The 3 different run times are seen below:

*Fig 1

Completed Applications (4)

Application ID	Name	Cores	Memory per Executor	Resources Per Executor	Submitted Time	User	State	Duration
app-20230222041615-0003	Page Rank	2	4.0 GiB		2023/02/22 04:16:15	ubuntu	FINISHED	19 min
app-20230222033150-0002	Partition Page Rank	4	4.0 GiB		2023/02/22 03:31:50	ubuntu	FINISHED	4.6 min
app-20230222032409-0001	Partition Page Rank	4	4.0 GiB		2023/02/22 03:24:09	ubuntu	FINISHED	4.4 min
app-20230222030819-0000	Page Rank	4	4.0 GiB		2023/02/22 03:08:19	ubuntu	FINISHED	11 min

As you can see the 3 different tasks have much different times. Task 1 is at the bottom and it is 11 minutes. Then with Task 2 breaking the dataset down into partitions drop the runtime to about 4.5 minutes (a mixture of the two middle run times). Finally in task 3 when we kill the second worker during the process the runtime increases to 19 minutes.

A big reason for the difference in runtime between Task 1 and Task 2 is the decrease in tasks that need to be accomplished.

*Fig 2

Active Jobs (1)

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
0	runJob at PythonRDD.scala:166 runJob at PythonRDD.scala:166 (kill)	2023/02/22 04:16:21	2.8 min	1/23	2/265 (2 running)

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

*Fig 3

Active Jobs (1)

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

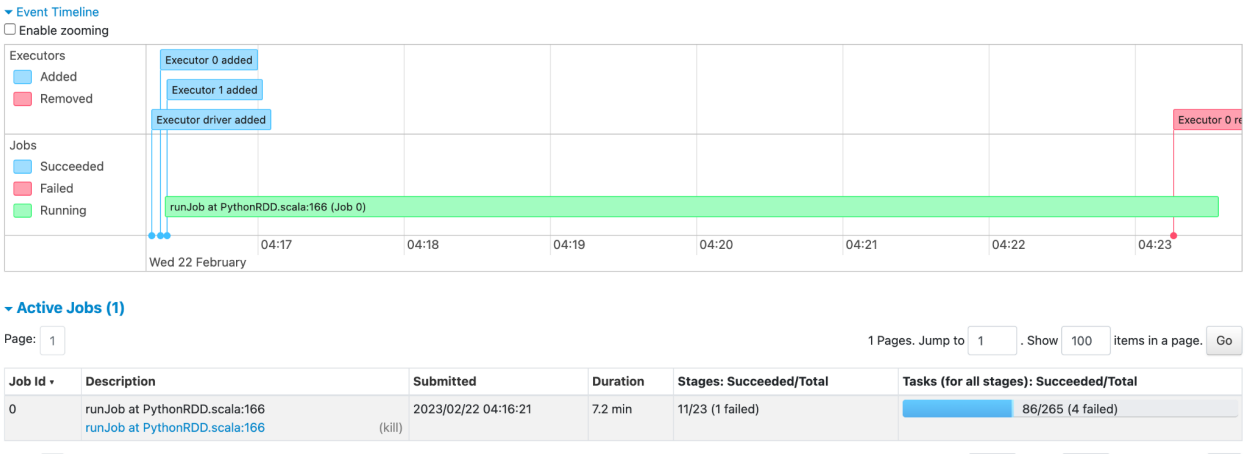
Job Id	Description	Submitted	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total
0	runJob at PythonRDD.scala:166 runJob at PythonRDD.scala:166 (kill)	2023/02/22 04:57:15	26 s	0/14	4/196 (2 running)

Page: 1 — 1 Pages. Jump to 1 . Show 100 items in a page. Go

As you can see from Figures 2 and 3 above is that in Fig 2 from task 1 there are 265 tasks to complete, but once you add in the partitions it drops down to 196 tasks in Fig 3.

For Task 3 we saw a big jump in time from Task 1 (19 minutes from 11 minutes, as seen in Figure 1). At first this would seem to be wrong because I used the same .py file for both of these tasks, and they both ahd the same number of tasks as seen in Fig 4.

*Fig 4



However, if we go back to FIg 1 and look at the number of Cores we see that in Task 1 at the bottom there are 4 cores, but at the top for task 3 there are only 2 cores because we kill the second worker. We can also see in Fig 4 the red flag where the worker is killed.

So, as we can see there are much different run times for the three tasks but there are obvious reasons for these differences.