Remote Visual Analysis on Large Turbulence Databases at Multiple Scales

Jesus Pulido[1,2], Daniel Livescu[2], Randal Burns[3], Curt Canada[2], James Ahrens[2], Bernd Hamann[1]

[1]*University of California, Davis* [2]*Los Alamos National Laboratory* [3]*Johns Hopkins University*

Remote analysis and visualization of raw large turbulence datasets is challenging. Current accurate direct numerical simulations (DNS) of turbulent flows generate datasets with billions of points per time-step and several thousand time-steps per simulation. The public Johns Hopkins Turbulence Database (JHTDB) simplifies access to multi-terabyte turbulence datasets and facilitates the computation of statistics and extraction of features through the use of commodity hardware. We present a framework that adds high-speed visualization of large datasets for this database cluster and wavelet support for multi-resolution analysis of turbulence. This framework enables remote access to tools available on supercomputers and visualization capabilities to over 230 terabytes of DNS data over the Web.

Introduction

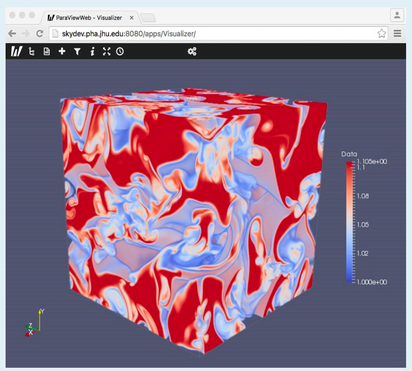
Extremely large datasets are becoming increasingly common in science and engineering, and it is often prohibitive to store an original massive dataset at multiple sites or to transmit over computer networks in its entirety. Regardless, such datasets represented tremendous scientific value for the broader scientific community. It is imperative to deploy effective technologies enabling the remote access to vast data archives for the purpose of having a large pool of scientists harness their value and make new discoveries. Our analysis framework presented here was driven specifically by the needs articulated by scientists from Johns Hopkins University (JHU) and Los Alamos National Laboratory.

Background and Method

JHTDB: The JHTDB also provides several remote tools that facilitate the analysis and retrieval of turbulence data. Data is partitioned spatially and temporally across the cluster and accessed through a database access server hosting a Web services module. This module allows for scheduling and divides user requests according to the partitioning of the data. Remote users originally interact with the database through web protocols via wrappers in multiple languages such as Matlab, Python, C and Fortran. (As seen in Fig. 2)

Wavelets: The cubic B-spline wavelet family is considered as the favored choice for compression and analysis based on previous work[1]. The wavelet basis allows for the capture of turbulence features and preservation of directionality. Wavelets are used for data reduction, analysis of turbulence, and enabling remote visualization.

Remote Visualization: The primary visualization tool used is ParaView, an open source visualization software that has been adapted for the JHTDB (Fig. 1). ParaView provides an extensive list of features and analysis tools, making it an ideal companion for this large database cluster. ParaViewWeb enables remote visualization via a light-weight web browser.



Database publically available at: http://turbulence.pha.jhu.edu

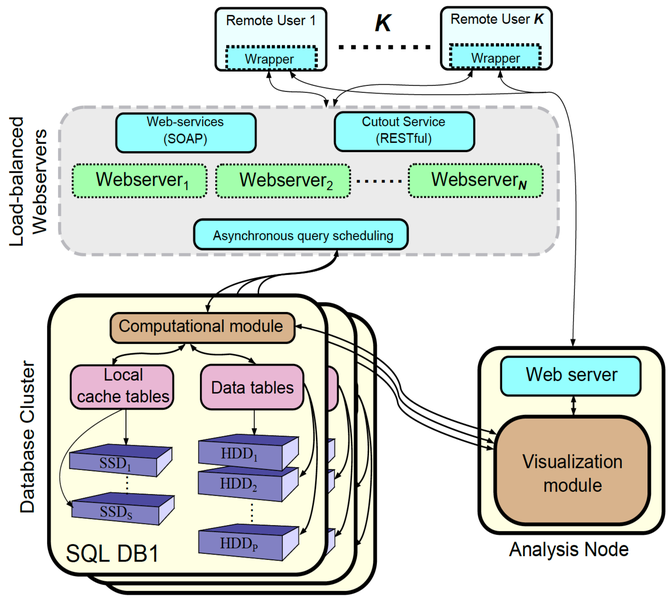


Fig.2. Architecture overview of the JHTDB. Multiple **N** server nodes create a database cluster to serve **K** users. The web services module provides an API where remote users may place requests to the database.

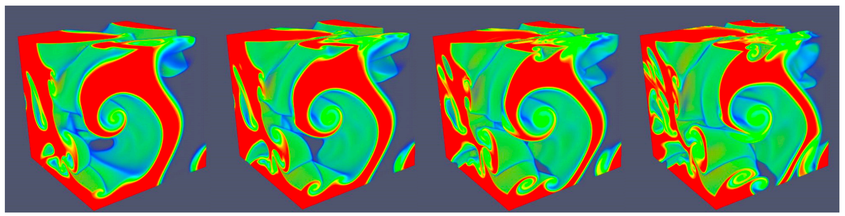


Fig.3. Temporal volume visualization of HBDT density. Wavelet integration and caching capabilities make it possible to quickly visualize multiple datasets to improve temporal understanding of a turbulence dataset. These four time steps represent t=60,75,90,105 of HBDT density.

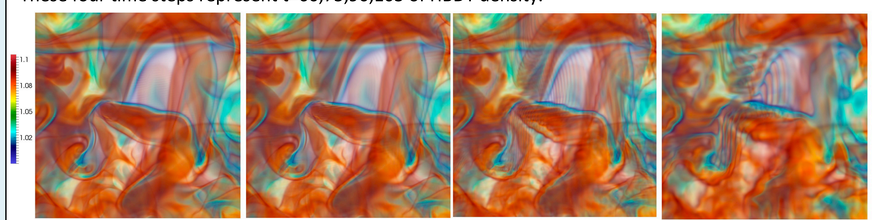


Fig.4. Volume visualization of HBDT density. A close-up of a feature shows the compression characteristics of cubic B-spline wavelets at the data level. From left to right: original, scale 1, scale 2, scale 3. Visible differences are minimal between the original and scale 1. Scale 2 begins to exhibit lossy features and by scale 3, most of the features begin to lose their finer structures around the edges.

Tests and Results

We perform a series of benchmarks given a sample case scenario where a single 1024^3 grid size dataset is accessed and visualized by a single user at multiple scales. The results can be seen in Table 1. A data analysis pipeline can put significant burden on a system depending on the size of the dataset. The benefits of data analysis on spatially smaller representations of the data is both a reduction in compute cost and memory requirements, those of which significantly impact the total time to perform this analysis pipeline. Temporal analysis is demonstrated in Fig. 3, and compression/quality comparisons in Fig.4. Although not shown, Scale-based wavelet analysis allows the creation of a hierarchy of vortex sizes and structure shapes with local and non-local interactions among the various scales.

TABLE I. PERFORMANCE RESULTS

Operation	1024^3 (Original)	512^3 (S1)	256^3 (S2)	128^3 (S3)
Data retrieval (DB)	45.5 s	43.5 s	9.8 s	1.3 s
Wavelet decompose (DB)	0 s	5.9 s	6.6 s	6.7 s
Wavelet reconstruct (Node)	0 s	6.2 s	0.8 s	0.09 s
Visualize volume (Node)	1.7 s	1.5 s	0.2 s	0.05 s
Visualize isosurfaces (Node)	5488 s	134.9 s	11.4 s	1.2 s
Total time (sequential)	5536 s	192 s	28.7 s	9.3 s
RAM used	24761 MB	4526 MB	865 MB	308 MB
Est. K concurrent users	(K<2) 1x	(K<14) 7x	(K<75) 37x	(K<212) 106x

Contributions

- Wavelet compression is introduced at the data-level to reduce access costs, reduce bandwidth and improve latency between database components.
- Remote visualization support is described for a multiterabyte database cluster supporting commodity hardware.
- Wavelet compression is introduced as a means to reduce the memory and compute footprint of datasets enabling support for many user groups at several, distributed remote sites.
- New analysis tools are demonstrated for two datasets for these types of turbulence: a) Homogeneous Buoyancy Driven Turbulence (HBDT) and b) Forced Magneto-Hydrodynamic turbulence (FMHDT).

1) J.Pulido et al. "Survey and Analysis of Multiresolution Methods for Turbulence Data." DOI: 10.1016/j.compfluid.2015.11.001