Vol. 9(10), pp. 544-548, November 2021
https://doi.org/10.14662/ijalis2021480

ISSN: 2360-7858
http://www.academicresearchjournals.org/IJALIS/Index.htm

International Journal of Academic Library and Information Science

Full Length Research

Best practices for efficient and effective database archiving

[1]NWORIE, Josiah Chukwumaobi, PhD, CLN, [2]YUSUF, Felicia O. PhD, CLN and [3]Dr. (Mrs.) Nkechi M. AMAECHI

[1]Senior Librarian, Alvan Ikoku Federal College of Education Owerri Imo State.
Corresponding author's E-mail: chimanworie@gmail.com / 08037319003
[2]University Librarian, Bells University of Technology, Ota Ogun State.
E-mail: foyusuf@bellsuniversity.edu.ng / 08033145024
[3]Deputy University Librarian, Federal University of Technology, Owerri, Imo State.
E-mail: ugegbe2015@gmail.com. / 08033600494

Accepted 31 November 2021

Day by day, data are being generated and in a very large volume, especially with the emergence of information and communication technologies (ICTs). Some of the data generated are for long term usage, while some are for short term usage. Some may be needed for a very long term but may not be actively needed, some space need to be created for them where they can be kept in safety for recall when they are needed. This is what database archiving is all about. At a very short space, technology keep evolving, new apps and newer models are manufactured, old ones go into extinction, and accessibility of the archived data becomes a problem etc. This paper therefore discusses those best practices that would enable data objects in databases being archived to continue to be relevant and accessible. The concept of database, types of database and benefits, types of database archiving, challenges of database archiving, benefits of database archiving etc are looked into.

Keywords: Archiving, database, DBMS, ICTs, Softwares

Cite This Article As: NWORIE, J. C., YUSUF, F.O., AMAECHI, N.M. (2021). Best practices for efficient and effective database archiving. Inter. J. Acad. Lib. Info. Sci. 9(10): 544- 548

INTRODUCTION

Human beings are wired to live and behave in certain ways based on their information contents, and information capacity. Information is packaged in various forms and shapes. When information is not yet processed, it can be called data, and when it is properly processed and put into a record, it is said to be knowledge. These processed data can be put into some forms and sources like books, journals, conference papers, dissertation and theses, patents, research reports, trademarks etc. The collection of these information sources together, organised on how they are related is called database. These sources can be further put into electronic format and be properly organised and that is what is called an electronic database. According to Bothma (2014), a database is an organised collection of data or information that is stored in records in electronic format. These organised collections of data can be accessed manually or electronically from computers. When they are runned electronically, they run through a system along with applications that are associated with them which is referred to as database management system (DBMS). These data need to be properly managed so that they can still be accessed for a very long time. When data are generated, processed and are not well taken care of, they can deteriorate and become useless and may not be accessed anymore (Nworie and

Nwosu, 2019). Accessibility is one of the important components that makes data useful and relevant. These data that are organised, need to be well and safe for a very long time so that they can still maintain their relevance and integrity, and how they are kept is very important. Very often, due to the presence of information and communication technologies (ICTs), data are being generated every now and then. Organisations and companies therefore keep battling with management of these data, yet, with limited storage capacities. Some of these data generated serve out their lifetime and are destroyed, while some can still be needed for a longer time even though they may not be actively needed. However, they may still be needed and can be recalled from safe place where they are domiciled, and that process of keeping the data that are not actively needed in a place of safety until when they are needed is what is called data archiving.

All the data in a database are not archived completely at once. Some data objects are selected for archiving, while some are retained for active usage. The problems of technological advancement and upgrades, obsolescence and damage are very prevalent and obvious in database archiving. It is imperative therefore to put in place best practices of how these data objects in a database can be archived in this 21st century thus the relevance of this discuss.

WHAT A DATABASE MEANS

A database can be defined as a huge collection of data or a collection of related files. It can also be explained as a collection of information organised in such a way that a computer programme can quickly select desired pieces of data. A database can be in a manual form or electronic form. It is manual, when it can be operated without the aid of electricity or computer. That means that a file record in an office built with wood or steel where records of people or events about a company or institution is stored, operated manually, is an example of a manual database. An electronic system on the other hand is an example of electronic database.

According to oracle (2021), a database is an organized collection of structured information, or data, typically stored electronically in a computer system. A database is usually controlled by a database management system (DBMS). Together, the data and the DBMS, along with the applications that are associated with them, are referred to as a database system, often shortened to just database.

Data within the most common types of databases in operation today is typically modeled in rows and columns in a series of tables to make processing and data querying efficient. The data can then be easily accessed, managed, modified, updated, controlled, and organized. Most databases use structured query language (SQL) for writing and querying data.

Types of Databases

There are many different types of databases. The description of the records contained in a database will often determine its type. However, the most common types of databases include:

- Bibliographic databases
- Full text databases
- Image databases
- Audio/video databases
- Mixed databases (a combination of any types of information)

As stated above, oracle (2021) states that a database is usually controlled by a database management system (DBMS). It is the software that is used to specify a logical organisation and access the database. Some examples of popular database software or DBMSs include mySQL, Microsoft Access, Microsoft SQL server, Filemaker pro, Oracle Database, and DBASE.

The DBMS contains (5) five important software components:

1. DBMS engine
2. Data definition soft system
3. Data manipulation system
4. Application generation subsystem
5. Data administrations of system.

There are four major types of DBMS:

- Those that organise data using a relational model
- Those that use hierarchical model
- Those that use network model
- Those that use object oriented model.

Benefits of Databases

- It helps one to have access to multiple files at one time
- The DBMS makes it easier and quicker to find information that one needs
- Redundancy is reduced through the DBMS maintenance

DATABASE ARCHIVING

Archiving according to Olson (2009) is the process of preserving and protecting artefacts for future use. These artefacts' have lived their useful life and are being kept solely for the purpose of satisfying future historical investigations or curiosity that might or might not occur. An archive is a place where these artefacts are stored for long periods of time. They are kept in a safe place and in a manner that when they are needed in the future, they

will still be available and accessible.

The term database archiving, however, can be problematic as it is used differently by different communities. According to Muller (2009), the most common definition of database archiving comes from the business and government communities as is defined as the process of removing selected records from operational databases that are not expected to be referenced again and storing them in an archive data store where they can be retrieved if needed. Following this definition, database archiving requires the active selection and appraisal of data records to ensure that those no longer deemed necessary for daily operational or reference purposes are moved to a separate data store for longer term retention. The data is maintained in the archive for as long as required based on legal or institutional requirements. If and when the data is no longer required, it is either transferred or destroyed.

When archiving scientific and reference data, database archiving is frequently regarded as maintaining a collection of database snapshots over time. This form of database archiving involves making off-line copies of the data and managing these copies efficiently. In addition to being able to reconstruct the database as it was at a certain point in time, database archives enable tracking and querying of the history of objects.

Database archiving is usually seen as a subset of data archiving. Database archiving focuses on archiving data that are maintained under the control of a database management system and structured under a database schema, e.g., a relational database. The primary goal of database archiving is to maintain access to data in case it is later requested for some particular purpose such as a Freedom of Information (FoI) request. In fact, compliance with government regulations on data preservation is the main driver for the majority of current data archiving efforts.

Types of documents to be archived depend on the company or people that are involved in using the data. Government and companies accumulate a large number of text documents that can take a form of Microsoft word files, acrobat PDFs, spread sheets, power point documents etc. Other documents that can be archived include XML documents, emails, multimedia and other complex documents and databases.

Types of Database Archiving

Basically, there are two major types of database archiving according to Olson (2009), the ***physical and virtual database archiving***.

a. The Physical Database Archiving involves removing data from the physical database storage and sending it to the archive. Here, there is no consideration or concern about the application programs that will be used to generate data in the relational tables.

b. Virtual Database Archiving - This happens when the archivist defines what the archive will look like by using a source of reliable data objects and their representation made possible through the application programs. These application programs have interfaces to extract data for specific data objects in a meaningful structure, to search for objects, and to delete objects. When the Archivist uses these application interfaces instead of the underlying tables, this is referred to as virtual database archiving.

Database Archiving tools and Softwares

Archiving software optimises the storage, discovery, and retrieval of corporate documents, emails and websites. Data retention is an important part of many industry regulatory requirements. You can analyse, visualise and archive files easily with the right archiving solution. Some of the archiving softwares are free based, while some are fee based. Some are mobile technology enabled while some are web based. According to Capterra (2021), there are over 800 categories of archiving softwares, ranging from Accounting to yoga studio management. Some examples include comet backup, swing seascape for notes, grax, widen collective, veeam backup free edition, nakivo backup & replication etc.

Benefits of Database Archiving

According to Namuag (2019), archiving your data provides many benefits, especially in terms of efficiency such as storage costs, optimizing data retrieval, data facility expenses, or payroll for skilled people to maintain your backup storage and its underlying hardware. To Muller (2009), there are a good number of businesses and research related benefits associated with database archiving. Some of these benefits include:

- Verification of scientific findings - Scientific data that are in the archives enable historical database to be retrieved. This can be used to verify scientific findings that were based on data.
- The history of the database can be queried. Database archiving helps tracking and querying the history of database objects. This capability to answer temporal queries is very valuable for reference databases. For instance, databases like the CIA world fact book can answer queries like, how did the energy consumption in China change over the past 15 years?
- Reduced cost of hardware requirements – it is noted that the bigger a production database, the more infrastructure, i.e., server and disk capacity it requires. Regular scheduled database archiving helps to free up disk capacity, and this saves money on hardware

upgrades. Throughout the information life cycle, maintaining active data online and selecting the most appropriate storage medium for archived data ensures cost effective balance of storage media.

- Improved application performance – An overloaded database can degrade performance. Database archiving ensures that production databases are maintained at a manageable size to improve performance and availability of critical systems. To remove rarely used data from production databases can free the processing power and improve application performance.
- Improved database administration – more administration efforts are needed for upgrade, migration and backup especially for large databases. This can potentially lead to long outages for critical applications.
- Legal and regulatory requirements – it is the expectation of several regulatory agencies that enterprises should store all related data for several years before deleting them, and archiving is the viable tool to storing such data and meeting compliance requirement.
- Customer service issue – one of the ways of resolving customers' issues is to maintain historic information in an archive. This enables the resolution of customer-related queries that may span several months or years

Challenges / Problems with Database Archiving

According to Olson (2009), the following pose as problems to database archiving.

1. Some data may need to be left behind while archiving some data. An entire database may not be entirely archived at one time, at least not typically. Some data are selected from within the database to be archived while leaving the unselected items behind. This type of archiving most often generates special problems. The archiving is expected to capture and maintain the metadata for each application as well as data itself. The metadata is required to explain the structure of the data as well as the semantic meaning of each data element'
2. The metadata and its corresponding data objects will change overtime for any specific application, because applications never remain the same. What this means is that data for specific application may have one structure for part of the data in the archive, and another that may seem similar for another part of the structure. Tracking the metadata evolution for an application becomes a major problem.
3. Another obstacle for database archiving can be the huge amount of data that accumulates in a single database. For instance, having the data records of credit card transactions of customers for just one year, imagine, the amount of data for that same company with a 25year retention requirement.

DATABASE ARCHIVING BEST PRACTICES

For an effective and efficient database archiving to take place, certain practices need to be put in place. This i classified into two major categories:

a. Policies
b. Strategies

a. **Policies –** For an effective and efficient database archiving to take place, an all inclusive archive policy should be developed. Policies are very important to facilitate the survival of database archiving. Policies help to set out goals to be achieved as well as guidelines for implementing them. Policies also facilitate creative allocation of funds, staff, strategies for database archiving. Agreement and advocacy rights issues are all taken care of in the policy guidelines (Nworie & Nwosu, 2019). There can be no serious database archiving without a policy that will guide effort at solving identified archived problems.

According to Namuag (2019), data archives are usually taken alongside through a backup life-cycle process. This life-cycle process has to be defined within their backup policy. Most backup policies have to undergo the process as listed below.

i. it has the process defined on which it has to be taken (daily, weekly, monthly),
ii. if it has to be a full backup or an incremental backup,
iii. the format of the backup if it has to be compressed or stored in an archived file format,
iv. if the data has to be encrypted or not,
v. its designated location to store the backup (locally stored on the same machine or over the local network),
vi. its secondary location to store the backup (cloud storage, or in a collo),
vii. and it's data retention on how old your data can be present until its end-of-life or destroyed.

b. Strategies - Having identified the policies that will guide the database archiving, the next important thing is the action strategies that will need to be followed for effective and long lasting database archiving. Strategies are action plans that bring dreams into reality. According to Namuag (2019), the following strategies or action plan are recommended for best practices when storing to the cloud.

i. It is important to identify the type of data to be archived – data backs are not the same as archived data. Data back ups are data that are active but are backed up and are recalled when needed, whereas, archived data are those data that are not active and are only needed at the long run. They have served out their life purposes but may be needed after a very long time, and they are infrequently accessed. You need to identify

first what are these, mark a tag or add a label to these archived data so you would be able to identify it when stored off-site.

ii. **Determine data access frequency -** the data to be archived needed to be identified how frequently they will need to be accessed when archived. Certain price can differ at the time the data have to be accessed. For example, Amazon S3 will charge higher if you avail for expedite retrieval using provisioned instead of on-demand, same thing with Microsoft Azure when you rehydrate archived data with a higher priority.

iii. Multiplication of Copies – ensure that multiple copies of either archived data or stagnant data are highly available and durable when needed. Some applications that are offered by some vendors have room on how data can be stored for efficiency and faster accessibility. Infact, when configuring the life cycle policy or back up policy of your data archiving, ensure that there is provision for multiple copies production or replicate your data in different region. It is interesting to note that most of these hitech giant cloud vendors store their archival cloud storage offerings with multiple zones to offer highly scalable and durable services in times of data retrieval, when it is requested.

iv. Data Compliance – Make sure that data compliance and regulations are followed accordingly and make it happen during initial phase and not later. Unless the data doesn't affect customer's profile and are just business logic data and history, it might be harmless when its destroyed but It is better to make things happen in one accord.

v. Maintain good Provider Standards – Walking the path of online data archiving and backup with an experienced service provider could save you from unrecoverable data loss. Therefore, it is good to choose the right cloud back up and data retention provider. The top tech giants like IBM, ALIBABA or Oracle Archiving Storage should be considered over everything. It can be best to try it out before making the final decision.

CONCLUSION

This study has been able to look at different types of databases, database archiving benefits and problems, and best practices for effective and efficient database archiving. It may seem difficult at the beginning on how to adopt the best practices, but with the right set of tools or bundled software, good policies and strategies, it can be easier to implement the process.

REFERENCES

Bothma, T., Cosijn,E., Fourie, I., Penzhorn, C., (2014). *Navigating information literacy: your information society survival toolkit.* Capetown: Pearson holdings Southern Africa (pty) Ltd.

Capterra (2021). Databse Archiving softwares. Retrieved on October 10, 2021 from http://www.capterra.com>archivingtools

Data Archiving best practice overview/redstor. Retrieved on September 15, 2021 from https://www.redstor.com

Müller, H. (2009). "Database Archiving". *DCC Briefing Papers: Introduction to Curation. Edinburgh: Digital Curation Centre. Handle: 1842/3346.* Available online: /resources/briefing-papers/introduction-curation

Namuag, P. (2019). Best Practices for Archiving your Database in the Cloud. Accessed on September 20, 2021 from *https://severalnines.com/database-blog/best-practices-archiving-your- database-cloud*

Nworie, J.C., Nwosu, O. (2019). Adoption of UNESCO Digital Preservation Guidelines for Measuring Preservation Policies of Digital Materials in University Libraries in South East, Nigeria*. International Journal of Academic Library and Information Science 7(3), 74-81*

Olson, J. (2009). *Database Archiving: how to keep lots of data for a very long time.* Massachusette: Morgan Kaufmana Publishers

Oracle, (2021). Databases. Retrieved on October 10, 2021 fromhttps://www.oracle.com/database/what-is-database/.