Distributed Computing in the Cloud

BOINC and Amazon's Elastic Compute Cloud (EC2)

Maximilian Palm
Rechenkraft.net e.V., Marburg, Germany
www.rechenkraft.net

22.03.2011

Table of contents

1 Preface

This document is intended for engaged volunteers in distributed computing (DC). Those people spent time and money to support charitable science projects and also to satisfy their own curiosity. The author will show a new way of supporting DC projects using the example of the Berkeley Open Infrastructure for Network Computing (BOINC)[1] and the Amazon Elastic Compute Cloud (EC2)[2].

2 Introduction to Amazon's Cloud

The Amazon Elastic Compute Cloud (EC2) offers flexible virtual servers called instances. As a customer, one pays only for the time each instance is running. Administration, e.g. starting or stopping instances, can be done using the browser based Amazon Management Console[3], a Firefox Plug-In[4] or talking directly to the Web Service API.[5]

2.1 Instance types

Several types of instance exist to meet demands of different applications:

- *Micro instance*: a cheap, tiny server with 613 MB of memory and little CPU resources

- *Standard instance*: balanced between memory and CPU resources in different sizes

- *High-Memory instance*: big servers with 17.1 GB, 32.2 GB or 68.4 GB of memory

- *High-CPU instance*: servers with emphasis on CPU resources

The CPU resources are measured in EC2 Compute Units (ECU). "One EC2 Compute Unit provides the equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor." [6]

2.2 Pricing models

Concerning the costs, there are three pricing models available:

- *On-demand instance*: This is the standard product. Customers pay the regular price for each hour an instance was running.

- *Reserved instance*: Customers can choose to pay an initial fee for an instance to get a discount on the hourly price for one or for three years.

- *Spot instance*: Customers can also bid on the price of otherwise unused EC2 resources. Amazon calculates the spot price for each type of instance regularly and everyone who holds at least that price gets an instance. Note that it is possible that a spot instance is shut down when the spot price exceeds the customer's maximum bid, so this offer involves a certain risk in terms of price and availability.

For all pricing models, costs also vary depending on the location of the instance. US East (N.Virginia) is usually the cheapest data centre, the other ones (e.g. in Ireland or Tokyo) are a bit more expensive.[7]

---

1 Website of the BOINC project: http://boinc.berkeley.edu/
2 Website of the Amazon Elastic Compute Cloud: http://aws.amazon.com/ec2/
3 With a registered account, you can find the Management Console at https://console.aws.amazon.com/ec2/home
4 For the Firefox Plug-In "Elasticfox, see http://aws.amazon.com/developertools/609
5 The API documentation is at http://docs.amazonwebservices.com/AWSEC2/latest/APIReference/
6 An overview of all instance types and resources can be found at http://aws.amazon.com/ec2/instance-types
7 To see all current pricing information, please see http://aws.amazon.com/ec2/pricing

When requesting an instance, customers can choose from several operating system images called Amazon Machine Images (AMI). Many Linux distributions as well as Microsoft Windows are available. It is possible to make an instance persistent even after it was terminated. This is done using the Amazon Elastic Block Store (EBS), but will create additional cost.[8]

3 Tests and results

Several instances where tested to find out how much performance EC2 servers have to offer and what the overall costs per CPU hour are. BOINC uses credits to compare the amount of contribution. Since many volunteers are eager to earn lots of credits, especially in races for their teams, the author also provides the cost per 1000 BOINC credits.

3.1 Software

The following software was used during the tests:

- Debian GNU/Linux "base" image (32 or 64 Bit, depending on instance type)

- BOINC 6.2.14 from the Debian repositories

  - The cc_config.xml file was used to make BOINC report finished results immediately. This setting is not recommended because it puts unnecessary load on the project server. However, for this test it was beneficial to know when each individual result was finished.[9]

- A shell script was used to get each server running BOINC as soon as possible.[10]

- World Community Grid's "Help Conquer Cancer" was chosen as the benchmarking project since it runs very stable and the workunits are rather short. All instances were run as spot instances at the US East data centre to achieve the lowest possible price.

3.2 Results

Each tested instance simply ran BOINC for a while. Afterwards, results were collected from the Amazon EC2 Account information website and the World Community Grid project website. All prices are given in US Dollars and without value added tax (VAT).

Type of instance	Price ($/h)	Uptime (h)	Actual CPU time (h)	Number of results	Earned credits
Micro	0.0100	77.76	3.06	3	72
Standard Small	0.0302	26.63	10.07	10	240
High-CPU Medium	0.0610	27.83	48.66	43	830
High-CPU xLarge	0.4600	17.56	126.14	104	2235

Table 1: Overview of the test results (Source: Compiled by the author)

There were no computing errors or invalid results. On the multi-core instances, there is always a little bit of wasted time at the end of each test because some started results do not get finished.

---

8 For more information on EBS, visit http://aws.amazon.com/ebs

9 To find out more about cc_config.xml in BOINC, visit http://boinc.berkeley.edu/wiki/Client_configuration

10 The shell script can be found in the appendix of this document.

4 Cost-benefit analysis

Goal of this chapter is to see what the EC2 can offer in terms of CPU resources and at which price. The following diagrams show the average cost per CPU hour and per 1000 BOINC credits respectively for each tested type of instance.
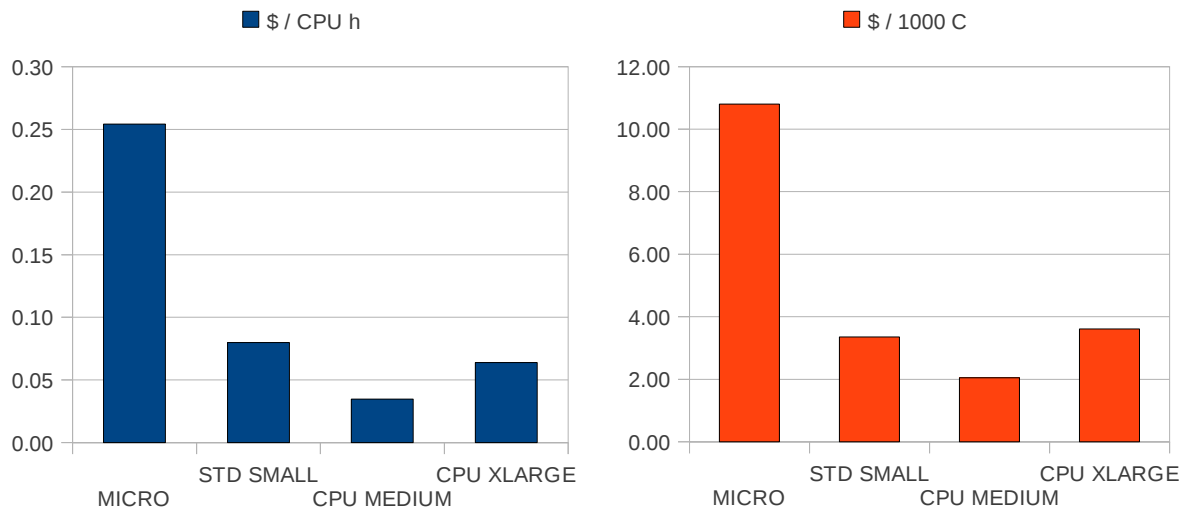

Figure 1: Cost of CPU resources for tested instance types (Source: compiled by the author)

Surprisingly, the High-CPU Medium instance is cheaper than the High-CPU xLarge.[11] This is due to high fluctuation in spot prices for this particular instance type in the US East data centre. The xLarge instance usually costs about $0.24/h, but there are peaks of up to $0.80/h, which is higher than the On-Demand price.
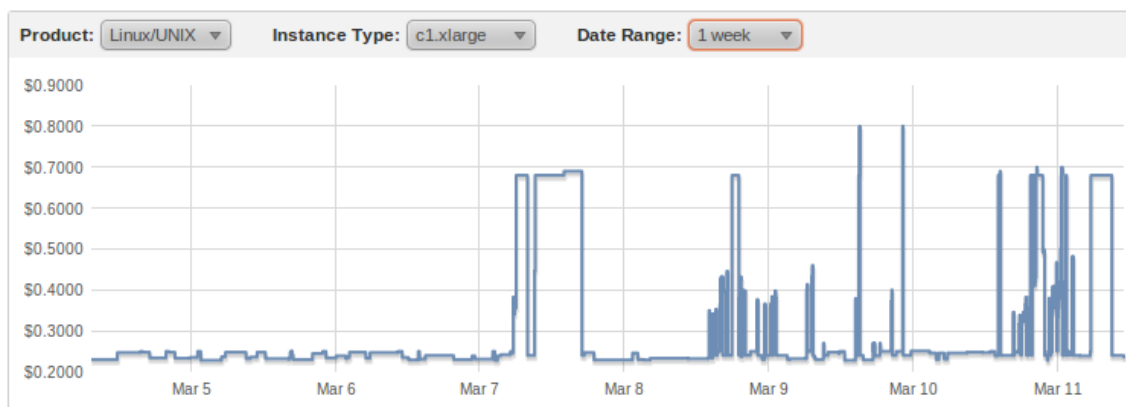

Figure 2: Spot price history for a High-CPU xLarge instance in US – East (Source: Amazon AWS Management Console, 11.03.2011)

At the cheapest alternative data centre in the region US West (N. California), the average spot price would have been 0.32$/h instead of the actual average of $0.46/h, that were paid by the author in US East. In terms of $/CPU hour, this is still more expensive than the High-CPU Medium instance. Only if the peak prices in the US East data centre disappeared, the average spot price of then $0.24/h would result in a lower cost per CPU hour ratio ($0.0334/CPU h).

---

11 Detailed results can be found in the appendix

The cheapest instance type in terms of CPU resources in the test was the High-CPU Medium instance in the US East data centre, which costs about $0.0349/CPU hour or $2.04/1000 credits.

5 Conclusion

Distributed Computing in the cloud is not only possible, but also easy to set up. However, with average cost of $10.20 per week and an output of about 5000 credits for one High-CPU Medium instance, the EC2 is still very expensive. Most home-made dedicated computing hosts or clusters can be run at a better cost per credit ratio, although this depends on energy prices, value added tax and so on. Even for really engaged volunteers, this way of contributing to a distributed computing project may only be interesting to temporarily expand computing capacities, e.g. for races.

6 Appendix

6.1 Note concerning the micro instance

The micro instance was not actually charged by Amazon since there is a one year free trial for it. Cost for this instance were taken from the spot price history, not from the billing information. The author will continue to run BOINC on the free micro instance during the one year period.

6.2 Server initialisation script

Please insert the desired project URL and the account key to be able to use this script.

```
#!/bin/bash -x


if [ $# -ne 2 ]; then
  echo "Usage: `basename $0` hostname private_key"
  exit 1
fi


SSH="/usr/bin/ssh -i $2 root@$1 "


$SSH "apt-get update"
$SSH "apt-get -y install boinc-client"


sleep 5


$SSH "/etc/init.d/boinc-client stop"
sleep 5
scp -i $2 ./cc_config.xml root@$1:/etc/boinc-client/cc_config.xml
sleep 1
$SSH "/etc/init.d/boinc-client start"
sleep 5
$SSH "boinccmd --project_attach <project_url> <account_key>"
```

6.3 Detailed data from the tests

Micro instance in US East

CPU hours:	**3.06**
real hours:	**77.76**
earned credits	**72**
price per hour	**0.01**
Total Price:	**0.78**
price per 1000 C	**10.80**
price per CPU hour	**0.2541**
Credits / h	**0.93**
Credits / cpu h	**23.53**
Utilisation (%)	**3.93**

Standard small instance in US East

CPU hours:	**10.07**
real hours:	**26.63**
earned credits	**239.9**
price per hour	**0.0302**
Total Price:	**0.80**
Price per 1000 C	**3.35**
price per CPU hour	**0.0799**
Credits / h	**9.01**
Credits / cpu h	**23.82**
Utilisation (%)	**37.82**

High-CPU Medium instance in US East

CPU hours:	**48.66**
real hours:	**27.83**
earned credits	**830.3**
price per hour	**0.061**
Total Price:	**1.70**
Price per 1000 C	**2.04**
price per CPU hou	**0.0349**
Credits / h	**29.83**
Credits / cpu h	**17.06**
Utilisation (%)	**174.83**

High-CPU xLarge instance in US East

CPU hours:	**126.14**
real hours:	**17.56**
earned credits	**2235.9**
price per hour	**0.46**
Total Price:	**8.08**
price per 1000 C	**3.61**
price per CPU hour	**0.0640**
Credits / h	**127.32**
Credits / cpu h	**17.73**
Utilisation (%)	**718.27**

Some of the CPUs the BOINC client reported during the tests:

- GenuineIntel Intel(R) Xeon(R) CPU E5506 @ 2.13GHz [Family 6 Model 26 Stepping 5]

- GenuineIntel Intel(R) Xeon(R) CPU E5410 @ 2.33GHz [Family 6 Model 23 Stepping 10]

- GenuineIntel Intel(R) Xeon(R) CPU E5430 @ 2.66GHz [Family 6 Model 23 Stepping 10]