

## Comparison of Cloud vs. Tape Backup Performance and Costs with Oracle Database

**Aljaž Zrnec**

*University of Ljubljana*

*Faculty of Computer and Information Science*

*aljaz.zrnec@fri.uni-lj.si*

**Dejan Lavbič**

*University of Ljubljana*

*Faculty of Computer and Information Science*

*dejan.lavbic@fri.uni-lj.si*

### Abstract

Current practice of backing up data is based on using backup tapes and remote locations for storing data. Nowadays, with the advent of cloud computing a new concept of database backup emerges. The paper presents the possibility of making backup copies of data in the cloud. We are mainly focused on performance and economic issues of making backups in the cloud in comparison to traditional backups. We tested the performance and overall costs of making backup copies of data in Oracle database using Amazon S3 and EC2 cloud services. The costs estimation was performed on the basis of the prices published on Amazon S3 and Amazon EC2 sites.

**Keywords:** computer center, bandwidth, service, cloud, performance, cost, remote location

### 1. Introduction

Cloud computing [5] allows users to tap into virtually unlimited pool of computational resources and data storage capacities via Internet. Compared with the traditional conception of the computers where users have full control over computing resources, cloud computing users have little or no access to and control over cloud computing infrastructure. So they have to interact with computational and data resources of the cloud using Web browser or appropriate APIs (API - Application Programming Interface) which have to be provided by cloud computing provider [1]. In return for these restrictions, cloud users receive more significant benefits such as ease of scalability, reliability, the independent and dynamic adaptation of the necessary resources and paying only for resources actually used.

The advantages offered by the cloud computing have enabled many service providers to offer different kinds of services over the Web. One of them which are especially interesting is the possibility of making backup copies of data in the cloud. Therefore, in Section 2 we will introduce distinction between traditional backup and data backup in the cloud. In Section 3 we will give an example of making backup in Oracle 11gR2 database, once without and secondly by using the cloud. In the context of that section we will also present a comparison of overall costs and speed between both aforementioned cases. In conclusion, we will present the findings and directions for further research.

### 2. Traditional backup and backup in the Cloud

#### 2.1. Traditional tape-based backup of data

Good practices in the field of database disaster recovery require backup storage of business critical data on a remote location, outside the business system [6]. Business system usually provides writing backups on magnetic tapes and sending them to remote location which is

expensive and complex process that requires special hardware, properly trained staff and procedures (regulations) to ensure that backups are regularly produced, protected and that the information contained in them can be obtained and used in the case of database accidents. Even if today it is customary for business systems to use outsourcing for the transport and protection of backup data, they still have to take care of data integrity in their backups and the aforementioned procedures.

## 2.2. Cloud backup

As an alternative to today's traditional backup of data, with the development of cloud computing, there are many providers of cloud services becoming available, that enable data backup in the clouds.

Backup in the cloud or so called online backup of data represents the way of making backups, where data from particular database are sent via the public or private network to a data server, located at a remote location [12]. Data server is managed by the provider of cloud service, who charges the customer for using data storage service, based on the required disk space, bandwidth or number of users of this service.

In our case the system for backing up data in the cloud is based on special application located at the user site of the cloud service. That application is launched with frequency (daily, weekly, etc.) defined in the contract on the use of service for backing up data in the cloud (SLA - Service Level Agreement). For instance, if the user (customer) has the contract for making daily backups, then the application gathers, compresses, encrypts and sends data to a data service provider's server every 24 hours. To reduce the bandwidth used for data transfer between the user and the cloud service provider, we may use incremental backups. According to the contract on the use of cloud service (SLA), the incremental backup is performed in intervals, only for changes of data in the original database. Since the data are transferred via the Internet, the bandwidth is usually relatively very limited. In addition, cloud service providers may also limit the bandwidth to prevent individual users from disproportionate use of resources in the cloud.

Based on the analysis of cloud service usage performed with Oracle database and Amazon's service for backing up data in the cloud - Amazon Simple Storage Service (hereinafter referred to as Amazon S3) [3], we found that the Amazon S3 service limits the data throughput in the context of individual sessions at 2.5 to 3.5 Mb/s. We obtained this results using special benchmark application, which was developed for testing data throughput of the connection between our test server and Amazon S3.

## 2.3. The advantages of Cloud backup

The main advantages of sending backup data over the Internet to the cloud are: flexibility of the cloud with regard to the performance needs, large amount of available storage space, and costs which are accounted only for actual use of resources. In addition, the use of cloud also significantly simplifies the user's own computational infrastructure, as there is no longer any need for own storage management (e.g. working with tapes, sending tapes to a remote location, etc.) and special hardware for making backups on tapes. Section 3.2 presents all the advantages offered by Oracle database for making backup copies of data in the cloud.

An important concern regarding data transfer to or from the cloud can be a limited bandwidth of the Internet which prevents fast transfers of a large amount of data (the problem of making full backup). Amazon service provider solved this problem by offering a special service (AWS Import/Export) [4], which allows movement of a full backup to or from the cloud and transportation with a portable hard drive. For example: after a disaster in the local operational database, the Amazon S3 service provider sends us full backup of data on portable hard drive by using express mail. In this way, data storage in the cloud is comparable to conventional data backup, especially when a remote location for storing backups is part of a business strategy, which includes both local backups and backups at remote location.

### 3. Backup in Oracle RDBMS

#### 3.1. Test environment

For the analysis we used the Oracle database 11gR2, which was installed on the computer server with 2.66 GHz Xeon processor, hard drive with 10.000 rotations per minute, 8 GB RAM and MS Windows Server 2008 x64 operating system. To back up data into the cloud, we selected Amazon S3 service from the Amazon service provider.

Amazon S3 is an essential Amazon's service for storing data in the cloud. The service provides through a simple web interface to store and transfer any amount of data to or from the Amazon's cloud. Amazon S3's main advantages are scalability which means that it acts independently of the number of users, reliability and speed. The price of service usage which is defined in the SLA is also very important, because it is based exclusively on the cloud resources, which are actually used. Amazon S3 service can be used for both classical textual and numeric data storage as well as serving multimedia content in real time.

The Oracle Cloud Backup Module (hereinafter referred to as CBM) enables an Oracle database to connect to Amazon S3 service and send its backups to the cloud. The module is compatible with all Oracle database versions 9iR2 and above. It requires a network connection to the Internet for its operation. CBM is a member of Oracle Secure Backup tools, used to create backup copies of data on traditional tapes or in the cloud. CBM can also be used when the database is running within the Amazon Elastic Cloud (hereinafter referred to as Amazon EC2), where the database is located on a virtual machine (server) inside the cloud [2]. In this case the CBM benefits from the higher internal network bandwidth and the lower cost of use by eliminating the costs for transferring data into and out of the Amazon S3. CBM is implemented using the Oracle Recovery Manager (hereinafter referred to as ORM), which enables easy integration with external libraries for making backup copies of data. In this way database administrators can continue to use existing tools for making backups. All previously described modules and backup process in the cloud is shown in Figure 1.

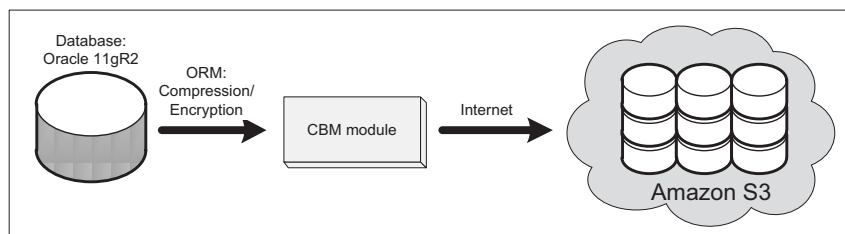


Figure 1. Oracle database backup in the cloud

#### 3.2. The advantages of using Oracle RDBMS for Cloud backup

Oracle's cloud backup functionality provides many advantages over traditional tape-based offsite backups:

- *Continuous accessibility*: Backups stored in the cloud are always accessible, much in the same way as local disk backups are. As such, in case of disaster, there is no need to call anyone and no need to ship or load tapes before a restore can be performed. Instead of this, administrators can initiate restore operations using their standard tools for disaster recovery (Enterprise Manager, script, etc.), just as if the offsite backup was stored locally. In this way restore activities can be performed faster and down time can be reduced from several days to a few hours or even just a few minutes. In case of large databases, when shipping a portable disk from the Amazon cloud service provider is required, a restore takes no longer than it would take to have a tape recalled from the remote location.
- *High level of reliability*: Storage clouds are disk based, so they are inherently more reliable than tapes. Additionally, cloud vendors typically keep multiple redundant

copies of data to offer a higher level of availability and scalability. For instance, Amazon S3 redundantly stores user's objects on multiple devices across multiple facilities. The service is designed to sustain concurrent device failures by quickly detecting and repairing any lost redundancy. When processing a request to store data, the service will redundantly store your object across multiple facilities before returning SUCCESS.

- *Unlimited increase in space and no upfront capital expenditure:* The cloud provides virtually unlimited capacity to store data with no upfront capital expenditure. This means that the cloud dynamically adjusts the size of storage space to hold the required backup data and users pay only for the space actually used.
- *Reduced usage of backup tapes and reduced offsite storage costs:* Clouds reduce or even eliminate the need for the tape based backup. This can lead to significant savings in buying the tape backup hardware and software and also savings in tape storage costs at remote locations.
- *Easy provisioning of test and development environments:* Cloud backups are accessible from anywhere via the Internet. They can be used to quickly clone databases to create custom test or development environments. For instance, cloud backup stored in the Amazon S3 can be cloned to virtual machine (virtual server) running in Amazon EC2 by running a simple script.

### 3.3. Assuring data security

In shared, publicly accessible environments such as the storage cloud, data security and privacy is particularly important. Therefore, when sending data to the cloud, Oracle CBM module uses special functionality of ORM component, which uses encryption to assure data security and privacy. In this way data are protected twice. One level of protection against unauthorized access is already ensured by cloud service provider, the second level is guaranteed by the above-mentioned encryption of backup data, before they are sent into the cloud. This reduces the risk of theft or unauthorized access to the data during the transport, as well as at storage of data in the cloud.

### 3.4. Data compression

Since the CBM module integrates into Oracle RDBMS, it can independently identify and skip unused space (blocks) in the database before the backup is made and sent into the cloud. At the same time the ORM component offers many possibilities for data compression, which has a direct impact on the speed of making backup copy. Relatively slow Internet connections have the greatest impact on the speed of cloud backup creation. Thus, within the analysis of the impact of data compression on the speed of data backup creation, we compared performance of making backup copy of data with and without using data compression.

### 3.5. Cloud backup performance

As already mentioned in Section 2.2, Amazon S3 may throttle the throughput of individual session from 2.5 Mb/s to 3.5 Mb/s to prevent individual users from consuming disproportional amounts of cloud resources. However according to Oracle, we found that it is possible to attain higher data throughput with the use of data compression and several concurrent data streams - CBM channels. We attained data throughput of 43 Mb/s to 55 Mb/s, which had a crucial impact on the speed of backup creation speed. Without compression we attained a data throughput of 10 Mb/s to 35 Mb/s.

In examining the performance of backup data in the cloud with the Oracle database, we performed several measurements, in which we determined the time needed for creation of backup. Firstly, we measured the duration of the backup creation in database that was located on our test server, and secondly on virtual server located in Amazon EC2 [8]. For each database we observed the impact of data compression on time for backup creation.

Measurements were performed for both, producing complete and incremental backups. The size of a full backup was 250 GB, and the size of an incremental backup was 10% of data changes made to the database. The results of the measurements are summarized in Table 1 (the first two lines). For better understanding and result comparison, Table 1 also includes the results of performing traditional backup on tapes (the third line).

Database location	Data throughput		Full backup time		Incremental backup time	
	Compression		Compression		Compression	
	No	Yes	No	Yes	No	Yes
Local test server	10 MB/s	43 MB/s	< 7 h	> 2 h	< 1 h	> 30 min
Virtual server in Amazon EC2	35 MB/s	55 MB/s	< 2 h	> 1 h	< 20 min	> 10 min
Local test server, transport to remote location and storage in the vault			(< 1:15 h) < 4:14 h	(> 15 min) > 3:15 h	(< 10 min) < 3:10 h	(> 1:42 min) > 3:01:40 h

Table 1. Oracle cloud backup and traditional tape based backup performance comparison

Some of the observations drawn from these results are:

- The time for full or incremental backup of the test database (without using compression of data) which is located on a local test server, is 3.5 times longer than if the database is located on a virtual server in Amazon EC2.
- The time to back up the test database located on a test server (using data compression), is two times longer (full backup) or three times longer (incremental backup) than if the database is located on a virtual server in Amazon EC2.
- The time for full backup of the test database which is located on a local test server, is at most 3.5 times longer without the use of compression, and in case of incremental backup it is at most two times longer without the use of compression.
- The time for full or incremental backup of the test database which is located on virtual server in Amazon EC2 is at most two times longer without the use of compression.

Additional comments:

- In both cases, when using local test server or virtual server in Amazon EC2, the compression can vary during the data transfer. The compression rate is dependent on data being backed up. If the compression rates lower, the times for data transfer increase. This is shown using "greater than" values in the table.
- In our tests data compression increased data throughput more than four times on our test server and by around 50 % on virtual server in Amazon EC2. So, the "greater than" values in the table present the minimum of the time used for creating backup. These times can increase to "lower than" values in the table, which represent the times for backing up data without compression. This is the extreme limit which would be achieved if we compressed already compressed data.

It can be concluded that both the throughput of the network (in this case Internet) and the level of data compression have great impact on the speed of making backups in the cloud. Furthermore we shall also mention, that Oracle database 11g versions and above use advanced compression mechanisms, which are significantly faster and more efficient in terms of CPU overhead than pre-11g compression.

Backup speed is also accelerated by the ORM module, which makes it possible to use multiple parallel transmission channels enabling to fully utilize the network. The highest performance in our tests was achieved with 64 simultaneous channels. Oracle database 11g

allows multiple channels to back up a single file in parallel, increasing the parallelism beyond the number of data files to be backed up.

### 3.6. Cloud backup cost assessment

When estimating the costs of backing up data in the cloud [3], [4], we derived from the costs of using Amazon cloud services (Amazon S3 and Amazon EC2). The cost of Amazon S3 service includes the price of 325 GB cloud storage (see description of cloud backup scenario below) and the price of Amazon S3 usage. The cost of Amazon EC2 service includes the price of virtual server image set up in Amazon EC2 and the price of data transfer into the cloud. We estimated the total costs for the period of one month, assuming that at start we had to transfer 250 GB of data to the cloud (full backup copy of the database), and then three times a month (weakly) we made incremental backup of database. The Internet connection speed was limited to 10 Mbit/s. The size of incremental backup was 25 GB. We used a special tool on Amazon S3's web page to calculate necessary costs of using cloud storage. We found out that using portable hard drive to transfer full backup to the cloud is not reasonable, because the costs of this kind of transfer amount to 235\$, versus transfer of the same amount of data over 10 Mbit Internet connection, which costs 25\$, and it lasts for 3 days and 6 hours. Besides, the transport of the portable hard drive to Amazon's computer center also lasts 3 to 4 days. Total costs of backing up data in the cloud (using Amazon S3 service) are presented in Table 2. For better understanding and result comparison, Table 2 also includes costs of performing traditional backup on tapes.

<b>Amazon S3 service</b>			
<b>Storage price (the price of the first 1 TB/month is 0,14\$/GB):</b>			
<b>Amount of data</b>	<b>Storage time (days)</b>	<b>Calculation: 0,00452\$/GB day * Storage time [days] * Amount of data [GB]</b>	<b>Price</b>
full backup - 250 GB	31 days	0,00452*31*250	35,00\$
1 <sup>st</sup> incremental backup	24 days	0,00452*24*25	2,71\$
2 <sup>nd</sup> incremental backup	17 days	0,00452*17*25	1,92\$
3 <sup>rd</sup> incremental backup	10 days	0,00452*10*25	1,13\$
<b>The price of data transfer to the cloud:</b>			
1 x 250 GB		1 x 25\$	25,00\$
3 x 25 GB		3 x 2,5\$	7,50\$
<b>TOTAL:</b>			<b>73,26\$</b>
<b>Traditional tape based backup</b>			
	<b>Unit price</b>	<b>Number of units</b>	<b>Price</b>
Tape drive	2750\$	1	2750\$
Transport of the tapes with courier service	40\$	4	160\$
Lease of the vault (1 year)	240\$	1	240\$
Cartridge (400 GB) for full backup	45\$	1	45\$
Cartridge (400 GB) for incremental backup	45\$	3	135\$

<b>TOTAL:</b>	<b>3330\$</b>
---------------	---------------

Table 2. Oracle cloud backup and traditional tape based backup cost estimation

When using the Amazon EC2 service, where database is located on a virtual server in the cloud, the price of backup is equal, since the data are also transferred to the Amazon S3 service. The advantage of using virtual server is mainly in the speed of backing up the data, because the data are transferred within the cloud (Amazon service provider), where we are not limited to a throughput of the Internet. So in our case, the full backup would be made in less than three days. We can also mention the cost of virtual server usage, which amounts to 0,62\$/h for Extra Large, High-memory instance of a server running Windows Server 2008 x64 operating system. The price for one month is therefore 461.28\$, as compared to the price of the physical server, nearly 10-fold less. For example, the price of our test server with equivalent operating system amounts to 4200\$.

**3.7. Traditional backup speed and cost estimation**

We measured the time of traditional backup so that we added the time for making backup on tape and the time for transportation of the tape to the remote location. Transportation time was estimated to 3 hours. That was the time needed to transport the tape on remote location which was 100 km away and put the tape in the vault. In doing so, we measured the time for full and incremental backup, without and with the use of data compression, similarly as described in Section 3.5. We used the tape drive for recording data on tape, courier service for transportation of the tapes to remote location and the service for tape storage in the vault at the remote location. In the selection of tape drive, we derived from the requirement that it must be able to record all data (full or incremental copy) on a single cartridge. The times to produce the full and incremental backups are shown in Table 1 (the third line). Times in brackets represent the backup times without transportation time.

From the full or incremental backup times given in Table 1, it can be concluded that:

- The minimum time required for full backup is 3 hours and 15 minutes. Compared to the time required for full backup in the cloud (2 hours), that is 62,5% more.
- The minimum time required for incremental backups is approximately 3 hours and 2 minutes, which is 6-fold more compared to the time required for incremental backup in the cloud (30 minutes).

In assessing the costs of traditional tape based backup, we assumed that we needed the backup tape drive, the appropriate number of tape cartridges, courier service for transportation of the tapes to remote location and the service for tape storage in the vault at the remote location. Costs were similarly to Section 3.6 evaluated for the period of one month, assuming that at start we had to make a full backup of the database, and then three times a month (weakly) we made incremental backup of database. Total costs of making tape based backups at the remote location are presented in Table 2.

The cost of tape based backup using the scenario mentioned above, would amount to 3330\$ in the first month, because we would have to buy a tape drive and because it is necessary to hire a vault for at least a year. In the following months (if we limit ourselves to a period of one year) the cost would reduce because of the price of tape drive and the price of the vault. Thus, the costs would reduce to 340\$. In spite of all that we can see that making such backup copies are still 4.6 times more expensive than data backups in the cloud.

**4. Conclusion**

Business systems can use very different scenarios for making backups. Therefore, in the light of obtained results, it doesn't make any sense to act without proper consideration and start to feverishly think about using a cloud. First, business systems should determine their

requirements about the performance regarding the speed of making backups and the costs associated with it. Only then they can find turning point on which to decide whether it makes more sense to use traditional way of backing up data or to use a cloud.

In addition to the performance and costs, we have discussed in this article, business systems should answer a number of other important questions on making backups in the cloud [11]: Can they trust the data to a cloud service provider [7]? What will happen to the data, if the service provider stops working? Can data be transferred "to another cloud" - to a different cloud computing provider? Are data really rejected after they become unusable? Are data in the cloud safe from a theft [10], [9]? Thus, there is a set of additional questions, which have to be answered in the context of further research, before the use of a cloud computing will really flourish.

## References

- [1] A. REED, S. G. BENNET, *Silver Clouds, Dark Linings: A Concise Guide to Cloud Computing*, Prentice hall, ISBN-13: 978-0-131-38869-7, 2010
- [2] Amazon Elastic Compute Cloud (EC2), <http://www.amazon.com/ec2/>
- [3] Amazon Simple Storage Service (Amazon S3), <http://aws.amazon.com/s3/>
- [4] AWS Import/Export, <http://aws.amazon.com/importexport/>
- [5] B. HAYES, Cloud Computing, *Communications of the ACM*, Vol. 51, No. 7, pp. 9-11, 2008
- [6] Bright Hub, The Next Generation of Server Backups - Sequential is OUT!, <http://www.brighthub.com/computing/hardware/articles/22328.aspx>, 2009
- [7] D. SVANTESSON, R. CLARKE, Privacy and consumer risks in cloud computing, *Computer Law & Security Review*, Vol. 26, No. 4, pp. 391-397, 2010
- [8] G. J. POPEK, R. P. GOLDBERG, Formal requirements for virtualizable third generation architectures, *Communications of ACM*, Vol. 17, No. 7, pp. 412-421, 1974
- [9] N. STINCHCOMBE, Cloud computing in the spotlight, Vol. 6, No. 6, pp. 30-33, 2009
- [10] S. SUBASHINI, V. KAVITHA, A survey on security issues in service delivery models of cloud computing, *Journal of Network and Computer applications*, Vol. 34, No. 1, pp. 1-11, 2010
- [11] S. VRHOVEC, R. RUPNIK, A model for resistance management in IT projects and programs, *Electrotechnical review*, Article in reviews
- [12] techs@work, Cloud backup vs local external backup!, <http://techsatwork.com/blog/?p=694>, 2010