


BOINC - an approach to grid (distributed) computing


Centre de Recherche Public Henri Tudor



BOINC – an approach to grid (distributed) computing

LinuxDays 2007 - marc.seil@tudor.lu

1




BOINC History

BOINC
Berkeley Open Infrastructure for Network computing

- ▶ started in February 2002
- ▶ first release 10th April 2002
- ▶ the first BOINC-based project, Predictor@home, started on the 9th June 2004
- ▶ the most known BOINC-based project, SETI@home, started in May 2004
- ▶ 89 developers worked on the sources

2



Success story

SETI@home
the goal of this project is to search for extraterrestrial intelligence

- ▶ first the signals were processed on dedicated single super computers
- ▶ in 1995 David Gedye proposed to use Internet-connected computers.
- ▶ since May 2004 the distributed computing environment switched to BOINC
- ▶ 1,3 million PCs (about 600.000 users)
- ▶ average of 278 TeraFLOP per second (volunteer computing)

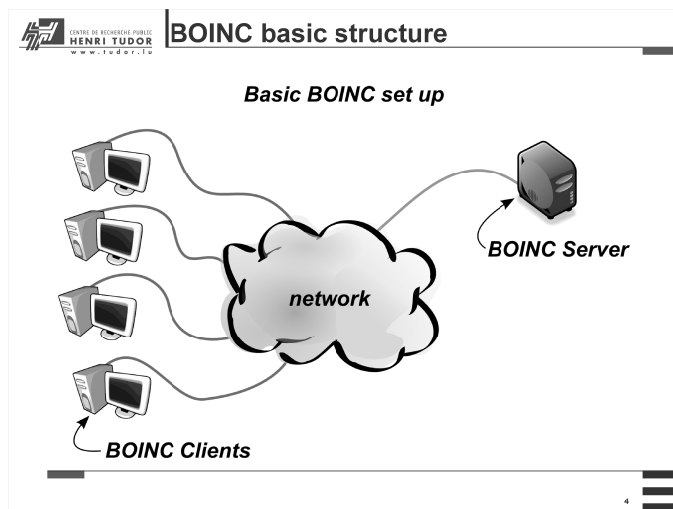
ref: SETI@home and Astronomie.de

3

LinuxDays 2007
marc.seil@tudor.lu

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor



BOINC basic structure

Which type of computing is BOINC?

- ▶ volunteer computing (seti@home - public resource computing)
- ▶ desktop grid computing

5

BOINC as VC environment

Volunteer Computing
the computing resources are donated by users

advantages

- + computing power
- + cheap
- + public interest

disadvantages

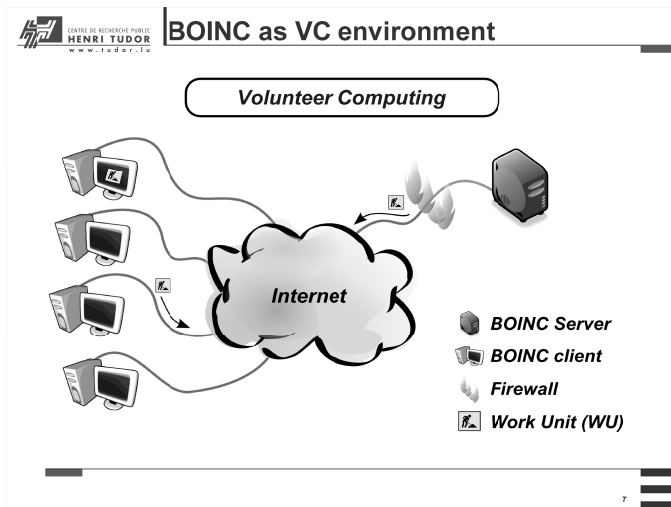
- security
- administration
- availability

- ref: <http://boinc.berkeley.edu/volunteer.php>

6

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor



BOINC as DGC environment

Desktop grid computing
the computing resources are based on existing IT material (like office or department PCs)

advantages

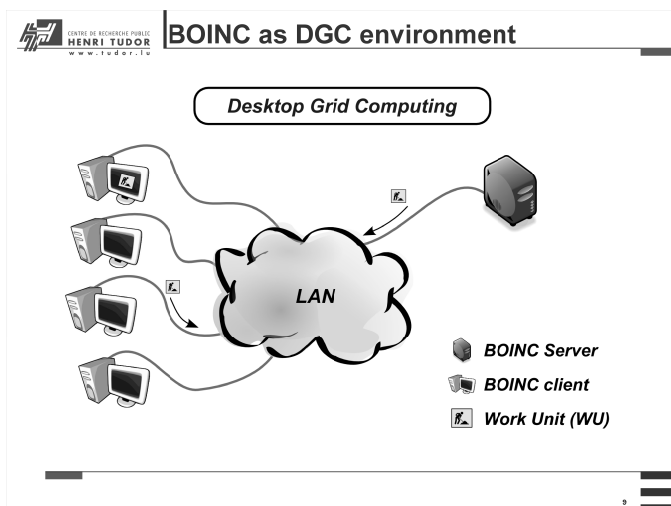
- + reliable resources
- + administration (single account)
- + security

disadvantages

- must be paid
- limited computing power


ref: <http://boinc.berkeley.edu/volunteer.php>

8



BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor


 **Software dependencies**

Which is the best OS to set up a BOINC server
It has to be the Linux® OS

There are some more server side dependencies

- > gnu dev tools (make, autoconf, gcc, automake, ...)
- > python
- > php
- > mysql
- > more dependencies can be checked on the website <http://boinc.berkeley.edu/build.php>

10

 **BOINC WU flow 1**

What is a BOINC WorkUnit (WU)?
A WU in BOINC is the task (job) which is send to the BOINC clients in order to be processed.


Consists of the Parts

- > application
- > dataset

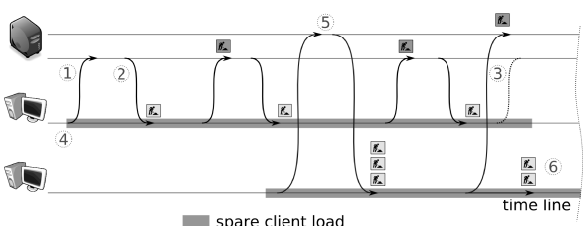
Description

- > work unit template
- > result template
- > identification name

11

 **BOINC WU flow 2**

BOINC WORK UNIT flow



1 2 3 4 5 6

time line


spare client load

BOINC Server Work Unit
BOINC Client Work Unit Result

12

BOINC - an approach to grid (distributed) computing


Centre de Recherche Public Henri Tudor

**BOINC WorkUnit flow 3/3**

Daemons

- **feeder:** is responsible to fill up the ready to send queue with job instances which can be fetched by clients.
- **transitioner:** checks for jobs where the state changed (eg. completed instance). Depending on the situation it may create new instances or flag a permanent error.
- **file_deleter:** removes output and input files which are no longer needed.
- **validator:** compares the result of different instances and defines out of the different instances a canonical instance.
- **assimilator:** handles the jobs that are "completed" and have a canonical instance or which have permanent error.

13


**Backend Software**

Apache2

- install through the adept package manager the tools
 - > apache2
 - > libapache2-mod-php5
- the apache2 configuration can be found in the directory /etc/apache2 and the root directory is /var/www . Logs can be found in /var/log/apache2.
- check if apache2 is running
 - > ps aux | grep apache
 - > with the browser
- check if the php module is loaded correctly with a file info.php which can be accessed through a browser

```
<?php
echo "hello world!";
phpinfo();
?>
```

14

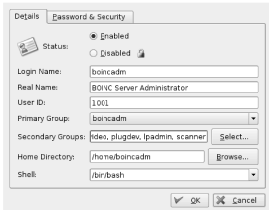
**Backend Software**

Dedicated User

create a user account which is dedicated to run the BOINC server side applications. This can be done in

K-menu->System Settings


- name : boincadm
- pwd: boincadmpwd (not secure)



15

BOINC - an approach to grid (distributed) computing


Centre de Recherche Public Henri Tudor

**Backend Software**

MySQL 1/2

- ▶ install through the adept package manager the tools
mysql-server-5.0, mysql-client-5.0, php5-mysql, php5-mysqli
- ▶ The configuration can be found in /etc/mysql
- ▶ check if mysql is running
 - `ps -aux | grep mysqld`
 - `netstat --list -tcp` (check for mysql or 3306 port)
- ▶ now set the mysql root password (the proposed password is not secure)
`mysqladmin -h localhost -u root password mysqlrootpwd`
- ▶ as user guest try to connect to the mysql server as user root
 - `mysql -h localhost -u root -p` (prompt for password)
 - > show databases;
 - > \q;


16

**Backend Software**

MySQL 2/2

- ▶ create a new mysql user which has the permission to create a database. (as root database user)
 - > `GRANT ALL ON *.* TO 'boincadm'@'localhost';`
 - > `SELECT user,create_priv,password from mysql.user;`
- ▶ define the password of this user to an empty string
 - > `SET PASSWORD FOR 'boincadm'@'localhost' = '';`

17

**Backend Software**

BOINC Sources (as user boincadm!!)

- ▶ now fetch the version 5_9_1 of the boinc core release sources from the cvs repository.
(alternative www.linuxdays.lu/downloads)

```
cvs -d \  
:pserver:anonymous@alien.ssl.berkeley.edu:/home/cvs/cvsroot \  
checkout -r boinc_core_release_5_9_1 boinc
```


BOINC build dependencies (based on kubuntu) 1/2

- ▶ identify the compiler version to use
`strings /usr/lib/libc.a | grep GCC`
or use
`objdump -j .comment -s /usr/lib/libc.a`

18

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor


**Backend Software**

BOINC build dependencies (based on kubuntu) 2/2
download and install the necessary kubuntu packages with the “adept manager” or use “apt-get install `cat package_list.txt`” with the downloaded package list (www.linuxdays.lu/downloads):

m4	autoconf	automake1.9
gcc-4.1	g++-4.1	libssl-dev
libmysql++-dev	libglut3-dev	glut3-dev
libglui-dev	libglitz-glx1-dev	libsdl1.2-dev
libcurl3-dev	freelut3-dev	libsm-dev
libice-dev	libxmu-dev	libxi-dev
libx11-dev	libjpeg62-dev	python-mysqldb
cvs	libtool	vim

http://boinc.berkeley.edu/compile_client.php
<http://boinc.berkeley.edu/build.php>

19


**Backend Software**

BOINC configure and build

- now run `./_autosetup` and `./configure` to set up the source tree to be compiled on the target kubuntu system
- BOINC can now be compiled with `make` (or `gmake`)
- run `./test/test_sanity.py` script to check if all crucial backend software packages are configured and compiled correctly.

The BOINC backend tools are now ready

20

**Backend Software**

Create a BOINC project 1/4

- create a new BOINC projects and key directory as user boincadm.


```
mkdir -p projects/test_seup  
mkdir -p projects/test_setup_keys
```
- fill up the project directory `./tools/make_project`

```
./tools/make_project --user_name boincadm \  
--drop_db_first --delete_prev_inst \  
--project_root $HOME/projects/test_setup \  
--key_dir $HOME/projects/test_setup_keys \  
--db_user boincadm --db_host localhost \  
test_setup 'Test Setup: first BOINC project'
```

21

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor

**Backend Software**


Create a BOINC project 2/4

- configure apache following point 1 of the make_project output and restart apache.
- adjust the group of the test_setup and test_setup_keys directory to apache group (as root)

```
ps aux | grep apache
groups www-data
chgrp -R www-data test_setup
chgrp -R www-data test_setup_keys
```

- check the project website http://localhost/test_setup

22

**Backend Software**

Create a BOINC project 3/4

the directory structure

apps: application directory

bin: applications which are used to manage the BOINC workflows


cgi-bin: stores the file_upload_handler

download: the files which will be send to the clients

templates: the template files describing the workunits

upload: files which are send by the clients

23

**Backend Software**

Create a BOINC project 4/4

some fine tuning

- locate the xml.gif file in the project directory and adjust the path in html/user/index.php


```
find ./ -name xml.gif
vi html/user/index.php
```

- recheck the project webpage

24

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor

**Backend Software**

Create a BOINC user account 1/3


a benefit of BOINC is the scalability and the build in autonomous user (node) management. There exist 3 different methods to create a user account in order to attach a client to a project.

- edit the config.xml file to the proper needs. The file can be found in the project root directory

```
toggle the disable_account_creation to 0
```

- add the phpmailer package to the test_setup/html/inc/phpmailer directory. phpmailer can be downloaded from sourceforge or www.linuxdays.lu

25

**Backend Software**


Create a BOINC user account 2/3

- now adjust some settings in the project.inc file, which is located in test_setup/html/project/ directory

add below the defines

```
define("EMAIL_FROM","boinc@company.net");
$USE_PHPMAILER=true;
$PHPMAILER_HOST="arthur.tudor.lu";
$PHPMAILER_MAILER="smtp";
```

26

**Backend Software**

Create a BOINC user account 3/3

- now create a user account through the web user account creation form.

!! Remember your password !!

```
http://localhost/test_setup/create_account_form
```


- after creation a mail will be send if not simply check the database

```
mysql -u boincadm test_setup
> SELECT email_addr,name,authenticator FROM user;
```

27

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor

**Backend Software**

Adding an application to the project 1/3


the BOINC core release includes a sample applications which can be used to test the set up.

- copy the uppercase application from apps to the test_setup/apps directory following the BOINC naming convention.

```
cd $HOME/projects/test_setup/apps
mkdir uppercase
cd uppercase
# now copy the sample application from the boinc source tree
cp $HOME/boinc/apps/upper_case \
  uppercase_5.0_i686-pc-linux-gnu
```

ref <http://boinc.berkeley.edu/platform.php>

28


**Backend Software**

Adding an application to the project 2/3

adjust the project.xml file (or download it from www.linuxdays.lu/downloads)

```
<boinc>
<platform>
  <name>i686-pc-linux-gnu</name>
  <user_friendly_name>Linux...</user_friendly_name>
</platform>
<platform>
  <name>anonymous</name>
  <user_friendly_name>anonymous</user_friendly_name>
</platform>
<app>
  <name>uppercase</name>
  <user_friendly_name>The ..</user_friendly_name>
</app>
</boinc>
```

29

**Backend Software**

Adding an application to the project 2/2


- now add the application to the project (database) and update the latest version with the commands `./bin/xadd` and `./bin/update_versions`
- check if the applications were added to the project `http://localhost/test_setup` or use the database

```
mysql -h localhost test_setup
> SELECT * FROM APP;
> SELECT * FROM APP_VERSION;
```

30

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor

**Backend Software**


Adding some work to the project 1/4

- create a work unit template file in the templates directory (wu_uppercase.xml)

```
<file_info>
  <number>0</number>
</file_info>
<workunit>
  <file_ref>
    <file_number>0</file_number>
    <open_name>in</open_name>
  </file_ref>
  <delay_bound>600</delay_bound>
</workunit>
```

available on www.linuxdays.lu/downloads

31


**Backend Software**

Adding some work to the project 2/4

- create a result template file in the templates directory (re_uppercase.xml)

```
<file_info>
  <name><OUTFILE_0/></name>
  <generated_locally/>
  <max_nbytes>2000</max_nbytes>
  <upload_when_present/>
  <url><UPLOAD_URL/></url>
</file_info>
<result>
  <file_ref>
    <file_name><OUTFILE_0/></file_name>
    <open_name>out</open_name>
  </file_ref>
</result>
```

32

**Backend Software**

Adding some work to the project 3/4

- copy an ascii input file to the download directory (or create a file)
- now add the WU to the project with the project tool
bin/create_work in the project root directory


```
./bin/create_work -appname uppercase
-wu_name wu_01 -wu_template templates/wu_uppercase.xml \
-result_template templates/re_uppercase.xml \
-min_quorum 1 -target_nresults 1 about.txt
```

available on www.linuxdays.lu/downloads

33

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor

**Backend Software**

Adding some work to the project 4/4

the prior steps can be checked in the database column workunit or they can be checked through the "Project Management" web interface.


Database

```
> SHOW COLUMNS FROM workunit;  
> SELECT id,name,xml_doc FROM workunit;
```

WEB frontend

http://localhost/test_setup_ops

34

**Backend Software**

Adjusting user and group rights


as some of the directories have to be writeable and readable by the owner who is running the apache server some adjustments have to be made

- ▶ change the group of the complete directory `project/` in respect to the primary group (gid) of the running apache process if these are not already configured

```
ps aux | grep apache # the first column  
id www-data  
chown boincadm:www-data -R projects/*
```
- ▶ the Set Group ID of the binary `./bin/feeder` has to be set to allow shared memory access to the group `www-data`

```
chmod g+s ./bin/feeder
```

35

**Backend Software**

Adjusting the project configuration


- ▶ Check the configuration file `config.xml` if all daemons are present to start the BOINC server side processes (feeder, file_deleter, transitioner)

```
<daemons>  
  <daemon><cmd>feeder -d 3</cmd></daemon>  
  <daemon><cmd>transitioner -d 3</cmd></daemon>  
  <daemon><cmd>file_deleter -d 3</cmd></daemon>  
</daemons>
```

36

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor




CENTRE DE RECHERCHE PUBLIC
HENRI TUDOR
www.tudor.lu

Backend Software

Starting the BOINC server processes

- ▶ The best method to start and stop the BOINC processes are `./bin/start` and `./bin/stop`
- ▶ The log files used by the daemons can be found in the `log_hostname/` directory
- ▶ a good monitoring in the beginning is
`tail -f log_*/*`

37




CENTRE DE RECHERCHE PUBLIC
HENRI TUDOR
www.tudor.lu

Backend Software

Attaching a client to the BOINC project

- ▶ Download the BOINC client application from the project website or from www.linuxdays.lu/downloads
- ▶ Unpack the client in a terminal with `sh boinc_*5.4.11*.sh` and a new directory `BOINC/` will be created with the client application
- ▶ Attach to the `test_setup` project with the prior created user account (`./run_client` or the GUI version `./run_manager`)

38



CENTRE DE RECHERCHE PUBLIC
HENRI TUDOR
www.tudor.lu

Backend Software

Securing the management web frontend

the authentication facility of apache2 can be used


- ▶ create a password file in the `html/ops` directory
`htpasswd2 -c password boincadm`
- ▶ now add a `.htaccess` file in the same directory with the following contents

```
AuthType Basic
AuthName "By Invitation Only"
AuthUserFile
/home/boincadm/projects/test_setup/html/ops/password
Require user boincadm
```

39

BOINC - an approach to grid (distributed) computing

Centre de Recherche Public Henri Tudor


**Backend Software**

Work and Debugging

update the config.xml file with some daemon applications

```
./bin/stop
<daemons>
  <daemon><cmd>
    make_work -wu_name name_of_existing_wu -cusion 5
  </cmd></daemon>
  <daemon><cmd>
    sample*_validator -d 3 -app uppercase
  </cmd></daemon>
  <daemon><cmd>
    sample_dummy_assimilator -d 3 -app uppercase
  </cmd></daemon>
</daemons>
```

40

**Gather infos**

Official

<http://boinc.berkeley.edu>

Unofficial Wiki

<http://boinc-wiki.ath.cx>

Message Board

<http://boinc.berkeley.edu/dev>

Bundled documents and papers

doc/ (bundled documents with source code)
<http://boinc.berkeley.edu/papers.php>

41