

Photo Storage

Facebook classifies photos into three categories, 'hot', 'warm' and 'cold' photos, and uses different mechanisms to process these images:

- **Hot: Popular, a lot of views (approx. 90% of views)**

CDN (Content Distribution Network)

- **Warm: Somewhat popular, but still a lot of views in aggregate**

Haystack (Facebook has designed its own storage called Haystack)

- **Cold: Unpopular, occasional views**

f4 (It is an “archival” storage designed by Facebook)

CDN

- **What is CDN**

1. CDN is a content delivery network;
2. A CDN is a cache, not a permanent store;
3. Content providers are CDN customers;

- **How does the CDN work?**

- CDN company (e.g., Akamai) installs thousands of servers throughout Internet (In large datacenters close to users)
- CDN replicates customers' content
- When provider updates content, CDN updates servers

- **Pros and cons of CDN**

- Pros: Very good performance
- Cons: no reliability guarantee

Haystack

- **What is Haystack**

- Designed for performance and reliability
- "Default" photo storage

- **Pros and cons of Haystack**

- Pros: Good throughput and reliability;
- Cons: Somewhat inefficient use of storage space (mainly due to replication).

- **Haystack Directory**

- Helps the URL construction for an image
- Logical & physical volumes

- **Haystack Cache & Store**

- Haystack cache

f4

- **f4's Replication**

- (n, k) Reed-Solomon code
- Parity example: XOR

- **f4: Cross-Datacenter**

- Additional parity block
- Overall average space usage per block:
2.1X
- With 2.1X space usage,

- **f4: Single Datacenter**

- Within a single data center, $(14, 10)$ Reed-Solomon code
- Distribute blocks across different racks

Content Storage

- **Architecture**

- **To served by the follwing steps**

1. Dedicated webservers,
2. Scribe–Hadoop Clusters
3. Hive–Hadoop
4. Mysql

- **Distributed systems components**

1. Two main components - Hadoop
 - Map–Reduce
 - Hadoop Distributed File System (HDFS)
2. Master nodes - Hadoop

Hadoop consists of multiple master nodes to avoid single point of failure in any environment.

3. The elements of master node

- Job Tracker
- Task tracker
- Name node (NN)
- Data Node (DN)
- Worker Nodes

- **Map Reduce (M-R)**

1. Index any data comes from HDFS and being divided into blocks.
2. Submit the M-R Job and its details to the Job tracker.
3. Mapper process data blocks and generates a list of key value pairs.
4. M-R merge list of key value pairs to generate final results.

- **HDFS**

5. Run on low-cost hardware
6. Highly fault-tolerance (as it supports block replication)
7. Store very large data sets
8. Reliability
9. High bandwidth
10. Ability to dynamically scale

Hadoop and Hive

In Facebook Hive is a data warehouse infrastructure built on top of Hadoop technology.

Role

1. **Easy data summarization;**
2. **Heavily reporting;**
3. **Adhoc querying;**
4. **Analysis of large datasets data stored;**
5. **HiveQL.**

A simple query language called HiveQL which is based on SQL and which enables users familiar with SQL to query this data.

Apache HBase

Facebook messaging system by the support of Apache HBase which is a database-like layer built on Hadoop designed to support billions of messages per day.

Memcached servers

Facebook, let Hadoop performing a random access workloads that provides low latency access to HDFS, by using a combination of large clusters of MySQL databases and caching tiers built using memcached ,that will be support a better in performance while all results from Hadoop are directed to MySQL or memcached for consumption by the web tier side.