Integrating Globus and MapReduce for Out-of-computer Analysis of Peta-scale CFD Data

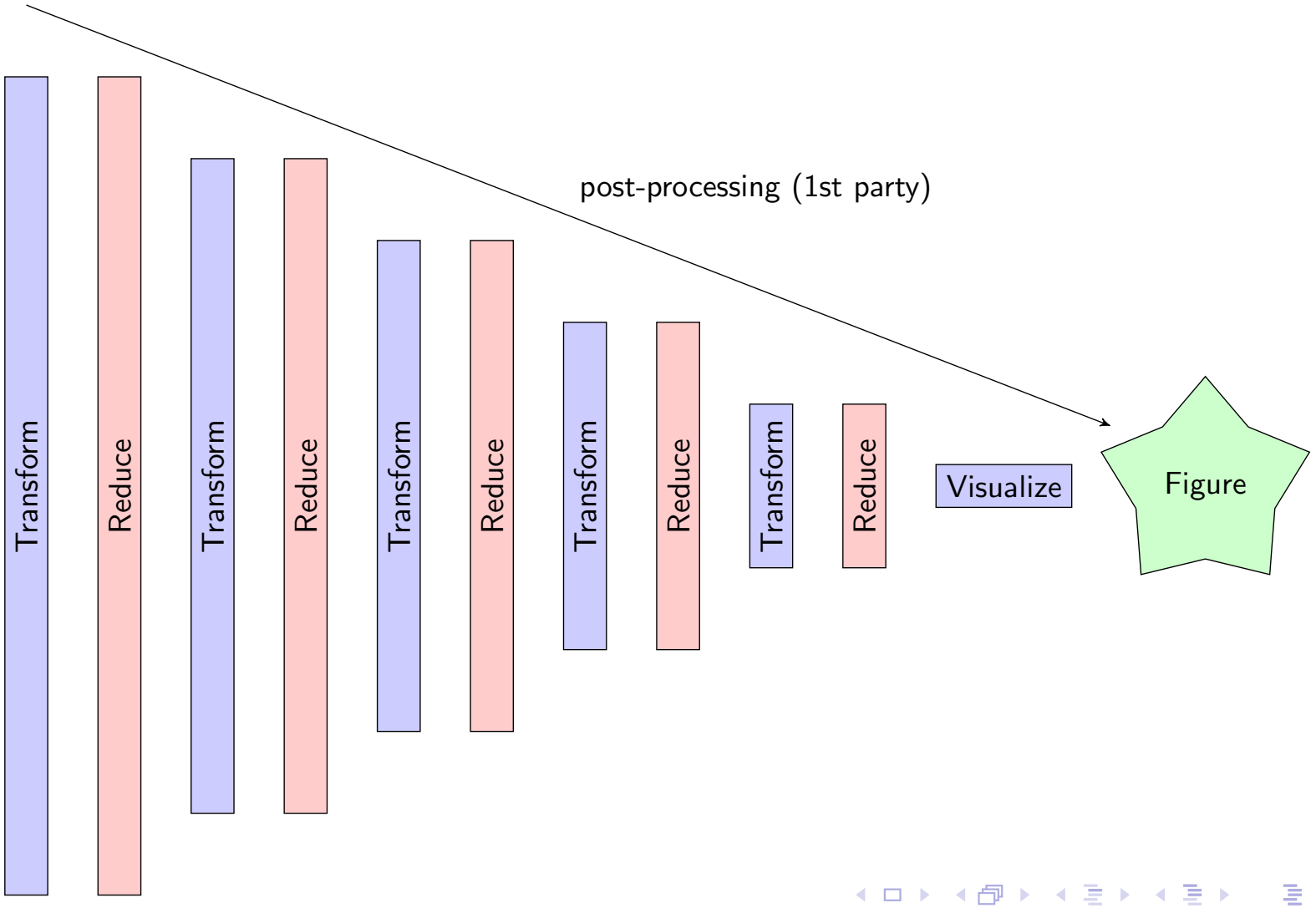
Max Hutchinson

University of Chicago

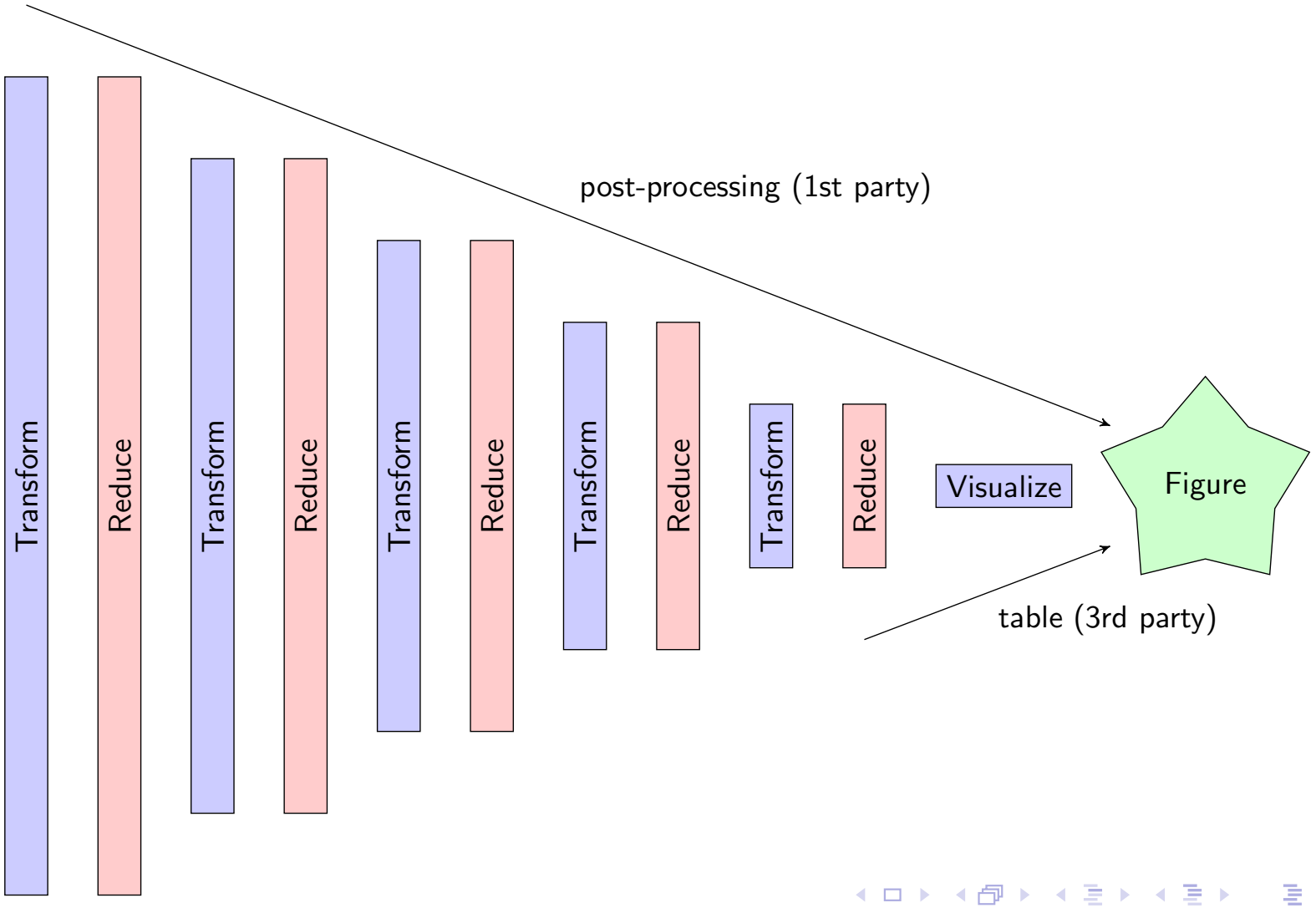
April 15th, 2015

THE UNIVERSITY OF CHICAGO

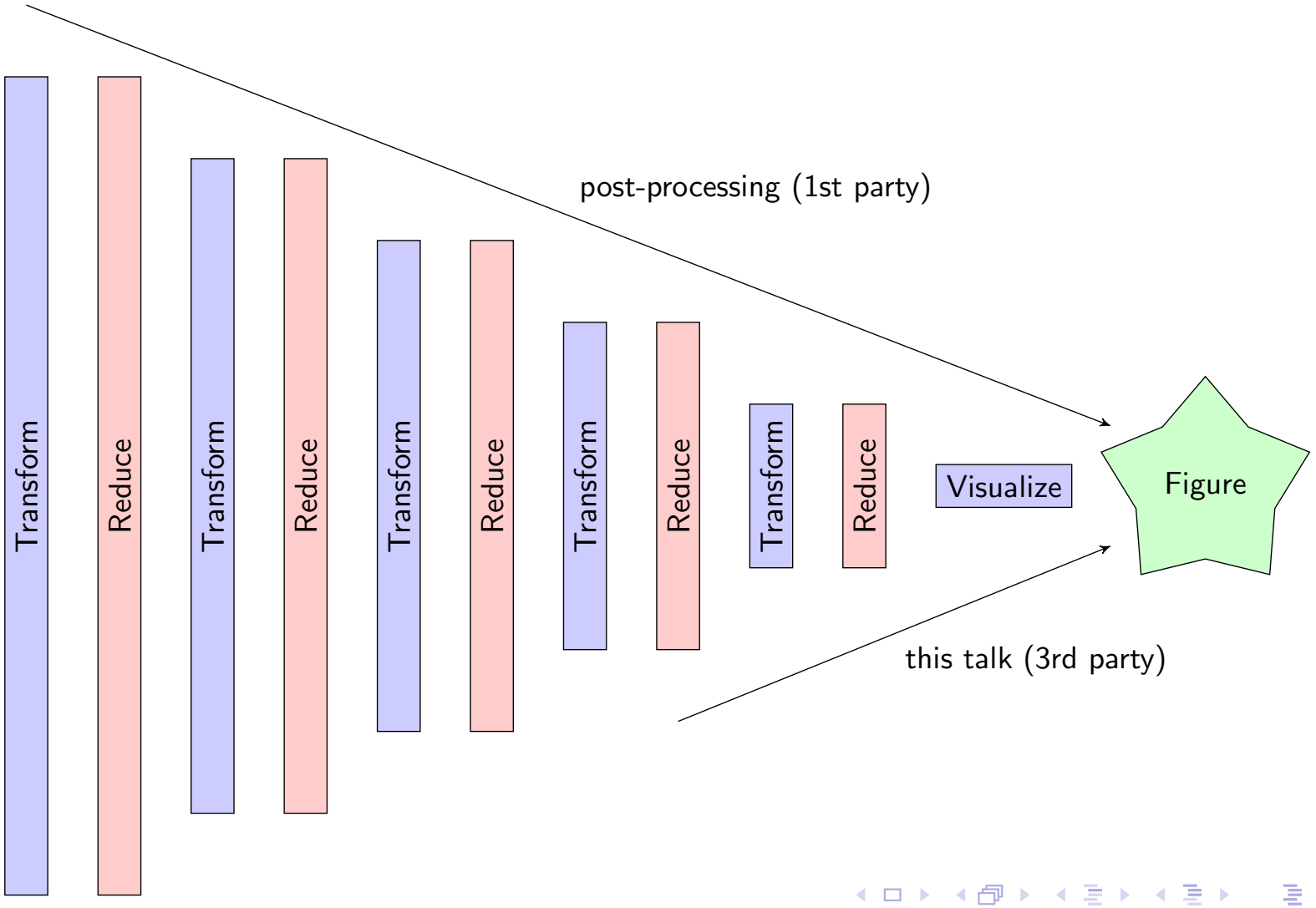
Post-processing is a series of transforms and reductions



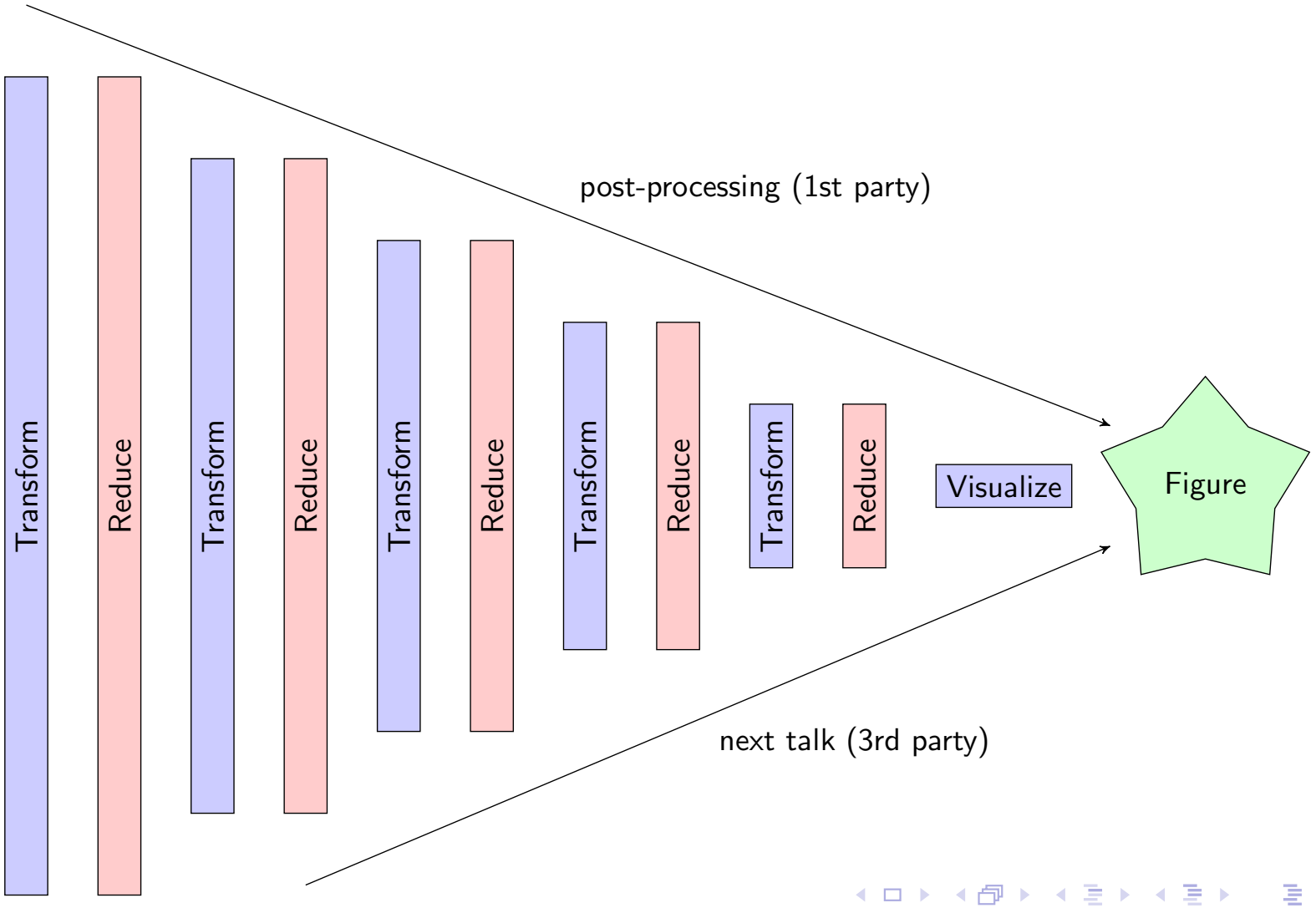
Post-processing is a series of transforms and reductions



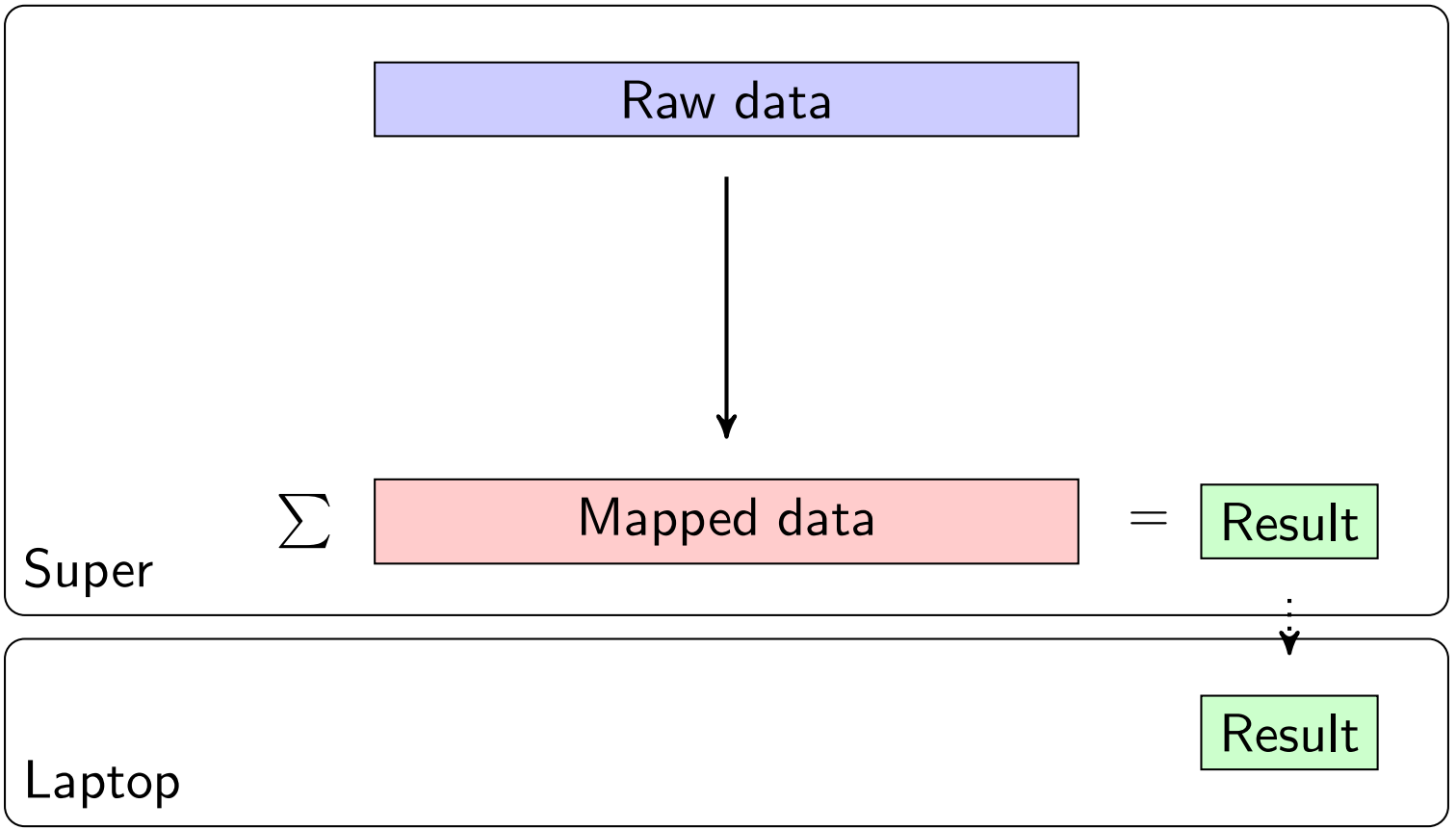
Post-processing is a series of transforms and reductions



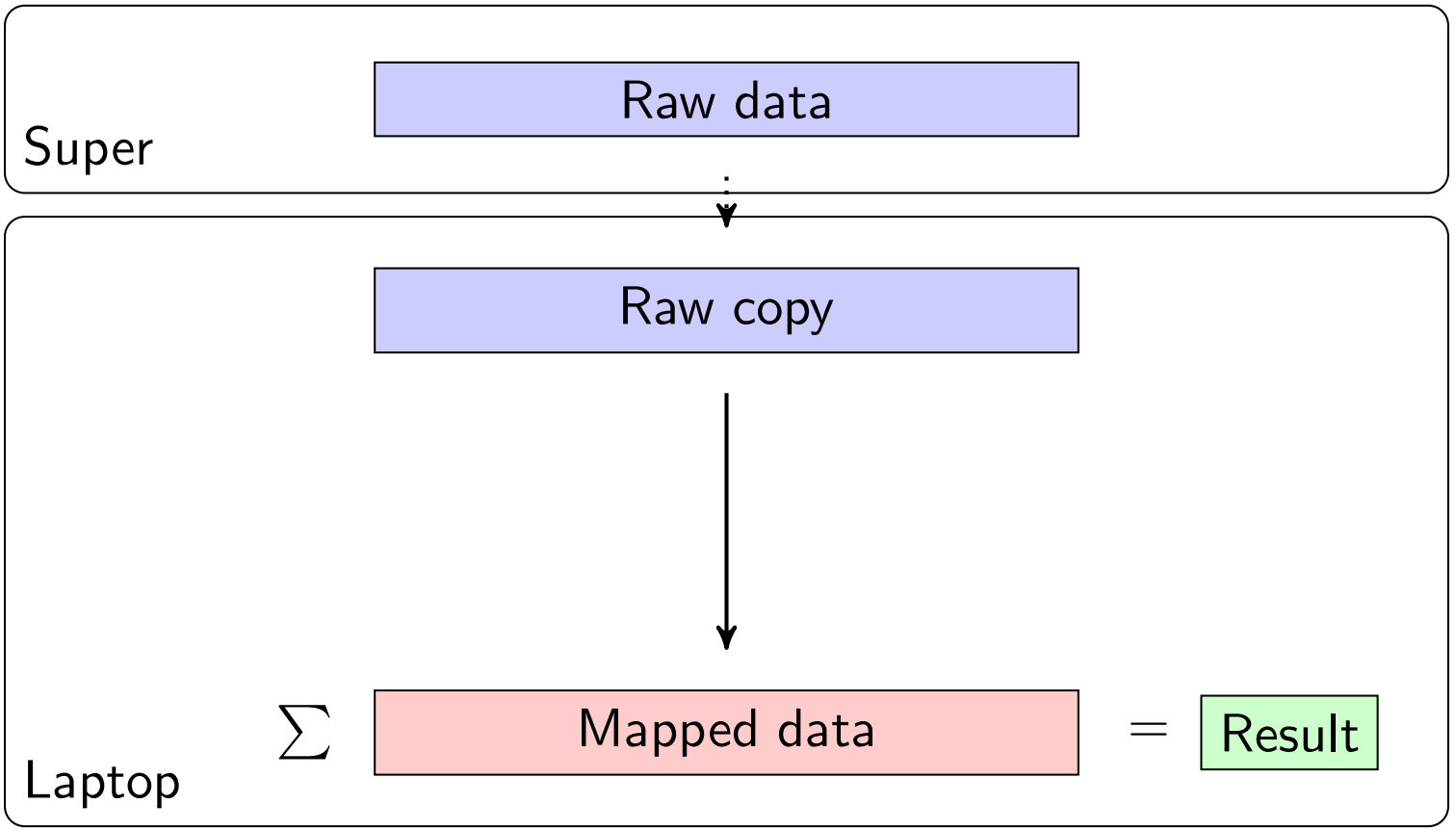
Post-processing is a series of transforms and reductions



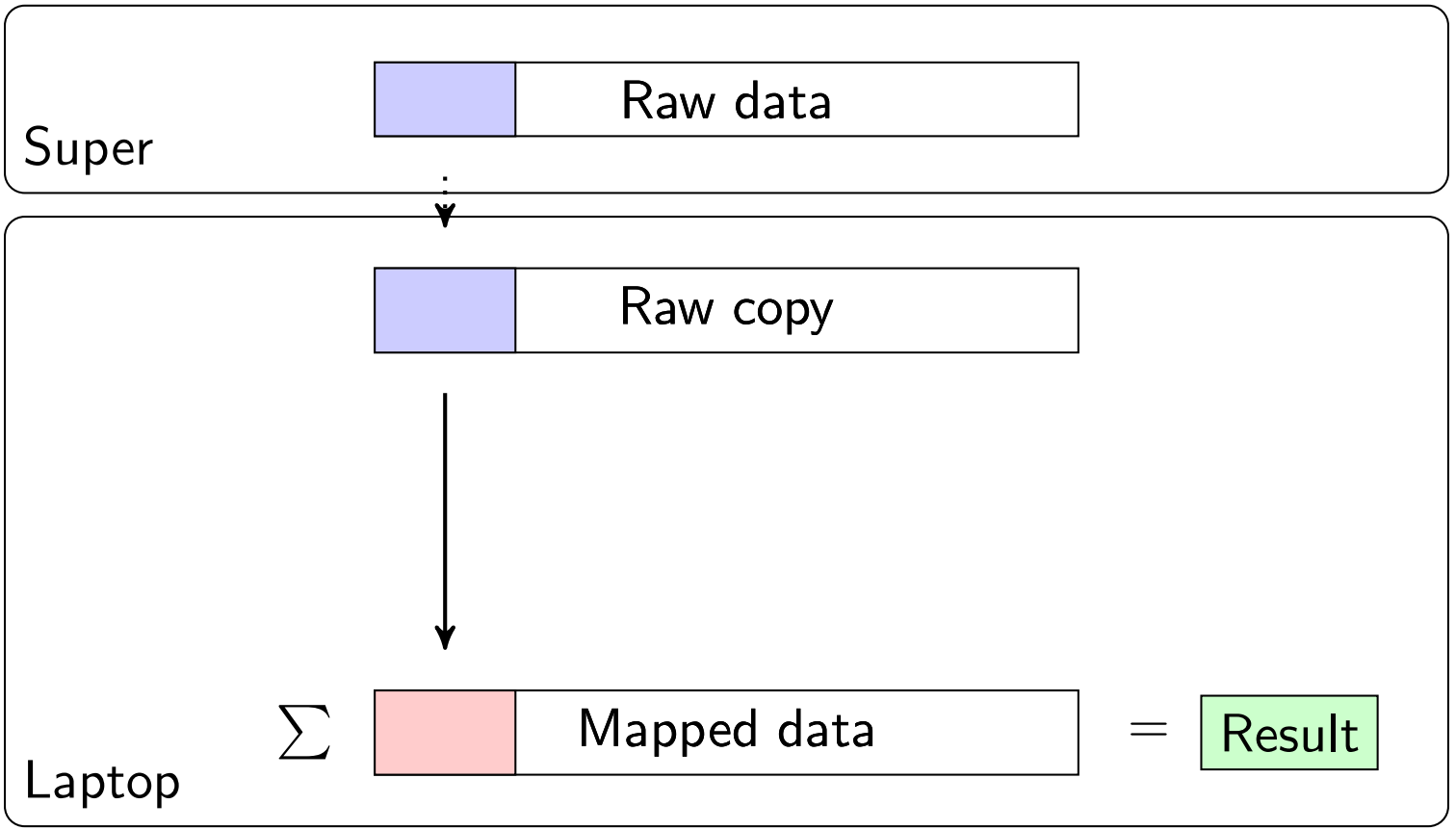
MapReduce post-processing



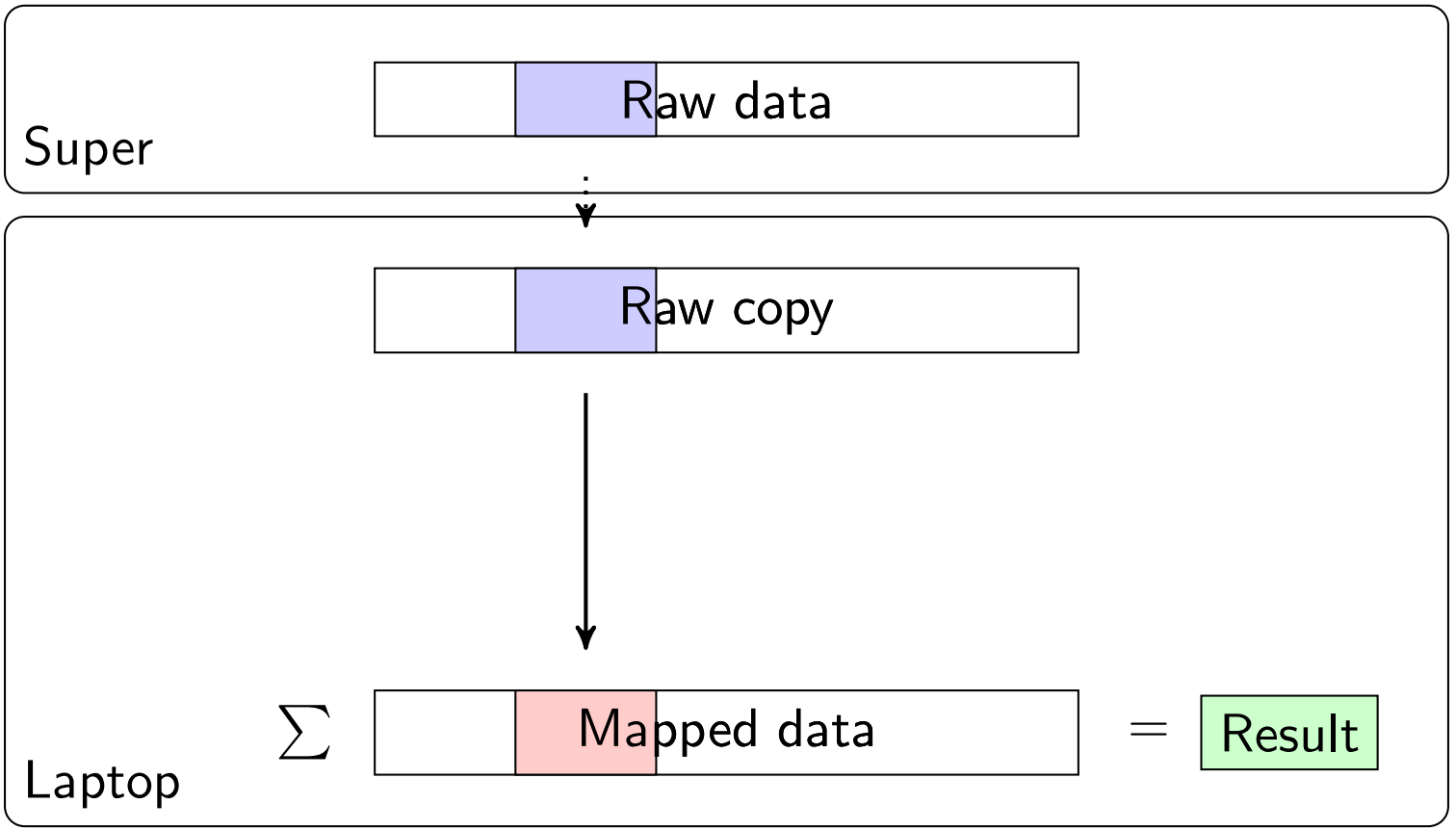
MapReduce post-processing



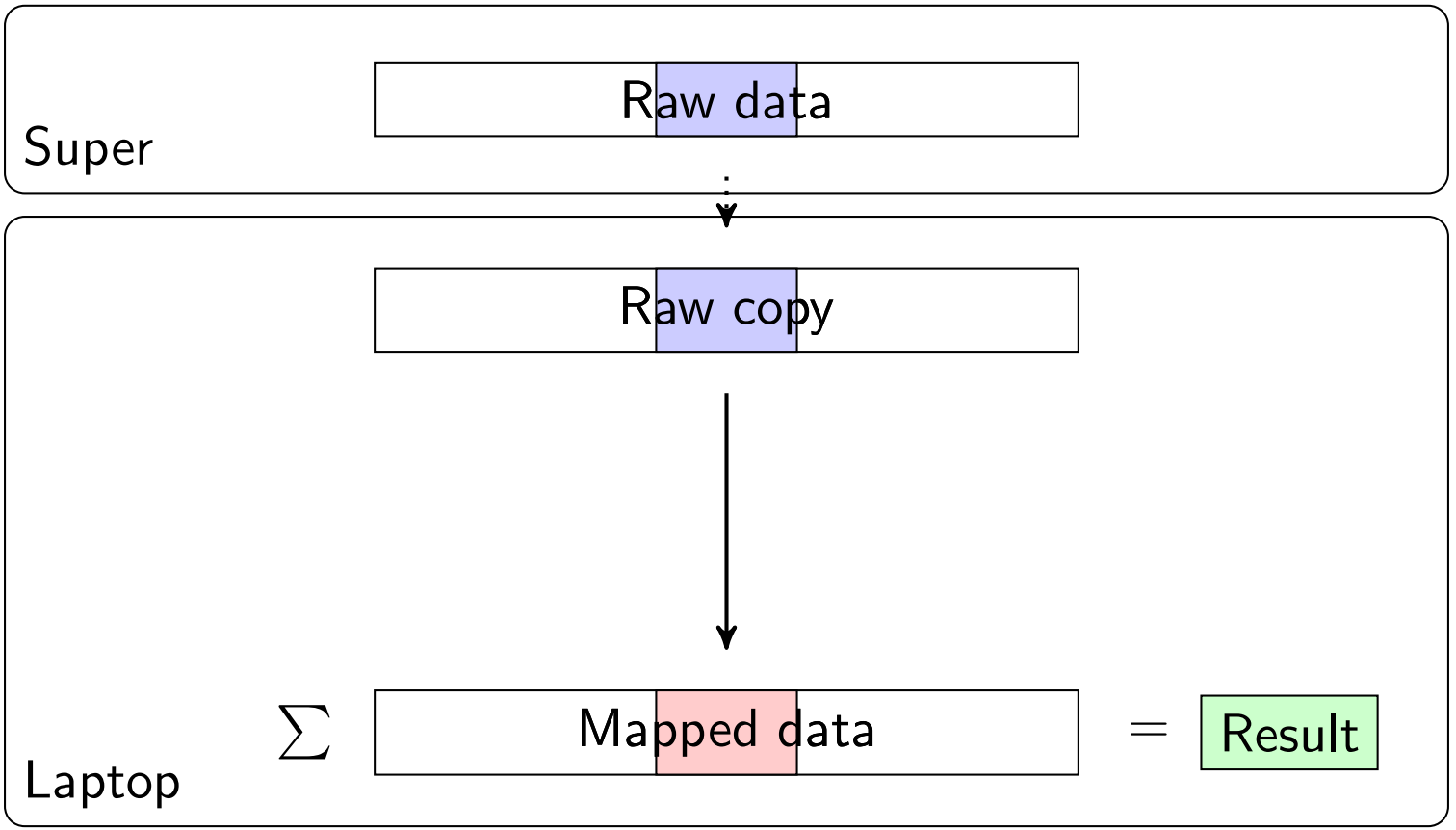
MapReduce post-processing



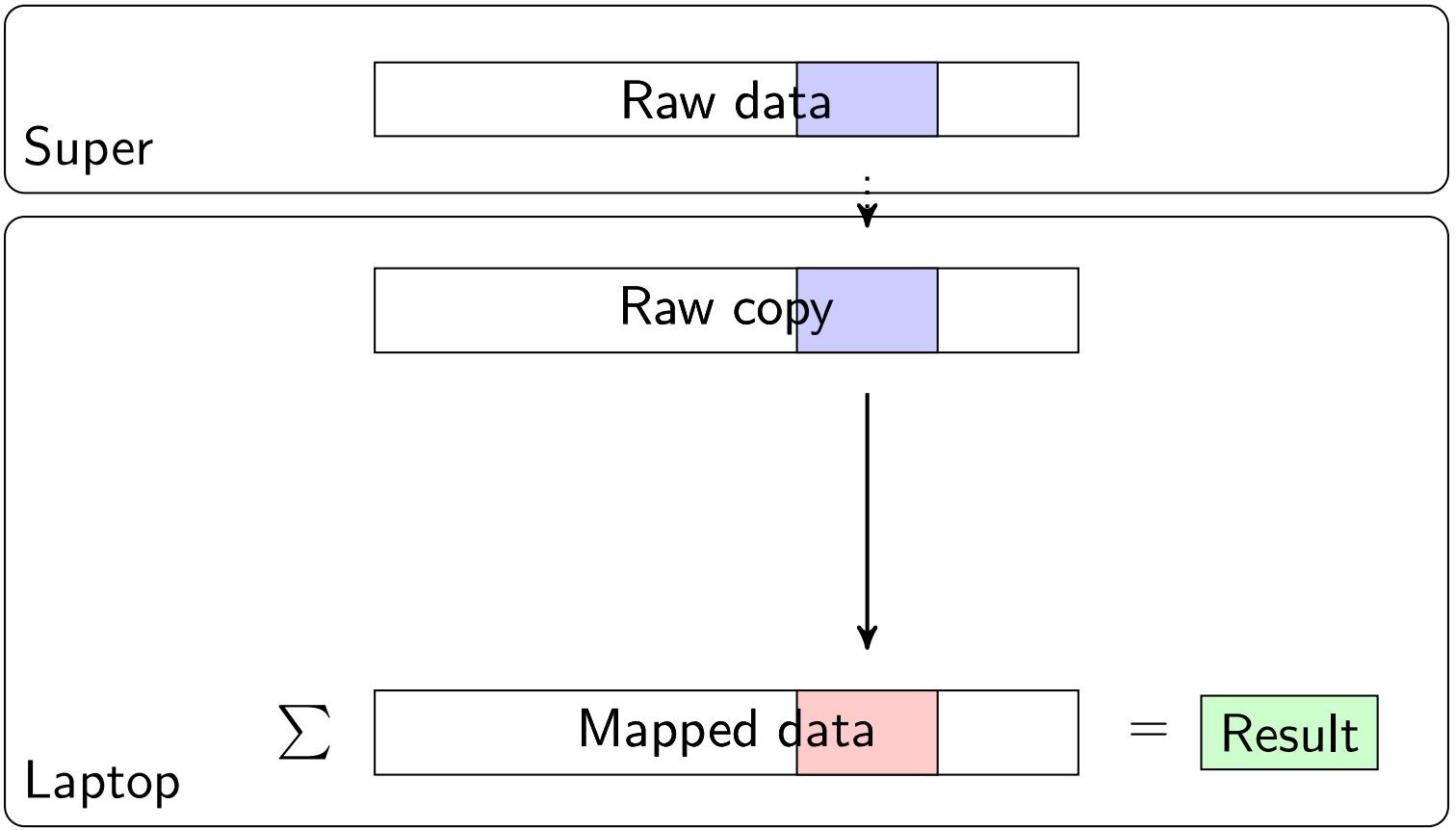
MapReduce post-processing



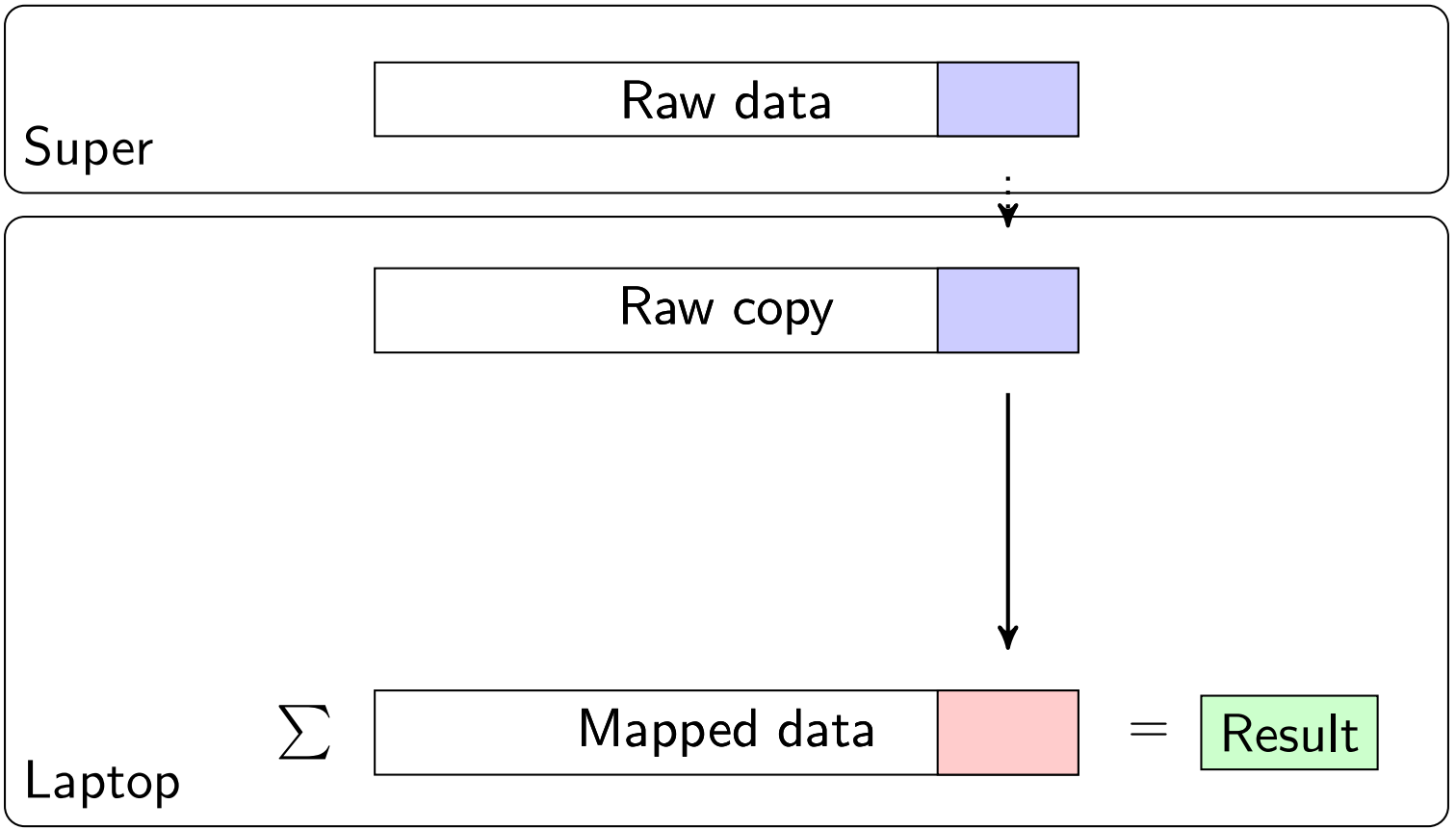
MapReduce post-processing



MapReduce post-processing



MapReduce post-processing



Capabilities

1 Frame	BW (Mbit/s)	Disk (GiB)	Wall (hr)	Core (hr)
Tukey (local)	n/a	256.	0.56	9.06
Tukey (remote)	6000	16.	0.65	10.45

8 Frames	BW (Mbit/s)	Disk (GiB)	Wall (hr)	Core (hr)
Tukey (local)	n/a	4096.	0.66	85.06
Tukey (remote)	6000	16.	5.23	83.61

[1] extrapolated

Capabilities

1 Frame	BW (Mbit/s)	Disk (GiB)	Wall (hr)	Core (hr)
Tukey (local)	n/a	256.	0.56	9.06
Tukey (remote)	6000	16.	0.65	10.45
Workstation	100	2.	6.22	< 12.44
Laptop[1]	15	2.	38.81	< 77.62

8 Frames	BW (Mbit/s)	Disk (GiB)	Wall (hr)	Core (hr)
Tukey (local)	n/a	4096.	0.66	85.06
Tukey (remote)	6000	16.	5.23	83.61
Workstation[1]	100	2.	49.79	< 99.58
Laptop[1]	15	2.	310.69	< 621.38

[1] extrapolated

Relative performance

1 Frame	BW (MBit/s)	Disk (GiB)	Wall (hr)
Copy + local	100	256.	6.75
Out-of-computer	100	16.	6.22

Out-of-computer faster for single-use

- Overlaps communication with computation

Thank you

Questions?

- `https://pypi.python.org/pypi/mapcombine`
- `https://pypi.python.org/pypi/glopen`
- `https://pypi.python.org/pypi/globussh`
- `https://pypi.python.org/pypi/chest`
- `https://pypi.python.org/pypi/slict`
- Also all on GitHub