Design and Implementation of the Tianhe-2 Data Storage and Management System

Yutong Lu, Peng Cheng, Zhiguang Chen

Lu YT, Cheng P, Chen ZG. Design and implementation of the Tianhe-2 data storage and management system. JOURNAL
OF COMPUTER SCIENCE AND TECHNOLOGY 35(1): 27–46 Jan. 2020. DOI 10.1007/s11390-020-9799-4

Research Objectives:

- Research Scope
  - High-performance computing (HPC)
  - Data management
  - Parallel file system

- Background and Motivation
  - Convergence of HPC, big data and artificial intelligence
  - “Triple use” systems are required
  - Formidable challenges in supporting converged applications

- Kernel Contributions:
  - Detail three data management challenges
  - File system optimizations in terms of metadata and small files
  - A spectrum of data management optimizations and application-specific optimizations

File system optimizations

- Hierarchical Storage architecture: Local storage + Shared storage
- H2FS: Hybrid virtual namespace + Predefined I/O modes
- Metadata throughput optimizations: Pre-allocation + Proxy Server
- Small files optimizations: Cuckoo Hash + Key-Value data structure

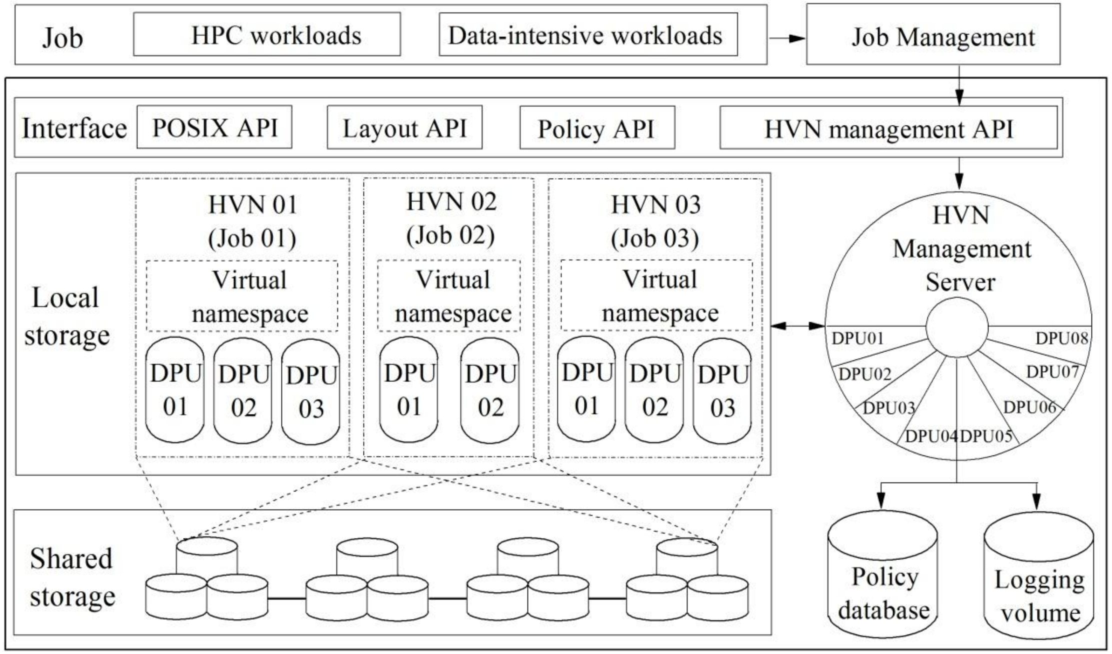

Fig.1. Architecture of H2FS.

- Compared with the original version, metadata throughput optimizations show up to 5x speedup for file open/create requests
- Compared with the original version, small files optimizations show up to 2.9x speedup for the file stat operation

Data Management Optimizations:

- Tiered data management
  - Workflow data access patterns
  - Customized data management strategies

- Data-Aware task scheduling
  - Pending tasks with locality labels
  - Bring computations to the data

- Indexing and query processing:
  - In-situ Indexing + bitmap-range indexes
  - Parallel query processing

- Intelligent storage optimization:
  - Collecting I/O records of scientific workflows
  - Train a classification model to make data placement decisions

Research Conclusions:

- **Challenges in storage and data management**
  - Exacerbated I/O bottleneck
  - Adaptive or intelligent data management optimizations
  - Unified data management for heterogeneous scientific data

- **Our solutions to embrace converged applications on HPC systems:**
  - File system optimizations in terms of metadata and small files
  - A spectrum of data management optimizations
  - Application-specific optimizations

- **Future challenges:**
  - Innovative storage architecture that can best utilize the emerging non-volatile storage devices
  - Parallel file system for next-generation exascale system
  - Data-centric programming models