SDM Center: Scientific Data Management Center

Norbert Podhorszki

ORNL,
Scientific Computing Group,
End-to-end team

Project Overview

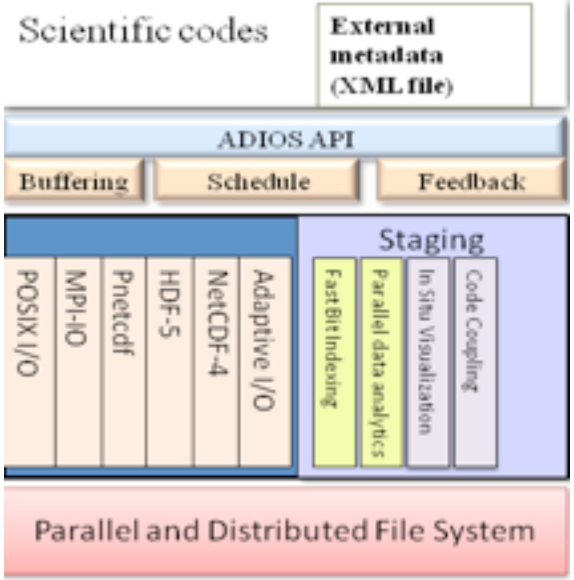
- http://sdmcenter.lbl.gov
- PI: Arie Shoshani, LBNL
- Generate, manage and analyze scientific data
  - Storage Efficient Access (SEA),
  - Data Mining and Analysis (DMA), and
  - Scientific Process Automation (SPA)

Key Technologies

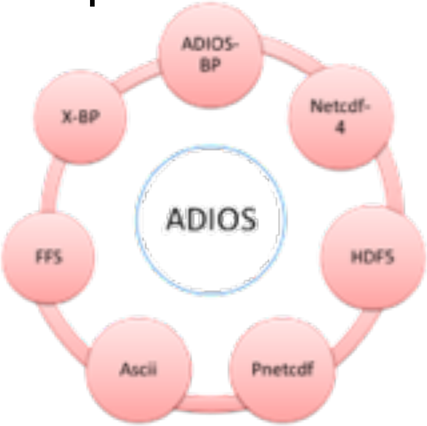
- Storage:
  - ROMIO (an MPI I/O implementation)
  - ADIOS: Adaptable I/O System
  - Parallel NetCDF
- Data Mining and Analysis
  - FastBit indexing
  - Sapphire mining software
  - Parallel R
  - ISABELA lossy compression
- Process Automation
  - Kepler: Sci. Workflow Management
  - eSiMon: simulation monitoring dashboard

ADIOS: Adaptable I/O System

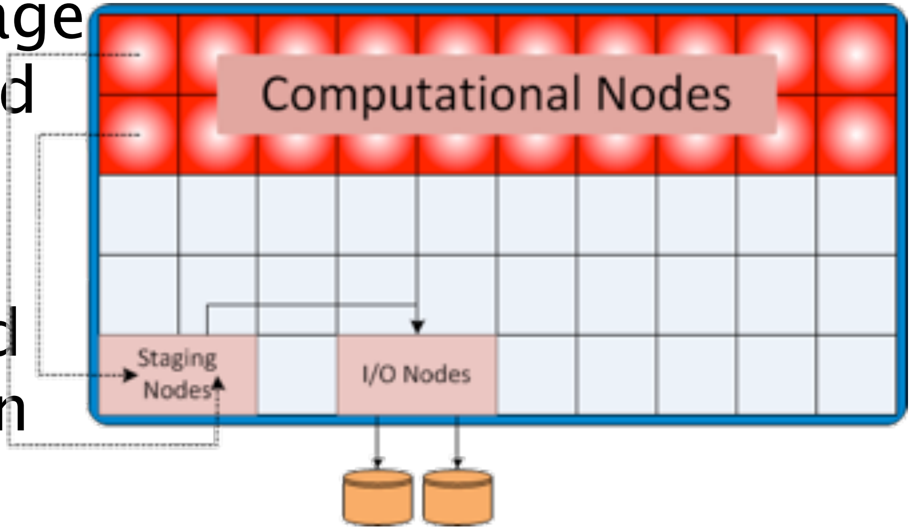
- Provides portable, fast, scalable, easy-to-use, metadata rich output with a simple API
- Change I/O method by changing XML
- Layered software architecture:
  - Allows plug-ins for different I/O implementations
  - Abstracts the API from the method used for I/O
- Open source:
  - http://www.olcf.ornl.gov/center-projects/adios/
- Research methods from many groups:
- S3D code: 32 GB/s with 96K cores, 1.9MB/core: 0.6% I/O overhead with ADIOS
- XGC1 code: 40 GB/s, SCEC code: 30 GB/s
- GTC code: 40 GB/s, GTS code: 35 GB/s



I/O componentization



Data Staging



- Reduces performance linkage between I/O subsystem and application
- Decouple file system performance variations and limitations from application run time
- Enables optimizations based on dynamic number of writers
- High bandwidth data extraction from application
- Scalable data movement with shared resources requires us to manage the transfers
- Scheduling properly can greatly reduce the impact of I/O

Why I am here, personally

- We support many applications at scale directly through INCITE and SciDAC programs
- The more technologies we know the better/faster we can help users
- I need to learn what people (should) do at large scale
- I also write pthreads+MPI apps and still use printf/gdb

Platforms

- ADIOS
  - MPI C/C++/Fortran90 applications
  - Platforms: Cray, Bluegene, Linux, OSX
- My staging method
  - MPI+Pthreads, RDMA
  - Infiniband currently, Portals/Gemini next