Accelerating Viz Pipelines Using Near-Data Computing: An Early Experience

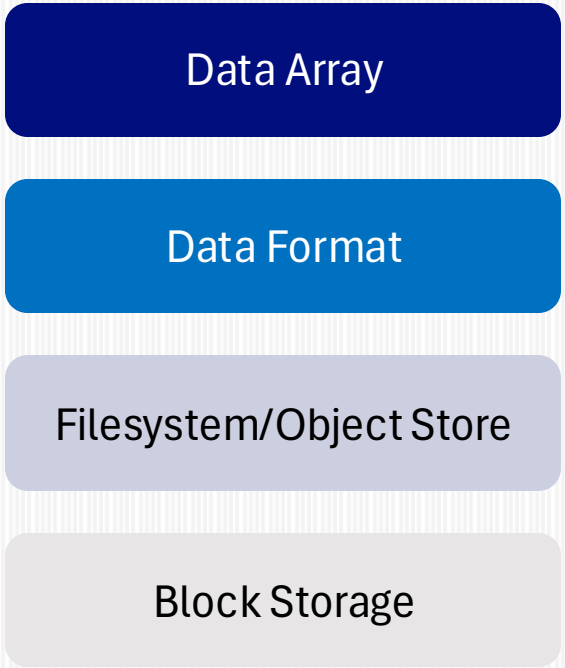
Qing Zheng, Los Alamos National Lab

qzheng@lanl.gov

LA-UR-24-31709

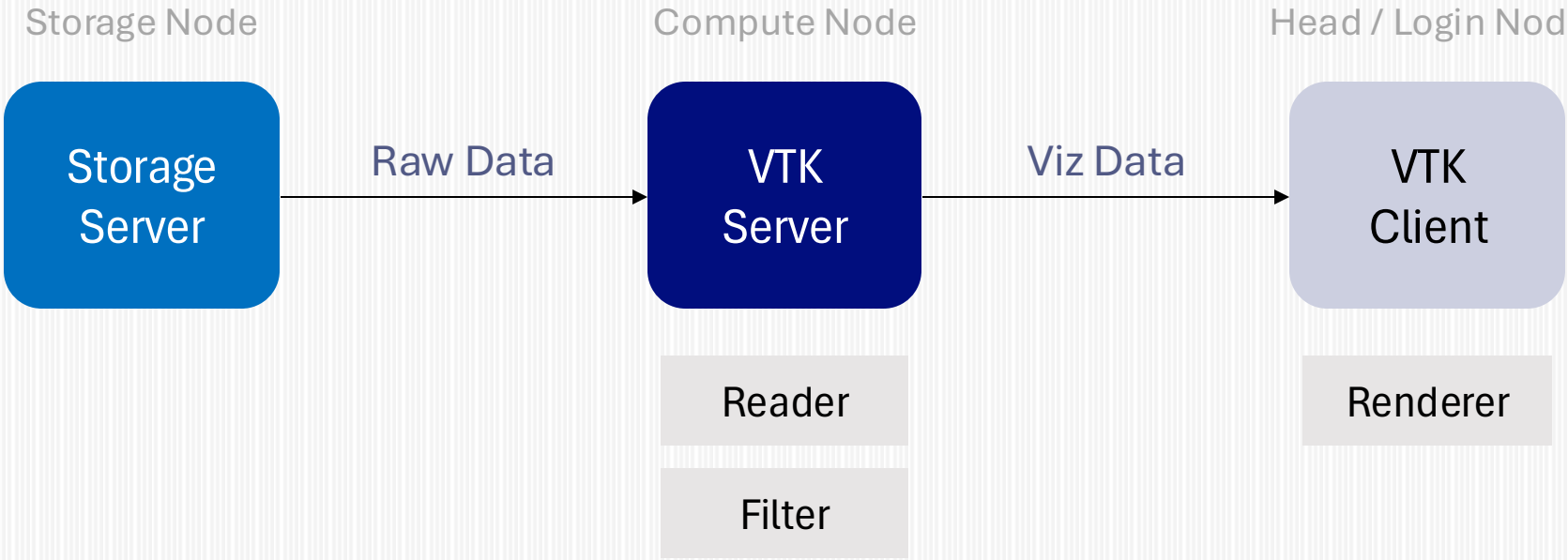
Scientific Data Storage

e.g., pressure, velocity, …



e.g., VTK, HDF5, …

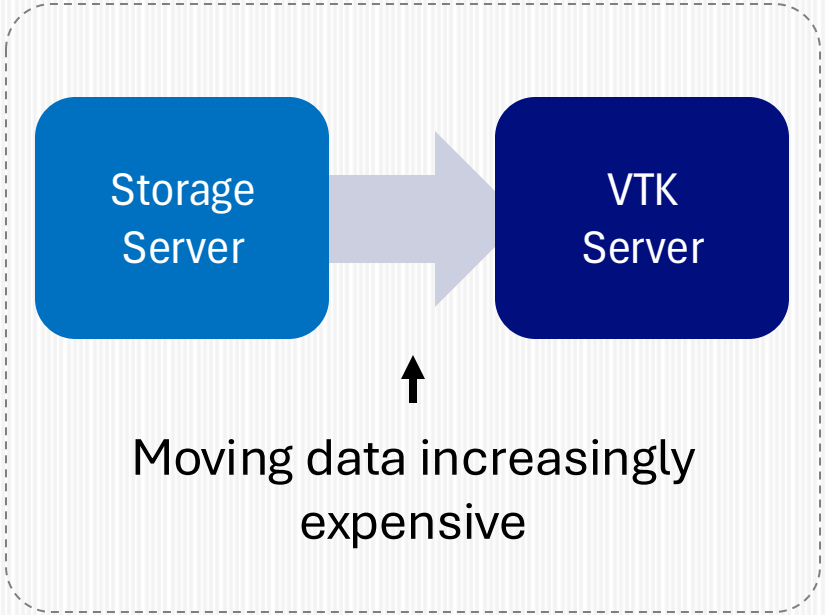
Viz Pipelines



e.g., contour filter

Two Existing Data Reduction Techniques

Problem:



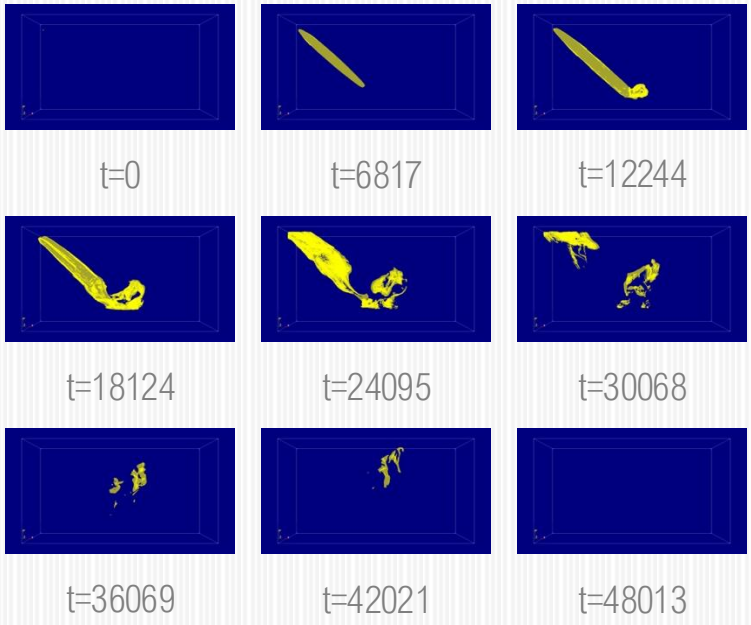
1. Data array selection

> Only read data arrays needed by the task at hand

2. Data compression

> Lower the size of each data array using GZ, LZ4, …

Example: Contour Analysis of Asteroid Impact



SC16 Scientific Viz Showcase

1. Data array selection

> There are 11 data arrays; only 1 is needed in this example: 91% data reduction

2. Data compression

> GZ: 86% - 99.8% data reduction depending on timesteps, LZ4: 84% - 99.6% depending on timesteps

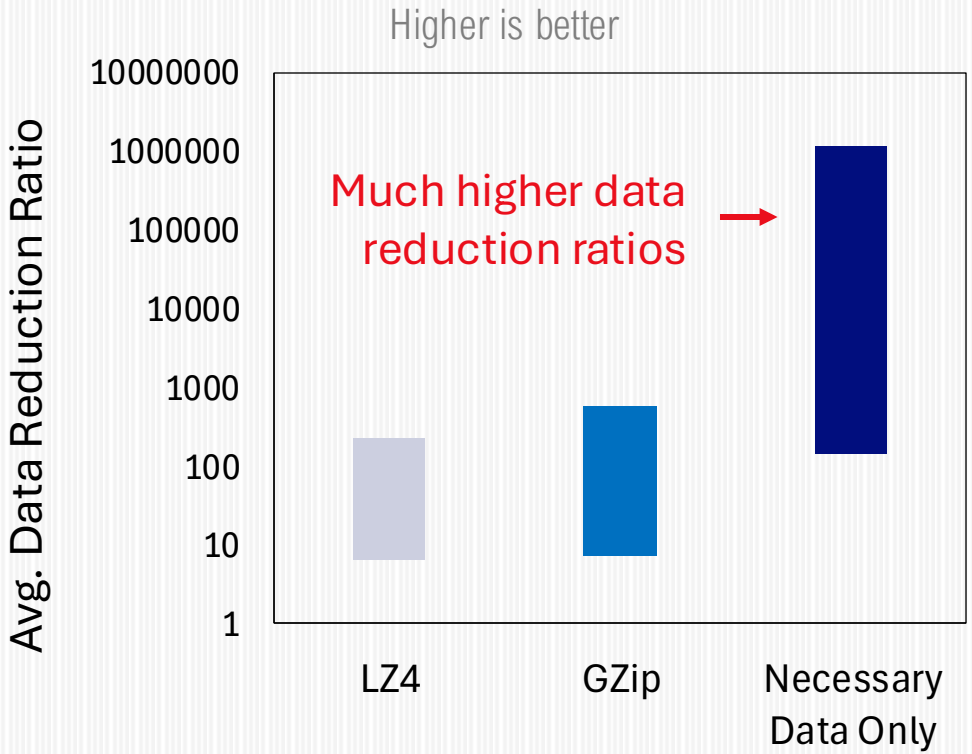
Can we do better?

Further Reducing Network Traffic

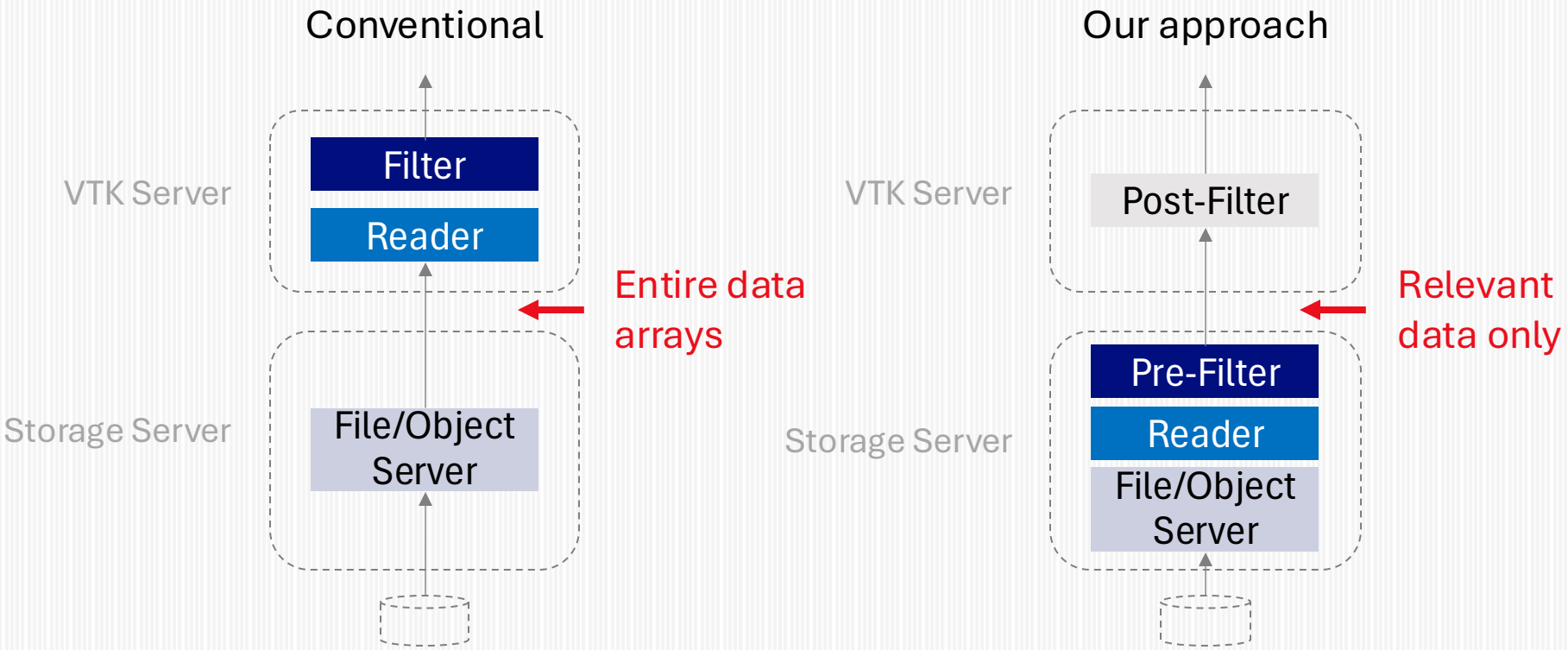
Not all array elements are needed during analysis

> Example: contour only requires data around the contour values

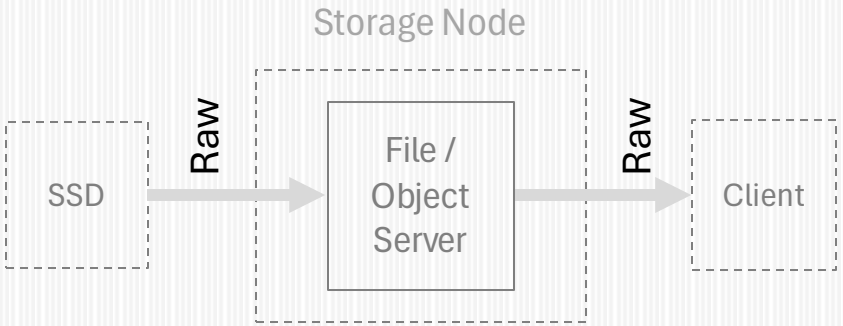
Send only necessary data by pre-processing and filtering data on storage nodes



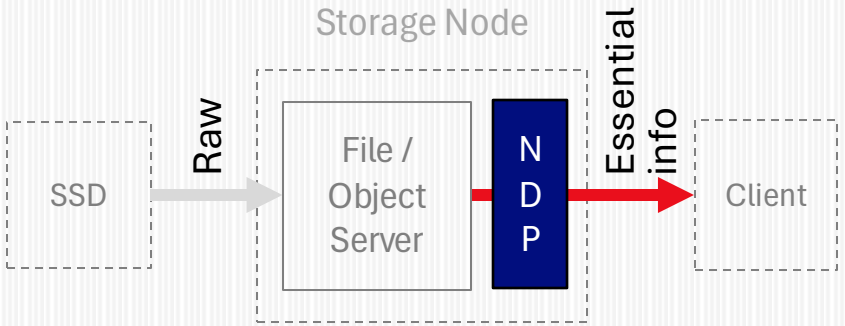
System Design



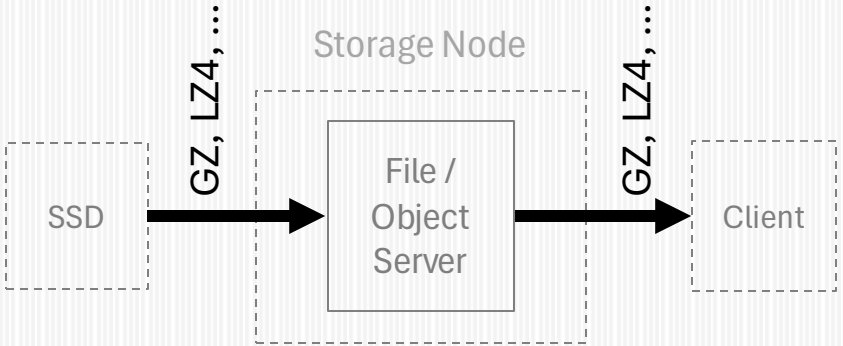
Working in Tandem with Compression



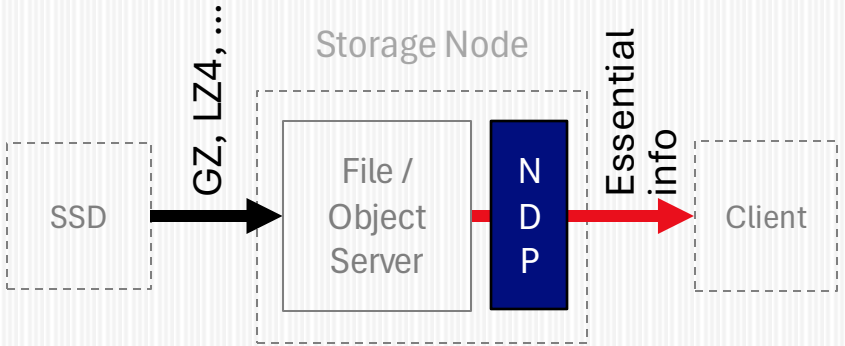
Baseline



Near-Data Processing (NDP) Only

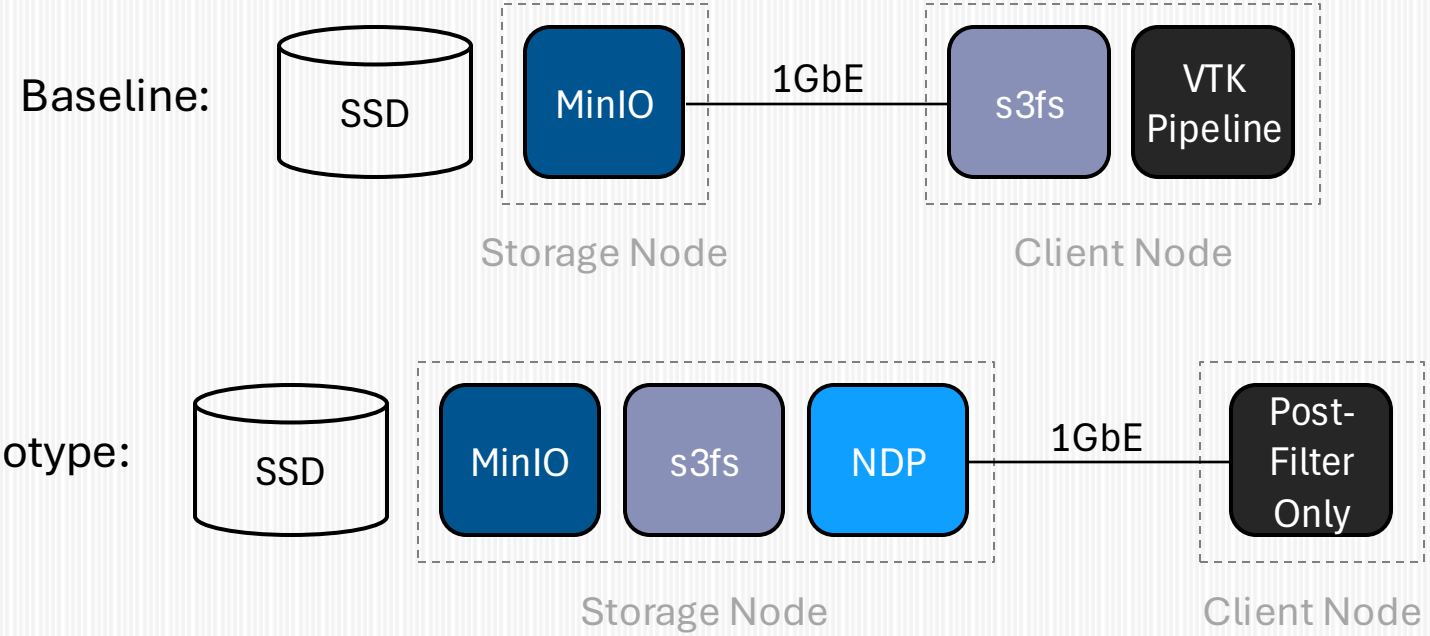


Compression Only

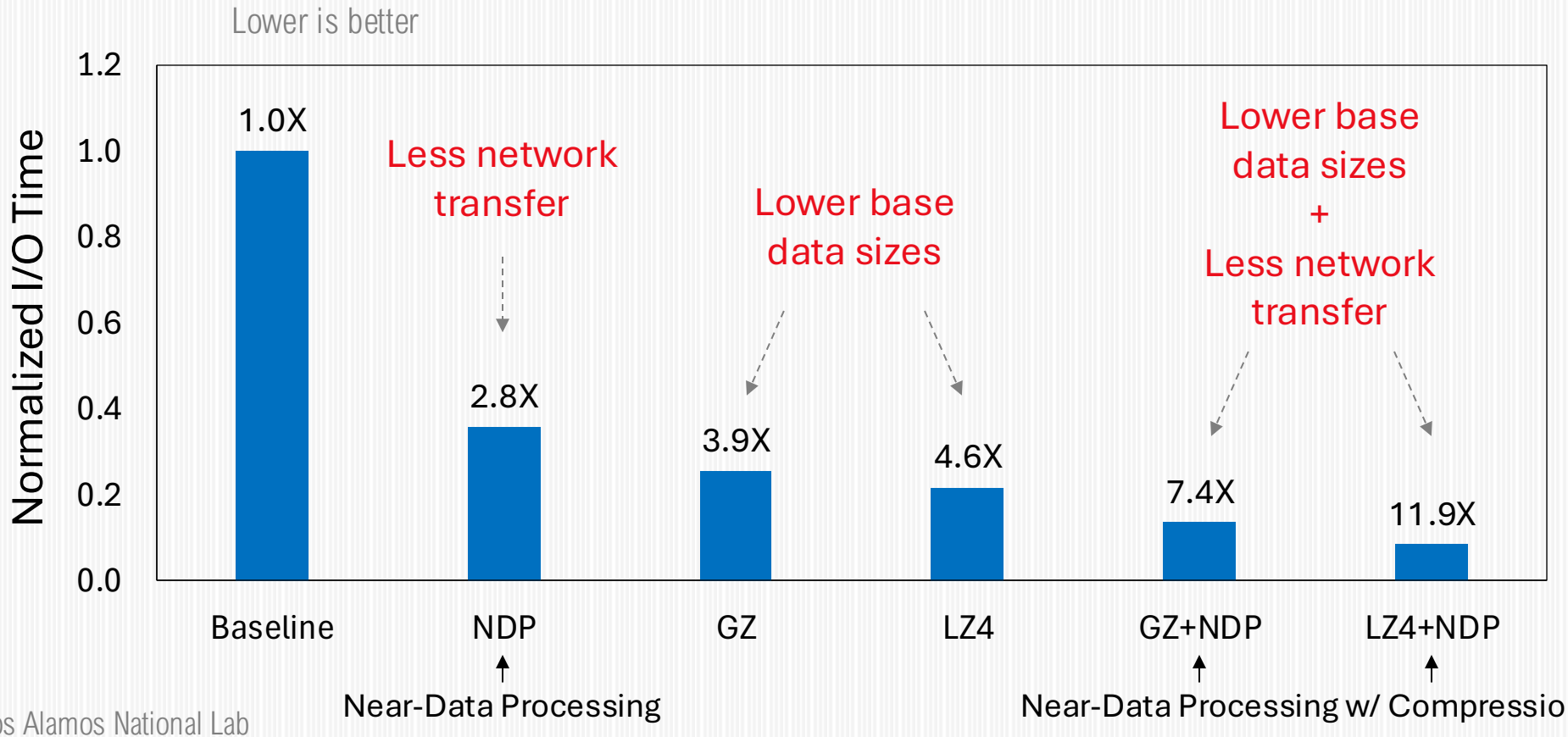


Compression + Near-Data Processing (NDP)

Experiment Setting



I/O Speedup Comparison

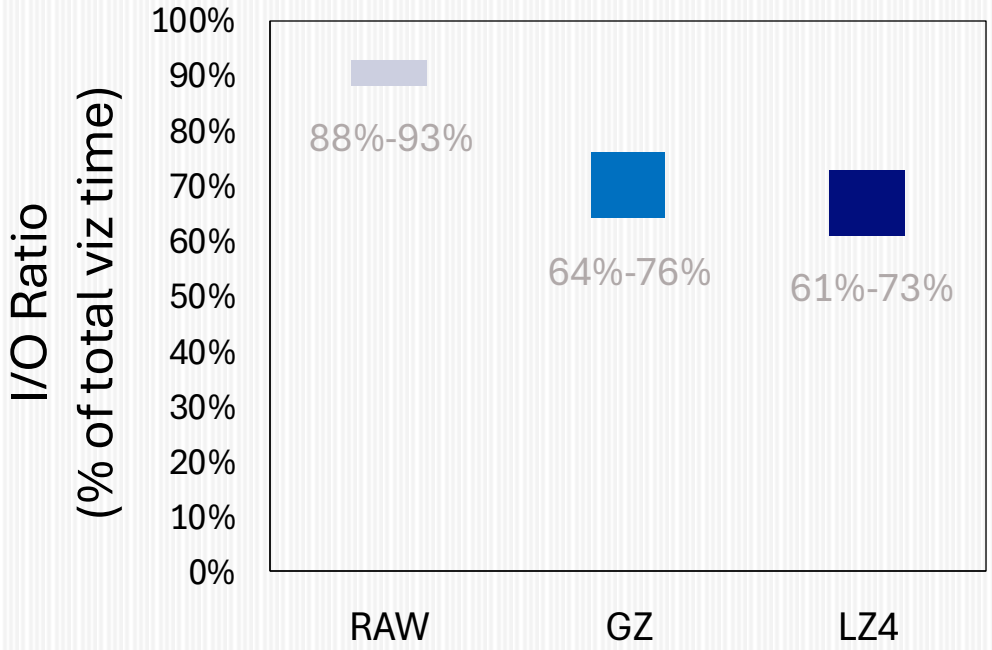


I/O Ratio Analysis

For uncompressed data, 90% of time is spent on I/O

When data is well compressed, 70% of time is spent on I/O operations

I/O remains an important bottleneck



Conclusion

Near-data processing: a new way of reducing network transfer volumes

Orthogonal to data compression (lossy or lossless)

Future work:

> More applications (currently 2)

> More mesh (currently only uniform structured grids) and filter types (currently only contours)

> Integration with storage services/devices

Please see our paper for more information!

Los Alamos

NATIONAL LABORATORY