Team 10: Sharan Vaitheeswaran, Sajan Gurung, Shristi Srivastava

Professor Munehiro Fukuda

CSS 534: Parallel Programming in Grid and Cloud

10 December 2023

Parallelizing String Matching with KMP Algorithm

The purpose of this project is to parallelize string matching applications with mpiJava, MapReduce, Spark and MASS and to analyze programmability and execution performance.

Application Overview

The Knuth-Morris-Pratt (KMP) algorithm serves as a sophisticated tool for string matching, adept at efficiently locating instances of a designated pattern within an extensive body of text. Distinguished by its capacity to circumvent needless character comparisons, this algorithm leverages intrinsic pattern information to streamline the search process.

Our application seamlessly integrates the KMP algorithm, requiring user-specified input of a pattern and an input folder. The outcome is a meticulously curated output.txt file, housing filenames alongside their corresponding indexes for the identified pattern within each respective file. The implementation of the KMP algorithm involves the meticulous computation of a Longest Proper Prefix Suffix (LPS) array and the strategic deployment of two index-pointing integers for pattern traversal.

In stark contrast to the brute force approach to string search, the KMP algorithm ensures that each character in the larger string is accessed only once, imbuing it with exceptional efficiency. The temporal complexity of pattern search via this algorithm is a commendable O(n + m), where 'n' denotes the length of the larger text and 'm' signifies the length of the pattern.

Parallelization with mpiJava

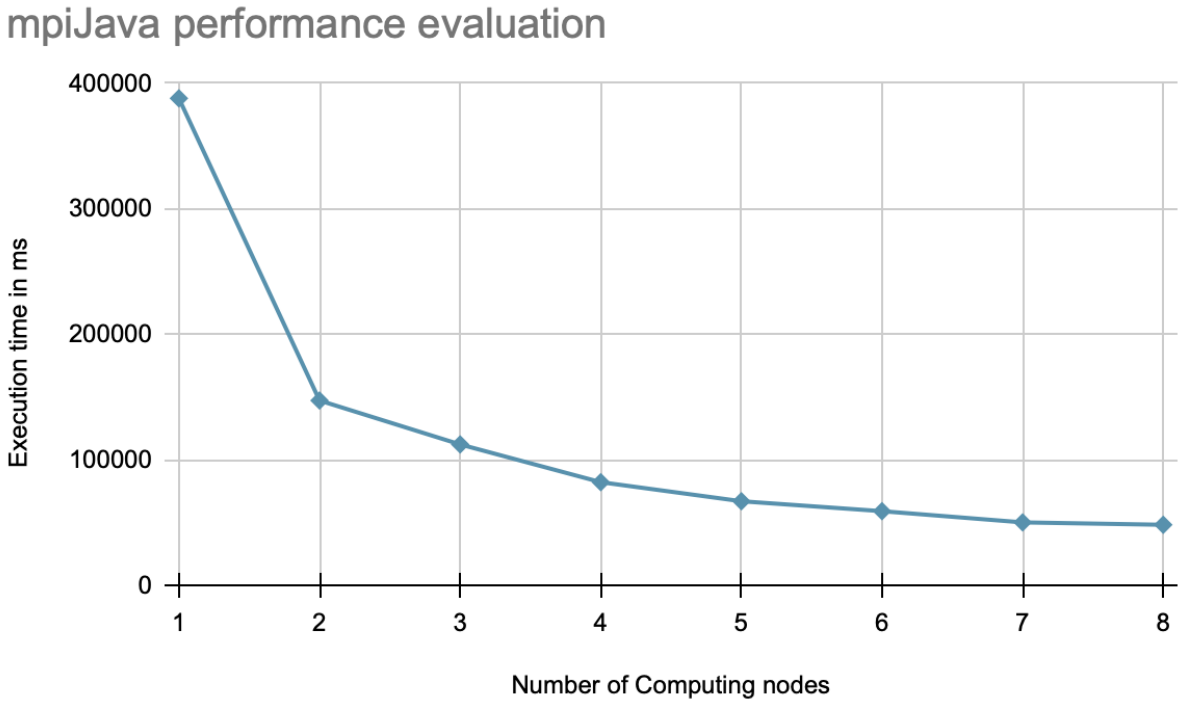
The MpiJava implementation of the program employs a divide-and-conquer strategy to distribute parallel tasks among processing nodes, each responsible for handling files within a specified range. After the master node initializes MPI, every node retrieves a list of filenames from the input folder and generates a sorted array of these filenames. Subsequently, each node utilizes its rank to compute the start and end indices for processing. For instance, in a scenario with 10 files and 5 computing nodes, rank 1 will process files from index 0 to 1, rank 2 from index 2 to 3, and so forth. In cases where the number of files isn't evenly distributed among nodes, the remainder is allocated, with one file assigned to each node starting from node 1.

Simultaneously, each node computes the Longest Prefix Suffix (LPS) array for the given pattern. Processing begins by accessing elements in the filenames array, starting from the respective start index. The node processes each file using the LPS array, recording the results—filename and found indexes—in an intermediate file specific to that node. As each node writes independently to its designated file, there is no contention for file access during the concurrent writing process. The intermediate file is continuously updated with filenames and occurrence indexes as each node processes its assigned files.

Moreover, MPI.COMM_WORLD.Barrier is employed to synchronize all computing nodes after completing the processing of their designated files. A local variable, localProcessingStatus, is utilized to monitor the processing status at each node. This integer variable can assume one of three possible values: -1 indicates no processing occurred, 0 signifies processing has commenced, and 1 denotes that processing is complete. MPI.COMM_WORLD.AllGather is then used to communicate these statuses to all nodes. The master node leverages the gathered statuses to combine the individual intermediate files into a consolidated output file containing all the indexes for the searched pattern.

Results of parallelizing using mpiJava. A folder with 3,784 files(228.5 MB) was used for evaluation.

Node Count	**Elapsed Time(ms)**	**Perf. Improvement**
1	387991	1
2	147201	2.64
3	112201	3.46
4	82115	4.72
5	66927	5.80
6	59008	6.58
7	50072	7.75
8	48163	8.06



From the above graph we can see that the performance improved drastically with 2 nodes but after that increase is gradual. We think this is because all nodes are connected to the network file system and as the number of nodes increases, performing IO over the same network is slower.

Parallelization with MapReduce

The code comprises two principal classes: KMPMapper and KMPReducer, designed to handle mapping and reducing tasks within the Hadoop MapReduce paradigm.

The KMPMapper class, inheriting from MapReduceBase and implementing the Mapper interface, forms the core of the algorithm. Its role involves reading input text files, performing

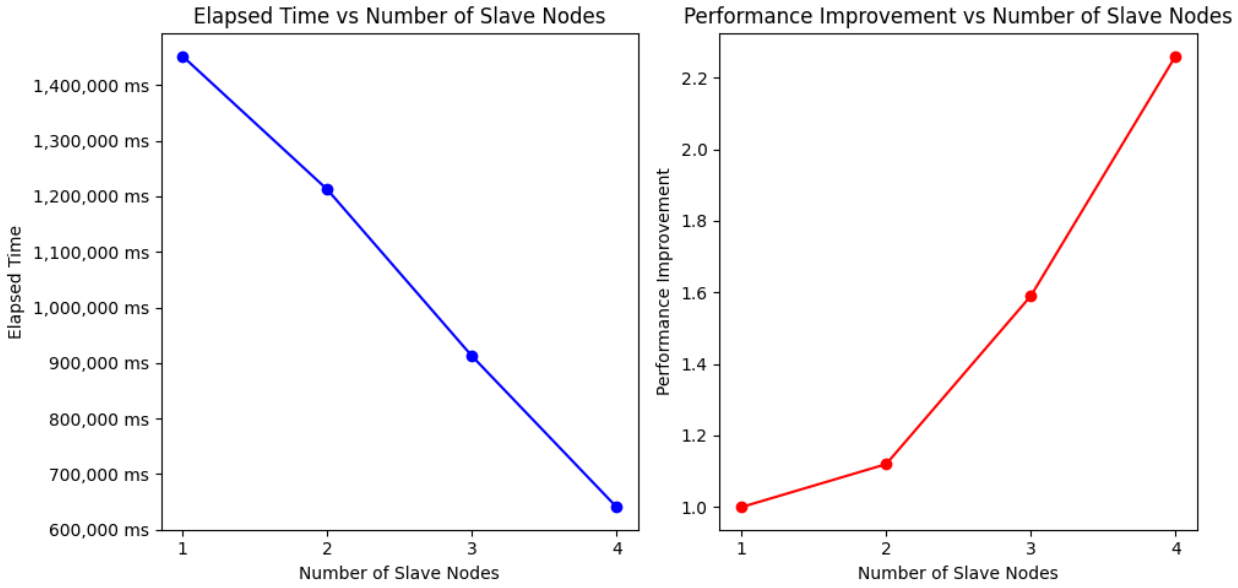
the KMP string matching algorithm to identify instances of the provided pattern, and emitting key-value pairs consisting of the filename and initial offset alongside the detected pattern occurrences. Conversely, the KMPReducer class, also derived from MapReduceBase and implementing the Reducer interface, consolidates the output received from mappers. It aggregates the identified pattern occurrences from different files, enabling a comprehensive view of the matches found in distributed data.

Furthermore, the main method orchestrates the Hadoop job configuration, encompassing key details such as the job's name, input/output classes, designated mapper and reducer functions, input/output formats, and the pattern under investigation. The execution of the MapReduce task is initiated using JobClient.runJob(conf), which triggers the algorithm to analyze the input text files and generate insights into the occurrences of the specified pattern across multiple files. This implementation stands as a robust solution for distributed pattern matching, leveraging the efficiency of Hadoop's MapReduce model to efficiently process extensive datasets and extract pattern occurrences in a distributed computing environment.

Results of parallelizing using MapReduce JAVA. A folder with 3,784 files(228.5 MB) was used for evaluation.

#no of slave nodes	Elapsed Time(ms)	Performance Improvement
1	1452442	1.00
2	1212334	1.12

3	912312	1.59
4	**640249**	**2.26**



Parallelization with Spark

Based on the code provided, here are some of the main parallelization techniques used to implement the KMP pattern search algorithm in Spark:

RDD Parallelism: The text files are loaded as a JavaPairRDD called textRDD, where the key is the file path and the value is the file contents. This allows the data to be distributed across executor nodes in the cluster.

Data Parallelism: The mapToPair transformation applies the KMP search function to each file in parallel, by searching for the pattern in each file contents independently. This is a form of data parallelism, since the data (text files) are processed in parallel.

Task Parallelism: Within each file, the actual KMP algorithm applies task parallelism. The search process generates many independent subtasks of comparing characters and checking for matches. These tasks are automatically parallelized within each file search.

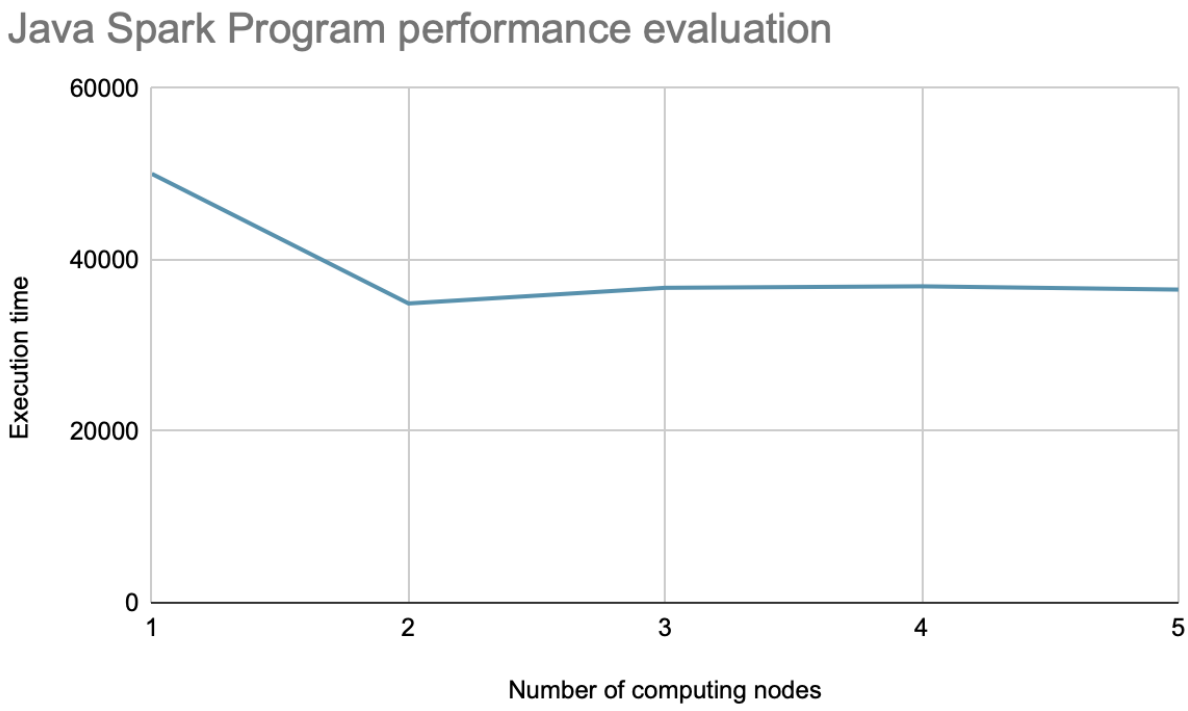
Embarrassingly Parallel: The search within each file does not depend on the processing of any other file. So the individual file searches are embarrassingly parallel across the nodes.

Partitioning: The final results RDD is coalesced to 1 partition to concatenate the outputs. This optimization helps shuffle the data to a single file efficiently.

So in summary, parallelism is achieved through RDD data distribution, data parallel file processing, parallel tasks within each search, and embarrassing parallelism across files. The Spark framework is leveraged to provide this distributed, parallel execution.

# of Comp nodes	Avg # of cores per node	Total # of cores	Elapsed time (Millisec)	Performance improvement
1	1	1	50021.000	1.000
1	2	2	36929.000	1.355

1	3	3	36816.000	1.359
1	4	4	37520.000	1.333
2	1	2	36876.000	1.356
2	2	4	34904.000	1.433
2	3	6	37922.000	1.319
2	4	8	43368.000	1.153
3	1	3	36729.000	1.362
3	2	6	41927.000	1.193
3	3	9	39495.000	1.267
3	4	12	42337.000	1.181
4	1	4	36902.000	1.356
4	2	8	38626.000	1.295
4	3	12	38655.000	1.294
4	4	16	40364.000	1.239
5	1	5	39418.000	1.269
5	2	10	40560.000	1.233
5	3	15	36524.000	1.370
5	4	20	44151.000	1.133



Graph is showing elapsed time of 1 computing Node with 1 core, 2 computing nodes with 4 cores, 3 computing nodes with 3 cores, 4 computing nodes with 4 cores and 5 computing nodes with 15 cores.

Parallelization with MASS Approach 1

MASS implementation of this program also utilizes a divide-and-conquer strategy. "Places" is initialized with the size of filenames. A subclass of Places, called FilesArray, is used as a "Place" in "Places." Upon initializing, each FilesArray object is populated with the LPS array using the pattern and file to process. FilesArray contains three functions: "init" to initialize each FilesArray object in place, "performSearch" to perform a search on the pattern using the LPS array, and

"getIndexes" to retrieve all indexes found in that place. "Places.CallAll" methods are used to execute each of the above functions in all the places, which returns an object of class Result to the driver program. The resultant array is then used to write an output file, which contains the file name and indexes found in that file.

Parallelization with MASS Approach 2

MASS implementation of this approach also utilizes a divide-and-conquer strategy. But in this approach we created both places and agents.

Parallelization

- Input text split across TextChunk places
- Search callMethod done in parallel across places
- Each KMPSearcher searches a portion of input

Programmability analysis

MPI : Parallelizing the application in javaMPI was very elegant since all nodes are connected to a network file system and a smaller task for performing search on individual files was very efficient. Only one synchronization is required between the parallel programs running at various nodes. Implementation of each node writing a specific file not only reduces IO operation competition between nodes but also provides an opportunity to add fault tolerance to the program. Meanwhile, implementing fault tolerance and recovery requires a lot of checks and distributing failed and remaining workload to other nodes. Similarly, since each node performs

search on the same number of files regardless of the file size, optimal load balancing was not achieved with the implemented approach. To further optimize the program, more checks with conditional statements and loops will be required to maintain the file assigned to each node rather than a simple array index used in the program. Therefore, parallelizing with MPI was able to improve execution speed but load balancing and optimization will require a lot of program changes.

MAP REDUCE: The code makes strategic design decisions to improve performance and scalability such as encapsulating the KMP search logic within mappers for independent processing, pre-computing the LPS array to optimize search time complexity and consolidating output to reduce fragmentation. However, tight coupling of the KMP algorithm to the mapper, reliance on native Hadoop types over abstractions and the possibility of reducer bottlenecks highlight some technical debt being accrued.

While these performance optimizations meet functional requirements well today, they tradeoff some flexibility, reusability, and portability of the implementation. Introducing greater modularity through more pluggable interfaces and component abstractions can help build longer-term resilience. In summary, the strategic thinking of separating core search execution from framework execution graphs shows promise from a programmability lens but further decoupling key modules like the algorithmic logic can maximize code quality and extensibility for future needs. Required tuning of partitioning and memory vs computing strategies also highlight the performance-portability tradeoffs.

SPARK:

Here is a programmability analysis of the key logic in the KmpSpark code:

1. It loads the text files from the input directory into a PairRDD, where the key is the file path and value is the file content. This allows parallel processing of each file.
2. The search pattern is obtained from the command line arguments. This allows the user to specify the search string without changing code.
3. The core search logic is implemented in the KMPSearch FlatMapFunction that takes the file content and returns a list of matching indices. This encapsulates the search logic for reusability and testing.
4. KMPSearch uses the computeLPSArray method to efficiently pre-process the pattern and construct the LPS array. This avoids re-computation of LPS for every search and improves performance.
5. The search logic itself correctly implements the KMP algorithm to find all occurrences of the pattern in the text. Use of LPS array allows faster matching.
6. The results are mapped back to filename and indices tuple and coalesced to one file for easy output. This provides a simple way to get search results.

In summary, the key aspects that enable efficient programmability are:

- Encapsulation of search logic
- Reusable LPS array computation
- Parallel processing per file

Here are some ways the KmpSpark code could be improved:

1. Computing the LPS array only once in the driver program.

2. Broadcasting the lps array.
3. Accessing the broadcasted LPS array in the KMPSearch class:
4. Removing the local computation of LPS array within KMPSearch.
5. Making sure Spark tasks do not recompute the LPS array on each executor.
6. Broadcasting the search pattern string as well
7. Using Accumulators to aggregate match counters
8. Custom partitioning to optimize pruning of data.

MASS:

We tried two approaches for MASS but not able to get result. Updated the code of both approaches.

Programmability Analysis of Approach 1:

Places Initialization:

- The program creates a Places instance named "FilesArray" with a handle of 1.
- The initial values for the Places instance include the search pattern and the array of file paths.

Parallel Execution:

- The program initializes all places by calling the callAll(0) method. The specific logic of initialization is implemented in the callMethod of each Place.
- It then performs the search on each place by calling callAll(2).

Programmability Analysis of Approach 2:

1. Initialize MASS - Sets up the parallel execution environment
2. Initialize Places (TextChunks) - Creates a shared array of TextChunk objects, one for each input text file
3. Initialize Agents (KMPSearchers) - Creates KMPSearcher agents across Places
4. Perform Search Cycles - Calls parallel search method of agents
5. Get Text Content - Calls method to provide text content to TextChunks
6. Get Results - Retrieves results from TextChunks

TextChunk Place

- Holds content of an input text file
- Provides content to agents on callMethod()
- Stores matches found by agents

KMPSearcher Agent

- Performs KMP string search
- Computes LPS array
- Searches text on callMethod()
- Updates match indices

Key Methods

- initializeTextChunks() - Initialize places
- callMethod() - Handle remote method calls

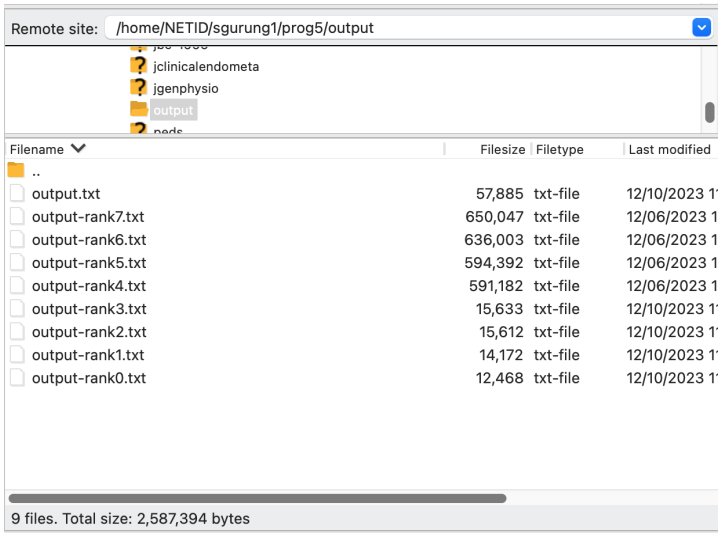
Execution Screenshoot

javaMpi:

Executing program with various node sizes:

```
[sgurung1@cssmpi1h ~]$ CSSmpdboot -n 8 -v
Preparing...
Starting Master's mpd Process...
Starting node: cssmpi2h.uwb.edu
Starting node: cssmpi3h.uwb.edu
Starting node: cssmpi4h.uwb.edu
Starting node: cssmpi5h.uwb.edu
Starting node: cssmpi6h.uwb.edu
Starting node: cssmpi7h.uwb.edu
Starting node: cssmpi8h.uwb.edu
Cluster built:
cssmpi1h.uwb.edu_38167 (10.158.82.61)
cssmpi8h.uwb.edu_41910 (10.158.82.68)
cssmpi7h.uwb.edu_33234 (10.158.82.67)
cssmpi6h.uwb.edu_34457 (10.158.82.66)
cssmpi5h.uwb.edu_37549 (10.158.82.65)
cssmpi4h.uwb.edu_45002 (10.158.82.64)
cssmpi3h.uwb.edu_38124 (10.158.82.63)
cssmpi2h.uwb.edu_35393 (10.158.82.62)
CSSmpdboot finished!
[sgurung1@cssmpi1h ~]$ cd prog5
[sgurung1@cssmpi1h prog5]$ ls
humanrep  input  jbc-1996  jclinicalendometa  jgenphysio  KmpMpi.class  KmpMpi.java  output  peds
[sgurung1@cssmpi1h prog5]$ mpirun -n 1 java KmpMpi therapy humanrep
therapy
0000000
Processing at rank 0 is Complete!
Output.txt Generation complete.
Elapsed time: 387991
[sgurung1@cssmpi1h prog5]$ mpirun -n 2 java KmpMpi therapy humanrep
therapy
0000000
Processing at rank 0 is Complete!
Processing at rank 1 is Complete!
Output.txt Generation complete.
Elapsed time: 147201
[sgurung1@cssmpi1h prog5]$ mpirun -n 3 java KmpMpi therapy humanrep
therapy
0000000
Processing at rank 0 is Complete!
Processing at rank 1 is Complete!
Processing at rank 2 is Complete!
Output.txt Generation complete.
Elapsed time: 112201
[sgurung1@cssmpi1h prog5]$ mpirun -n 4 java KmpMpi therapy humanrep
therapy
0000000
Processing at rank 0 is Complete!
Processing at rank 1 is Complete!
Processing at rank 2 is Complete!
Processing at rank 3 is Complete!
Output.txt Generation complete.
Elapsed time: 82115
```

Screenshot of output folder after program execution:



Screenshot of content of output.txt for pattern therapy

```
  GNU nano 2.3.1                          File: output.txt

10099964.html
2082 2377 3334 3557 6625 8339 9612 9678 16447 20503 21294 22254 23187 23403 23577 24304 31051 31835 36277

10099965.html
60226

10099967.html
38451 50244

10099968.html
56 2735 2776 3035 3102 3236 3302 5703 6093 9563 10724 10776 11327 11436 14195 14305 14404 14799 14933 15042 15186 15591 15657 157$

10099970.html
40961 62069

10099973.html
52368

10099977.html
16938 24448 40964 45871 48803

10099978.html
43306

10099980.html
15727 23292

10099985.html
55119

10099988.html
14062 54524

10099990.html
16668

10099991.html
161853

10099998.html
3825 4327 8122 34686 38066 43431 46655 51398 51805 56341 64635 68280 68571

10100000.html
41213

10100003.html
42499

10100005.html
7534 13610 13742 16622

10100007.html
84266

^G Get Help   ^O WriteOut   ^R Read File   ^Y Prev Page   ^K Cut Text   ^C Cur Pos
^X Exit       ^J Justify    ^W Where Is     ^V Next Page   ^U UnCut Text ^T To Spell
```

KMPMapReduce.java:

Execution output: Best performance when 4 slave nodes are used.

```
23/12/06 13:40:15 INFO mapred.JobClient:  map 95% reduce 31%
23/12/06 13:40:33 INFO mapred.JobClient:  map 96% reduce 31%
23/12/06 13:40:47 INFO mapred.JobClient:  map 96% reduce 32%
23/12/06 13:40:48 INFO mapred.JobClient:  map 97% reduce 32%
23/12/06 13:41:02 INFO mapred.JobClient:  map 98% reduce 32%
23/12/06 13:41:15 INFO mapred.JobClient:  map 99% reduce 32%
23/12/06 13:41:29 INFO mapred.JobClient:  map 100% reduce 33%
23/12/06 13:41:38 INFO mapred.JobClient:  map 100% reduce 100%
23/12/06 13:41:40 INFO mapred.JobClient: Job complete: job_202312061311_0001
23/12/06 13:41:40 INFO mapred.JobClient: Counters: 19
23/12/06 13:41:40 INFO mapred.JobClient:   Map-Reduce Framework
23/12/06 13:41:40 INFO mapred.JobClient:     Combine output records=0
23/12/06 13:41:40 INFO mapred.JobClient:     Spilled Records=4865744
23/12/06 13:41:40 INFO mapred.JobClient:     Reduce input records=2432872
23/12/06 13:41:40 INFO mapred.JobClient:     Reduce output records=1082
23/12/06 13:41:40 INFO mapred.JobClient:     Map input records=2432872
23/12/06 13:41:40 INFO mapred.JobClient:     Map output records=2432872
23/12/06 13:41:40 INFO mapred.JobClient:     Map output bytes=41371048
23/12/06 13:41:40 INFO mapred.JobClient:     Reduce shuffle bytes=46258044
23/12/06 13:41:40 INFO mapred.JobClient:     Combine input records=0
23/12/06 13:41:40 INFO mapred.JobClient:     Map input bytes=220682569
23/12/06 13:41:40 INFO mapred.JobClient:     Reduce input groups=3784
23/12/06 13:41:40 INFO mapred.JobClient:   FileSystemCounters
23/12/06 13:41:40 INFO mapred.JobClient:     HDFS_BYTES_READ=220682569
23/12/06 13:41:40 INFO mapred.JobClient:     FILE_BYTES_WRITTEN=92617400
23/12/06 13:41:40 INFO mapred.JobClient:     FILE_BYTES_READ=46236816
23/12/06 13:41:40 INFO mapred.JobClient:     HDFS_BYTES_WRITTEN=34686
23/12/06 13:41:40 INFO mapred.JobClient:   Job Counters
23/12/06 13:41:40 INFO mapred.JobClient:     Launched map tasks=3787
23/12/06 13:41:40 INFO mapred.JobClient:     Launched reduce tasks=1
23/12/06 13:41:40 INFO mapred.JobClient:     Rack-local map tasks=2
23/12/06 13:41:40 INFO mapred.JobClient:     Data-local map tasks=3785
Elapsed time = 640249 ms
```

Screenshot of output:

```
10099964.html
2082 2377 3334 3557 6625 8339 9612 9678 16447 20503 21294 22254 23187 23403 23577 24304 31051 31835 36277

10099965.html
60226

10099967.html
38451 50244

10099968.html
56 2735 2776 3035 3102 3236 3302 5703 6093 9563 10724 10776 11327 11436 14195 14305 14404 14799 14933 15042 15186 15591 15657 15796 16703 18195 19169 19530 20408 21558 23192 24587 24695 25245 25318 26030 30400 3$

10099970.html
40961 62069

10099973.html
52368

10099977.html
16938 24448 40964 45871 48803

10099978.html
43386

10099980.html
15727 23292

10099985.html
55119

10099988.html
14062 54524

10099990.html
16668

10099991.html
161853

10099998.html
3825 4327 8122 34686 38066 43431 46655 51398 51805 56341 64635 68280 68571

10100000.html
41213

10100003.html
42499

10100005.html
7534 13610 13742 16622

10100007.html
84266

10100008.html
17463

10100011.html
31002

10100013.html
2299 2634
```

KMPSpark Execution output:

Best performance with 2 computing node and 4 cores

1 Computing node with 1 Core

```
2023-12-06 04:05:25 INFO  TaskSetManager:54 - Starting task 0.0 in stage 0.0 (TID 0, 10.158.82.161, executor 0, partition 0, PROCESS_LOCAL, 2631
2023-12-06 04:05:25 INFO  BlockManagerMasterEndpoint:54 - Registering block manager 10.158.82.161:42581 with 366.3 MB RAM, BlockManagerId(0, 10.
2023-12-06 04:05:26 INFO  BlockManagerInfo:54 - Added broadcast_1_piece0 in memory on 10.158.82.161:42581 (size: 26.0 KB, free: 366.3 MB)
2023-12-06 04:05:26 INFO  BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on 10.158.82.161:42581 (size: 23.0 KB, free: 366.3 MB)
2023-12-06 04:06:04 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 38795 ms on 10.158.82.161 (executor 0) (1/1)
2023-12-06 04:06:04 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 04:06:04 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 47.991 s
2023-12-06 04:06:04 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 48.021855 s
2023-12-06 04:06:04 INFO  SparkHadoopWriter:54 - Job job_20231206040515_0004 committed.
Time = 50021
2023-12-06 04:06:04 INFO  AbstractConnector:318 - Stopped Spark@6d511b5f{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 04:06:04 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi1h.uwb.edu:4040
2023-12-06 04:06:04 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 04:06:04 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2023-12-06 04:06:04 INFO  MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
```

app-20231206040514-0011	KmpSpark	1	1024.0 MB	2023/12/06 04:05:14	shristi	FINISHED	50 s

1 Computing node with 2 Core

```
2023-12-06 12:38:12 INFO  BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on 10.158.82.161:39975 (size: 23.0 KB, free: 366.3 MB)
2023-12-06 12:38:46 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 34496 ms on 10.158.82.161 (executor 0) (1/1)
2023-12-06 12:38:46 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 12:38:46 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 34.566 s
2023-12-06 12:38:46 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 34.595813 s
2023-12-06 12:38:46 INFO  SparkHadoopWriter:54 - Job job_20231206123810_0004 committed.
Time = 36929
2023-12-06 12:38:46 INFO  AbstractConnector:318 - Stopped Spark@303dba45{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 12:38:46 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi17h.uwb.edu:4040
2023-12-06 12:38:46 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 12:38:46 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2023-12-06 12:38:46 INFO  MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
```

Workers (1)

Worker Id	Address	State	Cores	Memory
worker-20231206123727-10.158.82.161-38604	10.158.82.161:38604	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration

Completed Applications (1)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
app-20231206123809-0000	KmpSpark	2	1024.0 MB	2023/12/06 12:38:09	shristi	FINISHED	37 s

1 Computing node with 3 core

```
2023-12-06 12:41:05 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 34753 ms on 10.158.82.161 (executor 0) (1/1)
2023-12-06 12:41:05 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 12:41:05 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 34.810 s
2023-12-06 12:41:05 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 34.837197 s
2023-12-06 12:41:05 INFO  SparkHadoopWriter:54 - Job job_20231206124029_0004 committed.
Time = 36816
2023-12-06 12:41:05 INFO  AbstractConnector:318 - Stopped Spark@6d6bc158{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 12:41:05 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi17h.uwb.edu:4040
2023-12-06 12:41:05 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 12:41:05 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2023-12-06 12:41:05 INFO  MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
```

Workers (1)

Worker Id	Address	State	Cores	Memory
worker-20231206123727-10.158.82.161-38604	10.158.82.161:38604	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration

Completed Applications (2)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
app-20231206124028-0001	KmpSpark	3	1024.0 MB	2023/12/06 12:40:28	shristi	FINISHED	37 s
app-20231206123809-0000	KmpSpark	2	1024.0 MB	2023/12/06 12:38:09	shristi	FINISHED	37 s

1 computing node with 4 Cores

```
2023-12-06 05:03:26 INFO  BlockManagerInfo:54 - Added broadcast_1_piece0 in memory on 10.158.82.161:35843 (size: 26.0 KB, free: 366.3 MB)
2023-12-06 05:03:26 INFO  BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on 10.158.82.161:35843 (size: 23.0 KB, free: 366.3 MB)
2023-12-06 05:04:01 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 35226 ms on 10.158.82.161 (executor 0) (1/1)
2023-12-06 05:04:01 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 05:04:01 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 35.297 s
2023-12-06 05:04:01 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 35.333653 s
2023-12-06 05:04:01 INFO  SparkHadoopWriter:54 - Job job_20231206050324_0004 committed.
Time = 37520
2023-12-06 05:04:01 INFO  AbstractConnector:318 - Stopped Spark@6d511b5f{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 05:04:01 INFO  ContextCleaner:54 - Cleaned accumulator 13
2023-12-06 05:04:01 INFO  ContextCleaner:54 - Cleaned accumulator 17
2023-12-06 05:04:01 INFO  ContextCleaner:54 - Cleaned accumulator 24
2023-12-06 05:04:01 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi1h.uwb.edu:4040
2023-12-06 05:04:01 INFO  ContextCleaner:54 - Cleaned accumulator 2
```

Spark Master at spark://cssmpi17h.uwb.edu:58310

URL: spark://cssmpi17h.uwb.edu:58310
REST URL: spark://cssmpi17h.uwb.edu:6066 *(cluster mode)*
Alive Workers: 1
Cores in use: 4 Total, 0 Used
Memory in use: 14.5 GB Total, 0.0 B Used
Applications: 0 Running, 29 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers (1)

Worker Id	Address	State	Cores	Memory
worker-20231205193047-10.158.82.161-40766	10.158.82.161:40766	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration

Completed Applications (29)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
app-20231206050323-0028	KmpSpark	4	1024.0 MB	2023/12/06 05:03:23	shristi	FINISHED	38 s

2 computing nodes 2 Cores

```
2023-12-06 05:09:19 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Registered executor NettyRpcEndpointRef(spark-client://Executor) (10.158.82.162:39154) with ID 1
2023-12-06 05:09:19 INFO  BlockManagerMasterEndpoint:54 - Registering block manager 10.158.82.162:43658 with 366.3 MB RAM, BlockManagerId(1, 10.158.82.162, 43658, None)
2023-12-06 05:09:19 INFO  BlockManagerInfo:54 - Added broadcast_1_piece0 in memory on 10.158.82.161:39460 (size: 26.0 KB, free: 366.3 MB)
2023-12-06 05:09:20 INFO  BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on 10.158.82.161:39460 (size: 23.0 KB, free: 366.3 MB)
2023-12-06 05:09:54 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 34741 ms on 10.158.82.161 (executor 0) (1/1)
2023-12-06 05:09:54 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 05:09:54 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 34.827 s
2023-12-06 05:09:54 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 34.866735 s
2023-12-06 05:09:54 INFO  SparkHadoopWriter:54 - Job job_20231206050918_0004 committed.
Time = 36876
2023-12-06 05:09:54 INFO  AbstractConnector:318 - Stopped Spark@6d24600{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 05:09:54 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi1h.uwb.edu:4040
2023-12-06 05:09:54 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 05:09:54 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
```

2 computing nodes 4 Cores

```
2023-12-06 05:11:06 WARN  TaskSetManager:66 - Stage 0 contains a task of very large size (256 KB). The maximum recommended task size is 100 KB.
2023-12-06 05:11:06 INFO  TaskSetManager:54 - Starting task 0.0 in stage 0.0 (TID 0, 10.158.82.161, executor 0, partition 0, PROCESS_LOCAL, 263111 bytes)
2023-12-06 05:11:06 INFO  BlockManagerInfo:54 - Added broadcast_1_piece0 in memory on 10.158.82.161:43428 (size: 26.0 KB, free: 366.3 MB)
2023-12-06 05:11:07 INFO  BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on 10.158.82.161:43428 (size: 23.0 KB, free: 366.3 MB)
2023-12-06 05:11:39 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 32617 ms on 10.158.82.161 (executor 0) (1/1)
2023-12-06 05:11:39 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 05:11:39 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 32.690 s
2023-12-06 05:11:39 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 32.720289 s
2023-12-06 05:11:39 INFO  SparkHadoopWriter:54 - Job job_20231206051104_0004 committed.
Time = 34904
2023-12-06 05:11:39 INFO  AbstractConnector:318 - Stopped Spark@d714e5a{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 05:11:39 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi1h.uwb.edu:4040
2023-12-06 05:11:39 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 05:11:39 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
```

Workers (2)

Worker Id	Address	State	Cores	Memory
worker-20231205193047-10.158.82.161-40766	10.158.82.161:40766	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)
worker-20231206050758-10.158.82.162-34615	10.158.82.162:34615	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration

Completed Applications (32)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
app-20231206051103-0031	KmpSpark	4	1024.0 MB	2023/12/06 05:11:03	shristi	FINISHED	35 s

2 computing nodes with 8 Cores

```
2023-12-06 05:13:10 INFO  BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on 10.158.82.162:39682 (size: 23.0 KB, free: 366.3 MB)
2023-12-06 05:13:50 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 41185 ms on 10.158.82.162 (executor 1) (1/1)
2023-12-06 05:13:50 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 05:13:50 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 41.260 s
2023-12-06 05:13:50 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 41.343043 s
2023-12-06 05:13:50 INFO  SparkHadoopWriter:54 - Job job_20231206051308_0004 committed.
Time = 43368
2023-12-06 05:13:50 INFO  AbstractConnector:318 - Stopped Spark@6d511b5f{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 05:13:50 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi1h.uwb.edu:4040
2023-12-06 05:13:51 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 05:13:51 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2023-12-06 05:13:51 INFO  MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
2023-12-06 05:13:51 INFO  MemoryStore:54 - MemoryStore cleared
2023-12-06 05:13:51 INFO  BlockManager:54 - BlockManager stopped
```

Workers (2)

Worker Id	Address	State	Cores	Memory
worker-20231205193047-10.158.82.161-40766	10.158.82.161:40766	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)
worker-20231206050758-10.158.82.162-34615	10.158.82.162:34615	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration

Completed Applications (33)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
app-20231206051307-0032	KmpSpark	8	1024.0 MB	2023/12/06 05:13:07	shristi	FINISHED	44 s

3 computing nodes 3 cores

```
2023-12-06 11:29:52 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 34353 ms on 10.158.82.161 (executor 2) (1/1)
2023-12-06 11:29:52 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 11:29:52 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 34.423 s
2023-12-06 11:29:52 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 34.454983 s
2023-12-06 11:29:52 INFO  SparkHadoopWriter:54 - Job job_20231206112916_0004 committed.
Time = 36729
2023-12-06 11:29:52 INFO  AbstractConnector:318 - Stopped Spark@2e1b7a65{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 11:29:52 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi17h.uwb.edu:4040
2023-12-06 11:29:52 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 11:29:52 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
```

3 computing nodes 6 cores

```
2023-12-06 05:19:43 INFO  BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on 10.158.82.66:41308 (size: 23.0 KB, free: 366.3 MB)
2023-12-06 05:20:22 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 39716 ms on 10.158.82.66 (executor 2) (1/1)
2023-12-06 05:20:22 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 05:20:22 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 39.787 s
2023-12-06 05:20:22 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 39.821492 s
2023-12-06 05:20:22 INFO  SparkHadoopWriter:54 - Job job_20231206051941_0004 committed.
Time = 41927
2023-12-06 05:20:22 INFO  AbstractConnector:318 - Stopped Spark@6d511b5f{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 05:20:22 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi1h.uwb.edu:4040
2023-12-06 05:20:22 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 05:20:22 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
```

URL: spark://cssmpi17h.uwb.edu:58310
REST URL: spark://cssmpi17h.uwb.edu:6066 *(cluster mode)*
Alive Workers: 3
Cores in use: 12 Total, 0 Used
Memory in use: 47.5 GB Total, 0.0 B Used
Applications: 0 Running, 36 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers (3)

Worker Id	Address	State	Cores	Memory
worker-20231205193047-10.158.82.161-40766	10.158.82.161:40766	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)
worker-20231206050758-10.158.82.162-34615	10.158.82.162:34615	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)
worker-20231206051705-10.158.82.66-46207	10.158.82.66:46207	ALIVE	4 (0 Used)	18.5 GB (0.0 B Used)

Running Applications (0)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration

Completed Applications (36)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
app-20231206051940-0035	KmpSpark	6	1024.0 MB	2023/12/06 05:19:40	shristi	FINISHED	42 s

3 computing nodes with 9 cores

```
2023-12-06 11:36:27 INFO  BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on 10.158.82.163:38252 (size: 23.0 KB, free: 366.3 MB)
2023-12-06 11:37:04 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 36966 ms on 10.158.82.163 (executor 0) (1/1)
2023-12-06 11:37:04 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 11:37:04 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 37.023 s
2023-12-06 11:37:04 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 37.047538 s
2023-12-06 11:37:04 INFO  SparkHadoopWriter:54 - Job job_20231206113625_0004 committed.
Time = 39495
2023-12-06 11:37:04 INFO  AbstractConnector:318 - Stopped Spark@56cb9511{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 11:37:04 INFO  ContextCleaner:54 - Cleaned accumulator 5
2023-12-06 11:37:04 INFO  ContextCleaner:54 - Cleaned accumulator 4
2023-12-06 11:37:04 INFO  ContextCleaner:54 - Cleaned accumulator 15
```

3 computing nodes 12 cores

```
2023-12-06 05:22:03 INFO  BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on 10.158.82.66:33112 (size: 23.0 KB, free: 366.3 MB)
2023-12-06 05:22:42 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 40070 ms on 10.158.82.66 (executor 2) (1/1)
2023-12-06 05:22:42 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 05:22:42 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 40.143 s
2023-12-06 05:22:42 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 40.177183 s
2023-12-06 05:22:42 INFO  SparkHadoopWriter:54 - Job job_20231206052200_0004 committed.
Time = 42337
2023-12-06 05:22:42 INFO  AbstractConnector:318 - Stopped Spark@6d511b5f{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 05:22:42 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi1h.uwb.edu:4040
2023-12-06 05:22:42 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
```

4 nodes with 4 cores

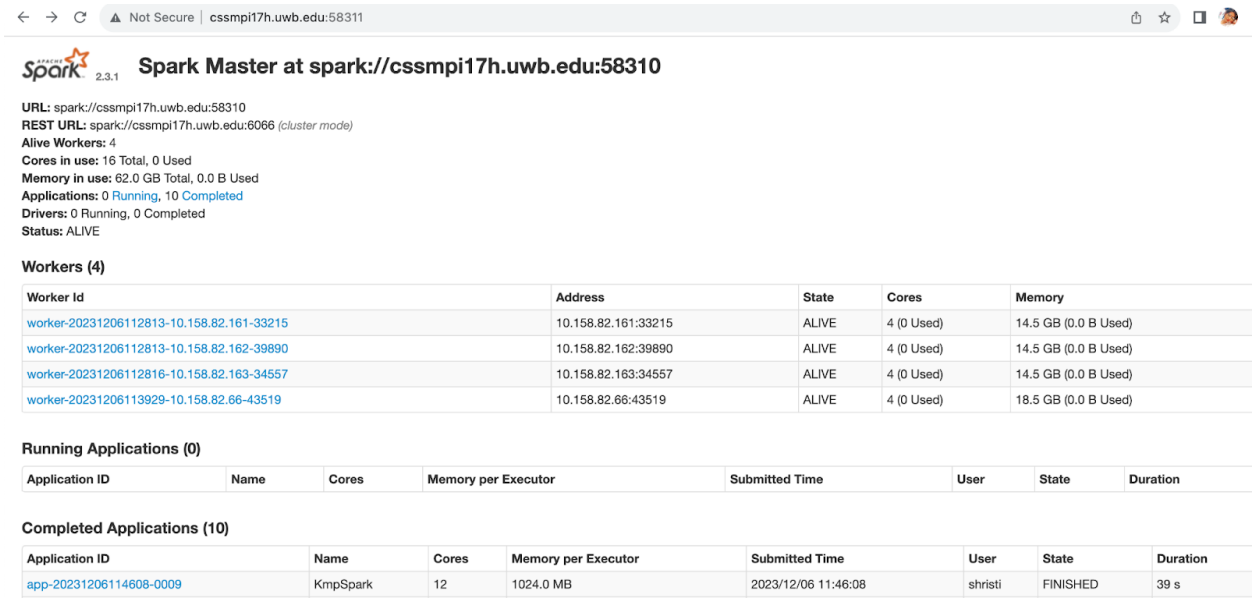
```
2023-12-06 11:41:02 INFO  ContextCleaner:54 - Cleaned accumulator 11
2023-12-06 11:41:02 INFO  ContextCleaner:54 - Cleaned accumulator 0
2023-12-06 11:41:02 INFO  ContextCleaner:54 - Cleaned accumulator 4
2023-12-06 11:41:02 INFO  SparkHadoopWriter:54 - Job job_20231206114026_0004 committed.
Time = 36902
2023-12-06 11:41:02 INFO  BlockManagerInfo:54 - Removed broadcast_1_piece0 on cssmpi17h.uwb.edu:38539 in memory (size: 26.0 KB, free: 366.3 MB)
2023-12-06 11:41:02 INFO  AbstractConnector:318 - Stopped Spark@2b1f72c7{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 11:41:02 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi17h.uwb.edu:4040
2023-12-06 11:41:02 INFO  BlockManagerInfo:54 - Removed broadcast_1_piece0 on 10.158.82.161:34695 in memory (size: 26.0 KB, free: 366.3 MB)
2023-12-06 11:41:02 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
```

4 nodes with 8 cores

```
2023-12-06 11:42:30 INFO  BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on 10.158.82.163:43525 (size: 23.0 KB, free: 366.3 MB)
2023-12-06 11:43:05 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 36301 ms on 10.158.82.163 (executor 0) (1/1)
2023-12-06 11:43:05 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 11:43:05 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 36.363 s
2023-12-06 11:43:05 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 36.393031 s
2023-12-06 11:43:05 INFO  SparkHadoopWriter:54 - Job job_20231206114227_0004 committed.
Time = 38626
2023-12-06 11:43:05 INFO  AbstractConnector:318 - Stopped Spark@10027fc9{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 11:43:05 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi17h.uwb.edu:4040
2023-12-06 11:43:05 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
```

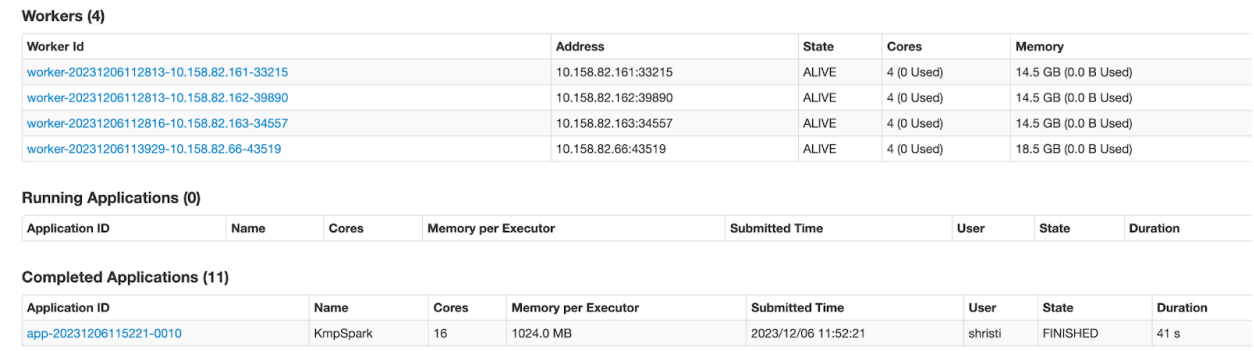
4 nodes with 12 cores

```
2023-12-06 11:46:47 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 36257 ms on 10.158.82.163 (executor 0) (1/1)
2023-12-06 11:46:47 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 11:46:47 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 36.315 s
2023-12-06 11:46:47 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 36.340145 s
2023-12-06 11:46:47 INFO  SparkHadoopWriter:54 - Job job_20231206114609_0004 committed.
Time = 38655
2023-12-06 11:46:47 INFO  AbstractConnector:318 - Stopped Spark@44140410{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 11:46:47 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi17h.uwb.edu:4040
2023-12-06 11:46:47 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 11:46:47 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2023-12-06 11:46:47 INFO  MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
2023-12-06 11:46:47 INFO  MemoryStore:54 - MemoryStore cleared
2023-12-06 11:46:47 INFO  BlockManager:54 - BlockManager stopped
2023-12-06 11:46:47 INFO  BlockManagerMaster:54 - BlockManagerMaster stopped
```



4 computing node with 16 cores

```
2023-12-06 11:53:02 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 37901 ms on 10.158.82.162 (executor 2) (1/1)
2023-12-06 11:53:02 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 11:53:02 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 37.958 s
2023-12-06 11:53:02 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 37.985370 s
2023-12-06 11:53:02 INFO  SparkHadoopWriter:54 - Job job_20231206115223_0004 committed.
Time = 40364
2023-12-06 11:53:02 INFO  AbstractConnector:318 - Stopped Spark@2440b67d{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 11:53:02 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi17h.uwb.edu:4040
2023-12-06 11:53:02 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 11:53:02 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2023-12-06 11:53:02 INFO  MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
2023-12-06 11:53:02 INFO  MemoryStore:54 - MemoryStore cleared
2023-12-06 11:53:02 INFO  BlockManager:54 - BlockManager stopped
2023-12-06 11:53:02 INFO  BlockManagerMaster:54 - BlockManagerMaster stopped
```



5 computing nodes with 5 core

```
2023-12-06 12:00:58 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 36844 ms on 10.158.82.163 (executor 0) (1/1)
2023-12-06 12:00:58 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 12:00:58 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 36.900 s
2023-12-06 12:00:58 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 36.927348 s
2023-12-06 12:00:58 INFO  SparkHadoopWriter:54 - Job job_20231206120020_0004 committed.
Time = 39418
2023-12-06 12:00:58 INFO  AbstractConnector:318 - Stopped Spark@303dba45{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 12:00:58 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi17h.uwb.edu:4040
2023-12-06 12:00:58 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 12:00:58 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2023-12-06 12:00:58 INFO  MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
2023-12-06 12:00:58 INFO  MemoryStore:54 - MemoryStore cleared
2023-12-06 12:00:58 INFO  BlockManager:54 - BlockManager stopped
2023-12-06 12:00:58 INFO  BlockManagerMaster:54 - BlockManagerMaster stopped
2023-12-06 12:00:58 INFO  OutputCommitCoordinator$OutputCommitCoordinatorEndpoint:54 - OutputCommitCoordinator stopped!
```

5 Computing nodes with 10 cores

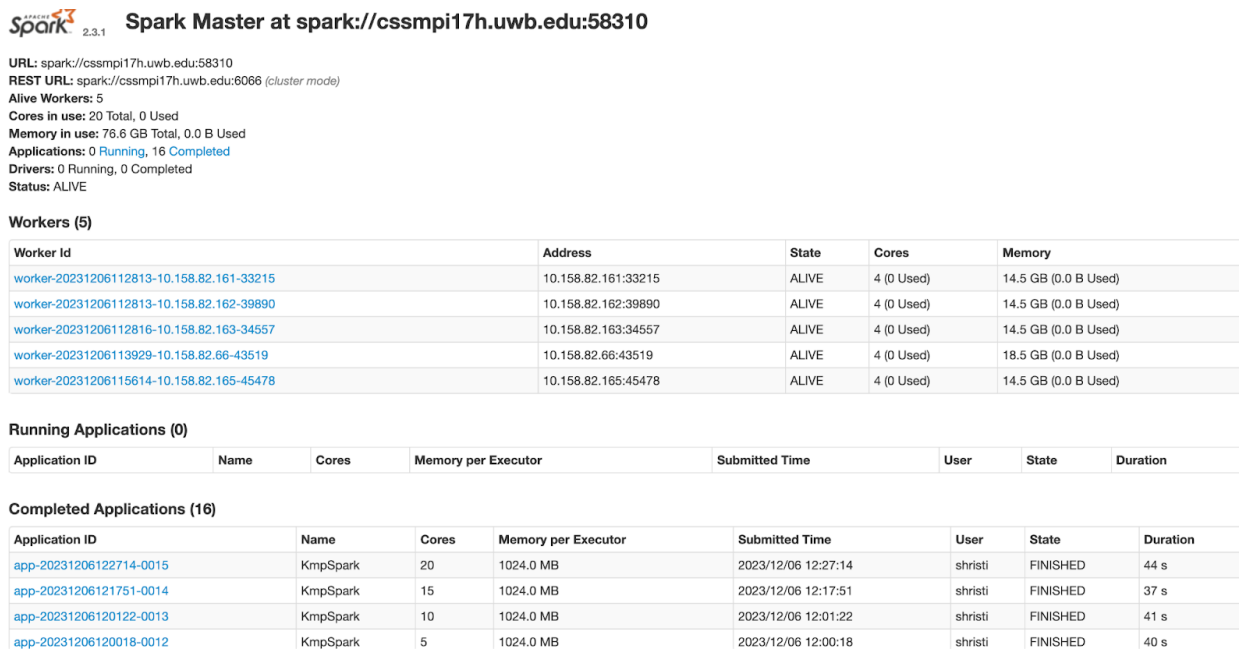
```
2023-12-06 12:01:25 INFO  BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on 10.158.82.162:44351 (size: 23.0 KB, free: 366.3 MB)
2023-12-06 12:02:02 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 37983 ms on 10.158.82.162 (executor 3) (1/1)
2023-12-06 12:02:02 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 12:02:02 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 38.054 s
2023-12-06 12:02:02 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 38.079082 s
2023-12-06 12:02:02 INFO  SparkHadoopWriter:54 - Job job_20231206120123_0004 committed.
Time = 40560
2023-12-06 12:02:02 INFO  AbstractConnector:318 - Stopped Spark@120ae5cb{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 12:02:02 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi17h.uwb.edu:4040
2023-12-06 12:02:02 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 12:02:02 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2023-12-06 12:02:02 INFO  MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
2023-12-06 12:02:03 INFO  MemoryStore:54 - MemoryStore cleared
2023-12-06 12:02:03 INFO  BlockManager:54 - BlockManager stopped
2023-12-06 12:02:03 INFO  BlockManagerMaster:54 - BlockManagerMaster stopped
```

5 computing nodes with 15 cores

```
2023-12-06 12:18:28 INFO  ContextCleaner:54 - Cleaned accumulator 4
2023-12-06 12:18:28 INFO  ContextCleaner:54 - Cleaned accumulator 14
2023-12-06 12:18:28 INFO  ContextCleaner:54 - Cleaned accumulator 11
2023-12-06 12:18:28 INFO  SparkHadoopWriter:54 - Job job_20231206121752_0004 committed.
Time = 36524
2023-12-06 12:18:28 INFO  AbstractConnector:318 - Stopped Spark@6d6bc158{HTTP/1.1,[http/1.1]}{0.0.0.0:4040}
2023-12-06 12:18:28 INFO  BlockManagerInfo:54 - Removed broadcast_1_piece0 on cssmpi17h.uwb.edu:36781 in memory (size: 26.0 KB, free: 366.3 MB)
2023-12-06 12:18:28 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi17h.uwb.edu:4040
2023-12-06 12:18:28 INFO  BlockManagerInfo:54 - Removed broadcast_1_piece0 on 10.158.82.161:46110 in memory (size: 26.0 KB, free: 366.3 MB)
2023-12-06 12:18:28 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 12:18:28 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2023-12-06 12:18:28 INFO  MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
2023-12-06 12:18:28 INFO  MemoryStore:54 - MemoryStore cleared
2023-12-06 12:18:28 INFO  BlockManager:54 - BlockManager stopped
2023-12-06 12:18:28 INFO  BlockManagerMaster:54 - BlockManagerMaster stopped
```

5 Computing nodes with 20 cores

```
2023-12-06 12:27:18 INFO  BlockManagerInfo:54 - Added broadcast_0_piece0 in memory on 10.158.82.165:34001 (size: 23.0 KB, free: 366.3 MB)
2023-12-06 12:27:58 INFO  TaskSetManager:54 - Finished task 0.0 in stage 0.0 (TID 0) in 40561 ms on 10.158.82.165 (executor 1) (1/1)
2023-12-06 12:27:58 INFO  TaskSchedulerImpl:54 - Removed TaskSet 0.0, whose tasks have all completed, from pool
2023-12-06 12:27:58 INFO  DAGScheduler:54 - ResultStage 0 (runJob at SparkHadoopWriter.scala:78) finished in 40.679 s
2023-12-06 12:27:58 INFO  DAGScheduler:54 - Job 0 finished: runJob at SparkHadoopWriter.scala:78, took 40.751095 s
2023-12-06 12:27:58 INFO  SparkHadoopWriter:54 - Job job_20231206122715_0004 committed.
Time = 44151
2023-12-06 12:27:58 INFO  AbstractConnector:318 - Stopped Spark@720971aa{HTTP/1.1,[http/1.1]}{0.0.0.0:4041}
2023-12-06 12:27:58 INFO  SparkUI:54 - Stopped Spark web UI at http://cssmpi1h.uwb.edu:4041
2023-12-06 12:27:58 INFO  StandaloneSchedulerBackend:54 - Shutting down all executors
2023-12-06 12:27:58 INFO  CoarseGrainedSchedulerBackend$DriverEndpoint:54 - Asking each executor to shut down
2023-12-06 12:27:58 INFO  MapOutputTrackerMasterEndpoint:54 - MapOutputTrackerMasterEndpoint stopped!
2023-12-06 12:27:58 INFO  MemoryStore:54 - MemoryStore cleared
2023-12-06 12:27:58 INFO  BlockManager:54 - BlockManager stopped
2023-12-06 12:27:58 INFO  BlockManagerMaster:54 - BlockManagerMaster stopped
```



Spark 2.3.1 **Spark Master at spark://cssmpi17h.uwb.edu:58310**

URL: spark://cssmpi17h.uwb.edu:58310
REST URL: spark://cssmpi17h.uwb.edu:6066 *(cluster mode)*
Alive Workers: 5
Cores in use: 20 Total, 0 Used
Memory in use: 76.6 GB Total, 0.0 B Used
Applications: 0 Running, 16 Completed
Drivers: 0 Running, 0 Completed
Status: ALIVE

Workers (5)

Worker Id	Address	State	Cores	Memory
worker-20231206112813-10.158.82.161-33215	10.158.82.161:33215	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)
worker-20231206112813-10.158.82.162-39890	10.158.82.162:39890	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)
worker-20231206112816-10.158.82.163-34557	10.158.82.163:34557	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)
worker-20231206113929-10.158.82.66-43519	10.158.82.66:43519	ALIVE	4 (0 Used)	18.5 GB (0.0 B Used)
worker-20231206115614-10.158.82.165-45478	10.158.82.165:45478	ALIVE	4 (0 Used)	14.5 GB (0.0 B Used)

Running Applications (0)

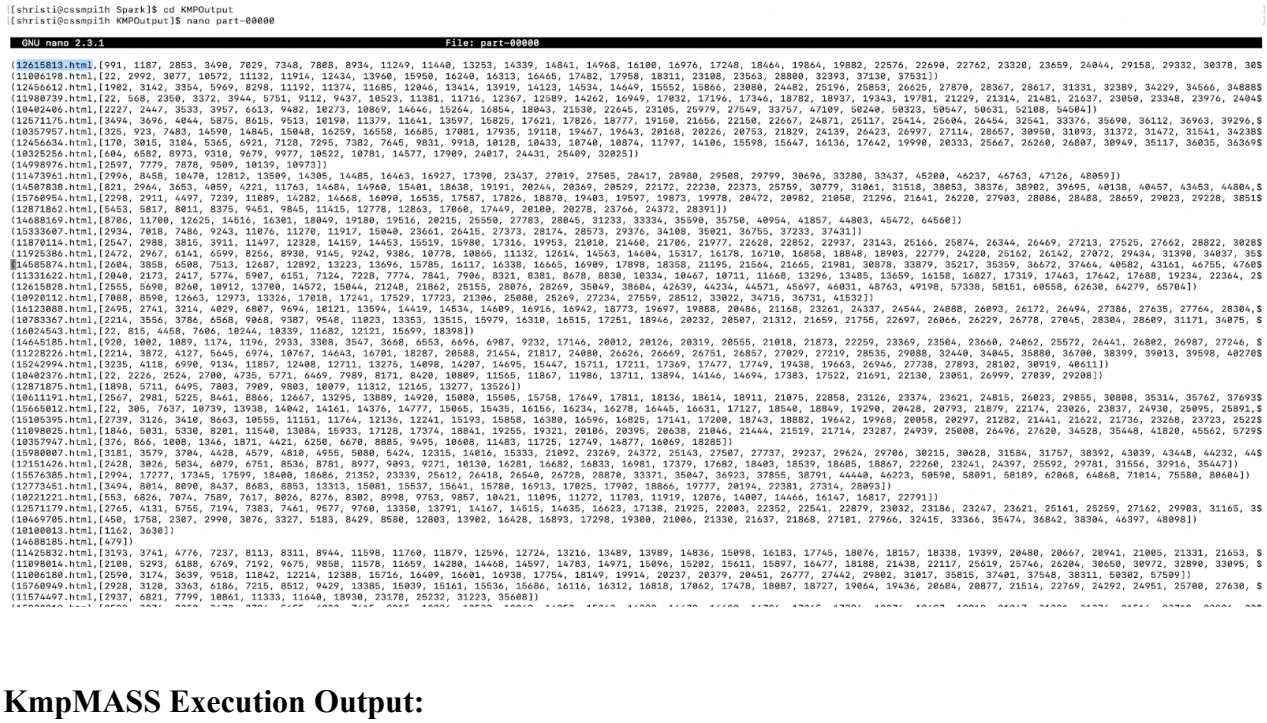
Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration

Completed Applications (16)

Application ID	Name	Cores	Memory per Executor	Submitted Time	User	State	Duration
app-20231206122714-0015	KmpSpark	20	1024.0 MB	2023/12/06 12:27:14	shristi	FINISHED	44 s
app-20231206121751-0014	KmpSpark	15	1024.0 MB	2023/12/06 12:17:51	shristi	FINISHED	37 s
app-20231206120122-0013	KmpSpark	10	1024.0 MB	2023/12/06 12:01:22	shristi	FINISHED	41 s
app-20231206120018-0012	KmpSpark	5	1024.0 MB	2023/12/06 12:00:18	shristi	FINISHED	40 s

Execution Output:

Output for Pattern: “the”



```
[shristi@cssmpi1h Spark]$ cd KMPOutput
[shristi@cssmpi1h KMPOutput]$ nano part-00000

 GNU nano 2.3.1                                File: part-00000

(12615813.html,[991, 1187, 2853, 3490, 7029, 7348, 7808, 8934, 11249, 11440, 13253, 14339, 14841, 14968, 16100, 16976, 17248, 18464, 19864, 19882, 22576, 22698, 22762, 23320, 23659, 24044, 29158, 29332, 30378, 30$
(11006198.html,[22, 2992, 3077, 10572, 11132, 11914, 12434, 13960, 15950, 16240, 16313, 16465, 17482, 17958, 18311, 23108, 23563, 28880, 32393, 37130, 37531])
(12456612.html,[1902, 3142, 3354, 5969, 8298, 11192, 11374, 11685, 12046, 13414, 13919, 14123, 14534, 14649, 15552, 15866, 23080, 24482, 25196, 25853, 26625, 27870, 28367, 28617, 31331, 32389, 34229, 34566, 34888$
(11980739.html,[22, 568, 2350, 3372, 3944, 5751, 9112, 9437, 10523, 11381, 11716, 12367, 12589, 14262, 16949, 17032, 17196, 17346, 18782, 18937, 19343, 19781, 21229, 21314, 21481, 21637, 23050, 23348, 23976, 2404$
(10402406.html,[2227, 2447, 3533, 3957, 6613, 9482, 10273, 10869, 14646, 15264, 16854, 18043, 21530, 22645, 23105, 25979, 27549, 33757, 47109, 50240, 50323, 50547, 50631, 52108, 54504])
(12571175.html,[3494, 3696, 4044, 5875, 8615, 9513, 10190, 11379, 11641, 13597, 15825, 17621, 17826, 18777, 19150, 21656, 22150, 22667, 24871, 25117, 25414, 25604, 26454, 32541, 33376, 35698, 36112, 36963, 39296,$
(10357957.html,[325, 923, 7483, 14590, 14845, 15048, 16259, 16558, 16685, 17081, 17935, 19118, 19467, 19643, 20168, 20226, 20753, 21829, 24139, 26423, 26997, 27114, 28657, 30950, 31893, 31372, 31472, 31541, 34238$
(12456634.html,[170, 3015, 3104, 5365, 6921, 7128, 7295, 7382, 7645, 9831, 9918, 10128, 10433, 10740, 10874, 11797, 14106, 15598, 15647, 16136, 17642, 19990, 20333, 25667, 26260, 26807, 30949, 35117, 36035, 36369$
(10325256.html,[604, 6582, 8973, 9310, 9679, 9977, 10522, 10781, 14577, 17909, 24017, 24431, 25409, 32025])
(14998076.html,[2597, 7779, 7878, 9509, 10139, 10973])
(11473961.html,[2996, 8458, 10470, 12812, 13509, 14305, 14485, 16463, 16927, 17390, 23437, 27019, 27505, 28417, 28980, 29508, 29799, 30696, 33280, 33437, 45200, 46237, 46763, 47126, 48059])
(14507838.html,[821, 2964, 3653, 4059, 4221, 11763, 14684, 14960, 15401, 18638, 19191, 20244, 20369, 20529, 22172, 22230, 22373, 25759, 30779, 31061, 31518, 38053, 38376, 38902, 39695, 40138, 40457, 43453, 44804,$
(15760954.html,[2298, 2911, 4497, 7239, 11089, 14282, 14668, 16090, 16535, 17587, 17826, 18870, 19403, 19597, 19873, 19978, 20472, 20982, 21050, 21296, 21641, 26220, 27903, 28086, 28488, 28659, 29023, 29228, 3851$
(12871862.html,[5453, 5817, 8011, 8375, 9451, 9845, 11415, 12778, 12863, 17060, 17449, 20100, 20278, 23766, 24372, 28391])
(14688169.html,[8706, 11700, 12625, 14516, 16301, 18049, 19180, 19516, 20215, 25550, 27783, 28045, 31233, 33334, 35590, 35750, 40954, 41857, 44803, 45472, 64560])
(15333607.html,[2934, 7018, 7486, 9243, 11076, 11278, 11917, 15040, 23661, 26415, 27373, 28174, 28573, 29376, 34108, 35021, 36755, 37233, 37431])
(11870114.html,[2547, 2988, 3815, 3911, 11497, 12328, 14159, 14453, 15519, 15980, 17316, 19953, 21010, 21460, 21706, 21977, 22628, 22852, 22937, 23143, 25166, 25874, 26344, 26469, 27213, 27525, 27662, 28822, 3028$
(11925386.html,[2472, 2967, 6141, 6599, 8256, 8930, 9145, 9242, 9386, 10778, 10865, 11132, 12614, 14563, 14604, 15317, 16178, 16710, 16858, 18848, 18903, 22779, 24220, 25162, 26142, 27072, 29434, 31390, 34037, 35$
(14585874.html,[2604, 3858, 6508, 7513, 12687, 12892, 13223, 13696, 15785, 16117, 16338, 16665, 16909, 17898, 18358, 21195, 21564, 21665, 21981, 30878, 33879, 35217, 35359, 36672, 37464, 40582, 43161, 46755, 4760$
(11331622.html,[2040, 2173, 2417, 5774, 5907, 6151, 7124, 7228, 7774, 7841, 7906, 8321, 8381, 8678, 8830, 10334, 10467, 10711, 11668, 13296, 13485, 13659, 16158, 16827, 17319, 17463, 17642, 17688, 19234, 22364, 2$
(12615828.html,[2555, 5690, 8260, 10912, 13700, 14572, 15044, 21248, 21862, 25155, 28076, 28269, 35849, 38604, 42639, 44234, 44571, 45697, 46031, 48763, 49198, 57338, 58151, 60558, 62630, 64279, 65704])
(10920112.html,[7088, 8590, 12663, 12973, 13326, 17018, 17241, 17529, 17723, 21306, 25080, 25269, 27234, 27559, 28512, 33022, 34715, 36731, 41532])
(16123088.html,[2495, 2741, 3214, 4029, 6807, 9694, 10121, 13594, 14419, 14534, 14609, 16916, 16942, 18773, 19697, 19888, 20486, 21168, 23261, 24337, 24544, 24888, 26893, 26172, 26494, 27386, 27635, 27764, 28304,$
(10783367.html,[2214, 3556, 3786, 6568, 9868, 9387, 9548, 11823, 13353, 13515, 15979, 16310, 16515, 17251, 18946, 20232, 20507, 21312, 21659, 21755, 22697, 26066, 26229, 26778, 27045, 28304, 28609, 31171, 34075, $
(16024543.html,[22, 815, 4458, 7606, 10244, 10339, 11682, 12121, 15699, 18398])
(14645185.html,[920, 1002, 1089, 1174, 1196, 2933, 3308, 3547, 3668, 6553, 6696, 6987, 9232, 17146, 20012, 20126, 20319, 20555, 21018, 21873, 22259, 23369, 23504, 23660, 24062, 25572, 26441, 26802, 26987, 27246, $
(11228226.html,[2214, 3872, 4127, 5645, 6974, 10767, 14643, 16701, 18287, 20588, 21454, 21817, 24080, 26626, 26669, 26751, 26857, 27029, 27219, 28535, 29088, 32440, 34045, 35880, 36708, 38399, 39013, 39598, 40270$
(15242994.html,[3235, 4118, 6990, 9134, 11857, 12408, 12711, 13275, 14098, 14207, 14695, 15447, 15711, 17211, 17369, 17477, 17749, 19438, 19663, 26946, 27738, 27893, 28102, 30919, 40611])
(10402376.html,[22, 2226, 2524, 2700, 4735, 5771, 6469, 7989, 8171, 8420, 10809, 11565, 11867, 11986, 13711, 13894, 14146, 14694, 17383, 17522, 21691, 22130, 23051, 26999, 27039, 29208])
(1287187.html,[1898, 5711, 6495, 7803, 7909, 9803, 10079, 11312, 12165, 13277, 13526])
(10611191.html,[2567, 2981, 5225, 8461, 8866, 12667, 13295, 13889, 14920, 15080, 15505, 15758, 17649, 17811, 18136, 18614, 18911, 21075, 22858, 23126, 23374, 23621, 24815, 26023, 29855, 30808, 35314, 35762, 37693$
(15665012.html,[22, 305, 7637, 10739, 13938, 14042, 14161, 14376, 14777, 15065, 15435, 16156, 16234, 16278, 16445, 16631, 17127, 18540, 18849, 19290, 20428, 20793, 21879, 22174, 23026, 23837, 24930, 25095, 25891,$
(15105395.html,[2739, 3126, 3410, 8663, 10555, 11151, 11764, 12136, 12241, 15193, 15858, 16380, 16596, 16825, 17141, 17200, 18743, 18882, 19642, 19968, 20058, 20297, 21282, 21441, 21622, 21736, 23268, 23723, 2522$
(11098025.html,[1846, 5031, 5330, 8201, 11540, 13084, 15933, 17128, 17374, 18841, 19255, 19321, 20106, 20395, 20638, 21046, 21444, 21519, 21714, 23287, 24939, 25008, 26496, 27620, 34528, 35448, 41820, 45562, 5729$
(10357947.html,[376, 866, 1008, 1346, 1871, 4421, 6250, 6670, 8885, 9495, 10608, 11483, 11725, 12749, 14877, 16869, 18285])
(15980007.html,[3181, 3579, 3704, 4428, 4579, 4818, 4955, 5080, 5424, 12315, 14016, 15333, 21092, 23269, 24372, 25143, 27507, 27737, 29237, 29624, 29706, 30215, 30628, 31584, 31757, 38392, 43039, 43448, 44232, 44$
(12151426.html,[2428, 3026, 5034, 6079, 6751, 8536, 8781, 8977, 9093, 9271, 10130, 16281, 16682, 16833, 16981, 17379, 17682, 18403, 18539, 18605, 18867, 22260, 23241, 24397, 25592, 29781, 31556, 32916, 35447])
(15576385.html,[2994, 17277, 17345, 17599, 18400, 18686, 21352, 23339, 25612, 26418, 26540, 26728, 28870, 33371, 35047, 36923, 37855, 38791, 44440, 46223, 50590, 58091, 58189, 62068, 64868, 71014, 75580, 80684])
(12773451.html,[3494, 8014, 8090, 8437, 8683, 8853, 13313, 15081, 15537, 15641, 15780, 16913, 17025, 17902, 18866, 19777, 20194, 22381, 27314, 28093])
(10221221.html,[553, 6826, 7074, 7589, 7617, 8026, 8276, 8302, 8998, 9753, 9857, 10421, 11095, 11272, 11703, 11919, 12076, 14007, 14466, 16147, 16817, 22791])
(12571179.html,[2765, 4131, 5755, 7194, 7383, 7461, 9577, 9760, 13350, 13791, 14167, 14515, 14635, 16623, 17138, 21925, 22003, 22352, 22541, 22879, 23032, 23186, 23247, 23621, 25161, 25259, 27162, 29903, 31165, 3$
(10469705.html,[450, 1758, 2307, 2990, 3076, 3327, 5183, 8429, 8580, 12803, 13902, 16428, 16893, 17298, 19300, 21006, 21330, 21637, 21868, 27101, 27966, 32415, 33366, 35474, 36842, 38304, 46397, 48098])
(10100013.html,[1162, 3630])
(14688185.html,[4791])
(11425832.html,[3193, 3741, 4776, 7237, 8113, 8311, 8944, 11598, 11760, 11879, 12596, 12724, 13216, 13489, 13989, 14836, 15098, 16183, 17745, 18076, 18157, 18338, 19399, 20480, 20667, 20941, 21005, 21331, 21653, $
(11098014.html,[2108, 5293, 6188, 6769, 7192, 9675, 9858, 11578, 11659, 14280, 14468, 14597, 14783, 14971, 15096, 15202, 15611, 15897, 16477, 18188, 21438, 22117, 25619, 25746, 26204, 30650, 30972, 32890, 33095, $
(11006180.html,[2590, 3174, 3639, 9518, 11842, 12214, 12388, 15716, 16409, 16601, 16938, 17754, 18149, 19914, 20237, 20379, 20451, 26777, 27442, 29802, 31017, 35815, 37401, 37548, 38311, 50302, 57509])
(15760949.html,[2928, 3120, 3363, 6186, 7215, 8512, 9429, 13385, 15039, 15161, 15536, 15686, 16116, 16312, 16818, 17062, 17478, 18087, 18727, 19064, 19436, 20684, 20877, 21514, 22769, 24292, 24951, 25700, 27630, $
(11574497.html,[2937, 6821, 7799, 10861, 11333, 11640, 18930, 23178, 25232, 31223, 35608])
```

KmpMASS Execution Output:

We tried to solve the runtime error we are getting in our code but not able to figure out what can be the exact cause of our error.

```
[INFO] --- maven-resources-plugin:2.6:resources (default-resources) @ KmpMASS ---
[WARNING] Using platform encoding (UTF-8 actually) to copy filtered resources, i.e. build is platform dependent!
[INFO] skip non existing resourceDirectory /home/NETID/shristi/mass_quickstart/Quickstart/src/main/resources
[INFO]
[INFO] --- maven-compiler-plugin:3.8.1:compile (default-compile) @ KmpMASS ---
[INFO] Changes detected - recompiling the module!
[INFO] Compiling 8 source files to /home/NETID/shristi/mass_quickstart/Quickstart/target/classes
[INFO] /home/NETID/shristi/mass_quickstart/Quickstart/src/main/java/edu/uwb/css534/QuickStart.java: /home/NETID/shristi/mass_quickstart/Quickstart/src/main/java/edu/uwb/css534/QuickStart.java uses or over
rides a deprecated API.
[INFO] /home/NETID/shristi/mass_quickstart/Quickstart/src/main/java/edu/uwb/css534/QuickStart.java: Recompile with -Xlint:deprecation for details.
[INFO] /home/NETID/shristi/mass_quickstart/Quickstart/src/main/java/edu/uwb/css534/KmpMASS.java: /home/NETID/shristi/mass_quickstart/Quickstart/src/main/java/edu/uwb/css534/KmpMASS.java uses unchecked or
unsafe operations.
[INFO] /home/NETID/shristi/mass_quickstart/Quickstart/src/main/java/edu/uwb/css534/KmpMASS.java: Recompile with -Xlint:unchecked for details.
[INFO]
[INFO] --- maven-resources-plugin:2.6:testResources (default-testResources) @ KmpMASS ---
[WARNING] Using platform encoding (UTF-8 actually) to copy filtered resources, i.e. build is platform dependent!
[INFO] skip non existing resourceDirectory /home/NETID/shristi/mass_quickstart/Quickstart/src/test/resources
[INFO]
[INFO] --- maven-compiler-plugin:3.8.1:testCompile (default-testCompile) @ KmpMASS ---
[INFO] No sources to compile
[INFO]
[INFO] --- maven-surefire-plugin:2.12.4:test (default-test) @ KmpMASS ---
[INFO] No tests to run.
[INFO]
[INFO] --- maven-jar-plugin:2.4:jar (default-jar) @ KmpMASS ---
[INFO] Building jar: /home/NETID/shristi/mass_quickstart/Quickstart/target/KmpMASS-1.0.0-SNAPSHOT.jar
[INFO]
[INFO] --- maven-assembly-plugin:3.1.0:single (default) @ KmpMASS ---
Downloading from netbeans: http://bits.netbeans.org/nexus/content/groups/netbeans/edu/ucar/netcdf4/5.1.0/netcdf4-5.1.0.pom
Downloading from netbeans: http://bits.netbeans.org/nexus/content/groups/netbeans/edu/ucar/cdm/5.1.0/cdm-5.1.0.pom
Downloading from netbeans: http://bits.netbeans.org/nexus/content/groups/netbeans/edu/ucar/udunits/5.1.0/udunits-5.1.0.pom
Downloading from netbeans: http://bits.netbeans.org/nexus/content/groups/netbeans/edu/ucar/httpservices/5.1.0/httpservices-5.1.0.pom
[WARNING] Artifact: edu.uwb.css534:KmpMASS:jar:1.0.0-SNAPSHOT references the same file as the assembly destination file. Moving it to a temporary location for inclusion.
[INFO] Building jar: /home/NETID/shristi/mass_quickstart/Quickstart/target/KmpMASS-1.0.0-SNAPSHOT.jar
[WARNING] Configuration option 'appendAssemblyId' is set to false.
Instead of attaching the assembly file: /home/NETID/shristi/mass_quickstart/Quickstart/target/KmpMASS-1.0.0-SNAPSHOT.jar, it will become the file for main project artifact.
NOTE: If multiple descriptors or descriptor-formats are provided for this project, the value of this file will be non-deterministic!
[WARNING] Replacing pre-existing project main-artifact file: /home/NETID/shristi/mass_quickstart/Quickstart/target/archive-tmp/KmpMASS-1.0.0-SNAPSHOT.jar
with assembly file: /home/NETID/shristi/mass_quickstart/Quickstart/target/KmpMASS-1.0.0-SNAPSHOT.jar
[INFO] ------------------------------------------------------------------------
[INFO] BUILD SUCCESS
[INFO] ------------------------------------------------------------------------
[INFO] Total time:  22.512 s
[INFO] Finished at: 2023-12-11T13:35:55-08:00
[INFO] ------------------------------------------------------------------------
[shristi@cssmpi1h Quickstart]$ cp target/KmpMASS-1.0.0-SNAPSHOT.jar .
[shristi@cssmpi1h Quickstart]$ java -jar target/KmpMASS-1.0.0-SNAPSHOT.jar input the
WARNING: An illegal reflective access operation has occurred
WARNING: Illegal reflective access by org.agrona.nio.TransportPoller (file:/home/NETID/shristi/mass_quickstart/Quickstart/target/KmpMASS-1.0.0-SNAPSHOT.jar) to field sun.nio.ch.SelectorImpl.selectedKeys
WARNING: Please consider reporting this to the maintainers of org.agrona.nio.TransportPoller
WARNING: Use --illegal-access=warn to enable warnings of further illegal reflective access operations
WARNING: All illegal access operations will be denied in a future release
MASS.init: done
Exception in thread "main" java.lang.NullPointerException
        at edu.uw.bothell.css.dsl.MASS.AgentsBase.<init>(AgentsBase.java:277)
        at edu.uw.bothell.css.dsl.MASS.Agents.<init>(Agents.java:64)
        at edu.uwb.css534.KmpMASS.main(KmpMASS.java:36)
```

	MPI	MapReduce	Spark	MASS
LOC	239	138	148	218

Loop#	7	3	3	4
Condition#	14	7	6	3