Cloud Data Storage

Dhatrika Nagalakshmi
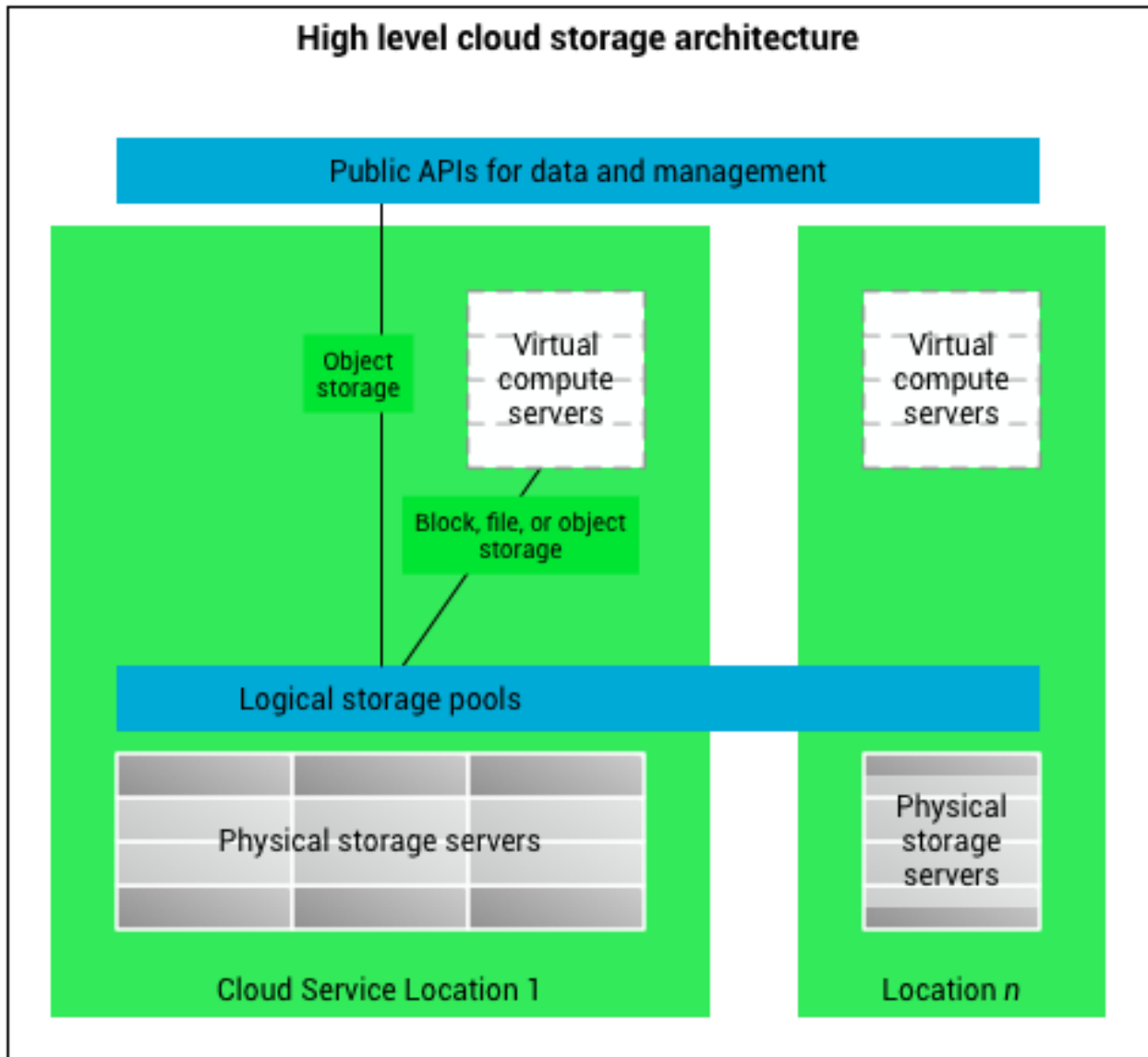**Illinois Institute of Technology**

ABSTRACT

Cloud storage is a model of data storage where 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. These cloud storage providers are responsible for keeping the data available and accessible, and the physical environment protected and running. People and organizations buy or lease storage capacity from the providers to store end user, organization, or application data.

INTRODUCTION

All giant companies like Google, amazon, Microsoft, apple, adobe etc. provide cloud services. Cloud services provided are either personalized or commercial. Examples for personalized cloud services are: Main example in this category would be google drive. Google drive enables us to upload our documents to a cloud, centralized data storage which can be accessed from anywhere, any device when you are logged in using your Gmail account. This is very useful as you don't have to keep copies of important documents on various devices or PCs and can have one place to store. Microsoft and Adobe are also offering similar services (Microsoft One-drive & Adobe Clear Cloud) when it comes to storing our personal information. Amazon also offers cloud services which can be used both personally as well as commercially. Amazon is a leading company when it comes to offering cloud services for commercial purposes.

Cloud storage is:

- Made up of many distributed resources, but still acts as one - often referred to as federated storage clouds.
- Highly fault tolerant through redundancy and distribution of data
- Highly durable through the creation of versioned copies
- Typically eventually consistent with regard to data replicas.

High level cloud storage architecture

Public APIs for data and management

Object storage

Virtual compute servers

Virtual compute servers

Block, file, or object storage

Logical storage pools

Physical storage servers

Physical storage servers

Cloud Service Location 1

Location *n*

Everything behind Cloud Storage

Cloud storage architecture comprises of a number of layers that work together to provide users with a seamless storage service. Generalized cloud storage architecture is shown:
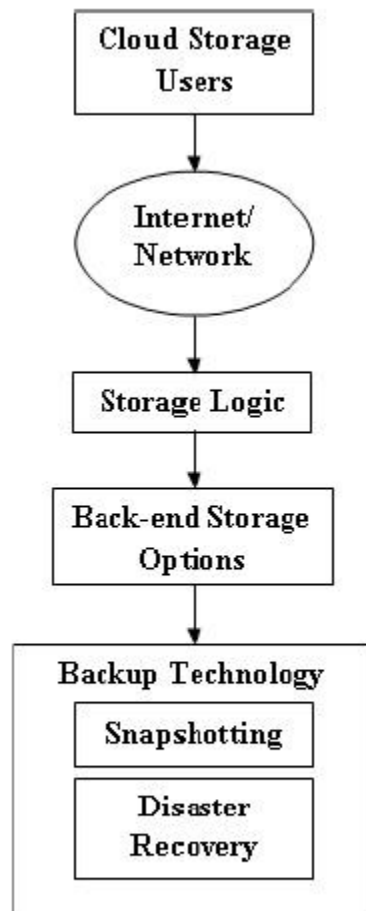
The different layers of cloud architecture are listed below:

Front End: This layer is exposed to end users and typically exposes APIs that allow access to the storage. A number of protocols are constantly being introduced to increase the supportability of cloud systems and include Web Service Front-ends using REST principles, file-based front ends. So for example, a user can use an App running on their desktop to perform basic functions such as creating folders, uploading and modifying files, as well as defining permissions and share data with other users.

Examples of Access methods and sample providers are listed below:

REST APIs: REST or Representational State Transfer is a stateless Web Architecture model that is built upon communications between clients and servers. Ex: Microsoft Windows Azure storage and Amazon Web Services Simple Storage Service (S3).

File-based Protocols: Protocols such as NFS and CIFS are supported by vendors like Nirvanix, Cleversafe and Zetta*.

Middleware: The middleware or Storage Logic layer supports a number of functions including data deduplication and reduction; as well as the placement and replication of data across geographical regions.

Back End: The back end layer is where the actual physical hardware is implemented and we refer to read and write instructions in the Hardware Abstraction Layer.

Additional Layers: Depending on the purpose of the technology, there may be a number of additional layers

Management Layer: this may support scripting and reporting capabilities to enhance automation and provisioning of storage.

Backup Layer: The cloud back end layer can be exposed directly to API calls from Snapshot and Backup services. For example Amazon's Elastic Block Store (EBS) service supports an incremental snapshot feature.

DR (Virtualization) Layer: DR Service providers can attach storage to a Virtual hypervisor, enabling cloud storage data to be accessed by Virtual Hosts that are activated in a DR scenario. For example the i365 cloud storage service automates the process of converting backups of server snapshots into a virtual DR environment in minutes.

Cloud & Database

A cloud database is a database that typically runs on a cloud computing platform, such as Amazon EC2, GoGrid, Salesforce, Rackspace, and Microsoft Azure. There are two common deployment models: users can run databases on the cloud independently, using a virtual machine image, or they can purchase access to a database service, maintained by a cloud database provider. Of the databases available on the cloud, some are SQL-based and some use a NoSQL data model.

The primary methods to run a database on the cloud are:

Virtual machine Image - cloud platforms allow users to purchase virtual machine instances for a limited time. It is possible to run a database on these virtual machines. Users can either upload their own machine image with a database installed on it, or use ready-made machine images that already include an optimized installation of a database. For example, Oracle provides a ready-made machine image with an installation of Oracle Database 11g Enterprise Edition on Amazon EC2 and on Microsoft Azure.

Database as a service (DBaaS) - some cloud platforms offer options for using a database as a service, without physically launching a virtual machine instance for the database. In this configuration, application owners do not have to install and maintain the database on their own. Instead, the database service provider takes responsibility for installing and maintaining the database, and application owners pay according to their usage. For example, Amazon Web Services provides three database services as part of its cloud offering, SimpleDB, a NoSQL key-value store, Amazon Relational Database Service, an SQL-based database service with a MySQL interface, and DynamoDB. Similarly, Microsoft offers the Azure SQL Database service as part of its cloud offering.

A third option is managed database hosting on the cloud, where the database is not offered as a service, but the cloud provider hosts the database and manages it on the application owner's behalf. For example, cloud provider Rackspace offers managed hosting for MySQL on dedicated and cloud architectures and NoSQL databases via Object Rocket's managed MongoDB service.

We can differentiate between cloud databases which are relational as opposed to non-relational or NoSQL:

SQL database, such as NuoDB, Oracle Database, Microsoft SQL Server, and MySQL, are one type of database which can be run on the cloud (either as a Virtual Machine Image or as a service, depending on the vendor). SQL databases are difficult to scale, meaning they are not natively suited to a cloud environment, although cloud database services based on SQL are attempting to address this challenge. **NoSQL databases**, such as Apache Cassandra, CouchDB and MongoDB, are another type of database which can run on the cloud. NoSQL databases are built to service heavy read/write loads and are able scale up and down easily, and therefore they are more natively suited to running on the cloud. However, most contemporary applications are built around an SQL data model, so working with NoSQL databases often requires a complete rewrite of application code.

Cases:

Amazon web services Simple Storage Service (S3)

Amazon Simple Storage Service (Amazon S3) is storage for the Internet. Amazon S3 is used to store and retrieve any amount of data at any time, from anywhere on the web. These tasks can be accomplished using the simple and intuitive web interface of the AWS Management Console.

AWS can be used in many ways:

- Backup and Storage – Provide data backup and storage services for others.
- Application Hosting – Provide services that deploy, install, and manage web applications.
- Media Hosting – Build a redundant, scalable, and highly available infrastructure that hosts video, photo, or music uploads and downloads.
- Software Delivery – Host your software applications that customers can download.
- Archiving - Documents can be archived on the cloud for very less cost. Encryption services are also available to keep the data safe.
- High performance computing: AWS can increase the speed of research by running high performance computing in the cloud by providing instant access to Cluster Compute, or Cluster GPU, servers, a no up-front cost payment model, and a global infrastructure to use.

Amazon Relational Database service (Amazon RDS)

Amazon Relational Database Service or Amazon RDS is a distributed relational database service by Amazon.com. It is a web service running "in the cloud" and provides a relational database for use in applications. It is aimed at simplifying the setup, operation, and scaling a relational database. Complex administration processes like patching the database software, backing up databases and enabling point-in-time recovery are managed automatically. Scaling storage and compute resources can be performed by a single API call. Amazon RDS was first released on 22 October 2009 supporting MySQL databases. This was followed by Oracle Database in June 2011, Microsoft SQL Server in May 2012 and PostgreSQL in November 2013.

Google cloud datastore (commercial)

Google cloud datastore enables clients to Use a managed, NoSQL, schemaless database for storing non-relational data. Cloud Datastore automatically scales as you need it and supports transactions as well as robust, SQL-like queries.
Its features are:

1. Schemaless access: with SQL-like querying no need to worry about data models and migration. Cloud datastore is a schemaless storage service that allows you to be agile by removing the need to think about the underlying structure of the data. Cloud datastore provides a robust query engine that allows you to search for data across multiple properties and sort as needed.

2. Managed database: Cloud Datastore is fully managed. Google automatically handles sharding and replication in order to provide you with a highly available and consistent database.

3. Autoscale with your users: Cloud Datastore automatically scales depending on your needs. This allows you to focus on building your application and not on worrying about provisioning and load anticipation.

4. ACID transactions: Cloud Datastore provides ACID transactions using optimistic concurrency control. Your application can execute multiple datastore operations in a single transaction in which either all succeed or all fail, ensuring the integrity of your data.

5. Built-in redundancy: With a single API call, data is automatically replicated across multiple data centers. High availability and durability are built in from the very core.

6. Access your data from anywhere: Build solutions that span App Engine and Compute Engine, and rely on Cloud Datastore as the integration point. With the RESTful interface that is exposed by Cloud datastore, data can easily be accessed by any deployment target.

References

1. http://en.wikipedia.org/wiki/Cloud_storage
2. http://aws.amazon.com/documentation/s3/
3. http://en.wikipedia.org/wiki/Cloud_computing_security
4. http://aws.amazon.com/documentation/
5. http://en.wikipedia.org/wiki/Amazon_S3
6. https://cloud.google.com/storage/
7. https://cloud.google.com/storage/docs/overview?csw=1
8. http://en.wikipedia.org/wiki/Cloud_database
9. http://news.microsoft.com/2014/02/19/microsoft-onedrive-launches-worldwide-free-cloud-storage-service-provides-one-place-for-everything-in-your-life/
10. http://cloudcomputing.sys-con.com/node/770174
11. http://theroadchimp.com/2013/03/17/cloud-architectures-storage-in-the-cloud/
12. http://aws.amazon.com/documentation/
13. http://en.wikipedia.org/wiki/Amazon_Relational_Database_Service
14. http://www.atmosonline.com/Blog/?Tag=cloud%20storage%20architecture
15. http://aws.amazon.com/what-is-cloud-computing/