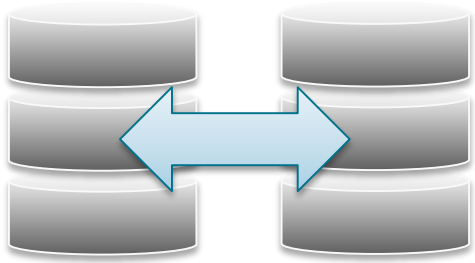
SDM storage/bandwidth

Pushing the requirements towards reasonable levels

Matthias Büchse Marko Thiele

NAFEMS Europe SPDM, 2018-11-29



SCALE

VAVID

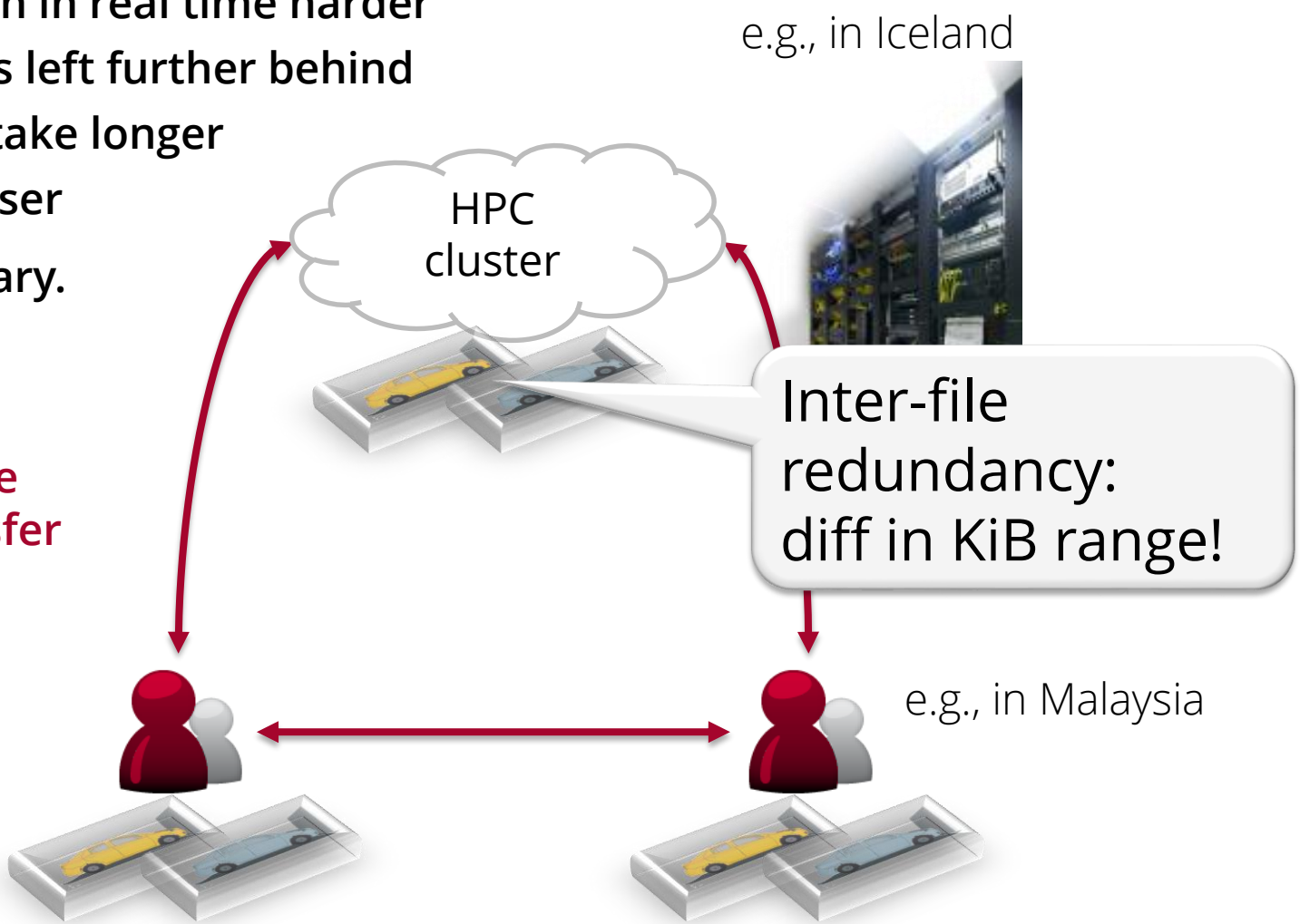
Today: Requirements significantly higher than necessary.

Therefore…

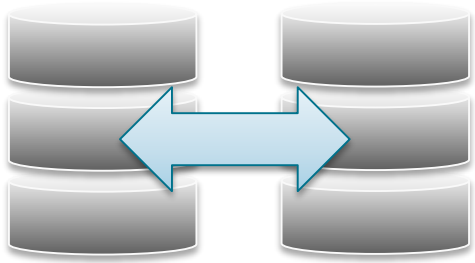
- **Collaboration in real time harder**
- **Remote sites left further behind**
- **Roundtrips take longer**
- **Models coarser**

… than necessary.

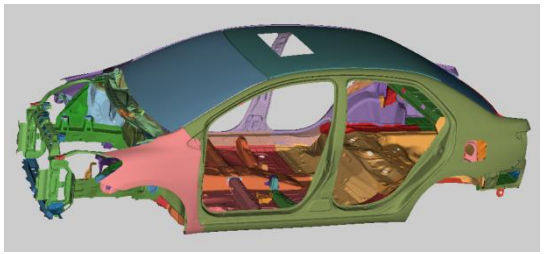
Goal: Cut storage per site and total transfer in half!

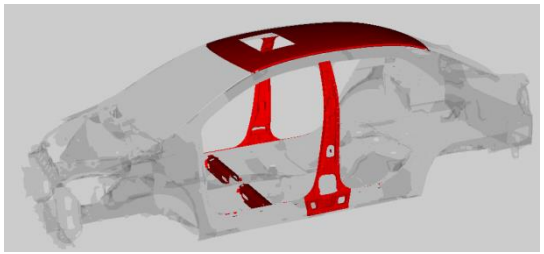


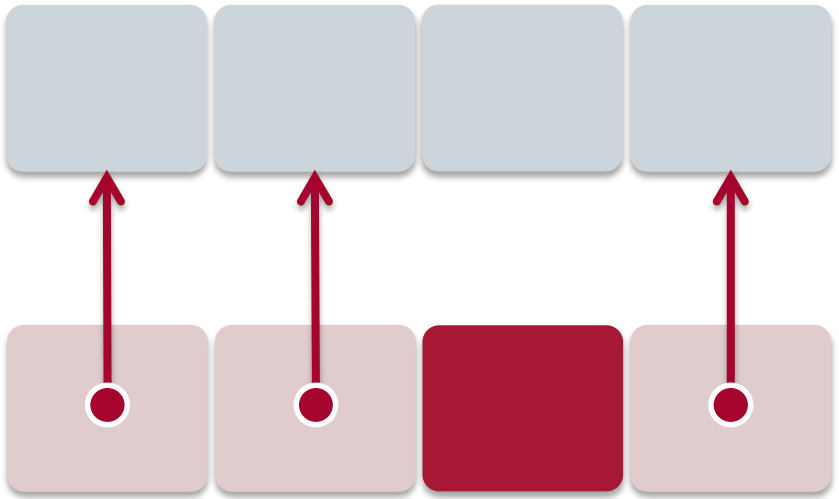
So the problem is clear. Time for solutions!



Data deduplication to the rescue!







Chunking: find block boundaries via rolling checksum
Indexing: identify each block with cryptographic hash

SDM requirements vs. redundancy solutions

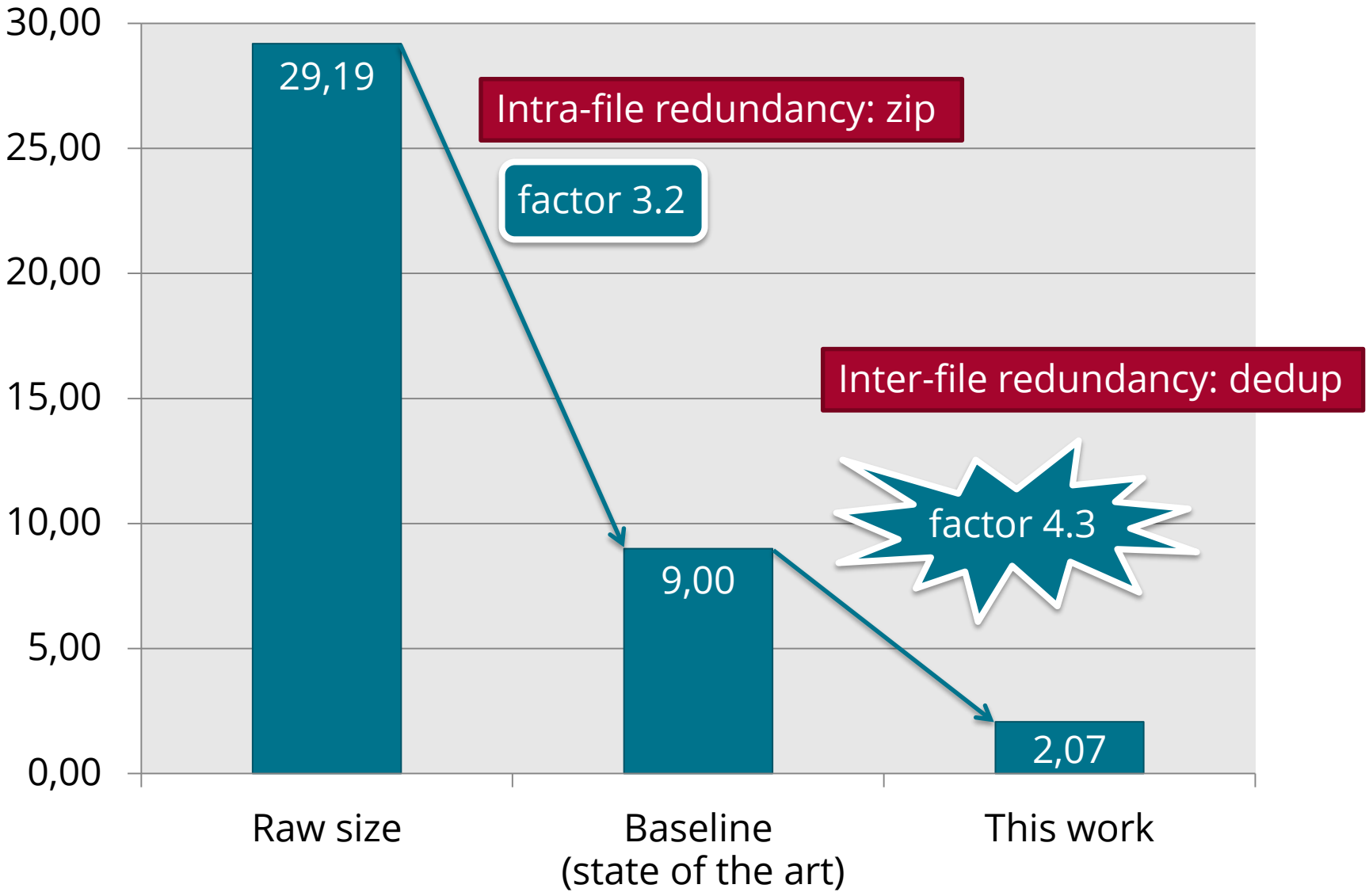
Requirement	Unix diff	Dedup FS	git/bup*	pcompress+
Random read access	X	√	?	?
Random append/delete	X	√	?	X
Concurrent access	√	√	?	?
Petabyte data	?	?	X	√
Data transfer	?	X	√	X
Encryption	X	?	X	X
No additional sys. req.	√	X	?	?

Re-implement pcompress, but with random append and delete, all the while preserving concurrent access and data integrity even after system crash (e.g., power outage).

+ pcompress: https://moinakg.wordpress.com/2013/06/22/high-performance-content-defined-chunking/

* bup: https://bup.github.io/

Results: *bulk data size, real-world input data collection [TiB]*



SCALE

Take-home message

- **Inter-file redundancy abounds in Simulation Data Management scenarios.**
- **Reduction of inter-file redundancy reduces storage and bandwidth requirements by factor 4.**
- **If your SDM system does not take care of inter-file redundancy, it wastes space and bandwidth.**
- **Otherwise, new possibilities arise…**
  - collaboration (almost) in real time,
  - collaboration with more remote sites,
  - shorter roundtrip times,
  - more elaborate simulation models.

Thank you!

The new technology was developed and integrated within the project VAVID (reference number: 01 IS 14005 C), which was partly funded by the German ministry of education and research (BMBF).