# Volunteer Computing at CERN

- SUSTAINABILITY MODEL PROPOSAL -

*Openlab Student:*
Daniela DORNEANU,
daniela.dorneanu@cern.ch

*Supervisors:*
Ben SEGAL,
Miika TUISKU,
Francois GREY

VERSION 0.1
September 12, 2011

# Contents

# 1 Executive summary

This report provides an analysis and evaluation of the possible sustainability models (business models) for Volunteer Cloud Computing. Currently on one side there are millions of volunteers willing to share their CPU power and on the other side there are scientists who need this CPU power but don't really know how to obtain this from the volunteers.

Thus the main purpose of this report is to suggest different institutional arrangements that would lower the entry barrier for scientists to Volunteer Cloud Computing, and in particular present such resources to them as a service that was fully compatible with the way they use their current computing infrastructure. In order to achieve this goal we first make an analysis of what Volunteer Computing represents today in terms of technology limitations and costs that should be taken into account when starting to use this infrastructure.

The method of analysis includes running surveys and having direct interviews with some of the main stakeholders in this domain. We have also used current research documentation and tools in order to compute costs, earnings and measure marketing outreach for different cases.

We found out that the four most important assets of Volunteer Cloud Computing that can be used to create sustainability are:

- The huge community of volunteers behind it, 6.6 million registered volunteers, many being technology very savvy.
- Almost free computation power, the overhead for accessing the volunteer resources being minimal.
- No major modifications needed to the current scientific project's workflow.
- Compatibility with the main commercial cloud, Amazon.

The last two of these assets have recently been developed thanks to a project called LHC@Home at CERN. The term Volunteer Cloud Computing, in the context of this report, refers to traditional volunteer computing combined with these two new benefits.

Based on these main assets we explored four possible Use Cases in which Volunteer Cloud Computing can achieve the dual purpose of lowering the cost of scientific computing and providing the scientists concerned with a significant outreach opportunity.

- Volunteer Cloud Computing for CERN (single institution implementation)
- Volunteer Cloud Computing for Industry (R&D type usage by large companies)
- Nonprofit Volunteer Cloud Computing (A foundation to help spread the technology)
- Commercial Volunteer Cloud Computing (A start-up company to provide the service)

For each Use Case we describe the main benefits for the customers (scientists) and other stakeholders, we formalize the business model using the 'The Business Model Canvas' and we make a SWOT analysis in order to estimate the likelihood of success of the Use Case.

The report finds that:

- There are several opportunities to implement Volunteer Cloud Computing as a sustainable model and states them in terms of the 4 possible Use Cases.
- The main barriers to do this are related to potential users/customers not being familiar with the Volunteer Cloud Computing concept and concern that volunteers would lose interest if they hear that a for-profit company is involved.

The most important recommendations for successful implementation of Volunteer Cloud Computing are:

- investing adequate effort to inform the volunteers clearly of the scientific purpose and results of the project;
- making the customers aware of the major outreach opportunity that a public project based on volunteers represents;
- making the customers and their funding sources more aware, in detailed quantitative terms, of the excellent cost/performance relation when using Volunteer Cloud Computing, compared with other options;
- making the customers aware of the wide range of profiles and motivations of the volunteers who contribute to projects, including a significant dedicated minority who will support a project with high level computer skills.

The information presented in this report has some limitations of accuracy due to the lack of rigorous documentation regarding current computational costs (hardware and personnel) and in some cases due to difficulties to get certain representatives to answer to our questions.

# 2 Introduction

## 2.1 What is Volunteer Computing?

Volunteer Computing is a concept that defines the process when users are voluntarily sharing their CPU for a certain project. The basic process between all the volunteer computing software has the following scheme: a client runs on the volunteer machine and from time to time contacts the project's server to report the computed results and requests new jobs. This model is mostly used for CPU intensive jobs which can be divided in smaller tasks which can be done by different hosts.

Usually the Volunteer Computing projects try to solve pressing health and environmental problems or physics experiments that would explain how the universe was created or to get in contact with extraterrestrial intelligence. Volunteers are interested to contribute to these projects mainly due to the problem or issue that these projects are interested to solve. Since 2002 the most known and used middleware for volunteer computing is BOINC - Berkeley Open Infrastructure for Network Computing which was also adopted by the largest public computing grid - World Community Grid. Currently there are 2.2 millions of registered BOINC users and over 6.5 millions connected hosts. This impressive number of hosts can compete in number of servers with any giant IT company like Amazon, Google, eBay, Yahoo, Microsoft, IBM, HP/EDS, GoDaddy. [15]

However there are some aspects about Volunteer Computing that are still considered weak points for this type of cloud compared with the commercial cloud alternative. These points are related to

- the anonymous status of the volunteers,
- the insecurity added by each of the users machine in the cloud
- malicious users that return bad results on purpose
- the workload should be fault tolerant as a work unit failure is not so rare
- workloads should have low IO rate - volunteers offer no guarantees regarding network bandwidth

All these factors cannot really be fully controlled like in a homogeneous server cluster. However considering that the BOINC community has currently more than 100 volunteer developers[17] and even more active testers we can consider this middleware as a robust open source software platform.

## 2.2 What is a Volunteer Cloud?

'Since 2000, over 100 scientific publications have documented real scientific results achieved on this platform'[9]. Considering this huge volunteer resource (2.2 millions BOINC users) first questions that arise are: Why is this resource not used by more scientific projects? or Why do scientific group still need to have huge computer centers?

It is proved that the issue currently is not on the number of volunteers but on the number of projects that use this resource. One reason for this is that Volunteer Computing power is mostly used up to now for simulation purposes and solving tasks that need CPU high computation rather than for managing and processing data. Therefore BOINC middleware is specialized in this kind of tasks. Another reason for not having more scientists using Volunteer Computing resources concerns the entry efforts for porting a project into this environment. The entry and maintenance costs for using BOINC involve the following factors [16]:

- adapting the software to each volunteer platform (portability) and linking it with BOINC libraries. (this can be a very heterogeneous environment - starting from different hardware support up to different operating systems)
- creating a system for job submission.
- recreating the software for the scientific projects every time the experiment's code changes.
- encapsulating the work in small tasks that can be sent and executed independently of each other
- acquiring and maintaining an infrastructure for hosting the project

These factors become even more challenging for CERN's projects as the applications developed for physics projects are very big ( 10 GB) and have very frequent update cycles ( days). Also these applications are very difficult to modify and have to run on a specific Linux flavor. Considering that any user can become a volunteer, another issue concerns data sensitivity. For solving some of these challenges CERN has initiated a project to encapsulate the tasks of the project inside virtual machines. This level of encapsulation solves many problems related with the heterogeneous environment of the users' machines, with porting the research project to BOINC infrastructure and with handling dynamic changes of the software transparently. Due to virtualization the physicist will just have to take into account the environment provided by the virtual machines and port their application to this environment. We consider that this offers a "cloud interface" to all applications, therefore we called this technology developed at CERN a "Volunteer Cloud".[16]

## 2.3 Volunteer Cloud characteristics

This section is meant to briefly describe the technology that integrates virtualization with BOINC. We include this paragraph here as we consider that this technology is an important step for the portability of the projects integrated with BOINC and for easing the integration process. Also we consider that virtualization encapsulation is the value added by CERN over the BOINC infrastructure and we will include this as an important assumption for our model proposals.
The main advantages offered by the Volunteer Cloud technology are:

- a framework compatible with the commercial cloud infrastructure of Amazon
- the portability issue is solved and the experiments' software supported by a virtualization system called "CernVM". So heterogeneous environment of the hosts is not an issue any more.
- the experiment code can be dynamically updated without bothering the volunteer
- the insecure nature of the volunteer medium is solved by using a trusted intermediate job management layer: "CoPilot".

The main achievements of this project until now are [16]: 2000 active users gained credit, over 3000 machines successfully handled, approximately 10000 jobs running on daily basis, a lot of feedback and media interest, a big interest from BOINC volunteers. Recently CERN assigned continuous manpower to the project (up to now the project has been developed only with summer students and part time working fellows),

Some more in-depth description of CernVM and CoPilot can be read in the Section 7.1. Much more detailed analysis of this technology has been provided by the developers and maintainers of CernVM and CoPilot and by the CTO of the Volunteer Cloud project at Cern - Ben Segal.

## 2.4 Case study objectives

One very important question that we will try to answer in this report is -'Can this big community of volunteers be used to create a sustainable facility for research projects, and if yes how?' As this expresses our goal in a very general way we will list our main objectives as answers to the following questions:

- What are the possible models for making a Volunteer Cloud initiative sustainable and grow it beyond its first main pilot LHC@home 2.0 version 1 ? Who could be the beneficiaries? Who are the stakeholders? Who are the customers? Who are the volunteers? Why would they invest their resources in this?
- Which are the possible areas to use Volunteer Cloud Computing in research environment and why would ICT companies be interested in investing in this project?
- How can CERN benefit from Volunteer Cloud Computing? What is CERN's expertise for initiating and implementing this project?
- What are the infrastructure and personnel investments for implementing a sustainable Volunteer Cloud Computing system?
- What could be the institutional way to implement this initiative inside CERN? What are CERN's benefits in terms of cost savings for implementing a Volunteer Cloud?

# 3 Analysis

For the analysis of the current situation we were interested to gather information about the following topics:

- What are the current costs to maintain and extend the current infrastructure of various sized research projects using Volunteer Cloud Computing, Amazon Cloud or Private clusters? For what type of computation can Volunteer Cloud Computing can be used?
- What are the profiles of the volunteers and if they would start sharing resources also for a commercial project?
- What are the interests and needs of the research groups and how do they see the Volunteer Computing power?
- Which are the current start-ups that involve volunteer computing or are using cloud computing for humanitarian purposes and what are the business models behind these?

For this we used the following methods: surveys and interviews for main stakeholders, brainstorming, consulting the statistics, research work and information that is already available on the web.

## 3.1 Surveys and interviews for main stakeholders

We can classify the persons we interviewed depending on the type of project they are currently working in as follows:

a. Current BOINC supporters - (David Anderson - BOINC, Matt Blumberg-GRID Republic, Bill Bovermann, Kevin Reed - World Community Grid)
We considered this target group in our research because we were interested to find out as many information as possible regarding what are the future perspectives of the BOINC representatives, what is their current load of work, how do they see marketing in this field, what were the lessons they learned and what is their opinion about this initiative.

b. Commercial start-ups - (Techila, E-science Central, CycleComputing, Wuala, CloudBroker, SuperDonate, XtremeWebCh, Pluraprocessing.com, )
There are many commercial start-ups out there from where we can have success or failure stories we can learn from. Therefore we tried to contact as many stakeholders and ventures initiators in this domain to ask about their business idea and their plan to have financial sustainability. Most of these approaches are quite different therefore this means that there are still a lot of approaches for Cloud Computing and Volunteer Computing to be experimented. The outcome of this research is presented in Table 4.1 Table 4.2 from the fourth use case we presented.

c. Universities and research groups - Research groups are the central beneficiaries the of this model so finding out what are their interest and what approach is more convenient for them is one of the most important goals of this research. Therefore besides the interviews we have also made an online survey to be filled by research group representatives. The results of the survey are presented in Section **??** and more in-depth details can be found in Section **??**.

d. CERN IT staff - CERN OpenLab representatives (Markus Schulz, Sverre Jarp, Ben Segal, Miika Tuisku)
CERN has a central role in this initiative and is also the place where the CernVM + CoPilot started development. Consulting CERN representatives is the starting point for searching different institutional ways to continue this project under a CERN initiative.

## 3.2 Volunteers Profiles

One of the main threats when involving companies to invest in a Volunteer Cloud Computing project is that once the volunteers will find out that commercial companies are also using their idle CPU they may not be willing to share their resources for these projects. During our research for this report we approached and debated this theme several times. We consider that taking into account their main motivation to share idle CPU the volunteers can be classified in the following way:

- technology geeks - these are the volunteers that contribute actively in the forum of the project with technical advice. Some of them are part also of the Boinc testers and developers community;
- followers of Science- attached to the scientific goal of the project;
- people who want to save the planet - attached to the humanitarian goal of the project;

- social networks - people that are motivated to contribute due to their social network; This category is very motivated by the credit points and being a member of the community.

For this classification - Figure 3.1 we received similar arguments from Matt Blumberg - the initiator of Grid Republic and Charity Engine. Therefore even if part of the users won't be motivated to join also commercial projects there will be still a very important community, from the 6.6 millions users, willing to be part in projects that test or implement new technologies.

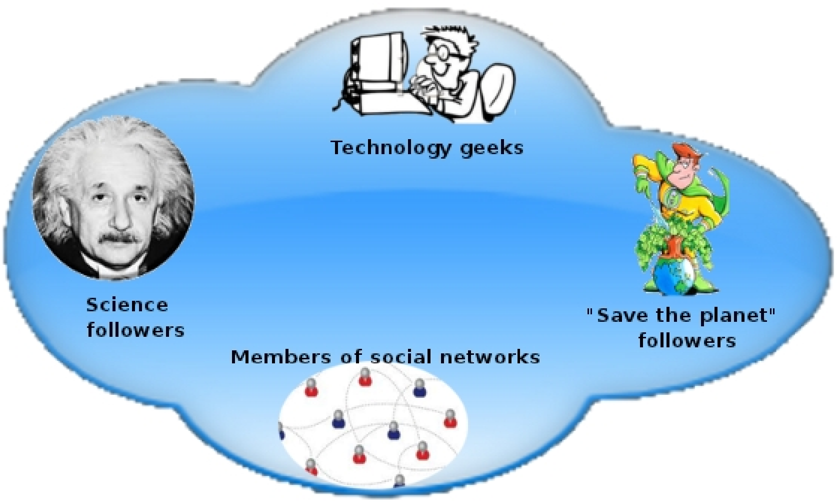



Figure 3.1: Volunteers' Profiles

## 3.3 Brainstorming

For the purposes of this report we have successfully used brainstorming sessions in which we have gathered ideas about the shape of each business model and we have evaluated the opportunities and threats for each use case. The techniques used during the business model design process were:

- face to face meetings and presentations where we expressed ideas using the Business Model Canvas, Story Telling, SWOT analysis, experience sharing, feedback. The participants to these brainstorming sessions were the initiators of this project and representatives of CCC and LHC@home 2.0.

- online brainstorming using: documents and drawings sharing and direct e-mails.

# 4 Use Cases

The models subject to this research are:

1. Volunteer Cloud for CERN: The Volunteer Computing server's administration would be maintained and run directly inside CERN for CERN's scientific projects.

2. Volunteer Cloud for Industry: For this context we will analyze possible situations big companies would be interested to use the research and development process from CERN to test the Volunteer Cloud environment - following the OpenLab model.

3. Nonprofit Volunteer Cloud: This model describes how a foundation can help spread the technology for scientific purposes. E.g: the Citizen Cyberscience Center

4. Commercial Volunteer Cloud: This model describes how a start-up company could use the technology to provide Volunteer Cloud as a service.

**Short guide to read our detailed model**

**- The Business Model Canvas**

We described each model using a very simple formalization scheme called - The Business Model Canvas. Few words about the steps to read our plan easier.

Step 1. The central element Value Proposition describes the bundle of products and services we offer to the Client Segment.

Step 2. The client perspective specifies exactly who are the Customers, what are the Relationships we are going to create with them and what are the Channels used for communicating with them.

Step 3. The activity perspective specifies the Partners we are counting on in our model, which are the unique Key Resources and what are the Key Activities for achieving the Value Proposition.

Step 4. The Financial Perspective very briefly describes the cost structure for our model and what will be the sources for the Revenue Streams.

**- The SWOT Analysis**

To evaluate the success rate of each model we used the SWOT analysis. SWOT analysis diagrams succinctly show the Strengths, Weaknesses, Opportunities and Threats that help or harm one's ability to reach a defined objective. Analyzing internal forces (Strengths/Weaknesses) and external forces (Opportunities/Threats) helps one define strategies that will help leverage positive forces and overcome obstacles in the course of reaching the objective.

## 4.1 Volunteer Cloud for CERN

In this context CERN would sustain the maintenance costs, and physics projects developed at CERN will be the direct beneficiaries of Volunteer Cloud. (e.g. CMS, Atlas, Theory Projects)

The main benefits CERN would have from a Volunteer Cloud are:

- It offers almost free computation power for CERN's research purposes.
- The administrative infrastructure already exists and can scale very easily.
- The maintenance costs are very attractive for big and medium size projects that run at least 3-4 months and have computational needs over 1500 CPUs. Comparative development costs are detailed below in Figure 4.2.
- Volunteer community identity and outreach value. Comparative media outreach is pictured in Figure 4.3.

In Figure 4.1 we formalized this model using The Business Model Canvas and in Figure 4.1 we analyzed this model using the SWOT analysis. (more details about The Business Model Canvas and SWOT analysis in the beginning of Chapter 4).

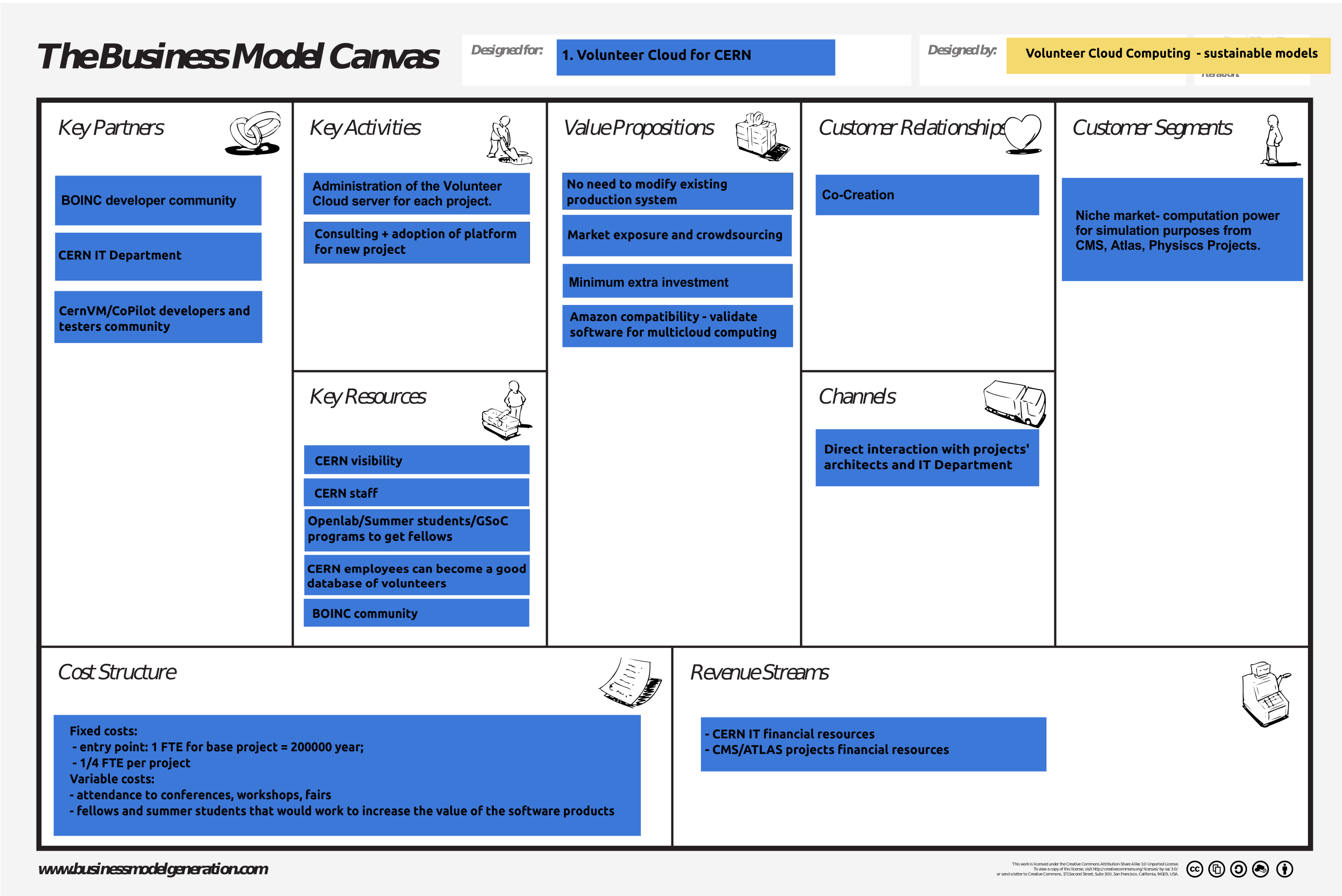



Figure 4.1: The Business Model Canvas - 1. Volunteer Cloud for Cern

To support our statement above regarding the cost advantage, we will provide a rough estimate of how much it costs the simulation power for one month using different alternatives:

- extending a new **T2 Grid**: refers to the costs for 2011 for maintainance and expansion cluster center in the Cern T2 Grid. T2 is the level where most of the simulation work is done and we considered that 30 % of the computation done here is for simulation purposes. We considered this for different countries in Europe. The price differences are due to personnel and cluster hosting costs. The data was provided after analyzing CERN's documentation regarding costs/performance and it represents a very rough approximation for now. The references we used for this were the CMS dash board [23] and reports about the expected costs per country. [24]

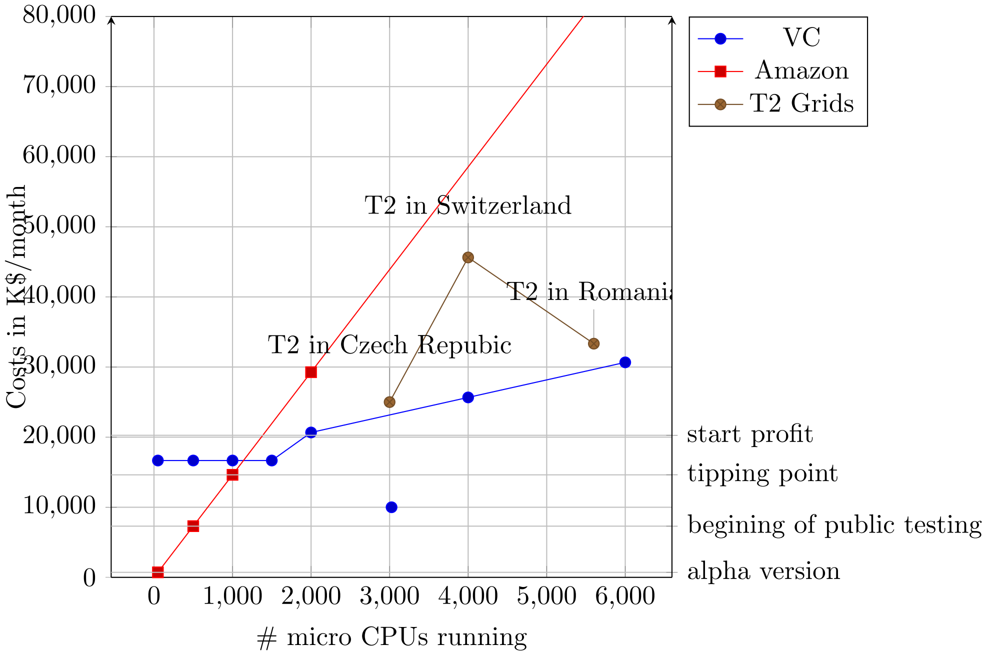



Figure 4.2: Comparative costs - micro CPU instances/month Volunteer Cloud vs Amazon vs T2 Grids

- buy computation from **Amazon**: means acquiring computation power from Amazon. For this information we used the monthly Amazon calculator.
- use the **Volunteer Cloud**: refers to using the infrastructure available through Volunteer Cloud. We considered in this case that the personnel needs would be 1 FTE for the basic system operation and then 1/4 FTE for setting up each new project. We ignored the infrastructure and marketing costs as we assume that we already have the necessary infrastructure and we will have enough marketing from CERN press releases. (more details about media coverage below).

As seen from Figure 4.2 it becomes interesting from a financial point of view to use the Volunteer Cloud if the necessary amount of simulation CPU power done in this this type of cloud exceeds 1500 CPUs running at 100% load factor.

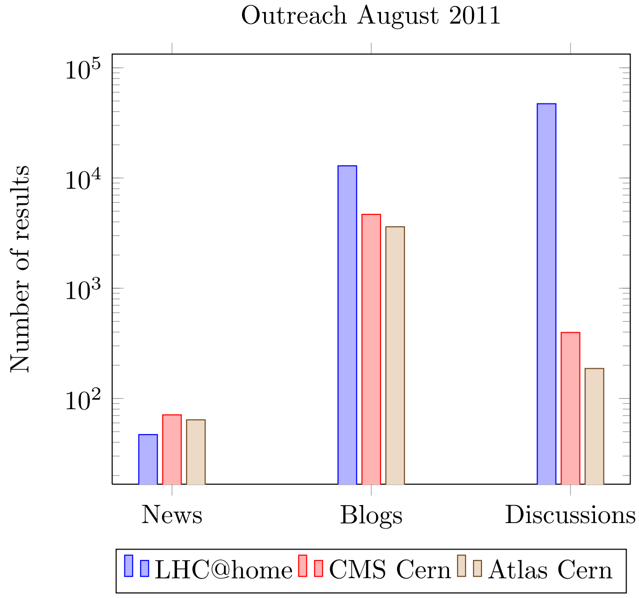



Figure 4.3: Comparative outreach August 2011 for - LHC@home, CMS Cern, Atlas Cern

One very interesting point for attaching CMS or Atlas projects to the Volunteer Cloud is the outreach these projects would gain. Being able to contribute directly every day with CPU power would form community of users that would identify themselves with a certain project. In the Figure 4.3 it can be seen that LHC@home has already half the number of online news bulletins compared with Atlas and CMS experiments (LHC@home - 47 news, CMS Cern - 71, Atlas Cern - 64 news).

However the number of blogs and discussions on the LHC@home topic is respectively 4 times, and 10 times bigger than for the Atlas and CMS experiment. This proves that having the volunteers contributing directly to the experiment leads to a very good market exposure for the project.

Next you can follow the SWOT analysis for implementing this model. As pictured in Figure 4.4 in order to increase the success rate of the project we have to make very good use of the financial and outreach strong points and find the best ways to approach Cern IT Department and the representatives of CMS and Atlas project to overcome the main threats.

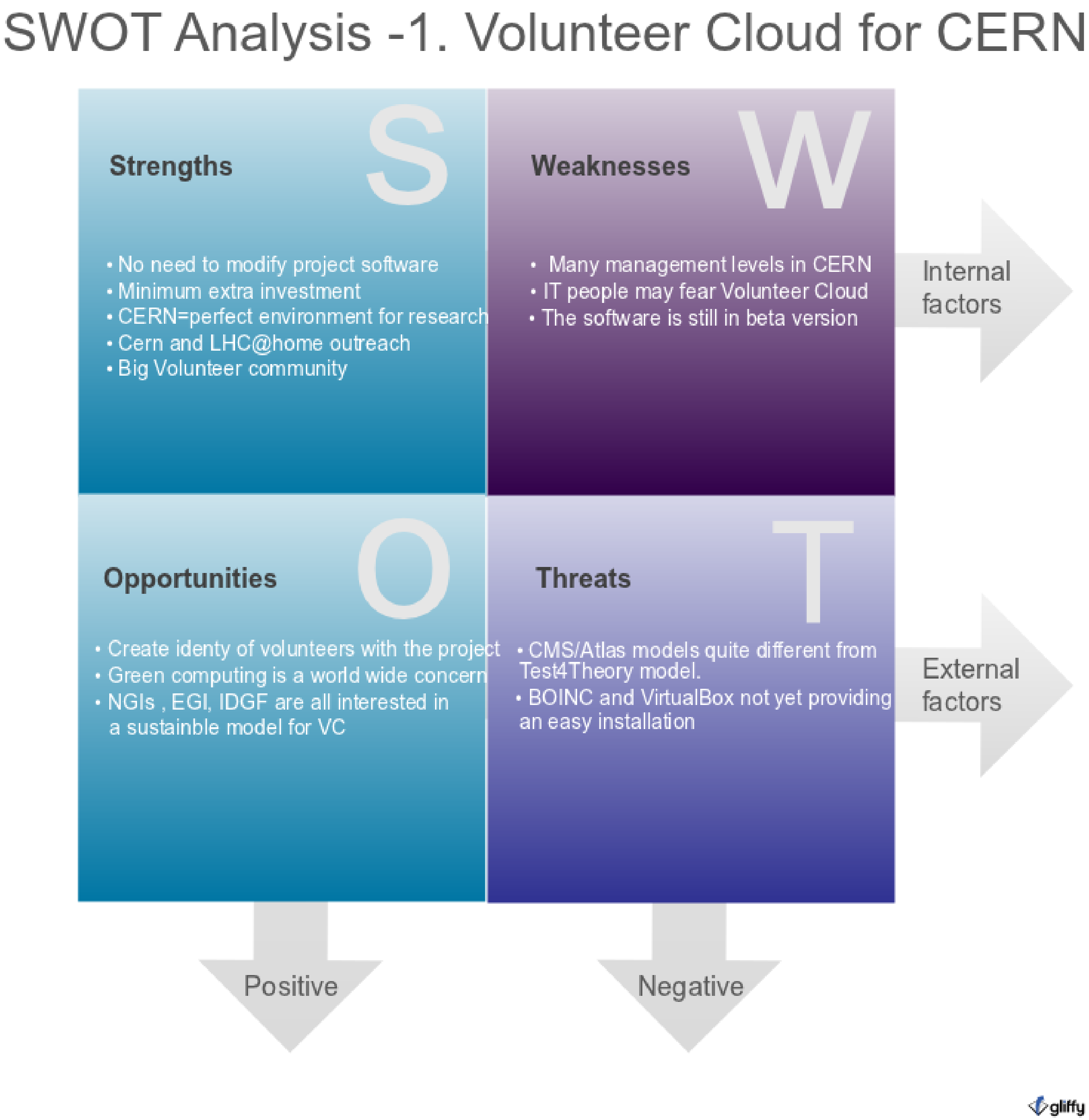



Figure 4.4: SWOT Analysis - 1. Volunteer Cloud for Cern

## 4.2 Volunteer Cloud for Industry - Openlab Model

In this context, Volunteer Cloud would be a CERN project aimed to gather collaboration with the ICT that are interested to use the research and development process from CERN to experiment the Volunteer Cloud environment. This model is already implemented successfully in Cern under the Openlab program and this model can become a new project of this program.

The main benefits CERN's partners would have for getting involved in the Volunteer Cloud project at CERN are:

- Use CERN's research environment to validate their project by doing a lot of stress testing for their cutting edge technology on the Volunteer Cloud infrastructure.
- Getting CERN's scientific outreach and a lot of attention from a specific target: group people interested in technology - physicists and computer scientists.
- The testing will validate the product also for Amazon Commercial Clouds as Volunteer Cloud is compliant with this infrastructure.

The main benefits CERN partners would have for implementing the Volunteer Cloud project following this model would be:

- Having experienced IT engineers developing and upgrading a product that started as a CERN initiative.
- Acquire personnel financed by the partners to get the software to a new level of robustness.
- Get more outreach from the projects in which they will be co-partners.

In Figure 4.5 we formalized this model using The Business Model Canvas and in Figure 4.6 we analyzed this model using the SWOT analysis. (more details about The Business Model Canvas and SWOT analysis in the beginning of Chapter 4).

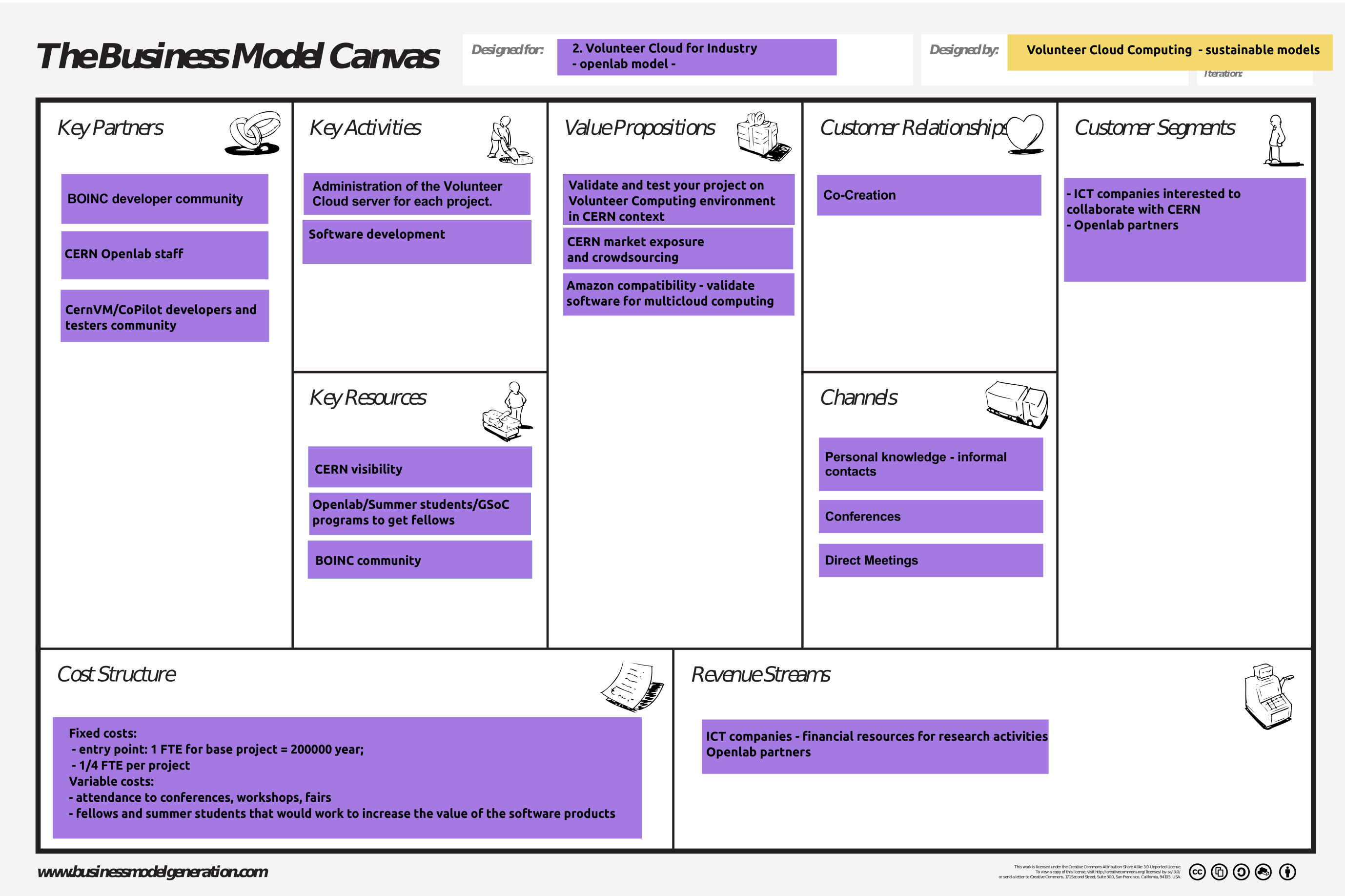



Figure 4.5: The Business Model Canvas - 2. Volunteer Cloud for Industry (Openlab Model)

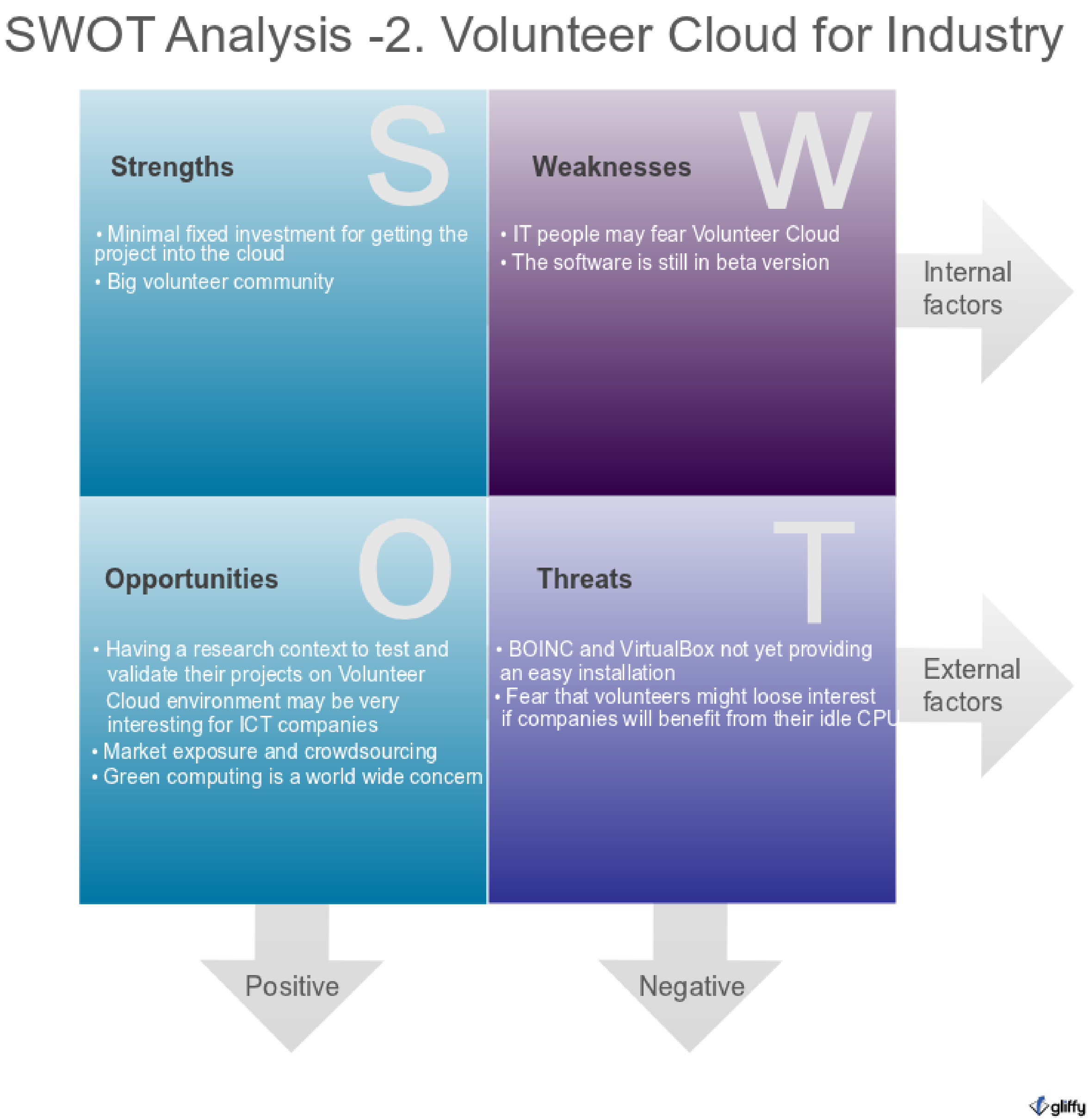



Figure 4.6: SWOT Analysis - 1. Volunteer Cloud for Industry (Openlab Model)

## 4.3 Nonprofit Volunteer Cloud

This model implies having a nonprofit foundation maintaining the Volunteer Cloud servers and providing direct services for the scientific projects that need distributed computation power. The main goal of this project is to lower the entry barriers of scientists into the Volunteer Cloud.

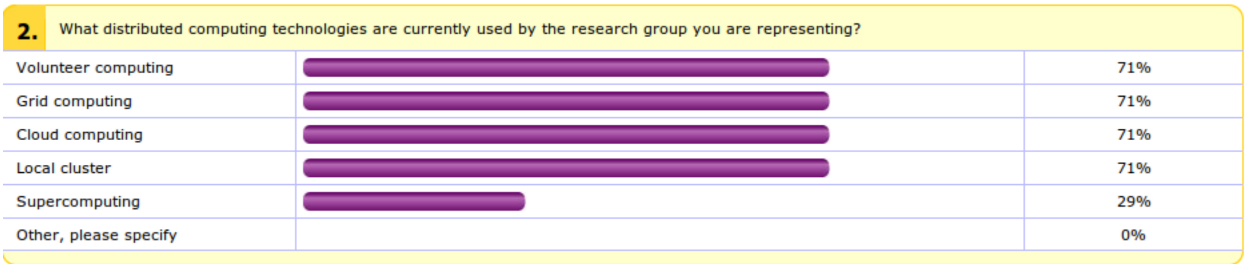
In this case the main customers are scientific projects and the product facilitated by the foundation offers direct access to the Volunteer Cloud for lower costs then acquiring computation from Amazon Clouds, or setting up a special cluster. This model can also accept ITC companies that are willing to contribute to the Volunteer Cloud's maintenance.

As the institution implementing this model is a nonprofit foundation, all the received funding will be used in the following ways:

- increasing the value of the Volunteer Cloud software;
- improving the services offered to the scientists and ICT partners;
- advertising more the Volunteer Cloud such that more scientists get familiar with this service;

In order to better understand the needs and expertise of the research institutes we have done a series of interviews and survey. (More details about the questions we asked and the persons that answered our question up to now are given in Section 7.2) Some interesting points we found out from this survey are:

- The research groups that answered our survey are using multiple distributed computing technologies and most of them are already using volunteer computing power as can be seen from Figure 4.7.



**2.** What distributed computing technologies are currently used by the research group you are representing?

| | |
|---|---|
| Volunteer computing | 71% |
| Grid computing | 71% |
| Cloud computing | 71% |
| Local cluster | 71% |
| Supercomputing | 29% |
| Other, please specify | 0% |

Figure 4.7: Survey - Q1 - Used technologies

- The amount of CPU hours that our respondents use varies very much (from 20K to 3M CPU hours per year). Considering our cost approximations these needs are high enough to justify using the Volunteer Cloud for their future needs.

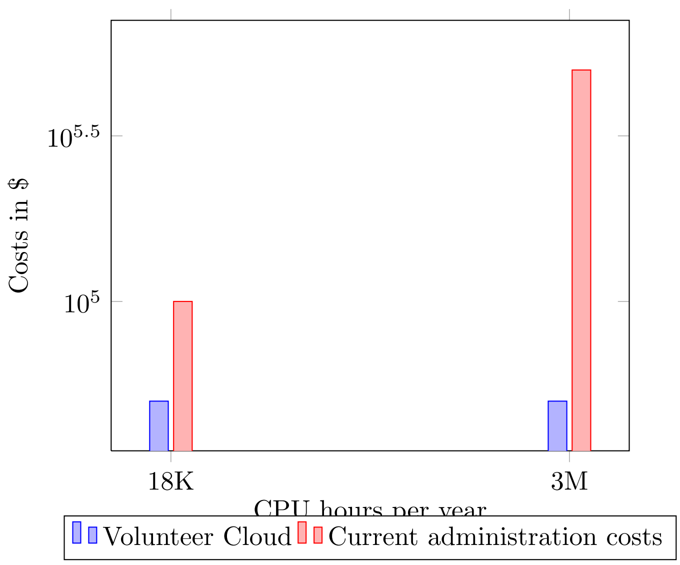



Figure 4.8: Research group CPU hours costs

- The amount of money invested yearly in personnel for administration and R&D is between 100K $ and 500K $, depending on the size of the project. We consider that for a project that needs between 18k and 3M CPU hours per year ( 1.5K CPU - 250K CPU hours per month) there are needed approximately 1/4 FTE manpower for operation purpose which means in rough terms approximately 50K $ per year. This need does not increase linearly with the running CPU hours. A good reference for this aspect is the World Community Grid example which computes yearly approximately 780M of CPU hours (at least 2M CPU hours per day [22]) and they have 6 FTE that manage this software project. This would mean 1/4 FTE for 3.2M of CPU hours. Therefore in Figure 4.8 we show that the operational costs for up to 3M CPU hours per year would be maintained at 50K$ with Volunteer Cloud while for the current infrastructure of the research groups they are paying up to 10 times more. We ignored the R&D expenses in both cases .

- We were also interested to see if research groups have projects that could benefit from Volunteer Cloud power but they haven't ported them for the BOINC infrastructure and we found out that some of them have this kind of projects so they might to interested to be a pilot project for the Nonprofit Volunteer Cloud model.

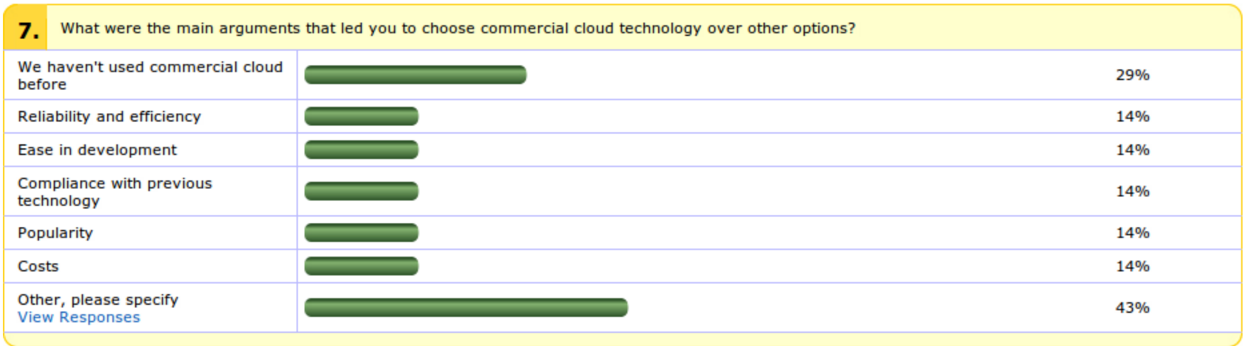



Figure 4.9: Survey - Q7 - Reasons for using Volunteer Cloud

- Research groups are not really reluctant to pay for some external computation power to commercial clouds providers like Amazon. However in the examples they provided the paid amounts are for very small projects or just for testing some prototypes. (under 5000 $). It also appears that the reasons that lead them to use commercial clouds are very diverse as can be seen from Figure 4.9

- The majority of the respondents wouldn't consider using the services of a commercial company for consultancy in Volunteer Cloud.

- From the replies we got up to now it seems that the main barriers to acquire volunteer computing power are

  - the complexity of the project make them difficult to parallelize under the limitations of Volunteer Cloud;
  - they lack the expertise and resources to integrate their project with BOINC libraries;

  There are opinions that consider virtualization as a very good step to make the entrance in the cloud easier. Also another used argument refers to education - having researcher informed of what exactly volunteer computing meas and what type of computation it can sustain successfully.

In Figure 4.10 we formalized this model using The Business Model Canvas and in Figure 4.11 we analyzed this model using the SWOT analysis. (more details about The Business Model Canvas and SWOT analysis in the beginning of Chapter 4).

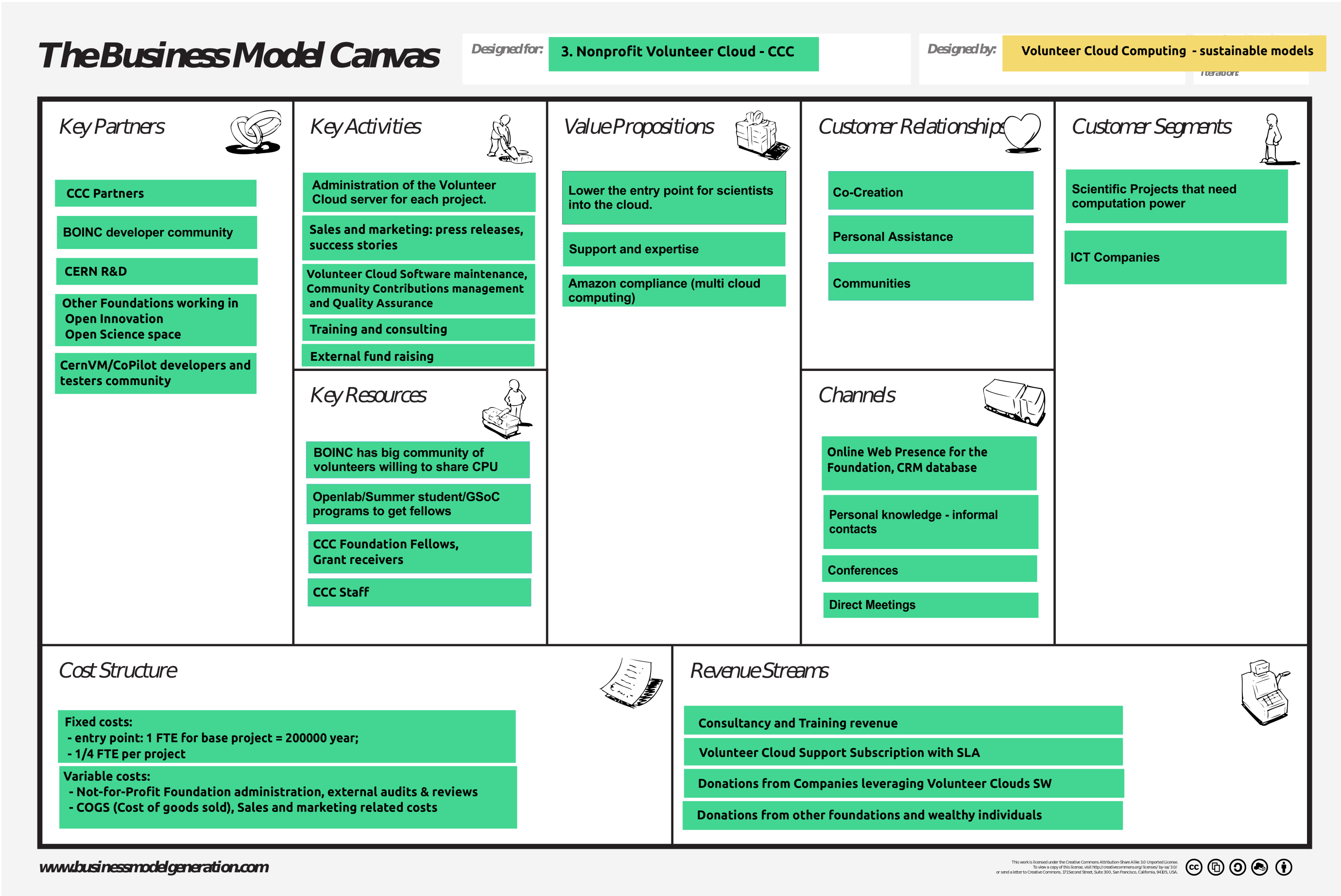



Figure 4.10: The Business Model Canvas - 3. Nonprofit Volunteer Cloud

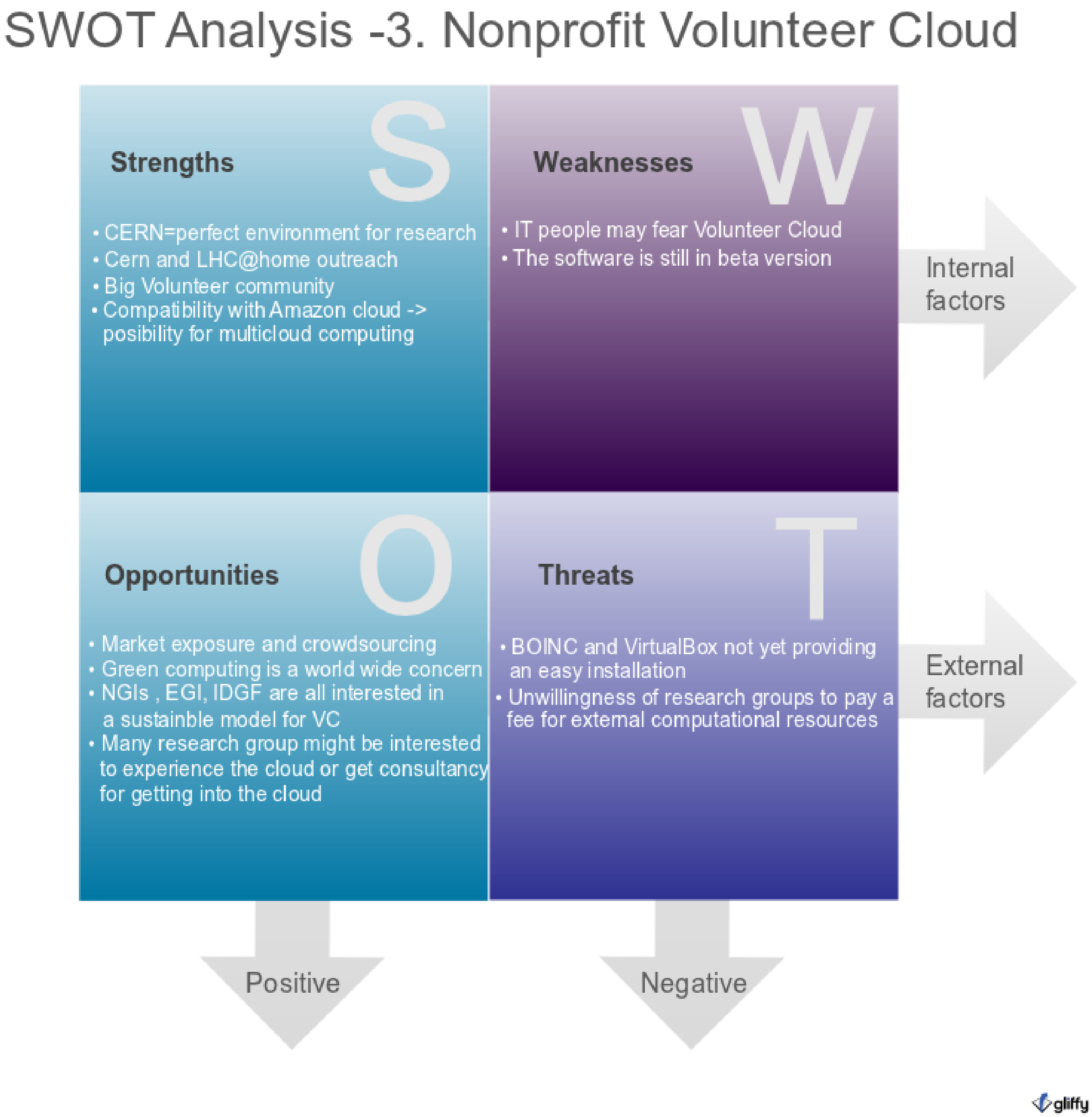



Figure 4.11: SWOT Analysis -3. Nonprofit Volunteer Cloud

## 4.4 Commercial Volunteer Cloud

This model describes how a start-up company could use the technology to provide Volunteer Cloud as a service. The main benefits customers will have are:

- Lower the entry point into the cloud
- Have support and expertise for fair price; pay for resources on-demand
- Amazon compatibility - opportunity for multicloud computing
- Benefit from the advantages of having an open-model,open data, open software;

**Volunteer Cloud Projects and Start-up**

We analyzed these cases because we wanted to be able to measure the efforts to have a project into the volunteer cloud. We also wanted to learn from the failures and successes lessons.

| | |
|---|---|
| world community grid. technology solving problems | - http://www.worldcommunitygrid.org/<br>- Contacted persons: Bill Bovermann (CEO WCG), Kevin Reed (CTO WCG)<br>- 6 FTE, 6 succesful projects, 567K users, 506K CPU years, 18 projects - 8 active, 6 completed, 4 Intermittent<br>- Focused on solving problems to benefit humanity for projects that are accelerated by grid computing - aprox 3 month to have the project on the volunteer computing infrastructure<br>- Success stories: 6 completed project, many research papers; IBM policy to have employees sharing idle CPU resources;<br>- Financing: Completely philanthropic project supported from IBM donations. |
| gridrepublic volunteer computing | - http://www.gridrepublic.org/<br>- Contacted persons: Matt Blumberg - project initiator;<br>- 8 Board Advisors, 280K users, 300K active PCs - 55 projects linked to GR<br>- Account manager for BOINC projects; Main purpose: market the projects maintain the communities<br>- Financing: donations and fellowships |
| charityengine | - http://charityengine.com/<br>- Contacted persons: Matt Blumberg - leader of the technical team;<br>- 4 co-founder, Matt Blumberg leader of the technical team<br>- Not launched yet. Currently has 2 customers(not official): 'one is a scientist, but the other is an online business'<br>- Financing (not official): £100k - from The Charity Engine - crunch base platform |
| Cloud Spree | - www.cloudspree.com<br>- Value Proposition 'Turn your social network into a supercomputer¡; Beta Project - 2 Employees<br>- Didn't respond to our interview yet. |
| Plura PROCESSING | - www.pluraprocessing.com/<br>- Value Proposition: Volunteer Computing as a service; 6 Employees<br>- No news since 2009; didn't respond to our interview yet.<br>- Financing (not official): $750k from Plura Processing crunchbase platorm |
| SuperDonate | - www.superdonate.org/<br>- Value Proposition: Sustain humanitarian project through donating 80 % of the profit;<br>- 7 sustained charities;<br>- No news since 2009; didn't respond to our interview yet. |

Table 4.1: Volunteer Cloud Projects and Start-up

**Multi Cloud Projects**

We analyzed these cases because we wanted to be able to measure the marketing and development efforts to offer consultancy and get a project into the cloud. As porting the application on Amazon or on Volunteer Computing Cloud platform will not be very different these efforts can be comparable. Two companies with which we were able to communicate successfully are: The Inkspot and Cloud broker; Other companies we have found and studied their model but we didn't communicate with yet are: Techila, CycleComputing, Wuala.

| | |
|---|---|
| www.inkspot.co | - http://www.inkspot.co<br>- Contacted persons: Paul Watson, Director of the North East Regional e-Science Centre at Newcastle University, UK.<br>- 6 FTE, 3 case studies (projects); focused in general on small projects;<br>- Value Proposition: consultancy, development and advice in many areas of Cloud Computing<br>- Financing: fees from the customers. |
| CB | - http://www.cloudbroker.com/<br>- Contacted persons: Peter Kunszt, Managing Partner, Academic Projects and Collaborations.<br>- Management team - 3 members; Main partner:IBM<br>- Value Proposition: multi-cloud web platform; consultancy and training; |

Table 4.2: Volunteer Cloud Projects and Start-up

In Figure 4.12 we formalized this model using The Business Model Canvas and in Figure 4.13 we analyzed this model using the SWOT analysis. (more details about The Business Model Canvas and SWOT analysis in the beginning of Chapter 4).

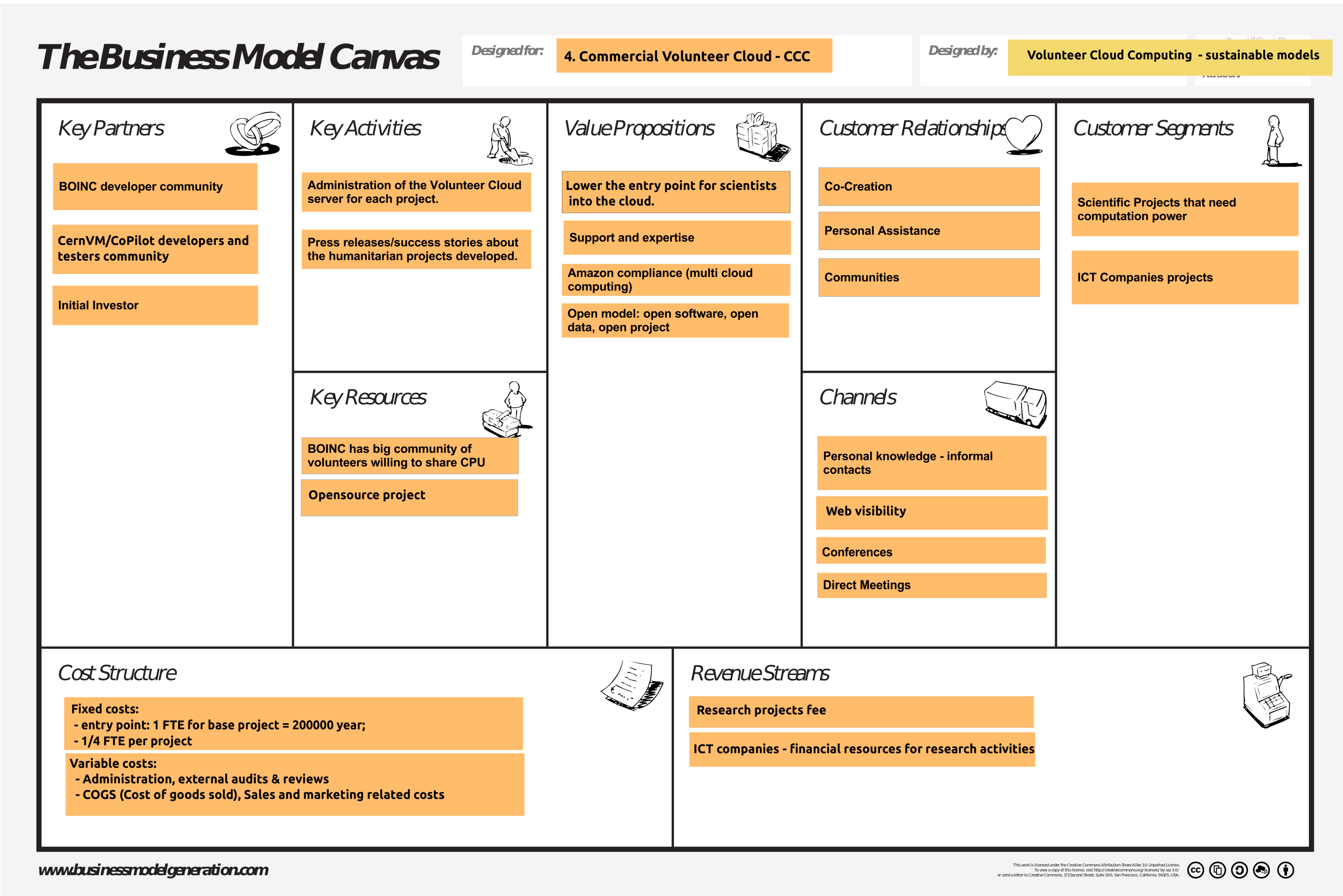



Figure 4.12: The Business Model Canvas - 4. Commercial Volunteer Cloud

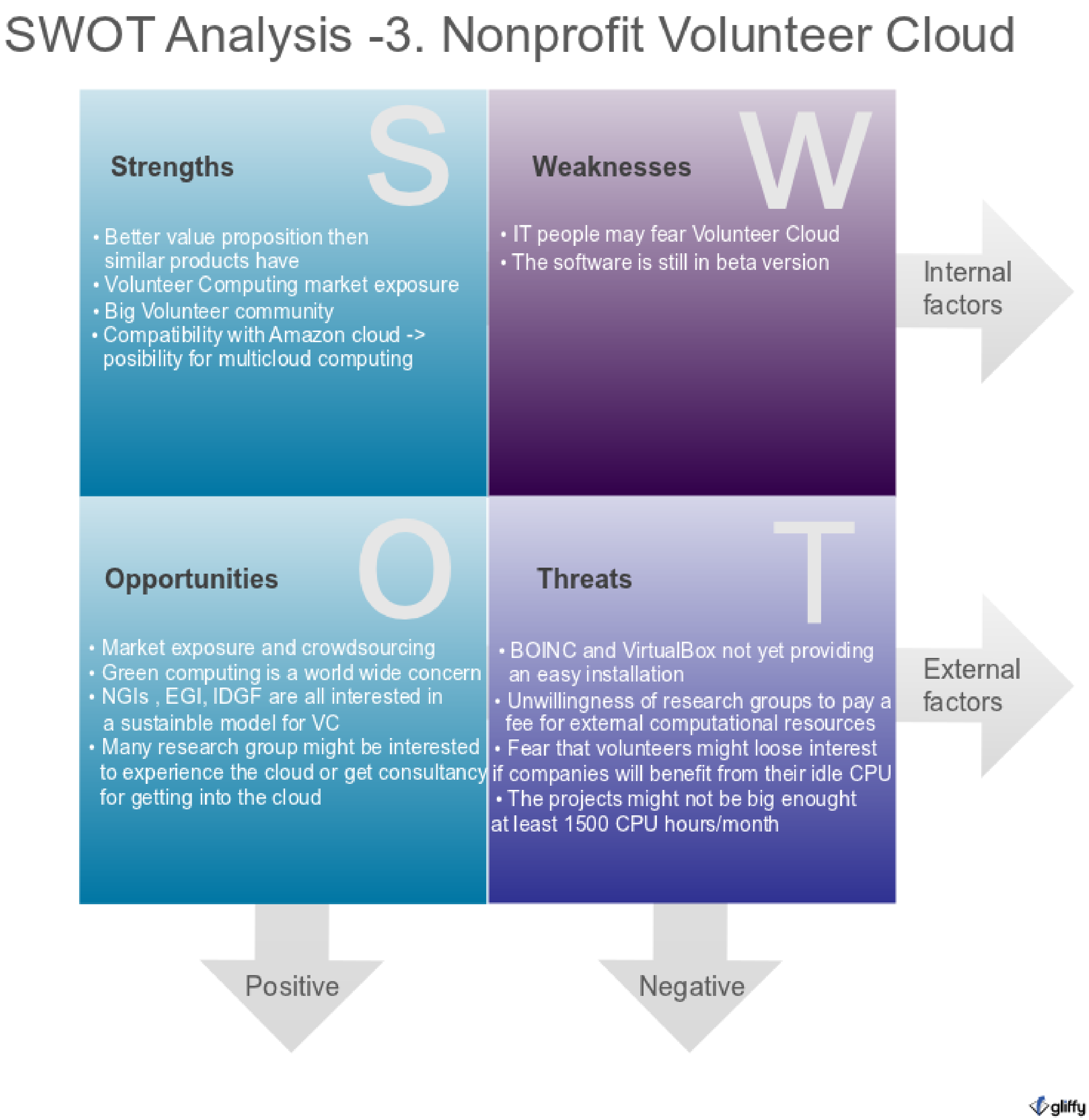



Figure 4.13: SWOT Analysis - 4. Commercial Volunteer Cloud

# 5 Conclusions and future steps

This research has made the first steps for implementing a sustainable Volunteer Computing Cloud project. The main achievements were to describe and formalize possible sustainable models for Volunteer Clouds under four Use Cases. Two Use Cases detail how the Volunteer Computing Cloud could develop in initiatives independent from Cern, implemented either by a nonprofit foundation or by a startup company. But as this is a Cern initiative we were also interested to study how Cern could benefit more from such a project, so we looked at two possible institutional ways to continue the development of Volunteer Clouds inside Cern.

The main aspects we have analyzed for each Use Case are: the value proposition, the client perspective, the activity perspective and the cost structure. For the client perspective we have emphasized who are the customers, the customer relationships and the possible communication channels with the client. For the activity perspective we pointed out the key activities needed to achieve the value proposition, the key resources we already have and the key partners of the project. For the financial structure costs and revenue sources we have compared the costs of using the Volunteer Computing Cloud with the costs of getting similar computation power from the Amazon EC2 Cloud or private clouds. After formalizing each model we have analyzed them by underlining the strengths, weaknesses, opportunities and threats for each project.

In order to gather information to support and validate our model we have consulted representatives from scientific research groups, start-ups that use Volunteer Computing or provide access to multi-cloud platforms, BOINC developers, and Cern management.

Our most important suggestions are based on enhancing the opportunities offered by Volunteer Cloud computing and seeing how to prevent and diminish possible threats.

The main opportunities are created by the huge number of volunteers that are currently contributing to BOINC, the value added by implementing virtualization within BOINC, and the need of research institutes to start using the Volunteer Computing Cloud. Another very important opportunity is the public outreach that a Volunteer project can obtain with minimal investments.

The most important threats relate to the fear that volunteers will not contribute to a commercial project, the fear to release private data on to volunteers' machines, and the fear of not having enough costing data to be able to estimate the cost of porting the project into the Cloud.

For the future, we recognize the need for improved cost approximations and comparisons with different ways of acquiring computation power. The future work that should follow this report is fully to formalize proposals into a Business Plan with detailed cost/revenue structure, scheduled milestones and activities, marketing strategy, success evaluation procedures and exit strategies. This report is meant to be the starting point to show possible directions Volunteer Computing Cloud has to take in order to become autonomous.

We also propose some possible strategies to validate and test these models:

- Pilot project - start with the foundation model and get a new research project into the Volunteer Cloud - to prove the costs and benefits. After this model is validated the commercial model can be accepted more easily and this can then work concurrently with the foundation. The foundation can take care of the software development part and the company of the services part.
- On the Cern side a pilot project could have a new LHC partner experimenting with the Volunteer Cloud as an Openlab summer project. This can show with minimal investment the potential of the project in terms of visibility, etc.

# 6 Acknowledgements

I would like to thank my supervisors Ben Segal, Miika Tusiku and Francois Grey for their advices and invaluable input to my work during my stay at CERN.

I also very appreciate the support from all the people that agreed to answer my interviews and questions: David Anderson, Matthew Blumberg, Bruce Allen, Juha Herrala, Kevin Reed, Bill Bovermann, Marko Niinimaki, Peter Kunszt, Rainer Wehkamp, Sverre Jarp, Peter Skands, Markus Schulz, Andrzej Nowak, Kacsuk Peter, Ad Emmen, Leslie Versweyveld, Robert Lovas, Adam Bazinet, M. F. Somers, Francisco Brasileiro, Peter Kacsuk, Nicolas Maire and Paul Watson.

I great support for developing this plan were the brainstorming sessions and direct feedback I received from: Matti Heikkurinen, Christian Pellegrini, Miguel Angel Marquina, Pedrag Buncic, Artem Harutyunyan, Ben Page, August Roos, Nicolas Maisonneuve, Anton Karneyeu,

I also want to express my gratitudes to my office colleagues that offered their journalistic expertise and very good collegial support Jacqueline Cherie Hayes and Adrian Giordani.

I would also like to thank Melissa Gaillard, Amalia Boari and Morag Hickman for devoted help with the administrative matters.

At last but not less important, I would like to leave my gratitude to all the other Openlab Summer Students 2011. They are Ioan Bucur, Ruggero Caravita, Goran Cetusic, Jiří Kunčar, Joao Faria, Urs Fassler, Georgi Hah, Martin Hellmich, Hallgeir Lien, Jurand Nogiec, Wojciech Ozga, Arsalaan Shaikh, Sonia Stan and Grace Young.

# 7 Annexes

## 7.1 Annex A - CernVM and CoPilot

CernVM[19] is a Virtual Software Appliance whose main purpose is to provide a simple interface for generating images on any end-user computer (laptop, desktop) as well as on the Grid and on Cloud resources, independently of host operating systems. The images are quite small ( 250 MB) and are based on a minimal Linux OS based on standard SLC5 (Scientific Linux CERN 5). The experiment software is delivered using the CernVM File System (CernVM-FS) that decouples the operating system from the experiment software life cycle.
CernVM Co-Pilot is a framework initially developed for the execution of the LHC experiments' pilot jobs on a wide variety of cloud computing infrastructure like comercial clouds (Amazon EC2) or scientific clouds (Nimbus) and the vounteer cloud (BOINC). The Co-Pilot plays the role of a gateway between different pilot job implementations and cloud resources. For each project a Co-Pilot adapter which links the experiment to the Co-Pilot gateway, and this adapter is the only addition such the the scientific project to be linked to the Volunteer Computing resources or other cloud infrastructure. All the necessary code to communicate with the cloud is included by the Co-Pilot and CernVM. The Co-Pilot framework consists of Agents (a minimal program hosted inside the virtual machines) and Services (the component which orchestrates the work of Co-Pilot Agents). These components communicate using Jabber/XMPP means through a special protocol. [20]
Even if for now this solution is also quite restrictive we consider this is an important step that can enhance faster development and adoption of Volunteer Computing. For example this infrastructure can only see one CPU, even if the volunteer's PC has more cores and also BOINC supports seing multiple cores.

## 7.2 Annex B - Detailed Surveys

Depending on the role and expertise domain of the persons we have interviewed we have addressed different questions. In this report we show some samples of these questions. The answers we received are used as a background information for this report. A detailed document with all the answers we received can be provided if you contact us. Samples from the questions we've asked to each of the main stakeholders in this domain clasified considering the main expertise domain of the person:

a. Current BOINC supporters (David Anderson - BOINC, Matt Blumberg-GRID Republic, Bill Bovermann, Kevin Reed - World Community Grid)

- What is the team size and composition of the development team?
- What complexity is behind deploying a new project in BOINC? Would virtualization speed up the process in most of the cases?
- What are the marketing strategies and costs allocated per project/per a certain number of volunteers?
- Are there any statistics, metrics or objectives to measure how much power is used for executing a job using Volunteer Computing power?
- What do you consider could be the possible directions to make this project into a sustainable model?

b. Commercial start-ups representatives (Techila, E-science Central, CycleComputing, Wuala, CloudBroker, SuperDonate, XtremeWebCh, Pluraprocessing.com, )

- What is the business model of your company?
- The software is open-source so part of the revenue will consist on offering support to customers and billing for the actual use. On the website there is not so much information about the billing system. Could you offer me some more information about this?
- What are the advantages of your product over the other commercial clouds: Amazon, Microsoft Azure?
- What is the financial plan for sustaining your business. Did you also consider gathering financing from donations for research projects?

c. Universities and research group representatives

- What distributed computing technologies are currently used by the research group you are representing?

  - Volunteer computing
  - Grid computing
  - Cloud computing
  - Local cluster
  - Supercomputing
  - Other (please specify)

- For each of the above, please specify type of project, name of infrastructure(s) you use, and approximate amount of computation carried out on the infrastructure(s) in the last year. (e.g. Cloud computing: protein folding, Amazon EC2, 10,000 hours).
- For each of the technologies your group uses, please specify roughly how much funding per year is required. If possible, please provide a rough breakdown according to manpower for development, manpower for operations, hardware and other running costs.
- Do you currently have scientific computing projects that could benefit from volunteer computing, but where you lack the resources to make the shift from another platform (eg local cluster)? If so, please specify what sorts of project, rough estimates of how much processing power they would require, and what sort of investment (time, money, manpower) is needed to develop them to volunteer computing projects.
- Have you ever used paid computation power from commercial clouds like Amazon? If so, roughly how much did you invest, how much computation did you do?
- What were the main arguments that led you to choose commercial cloud technology over other options? (cite top 3 arguments or specify if you haven't used this before)
  - We haven't used commercial cloud before
  - Reliability and efficiency
  - Ease in development
  - Compliance with previous technology
  - Popularity
  - Costs
  - Other (please explain)
- Would you consider using a commercial company to set up a volunteer computing project for your group, and if so, which of the following factors would be important in your decision? (cite top 3 arguments or specify if wouldn't consider using this)
  - No we would never consider using this
  - Cost relative to in-house development
  - Time to completion
  - Security of the solution
  - Reliability of the service
  - Quality of the resulting software
  - Other (please explain)
- Besides financial barriers to entry, what do you think are the most important challenges in the technical infrastructure or public messaging about volunteer computing that are preventing more scientific projects from using this type of resource? For example, do you consider that including virtualization support, and/or a Cloud interface, with BOINC would ease the process of porting your project to a volunteer computing infrastructure?

The persons who filled our survey until now are:

- Adam Bazinet,University of Maryland, Faculty Research Assistant
- M. F. Somers,Theoretical Chemistry; Leiden university, Lecturer & scientific programmer
- Francisco Brasileiro, Federal University of Campina Grande, Brazil, Professor
- Peter Kacsuk, Laboratory of Parallel and Distributed Systems, SZTAKI,Head of Laboratory
- Nicolas Maire, Swiss TPH, Project Leader Scientific Computing
- Paul Watson, Newcastle University, Professor of Computer Science
- Peter Kunszt, ETH Zurich, Project Lead SystemsX/SyBIT

d. CERN IT representatives, CERN OpenLab representatives ( Sverre Jarp, Markus Schulz, Ben Segal, Miika Tuisku)

- One of the considered barriers in Volunteer Computing is that ICT Companies are reluctant to use Volunteer Computing computation power as they don't want to spread their data in the user machines. However data sensitivity issue can be overcome in circumstances like the following.
  - Companies could test their application in the Volunteer Computing environment with some not sensitive data before they would go in the commercial cloud. (BOINC + CernVM +CoPilot is compliant with the Amazon Cloud Infrastructure)
  - Companies could test their applications in the Volunteer Computing environment before they consider to start giving some real data to compute through Volunteer Computing.
- How do you consider that these scenarios could fit with the Openlab current activities?
- How interested would Openlab's partners be to alpha/beta test some of their projects on the Volunteer Cloud environment implemented at CERN?
- Currently there are around 700K computers doing volunteer computing so it seems that the issue today is not to acquire new users sharing resources but to assign more efficiently this power to scientific projects. Currently CERN doesn't make a lot of use of this computation power for its own purposes. What do you think should be changed in the infrastructure, marketing idea or message of the Volunteer Computing such that more scientific projects would start using this resource? (inclusively here at CERN)
- What do you consider would be the possibilities of developing this project inside CERN (for the scenarios described above) in an independent project following the Openlab model?

## 7.3 Annex C - Detailed description of The Business Model Canvas

In this paragraph we will describe each of the bullets we included in The Business Model Canvas from each model we proposed. This formalization has been made after the 9 Building Blocks Business Model [1]. The business models can still be updated and their progress can be followed at this links [3].

We will start with details about the first model and then we will specify only the new items for next models.

**Model 1 - Volunteer Cloud for Cern**

- Offer Perspective
  - Value Proposition
    a. No need to modify the existing software and production system - The main software benefit of the BOIN+CernVM+Co-Pilot is the portability aspect - it eases the integration of cloud resources with the existing projects. The Co-Pilot agent can be customized to integrate a project with Volunteer Cloud project. Virtualization support is also a big asset that overcomes many problems in terms of: scalability, security and cross-platform.
    b. Market exposure and crow-sourcing
    c. Minimum extra investment
    d. Amazon Compatibility
- Client Perspective:
  - Customer Relationships
    a. Dedicated personal assistance - The main partners of OpenLab project are Intel, HP, Siemens, Oracle are currently using the CERN environment for testing their pilot projects on the CERN Infrastructure. We are interested in developing long time relationships with our partners for more than one project. The main assistance personnel will be technical fellows and students acquired through internships.
    b. Co-creation - Reporting and disseminating results is a way to explore and consolidate the customer-vendor relationship beyond the utility of the sold product. CERN has a great experience with communicating results through press-release, conferences in which each partner company its slot to present obtained result.
    c. Community - In order to grow and maintain the volunteer group motivated we will use the communities customer relationship model. It is a fact that actively maintaining a community of volunteers through a forum/discussion list is very important for retaining and adding new members to the contributors group therefore currently there already exists an active forum where people interested in contributing can ask/reply questions. [12]
  - Customer Segments

a. Niche market computation power for simulation purposes from CMS, Atlas, Physics Projects;

– Communication Channels

a. Direct communicating with projects' architects and IT Department

- Activity perspective

  – Key Activities

    a. Administration of the Volunteer Cloud server for each project

    b. Consulting

  – Key Resources

    a. CERN visibility

    b. BOINC community- From BOINC statistics the number of volunteers that participated until now in the projects sustained with the BOINC technology is 6.6 millions from 272 countries. The projects initiated by CERN even if they were running for a limited period of time they gathered around 15 percent of the total number of volunteers. The project LHC@Home version 1 is ranked as the 6th most popular project from all the BOINC projects. [7] The most popular Volunteer Computing projects are running on BOINC (WCG - IBM, SETI@home) and statistics show that the number of users increase every day with 300 to 500 users. [8] The BOINC community is the strongest community of users which are sharing their idle CPU resources for research purposes. This is a very strong advantage as the BOINC users are very fast when disseminating the word about a new research project they start sharing resources also for that project. Besides BOINC has also implemented an account manager which can share that idle CPU to different projects attached to that account without the user needing to subscribe every time to a certain project. Example- WCG User Account.
    Furthermore the volunteer community still has a high growing potential. The main motivations for new users to join are:

      * Green Computing Reasons - make the consumption of the power useful. This concept is very fashionable nowadays and there are lot of discussion upon different ways to acquire this existing computation powers. (There are also some start-ups out there that have tried to benefit from this).
      * Sympathy and attachment for the specific project or to the goal of the project. Example: WCG project, SETI@home
      * Users can be also interested to share resources for a commercial product. The reward for this sharing can be a virtual currency or the possibility to play some game or to participate to a jackpot type game.

    c. CERN Staff and contacts- Cern environment has a lot of resources in terms of hardware infrastructure, knowledge and good management structure.

    d. CERN emplyees can become a good database of volunteers.

    e. Green Computing Interest- Green computing and volunteer computing are 2 very hot topic in the computer science world nowadays and a lot of grid computing organizations and conferences are redirecting their interest toward cloud computing and as a special topic to explore the computation that is already available through Volunteer Computing.

    f. European Cloud Initiative In the context in which CERN and other European scientific organizations are interested in building a strategic plan for a Scientific Cloud Computing infrastructure for Europe[10] this kind of step could be one of the steps. Volunteer Computing is a concept that has the potential to gathers a lot interest because of the very low long term investment and because it is based on the idea of sharing knowledge.

  – Partners

    a. BOINC developer community- BOINC development community is well grounded community and has a well specified development process. [11]

    b. CERN IT Department- The IT Management is strong key resource for implementing the Volunteer Cloud administration in CERN. The team experience, expertise and autom department is organized to successfully support each project inside CERN.

    c. Research Foundation- Foundations interested in offering resources for research purposes which need big computing resources. (EGI, ShuttleWorth Foundation).

    d. ICT Companies- ICT Companies nowadays are very interested in using at least partially resources through the cloud. However as this is still a growing concept and the services they offer are not fully ready to be ported into a cloud environment ICT companies can be interested in using the Volunteer

Cloud environment from CERN to port and test their application before deploying it to the commercial cloud. It is an important asset that BOINC is complemented with virtualization support as this allows testing the application on the Volunteer Cloud and then (optionally) porting it to Amazon Cloud. Testing various methods and comparing benefits of each type of Cloud is a topic of high interest for the ICT companies and research institutes and current tests have showed that 'If cloud computing systems are to replace VC platforms, pay per-use costs would have to decrease by at least an order of magnitude'. [9]

# 8 References


[1 ] Business model analyzer and designer tool developed by http://bmdesigner.com/

[2 ] SWOT analysis model

[3 ] Link to the BMDesigner scheme project Account: username: cernaccount; passwd: cernaccount

[4 ] Student work placement

[5 ] The CERN Fellowship Program

[6 ] The most recent report from the OpenLab project

[7 ] BOINC project popularity statistics

[8 ] BOINC combined popularity

[9 ] Derrick Kondo1 , Bahman Javadi1 , Paul Malecot1 , Franck Cappello , David P. Anderson - Cost-Benefit Analysis of Cloud Computing versus Desktop Grids

[10 ] Strategic Plan for a Scientific Cloud Computing infrastructure for Europe

[11 ] BOINC developer community stats

[12 ] Test4Theory Message boards

[13 ] BOINC official forum - LHC@home thread BOINC official forum - LHC@home thread

[14 ] BOINC official statistics for Test@Theory Project

[15 ] http://www.datacenterknowledge.com/archives/2009/05/14/whos-got-the-most-web-servers/

[16 ] Ben Segal - Early Experience with CERN's Volunteer Cloud, BOINC Workshop 2011

[17 ] BOINC project personnel

[18 ] Attaching BOINC to Test4Theory@Home

[19 ] P Buncic, C Aguado Sanchez, J Blomer, L Franco, A Harutyunian, P Mato, Y Yao - CernVM – a virtual software appliance for LHC applications

[20 ] Pedrag Buncic, Artem Harutyunyan - Co-Pilot: The Distributed Job Execution Framework

[21 ] Exa-scale computing, software and simulation

[22 ] World Community Grid - Global Statistics

[23 ] CMS dashboart - Historical views

[24 ] Christoph Grab - Report from the CHIPP Computing Board