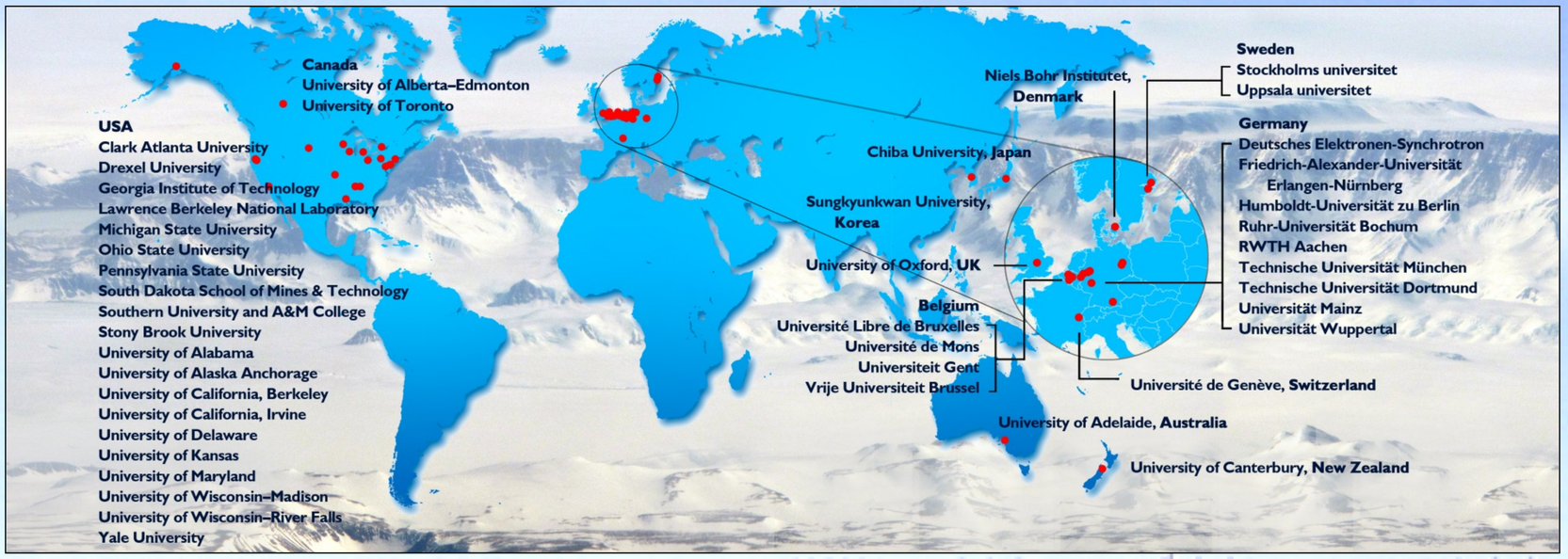
IceProd 2

David Delventhal
University of Wisconsin-Madison
IceCube Collaboration

IceCube Computing

Medium size collaboration

- 2 data centers and several smaller clusters
- Most CPU compute is opportunistic
- Diverse computing infrastructure



WIPAC

Requirements

1. Cannot require root permissions.
2. Need more than just glideins.
3. Must support multiple OS's and schedulers.
4. Local disk is usually not available.
5. Need to permanently record all job statistics.
6. Allow for both individual users and production.
7. Provide offline production environment.
8. Documentation of configuration parameters

IceProd 1

- Site-local python daemons backed by a central database
- Plugins for each batch system
- gridftp/http directly to/from main data center
- XMLRPC to communicate with jobs
- Store configuration and statistics for every job

IceProd 1 Issues

- A long series of patches to increase functionality
- Early design decisions hampering new ideas
- Multiple bottlenecks appearing:

  DB - locking issues, queries per second

  Queue - designed for 100 jobs per hour, not 10k

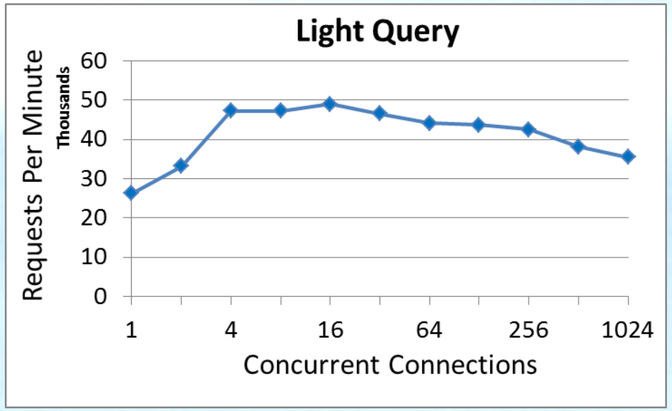
IceProd 2

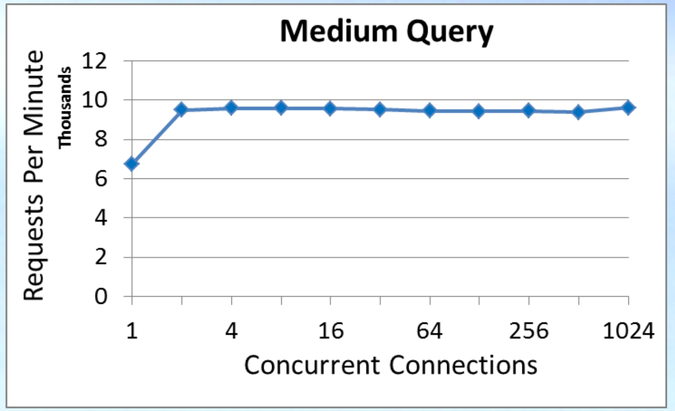
Complete Rewrite

- Support for Python 3
- Focus on unit tests
- Local SQLite database at each site
- JSON Web API
- HTML based UI
- User Accounts

Scalable Web Server

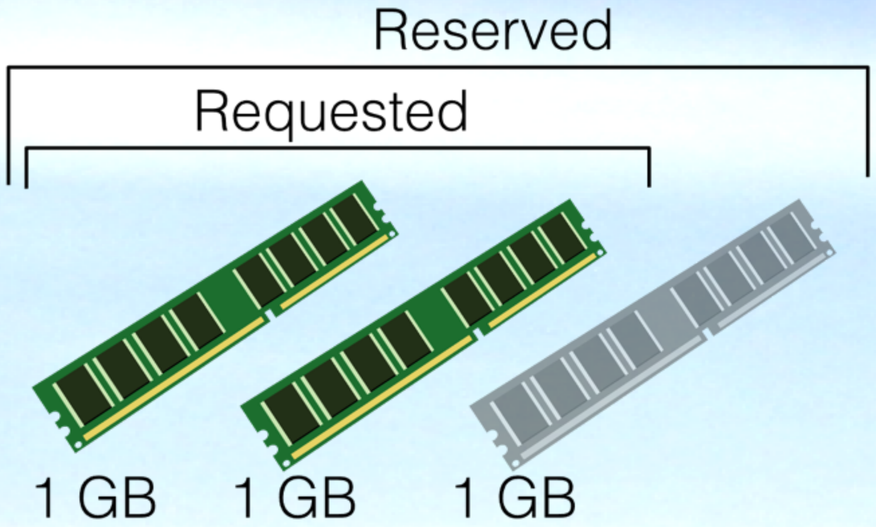
High-volume communications
Web server can handle more than 10000 connections per minute on a single thread.

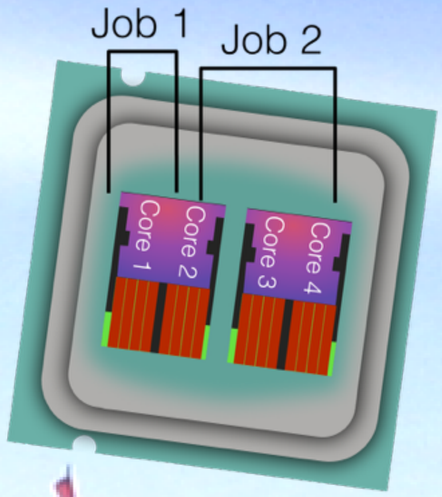




Pilot

Pilot allocates extra memory in case job exceeds memory usage
Pilot can also run multiple jobs on the same node





Logs

stdout

```
retrieving http://prod-exe.icecube.wisc.edu/corsika-v6900.md5sum...
retrieving http://prod-exe.icecube.wisc.edu/corsika-v6900.md5sum...
retrieving http://prod-exe.icecube.wisc.edu/corsika-v6900.md5sum...
retrieving http://prod-exe.icecube.wisc.edu/corsika-v6900.md5sum...
retrieving http://prod-exe.icecube.wisc.edu/corsika-v6900.md5sum...
```

stderr

```
INFO:CorsikaGenerator:summaryfile:summary.xml
INFO:CorsikaGenerator:length:1600.0
INFO:CorsikaGenerator:crtype:0
INFO:CorsikaGenerator:histogramfilename:None
INFO:CorsikaGenerator:model:sibyll
INFO:CorsikaGenerator:inputfile:
INFO:CorsikaGenerator:execute CorsikaGenerator: True
INFO:iceprod::Corsika:setting up working directory: /scratch/condor/dir_7546//dcors0
INFO:iceprod::Corsika:caching: 1
INFO:iceprod::Corsika:CVMFS configured to '/cvmfs/icecube.opensciencegrid.org/'
WARNING:iceprod::Corsika:No CVMFS repo. Downloading tarball from: 'http://prod-exe.icecube.wisc.edu'
Traceback (most recent call last):
File "/scratch/condor/dir_7546/local_temp/corsika.py", line 13, in <module>
cors.ExecuteOpts(stats)
File "/data/user/dschultz/simulation/build/lib/icecube/simprod/ipmodule.py", line 219, in ExecuteOpts
retval = self.Execute(stats)
File "/data/user/dschultz/simulation/build/lib/icecube/simprod/modules/corsika.py", line 330, in Execute
retval = cors.Execute(self.stats)
File "/data/user/dschultz/simulation/build/lib/icecube/simprod/modules/dcorsika.py", line 371, in Execute
tmpdir = self.stage()
File "/data/user/dschultz/simulation/build/lib/icecube/simprod/modules/dcorsika.py", line 319, in stage
fetch_tarball(meta,cachedir)
File "/data/user/dschultz/simulation/build/lib/icecube/simprod/modules/dcorsika.py", line 161, in fetch_tarball
wget(md5url,os.path.join(cwd,os.path.basename(md5url)))
File "/data/user/dschultz/simulation/build/lib/icecube/simprod/modules/dcorsika.py", line 91, in wget
f = urlopen(strip_auth(url,auth=auth))
File "/cvmfs/icecube.opensciencegrid.org/py2-v2/RHEL_6_x86_64/lib/python2.7/urllib2.py", line 154, in urlopen
return opener.open(url, data, timeout)
File "/cvmfs/icecube.opensciencegrid.org/py2-v2/RHEL_6_x86_64/lib/python2.7/urllib2.py", line 431, in open
response = self._open(req, data)
File "/cvmfs/icecube.opensciencegrid.org/py2-v2/RHEL_6_x86_64/lib/python2.7/urllib2.py", line 449, in _open
'_open', req)
```

Browsable logs on website
Includes links to entire logs.

IceProd 2: Status

Currently in internal beta

Plan to switch to production by year's end

Upcoming features:

- Monitoring
- UI improvements
- File pre-staging

Questions?