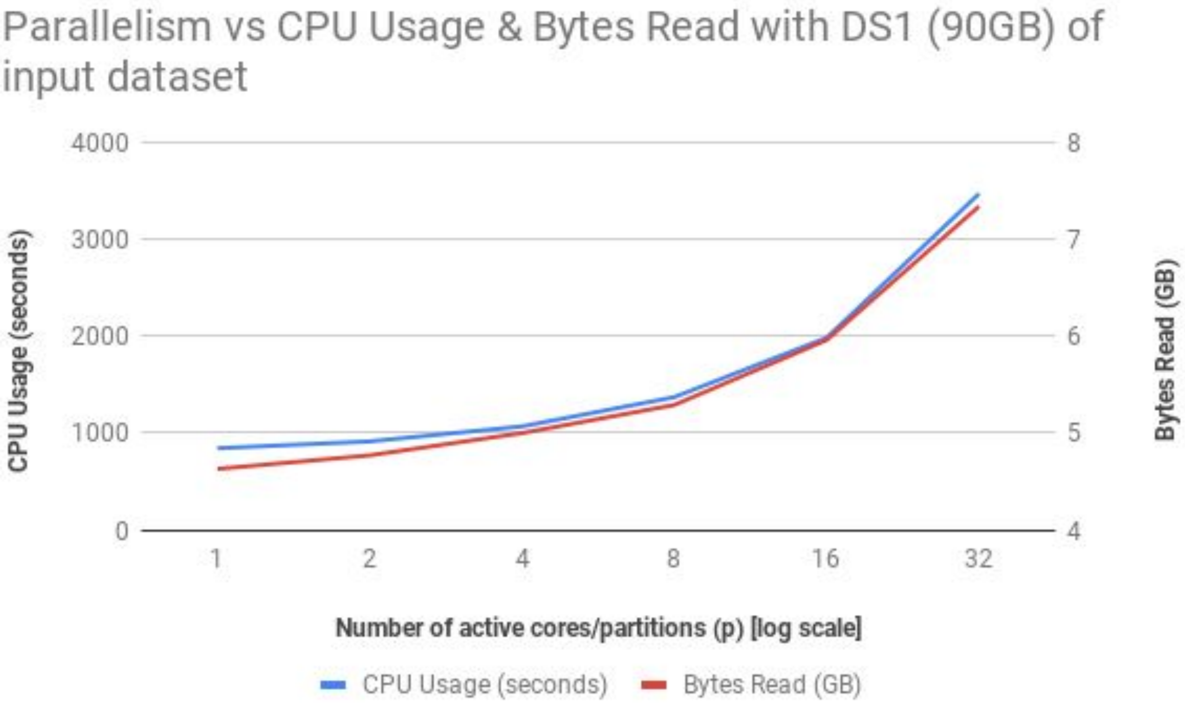
RDataFrame: task splitting

SWAN TOTEM Helix Nebula Project

Problem: more partitions, more data read



Credit: Prasanth Kothuri

Reading one file remotely (local tests)

- `collectd` to monitor data read through the network
- one single file from DS1: `TotemNTuple_9874.ntuple.root`
- partitioning by events

Number of partitions	**Total number of KB read**
1 partition	30.000 KB
2 partitions	46.000 KB
4 partitions	63.000 KB

Reading one file remotely (local tests)

- `collectd` to monitor data read through network
- one single file from DS1: `TotemNTuple_9874.ntuple.root`
- partitioning by events - **using ROOT clusters boundaries**

Number of partitions	**Total number of KB read**
1 partition	31.000 KB
2 partitions	46.000 KB
4 partitions	60.000 KB

Low-level behaviour

- same single file from DS1: `TotemNTuple_9874.ntuple.root`
- **same results reading from local disk**

Range of events	**Total number of KB read**
0 - 1	4200 KB
0 - 1400	4200 KB
0 - 148**7**	4200 KB
0 - 148**8**	52000 KB
1487 - 2975	4600 KB

To be understood: Pending to check with the IO experts