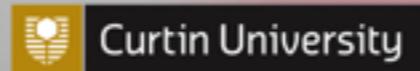




International
Centre for
Radio
Astronomy
Research

Citizen Science

Kevin Vinsen (ICRAR, UWA)
Astroinformatics 2013



THE UNIVERSITY OF
WESTERN AUSTRALIA
Achieving International Excellence



Happy to take questions as we go alone



Happy to take questions as we go alone



BUT

I'm "hard of hearing" - I wear hearing aids
so please speak clearly.

Ladies - I'm sorry I don't hear higher
frequencies very well at all.

I'm not ignoring you - I just can't hear you.



What is Citizen Science?

- Public participation in scientific research
- Long history
 - The Audubon Society's Christmas Bird Count, which began in 1900
 - The American Association of Variable Star Observers has gathered data on variable stars since 1911



More Recently

- Many famous astronomy projects
 - SETI@Home
 - GalaxyZoo
 - theSkyNet
 - theSkyNet POGS
 - Zooniverse

More Recently

- Many famous astronomy projects
 - SETI@Home
 - GalaxyZoo
 - theSkyNet
 - theSkyNet POGS
 - Zooniverse





Two Types

- Passive - just donate cycles
 - SETI@Home
 - theSkyNet POGS
- Active - user actively involved
 - Zooniverse

http://www.gala...oo.org/classify +
http://www.galaxyzoo.org/classify

Most Visited ICRAR Uni DET GWT Javadoc Banks Astronomy Virgin Wikipedia Slashdot The Australian New Scientist 128 km Perth Ra... Comics Bookmarks

EN · Galaxy Zoo is a ZOO NIVERSE project ...just like MOON ZOO

GALAXY ZOO HUBBLE

Home How To Take Part My Galaxies Contact Us Profile Logout



Invert galaxy image Add to my favourites

Take a Quiz

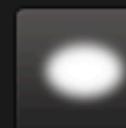
Would you like to take a Quiz?

Hide

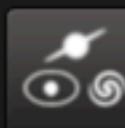
Classify galaxies

Answer the question below using the buttons provided.

Is the galaxy simply smooth and rounded, with no sign of a disk?



Smooth



Features or disk



Star or artifact

Galaxy Zoo Quick Links

- Classify
- How To Take Part
- Galaxy Zoo Forum
- Galaxy Zoo Blog

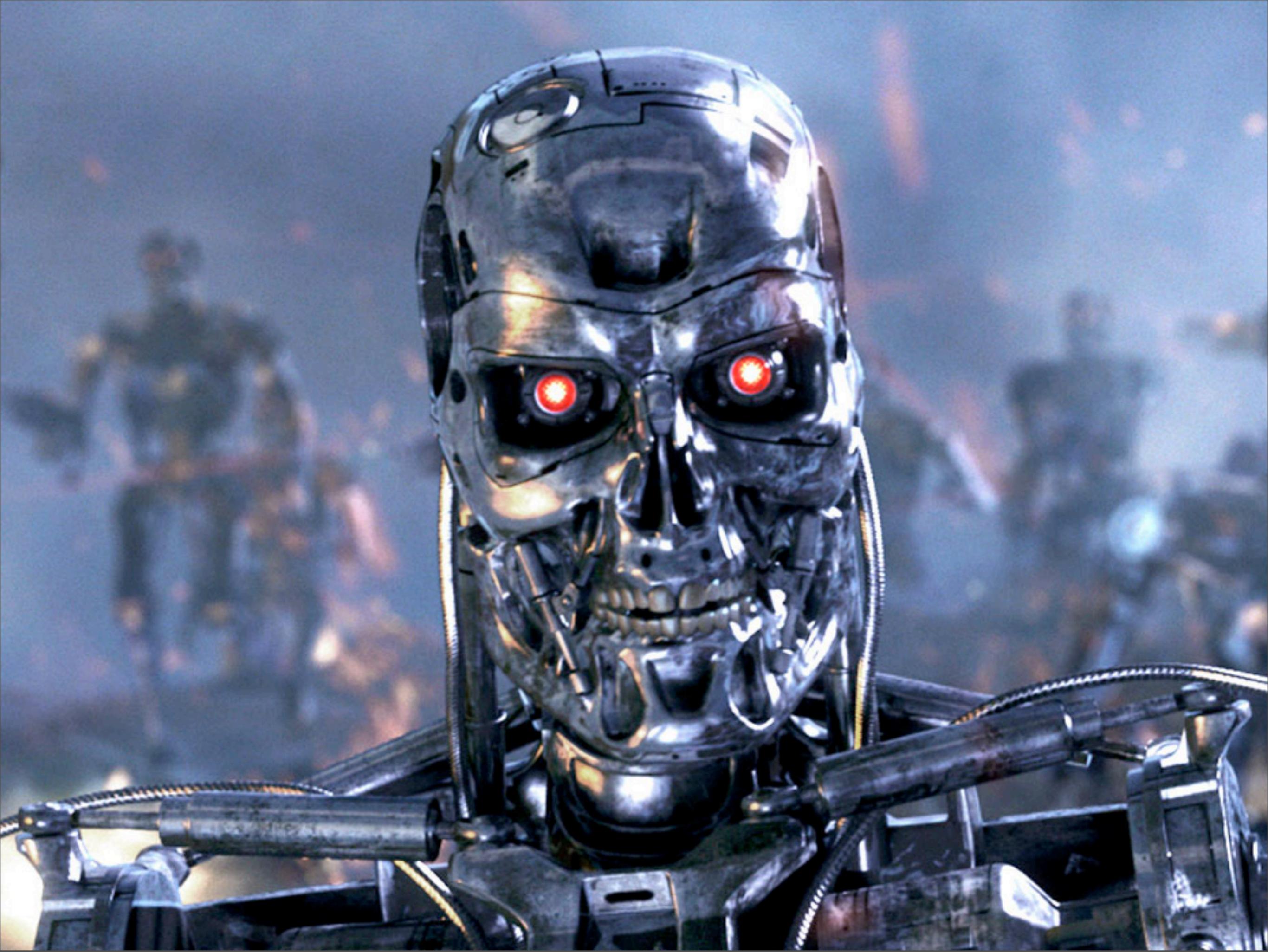
Astronomy Links

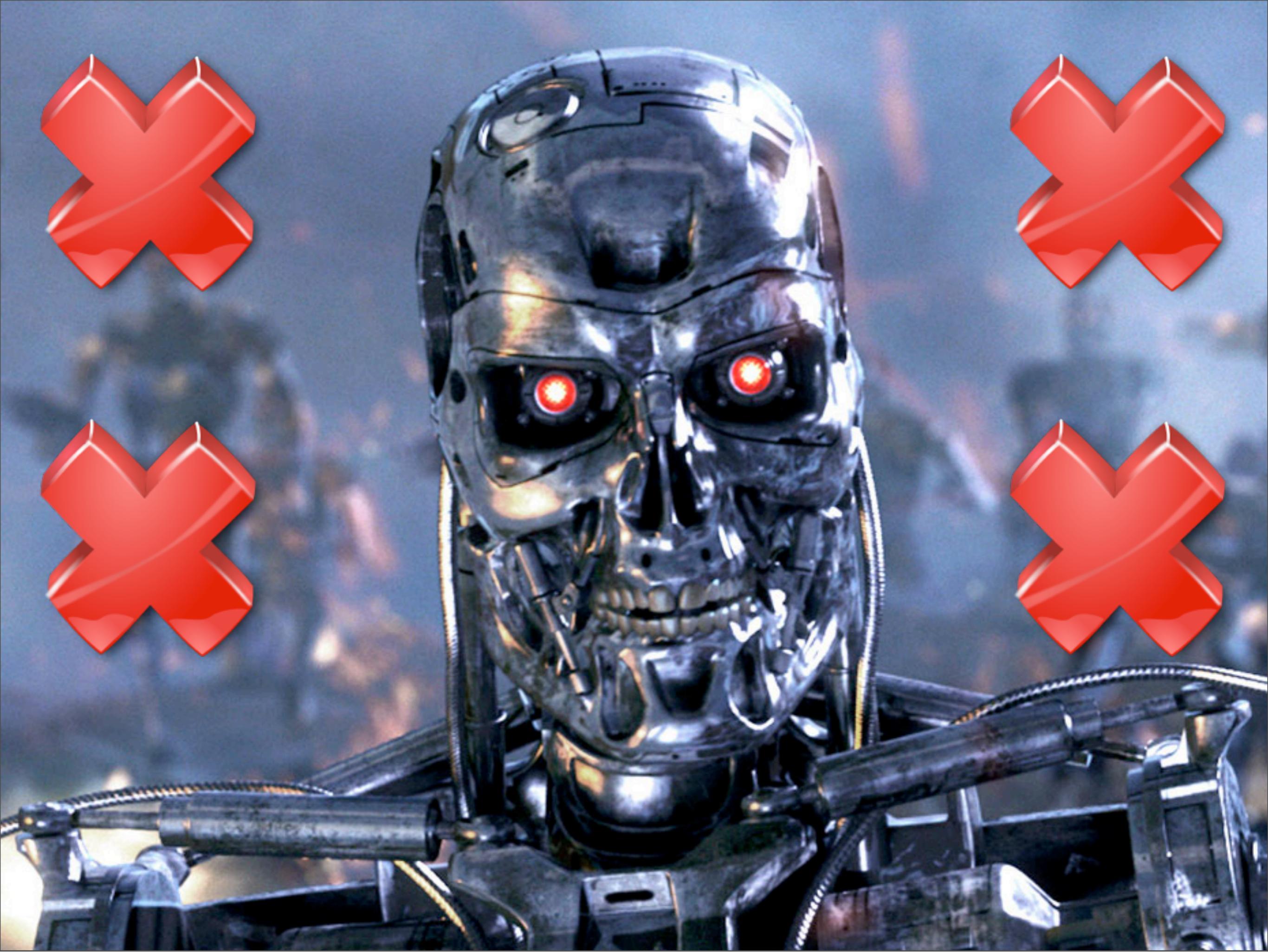
- Sloan Digital Sky Survey
- SDSS Database Access
- Oxford University
- University of Nottingham



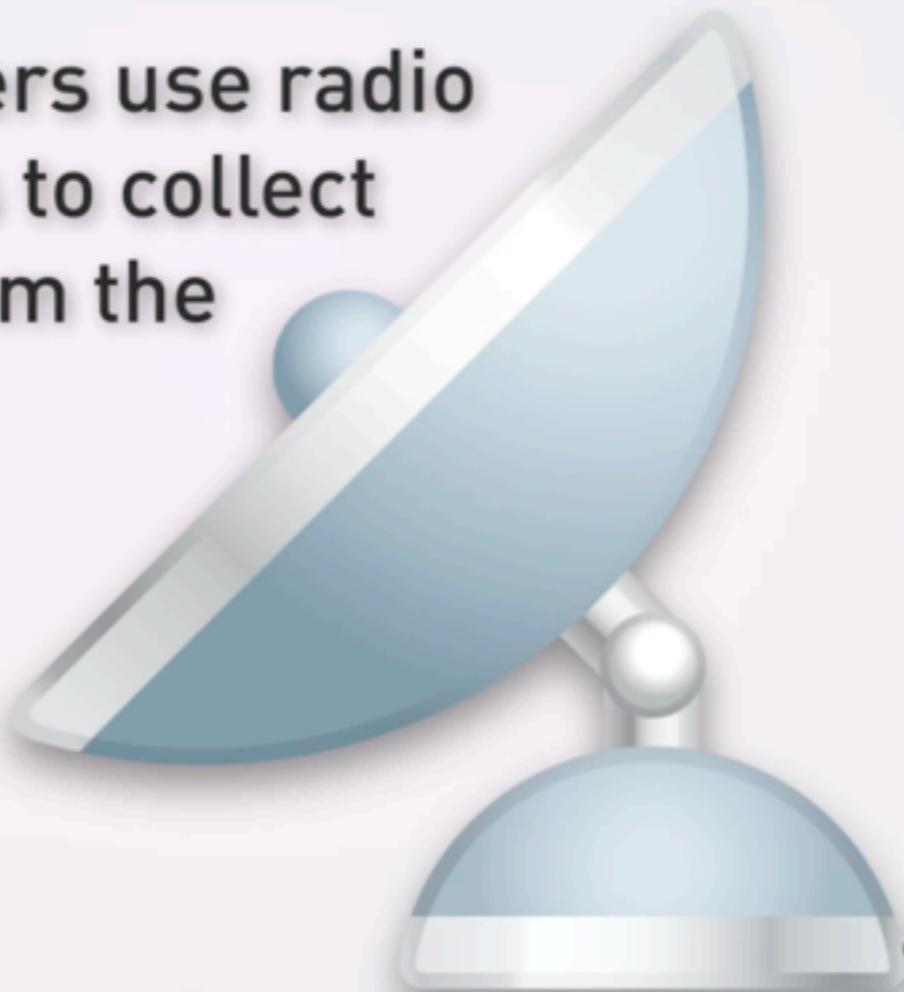
What is theSkyNet?

In Terminator, due to its massive computing needs and to protect itself from direct attack, theSkyNet utilised a large network of computers that would be nearly impossible to deactivate completely.





Astronomers use radio
telescopes to collect
signals from the
Universe





theSkyNet POGS

theSkyNet POGS - the PS1 Optical Galaxy Survey



About theSkyNet POGS - the PS1 Optical Galaxy Survey

theSkyNet POGS is a research project that uses Internet-connected computers to do research in astronomy. We will combine the spectral coverage of GALEX, Pan-STARRS1, and WISE to generate a multi-wavelength UV-optical-NIR galaxy atlas for the nearby Universe. We will measure physical parameters (such as stellar mass surface density, star formation rate surface density, attenuation, and first-order star formation history) on a resolved pixel-by-pixel basis using spectral energy distribution (SED) fitting techniques in a distributed computing mode. You can participate by downloading and running a free program on your computer.

theSkyNet POGS is based at The International Centre for Radio Astronomy Research.

- Images you have processed
- Images for all the Galaxies used in the survey
- [Link to page describing your research in detail]
- [Link to page listing project personnel, and an email address]

Join theSkyNet POGS - the PS1 Optical Galaxy Survey

- Read our rules and policies
- This project uses BOINC. If you're already running BOINC, select Add Project. If not, download BOINC.
- When prompted, enter <http://ec2-23-23-126-96.compute-1.amazonaws.com/pogs/>
- If you're running a command-line version of BOINC, create an account first.
- If you have any problems, get help here.

Returning participants

- Your account - view stats, modify preferences
- Server status
- Teams - create or join a team
- Certificate
- Applications

Community

User of the day



Pawn-D8-7

Born 1951 jun 22 at N69°05'33"E16°47'27"

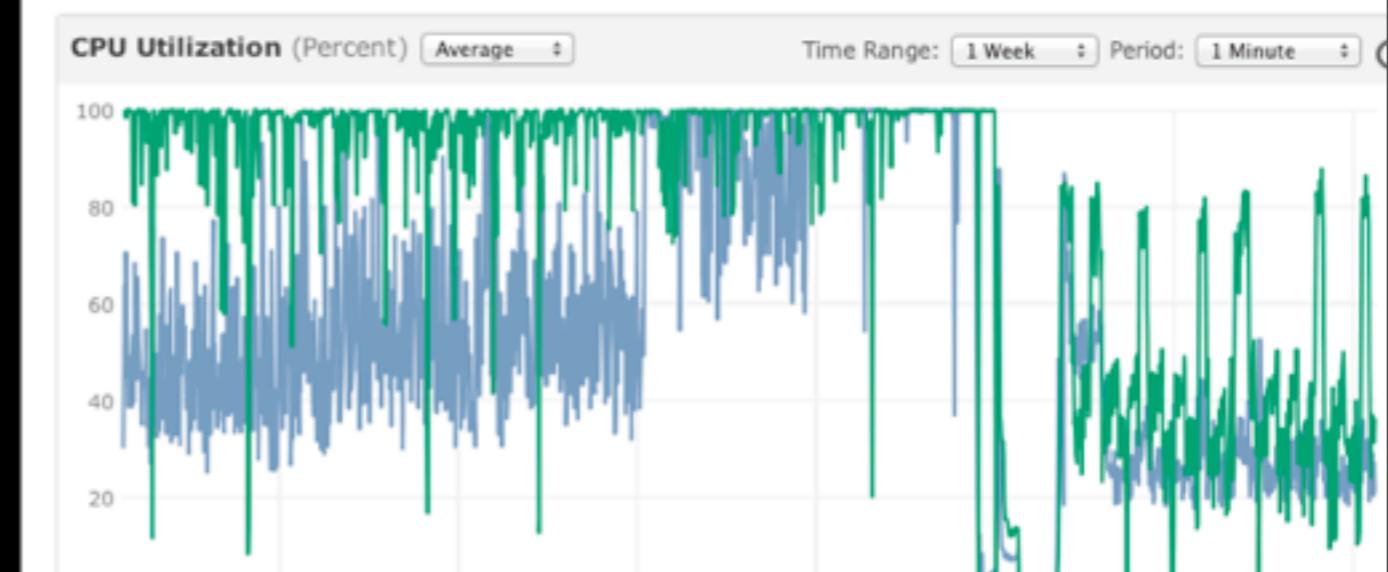
My profession is mechanical engineering
I am shooting as a hobby. Metallic...

News

More details on the disk crash

The CPU stats for the last week are shown in the graph before

CloudWatch Monitoring Details





BOINC

- Berkeley Open Infrastructure for Network Computing
- Open Source middleware system for volunteer computing
- Scalable up to 1.36 Million users (SETI@Home)
 - Total of 2.5 Million users on projects reporting to BOINC Stats
- *nix servers only
- All BOINC projects 8.3 PFlops of computing power

BOINC



Titan
US\$97 million
17.6 PFlops



Pan-STARRS1 Optical Galaxy Survey (POGS)

- Pixel-by-pixel spectral energy distribution fitting
 - Local stellar mass surface density
 - Star formation history
 - Age
 - Extinction
 - Dust attenuation
- Start with ~100 million pixel SEDs
 - Each pixel SED takes between 5 and 10 minutes
 - It would take between 950 and 1,900 years on a single core

Caveat

- This is not exhaustive - just to give you a flavour of what can be done



Every since they invented cloud computing,
I'm getting data stuck between my toes!



Scaling BOINC

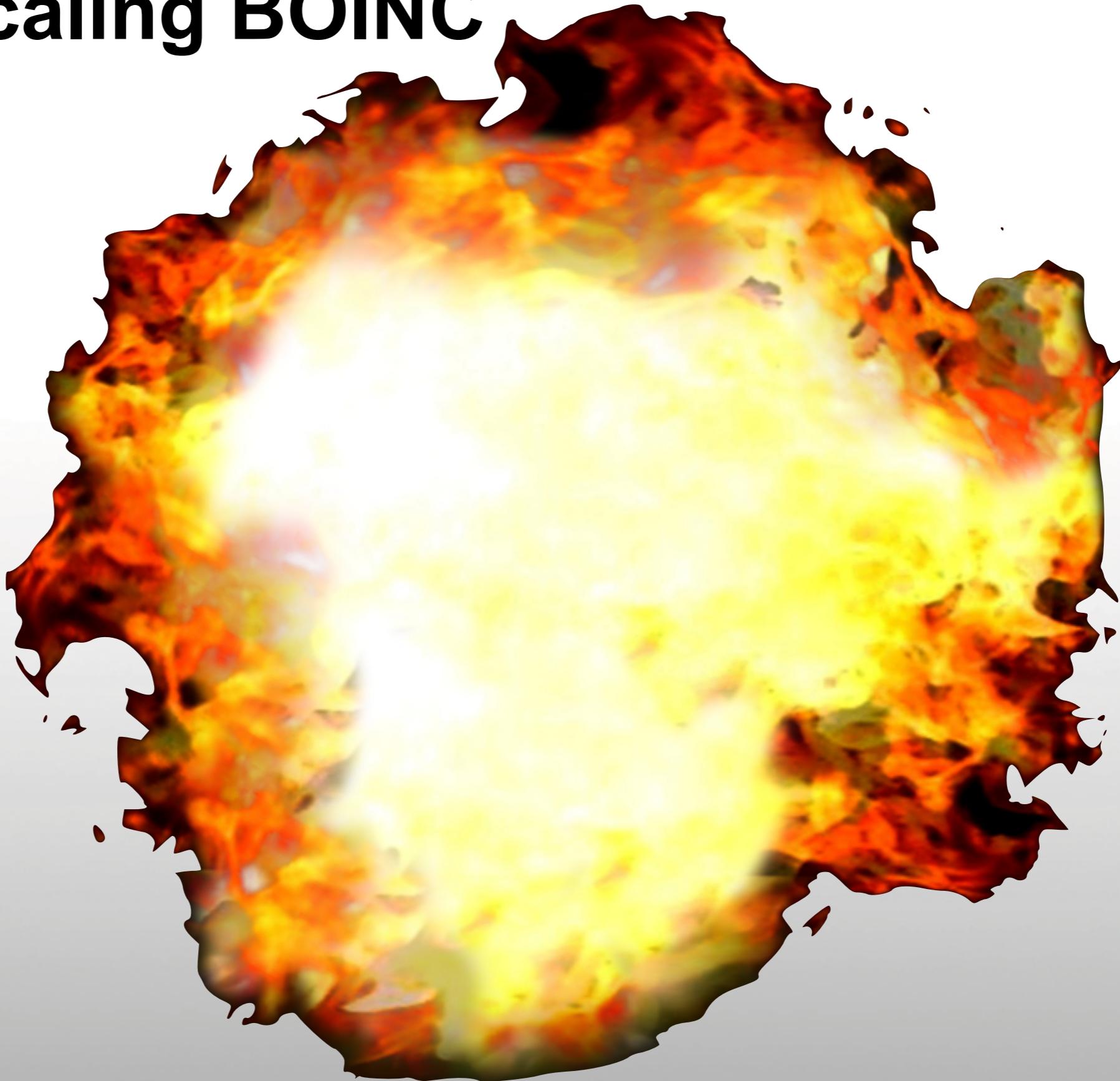


Scaling BOINC





Scaling BOINC





Scaling BOINC

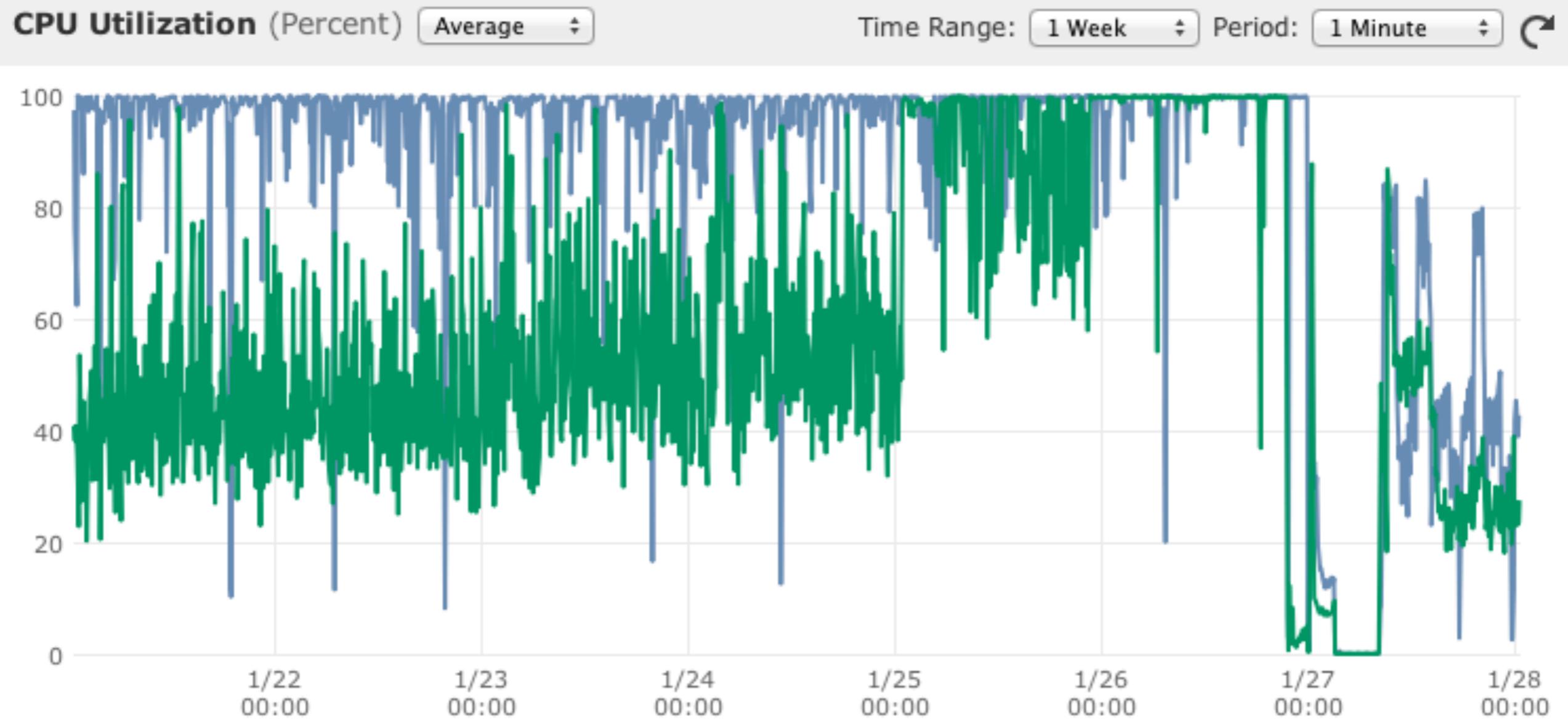
- You **MUST** think about scalability from the beginning
- Zooniverse - crashed due to load in the first 4 hours
- theSkyNet - crashed due to load in the first 6 hours
- theSkyNet POGS crashed due to load from a BOINC challenge (after 6 months) - was up and running again in 3 hours



Scaling BOINC

CloudWatch Monitoring Details

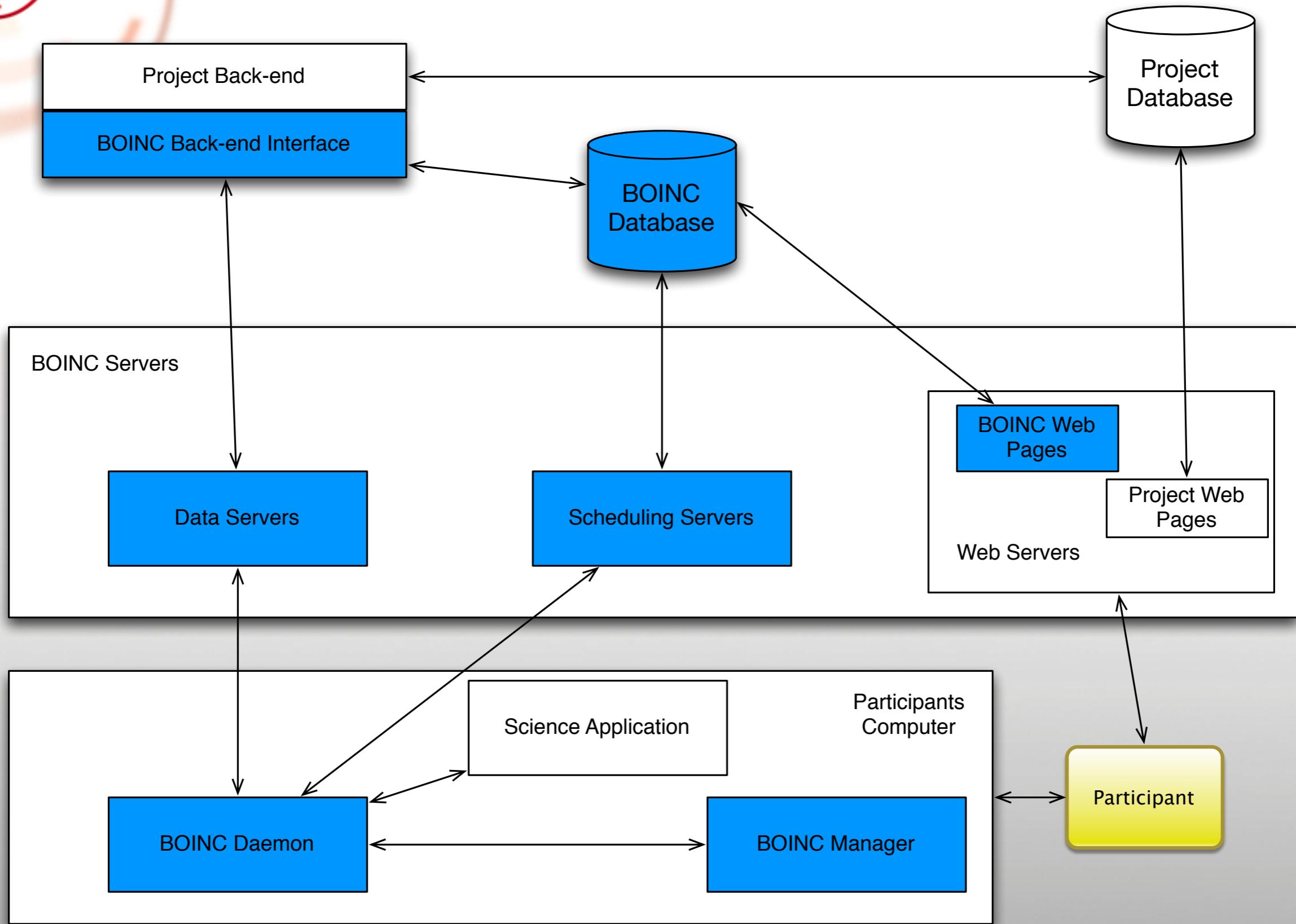
[Cancel](#)



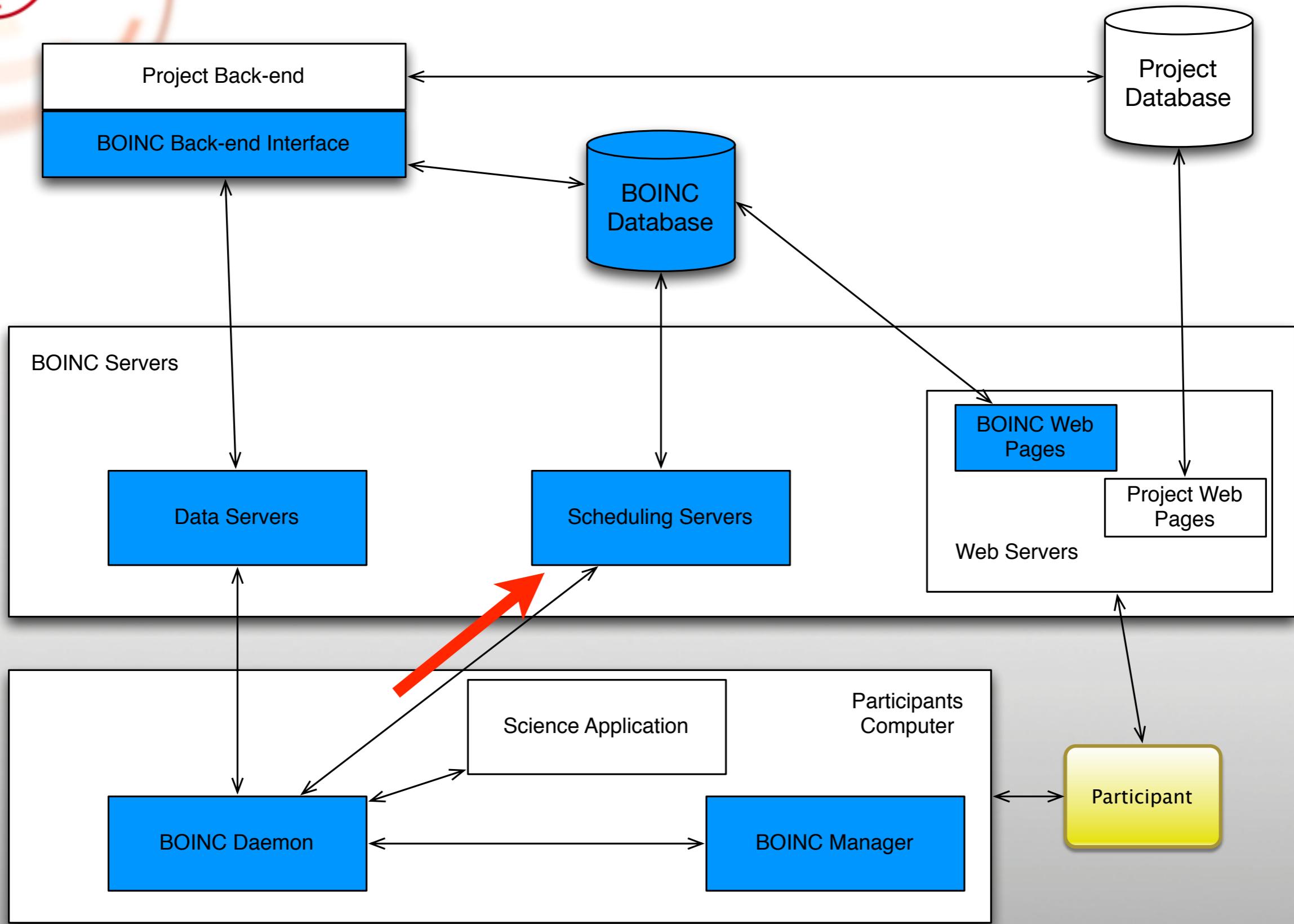
Monitored Instances: i-b89e41c7 i-06bab47e

Times are displayed in UTC.

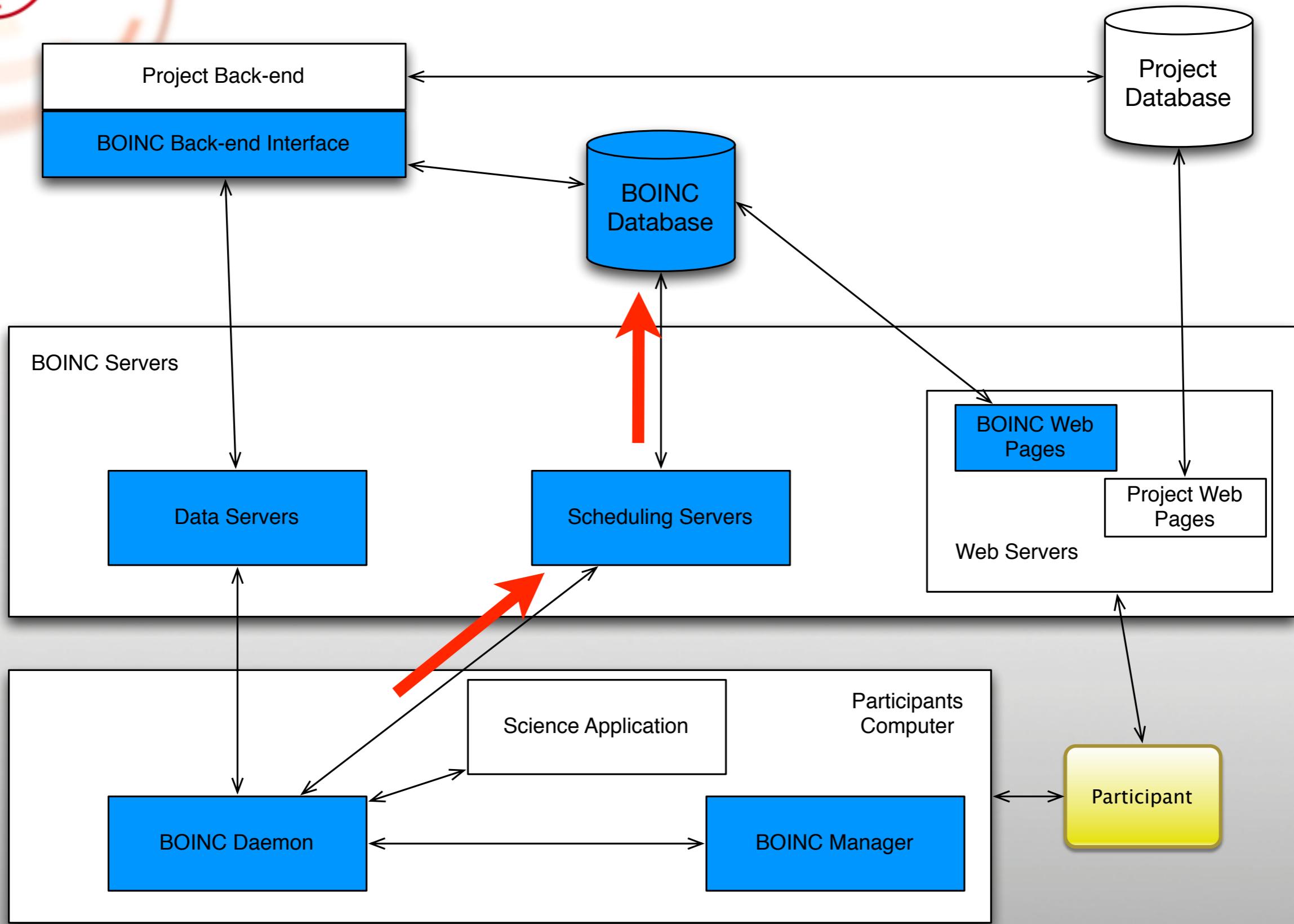
The BOINC Flow



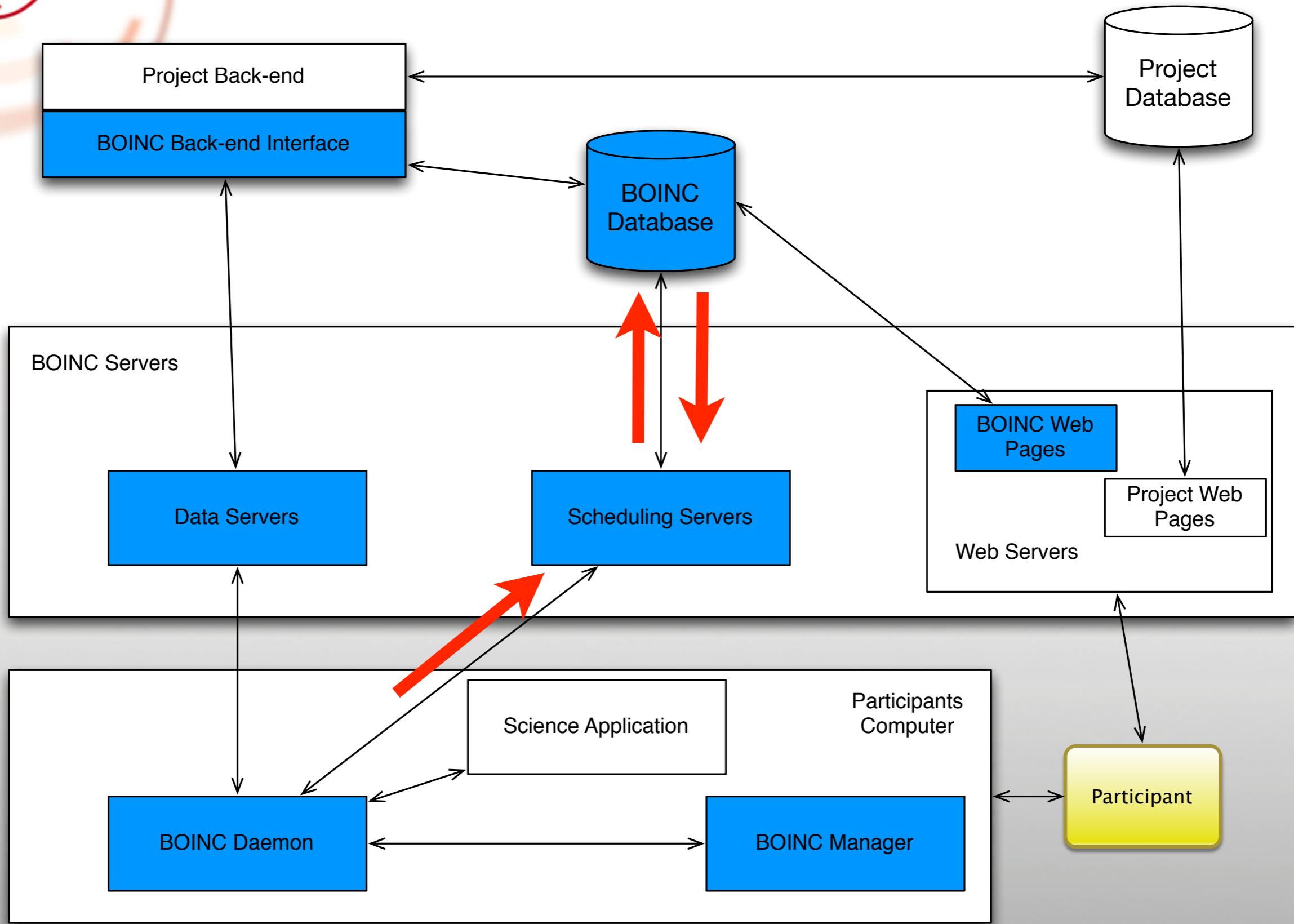
The BOINC Flow



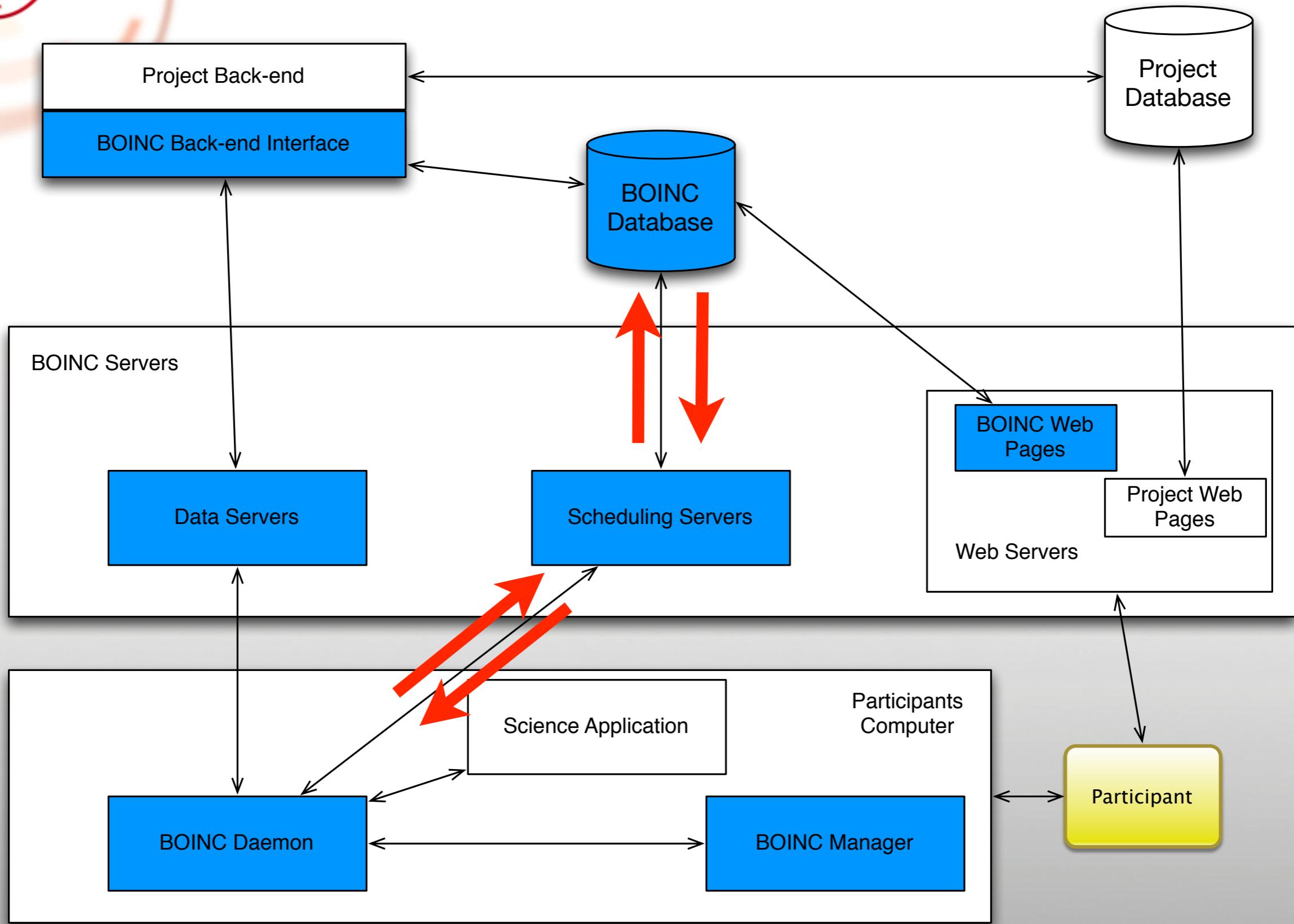
The BOINC Flow



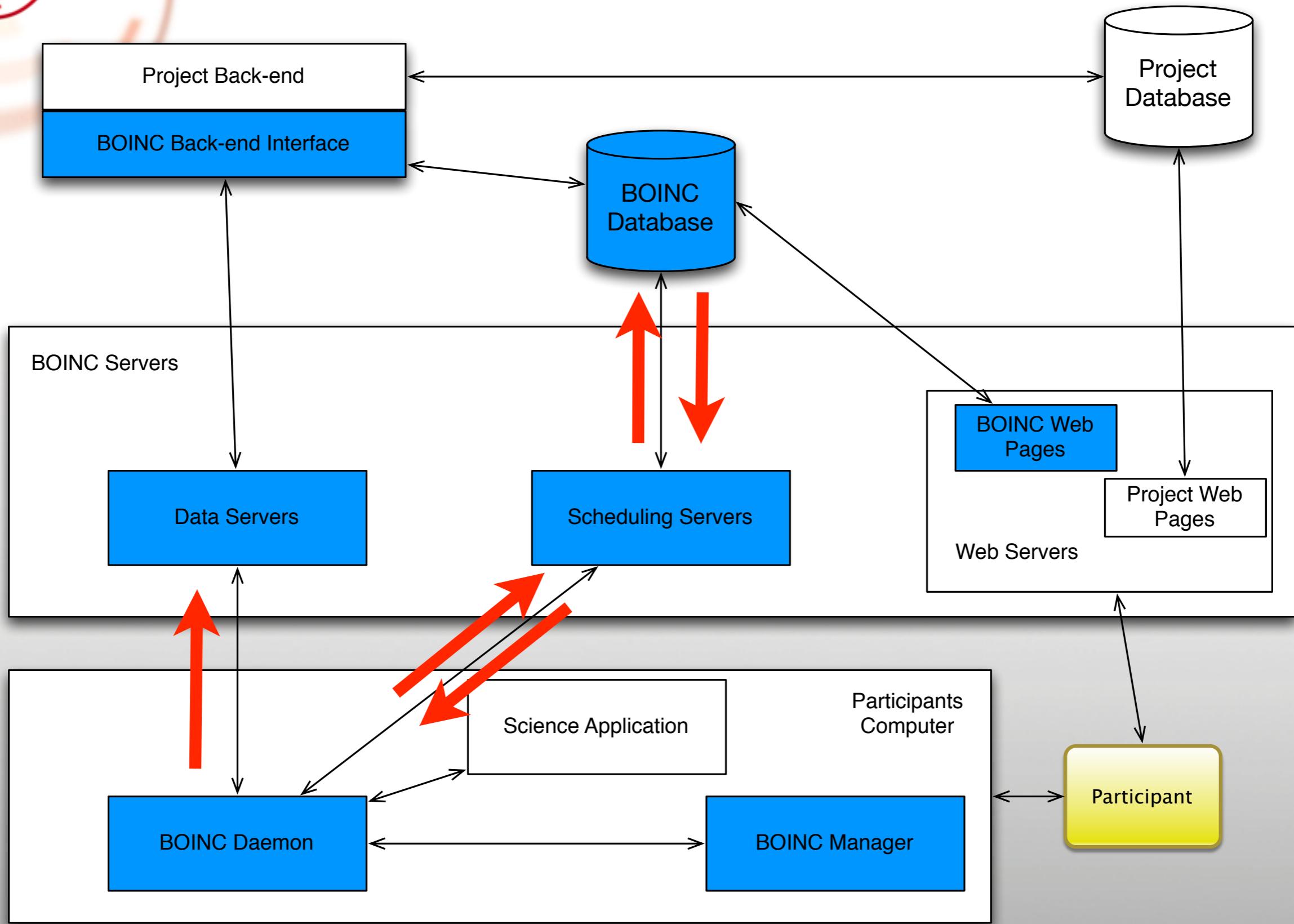
The BOINC Flow



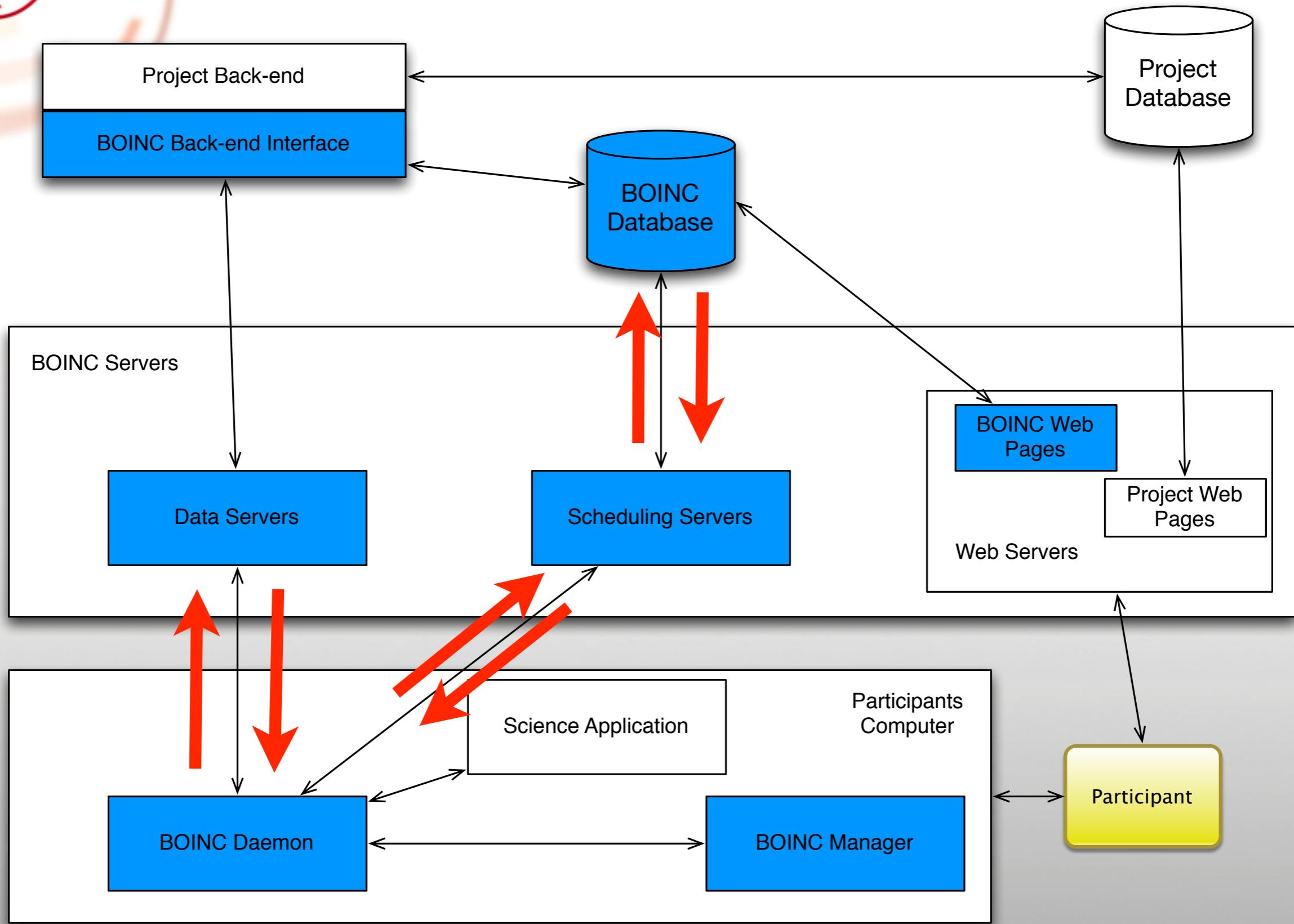
The BOINC Flow



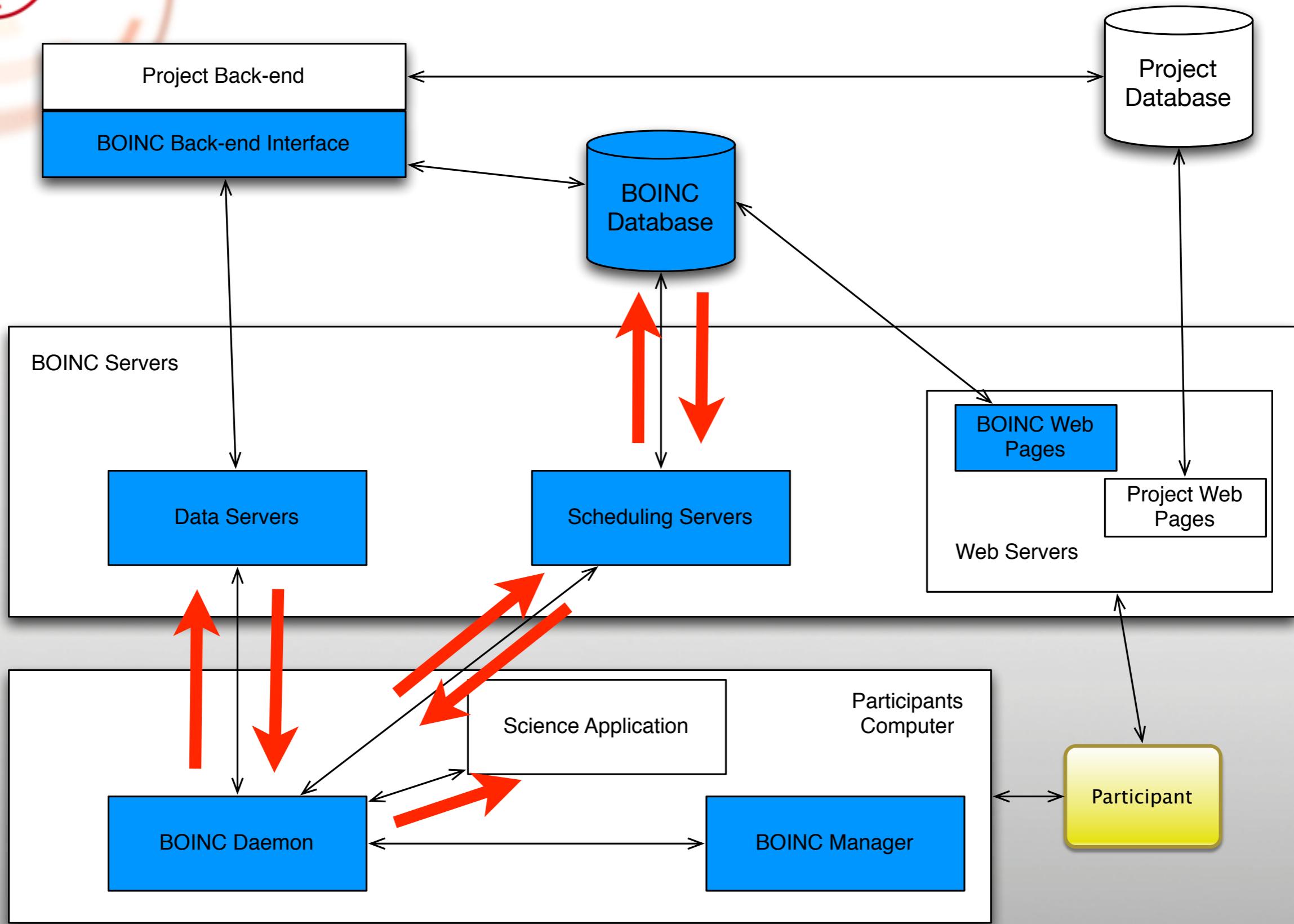
The BOINC Flow



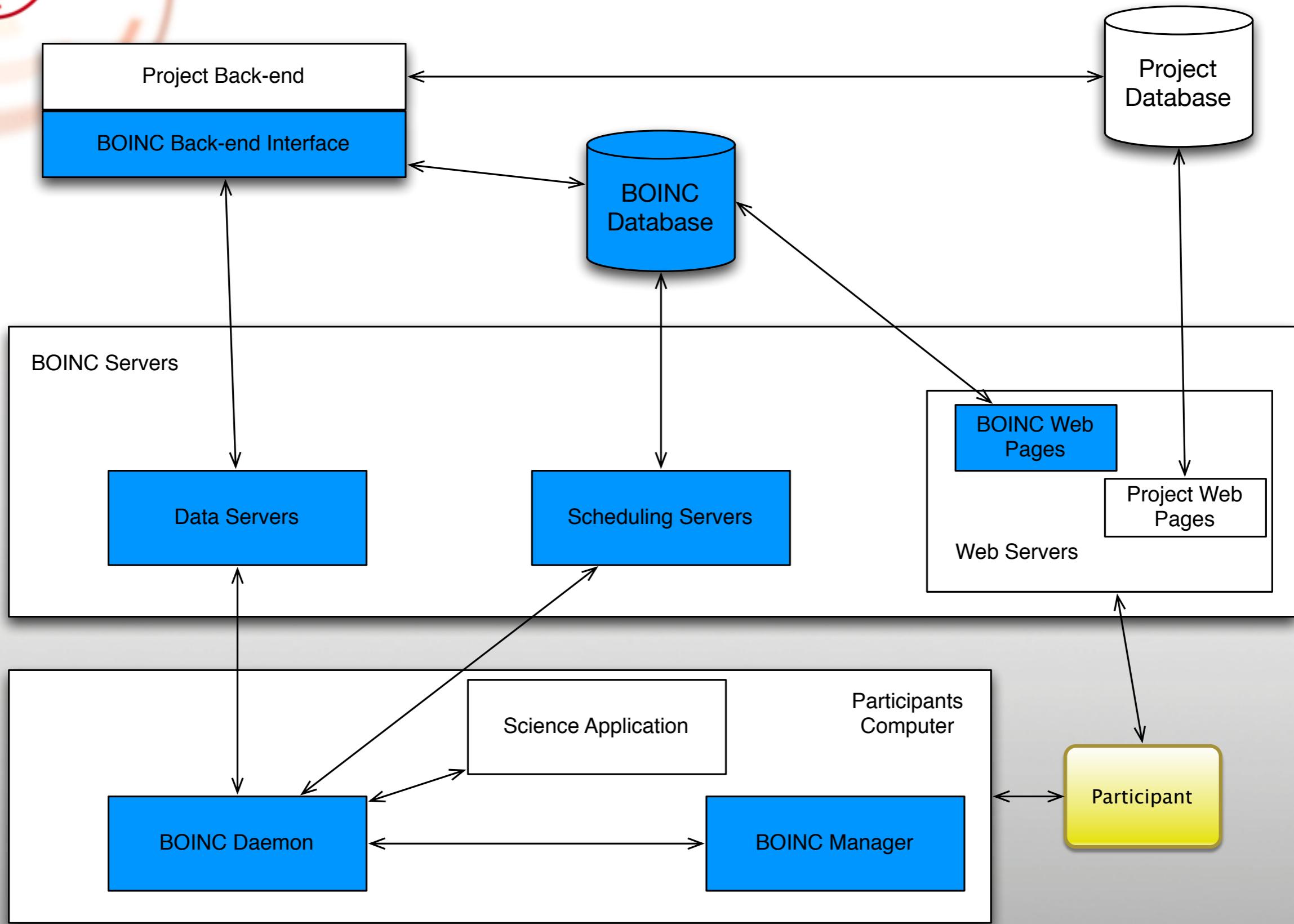
The BOINC Flow



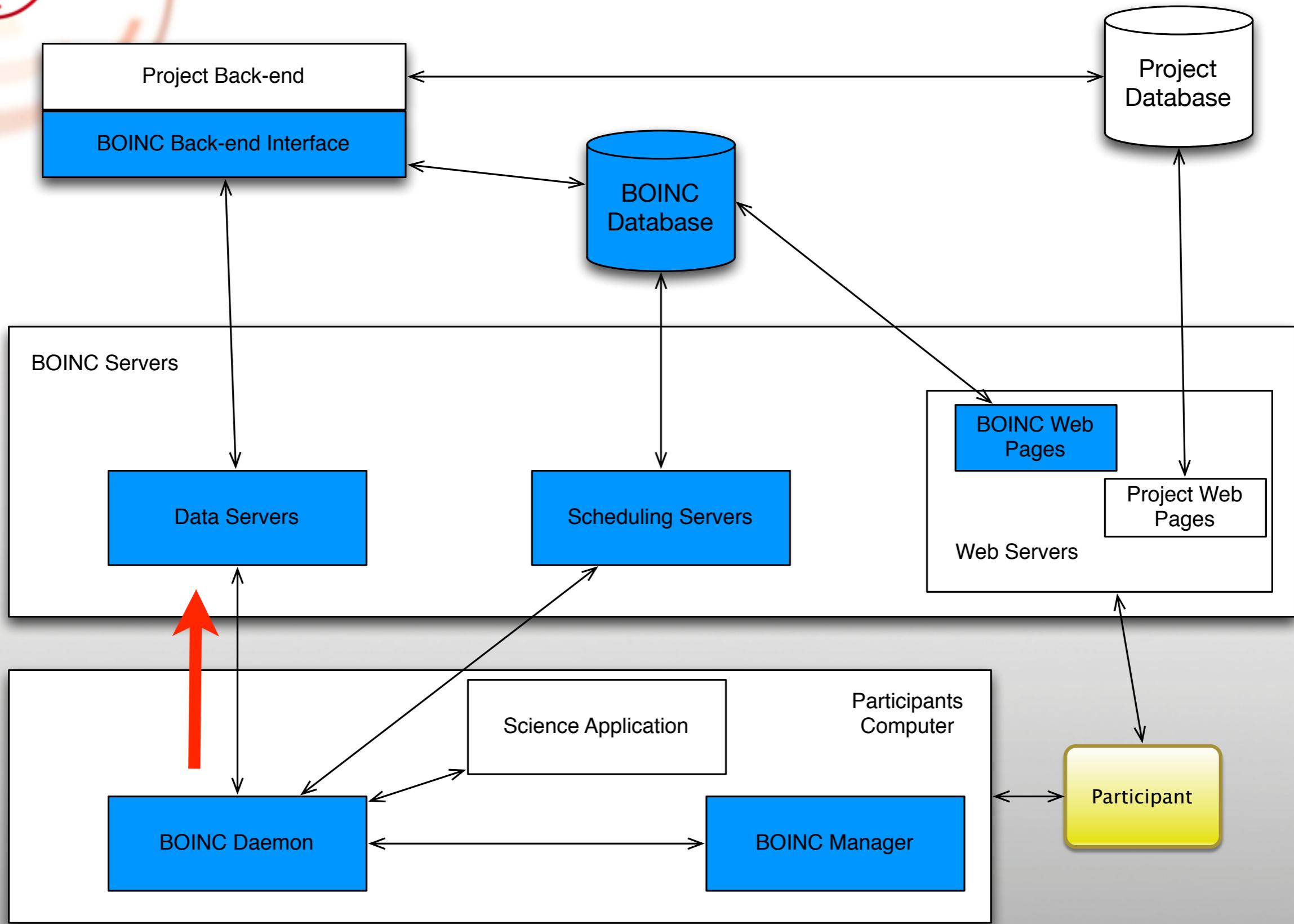
The BOINC Flow



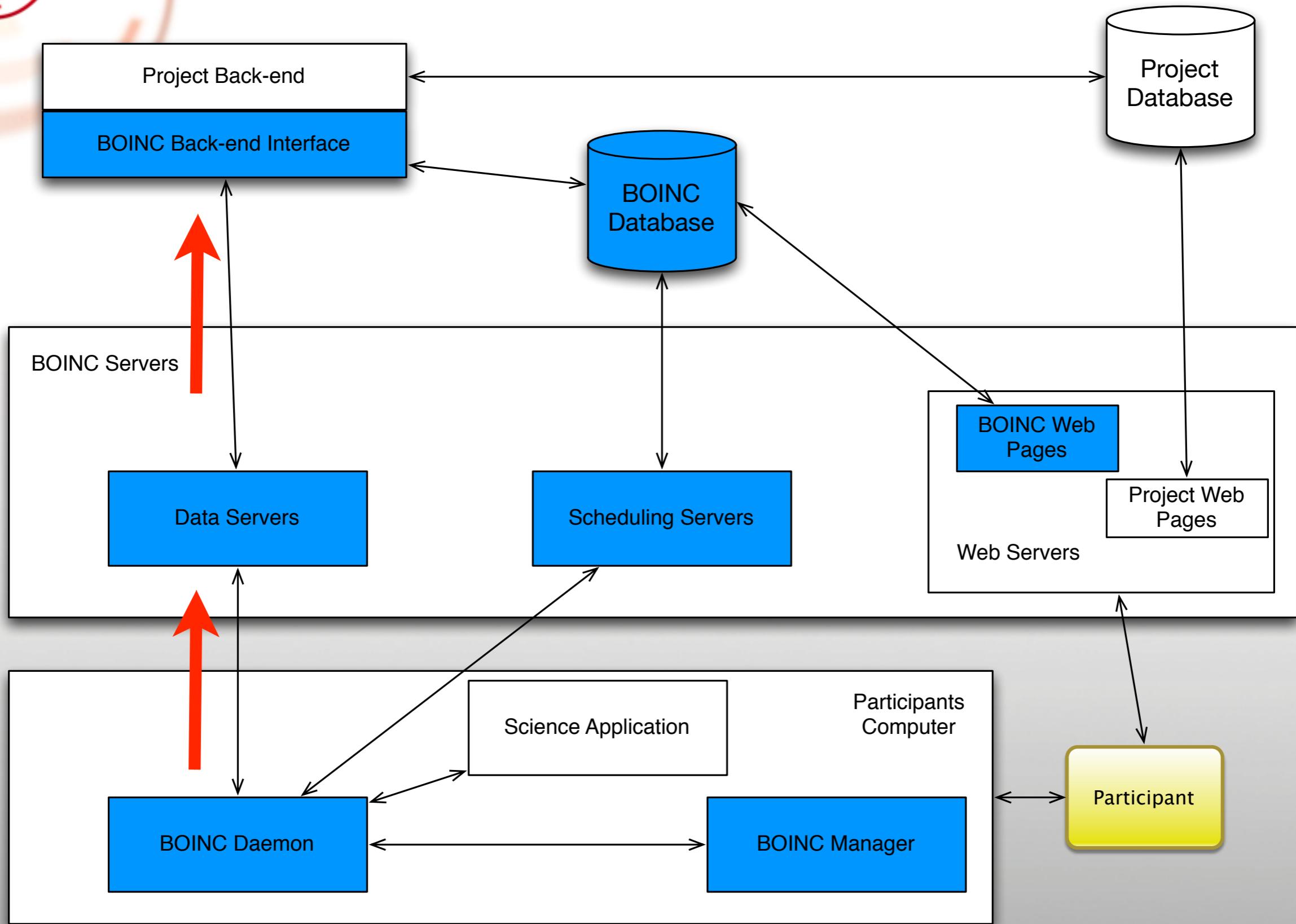
The BOINC Flow



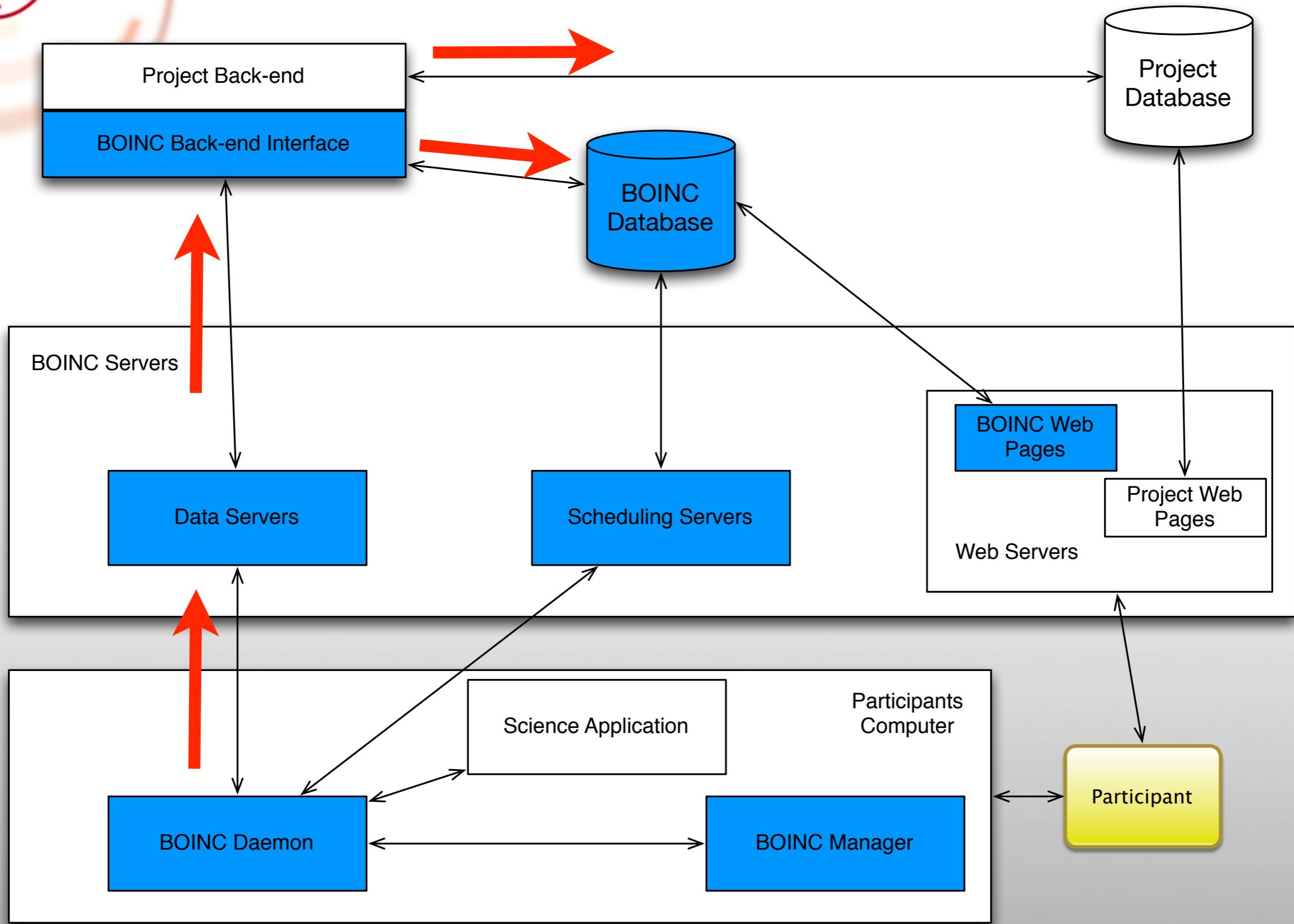
The BOINC Flow



The BOINC Flow



The BOINC Flow





POGS setup

- Designed to run under Amazon Web Services
- Uses fabric, boto (python) and puppet (ruby)
- 98% automated - only requires manual intervention at the very end
- <https://github.com/AstroinformaticsAU/AI2013/CitizenScience>
- <https://github.com/ICRAR/boinc-magphys/tree/master/machine-setup>

Fabric and Boto

```
# Setup NAGIOS
```

```
sudo('chkconfig nrpe on')
```

```
sudo('service nrpe start')
```

```
# Setup the HDF5
```

```
with cd('/usr/local/src'):
```

```
    sudo('wget http://www.hdfgroup.org/ftp/lib-external/szip/2.1/src/')
```

```
    sudo('tar -xvzf szip-2.1.tar.gz')
```

```
    sudo('wget http://www.hdfgroup.org/ftp/HDF5/current/src/hdf5-1.8.10.tar.gz')
```

```
    sudo('tar -xvzf hdf5-1.8.10.tar.gz')
```

```
    sudo('rm *.gz')
```



Basic setup

- Need git, puppet and xfsprogs installed for AWS

```
sudo yum --assumeyes --quiet install xfsprogs puppet git
```

- The other packages will be done by puppet later





Scaleability

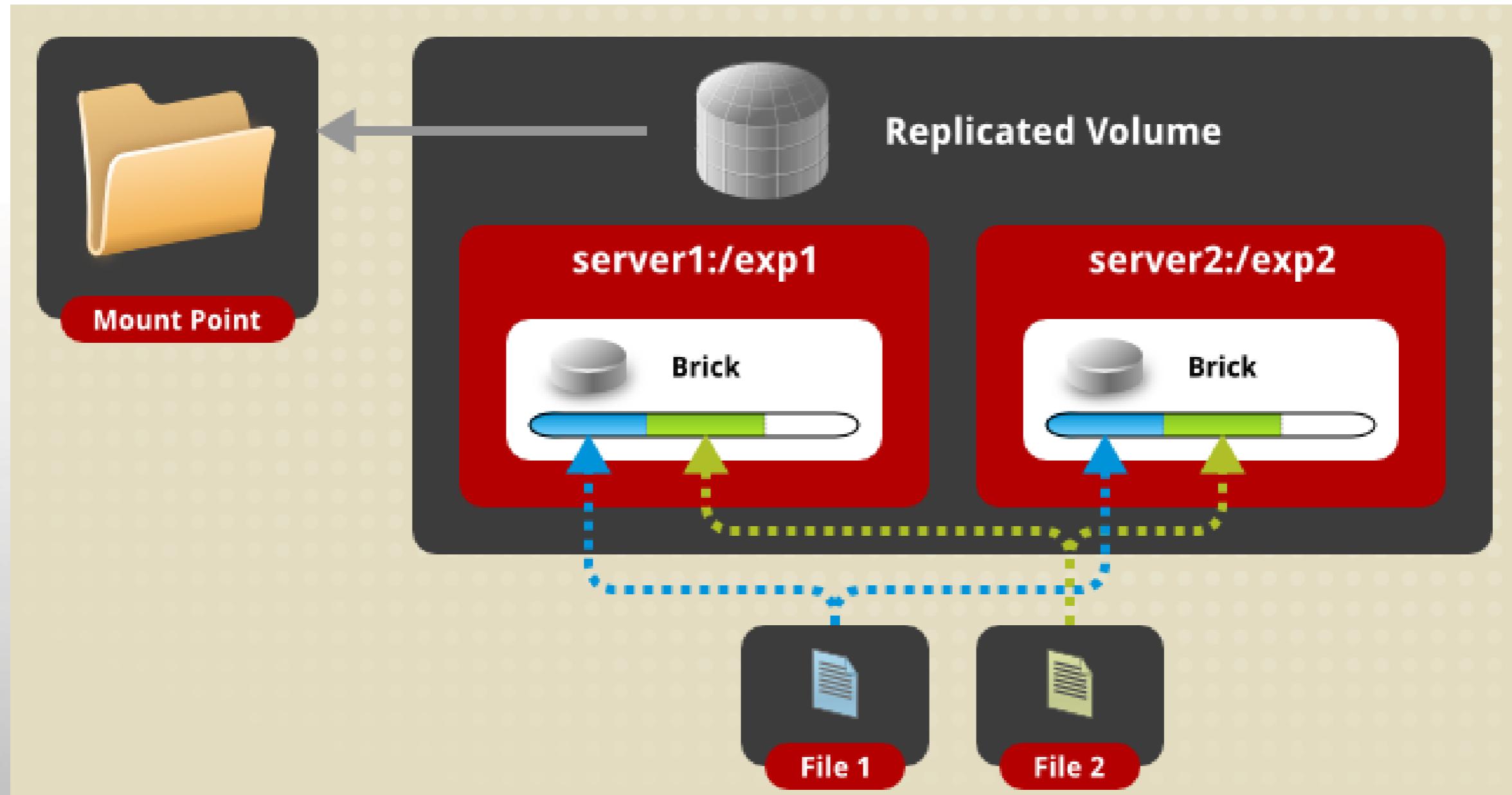
- To scale to multiple hosts in BOINC you need
 - Shared file system
 - Same project admin account (uid, gid)
 - 'ssh' to run commands on any other host without typing a password



Shared filesystem

- Path to the project directory relative to the project admin's home directory must be the same on all hosts.
~/projects/PROJECT_NAME
- NFS or GlusterFS
 - GlusterFS also offers file replication

Gluster Replicated Volume



Version Control

- Use it!
- We can deploy POGS easily using it
- Doesn't matter which git, subversion (git is better)

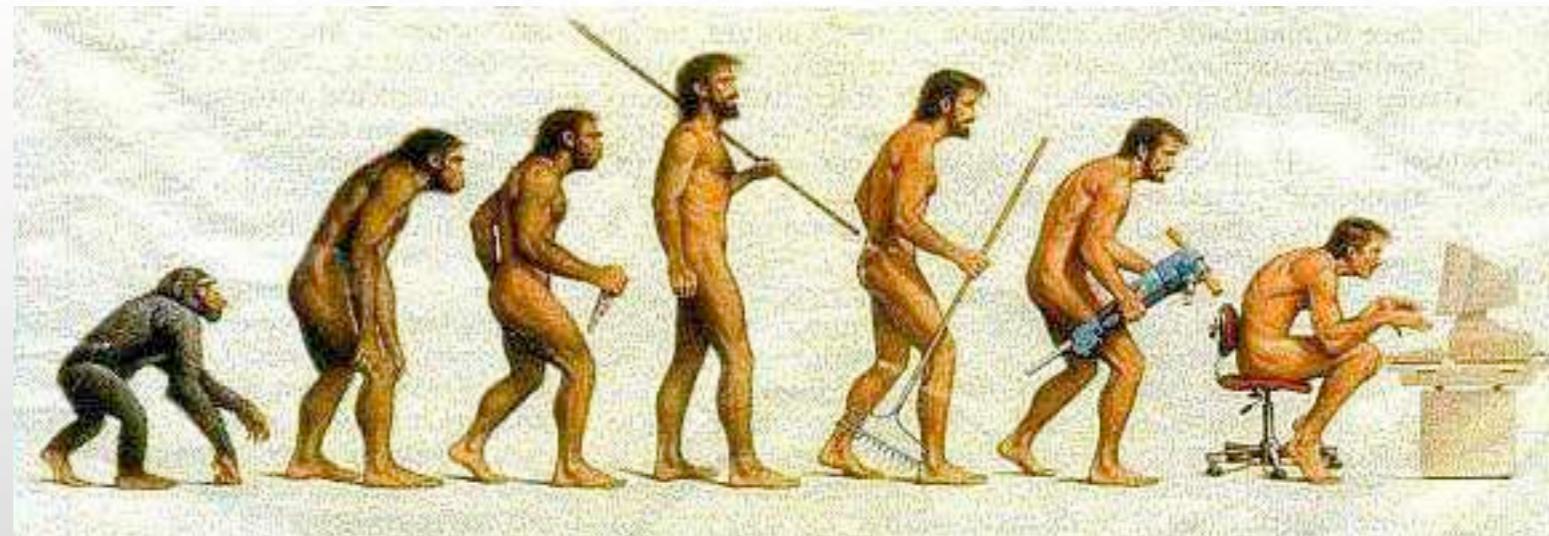


```
git clone git://github.com/ICRAR/boinc-magphys.git
```

Puppet

- Does all the yum updates you need and creates directories and users

```
package { 'httpd':  
    ensure => installed,  
}  
  
package { 'httpd-devel':  
    ensure => installed,  
}  
  
user { 'apache':  
    ensure  => present,  
    groups => ['ec2-user'],  
}  
  
service { 'httpd':  
    ensure => running,  
    enable => true,  
    require => Package['httpd'],  
}
```



Install BOINC Server

- We're ready for the BOINC system

```
svn co http://boinc.berkeley.edu/svn/trunk/boinc ~/boinc  
./_autosearch  
./configure --disable-client --disable-manager  
make
```

- The client is downloaded from Berkeley from the web site
- The trunk is under development and sometimes it won't compile.





BOINC Database & Project

- Has to be MySQL if you use their scripts
- Best to have it on a separate server
- Ideally with a read-replica for speed

```
./make_project -v --no_query --drop_db_first --url_base  
http://host --db_user <user> --db_host=<host>  
--db_passwd=<passwd> <project_name>
```

- This creates the project files needed, the web site and the database.
- The project will be in ~/projects/<project_name>

config.xml

- The master file telling the servers what to do when.

```
<boinc>  
  <config>  
    [ configuration options ]  
  </config>  
  <daemons>  
    [ list of daemons ]  
  </daemons>  
  <tasks>  
    [ list of periodic tasks ]  
  </tasks>  
</boinc>
```





Platforms

- A platform is a compilation target for BOINC applications
- Typically a combination of a CPU architecture and an operating system
- You will need to build your client code for each of the platforms you decide to support. POGS supports 5 + 1
 - Linux 32/64
 - Windows 32/64
 - OS X 10.6+
 - Android



Assume Nothing

- You need to build the executables as fat binaries
- You cannot assume the client machine will have:
 - Shared libraries
 - DLLs
 - executables
 - file system
 - command line
- Alternative - VirtualBox
 - hard to setup
 - *nix based

MAGPHYS Execution

- Requires the following:
 - observation data file
 - filter list file
 - redshift info file
 - optical model (BIG file)
 - infrared model (BIG file)





Creating Work Units

- A work unit is a chunk of work to be down
- Generates two or more results
- Specifies things like:
 - priority
 - delay bound (how long you can take)
 - minimum quorum
 - max errors
 - flops estimate
 - memory / disk estimates

Validators

- Make sure the answers make sense.
- You WILL get rubbish back
- Need to be careful with text
 - *nix end of line \n
 - Windows end of line \r\n
- Floating point numbers can be tricky, different processors can give different results



Copyright 2003 Randy Glasbergen. www.glasbergen.com



A good result file

Summary of statistics:

Detection threshold = 0.00054 Jy/beam

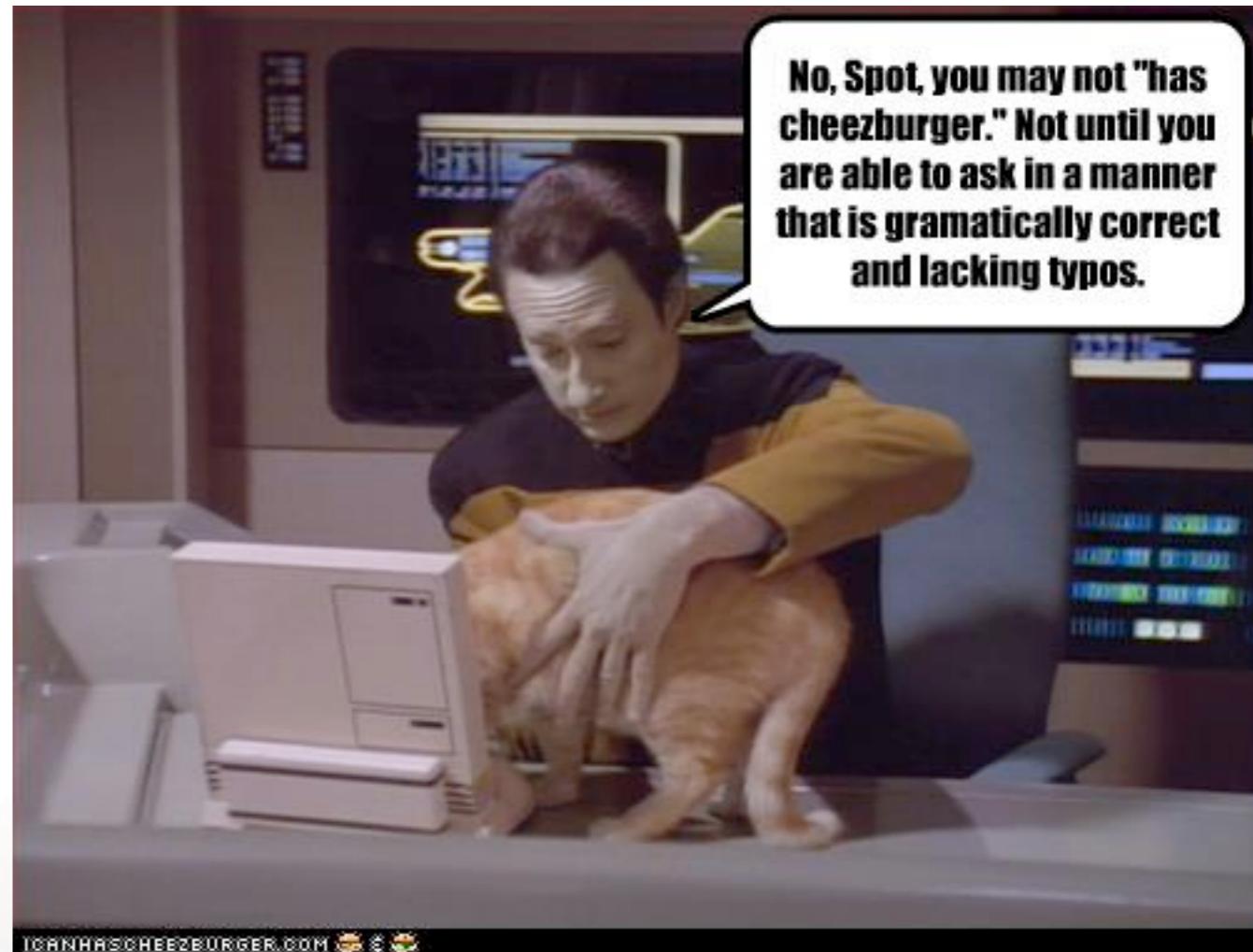
Not calculating full stats since threshold was provided directly.

Total number of detections = 100

Obj#	Name	X	Y	Z	RA	DEC	VEL [km/s]	w_RA [arcmin]	w_DEC [arcmin]	w_50 [km/s]	w_20 [km/s]	w_VEL [km/s]	[
1	J123103-441846	21.2	15.0	0.4	12:31:03.22	-44:18:46.57	78674.453	0.67	0.53	75.359	82.867	24.633	
2	J123013-441119	88.5	71.0	0.5	12:30:13.01	-44:11:19.48	78676.453	0.67	0.53	39.289	83.195	49.273	
3	J123111-441457	10.4	43.6	0.7	12:31:11.16	-44:14:57.75	78681.523	0.54	0.53	38.500	81.250	24.633	
4	J123053-440902	33.8	88.1	0.7	12:30:53.67	-44:09:02.42	78682.078	0.40	0.80	47.234	82.625	49.273	
5	J123035-441436	58.0	46.4	1.0	12:30:35.77	-44:14:36.34	78688.461	0.67	0.53	73.898	87.063	73.914	
6	J123021-441058	77.0	73.7	1.2	12:30:21.60	-44:10:58.22	78694.188	0.40	0.67	32.578	84.164	73.914	
7	J123034-441202	59.9	65.6	1.2	12:30:34.26	-44:12:02.50	78694.797	0.40	0.67	71.016	89.188	49.273	

Problems

- Localisation settings
- Number formats
 - Australia - 7,000.52
 - Europe - 7.000,52
- Dates
 - Australia - ASCII
 - Russia, Europe - UTF-8
 - China UTF-16



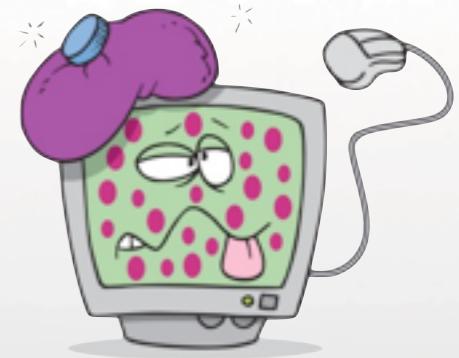
Problems

675	J123105-435742	17.1	93.0	160.5	12:31:05.88	-43:57:42.26	49566.715	0.67	0.40	168.297	190.945	167.93
676	J123044-440833	46.6	11.7	160.6	12:30:44.15	-44:08:33.82	49567.777	1.34	1.06	691.270	705.258	671.32
677	J123118-440420	0.0-	43.2	160.6	12:31:18.69	-44:04:20.32	49568.523	0.27	0.40	20.094	32.148	20.99
678	J123106-441003	16.0	0.4	160.8	12:31:06.94	-44:10:03.45	49573.555	0.40	0.27	88.379	131.531	62.91
679	J123008-440521	94.8	35.8	160.9	12:30:08.32	-44:05:21.65	49574.980	0.40	0.27	40.000	101.730	41.98

Host: sn0V*QC3AY116.gateway.messenger.live.com

Content-d=9e56f5d1-89cd-4578-8b11-701af1344d3a&cJDvL&COLxxx-W18&374; xidseq=14; wls=AIS0iY-t:a*m; E=P:VxmduEp3zog=:H+LN=63u4u1321979274351%267099%2611; LD=9e56f5d1-89cd-4578-8b11-701af1344d3a_00d7099b240_5725_1321979272635=E3135~ReadV~SMD:televisa.com.mx~MT:2~RMW:1019|E3135~ReadView.M2Launch~CLS:col112~BRW:IE~VRS:8&9e56f5d1-89cd-4578-8b11-701af1344

XvaxLCnuHmWoyjnupWOr6rfP/pJb3i4VHcF1ldikJSAILSCq7gVn0FINJ



387	J122947-463833	41.7	6.9	137.2	12:29:47.77	-46:38:33.35	64831.938	1.60	0.67	142.781	162.137	205.789
388	J122936-462848	57.0	80.0	137.8	12:29:36.01	-46:28:48.27	64844.184	0.40	0.40	60.625	141.340	68.613
389	J123001-463506	24.1	32.8	138.2	12:30:01.49	-46:35:06.46	64854.988	0.53	0.40	43.551	181.113	68.613
45.758	0.01307	0.00910	0.00140	2.58	93	96	67	68	140	142	15	E 94.7 67.5 140.8 94.0 67.4
396	J123019-463110	0.3	62.3	141.1	12:30:19.89	-46:31:10.16	64920.316	0.27	0.40	56.691	73.965	68.633
397	J122933-463818	59.9	8.8	141.9	12:29:33.67	-46:38:18.56	64939.586	0.40	0.40	79.328	133.059	68.641
2.96	39 41 74 76 138 139	10	-	39.6	74.6	138.5	39.6	74.6	138.8	39	75	139
393	J123007-463503	16.2	33.2	140.7	12:30:07.60	-46:35:03.43	64911.195	0.53	0.67	29.168	73.816	68.633
394	J123017-462933	3.2	74.4	140.8	12:30:17.64	-46:29:33.67	64912.598	0.93	0.67	408.656	419.180	343.281
395	J122907-463028	94.0	67.4	140.9	12:29:07.31	-46:30:28.41	64916.387	0.53	0.27	91.148	137.965	45.758

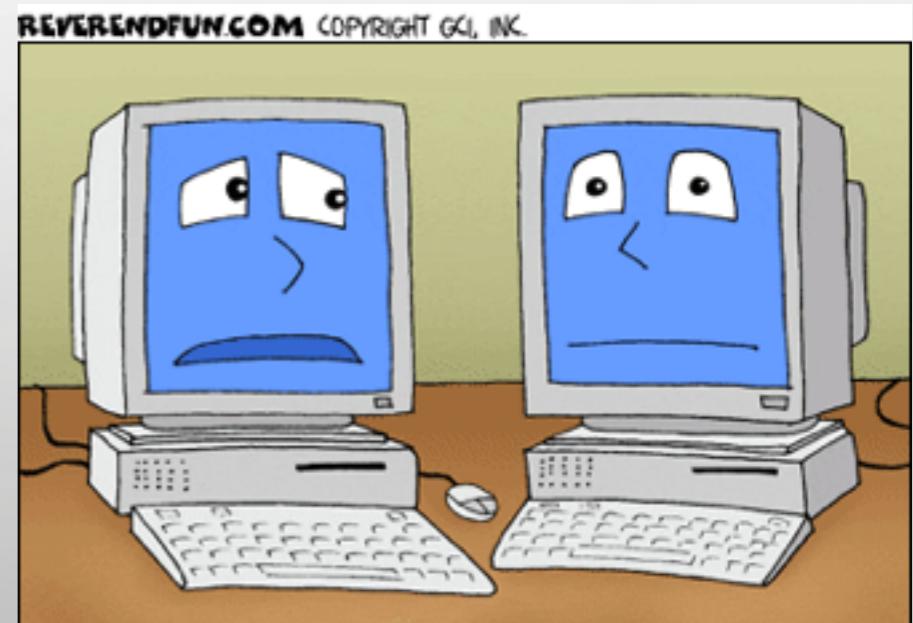


Validator

- Checks file is in correct format
- Compares two sets of results together to make sure the results are the same
 - Can set the quorum to any number
- LOTS of options here. Especially because Floating Point numbers will vary across chips
- If they don't match send it out again

Assimilator

- Parse the results and store them
- Issue credit to the users
- Credit is important to the BOINC community
- Needs to be fast



Thanks to Kerry Hodgkinson

11-29-2005

OH MY NO ... HUMANS DIDN'T INVENT US AT
ALL ... WE EVOLVED FROM PLASTIC, WIRES,
AND LITTLE MICRO THINGERS



BOINC Stats

- BOINC groups run challenges

Project name	Users	last day	Hosts	last day	Teams	last day	Countries	last day	Total credit	last day	Last update
BOINC combined	2,530,348	307	8,487,436	8,069	97,634	5	273	0	1,323,470,259,953	1,980,403,946	2013-02-07
DistrRTgen	13,866	8	32,076	17	766	0	150	0	134,630,588,397	681,118,687	2013-02-07
PrimeGrid	53,224	16	185,535	232	2,561	1	186	0	214,851,117,805	224,874,842	2013-02-07
GPUGRID	17,939	15	34,909	26	1,131	0	138	0	78,633,832,865	236,644,675	2013-02-07
MilkyWay@home	154,808	39	312,119	210	3,489	2	209	0	156,917,778,066	107,115,404	2013-02-07
World Community Grid	400,342	50	1,669,994	742	21,757	2	224	0	88,748,688,898	134,459,808	2013-02-07
POEM@HOME	38,839	18	104,030	51	1,416	2	168	0	37,013,708,480	120,346,830	2013-02-07
SETI@Home	1,360,222	82	3,339,444	221	61,188	0	233	0	174,054,546,275	67,868,567	2013-02-06
Collatz Conjecture	30,898	6	76,324	38	1,350	0	162	0	128,036,571,233	97,641,828	2013-02-07
Einstein@Home	336,566	73	3,485,883	6,689	10,587	1	222	0	68,807,540,657	121,591,631	2013-02-06
Moo! Wrapper	3,435	2	13,710	10	455	0	99	0	50,877,419,738	55,671,355	2013-02-07
Donate@Home	778	3	2,138	10	146	0	59	0	15,652,895,682	64,734,800	2013-02-07
Rosetta@Home	354,753	65	1,104,195	304	10,070	1	225	0	21,580,478,278	14,539,583	2013-02-07
Climate Prediction	266,998	72	548,454	131	7,666	0	221	0	19,598,920,242	7,114,357	2013-02-07
SIMAP	41,441	6	139,724	38	2,277	0	181	0	3,100,594,188	5,044,061	2013-02-07
Docking@Home	31,518	13	83,085	35	1,079	0	142	0	3,802,572,881	3,435,535	2013-02-07
Cosmology@Home	52,413	23	98,474	59	1,752	0	188	0	3,493,595,700	3,202,080	2013-02-06
Malaria Control	66,627	11	158,607	51	2,178	0	208	0	2,946,615,803	2,814,667	2013-02-06
LHC@Home Classic	110,117	40	286,614	136	4,600	0	193	0	1,453,585,746	6,599,653	2013-02-06
theSkyNet POGS	1,304	9	3,927	26	208	1	63	0	292,540,845	2,239,554	2013-02-07
Primaboinca	4,051	0	14,066	18	369	0	107	0	1,642,995,335	1,973,200	2013-02-06
Asteroids@home	3,156	55	6,851	88	281	3	84	0	120,627,653	3,529,200	2013-02-06
yoyo@home	14,800	6	52,617	21	790	0	120	0	2,086,274,614	2,218,429	2013-02-07
NFS@Home	7,266	2	20,602	3	587	0	125	0	1,071,178,189	962,353	2013-02-07
FreeHAL	17,495	4	61,814	45	781	0	137	0	3,659,508,925	365,202	2013-02-06

- They challenge each other to process most data



BOINC Stats

- The last challenge on theSkyNet POGS

Name	The deep-sky fishing
Status	Completed
Project	theSkyNet POGS
Issued by	Astronomy.Ru Forum
Start time	2013-01-25 01:00 UTC
End time	2013-02-01 01:00 UTC
Late entrants allowed?	Yes
Number of teams participating	» 13
Number of users participating	215

Team Name	Rank	Credit	-1:00	-2:00	-4:00	-8:00	-16:00	-32:00
BOINC@AUSTRALIA	1	13,190,323	13,155,498	13,101,752	13,003,969	12,773,001	12,245,090	10,919,997
Sicituradastra.	2	10,663,515	10,602,238	10,532,181	10,412,538	10,158,313	9,635,827	8,536,648
Astronomy.Ru Forum	3	1,514,244	1,511,058	1,506,128	1,498,907	1,480,897	1,413,681	1,223,194
BOINC@Poland	4	815,961	813,536	811,766	805,925	791,659	755,321	674,954
L'Alliance Francophone	5	521,132	519,761	516,088	511,167	494,441	467,404	408,100
Crunching Family	6	316,334	315,025	313,343	311,139	303,573	288,953	258,208
SETI.USA	7	145,830	144,981	142,556	137,839	130,361	124,626	112,342
The Scottish Boinc Team	8	135,971	135,272	135,016	129,122	127,263	122,980	107,235
AMD Users	9	121,820	121,634	121,634	121,130	119,971	116,661	99,987
SETIKAH@KOREA	10	72,048	71,490	71,136	70,189	69,030	67,340	58,844



Conclusions

- Cheap and cheerful way to build a TFlop/PFlop machine
- Need an embarrassingly parallel problem
- POGS
 - 2,000 galaxies processed
 - 1,100,000 areas
 - 19,083,012 pixels processed
 - 1,334 users
 - 4,037 computers
- First two papers in preparation



Links

- <http://www.theskynet.org>
- <http://23.23.126.96/pogs>
- <http://boinc.berkeley.edu>
- <https://github.com/ICRAR/boinc-magphys>
- <http://www.icrar.org>

Any Questions

- Remember I'm "hard of hearing" - I wear hearing aids so please speak clearly.
Ladies - I'm sorry I don't hear higher frequencies very well at all.
- Contact me at kevin.vinsen@icrar.org

