Pushing storage and bandwidth requirements of SDM towards reasonable levels

Matthias Büchse Marko Thiele

SCALE GmbH, Dresden

Today's vehicle development is faced with immense complexity—think product diversity, requirement diversity, job split, or distributed labor. Software systems for simulation data management (SDM) have proven an invaluable tool for coping with this complexity. However, SDM in general is prone to consuming a lot of storage and bandwidth, if care is not taken to reduce the abundant inter-file redundancy in (among others) simulation input data.[1] So far, the authors of this paper are not aware of any state-of-the-art SDM system to do just that.

As a result,

- collaboration in real time is harder,
- remote sites are left behind more,
- roundtrips take longer, and
- models are potentially coarser

than necessary.

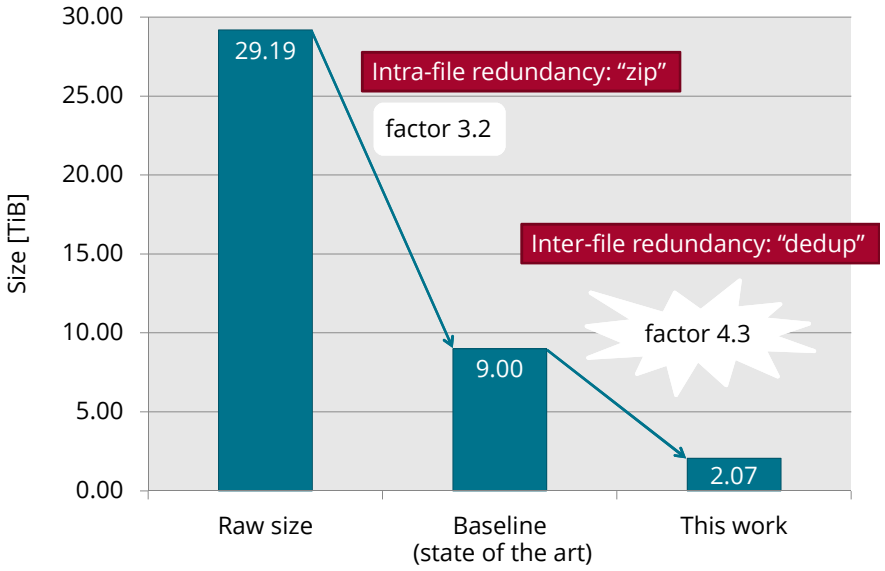

Figure 1: Storage savings on a real-world SDM dataset.

Intent on pushing the state of the art forward as well as storage and bandwidth requirements of SDM towards reasonable levels, the company SCALE integrated into their SDM solution a technology that easily lowers said requirements for real-world SDM datasets by factor 4 (cf. Fig. 1); for pure simulation input data even factor 14 is achieved.

This technology implements a well-known technique for reducing inter-file redundancy, called *data deduplication* [5], for the specific requirements of an SDM system, which are shown in Tab. 1, along with four existing implementations of this technique.

[1] This work's focus is on simulation input data. For simulation output data, see for example [4, 3].

Requirement	Unix diff	dedicated 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	?	?

Table 1: Storage/transfer requirements of an SDM system, as well as potential solutions.

As is evident from Tab. 1, none of the existing implementations satisfies every SDM requirement. Here, “dedicated FS“ refers to the several deduplicated file systems that are available, which cannot be used because they introduce additional system requirements and they do not address the matter of data transfer; “git/bup“ refers to the *git* version control system and the backup solution *bup* [1], which cannot be used because they do not scale to the petabyte range, and most other matters are unclear at best; and *pcompress* [2] refers to another backup solution, which stands out because it does scale to the petabyte range, but unfortunately, it does not satisfy many of the other requirements.

SCALE’s new technology is basically an intricate adaptation of the *pcompress* approach, but with random append and delete, all the while preserving concurrent access and data integrity even after system crash (e.g., power outage). As mentioned before, it has been integrated into the SDM solution, and it is being employed in productive installations at multiple OEMs, meeting all expectations.

Thanks to the lowered storage and bandwidth requirements, new possibilities in vehicle development arise, such as: collaboration (almost) in real time, collaboration with more remote sites, or even just working with more elaborate simulation models. Currently, SCALE is investigating slight adaptations of the technology to reduce storage requirements of other kinds of data, such as reports.

Acknowledgements 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).

References

[1] bup developers. *bup: it backs things up*. URL: https://bup.github.io/.

[2] M. Ghosh. *Scaling Deduplication in Pcompress – Petascale in RAM*. Apr. 2014. URL: https://moinakg.wordpress.com/2014/04/06/scaling-deduplication-in-pcompress-petascale-in-ram/.

[3] S. Mertler. “Neue Entwicklungen bei der Kompression von Simulationsergebnissen im Kontext von SDM-Systemen”. In: *SCALE-Informationstag Prozessautomatisierung und Simulationsdatenmanagement*. 2018.

[4] S. Mertler and S. P. Müller. “Reducing Storage Footprint and Bandwidth Requirements to a Minimum: Compressing Sets of Simulation Results”. In: *Proc. LS-DYNA-Forum 2016*. 2016.

[5] J. Paulo and J. Pereira. “A Survey and Classification of Storage Deduplication Systems”. In: *ACM Comput. Surv.* 47.1 (June 2014), 11:1–11:30. ISSN: 0360-0300.