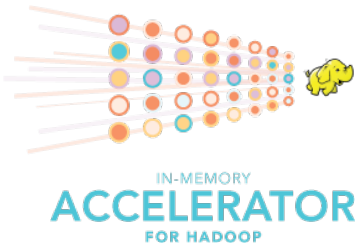
GridGain
IN-MEMORY COMPUTING

In-Memory Accelerator For Hadoop

White Paper
GridGain Systems, 2013



© Copyright 2007-2013 GridGain Systems, Inc.

Table of Contents:

© Copyright 2007-2013 GridGain Systems, Inc.

In-Memory Computing

What is In-Memory Computing?

Data volumes and ever decreasing SLAs have overwhelmed existing disk-based technologies for many operational and transactional data sets, requiring the industry to alter its perception of performance and scalability. In order to address these unprecedented data volumes and performance requirements a new solution is required.

> In-Memory Computing is characterized by using high-performance, integrated, distributed memory systems to compute and transact on large-scale data sets in real time, orders of magnitude faster than possible with traditional or hybrid disk-based technologies.

With the cost of system memory dropping 30% every 12 months In-Memory Computing is rapidly becoming the first choice for a variety of workloads across all industries. In fact, In-Memory Computing paves the way to a lower TCO for data processing systems while providing an undisputed performance advantage.

In-Memory Accelerator For Hadoop At A Glance

What is In-Memory Accelerator For Hadoop?

GridGain's In-Memory Accelerator For Hadoop is based upon the industry's first **dual-mode in-memory file system, GridGain File System (GGFS),** that is 100% compatible with Hadoop Distributed File System (HDFS) and **In-Memory MapReduce** implementation. GGFS is a plug-and-play, "no assembly required", alternative to disk-based HDFS enabling up to 100x faster performance for IO and network intensive Hadoop MapReduce jobs running on tens of hundreds of computers in a typical Hadoop cluster.

GridGain's In-Memory MapReduce allows to effectively parallelize the processing of in-memory data stored in GGFS. It eliminates the overhead associated with job tracker and task trackers in a standard Hadoop architecture while providing low-latency, HPC-style distributed processing.

In-Memory Accelerator For Hadoop is a first-of-its-kind Hadoop extension that works with your choice of Hadoop distribution, which can be any commercial or open source version of Hadoop available, including Hadoop 1.x and Hadoop 2.x distributions. Hadoop accelerator provides the same performance benefits whether you run Cloudera, HortonWorks, MapR, Apache, Intel, AWS, or any other distribution.

GGFS support for dual-mode allows it to work as either a standalone primary file system in the Hadoop cluster, or in tandem with HDFS, serving as an intelligent caching layer with HDFS configured as the primary file system. As a caching layer it provides highly tunable read-through and write-through behavior. In either case GGFS can be used as a drop-in alternative for, or an extension of, standard HDFS providing an instant performance increase.

© Copyright 2007-2013 GridGain Systems, Inc.

The unique “plug-in” architecture behind In-Memory Accelerator For Hadoop gives you the freedom to not only choose any Hadoop distribution but also use any of the dozens of Hadoop-based tools that your organization already utilizes without interruption because GGFS requires zero code change to existing MapReduce jobs. Whether you use standard tools such as HBase, Hive, Mahout, Oozie, Flume, Scoop, or Pig, or any of the commercial BI, data visualization or data analytics platforms, you can continue to use them without any change while enjoying an instant performance boost.

In the sections that follow we will look at the following questions:

- What problems does In-Memory Accelerator For Hadoop solve?
- What makes In-Memory Accelerator For Hadoop an unique solution?
- What are the In-Memory Accelerator For Hadoop key technical features?

In-Memory Accelerator For Hadoop Facilitates Fast Data

What problem does In-Memory Accelerator For Hadoop solve?

In today’s world, IT and business users alike are challenged with the need for better information and knowledge to differentiate, innovate and ultimately reshape their businesses. In a rapidly growing number of cases that process is being enabled by a move to Big Data.

Companies around the world are increasingly collecting vast quantities of real-time data from a variety of sources - from self-documenting online social media to highly structured transactional data, to data from embedded devices and the “Internet of Things”. Once collected, users or businesses are trying to make sense of the data for patterns and insights that can be used to drive better and optimized business decisions or actions.

Specialized new technologies like Hadoop are being used to store and process vast amounts of data in bulk in a predominately off-line batch-oriented mode. Consequently, most of the focus on Big Data to date has been on “low hanging fruit” analytics (i.e. traditional OLAP) use cases where the data being processed is relatively static—meaning that it has already been collected and stored in Hadoop and will never be updated.

This is where Fast Data comes into play. Fast Data is a complementary technology to Big Data where the focus is shifted toward processing large operational (i.e. traditional OLTP) and/or streaming data sets with low-latency, in real-time. Fast Data focuses on delivering instant awareness and instant actions to businesses and users. It often relies on and leverages Big Data sources but adds the distinct real-time capabilities by providing instant actionable results to businesses based on live, up-to-the-second data.

GridGain’s In-Memory Accelerator For Hadoop enhances existing Hadoop technology to enable Fast Data processing using the tools and technology your organization is already using today.

© Copyright 2007-2013 GridGain Systems, Inc.

Why Fast Data Matters To Your Organization or Project

The best way to answer this question is to examine how Fast Data is used in different industries by some of GridGain's customers today.

An electric power plant uses Fast Data technology to make real time decisions when demand spikes. In such a case the power plant has to decide whether to turn on additional production capacity, buy the required power on the spot market, or let someone else fulfill the demand. That decision depends on a multitude of factors such as current weather forecast, historical trends regarding demand and usage, immediate cost-benefit analysis, and current prices on the energy spot market. A Fast Data system can collect live streaming information and aggregate it with existing historical data stored in a Big Data system to provide real-time decision-making capability.

A wireless telecommunication provider is using Fast Data to help manage its resources more effectively. This starts with optimizing capital expenditure (CAPEX) on network infrastructure while lowering or maintaining operational expenditure (OPEX). To achieve this requires the ability to develop real-time insights to understand allocation of network resources based on traffic, specific application requirements and network usage patterns. Ultimately, Fast Data can help gain real-time insights derived from the live data instead of relying solely on approximate trending on historical data.

An analytics company that provides big data analytics services and software to their clients uses Fast Data to accelerate their interactive/ad hoc analytics. Their users interactively analyze customer interactions from a variety of sources such as Integrated Voice Response (IVR) data and authenticated and non-authenticated browsing history from their customers' corporate web properties for insight and patterns as to the effectiveness of marketing, support and promotional materials. By providing Fast Data to their end users they are able to empower their users to explore and discover, in real-time, previously hidden insights into what makes a successful customer interactions.

Finally, an online ad serving company uses Fast Data technology to fuse live clickstreams with pre-built insights and a behavioral data set that is based on collected historical data. Fast Data can process real-time information about millions of events per second into business intelligence and insights. These insights in turn help drive optimized and personalized ad placement based on each customer's experience. Fast Data collects data about what customers are currently doing, or how they have recently interacted with the company through other various channels, including purchasing, social media and email - leading to an understanding of the total customer experience and, ultimately, better conversion rates.

These are just a few of the many ways in which organizations are using, and will be using, Fast Data to augment the power of Big Data. Fast Data is fast becoming one of the top tools for organizations trying to keep up with information coming from various sources and make real-time decisions that serve the need of their customers and the business.

© Copyright 2007-2013 GridGain Systems, Inc.

In-Memory Accelerator For Hadoop vs. Other Solutions

What makes In-Memory Accelerator For Hadoop an unique solution?

To understand how In-Memory Accelerator For Hadoop differs from other solutions, let's examine three different approaches that an organization can take to introduce Fast Data into an existing Big Data system based on Hadoop.

External System To Hadoop

The unique characteristic of these solutions is their reliance on an external system that stores, often temporarily, an operational subset of the Big Data that needs to be processed faster and with lower latency. Typically this data subset is defined as a sliding, time-based window of data such as "last 24 hours of activity", "last month of sales", "last quarter of inventory", and so on.

Typically these systems are deployed alongside Hadoop in an up-stream or down-stream fashion. In the up-stream scenario these systems usually process incoming streaming data in a real-time context while asynchronously storing this data into long-term durable storage in Hadoop. In the down-stream scenario the data subset is manually or automatically ETL-ed from Hadoop into the system when it needs to be processed in real-time.

Solutions in this category are represented by a variety of products including standard SQL, MPP DBMS, new NoSQL and NewSQL DBMS, in-memory DBMS, and Streaming Processing systems. There are many products in this category with all of them invariably requiring a trade off between the time and material cost of implementation vs. optimization for high performance and low latency processing.

And despite the fact that these solutions can provide the true real-time and low latency processing required by Fast Data (and limited to a subset of the overall data stored in Hadoop) - they do require a substantial additional development effort to the existing Hadoop-based system which may limit their practical applicability.

System On Top Of Hadoop

The defining characteristic of these solutions is that they use Hadoop as a primary storage system and provide faster data processing capabilities on top of existing data stored in Hadoop HDFS without a need to move the data, even temporarily, elsewhere.

Examples in this category include HBase and HadoopDB - OLAP databases based on top of Hadoop - and various SQL interfaces for Hadoop like Cloudera Impala, DrawnToScale, as well as extensions to standard Hadoop Pig and Hive such as HortonWorks Stinger initiative or Apache Drill project. These products and projects employ a combination of sophisticated distributed indexing, MPP-style query optimization, relaxed consistency models or in-memory processing to gain high performance and low latency capabilities for processing data stored directly in Hadoop.

Solutions in this category strike a different balance than the ones we discussed above. While they require less integration and development, they also provide significantly smaller

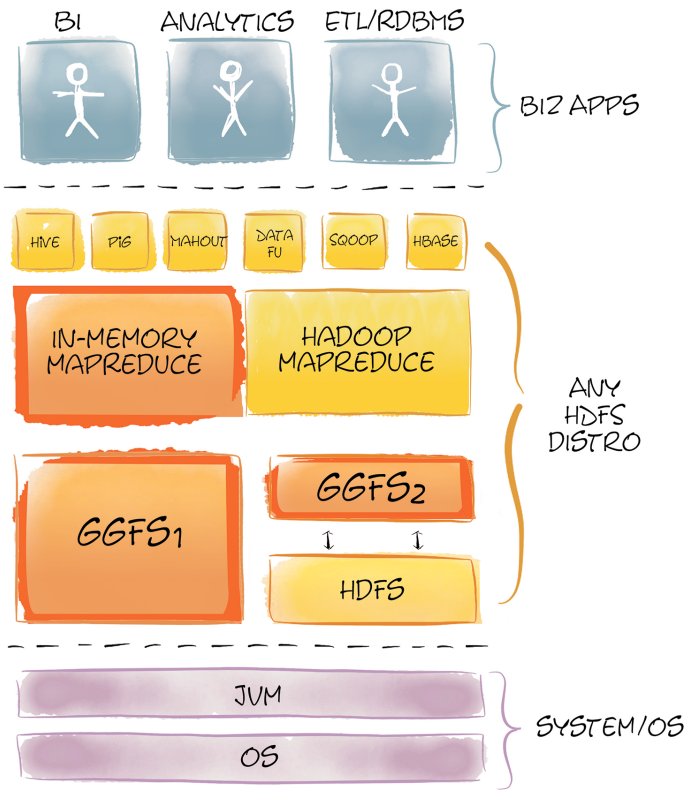
© Copyright 2007-2013 GridGain Systems, Inc.

performance gain. While they can be used to achieve Fast Data processing, these solutions are often primarily selected for other reasons such as a familiar SQL interface or desired OLAP database functionality. Systems on top of Hadoop are also not a good choice for Streaming Processing due to the fact that they are still limited by standard HDFS - the underlying storage technology in Hadoop.

Plug-n-Play Hadoop Accelerator

In-Memory Accelerator For Hadoop was designed to eliminate the trade-offs when adding Fast Data capabilities to existing Hadoop systems. Compared to external systems and systems on top of Hadoop, GridGain's In-Memory Accelerator For Hadoop delivers the three unique characteristics:

1. It requires **minimal or no additional integration** or development. It requires only minimal configuration change to existing Hadoop clusters and simple integration for In-Memory MapReduce. It works out-of-the-box with hundreds of projects in the Hadoop eco-system including standard HBase, Hive, Mahout, Oozie, Flume, Scoop, and Pig.
2. It works with **any existing Hadoop distribution**, open source or commercial - no need to roll out yet another proprietary distribution. You can continue to use your existing Apache, Pivotal, Intel, Cloudera, HortonWorks, MapR, AWS, or any other Hadoop 1.0 or Hadoop 2.0 (YARN) distributions.
3. It provides up to a **100x performance** increase for IO, network or CPU intensive Hadoop MapReduce job and HDFS operation - delivering easy and quick acceleration to existing Hadoop-based systems and products.



One of the key elements of In-Memory Accelerator For Hadoop's architecture is GGFS - a dual-mode, high performance in-memory file system. Due to its dual-mode design, GGFS can work as either a standalone primary file system in the Hadoop cluster, or work in tandem with existing HDFS, providing an intelligent caching layer for the primary HDFS. When GGFS is used as a standalone primary file system it brings a host of its own unique additional benefits to the Hadoop cluster:

1.**Simplified Deployment.** Unlike the Hadoop master-slave architecture, GridGain's In-Memory Accelerator For Hadoop is based on peer-to-peer topology and does away with master-slave failover, zookeeper installation or NFS setup for secondary NameNode.

© Copyright 2007-2013 GridGain Systems, Inc.

Specifically, In-Memory Accelerator For Hadoop in a standalone mode eliminates the need for three Hadoop components: NameNode, Secondary NameNode and DataNode, which significantly simplifies Hadoop configuration and deployment.

2. **Automatic Failover Without Shared Storage.** Unlike a standard Hadoop installation that requires shared storage for primary and secondary NameNodes which is usually implemented with a complex NFS setup mounted on each NameNode machine, In-Memory Accelerator For Hadoop seamlessly utilizes GridGain's In-Memory Data Grid that provides completely automatic scaling and failover without any need for additional shared storage or risky Single Point Of Failure (SPOF) architectures.

3. **Improved Scalability and Availability.** Unlike Hadoop's master-slave topology (specifically a NameNode component) that prevents it from linear runtime scaling when adding new nodes, GridGain's In-Memory Accelerator For Hadoop is built on a highly scalable, natively distributed partitioned data grid that provides linear scalability and auto-discovery of new nodes. It was independently tested to provide linear scalability with over 2000 nodes in the cluster.

100x Faster, Minimal Integration, Any Distribution

What are the key In-Memory Accelerator For Hadoop features comparing to other vendors products?

GridGain's In-Memory Accelerator For Hadoop architecture is based on the industry's first **dual-mode in-memory file system** that is 100% compatible with Hadoop Distributed File System (HDFS) and **In-Memory MapReduce**.

GridGain File System is a plug-and-play alternative to the disk-based HDFS enabling up to 100x faster performance for IO, CPU or network intensive Hadoop MapReduce jobs running on tens and hundreds of computers in a typical Hadoop cluster.

It is important to note that In-Memory Accelerator For Hadoop is built on top of two of GridGain's core platform products: **In-Memory HPC** and **In-Memory Data Grid.**

These products provide infrastructure services and functionality such as cluster and resource management, high-performance distributed partitioning and fully replicated caching with HyperLocking and off-heap memory support, a high-performance execution framework, cluster-aware peer-to-peer zero Java deployment and provisioning, comprehensive security, a SPI-architecture for pluggable system services, advanced load balancing and pluggable fault tolerance.

No ETL Required

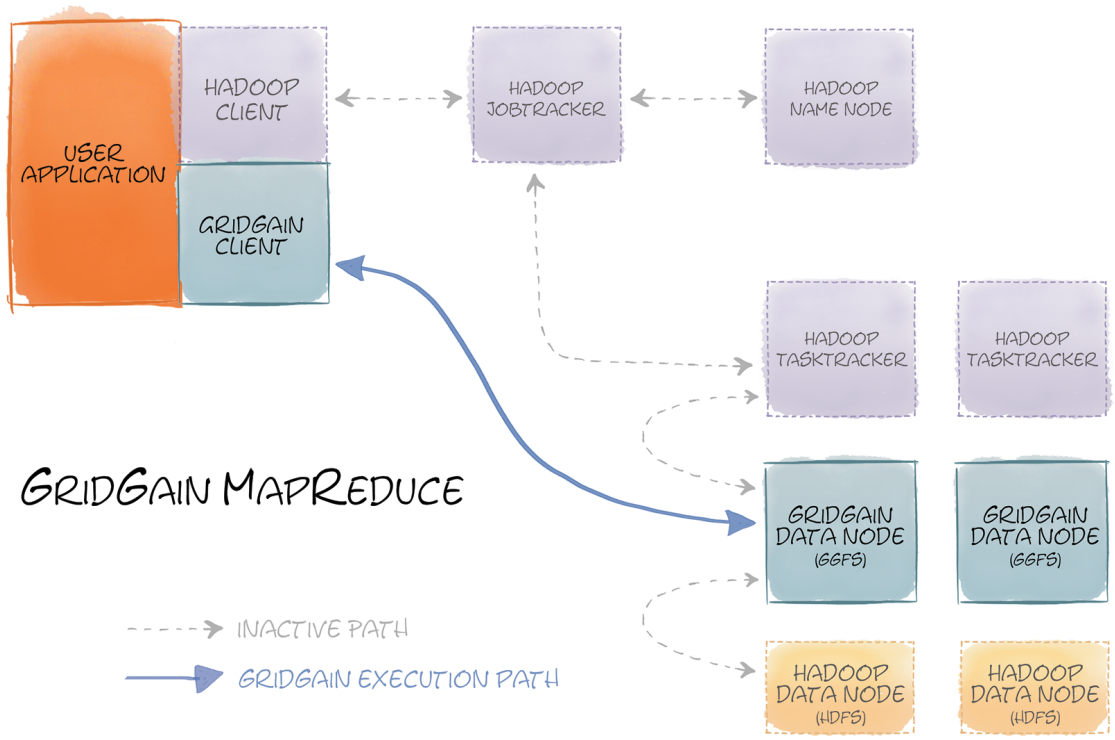
GridGain Hadoop Accelerator's unique In-Memory File System allows it to work with data that is stored directly in Hadoop. Whether In-Memory File System is used in primary mode, or in secondary mode acting as an intelligent caching layer over the primary disk-based HDFS, it completely eliminates the time consuming and costly process of extracting, loading and transforming (ETL) data to and from Hadoop.

© Copyright 2007-2013 GridGain Systems, Inc.

The ETL-free architecture of GridGain's In-Memory Hadoop Accelerator enables companies to process live data in Hadoop without the need to offload it to other downstream system to gain the in-memory computing performance advantage. GridGain's accelerator avoids duplication of data and eliminates unnecessary data movement that typically clogs the network and I/O subsystems.

Eliminate Hadoop MapReduce Overhead

For CPU-intensive and real-time use cases In-Memory Accelerator for Hadoop relies on in-memory MapReduce implementation that eliminates standard overhead associated with typical Hadoop's job tracker polling, task tracker process creation, deployment and provisioning. GridGain's in-memory MapReduce is a highly optimized HPC-based implementation of MapReduce concept enabling true low-latency data processing of data stored in GGFS:



Boost HDFS Performance

In-Memory Accelerator For Hadoop ships with transparent benchmarks that compare GGFS and HDFS performance against the same set of operations. These benchmarks indicate an average of 10x performance increase for file system operations.

The following tests were performed on a 10-node cluster of Dell R610 blades with Dual 8-core CPUs, running Ubuntu 12.4 OS, 10GBE network fabric and stock unmodified Apache Hadoop 2.x distribution:

© Copyright 2007-2013 GridGain Systems, Inc.

Benchmark	GGFS, ms	HDFS, ms	Boost, %
File Scan	27	667	2470%
File Create	96	961	1001%
File Random Access	413	2931	710%
File Delete	185	1234	667%

Hadoop 1.x and Hadoop 2.x Support
GridGain In-Memory Accelerator For Hadoop provides out-of-the-box support for both legacy Hadoop 1.x and new Hadoop 2.x (YARN) distributions. This allows organizations that utilize older Hadoop 1.x distributions to get full benefits of In-Memory HDFS while preserving their investment when moving to the latest Hadoop 2.x distributions.

Speed Up Java/Scala/C/C++/Python MapReduce Jobs
Hadoop Accelerator architecture allows it to speed MapReduce jobs written in any Hadoop supported language and not only in native Java or Scala. Developers can easily reuse existing C/C++/Python or any other existing MapReduce code with In-Memory Accelerator for Hadoop to gain significant performance boost.

Any Hadoop Distribution
As mentioned above, In-Memory Accelerator For Hadoop isn't another proprietary Hadoop distribution, rather, i's a first-of-its-kind Hadoop accelerator that works with your choice of Hadoop distribution, which can be any commercial or open source version of Hadoop available, and provides the same performance benefits whether you run Cloudera, HortonWorks, MapR, Apache, Intel, AWS, or any other distribution.

Plug-n-Play 100% Compatible With HDFS
As the industry's first plug-n-play Hadoop accelerator, In-Memory Accelerator For Hadoop does not require any code change to use its in-memory file system. The unique "plug-in" architecture gives you the freedom to not only chose any Hadoop distribution but also use any of the dozens of Hadoop-based tools that your organization already has in place . Whether you use standard tools such as HBase, Hive, Mahout, Oozie, Flume, Scoop, or Pig, or any of the commercial BI, data visualization or data analytics platforms, you can continue to use them without any change while realizing an instant performance boost.

Dual-Mode Operation
As mentioned above, In-Memory Accelerator For Hadoop's GGFS can work in dual mode:
1. Primary 100% HDFS compatible in-memory standalone file system
2. Intelligent caching layer for the primary HDFS

To support these two use cases any number of subsets of files or directories in GGFS can be configured with one of the following four different operation modes:

© Copyright 2007-2013 GridGain Systems, Inc.

PRIMARY	In this mode GGFS serves as a primary, standalone, distributed in-memory file system. All files and directories with this mode will be stored in memory only.
PROXY	In this mode GGFS serves as a proxy which will always delegate to HDFS without caching anything in memory. Files and directories configured with this mode will be stored on underlying HDFS via pass through read and write.
DUAL_SYNC	In this mode GGFS will synchronously read-through from HDFS whenever data is requested and is not cached in memory, and synchronously written through to HDFS whenever data is updated or created in GGFS. Essentially, in this case GGFS serves as an intelligent caching layer on top of HDFS. All files and directories configured with this mode will be simultaneously stored in memory as well as in underlying HDFS.
DUAL_ASYNC	In this mode GGFS will synchronously read-through from HDFS whenever data is requested and is not cached in memory (just like in DUAL_SYNC mode), and asynchronously written-through to HDFS whenever data is updated or created in GGFS. Since data is modified in HDFS asynchronously, there is a lag between GGFS updates and HDFS updates, however the performance of updates is significantly faster than using HDFS directly. Essentially, in this case GGFS again serves as an intelligent caching layer on top of HDFS.

File-Based and Block-Based LRU Eviction

GGFS supports data sets that are significantly larger than the memory available on all participating cluster nodes. For example, if you have 10 nodes with 120GB of RAM each, then GGFS will be able to cache about 1TB of data total across 10 grid nodes, however, the total working set managed by In-Memory HDFS may be in petabytes. Such behavior is achieved via smart eviction policies implemented within GGFS.

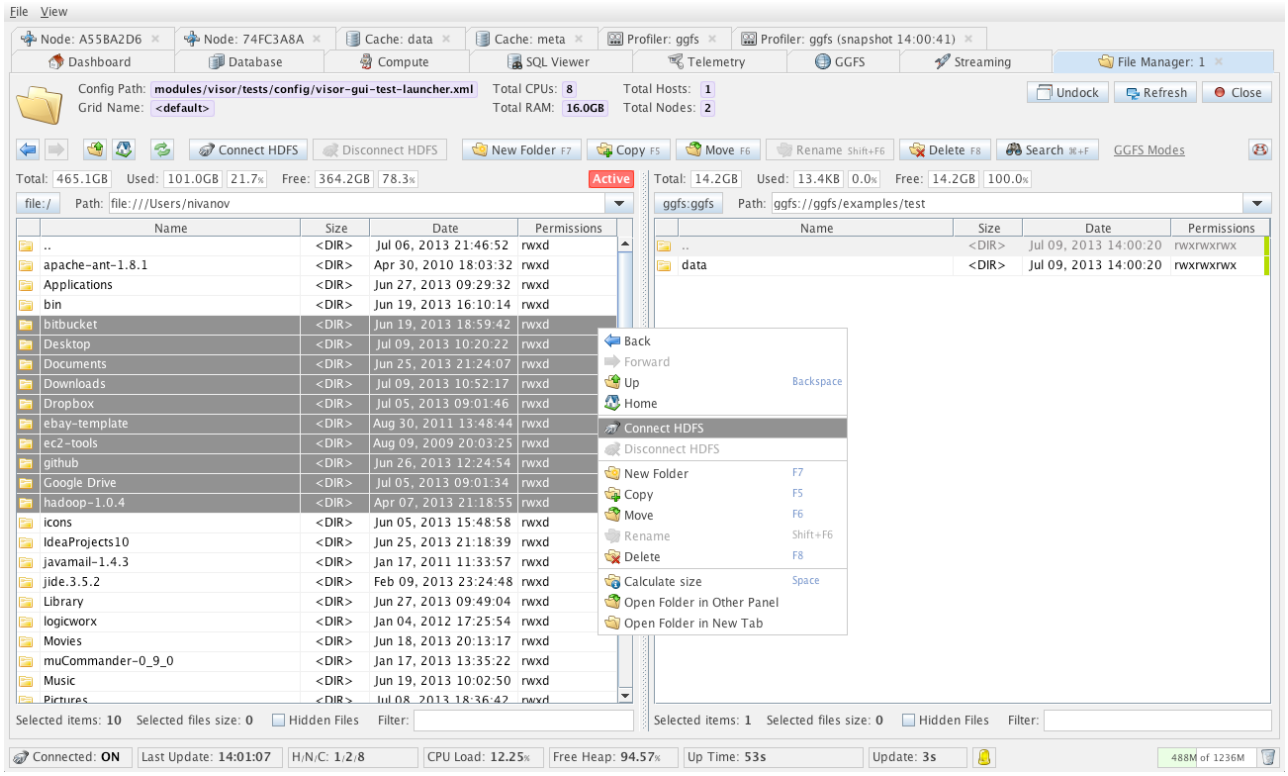
Whenever read-through or write-through is enabled, GGFS utilizes a block-based eviction policy and evicts the oldest unused file blocks from memory. This is safe because whenever evicted data will be requested in future, it will be automatically read-through from underlying disk-based HDFS file system.

If GGFS works as a primary file system, without HDFS underneath, then evicting individual file blocks would essentially corrupt the in-memory file. In this case GGFS will evict the oldest unused file as a whole, instead of individual blocks, to free up memory for new data. However,

© Copyright 2007-2013 GridGain Systems, Inc.

even in this case, the configuration is flexible and you can configure any subset of files to never be evicted at all.

GUI-Based File Management



In-Memory Accelerator For Hadoop comes with a comprehensive unified GUI-based management and monitoring tool called GridGain Visor. It provides deep administration capabilities including an operations & telemetry dashboard, data grid and compute grid management, as well as GGFS management that provides GGFS monitoring and file management between HDFS, local and GGFS file systems.

Pre-Fetching And Streaming

To ensure seamless and continuous performance during MapReduce file scanning, GGFS does smart data prefetching via streaming data that is expected to be read in the nearest future to the MapReduce task ahead of time.

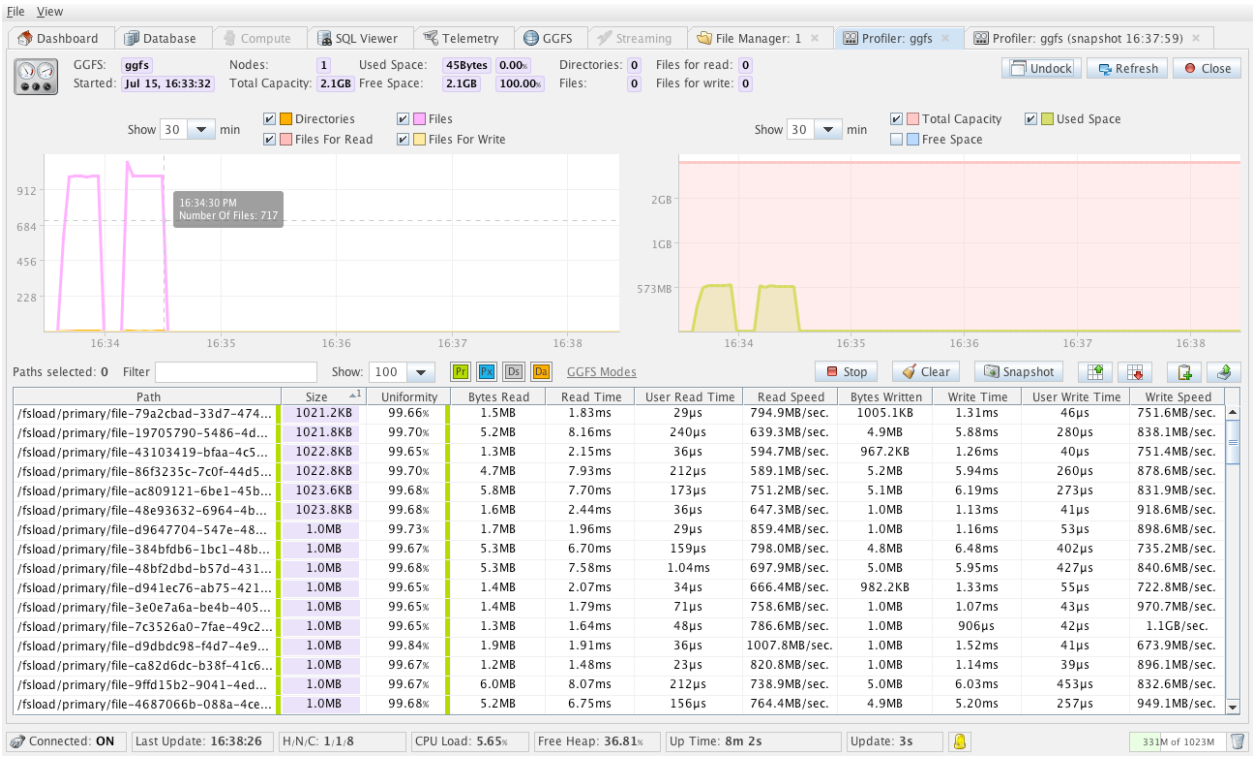
By doing so, GGFS ensures that whenever a MapReduce task finishes reading a file block, the next file block is already available in memory. A significant performance boost is achieved here due to proprietary Inter-Process Communication (IPC) implementation which allows GGFS to achieve throughput up to 30 Gigabit/second between two processes.

GUI-Based HDFS/GGFS Profiler

As part of GridGain Visor In-Memory Accelerator For Hadoop also comes with a GUI-based file system profiler, which allows to keep track of all operations your GGFS or HDFS file systems

© Copyright 2007-2013 GridGain Systems, Inc.

make and identifies potential hot spots. GGFS profiler tracks speed and throughput of reads, writes, various directory operations, for all files and displays these metrics in a convenient view which allows you to sort based on any profiled criteria, e.g. from slowest write to fastest.



Profiler also makes suggestions whenever it is possible to gain performance by loading file data into in-memory GGFS.

Read-Through And Write-Through HDFS Caching

When GGFS is used as an intelligent caching layer on top of the primary HDFS, if data is read from GGFS and this data is not currently in memory it will be read-through from HDFS on first access and then remain in the memory cache. The reverse happens whenever a write occurs - the data will get written into memory and it will also get either synchronously or asynchronously written to HDFS.

When reading and writing to and from HDFS is enabled, data cached in GGFS always gets persisted to the on-disk file system. This means that the data will survive any GGFS failures and data will be reloaded transparently on-demand whenever GGFS is restarted.

End-to-End Stack & Total Integration

What are different GridGain editions?

© Copyright 2007-2013 GridGain Systems, Inc.

GridGain provides full end-to-end stack for in-memory computing: from high performance computing, streaming, and data grid to Hadoop accelerators, GridGain delivers a complete platform for low-latency, high performance computing for each and every category of payloads and data processing requirements. Total integration is further extended with a single unified management and monitoring console.

Platform Products
GridGain's platform products are designed to provide uncompromised performance by providing developers with a comprehensive set of APIs. Developed for the most demanding use cases, including sub-millisecond SLAs, platform products allow to programmatically fine-tune large and super-large topologies with hundreds to thousands of nodes.

In-Memory HPC	Highly scalable distributed framework for parallel High Performance Computing (HPC).
In-Memory Data Grid	Natively distributed, ACID transactional, SQL and MapReduce based, in-memory object key-value store.
In-Memory Streaming	Massively distributed CEP and Stream Processing system with workflow and windowing support.

GridGain Foundation Layer

What are the common components across all GridGain editions?

GridGain foundation layer is a set of components shared across all GridGain products and editions. It provides a common set of functionality available to the end user such clustering, high performance distributed messaging, zero-deployment, security, etc. These components server as an extensive foundation layer for all products designed by GridGain.

Hyper Clustering®
GridGain provides one of the most sophisticated clustering technologies on Java Virtual Machine (JVM) based on its Hyper Clustering® technology. The ability to connect and manage a heterogenous set of computing devices is at the core GridGain's distributed processing capabilities.

Clustering capabilities are fully exposed to the end user. The developers have full control with the following advanced features:

- Pluggable cluster topology management and various consistency strategies
- Pluggable automatic discovery on LAN, WAN, and AWS
- Pluggable "split-brain" cluster segmentation resolution
- Pluggable unicast, broadcast, and Actor-based cluster-wide message exchange
- Pluggable event storage
- Cluster-aware versioning
- Support for complex leader election algorithms

© Copyright 2007-2013 GridGain Systems, Inc.

- On-demand and direct deployment
- Support for virtual clusters and node groupings

Zero Deployment®
The zero deployment feature means that you don't have to deploy anything on the grid – all code together with resources gets deployed automatically. This feature is especially useful during development as it removes lengthy Ant or Maven rebuild routines or copying of ZIP/JAR files. The philosophy is very simple: write your code, hit a run button in the IDE or text editor of your choice and the code will be automatically be deployed on all running grid nodes. Note that you can change existing code as well, in which case old code will be undeployed and new code will be deployed while maintaining proper versioning.

Advanced Security
GridGain security component provides two levels by which security is enforced: cluster topology and client connectivity. When cluster-level security is turned on, unauthenticated nodes are not allowed to join the cluster. When client security is turned on, remote clients will not be able to connect to the grid unless they have been authenticated.

SPI Architecture And PnP Extensibility
Service Provider Interface (SPI) architecture is at the core of every GridGain product. It allows GridGain to abstract various system level implementations from their common reusable interfaces. Essentially, instead of hard coding every decision about internal implementation of the product, GridGain instead exposes a set of interfaces that define the GridGain's internal view on its various subsystem. Users then can use either provided built-in implementations or roll out their own when they need different functionality.

GridGain provides SPIs for 14 different subsystems all of which can be freely customized:

- Cluster discovery
- Cluster communication
- Deployment
- Failover
- Load balancing
- Authentication
- Task checkpoints
- Task topology resolution
- Resource collision resolution
- Event storage
- Metrics collection
- Secure session
- Swap space
- Indexing

Having ability to change the implementation of each of these subsystems provides tremendous flexibility to how GridGain can be used in a real-world environment. Instead of demanding that

© Copyright 2007-2013 GridGain Systems, Inc.

other software should accommodate GridGain, GridGain software blends naturally in almost any environment and integrates easily with practically any host eco-system.

Remote Connectivity
GridGain products come with a number of Remote Client APIs that allow users to remotely connect to the GridGain cluster. Remote Clients come for multiple programming languages including Java, C++, REST and .NET C#. Among many features the Remote Clients provide a rich set of functionality that can be used without a client runtime being part of the GridGain cluster: run computational tasks, access clustering features, perform affinity-aware routing of tasks, or access in-memory data grid.

Summary

The case for in-memory computing is actively winning converts. Analyst firm Gartner says that in 2012, 10% of large and medium-sized organizations had adopted in-memory computing in some capacity. By 2015, that figure will have more than tripled to 35%.

GridGain's new In-Memory Accelerator For Hadoop product extends the performance value chain to Hadoop distributions while also significantly cutting an organization's storage costs. It's flexibility, scalability and plug-n-play architecture allow for seamless integration and improved velocity of analytics and reporting.

###

© Copyright 2007-2013 GridGain Systems, Inc.