Exascale Computing Project (ECP) Data Analytics and Visualization



James Ahrens Data and Visualization Lead Los Alamos National Laboratory <u>Ahrens@lanl.gov</u>





Overview of ECP Data and Visualization

Applications running on exascale supercomputer

In situ visualization and analysis

Data services

Checkpoint restart, compression, coupling, storage performance tracking

Storage

Data collection, reduction, transform and workflow

Scalable storage software infrastructure Interactive post-hoc approaches



ECP Data and Visualization : Context for the portfolio

Vision	Supporting achieving exascale by addressing exascale data and visualization challenges		
Challenges	Exascale system concurrency is expected to grow by five or six orders of magnitude, yet system memory and I/O bandwidth/persistent capacity are only expected to grow by one and two orders of magnitude		
Mission	Deliver exascale-ready storage, data services and in situ visualization solutions for applications		
Objective	Produce data and visualization capabilities, integrate these capabilities into ECP applications, demonstrate solutions, deliver software as part of SDK		
Starting Point	Existing packages including storage tools such as ADIOS, MPI-IO, HDF5, services such as SCR for checkpoint restart, and post-processing visualization tools such as ParaView and Visit		
Portfolio Goals	Storage	 Deliver via HDF5 API, focus on burst buffer and backends Includes top DOE HPC storage teams - ADIOS, MPI-IO, HDF5, PnetCDF 	
	Services	Deliver data services (What not how) such as scientific data compression (ZFP, SZ), checkpoint restart (VeloC), storage performance tracking (Darshan)	
	Visualization	Deliver new exascale-oriented in situ visualization and analysis workflow via ALPINE, VTK-m and Cinema products	

ECP Data and Visualization

Project Name	PI Name, Inst	Short Description/Objective
Data & Vis SDK	Chuck Atkins, Kitware	Support the deployment, testing and usage of ECP Data and Visualization packages.
ADIOS	Scott Klasky, ORNL	ADIOS supports efficient I/O and code coupling services.
DataLib	Rob Ross, ANL	Supports efficient I/O including MPI-IO (ROMIO) and PNetCDF, as well as I/O monitoring (Darshan) and data services (Mochi)
ECP/VTK-m	Ken Moreland, SNL	Comprehensive effort to provide VTK-based scientific visualization software that supports shared memory parallelism.
VeloC	Franck Cappello, ANL	VeloC-SZ develops two software products: VeloC (checkpoint restart) and SZ (lossy compression with strict error bounds).
ExalO	Suren Byna, LBNL	Efficient system topology and storage hierarchy-aware HDF5 and Unify (node local) parallel I/O libraries
ALPINE	James Ahrens, LANL	Deliver in situ visualization and analysis algorithms, infrastructure (ALPINE) and data reduction of floating-point arrays for reduced memory, communication, I/O, and offline storage via advanced data compression (ZFP).



Number of particles (128 nodes)

ZFP: 10x compression of simulation state in GENE fusion code with acceptable loss



