MapReduce Job Analysis

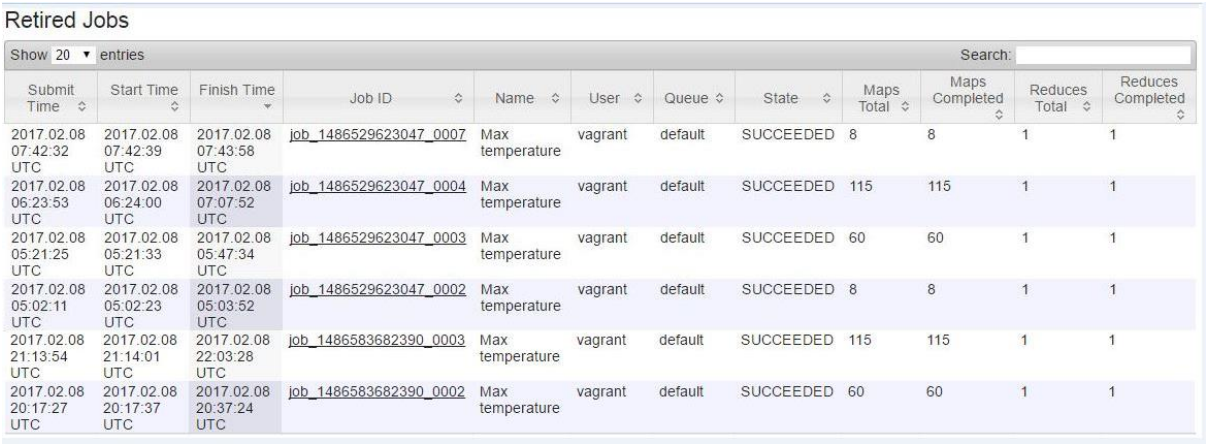
In this part, we must analyse the time taken to run the MaxTemperature and MaxTemperatureWithCombiner java files using MapReduce framework. From the below job chart it is evident that the time taken to process jobs mentioned in below sets, is increasing depending upon the size of the input file.

Set 1: 1990

Set 2: 1990 and 1992

Set 3: 1990, 1991, 1992 and 1993

We ran the above 3 sets of data using MaxTemperature and MaxTemperatureWithCombiner java files, and below is the output for the same.



Retired Jobs

Show 20 entries Search:

Submit Time	Start Time	Finish Time	Job ID	Name	User	Queue	State	Maps Total	Maps Completed	Reduces Total	Reduces Completed
2017.02.08 07:42:32 UTC	2017.02.08 07:42:39 UTC	2017.02.08 07:43:58 UTC	job_1486529623047_0007	Max temperature	vagrant	default	SUCCEEDED	8	8	1	1
2017.02.08 06:23:53 UTC	2017.02.08 06:24:00 UTC	2017.02.08 07:07:52 UTC	job_1486529623047_0004	Max temperature	vagrant	default	SUCCEEDED	115	115	1	1
2017.02.08 05:21:25 UTC	2017.02.08 05:21:33 UTC	2017.02.08 05:47:34 UTC	job_1486529623047_0003	Max temperature	vagrant	default	SUCCEEDED	60	60	1	1
2017.02.08 05:02:11 UTC	2017.02.08 05:02:23 UTC	2017.02.08 05:03:52 UTC	job_1486529623047_0002	Max temperature	vagrant	default	SUCCEEDED	8	8	1	1
2017.02.08 21:13:54 UTC	2017.02.08 21:14:01 UTC	2017.02.08 22:03:28 UTC	job_1486583682390_0003	Max temperature	vagrant	default	SUCCEEDED	115	115	1	1
2017.02.08 20:17:27 UTC	2017.02.08 20:17:37 UTC	2017.02.08 20:37:24 UTC	job_1486583682390_0002	Max temperature	vagrant	default	SUCCEEDED	60	60	1	1

Hadoop MapReduce Job Details

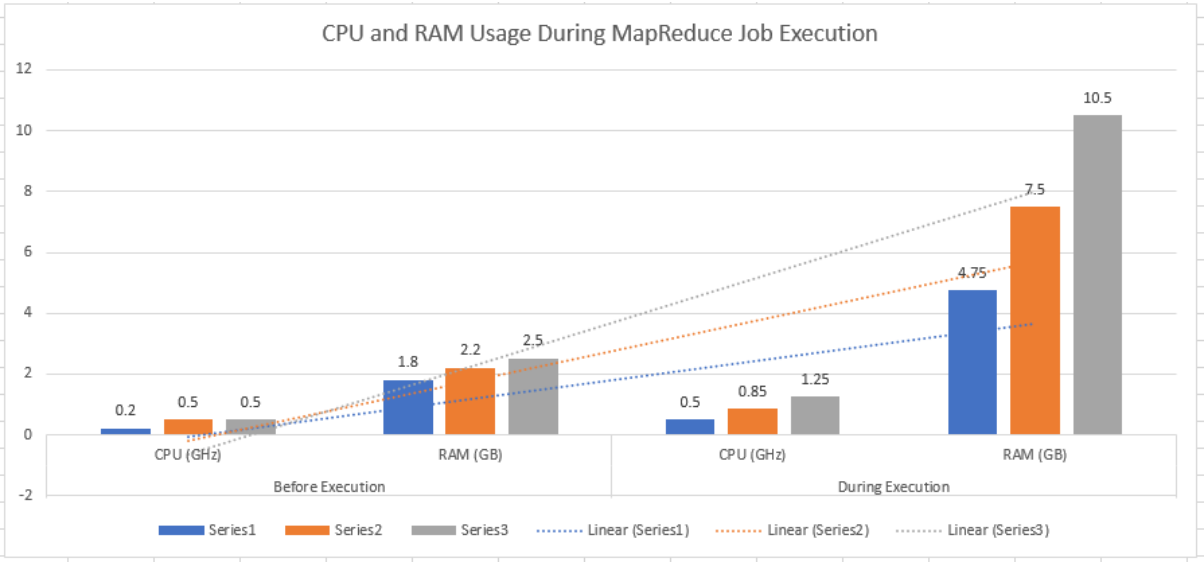
The above calculation was performed on the machine with following configuration:

Processor	Intel i5, 2.2 GHz
Installed RAM	12 GB
Operating System	Windows 64 bit

The below table gives information about the CPU and RAM consumption happened while running the above MapReduce jobs.

		Before Execution		During Execution	
Data Sets	Maps	CPU (GHz)	RAM (GB)	CPU (GHz)	RAM (GB)
Set 1 (1990)	8	0.2	1.8	0.5	4.75
Set 2 (1990 and 1992)	60	0.5	2.2	0.85	7.5
Set 3 (1990, 1991, 1992 and 1993)	115	0.5	2.5	1.25	10.5

One can make out from the below graph that the CPU and RAM consumption has gradually increased from set 1 to set 3 linearly. Series 1 represents Set1, Series 2 represents Set2 and Series 3 represents Set 3. Set 1 has 8 maps, Set 2 has 60 maps and Set 3 has 115 maps.



The below table displays the start and end time of the above-mentioned MapReduce jobs along with total execution time. From the graph below, we can see that the time taken has gradually increased along with the size of the data sets.

Job_id	Job Name	Start Time (HH:MM:SS)	End Time (HH:MM:SS)	Total Execution Time (Minutes)	Total Maps	Total Reduce
job_1486529623047_0002	MaxTemperature	05:02:23	05:03:52	1.48	8	1
job_1486529623047_0003	MaxTemperature	05:21:33	05:47:34	26.02	60	1
job_1486529623047_0004	MaxTemperature	06:24:00	07:07:52	43.87	115	1
job_1486529623047_0007	MaxTemperatureWithCombiner	07:42:39	07:43:58	1.32	8	1
job_1486583682390_0002	MaxTemperatureWithCombiner	20:17:37	20:37:24	19.78	60	1
job_1486583682390_0003	MaxTemperatureWithCombiner	21:14:01	22:03:28	49.45	115	1