Yandex School of Data Analysis

SHiP
Search for Hidden Particles

Skygrid

Alexander Baranov, Konstantin Nikitin, Andrey Ustyuzhanin

Background

- Yandex-CERN group
- Analyzes involving Machine Learning (LHCb)
- Monte-Carlo generation (SHiP)
- Experiment reproducibility: fixing computations pipeline, so you can repeat it at any time.

Background

- Yandex-CERN group
- Analyzes involving Machine Learning (LHCb)
- Monte-Carlo generation (SHiP)
- Experiment reproducibility: fixing computations pipeline, so you can repeat it at any time.

On cluster

Problems

- Want to compute on X while knowing how to support Y
- Run well-tested old software on old platforms
- Strictly fix environment where computations run

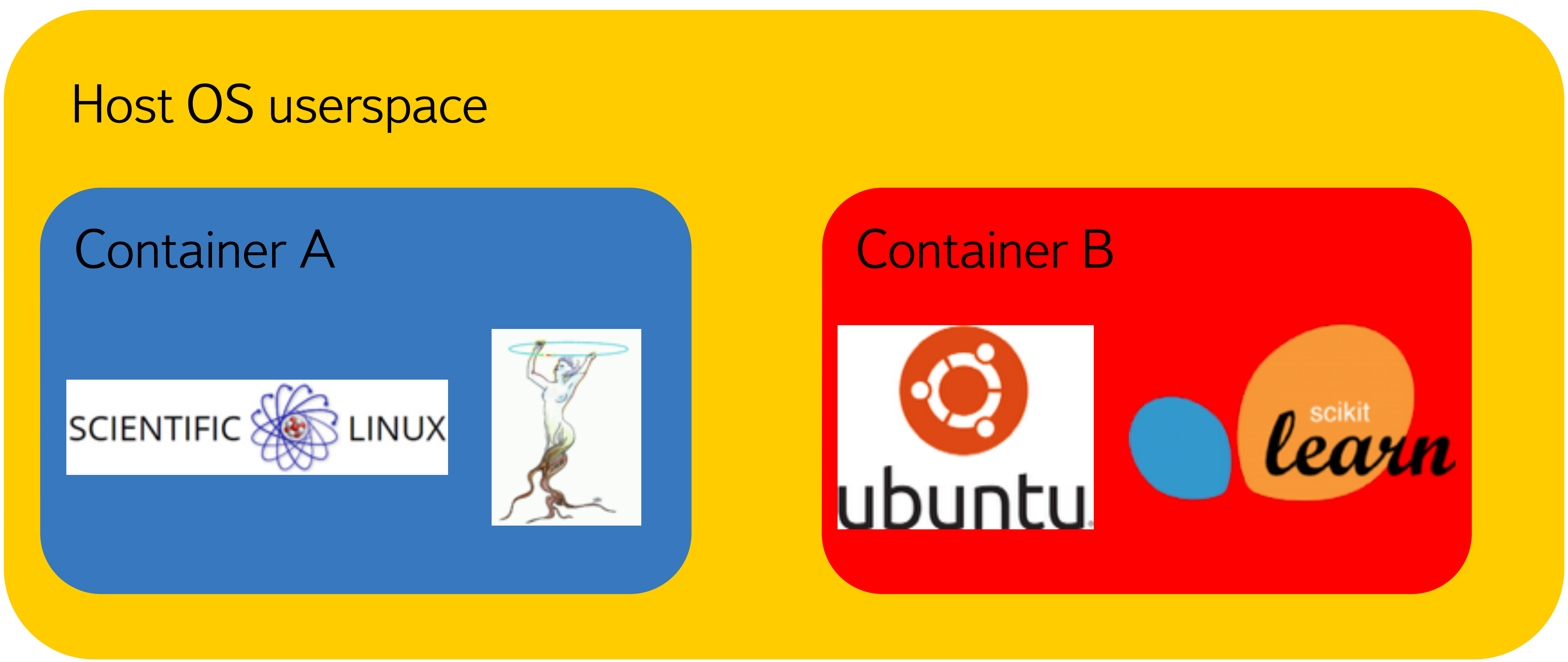
What is Skygrid?

- System for distributed computing
- Currently is being developed at Yandex
- Uses docker for job sanboxing
- Provides nice HTTP APIs to connect with existing services
- «Docker Executor as a Service»

Docker container execution

Docker:

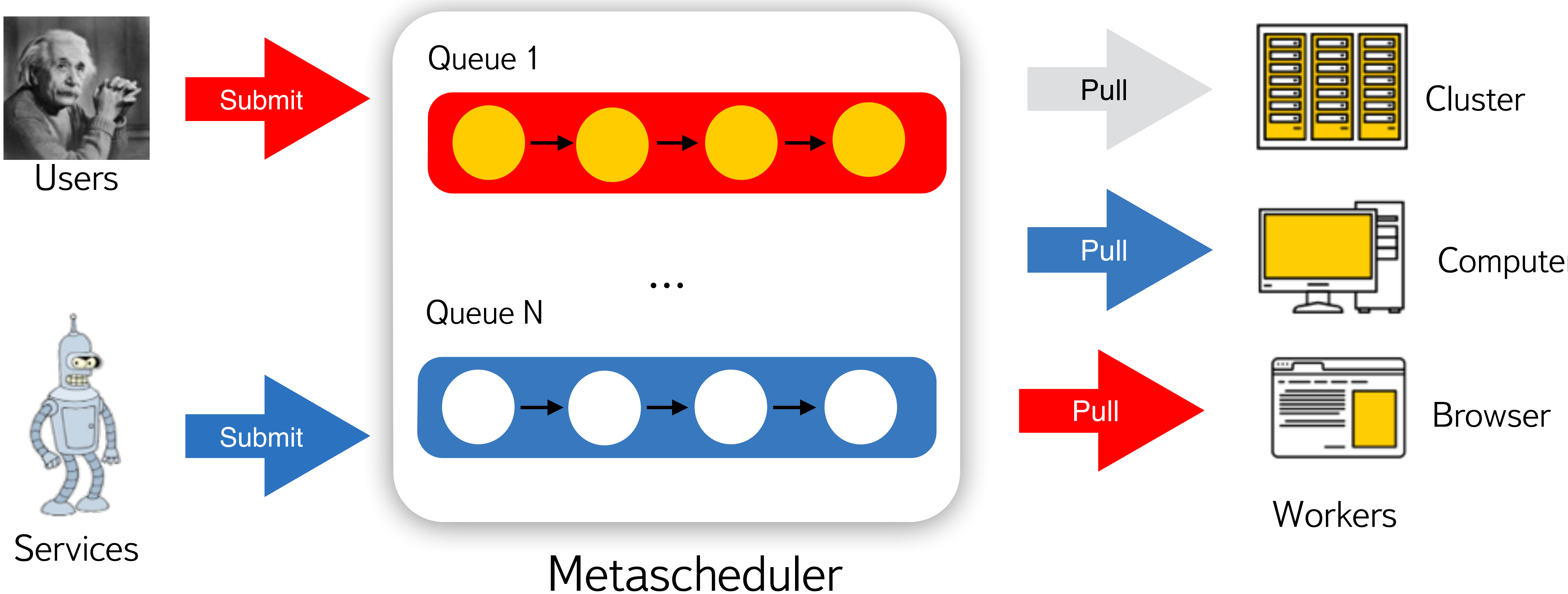
- Lightweight containers (LXC + cgroups + AUFS)
- Container = packed environment (bins, libs, data)
- Lets you execute your container as user process on any environment



Skygrid architecture

«Job → Queue → Worker» scheme:

- Job = JSON description of what to do
- Queue = Metascheduler service (matches jobs to resources)
- Worker = Basically anything (single computer, cluster, browser, ...)



Docker container executor

- Pulls docker image
- Downloads input data and mounts it into container
- Runs container
- Uploads results

Advanced things:

- Split jobs into small chunks (YARN)
- Container communication and coordination (MPI, Zookeeper)

In-browser volunteer computing

- Skygrid provides API so your browser can pull a job
- Helpful for volunteer computing
- Prototype: generate MC in browser (Pythia→ LLVM → JavaScript)
- Details at poster session on DiBroCop — Distributed inBrowser ComPutations

Seriously, you can run it on your phone:

Current status

- Developing at https://github.com/anaderi/skygrid
- Working prototype with 60-machine cluster
- YARN worker prototype

Roadmap

- Advanced scheduling schemes (Auction)
- Integration with existing applications (IPython, etc)
- Jobs coordination(MPI + etcd)
- In-cluster job splitting
- Token-based authentication

Questions and Answers

Contact me via a.baranov@cern.ch

Fork us on github.com/anaderi/skygrid

Backup

Yandex infrastructure

- OpenStack
- Ubuntu-based
- Everything is packaged in deb
- Home-brewed deploy and monitoring
- No one wants to bother with SLC

Metascheduler

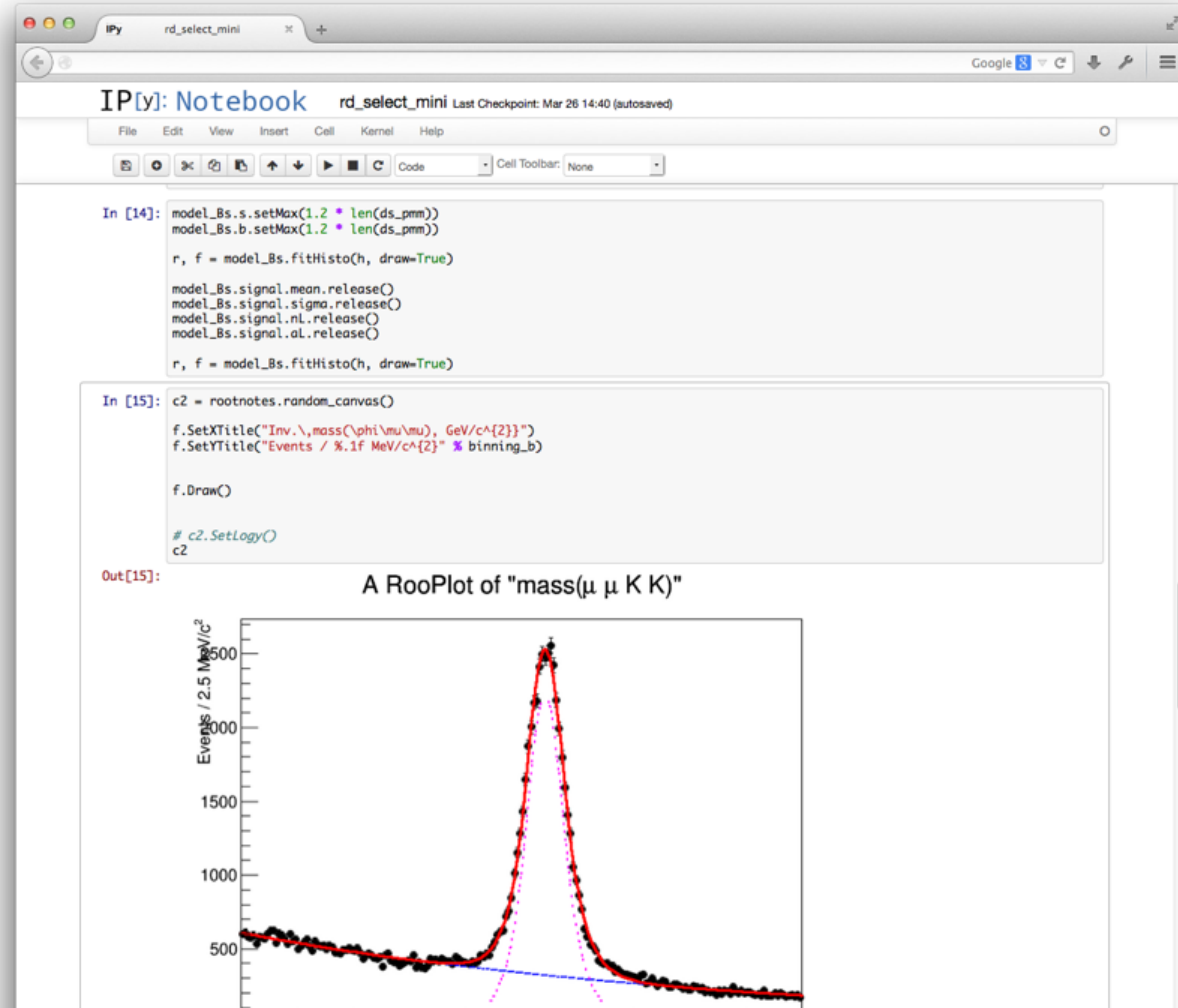
- Basically the global queue service
- Queues can be created with different scheduling schemes(FIFO, Round-Robbin, Auction)
- Saves jobs persistently
- Schedules jobs to execution when pulled by workers
- Stores job UID, descriptor, status and output URIs

Workers

Pulls jobs from metascheduler and executes them on available CPUs and memory

- Can be single computer i.e. «Flat» worker
- Or cluster i.e. YARN
- Can perform internal job splitting and coordination

Integration with IPython



In-browser computing performance