Costs of Cloud Computing for a Biometry Department*

A Case Study

J. Knaus1; S. Hieke1,2; H. Binder1,2,3; G. Schwarzer1

1Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Freiburg, Germany;
2Freiburg Center for Data Analysis and Modeling, University of Freiburg, Freiburg, Germany;
3Institute of Medical Biometry, Epidemiology and Informatics, University Medical Center, Johannes Gutenberg University Mainz, Mainz, Germany

Keywords
Costs and cost analysis, biostatistics, mathematical computing, cloud computing, Amazon Web Services

Summary
Background: “Cloud” computing providers, such as the Amazon Web Services (AWS), offer stable and scalable computational resources based on hardware virtualization, with short, usually hourly, billing periods. The idea of pay-as-you-use seems appealing for biometry research units which have only limited access to university or corporate data center resources or grids.

Objectives: This case study compares the costs of an existing heterogeneous on-site hardware pool in a Medical Biometry and Statistics department to a comparable AWS offer.

Methods: The “total cost of ownership”, including all direct costs, is determined for the on-site hardware, and hourly prices are derived, based on actual system utilization during the year 2011. Indirect costs, which are difficult to quantify are not included in this comparison, but nevertheless some rough guidance from our experience is given. To indicate the scale of costs for a methodological research project, a simulation study of a permutation-based statistical approach is performed using AWS and on-site hardware.

Results: In the presented case, with a system utilization of 25–30 percent and 3–5-year amortization, on-site hardware can result in smaller costs, compared to hourly rental in the cloud dependent on the instance chosen. Renting cloud instances with sufficient main memory is a deciding factor in this comparison.

Conclusions: Costs for on-site hardware may vary, depending on the specific infrastructure at a research unit, but have only moderate impact on the overall comparison and subsequent decision for obtaining affordable scientific computing resources. Overall utilization has a much stronger impact as it determines the actual computing hours needed per year. Taking this into account, cloud computing might still be a viable option for projects with limited maturity, or as a supplement for short peaks in demand.

1. Introduction

During the last three decades the invention and widespread availability of personal computers and workstations has a major impact on the use and development of statistical methods. Contemporaneously the computational performance of a single computer and the computational complexity of statistical methods as well as the size of datasets increased over time. Accordingly, the provision of additional computing resources constitutes a recurring business decision at a Statistics department. Until recently this decision boiled down to the question which on-site hardware to purchase, e.g. either powerful personal computers or dedicated compute servers. With the advent of “cloud computing” during the last 5–10 years, a promising alternative approach for statistical computing emerged.

“Cloud” providers like the Amazon Web Services (AWS) offer stable and scalable computational power based on virtualization with short, usually hourly, billing periods.

This means a user can start some virtual machines through the service provider which basically behaves like physical computers. These machines run a user selected and configured operating system and customer user installed software and can be turned on and off by demand. The idea of pay-as-you-use is appealing to smaller institutes or working groups, which have no access to large university or corporate data center resources or grids.

Like many other biostatistical departments, the Department of Medical Biometry and Statistics of the Institute of Medical Biometry and Medical Informatics in Freiburg (IMBI) has a decent, but limited annual budget for IT services. In the year 2005, our department decided to buy dedicated hardware (rack servers) to supply on average ten researchers with computing power for different statistical
purposes. For routine desktop work as well as smaller statistical analyses other server hardware (both Linux and Windows terminal servers) are provided. The department's focus on providing additional computing resources changed during the last five years. With the a new focus on statistical analyses of genomic data, demand for memory grew faster than our demand for processing cores at our department. Although our department is able to buy new hardware for statistical computing almost annually, the question arose if it was more cost-effective to rent necessary compute resources from cloud providers instead of buying new dedicated server hardware.

There are many studies on the costs of cloud computing in scientific environments, but most describe data center installations or smaller examples (e.g. [1-5]). As these publications do not provide the necessary information for an informed business decision for our department, we decided to conduct a cost analysis for our setting.

This article compares the costs of an existing heterogeneous on-site hardware pool from 2003 to 2011 in our biostatistical department to an comparable offer from Amazon Web Services (AWS).

2. Methods

A cost comparison between on-site hardware and the rental of cloud instances over the whole usage time of a dedicated server (which is usually 3-5 years [6]) would be incorrect, as this deviates from the idea of using cloud services on-demand. Therefore only hourly costs for the actual utilization of the department's servers are calculated. These hourly costs are compared to costs for using the Amazon Web Services.

Calculation of hourly costs of owned hardware starts with the estimation of the "Total Cost of Ownership" (TCO) [7] with costs divided into direct and indirect costs. Direct costs include the server buying price, costs for attached infrastructure, housing, electricity, software licenses, and salaries. Indirect costs mainly occur as working time costs for teaching, data management, support etc. In the following analysis only direct costs will be considered, as indirect costs can highly depending on the individual situation. However, in our experience indirect costs resulting from using cloud services are higher than those from using on-site hardware. For example, our terminal and compute servers share the same home directories simplifying data management.

2.1 Cost for Computing Using Hardware Owned by our Department

Direct costs depend on the use period of a machine. While in industry use period is often determined by depreciation, which is usually three years, we noticed that most academic institutions use their machines longer. Depending on the hardware performance this can be beyond five years. CPU development during the last ten years mainly consisted of adding more calculation cores, but without coercible faster processing units. Accordingly, older servers often have fewer cores, but are otherwise comparable in raw core power to current hardware.

▶ Table 1 shows the total expenses for our computing infrastructure between Nov. 2006 and Dec. 2011 (R program for cost calculation is given in the ▶ Supplementary Online File). Nine dedicated compute servers (max. two per year) were bought during these five years. Their technical specifications range from a 4-core machine with 16 Gigabytes of RAM bought in 2006 to one with 24 cores with 128 Gigabytes bought in 2010 with prices ranging from 3,500 to 10,000 Euros. Over time, machines with more RAM were bought enabling the analysis of high-dimensional applications (four machines have 64 Gigabytes or more of RAM). Overall, the department has 84 (real, non hyper-threading) cores with 416 Gigabytes of RAM. Three machines with a total of 24 cores have only 2 Gigabytes per core, the other machines have 4-9 Gigabytes per core.

A complete renovation of the server room took place in the year 2008. Total costs for this modernization include costs for server racks, the air conditioning system, network switches, and wiring. The server room does not only host the nine compute servers, but another eleven machines for other tasks like file-, web-, mail-, and terminal services. The life time of this modernized server room is assumed to be at least 10 years resulting in annual costs of 200 Euros per server. We cannot quote the rental costs for the department's server room (9 m² cella room), as the building is owned by the university. This reduces our costs and is a common situation in academia. Accordingly, no rental room costs are considered.

Similar to costs for server room usage, personal costs for system administrators only partly apply to the administration of compute servers. Administration tasks for compute servers are minimal as i) all compute servers have the same operating system (Debian GNU/Linux), ii) home directories as well as the statistical software environment "R" [8] are provided on shared

<table>
<thead>
<tr>
<th>Table 1 Total costs of IMBI compute infrastructure since 2006</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fixed costs onetime per machine (average)</strong></td>
</tr>
<tr>
<td>Compute servers (9 servers)</td>
</tr>
<tr>
<td>Overall costs: 51,942 €</td>
</tr>
<tr>
<td>Server room (air conditioning, racks, door), overall</td>
</tr>
<tr>
<td>Overall costs: ~ 40,000 €</td>
</tr>
<tr>
<td>Administration: 10h Installation</td>
</tr>
<tr>
<td>Administration: 0.3 h/month maintenance</td>
</tr>
<tr>
<td>Energy (250 W average, 0.15 €/kWh)</td>
</tr>
<tr>
<td>Cooling energy (854 Btu/h, EER 3 cooling unit)</td>
</tr>
<tr>
<td>Sum</td>
</tr>
</tbody>
</table>
network file systems, iii) accounts are managed centrally via a LDAP directory. Costs for administration divide into nonrecurring initial installation costs (including rackmounting the new server) and regular maintenance tasks like updates of the statistical software environment R and the operating system.

Costs for electricity do not appear on the department’s bill. However, as this is an important cost item, we try to approximate these costs. We assume that each compute server consumes on average 250 Watt with a price of 15 cent per kWh. Some machines use up to 500 Watt if used to capacity, but also down to less than 100 Watt during idle times while others use below 350 Watt on maximum. The rough average of 250 Watt is approximated by the power consumption of one server with an integrated power indicator over a period of one year (more details on power consumption can be found in the ▶Supplementary Online File). According to specification information from the manufacturer, the air conditioning system has an Energy Efficiency Ratio (E.E.R.) of 3.01 [9], i.e. for every 3.01 watts consumed by a computer server, one watt is used by air conditioner to cool the server. Accordingly, we assume that the air conditioner consumes on average 250/3 = 83 Watt per server. We do not use any Uninterrupted Power Supply (USV) for our compute servers. Running the compute server infrastructure on Debian GNU/Linux and using R in computational demanding statistical applications, no software license costs are considered. Note, besides R, also the statistical packages SAS [10] and Stata [11] are utilised at our department for statistical analyses, however, additional compute resources have been bought during the last years to fit the increasing computational needs for R.

Based on information given in ▶Table 1, the theoretical annual budget spend for computing resources at our department – ignoring costs for system administration which also apply to cloud services – is 16,131 €, i.e. overall hardware costs per year + costs for server room utilization + energy costs for compute servers and cooling = 51,942 €/5 + 9 × 200 € + 9 × (328.5 € + 109.58 €). As costs for electricity do not appear on the department’s bill the annual costs reduce to about 12,000 €. This amount of money could have been spent annually for cloud services by our department during the last five years. As many biostatistical departments do not have a dedicated server room, annual costs would reduce this number to about 10,000 €.

### 2.2 Estimating System Utilization and Hourly Costs

System utilization is very important to assess the cost of owned hardware. Although nearly all computing cores are used during peak times, in general this is not the case. Often, single machines are only used partly or even idling. As not all statistical computations on our compute servers are managed by a batch system, we can only approximate the real usage for computing. Using the monitoring system Xymon [12] which is used routinely for all servers at our department we measured the average annual load in the year 2011 for all nine compute servers (see corresponding figure in the ▶Supplementary Online File). System utilization of IMBI compute servers in 2011 ranges from 10 to 37.5 percent with an average of 24.3 percent assuming that 10 percent of a single core is used by the operating system for administrative purposes. Overall, based on the complete amount of CPU cores we have a demand for around 182,000 computing hours per year.

Based on system utilization for load, we calculate an hourly price for machines owned by our department by assuming that a compute server is running under full load for a certain time of the year and switched off afterwards for the rest of the year. For example, a system utilization of 50% means that all cores of a compute server are running under full load for six months (4,383 hours) and switched off afterwards. This is not a real world scenario, but makes hourly prices comparable to the cloud offers.

Given an average price of 5,771 Euros for a compute server purchased during the last five years, ▶Table 2 shows the hourly costs for this hypothetical machine (with 9.3 cores and 45.5 Gigabytes RAM) by system load and operating time. An average load of 100% will rarely occur in practice, but is added as the possible upper limit of system utilization. Note, for this calculation the hourly prices are not dependent on the amount of cores and memory in a machine but only depend on costs and utilization (but of course, purchase price depends on technical specification). A machine with twice the number of cores also needs on average twice as many busy cores to achieve the same utilization level. But power consumption is higher on a full server. In the appendix two examples of hourly instance prices of single machines with their individual power consumption limits are given.

Another important aspect of system utilization not considered in calculations in ▶Table 2 is the amount of memory used in statistical computations. The monitoring system Xymon also provides a metrics report on memory utilization. For the year 2011, the average memory utilization of the nine compute servers is only 7.9% with a range from 2.6% to 18.8%. However, time periods of idle compute servers are included in these values resulting in too small values of memory usage during statistical

### Table 2

Example for hourly usage prices depending on usage for TCO (Server price: 5,771 Euros)

<table>
<thead>
<tr>
<th>System utilization</th>
<th>1 year operating time</th>
<th>3-year operating time</th>
<th>5-year operating time</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Hours absolute</td>
<td>Price per hour (Euro)</td>
<td>Hours absolute</td>
</tr>
<tr>
<td>10%</td>
<td>876.6</td>
<td>7.76</td>
<td>2629.8</td>
</tr>
<tr>
<td>20%</td>
<td>1753.2</td>
<td>3.88</td>
<td>5259.6</td>
</tr>
<tr>
<td>25%</td>
<td>2191.5</td>
<td>3.11</td>
<td>6574.5</td>
</tr>
<tr>
<td>30%</td>
<td>2629.8</td>
<td>2.59</td>
<td>7889.4</td>
</tr>
<tr>
<td>50%</td>
<td>4383</td>
<td>1.55</td>
<td>13149</td>
</tr>
<tr>
<td>100.00%</td>
<td>8766</td>
<td>0.776</td>
<td>26298</td>
</tr>
</tbody>
</table>

© Schattauer 2013
computations. Dividing the percentage of memory usage by the percentage of load for each compute server, respectively, we get a rough estimate on the memory utilization during statistical computations resulting in an average memory usage of 35.6% with a range from 10.4% to 83.6%.

### 2.3 Cost of Cloud Instances

In order to evaluate the "total cost of renting" using the AWS Elastic Compute Cloud (EC2 [13]) and Elastic Block Store (EBS [14]) not only costs for virtual machines (“instances”) have to be considered but also costs for data storage, reading and writing access to stored data, and outgoing network traffic. Instances are paid on an hourly basis ranging from 0.095 to 2.28 US$ per hour (all prices given for the Amazon EU West data center in Ireland, as of September 2011). Data storage is payed on a monthly basis (0.11–0.14 $ per GB/month) and all access procedures on a volume based amount (0.11 $ per 1 Million I/O operations). As an institution located in Europe and working with clinical data we are bound to the European data privacy guidelines [15] and therefore we chose the Amazon data center in Ireland. Prices vary between these data centers and are lower for the US data centers.

With our usage of EC2 instances as plain compute servers, instance costs outweigh other costs – they can mostly be disregarded (this effect also applies for very data intensive compute applications [16], further information also in the Supplementary Online File). The costs for EC2 instance hours can be reduced by “reservation” of computing time for one or three years. By paying a fixed amount upfront, the hourly price per instance drops massively. For example, a three year reservation of the EC2 instance type m1.4xlarge for 8,000 $ results in a drop of the hourly price from 2.28 $ to 0.96 $ and a reservation of type m1.large for 2,800 $ results in a drop from 0.76 $ to 0.32 $. Note, reservation costs have to be applied to each instance type separately.

Some direct cost factors naturally are not applicable for cloud computing, e.g. purchase price, infrastructure, and power consumption. However, costs for system administration still apply. We assume that the same administration time is necessary for EC2 instances as for physical servers.

### 3. Results

#### 3.1 General Comparison

Even in it’s European data center in Ireland Amazon is charging in US Dollars. Therefore all prizes are exchanged to Euro (exchange rate 1.35) and 19% tax is added (German VAT).

In order to compare costs for on-site machines and EC2 instances, two compute servers which are fairly similar to existing EC2 instances with respect to the number of cores and RAM were chosen: a machine bought in November 2006 (4 cores, Xeon, 3 GHz, 16 GB RAM, purchase price: 6,680 €) and a machine bought in December 2009 (8 cores, Opteron, 2.5 GHz, 72 GB, purchase price: 7,057 €) compared to EC2 instances m1.xlarge (4 cores with 2 EC2 Compute Units each, 15 GB, hourly price: 0.76 $/h) and m2.4xlarge (8 cores with 3.25 EC2 Compute Units each, 68.4 GB, hourly price: 2.28$/h). As the EC2 instances are virtual machines hosted on varying hardware, Amazon describes their speed with “EC2 Compute Units” which “provides the equivalent CPU capacity of a 1.0–1.2 GHz 2007 Opteron or 2007 Xeon processor” [13].

Figure 1 shows a price comparison between two IMBI compute servers and corresponding EC2 instances assuming an annual system utilization of 25 percent/2191.5 hours (imb7) and 30 percent/2,629.8 hours (imb1). Both physical machines are faster than their EC2 counterparts which necessitates the purchase of more computing hours for the same number of calculations (70% for left comparison, 40% for right comparison, benchmarks described in [17], also further details in the Supplementary Online File). Accordingly, we also report costs adjusted for the speed difference assuming that EC2 instances m1.xlarge and m2.4xlarge are utilized for 4,425 and 3,527 hours per year, respectively. Power consumption of the on-site servers is included in the plot with the approximated average on the given utilization level (further details can be found in the Supplementary Online File). As both on-site hardware and EC2 instances have the same amount of memory, full memory usage over the utilized time is assumed for this comparison.

One year reservation is even more expensive than three year reservation. Reservation is getting cheaper than on-demand usage if more than 4,000 hours per year (one year reservation) or 6,000 hours in three years (three year reservation) are rented. An alternative to reservation for even further price reduction can be gained by using “Spot” instances, which basically are auctioned server hours. This can reduce overall price [18], but requires specifically modified programs and leads to uncertain runtimes, as instances can be automatically started and stopped depending on auction biddings. EC2 prices dropped between 15 and 19% over the last five years, but the comparison is made with current cheaper prices. That means in a real three year rental period, renting instances would have been higher than shown [19, 20].

At first glance, both plots in Figure 1 look rather different. In the left plot the EC2 instances seem to be more cost-effective up to four years than the physical server with 4 cores and 16 Gigabytes RAM. However, comparing a hardware price from 2006 with current EC2 prices is rather inadequate. Considering a current street price for a server with 4 cores and 16 Gigabytes RAM of 2,000 Euro (grey line with open circles), both plots provide very similar results.

Using on-site hardware is more cost-effective after one to two years than renting EC2 instances in this setting – even without adjustment for the speed difference. A price drop for EC2 instances over time is not considered in Figure 1. However, a substantial price drop within the first year would be necessary in order to make EC2 instances more cost-effective in this setting.

The given comparison assumes full usage of the memory, which is true for individual programs, but often machines under full load only use smaller shares of the overall memory. For the 4-core machine the average memory use under load is only 10.4% (1.7 GB), for the 8-core machine it is 43% (31.0 GB). Therefore, for many applications, they could be replaced.
by instances with fewer memory reducing costs. As an example, in Figure 2 we compare the same on-site machines with instances c1.medium (2 cores, 1.7 GB RAM, 92% slower than 4-core on-site) respective m2.2xlarge (8 cores, 34.2 GB, 41% slower than on-site counterpart).

It shows if memory consumption is low (in this case maximum memory is average memory on load from on-site) even much slower instances are cheaper than on-site hardware.

3.2 Application: Min P Test

To illustrate some cost effects in the cloud for a single application, we consider simplistic single nucleotide polymorphism (SNP) data simulation scenarios to compute gene region-level summaries.

In recent years, advanced technologies enable genotyping of thousands of SNPs, where subsequent analysis faces a severe multiple comparison problem since thousands to millions of statistical tests are conducted separately over multiple loci. Genetic case-control studies consider whether SNPs from such technologies are associated with some disease of interest using univariate statistical tests, i.e. tests for trend. Association analysis guided by a candidate gene approach utilizing candidate gene region-level summaries can be considered to ameliorate the multiple testing problem. Candidate gene region-level summaries, such as the min P test [21, 22], integrate all single locus tests within a gene into a single test statistic. The min P test combines the p-values from separately over multiple loci conducted statistical tests for trend into one number, minimum p-value, that represents the gene-level association.

The min P test is a permutation-based method that can be based on different univariate trend tests per SNP. In permutation resampling, the observed variable (case/control status) is randomly reassigned without replacement to “pseudo case/control status”. Then, a test statistic is recomputed using the “pseudo” data and compared to the marginal test statistic of the original data set. This procedure is repeated B times.

Calculations of permutation-based p-values, such as the min P test, can be time-consuming for large data sets. Fortunately, the procedure is predestined for parallel execution as each permuted data set is considered independently. However in our real application using the min P test a somewhat time-consuming initialization which can not be parallelized is necessary before the parallel part. Further details about the setting and the simulation program can be found in the Supplementary Online File. Table 3 shows the costs of running the simulation on three different instance types using the min P test package [23]. By choosing the appropriate instance type
4. Discussion

Our cost evaluation demonstrates that buying server hardware with a planned usage of three to five years for computing can be at the moment more cost-effective than investing in the cloud/EC2 instances for our department with an average system utilization of 25–30 percent. As prices for cloud services and server hardware as well as average system utilization will change over time our department will periodically re-evaluate the business decision on how to provision additional computing resources using the method described in this paper.

As we pointed out in the results section, our department’s on-site hardware may be more cost-effective if prices for EC2 instances drop over time. Furthermore, the argument that cloud computing gets cheaper in the future also is true for future hardware purchases where the same money will buy more cores and memory. Also, our department in general is buying industry-standard servers with redundant hard disks and power supplies as well as five year hardware warranty. The cost advantage of on-site hardware would be even larger by buying commodity white box hardware [24] as system stability and high-availability are not very important for our compute servers. Another useful, however not essential, item in our cost calculation is a dedicated server room with air conditioning system. In many departments powerful desktop computers also serve as computing infrastructure which also reduces costs for computing substantially on the cost of higher administration for these many machines.

Only CPU utilization excluding memory consumption is considered in cost calculations in Tables 1 and 2 and Figure 1 which leads to an underestimating bias, as often a single process is running on a machine consuming all the memory which also means the machine is fully utilized. Similarly, the comparison suggests a 100% utilization of the AWS instances, which is not realistic in practice. In our SNP example with a longer sequential setup part not all cores of an EC2 instance are used constantly. Due to technical issues, we stick to only CPU load here.

But in general also the opposite is true: if all processors are under full load, often only small parts of the memory are used. Therefore, for many jobs, machines with fewer memory will be sufficient, reducing the price of the cloud offers by a large amount.

Indirect costs are typically hard to estimate, however, it can be safe to assume that
indirect costs of using the compute infrastructure at our department are smaller than using EC2 instances for IMBI researchers. At our department, the same operating system is used for desktops and compute servers (Debian GNU/Linux). Accordingly, the working environment is familiar. Furthermore, users mostly run jobs over a convenient batch system sfCluster [25], so essentially no time is needed for cluster setup and managing of parallel R sessions using on-site hardware. All necessary data for statistical analyses are available on shared network directories without any additional action by the user. Therefore indirect costs of cloud computing are higher. The large impact of choice of an EC2 instance type on overall costs was described in the results section for the SNP example. Furthermore, users must understand some technical specialities of EC2 instances like the lack of swap space on most Amazon Machine Images (AMI) or a more complicated cluster management. Moreover, any test run or repeated computations will inevitably increase the amount of instance hours which will potentially influence the software development process strongly. For example with limited financial resources, testing software using a small dataset on local hardware and conducting the analysis in the cloud with the real data could be a sensible but more complex approach. Additionally, data management is more complex in the cloud.

We assume one needs at least 40–50 working hours to get started with the AWS concepts, working with the data management, evaluating and modifying suitable AMIs. Although there are many pre-configured and free to use AMI with operating systems and applications for statistical computing [26, 27], for real life purposes these need to be modified or tweaked to create a working environment. Although the general administration of the machines themselves is not required, the used AMIs still have to be maintained for updates on the operating system, programs etc. Introduction of cloud computing to working groups is, in our opinion, not possible without further administrative software or scripts, especially for cost observation – after all you are paying for each instance started until it is terminated. Also, if rented instances are not fully utilized, costs dramatically increase compared to on-site hardware, as payment is made for instances, not cores. Single users therefore can accumulate huge costs without any benefit very fast. Monitoring of cloud usage on the institution level is another problem. Commercial third party companies may help in this case, but come at additional cost.

The use of commercial statistical software in the cloud can be very expensive, as in the current situation the user must provide licenses for each used machine or processor core, depending on the license model of the specific software vendor. Therefore the scalability of the cloud services by just adding instances can quickly become due to the rise of license costs. In these cases it can be far cheaper to use single big instances instead of many smaller ones. Cost control is a real concern for cloud computing, especially with a limited yearly budget and possible traps discussed above. Also, in scientific settings the exact amount of needed computing hours may be unpredictable with changing demands over time or incalculable demand. Accordingly, there is a certain risk of running out of budget, as costs transform from investment (on-site) to variable costs (cloud). In addition, usage of AWS is currently complex as purchase on account is only possible for monthly expenditures over 2,000 $. Otherwise, payment by credit card is the only possibility which is very uncommon at least for academia in Germany. Furthermore, no individual cost limits can be set for users which might result in unpredictable costs, especially if several members of a working group are using cloud computing.

The major advantage of renting computers in the cloud is the ability to directly map utilization to cost. That means computers are only rented on demand, so in times without demand, no costs apply. Also, the scalability of the services allows one to instantiate as much virtual machines as you need in peak times (up to a certain degree). With suitable parallel applications, the runtime can be shortened by ordering more instances running for a shorter time overall getting results faster with a slight cost overhead. Combining the advantages of on-site (cheaper) and cloud computing (scalability) seems to be appealing at first glance, but with limited budget this could lead to a slow displacement of on-site hardware, as less money can buy fewer or slower machines.

Also, a big advantage of EC2 is the flexibility to shift individual priority between execution time and costs. As shown in Table 3, by choosing appropriate instance types, the user can trade longer overall execution times for far cheaper cost. Also the opposite is true: spending more money can lead to far shorter execution times. But as the reservation costs have to be paid for each instance type, this flexibility is maybe limited on some instance types on heavy usage.

In summary, our cost evaluation based on the system utilization in the year 2011 shows that at the moment on-site hardware, in most cases, is more cost-effective than using EC2 instances for our biostatistics department. For smaller working groups without appropriate computing infrastructure, applications with very low memory demands or only occasional demand of compute resources, e.g. limited time-constrained projects, cloud comput-
ing might be a useful option. The method described in this paper can be used to get the necessary information for this business decision.

References