Ceph

* Storage technology
* It’s open source software
* It is a reliable, easy to manage, object store that provides storage of unstructured data for application.
  + RADOS(Reliable Autonomic Distributed Object Store )
    - Ceph is built on top of it
    - It’s reliable, autonomous, distributed object store consist of self healing, self-managing, intelligent storage nodes.
* Ceph components:
  + Client
  + Metadata
  + Object storage

Hive

* Data warehousing package(storing data which is generated from different sources in one place) built on top of Hadoop
* It is used for data analysis
* It is based on SQL
  + History
    - In facebook, data was collected by using cron Jobs(means scheduled jobs) to Oracle database
    - They have to do lots of coding in Java, thus they designed Hive
  + Uses
    - Data mining(find the pattern of buying)
    - Document indexing
  + Components
    - Shell
    - Driver
    - Compiler
    - Execution engine
    - Metastore

Blob storage

* Unstructured DB
* It’s a services hosted inside the cloud
* Three types
  + Block
    - Text
    - Binary Files
    - Document
    - Media Files
  + Append
    - Optimized for append operation
    - Useful for logging scenarios
  + Page
    - Efficient I/O operation
    - Used as Data disk and OS
* Examples
  + Windows Azure Blob storage
  + Ambry LinkedIn

TEZ

* Tez is an extensible framework for building high performance batch and interactive data processing applications, coordinated by YARN in Apache Hadoop. Tez improves the MapReduce paradigm by dramatically improving its speed, while maintaining Map Reduce ability to scale to petabytes of data.
* It has a flexible input processor output runtime mode. It has a data type diagnostic that allows to be able to treat all the data types in a similar way. It has an easy to deploy simplifications so that actually, allocating resources to this is made easier.

SWIFT

* The OpenStack Object Store project, known as Swift, offers cloud storage software so that one can store and retrieve lots of data with a simple API.
* It's built for scale and optimized for durability, availability, and concurrency across the entire data set.
* It is ideal for storing unstructured data that can grow without bound.
* It is designed to manage the storage of large amounts of data cost-effectively on a long-term basis across clusters of standard server hardware.

Amazon S3 BLOB Storage

* Amazon Simple Storage Service (Amazon S3) is an object storage service that offers industry-leading scalability, data availability, security, and performance.
* It used to store and protect any amount of data for a range of use cases, such as websites, mobile applications, backup and restore, archive, enterprise applications, IoT devices, and big data analytics.
* Amazon S3 provides easy-to-use management features so you can organize your data and configure finely-tuned access controls to meet your specific business, organizational, and compliance requirements.

Amazon AWS EBS Storage

* Amazon Elastic Block Store (Amazon EBS) provides persistent block storage volumes for use with Amazon EC2 instances in the AWS Cloud.
* Each Amazon EBS volume is automatically replicated within its Availability Zone to protect one from component failure, offering high availability and durability.
* Amazon EBS volumes offer the consistent and low-latency performance needed to run the workloads.

AWS Glacier

* Amazon Glacier is an online file storage web service that provides storage for data archiving and backup.
* Glacier is part of the Amazon Web Services suite of cloud computing services, and is designed for long-term storage of data that is infrequently accessed and for which retrieval latency times of 3 to 5 hours are acceptable.
* Storage costs are a consistent $0.004 per gigabyte per month, which is substantially cheaper than Amazon's own Simple Storage Service (S3).

Amazon AWS Elastic File System

* Amazon Elastic File System (Amazon EFS) provides a simple, scalable, elastic file system for Linux-based workloads for use with AWS Cloud services and on-premises resources.
* It is built to scale on demand to petabytes without disrupting applications, growing and shrinking automatically as one add and remove files, so it applications have the storage they need – when they need it.

Dropbox API

* Dropbox itself keeps metadata of files in its own servers, which is probably small amounts of information compared to files themselves.
* Dropbox itself keeps metadata in its own servers, it stores the files on Amazon S3.
* managers their own metadata.