

List of Figures

Figure 2.1 Big Data Characteristics.....	10
Figure 2.2 Distributed Computing System Architecture	16
Figure 2.3 Heterogeneity of Nodes	17
Figure 2.4 (A) Vertical Scaling (B) Horizontal Scaling.....	21
Figure 2.5 Hierarchical classification of scheduling algorithms.....	22
Figure 3.1 Hadoop Architecture.....	26
Figure 3.2 HDFS Architecture	28
Figure 3.3 YARN Architecture.....	30
Figure 3.4 MapReduce Model.....	31
Figure 3.5 HDFS Block Placement Policy.....	32
Figure 3.6 FIFO Scheduler	34
Figure 3.7 Capacity Scheduler.....	35
Figure 3.8 Fair Scheduler.....	36
Figure 3.9 Node Label Configuration for Queues	38
Figure 3.10 Application Submission to Queues	39
Figure 4.1 Objective-based Classification of Scheduling Algorithms.....	45
Figure 4.2 (A) Average Total Time (B) Average Turnaround Time (Single Queue) ...	56
Figure 4.3 CPU & Memory Usage (Single Queue)	57
Figure 4.4 (A) Average Total Time (B) Average Turnaround Time (Multi-Queue)	58
Figure 4.5 CPU & Memory Usage (Multi-Queue)	58
Figure 4.6 (A) Average Total Time (B) Average Turnaround Time (Mixed-Multi Queue).....	59
Figure 4.7 CPU & Memory Usage (Mixed-Multi Queue).....	59
Figure 5.1 HDFS Default Block Placement Example.....	66
Figure 5.2 “Saksham” Approach: Group-1: Priority=2; Group-2: Priority=1	67
Figure 5.3 Proposed “Saksham” Algorithm	69
Figure 5.4 Flow Diagram of “Saksham” algorithm.....	70
Figure 5.5 Proposed “Saksham” Model for Load Balancing	72
Figure 6.1 Test Cycle for All Scenarios	77

Figure 6.2 Saksham Balancing: Text Dataset (A) Size-15 GB (B) Size- 30 GB	81
Figure 6.3 Saksham Balancing: TeraGen Dataset (A) Size- 15 GB (B) Size-30 GB.....	82
Figure 6.4 Saksham Block Rearrangement Time	83
Figure 6.5 Comparison of Data Locality Achieved	85
Figure 6.6 Job Execution Time using Capacity Scheduler	85
Figure 6.7 Job Execution Time using Fair-FIFO Scheduler	86
Figure 6.8 Job Execution Time using Fair-Fair Scheduler.....	86
Figure 6.9 Job Execution Time using Fair-DRF Scheduler	86
Figure 6.10 Job Execution Time of all Scheduling Policies	87
Figure 6.11 Total time (Execution Time + Rearrangement Time) using Fair-Fair Policy	88
Figure 6.12 Average CPU Utilization	88
Figure 6.13 Average Memory Utilization.....	89
Figure 6.14 Data Locality for DNA Pattern Search Dataset	91
Figure 6.15 Job Execution Time for DNA Pattern Search.....	91
Figure 6.16 Total time (Execution Time + Rearrangement Time) for All Policies	92