Marcel Trattner, Tim Oelkers

Radio2022 Symposium and GLOW Annual Assembly

Memory-based Computing in Astronomy

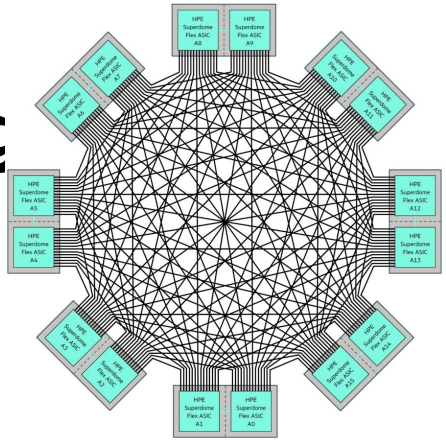
Hochschule für Technik und Wirtschaft Berlin
University of Applied Sciences

Big Data @ SKA

- Square Kilometre Array (SKA)
- Raw data: 1.3 Exabyte per day
- ~1 Petabyte per single image
- Novel demands on computing

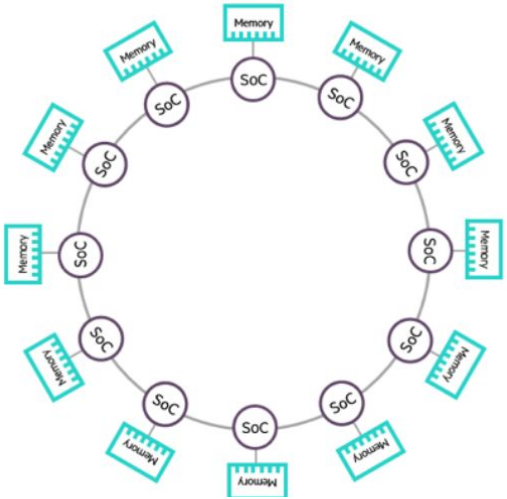
[1]

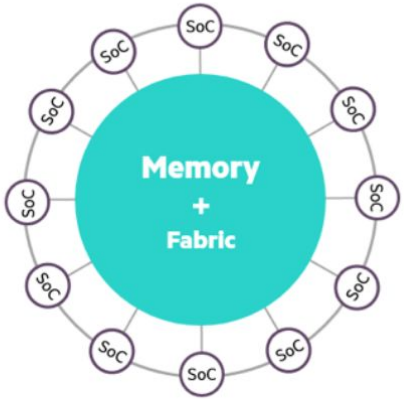
Memory-Based Computing



[2]

- Challenge: Memory Wall Problem
- Memory is in the center the architecture instead the processor
- Managed by **Librarian File System** (LFS)

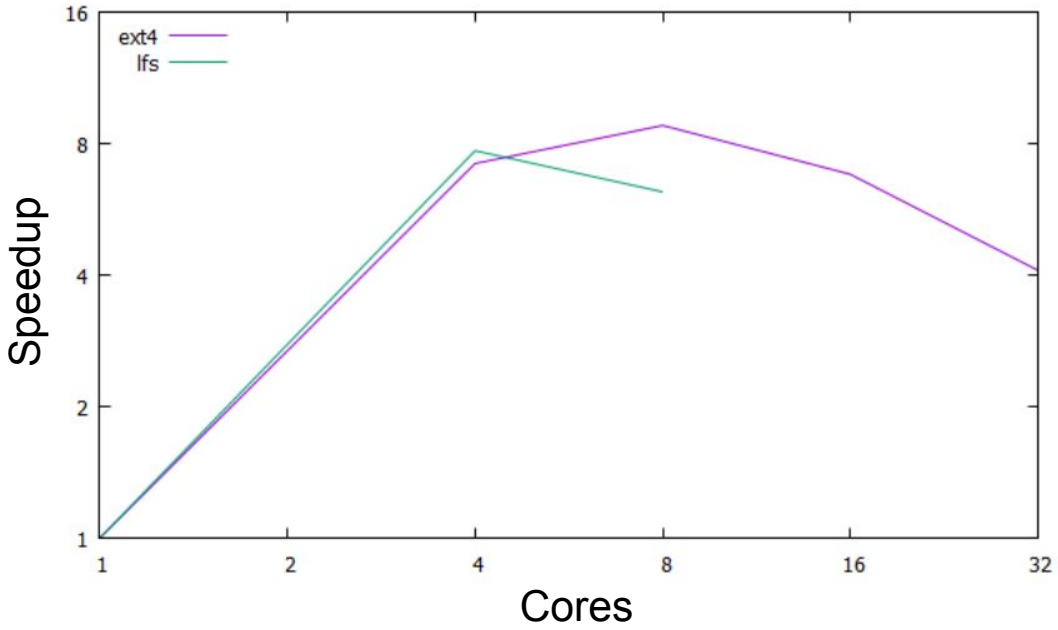




[3]

Performance of CASA 5.1

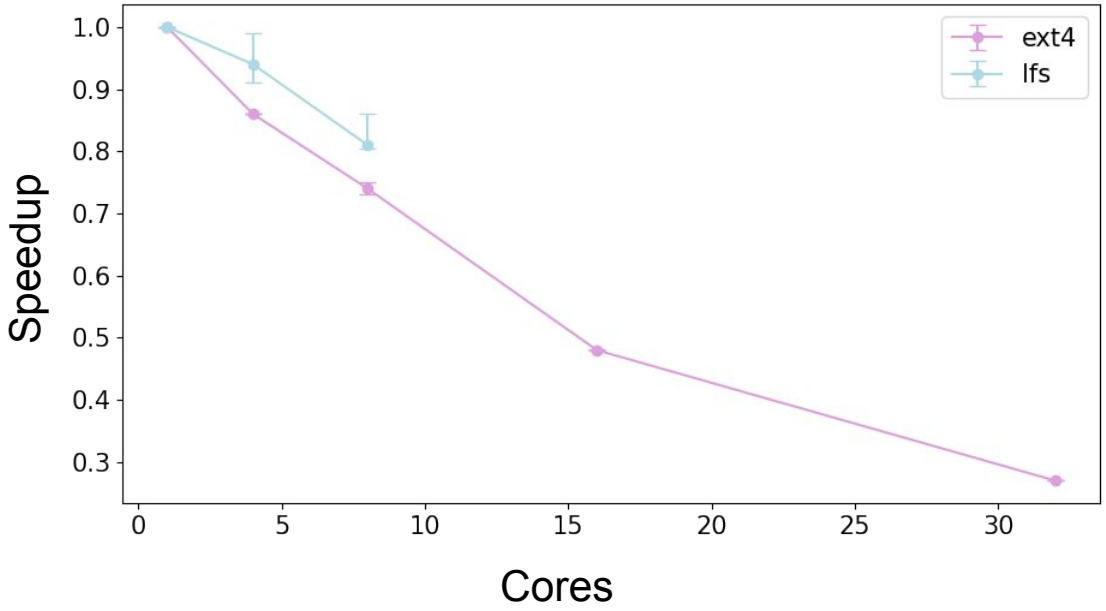
- Running tclean on Sandbox
  - Continuum imaging
  - No parallelization of minor cycle
- Speedup: S(n) = T(1)/T(n)
  - ext4: maximum at n = 8 cores
  - LFS: CASA crashes for n > 8 cores



[4]

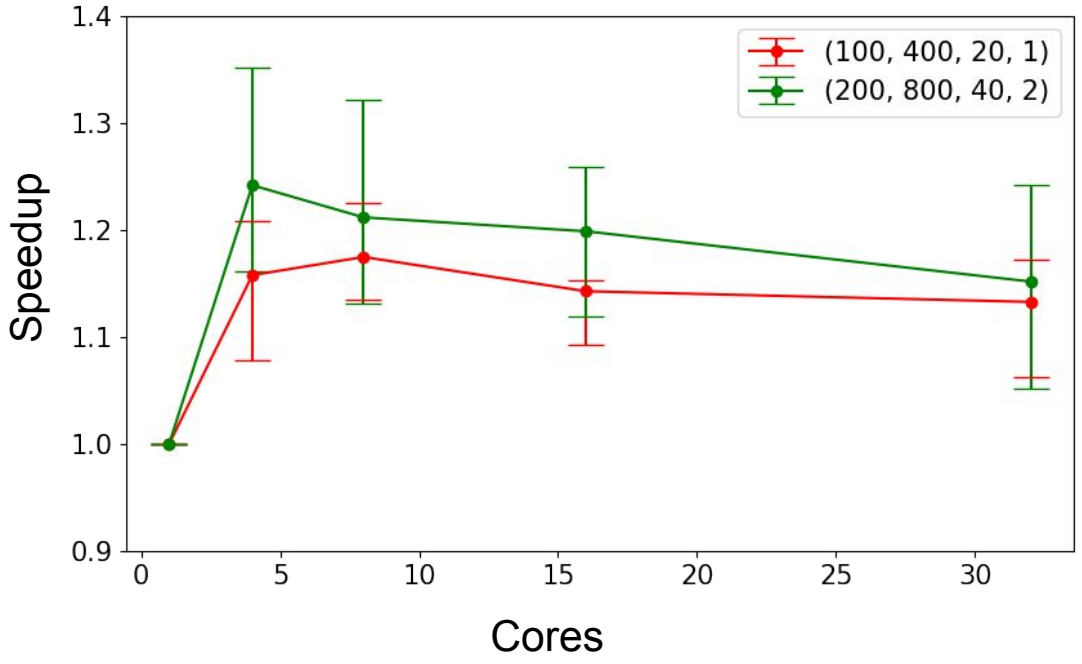
Performance of CASA 6.4

- CASA 6.4: parallelization revised (e.g cube imaging)
- Running tclean on Sandbox
  - Continuum imaging
  - no parallelization of minor cycle
- Speedup: S(n) = T(1)/T(n)
  - ext4: no speedup
  - LFS: CASA crashes for n > 8 cores



Performance of CASA Next Generation Infrastructure (CNGI)

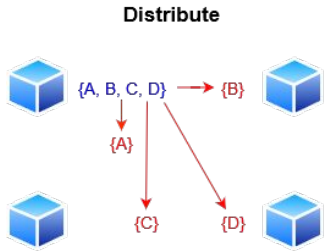
- CNGI: work in progress
- Running the subcomponent convert_ms on Sandbox
  - Conversion of MS to ZARR format
  - Parameters: (time, baseline, channel, polarization)
  - Almost no speedup

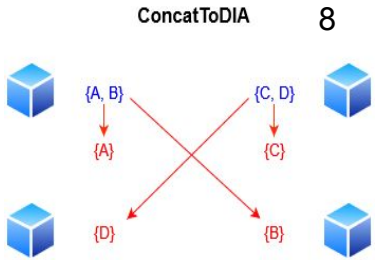


Thrill

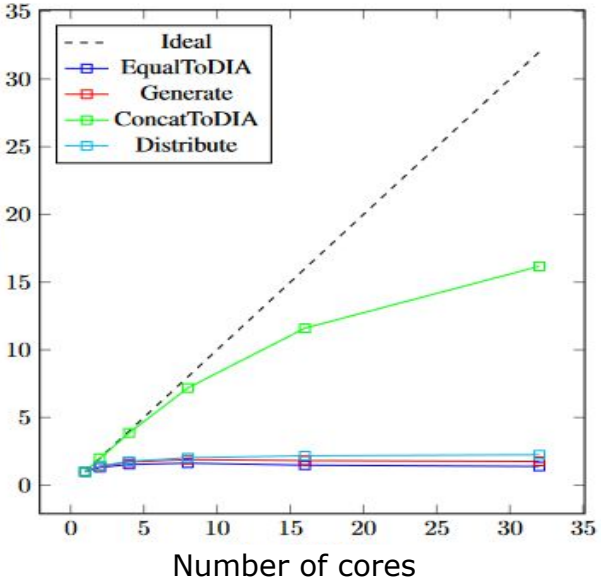
- Parallel, distributed processing of data based on the MapReduce paradigm
- OpenSource project (BSD-2 License) developed at KIT
- Experimental status, missing features (fault tolerance)
- Similar Frameworks: Apache Hadoop, Flink, Spark
  - Thrill several times faster

Thrill - Evaluation

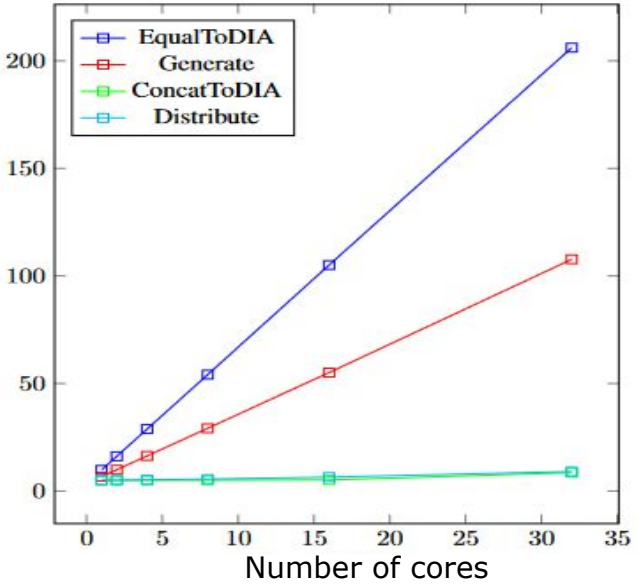




Speedup (T_1 / T_n)



Maximum memory usage in GB



Sources

Images:

[1] SKA Observatory. SKAO Prospectus 2021. 2022. URL: https://www.skatelescope.org/wp-content/uploads/2022/01/SKAO-Prospectus-Jan-2022-reduced.pdf

[2] https://www.hpe.com/psnow/doc/A00036491ENW.pdf?jumpid=in_lit-psnow-getpdf

[3] https://www.hpe.com/us/en/newsroom/blog-post/2017/05/memory-driven-computing-explained.html

[4] Elsa Buchholz. “Untersuchung eines astronomischen Workflows im memory-based Computing”. MA thesis. University of Applied Sciences, 2020.

[5] Timo Bingmann, Thrill Tutorial, https://panthema.net/2020/0601-thrill-tutorial/slides-20200601-thrill-tutorial.pdf

Hochschule für Technik
und Wirtschaft Berlin

University of Applied Sciences

www.htw-berlin.de