



Volunteer Computing in ALICE?

Ioannis Charalampidis, PH-SFT



Contents



1. **Introduction** to Volunteer Computing



2. **Implementation** of Volunteer Computing



3. **Future** of Volunteer Computing

But the main idea is: Learn from our mistakes and pick the best VC solution for ALICE!



Volunteer Computing

- Ordinary people voluntarily running scientific tasks on their PCs

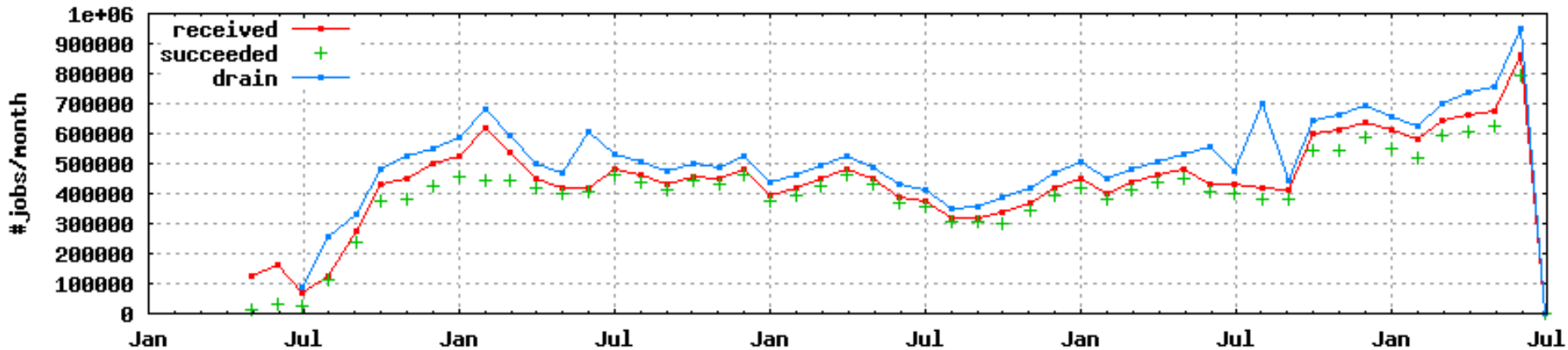




Volunteer Computing at CERN

1. Virtual LHC@Home (Test4Theory)

- Operates mainly on volunteer computing
- About 1k parallel jobs
- 2 Trillion events simulated since 2011





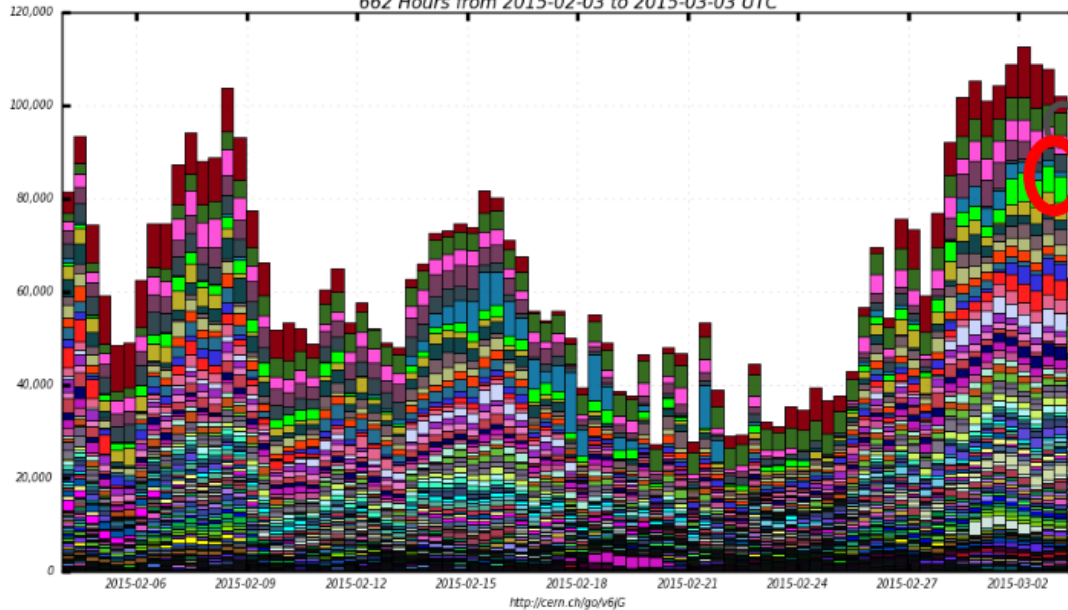
Volunteer Computing at CERN

2. ATLAS@Home

- Volunteer computing is their 2nd largest simulation site
- 4-5k parallel jobs

dashboard

Slots of Running Jobs
662 Hours from 2015-02-03 to 2015-03-03 UTC



Maximum: 112,630, Minimum: 0.00, Average: 63,358, Current: 62,935

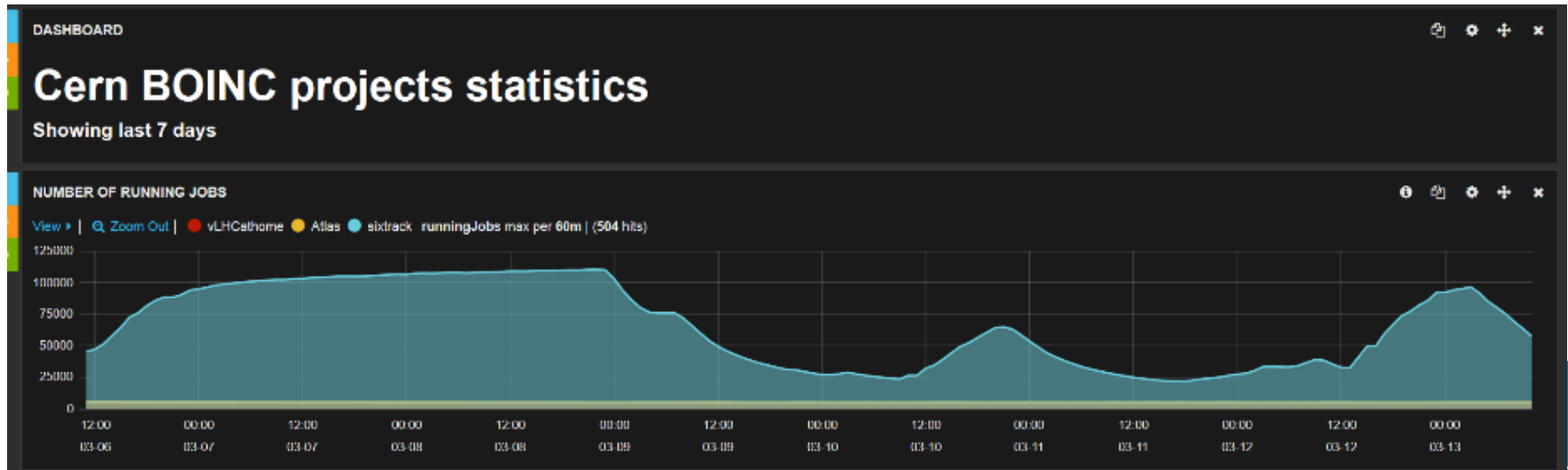




Volunteer Computing at CERN

3. Sixtrack

- Operates only on volunteer computing
- Up to 100k parallel jobs





Why Volunteer Computing?

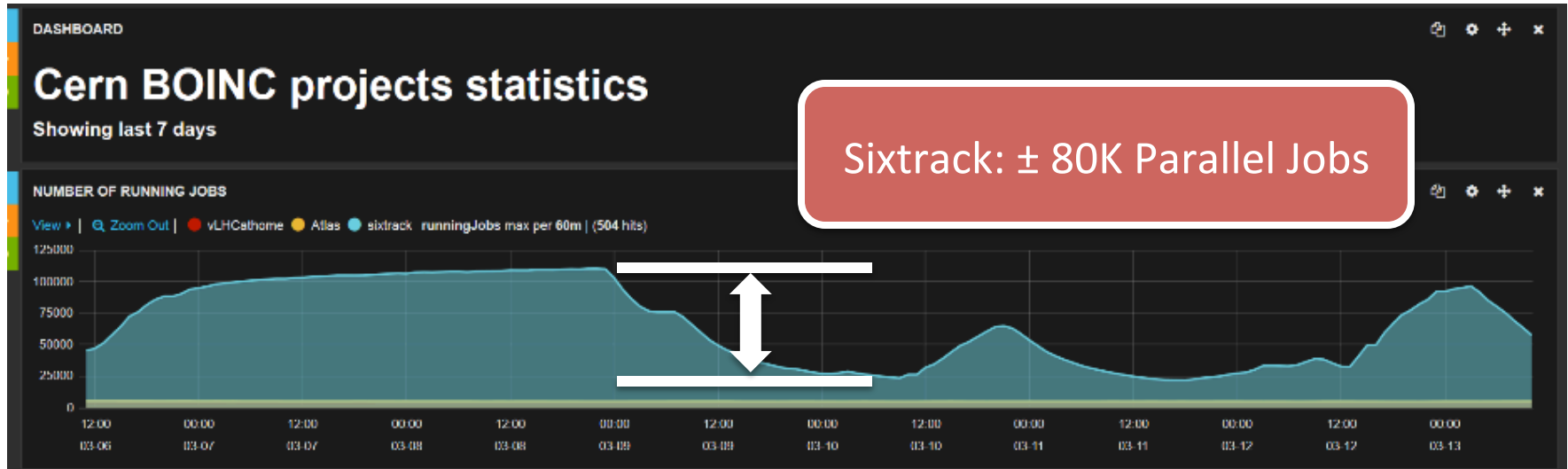
- **Free*** resources
 - 100K+ hosts achievable for large projects
 - Actual core count is higher
- **Engage** with communities outside HEP
 - Outreach and publicity for HEP and science





Understand the Environment

X Uncoordinated, opportunistic resources



- Good PR = More resources





Understand the Environment

X Inconsistent network configurations

- Assume worst-case scenarios
(bad broadband or mobile connection)

X Untrusted environment

- Assume user has full control on your job binaries,
and **not** always with good intentions
(no matter how hard you try to secure it)

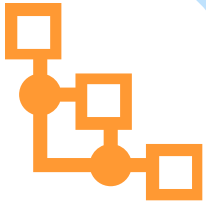




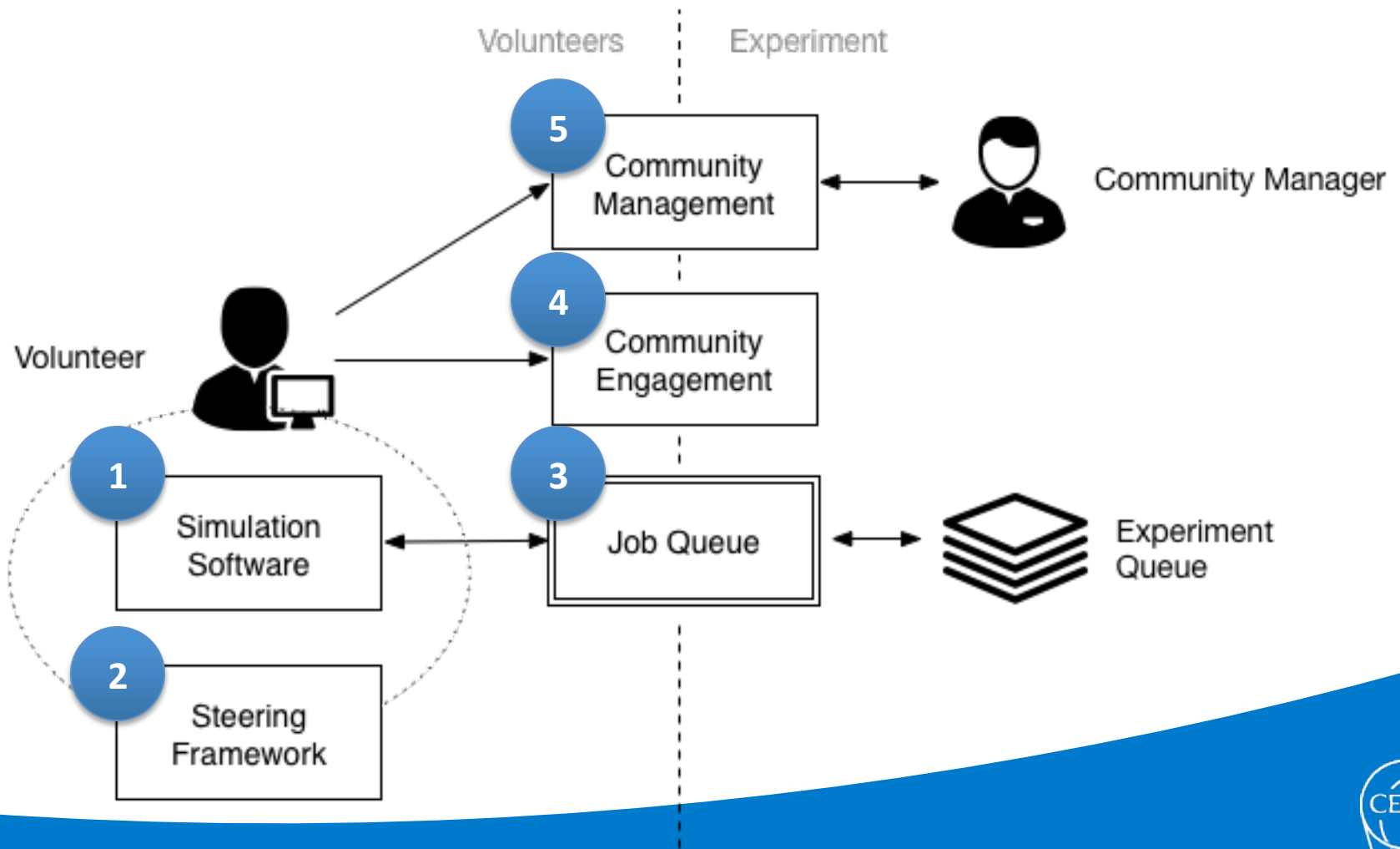
Volunteer Computing ...

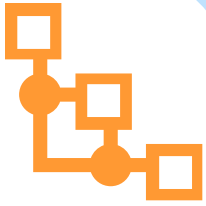
- Works well for
 - ✓ **Low-priority** jobs (ex. M-C simulations)
Assume the job will be processed in 'best-effort' manner
 - ✓ **Low I/O** jobs
Assume the bandwidth is limited
 - ✓ **Reproducible** jobs
You may need to run the same job multiple times for integrity checks





Ingredients of a V/C project





Pure BOINC Implementation

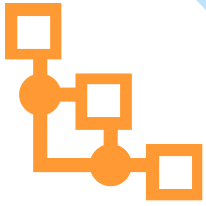
(Sixtrack)

Component	Implementation
1. Simulation Software	Cross-compile simulation binaries for every operating system (Windows, Linux, Mac OS)
2. Steering Framework	BOINC Client
3. Job Queue	BOINC Scheduler
4. Community Engagement	BOINC Website
5. Community Management	Common BOINC-Project CERN Website

✓ **Pros:** Works out of the box

✗ **Cons:** Cross-compiling for every platform!

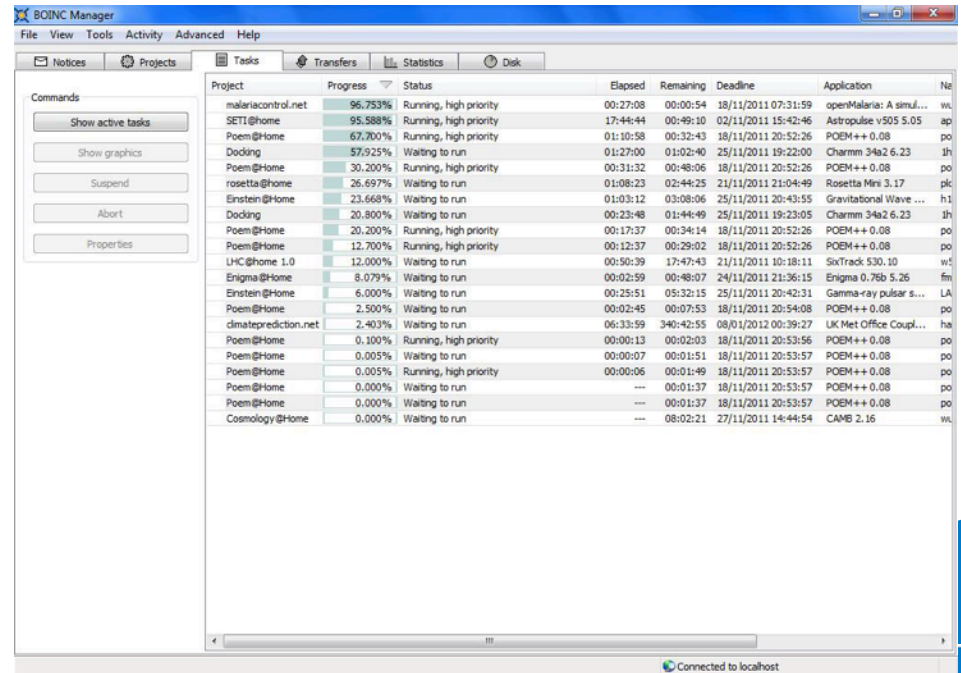
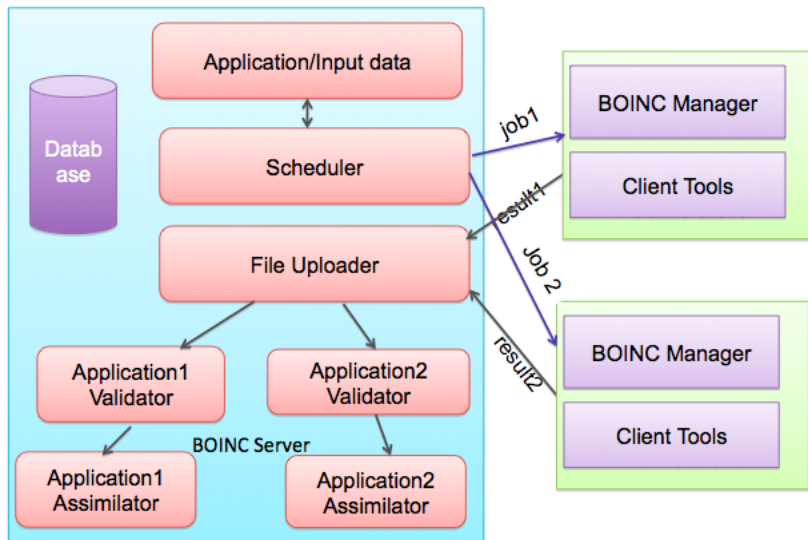


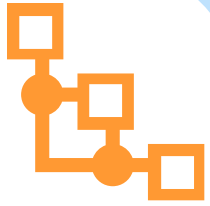


(What is BOINC?)

- Berkeley Open Infrastructure for Network Computing (**BOINC**)

BOINC Architecture



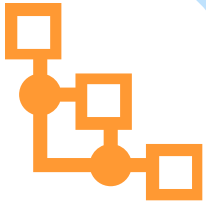


(What is BOINC?)

- **But...**

- National Science Foundation (NSF) **ceased funding** of BOINC at May 2015
- BOINC software moving towards a **community support** model
- *There are some positive signals for new round of NSF funding...*





BOINC + Virtualization

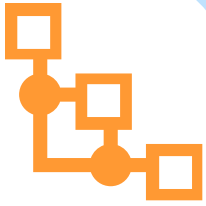
(ATLAS@Home)

Component	Implementation
1. Simulation Software	VirtualBox + CernVM + Standard Linux Binaries
2. Steering Framework	BOINC Client
3. Job Queue	BOINC Scheduler
4. Community Engagement	BOINC Website
5. Community Management	Common BOINC-Project CERN Website

✓ **Pros:** Works ootb, **No need to cross-compile**

✗ **Cons:** BOINC Spawns 1 VM per task, BOINC scheduler not designed for heavy I/O





BOINC + Virtualization + Co-Pilot

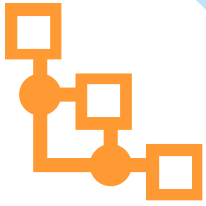
(Test4Theory)

Component	Implementation
1. Simulation Software	VirtualBox + CernVM + Standard Linux Binaries
2. Steering Framework	BOINC Client
3. Job Queue	Co-Pilot
4. Community Engagement	BOINC Website
5. Community Management	Common BOINC-Project CERN Website

✓ **Pros:** Works ootb, No need to cc, **One VM per project**

✗ **Cons:** BOINC Credits not working, Co-Pilot not maintained and not supported by CERN IT





BOINC + Virtualization + DataBridge

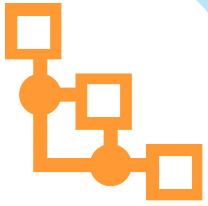
(CMS@Home)

Component	Implementation
1. Simulation Software	VirtualBox + CernVM + Standard Linux Binaries
2. Steering Framework	BOINC Client
3. Job Queue	Data-Bridge
4. Community Engagement	BOINC Website
5. Community Management	Common BOINC-Project CERN Website

✓ **Pros:** Same as before + **supported by CERN IT**

✗ **Cons:** BOINC Credits not working

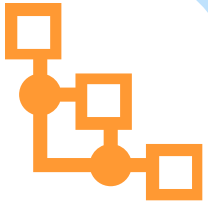




IT-Supported Services

- **BOINC** Infrastructure
 - Installation, Maintenance, Monitoring of software
 - MySQL Database + NFS Storage
- **Drupal** Portal
 - Project Website & Community Management tools
- ***Ongoing: Working towards a common Job Queue solution***
 - *Feedback is welcome*





Implementation Effort

- **Package** your software
 - Put your software in CVMFS
- **Integrate** to the experiment's scheduler
 - Write an adapter for connecting to the VC queue
- Find a **community manager**
 - Someone has to monitor the forums and reply to questions!





Towards a Better V/C Experience

- There are a couple of **loose ends** with the BOINC implementation
 - BOINC was not meant for Virtualization
 - Credits are not properly counted
 - Virtual Machines are started in place of processes
 - BOINC is only used for launching the VM
 - BOINC targets more ‘geeky’ audience
 - We are restricting the overall audience





Towards a Better V/C Experience

- Why couldn't we just ...
 - Visit a website and have a seamless **click-to-join** experience?
 - Launch **one VM**, and spawn multiple projects inside it?
 - Let projects decide the way they want to give **credits** to the users?






CERN Computing Challenge

The Virtual Machine is ready Log-in and keep track of your progress Public Beta

Event Rate	Progress	Jobs Completed	Ranking
How many events (particle collisions) your virtual machine is simulating per minute.	What fraction of the current job is completed. (Typically, 1 job = 100,000 events)	How many simulation jobs your virtual machine has processed.	Your ranking among the volunteers in this challenge, by number of jobs completed.
500 0	1 % 0 %	1 0	1 0

You are now ready to start computing

By clicking the *Start* button below, we are going to start a Virtual Machine in your computer, which is going to start performing virtual collisions and sending the statistics back to CERN. Click **Log-in to track your progress** on top right corner keep track of your progress!



Start ⚙️ 🗑️

Credits





CERN Computing Challenge

- A test-bed for the **latest** VC technologies:
 - **CernVM WebAPI** : Control VMs from the web
<https://github.com/wavesoft/cernvm-webapi/wiki>
 - **DumbQ** : A ‘Dumb’ scheduler for running multiple projects in the same Virtual Machine
<https://github.com/wavesoft/dumbq>
 - **CreditPiggy** : An on-line piggy-bank for keeping the user’s contribution
<http://creditpiggy.cern.ch>





CERN Computing Challenge

Component	Implementation
1. Simulation Software	VirtualBox + CernVM + Standard Linux Binaries
2. Steering Framework	CernVM WebAPI + DumbQ
3. Job Queue	Data-Bridge
4. Community Engagement	CreditPiggy
5. Community Management	Project CERN Website

✓ **Pros:** Same as before + **smooth experience for all users**

✗ **Cons:** Not using BOINC

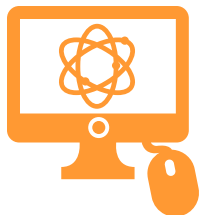




Questions ?

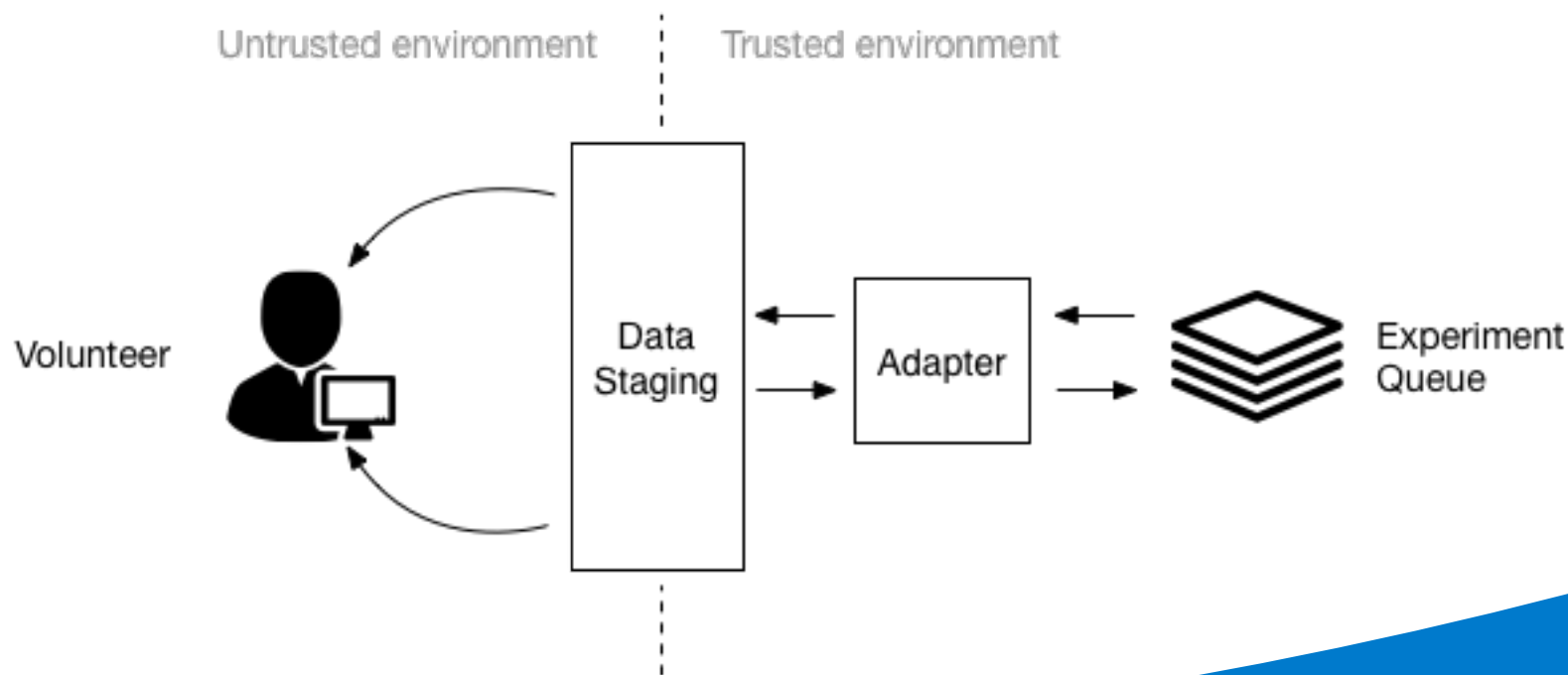
Spare Slides

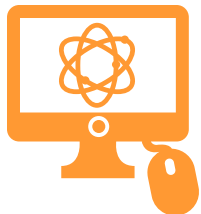




Detailed look on the Queue

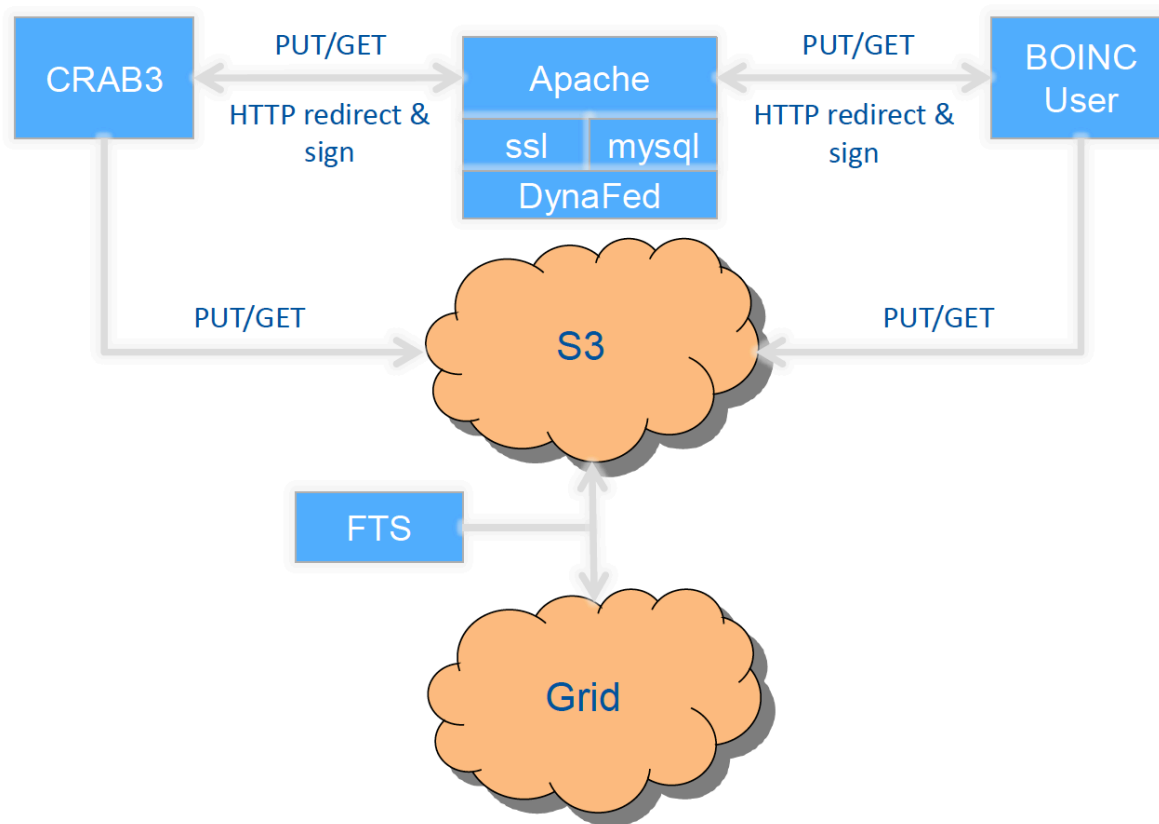
- Queue **components** in Volunteer Computing

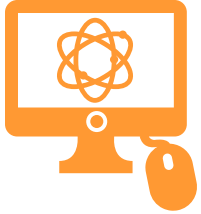




Detailed look on the Queue

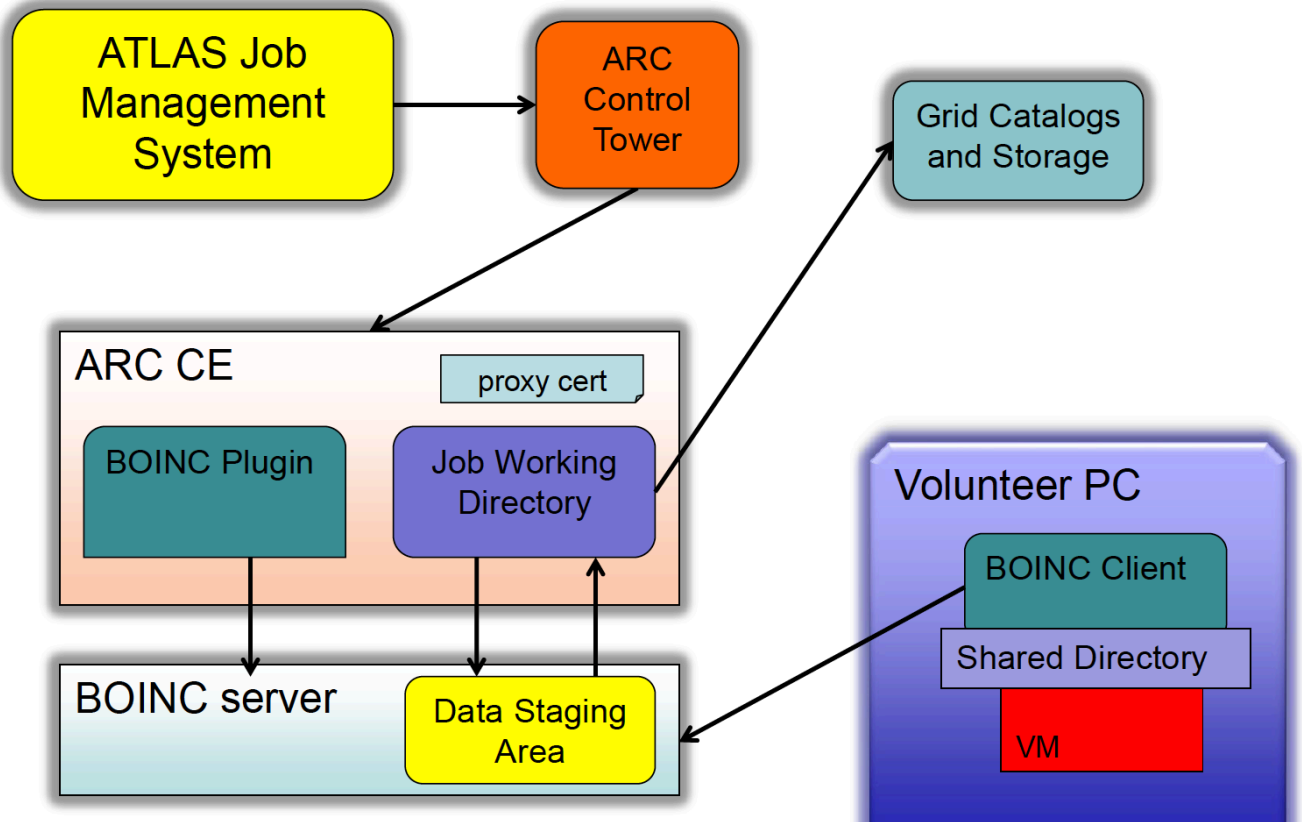
- CMS implementation using **DataBridge**

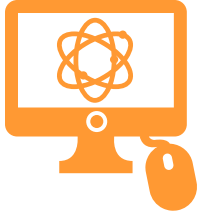




Detailed look on the Queue

- ATLAS implementation using **BOINC**





The CERN60 Challenge

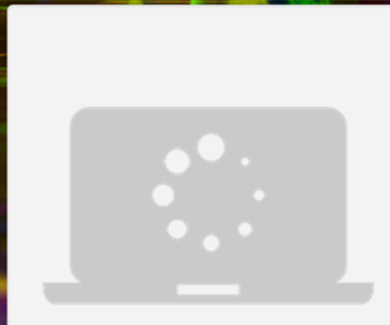
The Virtual Machine is booting

Event Rate

How many events (particle collisions) your virtual machine is simulating per minute.

500

0



Starting virtual event generator

John Charalampidis Progress details Public Beta

Event Rate	Progress	Jobs Completed	Ranking
How many events (particle collisions) your virtual machine is simulating per minute. 500 0	What fraction of the current job is completed. (Typically, 1 job = 100,000 events) 1 % 0 %	How many simulation jobs your virtual machine has processed. 1 0	Your ranking among the volunteers in this challenge, by number of jobs completed. 1 0

Learn more:

- Introduction to high energy physics simulations
- See the simulations produced by your computer
- Learn about the software that does the simulations

Generator: pythia8

Collisions analyzed with: ATLAS_2011_S9131140

Beams: Starting virtual event generator Energy: 2000.00 GeV Producing: 100,000 events

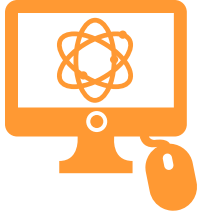
While waiting... wanna help ATLAS look for Higgs?

Stop [Settings] [Trash]

<http://test4theory.cern.ch/vlh>

“Let’s reach the **LHC** Event Rate”





The CERN60 Challenge

HALL OF FAME

Here is a name cloud of all the volunteers who signed in and contributed at least one job. Thanks to you, and t

Patrick Werber Hendrik Richter Michael Claes Indy Gilstrøm Jaime Farill Mike Hamilton Kirsten
W Hallam Albert Booth Evert adsgafg Ryder Bluhm Harrison Totty Nicolò Gottardello Neil M
Katz Wade Gillingham nekdo z jamian Dimock Michelle Greenlee Lex Imperatoris Billy Vier
Damme Justin Pekular Jason Lorsung Joshua Lee AXfactOR AXfactOR Giovanni Siragusa Carson T
Marek upinsmoke1973 Maelstrom Scott McDermott Ryan Ford Naomi Cathcart Brian Bond Geoff
Blapkinz Ian Peter Braun RJ Hill Bertil Spolander Cody Wang Carl Michael Stojanovic duG Varrette Arturo Saura Ben Segal Eraip Ersoy Alex
Copero Tim Weinert Jan Füsting thomas noé Leo Wright Joe Tursi david9000 Borja González Herrero Benjamin Bertrand Pieter
Orlando Andrew Smith John Jones Rouslan Korneychuk Jasper Homann Matthew Snow Sylvain
Manzi Elemental Brain Adam Paugh Ninette Kelly Cameron Phillips Erick Erickson Emma
Bidema Ex Cool Francois Grey Clara Á. Luna Jim 'Artless' Merrill Mikey Babb Chris I
Engebreetsen Adrian Ellingsgaard Jonathan Chan Aika Code Arthur Molnar Clemens Der-ganze
Anhgarin Phil Webb Greg Lockett Dubois Alexandre Rohan Kundu Kostas Rakitzis Ayush Jha José Rafael
(zeroXten) Amkazan Amkazan Jan-niclas Graumann Not Erison Veshi Matthew Gregg Simone G
Poe Felix Wustrack Ivan Koeff Daniel Reynolds Jacob Wolf Nel-gez Andreas Roy van Olenbaekes ymash101 Robby Wilson Johan
Torselius Alexandr Samsonov Steve Kuntz Mike Bentley Daniel Lindmark mininukewarrior Ivan Bisol Frederik Vander Biest Thomas
Kristensen Marco Nardes Craig Pekar Mike Wilkinson John Smith Bernie Telalovic Kilazur Dargnaith Laurence Woolford Guillaume
Toussaint Adam Wheeler Jade Woods Nicolas Segal Anders Hvidberg Frandsen James Alexander Gardner Efi Psomopoulou Kalle
Bexhorn Joseph Flagler Ben Vaughn Anton Stensgaard Jakob Rigsby Eric Plummer Josh Robbins Ryan Moodey Kristof Ballet Jack Eden
Patrick Dan Stenhouse Michael Ben Drake
Isaza
Boch Karevinads Kristensen Conny francis Marco Vithuuh bunbv Joshua Karim (Samintell) Kon
M

16.000 sessions

8.000 users

108 countries – 90 languages

Avg. of 400 sessions per day

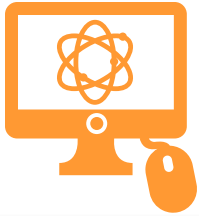
5,400 people booted a VM

1,100 had problems

(about 80%: slow network)

During the challenge detailed analytics were collected





The CERN60 Challenge

