especially from younger researchers

- Obtaining visas has proven non-trivial (partially due to policy changes that came with/during COVID), but we now understand the process pretty well

- It is important to take the time to understand local/regional practices and models for doing research
  - For example, we did not understand the funding model for students and postdoc when we started
  - Doing site visits has proven to be a good approach to meeting and understanding better opportunities and constraints

- Local interest (and funding) are essential for organizing events such as software workshops.

Thanks for listening

- HSF-India is a still a new project. We hope can catalyze global collaboration in research software in Physics.

- We are here to discuss ways that this project work with you to help further HSF goals and initiatives
  - http://research-software-collaborations.org