

A complex network diagram with numerous nodes and connecting lines, forming a dense web-like structure that fills the background of the slide.

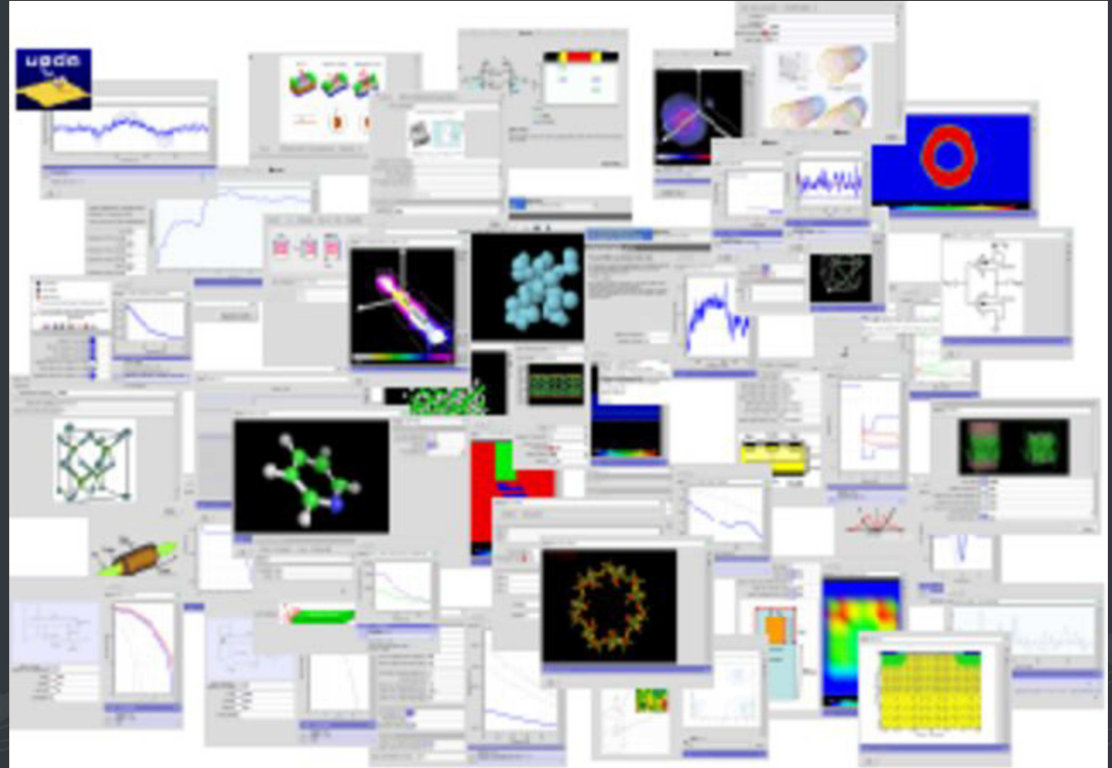
# **nanoHUB@home Expanding nanoHUB Through Volunteer Computing**

---

Steven Clark, Ben Haley, Martin Hunt, Nathan Denny: [HUBzero.org](https://HUBzero.org)  
David Anderson: UC Berkeley



- Collaborate
- **Simulate**
- Explore
- Learn



# PLATFORM FOR SCIENTIFIC COMPUTATION

- 1.5 million users annually
- 175000+ registered users
- 600+ published tools
- Varying resource requirements for tool execution
  - Mostly single core reasonably short duration
  - Several require multinode long walltime cluster access



# PLATFORM FOR SCIENTIFIC COMPUTATION

- GUI development environments
  - RAPPTURE
  - Jupyter Notebooks
- Source code management - GIT/subversion
- Rich Linux based development platform
- Powerful computing resources
  - Purdue community clusters
  - Stampede2 (TACC)
  - Comet (SDSC)
  - Open Science Grid (OSG)





---

# SIMULATION USE CASES

- **On demand**

- UI used to declare inputs for simulation
- Command line
- Single simulation or parametric sweep



---

# SIMULATION USE CASES

- **Uncertainty quantification**

- Inputs declared as distributions
- Statistical methods used to determine meaningful input combinations
- A simulation is run for each input combination
- Results include a response surface model which can be used to approximate simulation, sensitivity analysis, and probability distribution function (PDF) for outputs.



---

# SIMULATION USE CASES

- **Exploratory simulation**

- Automatically generate simulation input samples covering the space
- Allow for interactive selection of multidimensional input space
- Execute simulation for each sample



---

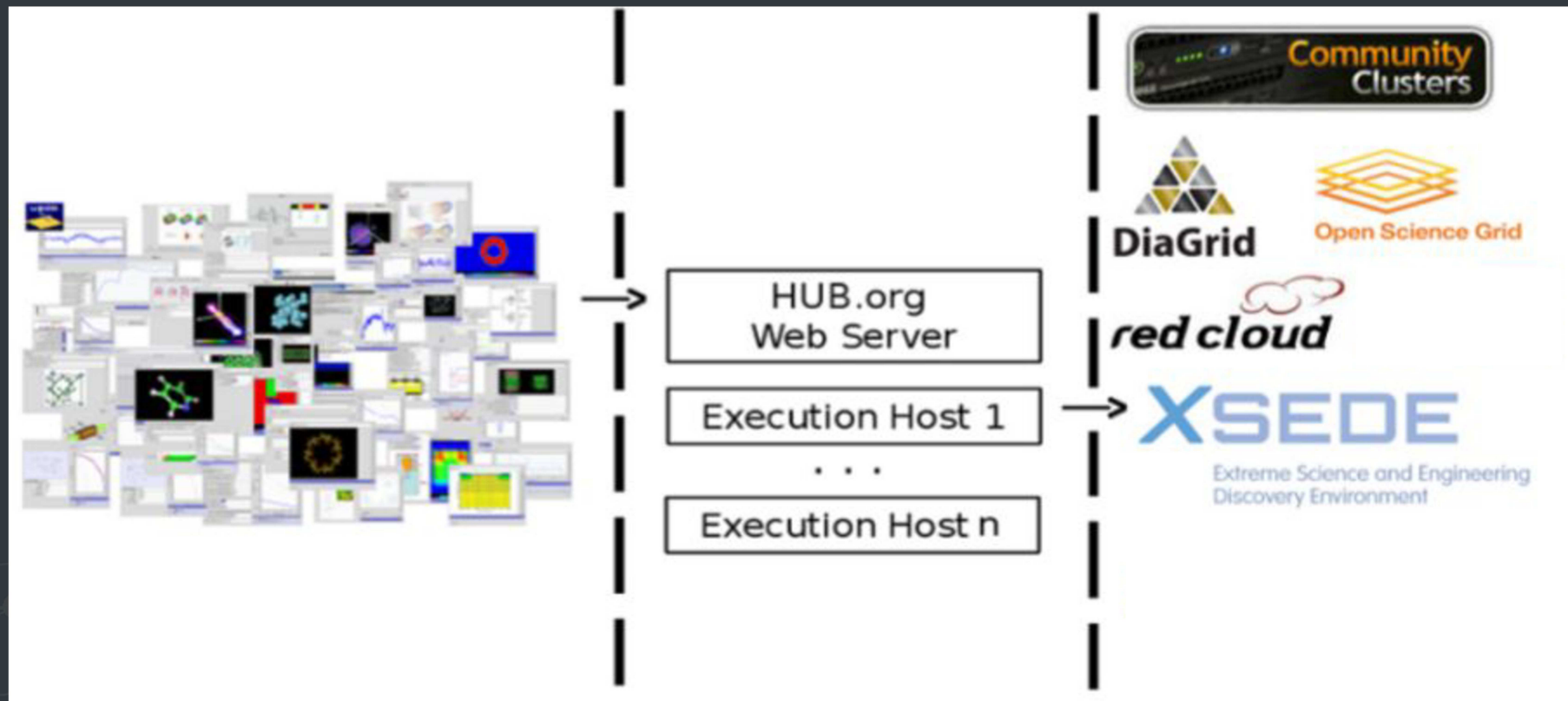
# SIMULATION USE CASES

- **Machine Learning**

- Examine simulation results discern patterns
- Build approximate simplified models
- Requires large quantity of simulation results



# TRADITIONAL COMPUTATION WORKFLOW





---

# SIMULATION RESOURCE DEMAND

- **Interactive Users**

- 100's active at any given time
- Short lived execution
- Time between runs for analysis

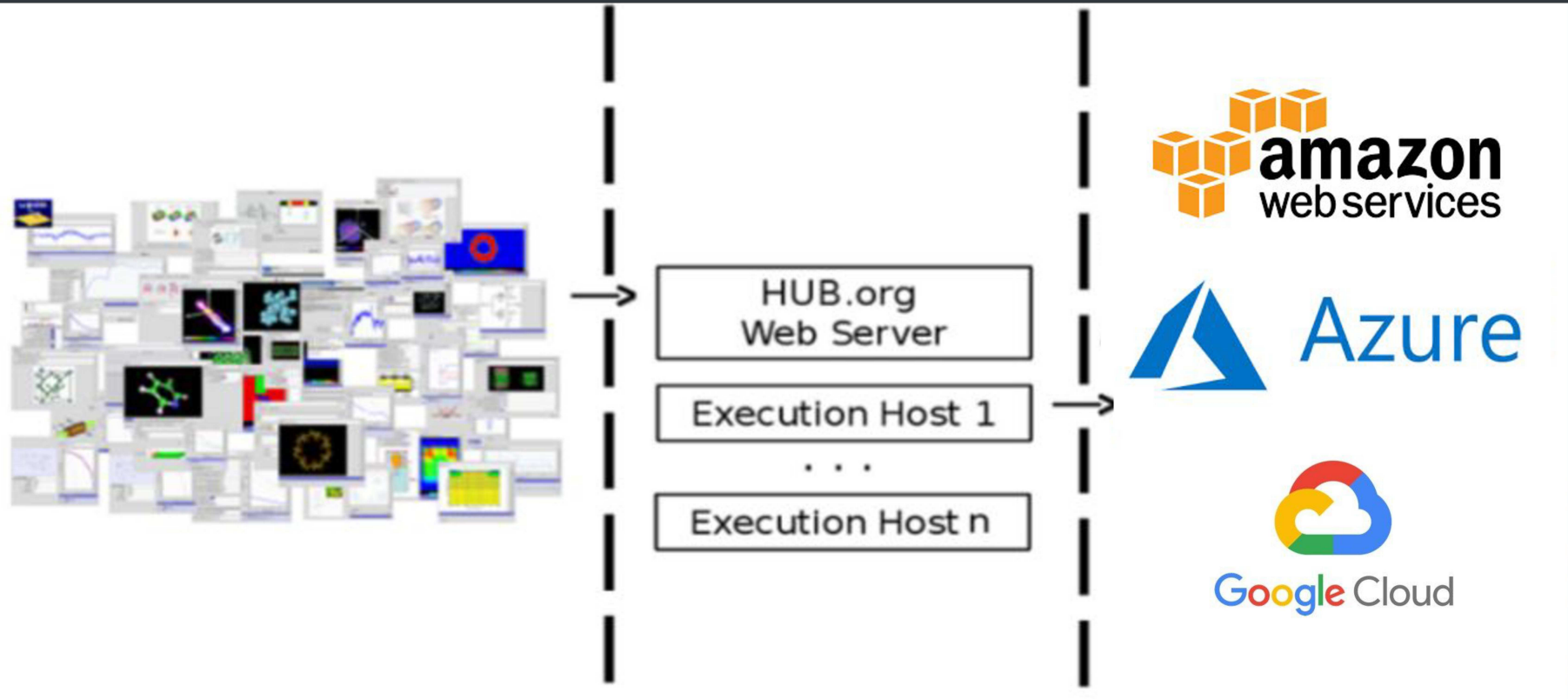
- **UQ**

- 10's - 100's of simulations requested for single run
- May be staged serially or run in parallel
- Long turn around pushes limits of user patience

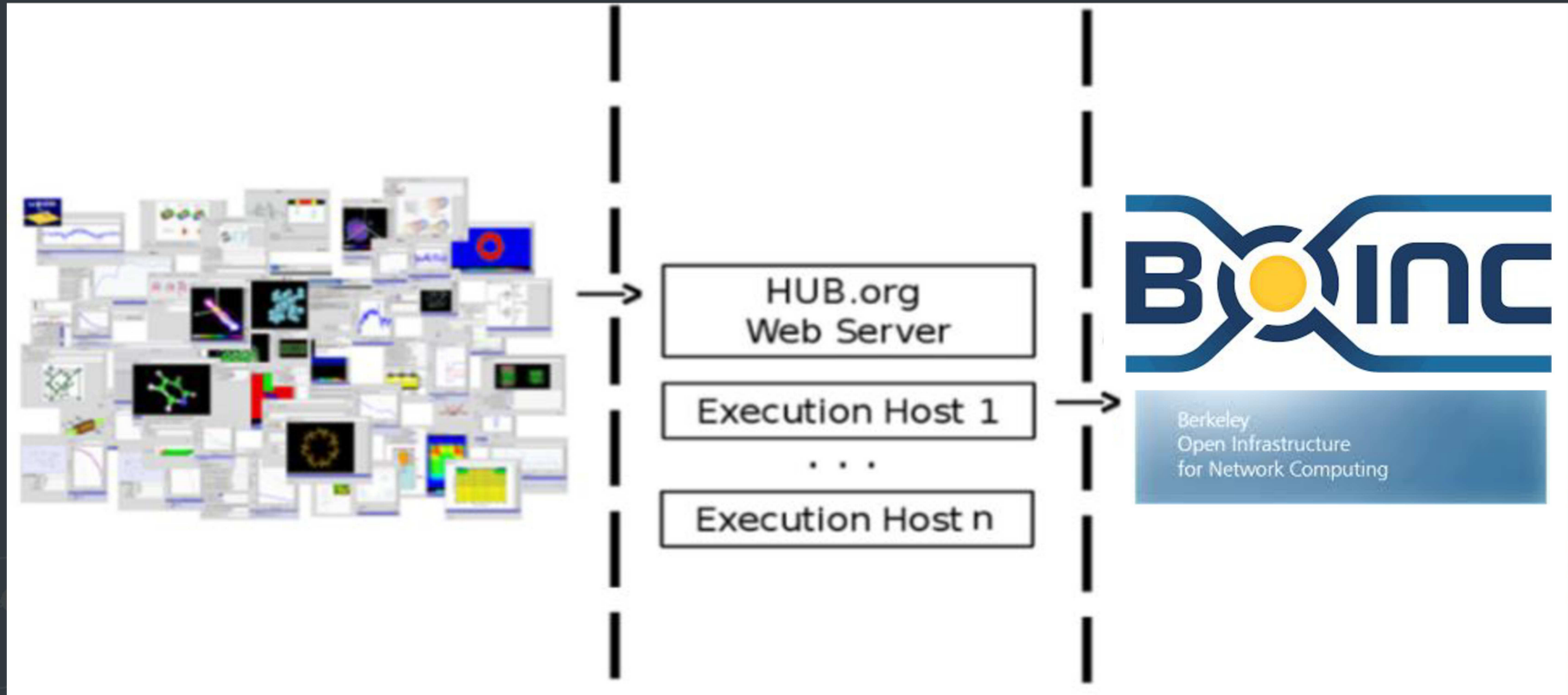
- **Exploratory, Machine Learning**

- 100's - 1000's of simulations requested as part of single study or experiment

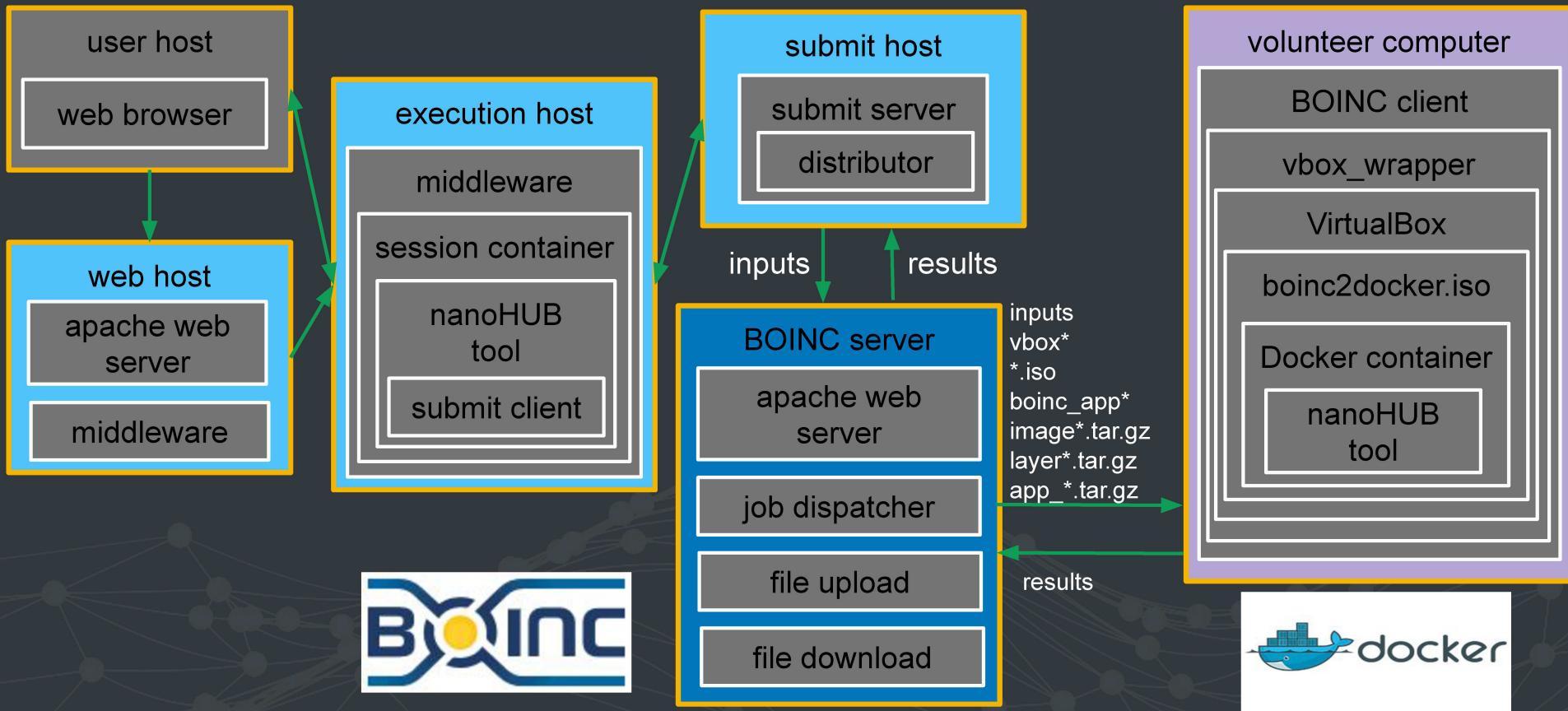
# CLOUD BASED COMPUTATIONAL WORKFLOW



# VOLUNTEER BASED COMPUTATIONAL WORKFLOW



# SUBMIT ARCHITECTURE - nanoHUB/BOINC



# SUBMIT/BOINC - INTEGRATION

- **nanoHUB Application**

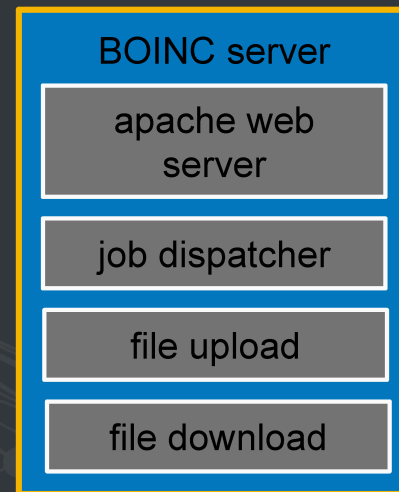
- UI
- User supplied data
- Application files

- **submit Server**

- Common interface between local and all remote resources (HPC + HTC)

- **BOINC Project Server**

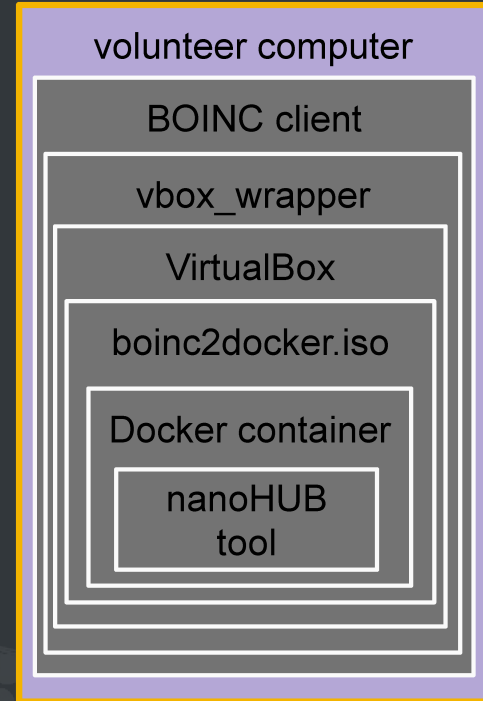
- Job execution manager for all BOINC submissions - submit\_api batch submission
- stage\_docker\_image - combines docker save and stage\_file to place tarballs in the download directory. Also creates nanoHUB specific VirtualBox initialization file and boinc\_app startup script.



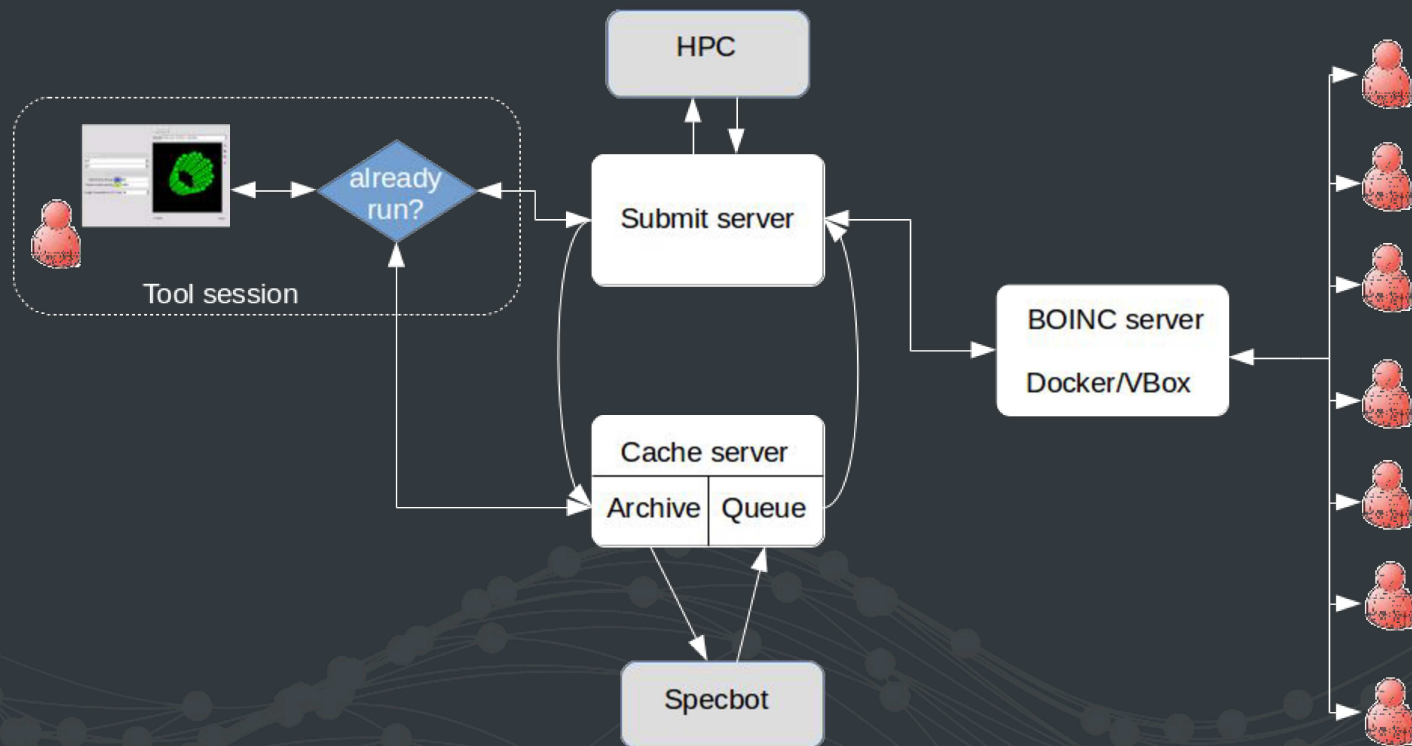
# SUBMIT/BOINC - INTEGRATION

- **Volunteer Host**

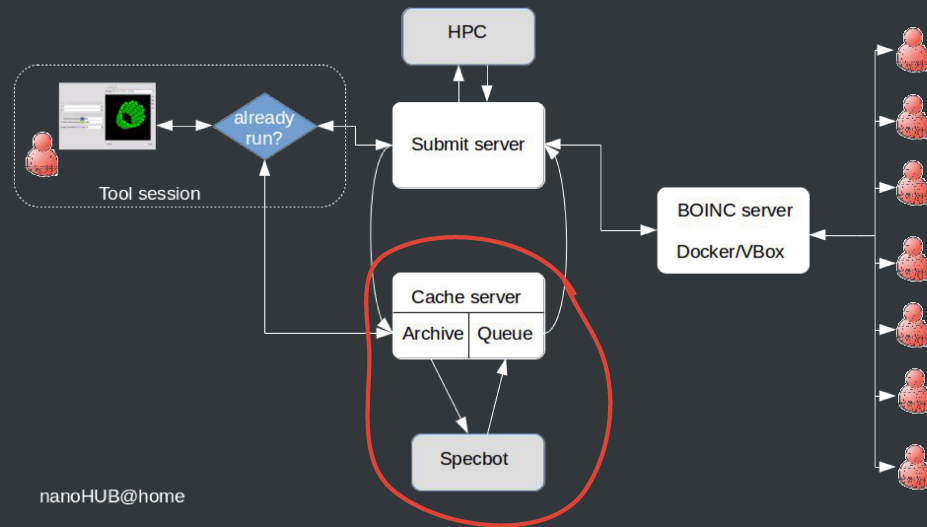
- BOINC client - download from BOINC website or system package install
- VirtualBox - nanoHUB applications run in Linux environment. VirtualBox provides access to Linux, Windows and MAC volunteer hosts.
- boinc2docker - introduction of docker containers allows simpler change management. Nine docker containers support 316 nanoHUB applications.
- Mounted volumes - allow for reduced memory requirement when loading docker container and greater flexibility creating new applications.
- nanoHUB application files sent as tar balls and are not removed at job completion to reduce bandwidth requirement
- User supplied data is also sent as a tarball but is job specific and is removed at job completion



# Exploratory Simulation - Architecture



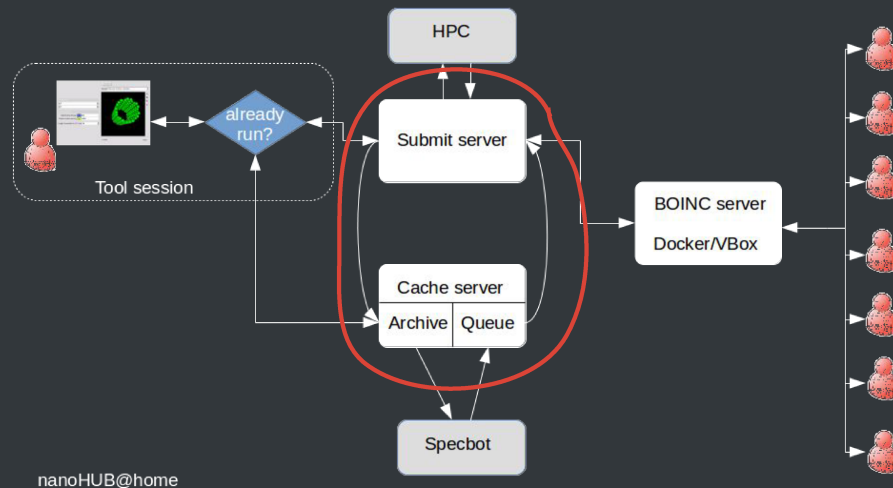
# Exploratory Simulation



- Rappture XML schema defines simulation inputs
  - Scalar numeric inputs
  - Fixed lower and upper bounds
  - Automatically generate simulation input samples spanning the range
- Input driver files are uploaded to cache server queue



# Exploratory Simulation



- Automated process pulls driver file inputs from cache server queue
- Assemble driver input files into BOINC batches
- Employ submit to execute BOINC batches
- When BOINC batch is complete publish successful results to cache server archive

---

# Exploratory Simulation - 10,000,000 jobs/yr ?

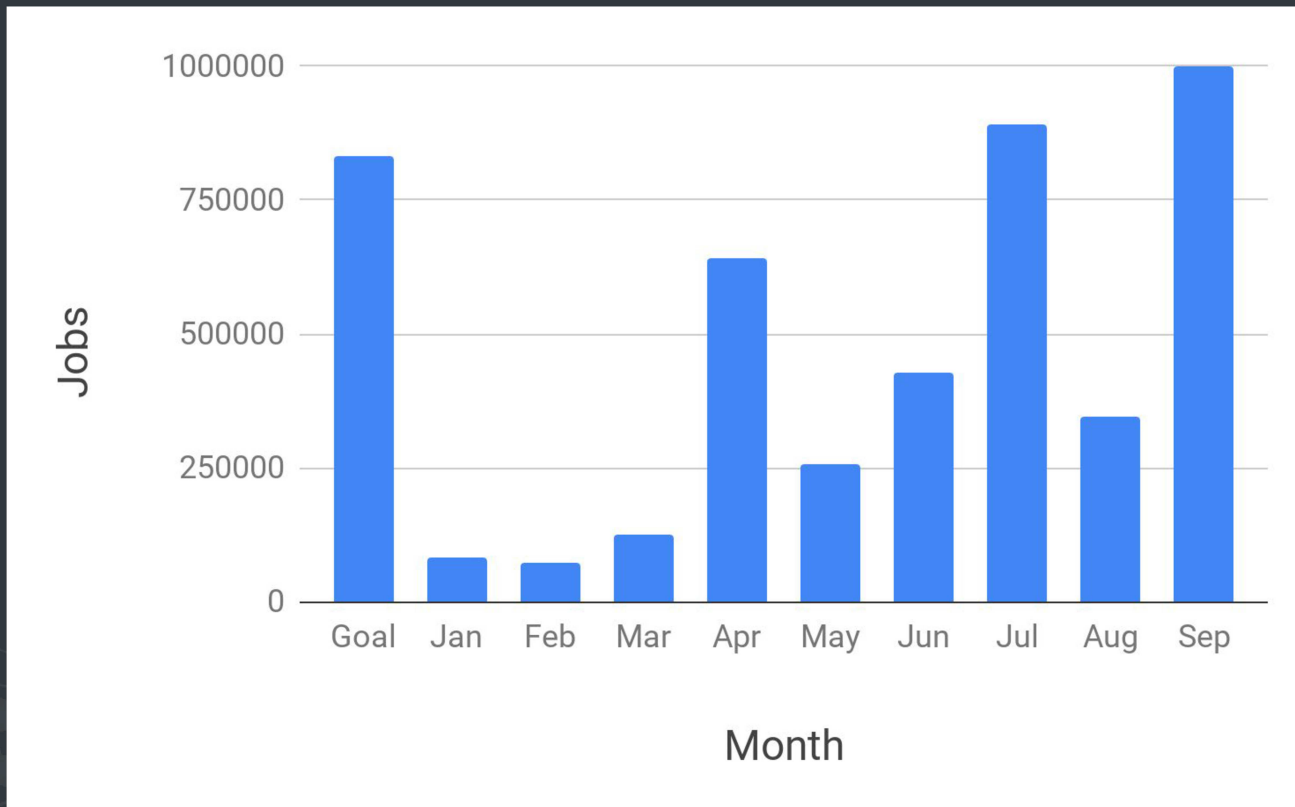
- **Infrastructure Problems**

- Number of Processes
- Disk quota
- Memory
- Database transaction processing

- **BOINC Problems**

- Large variation in nanoHUB tool requirements executing as a single BOINC application
  - Memory
  - Disk
  - Time

# SPECULATIVE SIMULATION RATE - jobs/month



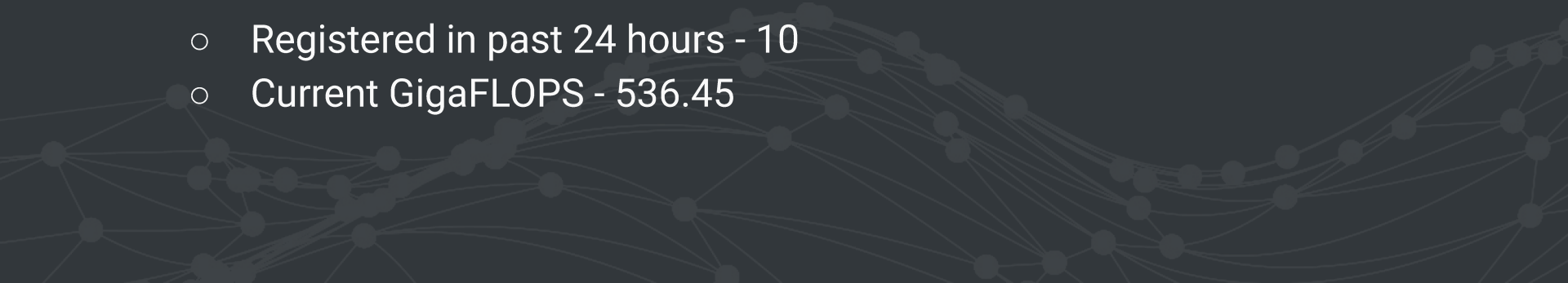
---

# BOINC COMMUNITY - SIZE

- **Users**

- With credit - 530
- With recent credit - 210
- Registered in past 24 hours - 41

- **Computers**

- With credit - 1192
  - With recent credit - 460
  - Registered in past 24 hours - 10
  - Current GigaFLOPS - 536.45
- 
- A decorative network diagram at the bottom of the slide, consisting of numerous dark grey circular nodes connected by thin, light grey lines, forming a complex web-like structure that spans the width of the slide.

The background is a dark blue-grey color with a complex, abstract pattern of thin, light grey lines and small dots. These lines and dots form a network-like structure, with some areas being more densely connected than others, creating a sense of depth and complexity. The overall effect is that of a digital or scientific network visualization.

QUESTIONS & ANSWERS

?