Study And Analysis Of Different Cloud Storage Platform

S Aditi Apurva, Dept. Of Computer Science And Engineering,

KIIT University ,Bhubaneshwar, India.

---***---

Abstract - Cloud Storage is becoming the most sought after storage , be it music files, videos, photos or even general files people are switching over from storage on their local hard disks to storage in the cloud.

Google Cloud Storage offers developers and IT organizations durable and highly available object storage. **Cloud storage** is a model of data storage in which the digital data is stored in logical pools, the physical storage spans multiple servers (and often locations), and the physical environment is typically owned and managed by a hosting company.

Analysis of cloud storage can be problem specific such as for one kind of files like YouTube or generic files like google drive and can have different performances measurements .

Here is the analysis of cloud storage based on Google's paper on Google Drive,One Drive, Drobox, Big table , Facebooks Cassandra This will provide an overview of how the cloud storage works and the design principle .

Fig 1: Layers of cloud computing.

1.INTRODUCTION

The term cloud is the metaphor for internet. The network of servers and connection are collectively known as Cloud .Cloud computing emerges as a new computing paradigm that aims to provide reliable, customized and quality of service guaranteed computation environments for cloud users. Applications and databases are moved to the large centralized data centers, called cloud.
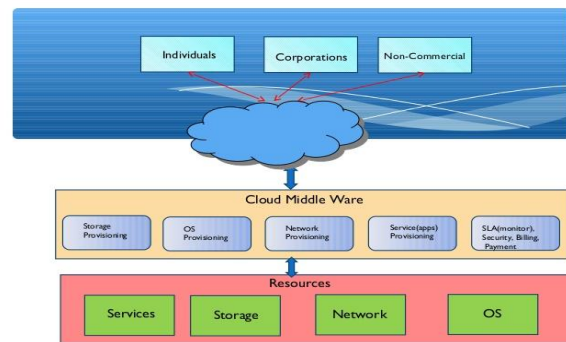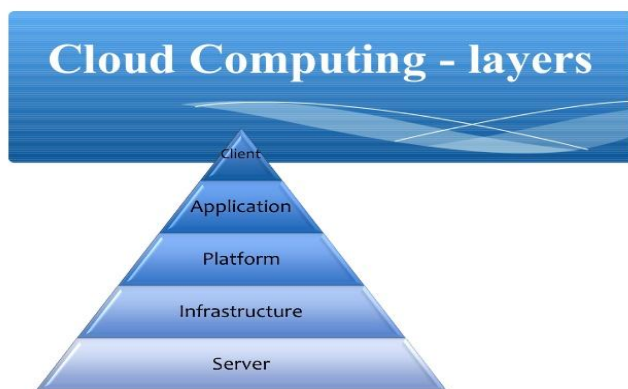

Fig 2: Cloud storage working.
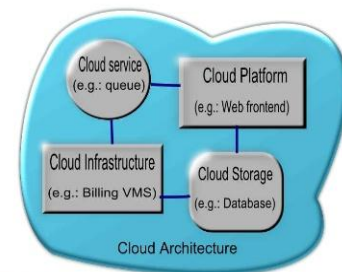
Fig 3: Architecture of cloud computing.

Why Cloud Computing

⇨ Hide complexity of IT infrastructure management
⇨ Massive configurability
⇨ Reliability
⇨ High Performance
⇨ Specifiable configurability
⇨ Low cost compared to dedicated infrastructure.

1.1 Big Table

Big table is a distributed storage system for managing structured data that is designed to scale to a very large size: petabytes of data across thousands of commodity servers. Many projects at Google store data in Big table, including web indexing, Google Earth, and Google Finance. These applications place very different demands on Big table, both in terms of data size (from URLs to web pages to satellite imagery) and latency requirements (from backend bulk processing to real-time data serving). Despite these varied demands, Big table has successfully provided a flexible, high-performance solution for all of these Google products. In this paper we describe the simple data model provided by Big table, which gives clients dynamic control over data layout and format, and we describe the design and implementation of Big table.

1.2 Purpose

Big table is built on several other pieces of Google infrastructure. Bigtable uses the distributed Google File System (GFS) to store log and data files. A Bigtable cluster typically operates in a shared pool of machines that run a wide variety of other distributed applications, and Big table processes often share the same machines with processes from other applications. Big table depends on a cluster management system for scheduling jobs, managing resources on shared machines, dealing with machine failures, and monitoring machine status.

1.3 DATA MODEL

A Bigtable is a sparse, distributed, persistent multi-dimensional sorted map. The map is indexed by a row key, column key, and a timestamp; each value in the map is an uninterested array of bytes.

1.3.1 ROW KEY:

The row keys in a table are arbitrary strings (currently up to 64KB in size, although 10-100 bytes is a typical size for most of our users). Every read or write of data under a single row key is atomic (regardless of the number of different columns being read or written in the row), a design decision that makes it easier for clients to reason about the system's behavior in the presence of concurrent updates to the same row.

1.3.2 COLOUM KEY

Column keys are grouped into sets called *column families*, which form the basic unit of access control. All data stored in a column family is usually of the same type. A column family must be created before data can be stored under any column key in that family; after a family has been created, any column key within the family can be used.

1.3.3 TIMESTAMP

Each cell in a Bigtable can contain multiple versions of the same data; these versions are indexed by timestamp. Bigtable timestamps are 64-bit integers. They can be assigned by Bigtable, in which case they represent "real time" in microseconds, or be explicitly assigned by client applications. Applications that need to avoid collisions must generate unique timestamps themselves. Different versions of a cell are stored in decreasing timestamp order, so that the most recent versions can be read first.

1.4 BIG TABLE API

The Big Table API provides various functions such as

➔Creating Tables

➔Deleting Tables

➔Column Families

→Changing Cluster

→Changing Table

→Column Family Meta Data
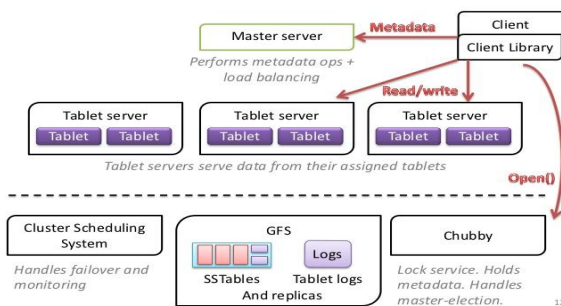
1.5 Architecture.

Fig 4: Bigtable system architecture.
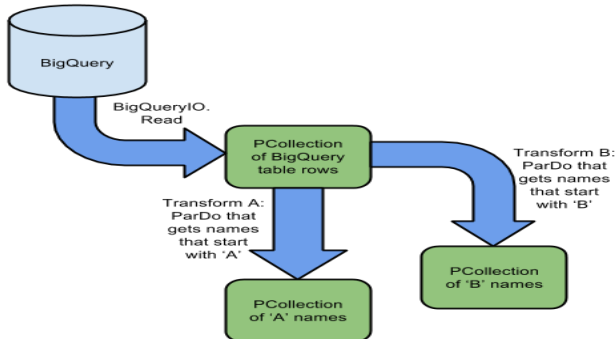
1.6 WORKING

Fig 5: Working of bigtable.

2. FACEBOOK CASSANDRA

Cassandra is a distributed storage system. It manages very large amounts of structured data spread out across many commodity servers, providing highly available service with no single point of failure. Aims to run on top of an infrastructure of hundreds of nodes. The way Cassandra manages the persistent state in the face of these failures drives the reliability and scalability of the software systems relying on this service. While in many ways Cassandra resembles a database and shares many design and implementation strategies therewith, Cassandra does not support a full relational data model; instead, it provides clients with a simple data model that supports dynamic control over data layout and format. Cassandra system was designed to run on cheap commodity hardware and handle high write throughput while not sacrificing read efficiency.
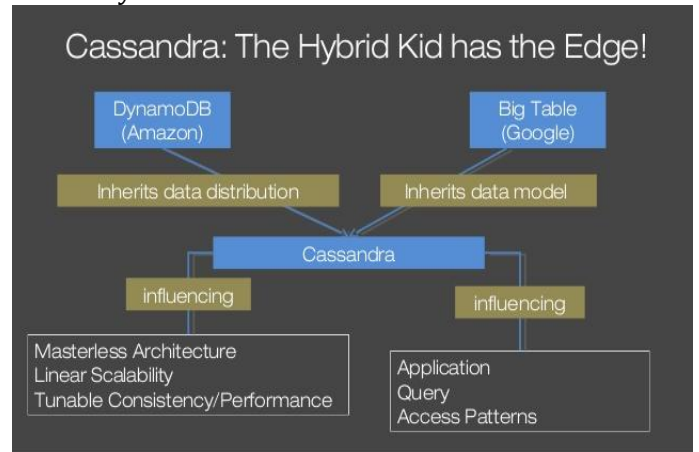

Fig 6: Casandra Outline

2.2 PURPOSE

First deployment of Cassandra system within Facebook was for the Inbox search system. The system currently stores TB's of indexes across a cluster of 600+ cores and 120+ TB of disk space. Performance of the system has been well within our SLA requirements and more applications are in the pipeline to use the Cassandra system as their storage engine.

2.3 DATA MODEL

A table in Cassandra is a distributed multi-dimensional map indexed by a key. The value is an object which is highly structured. The row key in a table is a string with no size restrictions, although typically 16 to 36 bytes long. Every operation under a single row key is atomic per replica no matter how many columns are being read or written into. Columns are grouped together into sets called column families very much similar to what happens in the Bigtable system. Cassandra exposes two kinds of columns families, Simple and Super column families. Super column families can be visualized as a column family within a column family

2.4 CASSANDRA API

The Cassandra API consists of the three simple methods.

• insert (table, key,rowMutation)

• get(table,key,columnName)

• delete(table,key,columnName)

2.5 ARCHITECTURE

The system needs to have the following characteristics; scalable and robust solutions for load balancing, membership and failure detection, failure recovery, replica synchronization, overload handling, state transfer, concurrency and job scheduling, request marshalling, request routing, system monitoring and alarming, and configuration management. Describing the details of each of the solutions is beyond the scope of this paper, so we will focus on the core distributed systems techniques used in Cassandra: partitioning, replication, membership, failure handling and scaling. All these modules work in synchrony to handle read/write requests.
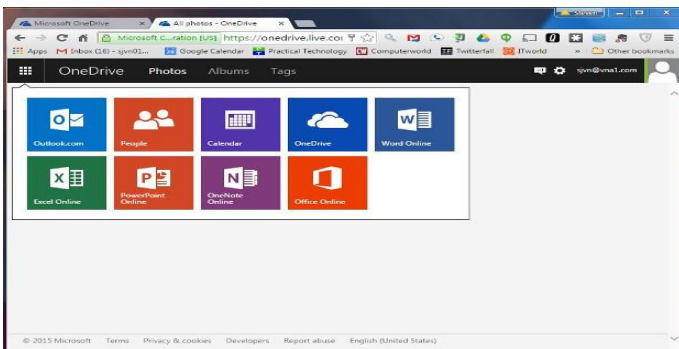
3. ONE DRIVE.

Fig 7: View Of One Drive.

3.1 INTRDUCTION

One Drive is the place where we store photos, and files .The company is working on technology that will eventually sort all of the photos taken based on how important and meaningful they are.

3.2 PURPOSE

• Works seamlessly with Windows devices because it's built in to the Windows operating system.

• It's easy to open and edit files from One Drive in Microsoft's other applications, such as Word or Excel.

• Signing up for One Drive gets you a Microsoft account, which gives you access to Outlook, Xbox Live, and other Microsoft services.

3.3 DRAWBACK

• It doesn't always put files in the correct folders.
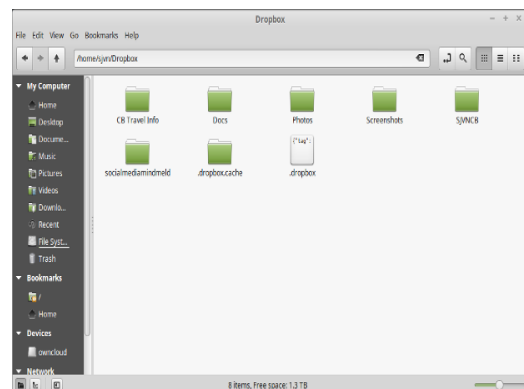
4. DROPBOX

Fig 8: View Of DropBox

4.1 INTRODUCTION

It is the favourite in the cloud storage in the world due to itsreliability, easy to use and breeze to set up. Your files live in the cloud and you can get to them at any time from Dropbox's website, desktop applications for Mac, Windows and Linux (Ubuntu, Debi an, Fedora or compile your own), or the iOS, Android, BlackBerry and Kindle Fire mobile apps.

4.2 PURPOSE

- Dropbox works equally well on PCs and Macs, Android and iOS.

- The service is so simple and elegantly designed, that it's easy for anyone to master. Its desktop applications seamlessly blend with your computer's file system.

4.3 DRAWBACK

No controls on how the files are displayed.
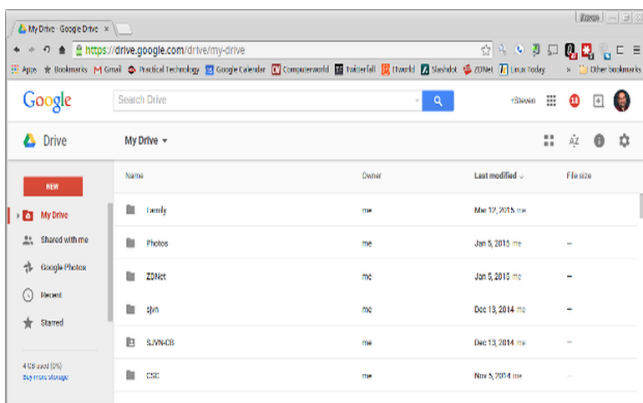
5. GOOGLE DRIVE

Fig 9: View Of Google Drive

5.1 INTRODUCTION

Google combines a complete set of office tools with cloud storage in Drive, a little bit of everything with this service, including a word processor, spreadsheet application, and presentation builder, plus 15GB of free storage space.

5.2 PURPOSE

15GB of storage for anything uploaded to Drive, including photos, videos, documents, Photoshop files and more. However, have to share that 15GB with your Gmail account, photos you upload to Google+, and any documents you create in Google Drive.

Table -1: Cloud storage comparison

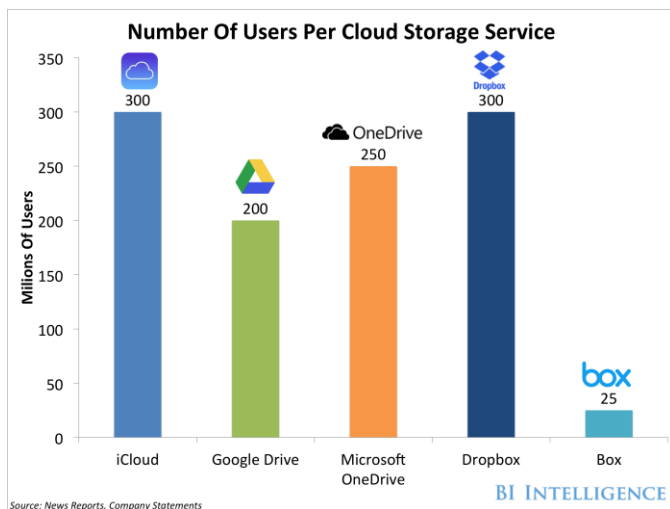|	One Dive	Drop Box	Google Drive
File Size Restriction	10GB	10GB with website, none with Drop box apps	**5TB**
Free Storage	5GB	2GB	15GB
Extra Free Storage	NO	YES	NO
OS Support	Windows, Mac, Android, iOS, Windows Phone	Windows, Mac, Android, iOS, Linux Windows Phone	Windows, Mac, Android, iOS, Windows Phone

Fig 10: Graph showing users per cloud storage Service.

6.CONCLUSION

Dropbox gives a free storage is only 2 GBs, but can be usedon any platform. We can get to files from Dropbox's web site, desktop applications for Mac, Windows and Linux, their native files systems and the iOS, Android, BlackBerry and Kindle Fire mobile apps. It's a snap to set up, and don't need to worry about syncing files for a second.

Google account gives 15GBs of free storage and an excellent office suite. It's good enough that many businesses and every Chromebook user is now using it as their complete cloud-based office.

OneDrive comes with 15GBs of free storage. On top of that, for each friend who signs into OneDrive as a new customer, both will receive an extra 0.5 GB of free storage up to a maximum of 5 GBs.

Big table is a distributed storage system for managing structured data that is designed to scale to a very large size: petabytes of data across thousands of commodity servers. Many projects at Google store data in Big table, including web indexing, Google Earth, and Google Finance.

Cassandra is a distributed storage system. It manages very large amounts of structured data spread out across many commodity servers, providing highly available service with no single point of failure.

REFERRENCES

1.Big table: A Distributed Storage System for Structured Data   Fay Chang, Jeffrey Dean, Sanjay Ghemawat, Wilson C. Hsieh, Deborah A. Wallach

Mike Burrows, Tushar Chandra, Andrew Fikes, Robert E. Gruber

2.Cassandra - A Decentralized Structured Storage System Avinash Lakshman Facebook andPrashant Malik Facebook Inside Dropbox: Understanding Personal Cloud Storage Services