Cloud-Based Batch Processing with Job Scheduling

Kean Main Campus
Union, N.J
NEW YORK CITY

W05 GROUP5
Members:
Dong Shuaiwu
Jiao Luyao
Hou Dianzhe
Advisor:
Djigal Hamza

1 OVERVIEW

This project focuses on developing a cloud-based batch processing system to efficiently schedule and process large datasets using algorithms like FCFS, SJF, and Priority Scheduling. Key objectives include comparing scheduling strategies based on job completion time and resource utilization. This exploration provides insights into optimizing batch systems for real-world applications, finding effective scheduling in cloud computing.

What is Batch Process?

Batch Processing is a computing mode that is mainly used to process a large number of tasks or data at the same time, usually in the background in a non-interactive manner. This mode is very useful when you need to process large-scale data or run tasks that take a long time.

2 INTRODUCTION

Efficient batch processing of large datasets is essential in today's data-driven world. Cloud-based systems, such as those on Alibaba Cloud, offer scalable solutions, but their effectiveness heavily depends on the choice of job scheduling algorithms. These algorithms dictate the order and allocation of tasks to resources, directly impacting processing time, resource usage, and cost. By simulating and optimizing these algorithms, we can identify strategies that save time and reduce operational costs, demonstrating how effective scheduling is crucial for maximizing cloud resources' potential and ensuring cost-efficiency in large-scale data processing.

4 Methodology

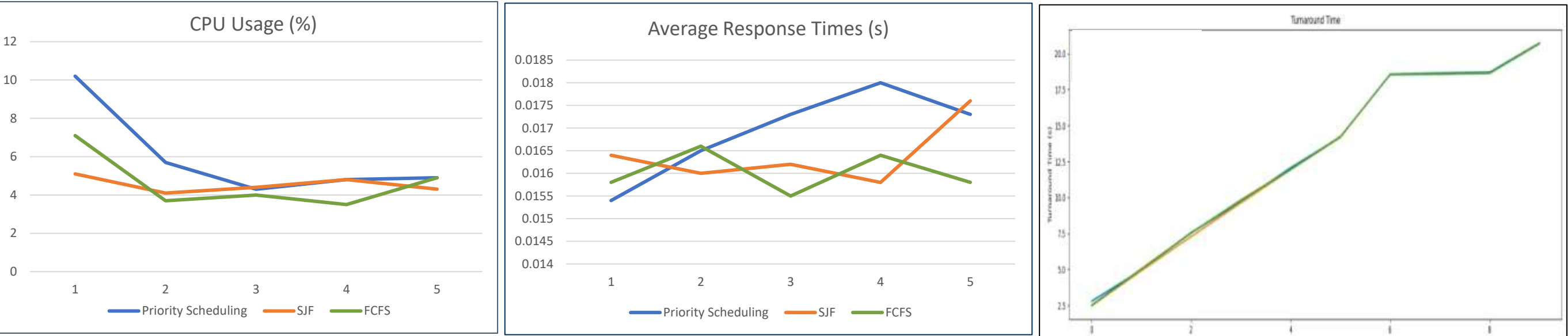| Our Work flow | Simulate the generation of task queues | Implement three task scheduling algorithms: FCFS, SJF, and Priority Scheduling. | Submit tasks to the intranet server. | Monitor task status and record results. |

3 The Algorithm we used

Process	Arrival time	Execution time	Priority
P1	0	5	2
P2	2	4	1
P3	3	7	3
P4	5	6	4

1. FCFS (First-Come, First-Served)
   Just consider arrival time, less arrival time first

P1	P2	P3	P4
0	5	9	16	22

2. SJF (Shortest Job First)
   Just consider execution time, less execution time first

P1	P2	P4	P3
0	5	9	15	22

3. Priority Scheduling
   FCFS, but preemptive when higher priority come

P1	P2	P1	P3	P4
0	2	6	9	16	22

5 Results

CPU Usage (%)

Average Response Times (s)

Turnaround Time

Priority Scheduling — SJF — FCFS

6 Conclusion

Based on data, algorithms impact resource usage and task efficiency in batch processing systems.

①Priority Scheduling utilizes the CPU most effectively, achieving higher utilization but slightly longer response times.

②SJF minimizes average response times due to its focus on shorter tasks, making it optimal for time-sensitive operations with predictable task lengths.

③FCFS maintains fairness and simplicity, offering balanced performance across metrics.

7 Reference

[1] Alworafi, M. A., Dhari, A., Al-Hashmi, A. A., & Darem, A. B. (2016, December). An improved SJF scheduling algorithm in cloud computing environment. In *2016 International Conference on Electrical, Electronics, Communication, Computer and Optimization Techniques (ICEECCOT)* (pp. 208-212). IEEE.
[2] Pandit, R., & Dwivedi, M. (2022). Resources Load Sharing using RR, FCFS, SJF, MSJF, GP Algorithms in Cloud Computing. *NeuroQuantology*, 20(11), 7853.
[3] Arunarani, A. R., Manjula, D., & Sugumaran, V. (2019). Task scheduling techniques in cloud computing: A literature survey. *Future Generation Computer Systems*, 91, 407-415.