**Presentation**

* Good morning, In this presentation, I discuss and compare different data management and system optimization techniques.
* This is the outline, first, we will discuss each paper in detail, and finally, I will compare them all.

**Paper 1**

* While storing small files the storage gets wasted because of data holes and high metadata content.
* To solve this problem the paper proposes a method, that says instead of storing the files individually we can directly merge data files during IO operation.
* To merge the files we use the CRUSH algorithm which improves the IO performance.
* To avoid data holes while deleting the small files, there is a garbage collection method used.

**Paper 2**

* While designing software one of the most important aspects is flexibility and to maintain that we can use microservice architecture.
* If there are more services data consistency can be compromised which could be a problem.
* So to improve the response time and handle large datasets the paper proposes a solution.
* Instead of using the same database for all services, we can use different types of databases based on service requirements.
* There are databases like SQL, NoSQL, or hybrid can be used based on the needs and priorities of the service.

**Paper 3**

* This paper develops an application that manages short video content efficiently storing data in an SQL database and maintaining system security.
* This application uses technologies like HTML, CSS, and JS for the front end and Springboot framework which is based on Java for the backend.

**Paper 4**

* This paper gives the methods to keep the data more secure in the cloud.
* We can use encryption but it has a drawback that if the encryption key is leaked then the data can get compromised.
* So we can use random pattern fragmentation techniques.
* According to this we divide the data into small chunks, shuffle them, and store them in different nodes.
* We also store the metadata containing important information in the split file which will help while fetching the data and reconstructing it again.

**Paper 5**

* This paper examines the fundamental trafe-offs in fault-tolerant distributed systems and replication databases.
* According to the CAP theorem, only 2 out of availability, consistency, and partition tolerance can be preserved at a time.
* There are sub-systems like AP system which compromises consistency and the CP system compromises availability.
* Different trade-offs can be used in different applications based on priorities for example in large systems like Facebook availability and latency are chosen over consistency.

**Paper 6**

* This paper tackles challenges in distributed database query processing, aiming to reduce communication overhead and response time.
* It introduces the parallel CART decision tree method, which uses vertical data segmentation to minimize communication costs.
* Additionally, it utilizes the Fayyad algorithm to improve decision tree learning and edge point classification.
* These techniques collectively enhance performance and efficiency in distributed database systems.

**Comparison**

First, let’s compare based on speed

* The speed of papers 2 and 5 methods depends on the database.
* In 3rd paper, the startup time is reduced but the scalability is not guaranteed.
* Papers 4 and 6 use parallel processing which improves the speed and reduces the communication overhead.

Talking about the accuracy

* 1st, 4th, and 6th paper has an increase in accuracy when compared to the remaining papers.

Next is space utilization

* Space utilization in papers 1 and 4 has improved but in paper 2 there is less space efficiency because it uses a relational database model like SQL

Last is the Performance and resource utilization

* In paper 1 the write performance has increased and the read performance is not guaranteed.
* In the 3rd, 4th, 5th, and 6th papers also the performance has improved due to various reasons.