

Paper Title:**Design of Distributed Network Mass Data Processing System Based on Cloud Computing Technology****Paper Link:**

<https://ieeexplore.ieee.org/document/9452963>

1 - Summary:**1.1 Motivation:**

This paper delves into distributed data processing systems based on cloud computing technology. With the exponential growth of data that needs to be processed, distributed computing, cloud computing, and other technologies have gained prominence. Cloud computing security has become a focal point for researchers due to the increasing emphasis on information security and personal privacy.

1.2 Contribution: Enhanced data processing capabilities: Cloud computing offers a scalable and elastic infrastructure that can handle massive amounts of data efficiently. Distributed data processing systems can leverage this infrastructure to process large datasets in a timely manner, enabling real-time analytics and decision-making.

1.3 Methodology: The proposed methodology is divided into three divisions: Mapreduce of iterative computing, design of Distributed Task Scheduling System and Distributed collaborative Computing computing mechanism. The author describes in detail about the taken approach and suggests to opt for a hybrid structure that combines centralize and decentralize architectures.

1.4 Conclusion: This study proposes a cloud computing security model that combines homomorphic encryption protocol with MapReduce distributed computing framework to achieve privacy protection in cloud computing. The model has been tested and proven to be efficient and secure to some extent.

2 - Limitations:

2.1 First Limitation: Managing and maintaining distributed systems can be complex, requiring expertise in distributed computing, cloud technologies, and data management. The overhead associated with coordinating data distribution and processing across multiple nodes can also impact performance.

3 - Synthesis: Distributed data processing systems based on cloud computing technology have emerged as a powerful approach for managing and processing massive amounts of data. These systems leverage the scalability, elasticity, and cost-effectiveness of cloud computing infrastructure to handle large datasets efficiently and enable real-time data analytics. By distributing data processing tasks across multiple nodes, these systems can significantly improve performance and reduce latency.