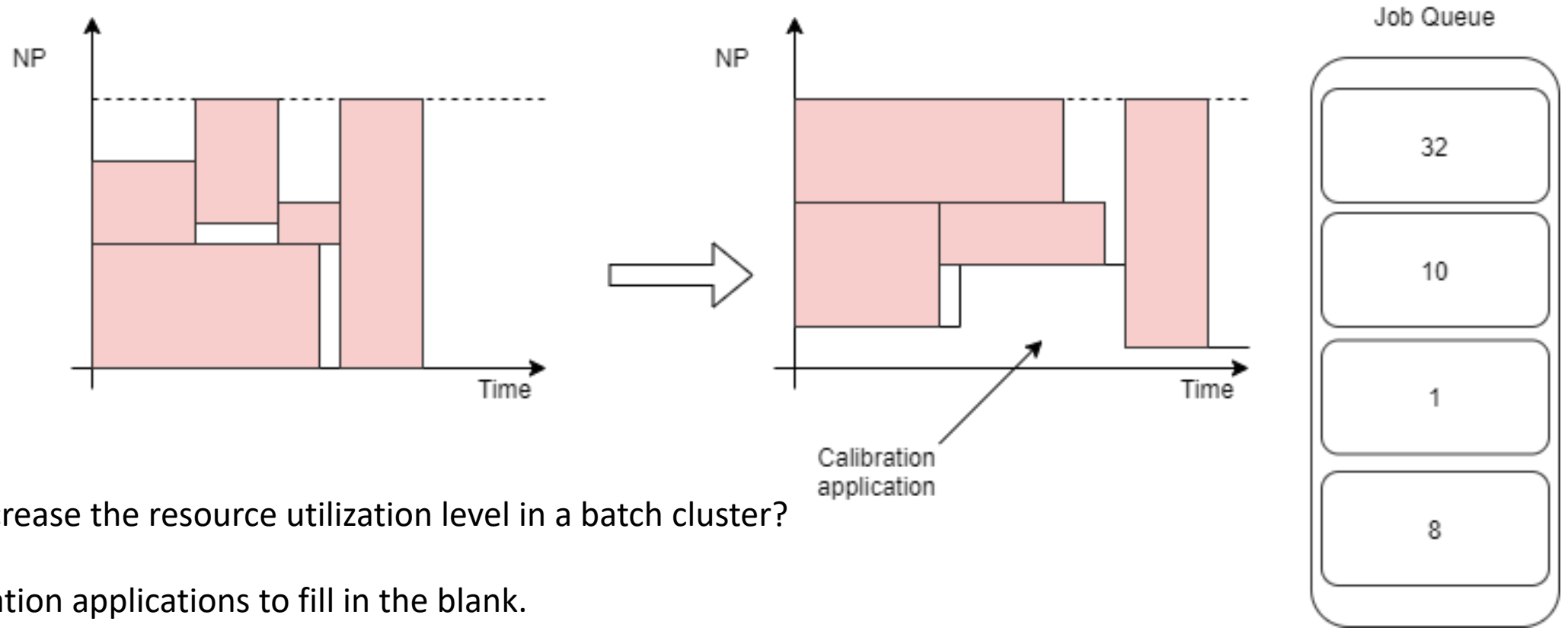
Self adjusted auto provision system at resource level

Weekly report 8st July

You Hu

Recap – issue and research question



How to increase the resource utilization level in a batch cluster?

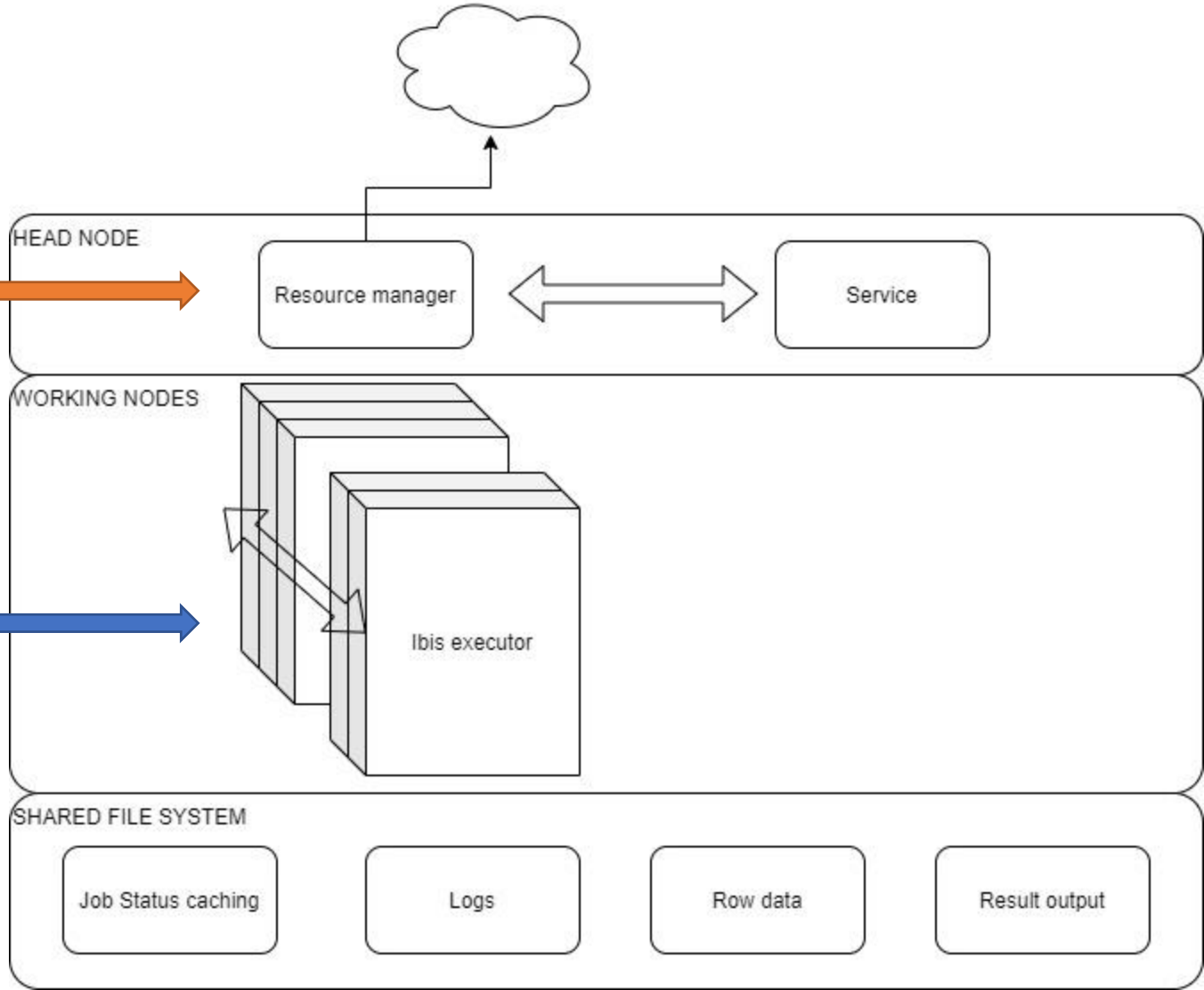
Use calibration applications to fill in the blank.

It requires: auto scaling/provisioning;

Design-current layout

Auto scaling algo

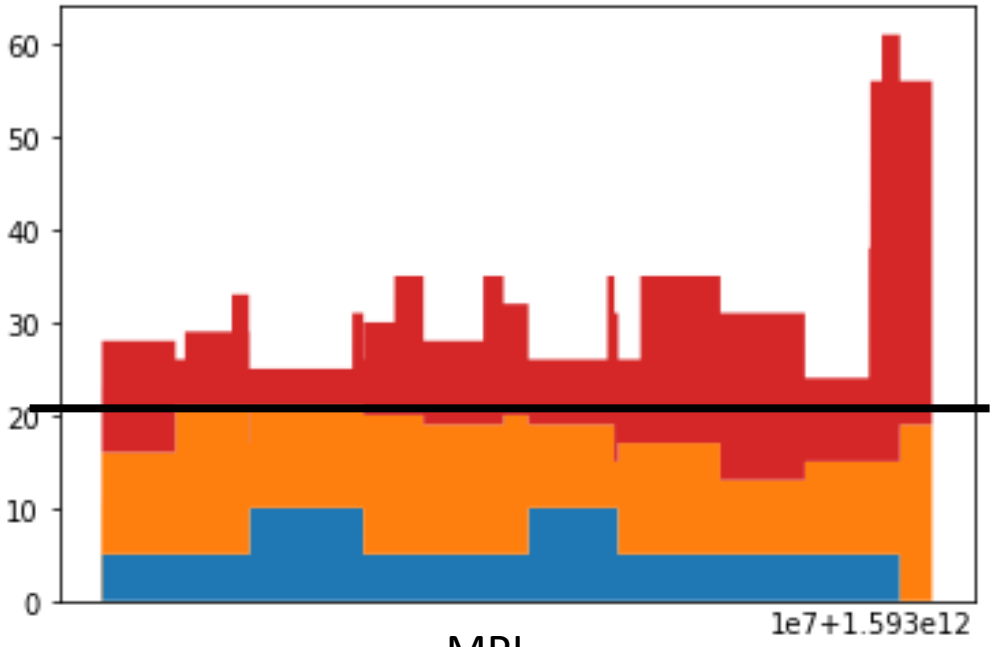
Elastic distributed computing
(Master/Worker)



Auto scaling simulation

- Calibration resource: 45.2% increase
- Other resource: 7.0% decrease
- Total resource: 10.6% increase

In total check point:39012
In total Calibration use:235280.0 AVG:6.030964831333948
In total Normal use:464868 AVG:11.91602583820363
In total use:700148.0 AVG:17.946990669537577



MPI

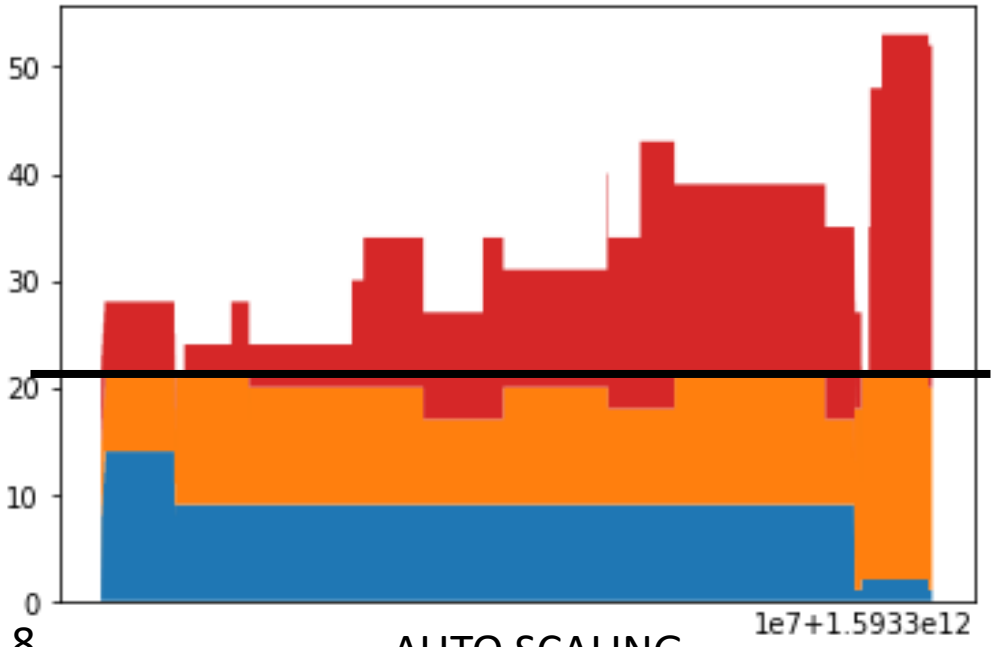
Pending jobs

Other Jobs

Calibration

Fixed mini node = 8

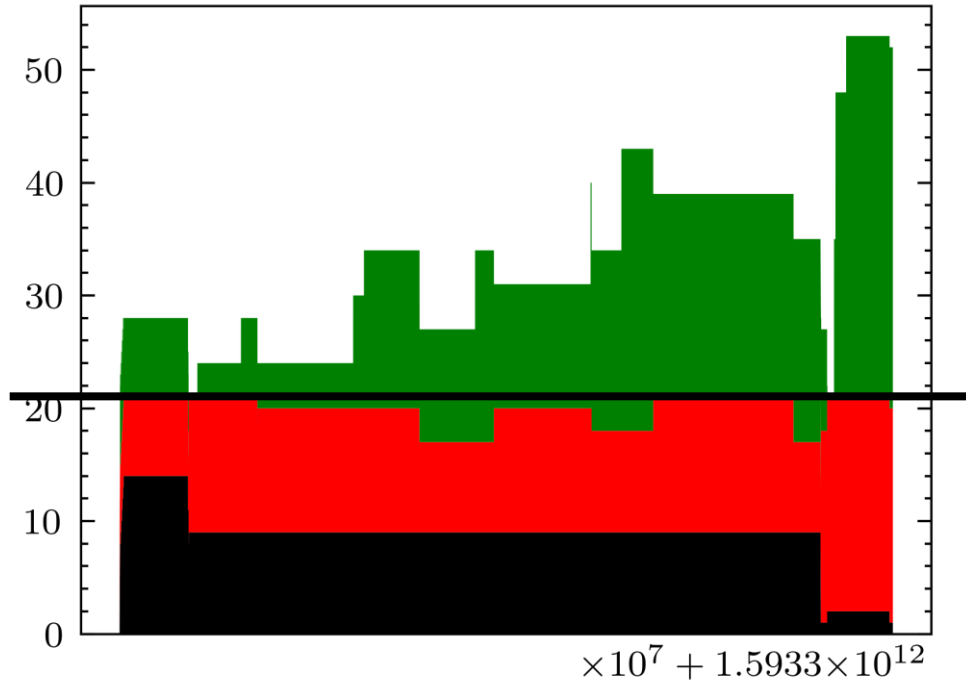
In total check point:39062
In total Calibration use:342067.0 AVG:8.757027289949312
In total Normal use:433084 AVG:11.087092314781628
In total use:775151.0 AVG:19.844119604730942



AUTO SCALING

Auto scaling simulation

- Issues:
  - The one last node is not taken by the calibration application
  - Reduce the limit of (Xenon) job running time, does not help to increase the performance(node/sec)
  - Reduce the mini number does not help as well
- Possible reasons
  - Auto-scaling algorithm still hidden problem
  - The given test case can not give a chance to show the mechanism



AUTO SCALING

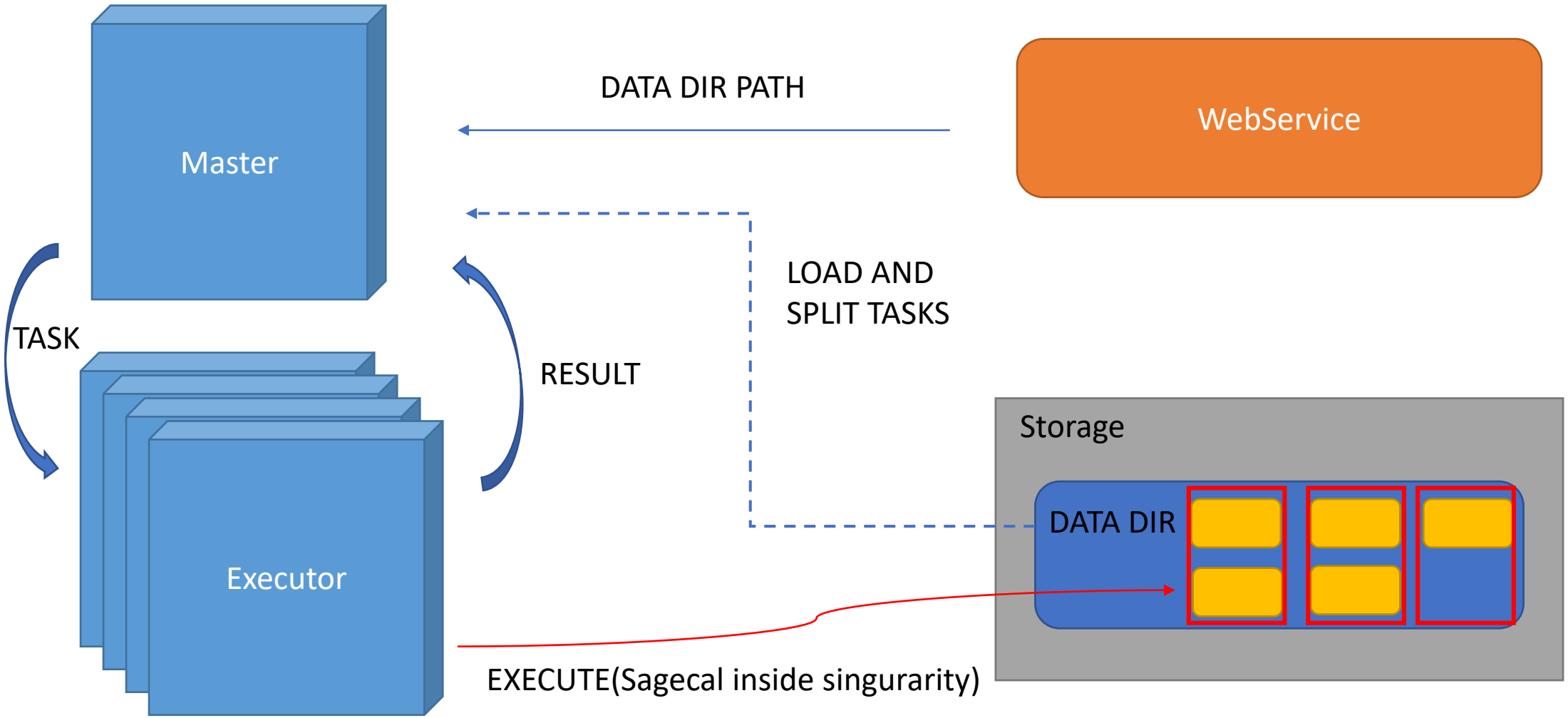
How data processed

```
→  Calibration time singularity exec ~/DynPrvDriver/AppContainers/Sagecal/SagecalContainer.simg /opt/sagecal/bin/sagecal -d DATA0 -s 3c196.sky.txt -c 3c196.sky.txt.cluster -n 4 -t 10 -p sm.ms.solutions -e 4 -g 2 -l 10 -m 7 -x 30 -F 1 -j 5  -k -1 -B 1 -w
SAGECal 0.7.1 (C) 2011-2020 Sarod Yatawatta
 MS: DATA0
Selecting baselines > 30 and < 1e+08 wavelengths.
Using Robust noise model for solver with degrees of freedom [2,30].
Stations: 61 Baselines: 1830
Integration Time: 10.0139 s, Total timeslots: 125
Phase center (2.15374, 0.841552)
Only one MS
Got 2 clusters
Total effective clusters: 3
For 10 samples, solution time interval (s): 100.139
Freq: 152.733 MHz, Chan: 1 Bandwidth: 0.183105 MHz
nu=30
Timeslot: 10 Residual: initial=0.0870613,final=3.73346e-07, Time spent=0.0666667 minutes
nu=30
Timeslot: 20 Residual: initial=5.83557e-07,final=1.91819e-07, Time spent=0.0333333 minutes
nu=30
Timeslot: 30 Residual: initial=2.49905e-07,final=1.48017e-07, Time spent=0.0166667 minutes
nu=30
Timeslot: 40 Residual: initial=1.53844e-07,final=1.43131e-07, Time spent=0.0333333 minutes
nu=30
Timeslot: 50 Residual: initial=1.50543e-07,final=1.41432e-07, Time spent=0.0333333 minutes
nu=30
Timeslot: 60 Residual: initial=1.47163e-07,final=1.37958e-07, Time spent=0.0166667 minutes
nu=30
Timeslot: 70 Residual: initial=1.49591e-07,final=1.39044e-07, Time spent=0.0166667 minutes
nu=30
Timeslot: 80 Residual: initial=1.45653e-07,final=1.41048e-07, Time spent=0.0333333 minutes
nu=30
Timeslot: 90 Residual: initial=1.4735e-07,final=1.43614e-07, Time spent=0.0166667 minutes
nu=30
Timeslot: 100 Residual: initial=1.45031e-07,final=1.42376e-07, Time spent=0.0333333 minutes
nu=30
Timeslot: 110 Residual: initial=1.45285e-07,final=1.42938e-07, Time spent=0.0166667 minutes
nu=30
Timeslot: 120 Residual: initial=1.4827e-07,final=1.44877e-07, Time spent=0.0333333 minutes
Warning: Missing rows, got 9455 expect 18300 +- 610. (probably the last time interval, so not a big issue).
nu=30
Timeslot: 130 Residual: initial=4.23302e-08,final=2.58443e-08, Time spent=0.0166667 minutes
Done.
singularity exec ~/DynPrvDriver/AppContainers/Sagecal/SagecalContainer.simg    27.88s user 1.82s system 126% cpu 23.493 total
```

Current data & speed

- Dataset sm.ms https://github.com/nlesc-dirac/data
  - Size 69M
- Speed:
  - -n 2: 31s
  - -n 4: 26s
  - -n 6: 20.7s
- Work around:
  - Big/mid/small data set = duplicate sm.ms x 500/150/50
  - The assumption behind: an observation can be divided into sub data sets
- Limit:
  - 40 GB disk quotum on DAS5
  - No Sky Map for sagecal

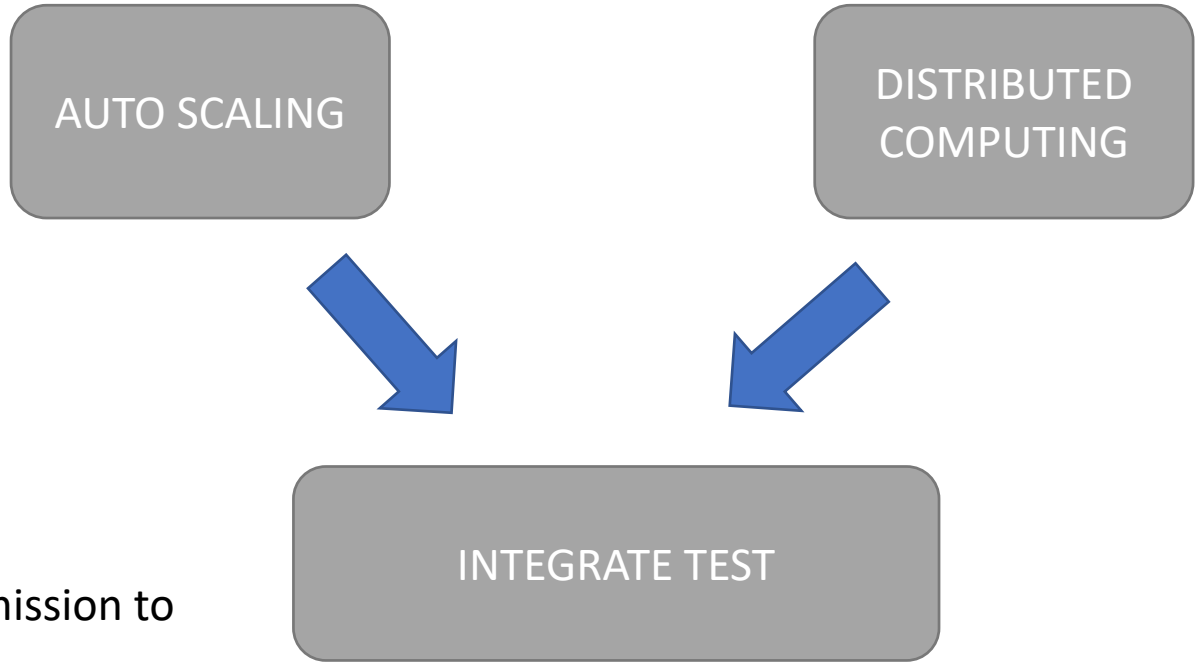
DynPrvDriver – compute layer



DynPrvDriver – compute layer

- Issues:
  - Waste of compute resource: one node as master
    - Possible solution: if it is a master, create a new thread or process as Ibis executor(be careful about limiting the CPU usage)
  - How AssumeDied works?
    - Currently, AssumeDied+signalhandler is used for fast recover, but the function is not clear

Next phase: performance test



Integrate test:

- Dynamic mini node
- Log the time from submission to the end

Overall problems

- Data set(from Souley or fetch it from LOAFR website by myself)
  - Including: observation and sky map
  - L246909_SAP006_SB486_uv_001.MS_02134b61.tar + 3c196.sky.txt?
  - L250048_SB342_uv.dppp.MS_32c7efc3.tar
- Capability of singularity:
  - LU site: 3.4 headnode,2.6 work node
  - VU site: 3.4 head and work node (8888 port occupied)
- DDL: 31st August for grading or submission?
  - I can extend for 2 months, new rules, Non-EU student 2K a year now
  - Second reader?
- Contract with eScience Center
  - At least a record from eScience Center, showing I used to be there ☺