Building a Tiered Digital Storage Environment Based on User-Defined Metadata to Enable eResearch

David Fellinger

iRODS Consortium, Chapel Hill, North Carolina, USA, davef@renci.org

The history of digital storage is as interesting as the history of high performance computing (HPC). In fact, the technology advancements of storage have driven HPC with respect to performance and data gathering. Succeeding vacuum tube storage some of the earliest random access storage (RAM) consisted of magnetic cores which could be magnetized or not by a current through a small winding on the core. Each core represented one digital bit and the state indicated either a zero or one. This core storage was non-volatile so it could maintain a state when power was removed. The problem was that "reading" a 1 reset the bit which then had to be re-written. Reading a 0 did not reset the bit so the access time depended on the state of the core. The other problems were both size and cost. The cost for core storage decreased through the 1960's from 1 dollar to 1 cent per bit but it remained a necessary part of high speed computing for over 10 years regardless of the expense. The performance of this memory was about 1 uS access for both "READS" and "WRITES". In the late 1950's IBM saw the need for a slower but cheaper form of storage and introduced the first magnetic disk drive. The IBM 350 used 50, 24 inch platters and wrote on 100 sides with a capacity of 5 mB and an access time of about 3 mS with a capacity of 5 mB [1]. Even then the storage was used for very different forms of data. The random access core memory was used for executable programs while the output data was usually stored on disk. Magnetic tape actually preceded other forms of data storage with the earliest example dating back to 1951 on the UNIVAC machine [2]. The earliest forms were open reel using 10.5 inch reels for local storage and 7 inch reels when there was a necessity to send a large amount of data to another facility. Even in these early days of computation, tape was considered an archival storage type or a way of data distribution. Access was sequential followed by block searches. The development of field effect transistor technology had a huge impact on both storage and computation. Today, both static random access memory (SRAM) and dynamic random access memory (DRAM) are the primary forms of storage within compute nodes but we still have modernized discs and tape for external storage. Digital storage still varies widely in performance and cost.

Moving data across storage layers

Early HPC users immediately saw the need to store data in the appropriate hardware depending on need with cost being a major consideration followed by retrieval latency. As the number of storage types increased, the need to apportion data to optimized hardware became critical. The software for doing this operation was termed hierarchical storage management (HSM) and IBM was first again to introduce this. It is interesting to note that CSIRO developed one of the first HSMs in the 1960's. This was part of the Drum and Display (DAD) operating system and ran on various CSIRO platforms [3]. These initial HSMs systems worked well but they all had one common characteristic; data placement was based on either data type or age based upon the date of creation. In other words data placement or movement was based only on the handle that was initially assigned to a file, primarily the file extension and the date of creation. HSM systems have shared that one commonality for over 60 years.

Storage hierarchy based upon extracted and user-defined metadata

The widespread use of various sensors has greatly increased the volume of data that must be handled. Organizations such as the Wellcome Sanger Institute in the UK have a number of genomic sequencers producing gigabytes of data. A project by the Victoria Department of Agriculture uses hundreds of sensors on cattle and in various places to create data driven Smart Farms. NIWA in New Zealand has data produced by a great many climate sensors and they share data with other organizations across the world. In fact the term "Big Data" has been coined to represent these data types that are characterized by volume, variety, velocity, and veracity. Traditional HSM systems were designed simply to move data from active to more passive storage for archival purposes. Today data pools are created in file systems covering many aspects of HPC. In a data gathering model, there is usually a file system where the incoming data is stored. There is a file system attached through storage gateways within an HPC cluster to a parallel file system and finally, there are numerous file systems used for distribution as well as archives. This data is generally identified by specific extracted or user-defined metadata.

iRODS as a data manager

The Integrated Rule-Oriented Data System (iRODS) is an open source software product of the iRODS Consortium. It has been designed to enable data virtualization and data discovery while enabling workflow automation and secure collaboration. Recent development efforts have been focused on storage tiering which allows iRODS to be an ideal HSM with data movement and retention criteria based on metadata rather than file attributes. The iRODS software can automatically "read" a file header or the entire file to gather and catalog descriptive metadata. For example, Wellcome Sanger collects genomic sequence files that have various anomalies. These files are ingested and stored within a file system known as a "landing zone". The iRODS software can read the file headers and/or the machine annotations and the files are then stored in file system locations based on the anomalies. An administrator can specify a number of files with similar anomalies that must be collected before a data reduction operation is performed. When that number has been reached iRODS can migrate the data to a parallel file system that is directly connected to a compute cluster. The iRODS software can even notify a machine scheduler that the required data is available. At the conclusion of the process, iRODS can purge the data from the parallel file system to reduce file system traffic and migrate it to a data distribution file system. Of course, numerous data distribution file systems may be available depending of the nature of the data and the distribution requirements. In some cases, the data may be critical such that two copies are created in two separate file systems for redundancy and disaster recovery. That copy operation can be handled completely by iRODS and the data locations are tracked such that the data can be utilized by researchers. The specific metadata is stored in a database and is searchable and discoverable. A researcher wishing to understand a specific anomaly or characteristic merely searches the database with a simple tool set and can retrieve all of the relevant material quite easily. All of these operations are based on collected metadata regardless of file creation attributes. The database or catalog can also track usage data which can be automatically reviewed. If a specific file has not be accessed for a period of time determined by the administrator, it can be migrated by iRODS to an even less expensive form of storage such as tape. This same process can be used in many areas of research to group and categorize similar data types and process them based on relevant metadata.

The iRODS software is very flexible with respect to changes based on new circumstances or new sensors. Conditions may dictate various priorities. For example, drought conditions may change the priority of sensor information on a Smart

Farm. Weather conditions may change the priority and analysis requirements of various atmospheric sensors in a climate research facility. In all cases, iRODS can be re-configured easily to accommodate changing requirements.

Data ingestion is far more complex than a simple file create operation. It is, in fact, the first operation in a process workflow that must be metadata driven to enable efficiency of analysis. The data must be properly cataloged, and apportioned to diminish research cycle time. HSM systems are simply designed to build archives. The use of iRODS to manage tiered storage elements enables both the process and the subsequent distribution operations allowing researchers in many areas to gather, process, and discover data of interest.

About the author

Dave Fellinger is a Data Management Technologist and Storage Scientist with the iRODS Consortium. He has over three decades of engineering and research experience including film systems, video processing devices, ASIC design and development, GaAs semiconductor manufacture, RAID and storage systems, and file systems. As Chief Scientist of DataDirect Networks, Inc. he focused on building an intellectual property portfolio and presenting the technology of the company at conferences with a storage focus worldwide.

In his role at the iRODS Consortium, Dave is working with users in research sites and high performance computer centers to confirm that a broad range of use cases can be fully addressed by the iRODS feature set. He helped to launch the iRODS Consortium and was a member of the founding board.

He attended Carnegie-Mellon University and holds patents in diverse areas of technology.

References

1. IBM 350 Disk Storage Unit. Available from: https://www.ibm.com/ibm/history/exhibits/storage/storage_350.html , accessed 3 May 2019
2. UNIVAC Magnetic Tape. Available from: http://www.ricomputermuseum.org/Home/interesting_computer_items/univac-magnetic-tape , accessed 4 May 2019
3. CSIROpedia New Systems, Drum and Display, Available from: https://csiropedia.csiro.au/systems-developed/#Systemsdeveloped-D , accessed 4 May 2919