

# Medical Image Handling in the Cloud based on Document Databases

D. Revina Rebecca  
Research Scholar  
Department of Computer Science  
Avinashilingam University  
Coimbatore, India

I. Elizabeth Shanthi  
Associate Professor  
Department of Computer Science  
Avinashilingam University  
Coimbatore, India

## ABSTRACT

Medical images are the key to healthcare industry. Medical images are acquired from various different modalities and they produce extremely large data files, and the modalities used to create them are constantly evolving. As the medical images need to be archived for future references. Archiving medical images locally is a huge challenge, which involves huge investment by the health care providers. A better solution would be moving the medical images to a cloud environment, which provides lot of flexibility in archiving as well as retrieving the images. A Database as a Service will be more advantageous in moving medical Images o the cloud. The NoSQL databases are robust in handling Data in the cloud. The suitability of NoSQL databases in storing the medical images is considered and it is found that the document databases to be suitable[3]. In this paper a performance based study is performed on two document databases in handling huge medical images. The various performance metrics analysed can be the foundation to fix up the right database in developing an framework in moving medical Images to the cloud .

## Keywords

Big Data, Medical Images, Cloud Computing, NoSQL Databases, Document Databases.

## 1. INTRODUCTION

The number of medical images generated per annum is estimated to be in millions, with an every year increment of approximately 20-40 percent, making it a great struggle for healthcare organisations to manage, share and access this data while trying to be economical. It is estimated that in 2020, 30% of the world's data stored would be only medical images.

Picture Archiving and Communication Systems(PACS) used by radiologists, and the available data storage capabilities being the currently used technologies, radically limits the efforts to harness the huge amount of this medical image data as they don't scale. And cloud computing is now coming to the picture as it offers to be an attractive alternative for management of medical imaging data. Also at the same time

Storing, archiving, sharing and accessing images in the cloud allows the industry to manage data more efficiently and cost-effectively while overcoming many of the legal, regulatory and technical challenges that data requirements pose.

## 2. BACKGROUND

### 2.1 Cloud Computing

Cloud computing is as good as an emerging player in the field challenging all the so called limitations and provides features like capacity to manipulate data with software packages and computational algorithms along with storage, synthesis,

retrieval and aggregation of data. It provides on-demand a scalable storage and computational platform which supports remote accessing. It reduces the local storage requirements significantly meanwhile facilitating data sharing in a further more effective way. Cloud will be a better option when it comes to transfer and interchange of data over the internet and to providing complex services, since the patient details will be stored in virtual archives which will be accessible to different healthcare providers rather than its physical counterpart.

Cloud provides a better storage option for the Medical Images, as medical images are huge in size. There are many advantages of storing data in the cloud. The various advantages of moving the medical images to the cloud being

- More efficient in handling images of huge size.
- Provides elasticity in storage
- It manages authentication, encryption and security protocols
- System-wide application upgrades can be efficiently conducted

### 2.2 NoSQL

The term 'NoSQL' collectively refers to the database Models which are Non-Relational. These Databases are known to handle Unstructured and Semi-Structured data effectively. NoSQL databases are highly availability through replication and are scalable, which are essentially needed for big data handling. Scaling is achieved by partitioning data over many nodes, thereby achieving stable and fast read/write operations of massive data [6], [7].

NoSQL databases are broadly classified into three major types, based on their data models. They are key-value stores, document databases, and column-oriented databases.

In a key-value store, all the data instances are stored in the

form of key-value pair. Document databases stores data as documents which are grouped into collections.

Column-oriented databases group similar columns as column families. Each column in a Record is identified using three keys, A row Key, A Column family Key and A Column key.. This type of database is efficient when the majority of the database operations requires intensive column-oriented calculations like aggregation.

Table I- NoSQL Database Classification

| Key-Value Data bases   | Column Databases            | Document Databases          |
|------------------------|-----------------------------|-----------------------------|
| Redis, Riak, Voldemort | Bigtable, Cassandra, HBase, | MongoDB, CouchDB, Couchbase |

Table I shows representative NoSQL database solutions of each type. Eventhough it is not possible to replace the more mature Relational databases by NoSQL in the near future, NoSQL will be a better choice and will be adopted for some specific use cases such as handling unstructured massive data which requires high level of scalability. There is a wide variety of NoSQL databases products available. These are built to fit in for specific Usecases.

Therefore, NoSQL databases will have their own niches, which are even expanding rapidly according to growing needs for big data processing in various fields. NoSQL databases are very much use case specific.

### 3. MEDICAL IMAGE STORAGE

Various methods employed in storing medical images

- i) PACS(Flat files)
- ii) RDBMS
- iii) NoSQL

#### 3.1 PACS Storage:

Digitized medical images can be stored in different ways. Medical images are stored in DICOM(Digital Imaging Communication in Medicine) format as a flat file. These images are accessed through PACS(Picture Archive Communication Systems), which is RDBMS based storage. The major drawback in a PACS based system is the inability to scale. The PACS system has some limitations in scaling.[1]

DICOM (Digital Imaging and Communication in Medicine) being an international standard which defines network services, file formats and requirements for storage and transmission of digital medical images and reports, forms a component of most of PACS radiological imaging data in the form of DICOM complaint objects. A header file (Identifies the file as a DICOM file), a metadata portion (image related information) and a pixel data portion (which contains the actual image pixels) all together constitutes a DICOM object. This structuring makes images self-contained, tags or data elements that encode patient, image and study related attributes are found in its metadata.

#### 3.2 RDBMS based Storage:

The problem comes in to picture when it comes to loading of these DICOM objects in to RDBMS data base since, much of the metadata which are freely structured and inherently heterogeneous in nature will be stripped out so as to support the predefined database schema, thus compromising the data.

The various drawbacks of storing medical images in a RDBMS based data storage is well discussed in [4]. The suitability and performance of a document database is better compared with many other data models. [5].

#### 3.3 NoSQL Based storage:

NoSQL can be used to store medical images which is experimented by many researchers. A CouchDB based medical archiving system was developed Rascovsky et al. The authors suggest that document databases are highly suitable to store, retrieve and query DICOM files.[3].

The medical images can be stored on to a NoSQL data store before they are moved to the cloud. There are many advantages in storing a medical image in a NoSQL database.

Advantages of moving Medical images to a NoSQL Database

- Simplifies managing images
- Portability and interoperability

- Enhanced security
- Efficient when move to the Cloud
- Handles huge data with ease.
- Faster querying with indexing.

### 3.4 DATA STORAGE IN THE CLOUD

The challenges faced in storing the DICOM images can be solved by moving the medical images to a Document based NoSQL database. In this paper an performance based study carried out to find a better document database to move to the cloud. The performances of two document databases were studied. The various performance metrics are considered and it is tested in a private Cloud. A framework to store the DICOM images on to the cloud is proposed by storing it to a Document based NoSQL database. This simplifies the process of moving medical images to the cloud.

#### 3.4.1 Document Databases

The problems posed by RDBMS database can be overcome by using a Document-Based database which are schema-less thus enabling us to store DICOM objects without stripping or compromising the data. A document-based database is designed to store metadata without modifying it. There are a few popular Document Databases, MongoDB,CouchBase and CouchDB. Two popular document databases MongoDB and CouchDB were considered in storing medical images. The better Document database will be considered for further study.

#### 3.4.2 Couch DB

Couch DB was developed by the Apache software foundation as an open-source document-based database management system. CouchDB is a Document database,can store **documents**. Every document has a unique name, and it provides a RESTful HTTP API for reading and updating (add, edit, delete) database documents.

It is suitable for the current needs of medical imaging practices, Medical Images as attachments Couch DB can associate attachments with documents and store them. Attachments can be referenced using the id used for storing it in the database.

What makes Couch DB special in comparison with other document-based database management systems is that the DICOM pixel data can be stored as an attachment thus facilitating the retrieval of image data along with DICOM metadata

#### 3.4.3 MongoDB

MongoDB, developed by 10gen, is a document-oriented NoSQL database that offers high performance and scalability. MongoDb provides horizontal scaling through auto-sharding. The data gets distributed automatically over a number of servers.[9][10].

There have been various research on the performance of MongoDB. Nyati et al. [21] compared the insertion/searching performance of MongoDB to MySQL in a single machine, showing that MongoDB outperformed MySQL.

In [12] the authors have done a time based performance study on few NoSQL databases including MongoDB and CouchDB for handling medical Images. MongoDB had a lesser I/O time comparatively. But few more performance based metrics need to be analysed to come to a conclusion that MongoDB is the best suited for storing Medical Images. In this paper MongoDB and CouchDB is been compared on various performance metrics to arrive at a conclusion.

#### 4. PERFORMANCE BASED COMPARISON BETWEEN MONGODB AND COUCHDB

As Medical Images are huge and it can be represented better in document databases, it is needed to further probe and find a better document database among the document databases. MongoDB is popular, but it is needed to analyze the performances of both the databases to fix up with one document database for further research and moving the same to the cloud. The various parameters analysed are

- No of Concurrent users vs Latency time
- Data Size vs Latency
- No of Cores vs Latency
- No of Users Vs Throughput
- Actual Data size vs Storage Size

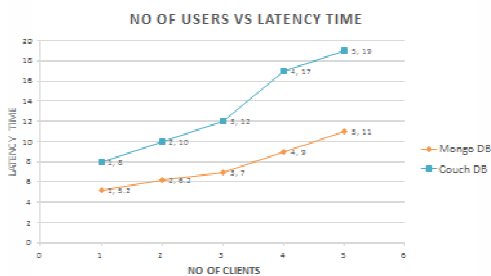
##### 4.1 Experimental Setup:

The Experimental setup was done on Ubuntu 16.04, MongoDB 3.4 version and CouchDB 1.6 was used.

##### 4.2 Experiments and the Results

The performance variations in Latency Time and throughput was checked for the increasing the number of concurrent users, it was found MongoDB performed better when the number if concurrent users was increased. It is inferred from Fig 1 that MongoDB has a lesser Latency time than CouchDB as the number of concurrent users were increased.

##### Performance – varying Number of Users

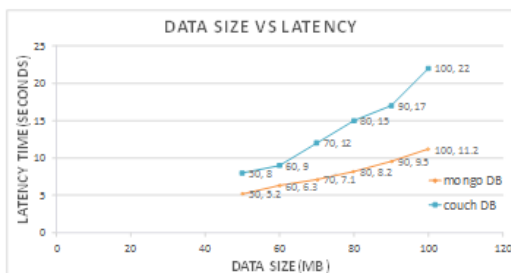


MongoDB performed better than CouchDB

Figure 1 –Performance with varying concurrent Users

The performance of MongoDB and CouchDB was analysed for different Data sizes. It was found that MongoDB performance for Larger files was better compared with CouchDB. Figure 2 shows the results.

##### Data size Vs Latency

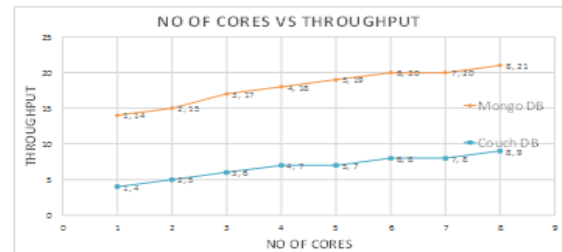


MongoDB maintains slower increase in Latency comparatively

Figure 2-Performance with different data sizes.

The analysis was done for varying number of Cores/CPU's for throughput(Number of Images retrieved per minute) a given Query, where MongoDB's throughput was better compared with CouchDB. It is inferred from Figure 3 the throughput of MongoDB was much better than couchDB.

##### Number of Cores Vs Throughput



MongoDB Showed a significant rise in Throughput Comparatively

Figure 3. Throughput with varying number of Cores.

The throughput was analysed for varying number of Users, where MongoDB's Throughput was comparatively constant as the number of concurrent users were increased. From Fig 4 it is inferred that the performance of MongoDB showed a better throughput with increase in number of concurrent users than CouchDB.

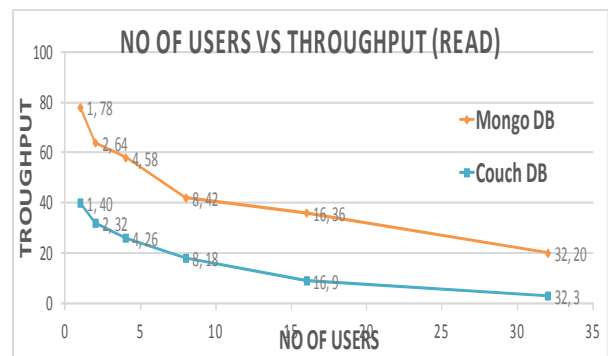
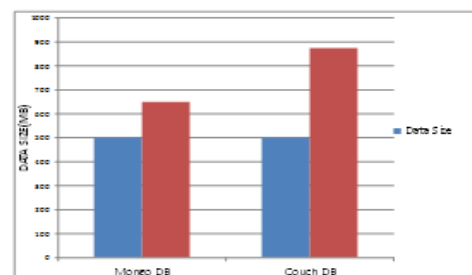


Figure 4. Throughput- Varying number of Users

##### Actual Data Size VS Storage Size



Couch DB Occupied More space than Mongo DB

Figure 5. Actual size Vs Storage size

The storage size was measured against the actual size, the amount of storage occupied for a given fixed datasize varies due to indexing and some meta data stored along with th data. CouchDB occupied more storage than MongoDB. Figure 5 indicates the difference between the Actual Data Size and Storage used for the data size.

## 5. RESULTS AND DISCUSSIONS

The various performance metrics studied clearly indicated that MongoDB outperformed CouchDB. These performance based study clearly indicates that MongoDB can be used to store Medical Images in the cloud. The same can be applied to any Big Data which is stored in the Cloud. Cloud Computing offers better benefits, and these benefits can be enhanced using a Cloud Database.

## 6. CONCLUSION & FUTURE WORK

As medical Images are moved to the Cloud, there is a tremendous need for a better infrastructure to handle them. The present study has proved that MongoDB is a better candidate to handling Huge Medical Images and Big Data in the cloud. Big Data handling in the Cloud will become simpler and performance wise better with MongoDB. In the search for a better Cloud Database, MongoDB can be considered. This Study can be further continued with MongoDB to find an Optimization method for storing Medical Images and Big Data in the Cloud. It is possible to arrive at a Performance framework based on MongoDB to handle Big Data in the Cloud.

## 7. ACKNOWLEDGMENTS

The Authors would like to thank Mr.Binu, Student of REVA University, Bangalore for his timely help.

## 8. REFERENCES

- [1] Steve G. Langer, Challenges for Data Storage in Medical Imaging Research, Volume 24, Issue 2, pp 203-207, April 2011, Journal of Digital Imaging.
- [2] Kagadis et al.: Cloud computing in medical imaging, 70901-1 Med. Phys. 40 (7), July 2013
- [3] Simón J. Rascovsky, Use of CouchDB for Document-based Storage of DICOM Objects, RadioGraphics 2012; 32:913–927 • Published online 10.1148/rg.
- [4] D. Revina Rebecca, Dr. I. Elizabeth Shanthi, Analysing the suitability of storing Medical Images in NoSQL Databases, ISSN 2229-5518, International Journal of Scientific & Engineering Research, Volume 7, Issue 8, August-2016.
- [5] Luís A. Bastião Silva, Louis Beroud, Carlos Costa and José Luis Oliveira, Medical imaging archiving: a comparison between several NoSQL, 978-1-4799-2131-7/14/\$31.00 ©2014 IEEE.
- [6] C. Strauch, U. L. S. Sites, and W. Kriha, NoSQL Databases, Lecture Notes. Stuttgart, Germany: Stuttgart Media Univ., 2011.
- [7] R. Cattell, “Scalable SQL and NoSQL data stores,” ACM SIGMOD Rec., vol. 39, no. 4, pp. 12–27, 2011.
- [8] N. Leavitt, “Will NoSQL databases live up to their promise?” Computer, vol. 43, no. 2, pp. 12–14, Feb. 2010.
- [9] R. Copeland, MongoDB Applied Design Patterns. Newton, MA, USA: O’Reilly Media, Inc., 2013.
- [10] K. Chodorow, MongoDB: The Definitive Guide. Newton, MA, USA: O’Reilly Media, Inc., 2013.
- [11] Yong-Shin Kang, et al, MongoDB-Based Repository Design for IoT-Generated RFID/Sensor Big Data
- [12] John Klein et.al, Application Specific NoSQL Databases, DOI:10.1109/BigDataCongress.2015.83, IEEE Explore
- [13] Bruno Guazzelli Batista et. Al, Performance Evaluation of Resource Management in Cloud Computing environments
- [14] Ercan, M. Z., Lane, M.: An evaluation of NoSQL databases for EHR systems. In Proceedings of the 25th Australasian Conference on Information Systems (2014)
- [15] A. Gandini et.al, Performance evaluation of NoSQL databases, DOI: 10.1007/978-3-319-10885-8\_2, Springer 2014
- [16] Gang Chen et.al, Federation in Cloud Data Management: Challenges and Opportunities, IEEE Explore, 2014
- [17] Daniel J. Abadi, Data Management in the Cloud: Limitations and opportunities, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering (2009).
- [18] M. Stone braker et.al. The end of an architectural era (it’s time for a complete rewrite). In VLDB, Vienna, Austria, 2007
- [19] S. S. Nyati, S. Pawar, and R. Ingle, “Performance evaluation of unstructured NoSQL data over distributed framework,” in Proc. IEEE ICACCI, Aug. 2013, pp. 1623–1627
- [20] Alexandros Antoniadis et.al, Tossing the NoSQL-Databases out to the public cloud, dl.acm.org/citation.cfm?id=2760038, 2014
- [21] Luís A. Bastião Silva, Louis Beroud, Carlos Costa and José Luis Oliveira, Medical imaging archiving: a comparison between several NoSQL, 978-1-4799-2131-7/14/\$31.00 ©2014 IEEE
- [22] S. S. Nyati, S. Pawar, and R. Ingle, “Performance evaluation of unstructured NoSQL data over distributed framework,” in Proc. IEEE ICACCI, Aug. 2013, pp. 1623–1627