

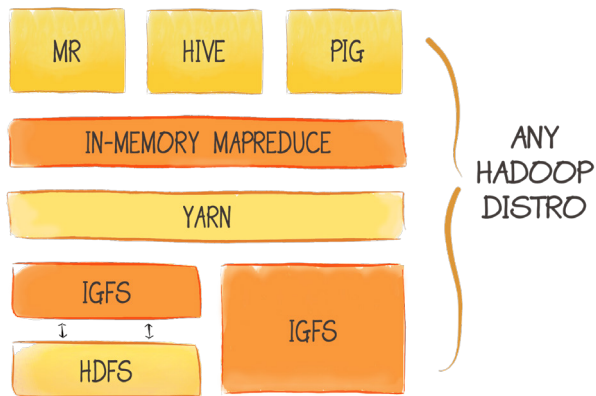
# GRIDGAIN IN-MEMORY ACCELERATOR FOR HADOOP® / SPARK™

POWERED BY Ignite™

## 10X FASTER IN 10 MINUTES – ZERO CODE CHANGE

The GridGain In-Memory Accelerator for Hadoop is a purpose-built product developed on top of the GridGain In-Memory Data Fabric Powered by Apache Ignite™. It is based on the industry’s first dual-mode, high-performance In-Memory File System, which is 100% compatible with Hadoop HDFS, as well as a fully plug-and-play and optimized in-memory implementation of MapReduce for YARN. In-memory HDFS and In-Memory MapReduce provide easy to use extension to disk-based HDFS and traditional MapReduce delivering up to 10x faster performance on selected payloads. The In-Memory Accelerator for Hadoop also provides easy, plug-and-play acceleration for HIVE jobs.

The In-Memory Accelerator for Hadoop provides plug-and-play integration, requires no code change, and works with Apache™ open source and commercial Hadoop 2.x distributions.



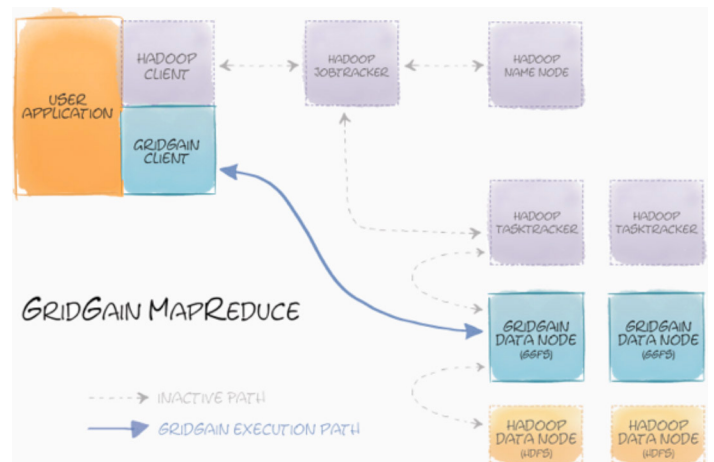
## Key Features

### NO ETL REQUIRED

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

## ELIMINATE HADOOP MAPREDUCE OVERHEAD

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



## SPEED UP JAVA/SCALA/C/C++/PYTHON MAPREDUCE JOBS

The architecture of the In-Memory Accelerator for Hadoop 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 the In-Memory Accelerator for Hadoop to gain significant performance boost.

## SHARED IN-MEMORY STORAGE FOR APACHE SPARK™

GridGain provides shared in-memory storage for Spark jobs. It supports it via a HDFS-compliant In-Memory File System and a native Shared RDD implementation. Based on the Apache Ignite In-Memory Data Fabric, the shared in-memory storage allows Spark jobs to exchange data up to 100x faster than what is possible with traditional approaches. Ignite also accelerates SQL execution dramatically for Spark jobs, and like Spark, is a Top Level Project of the Apache Software Foundation.



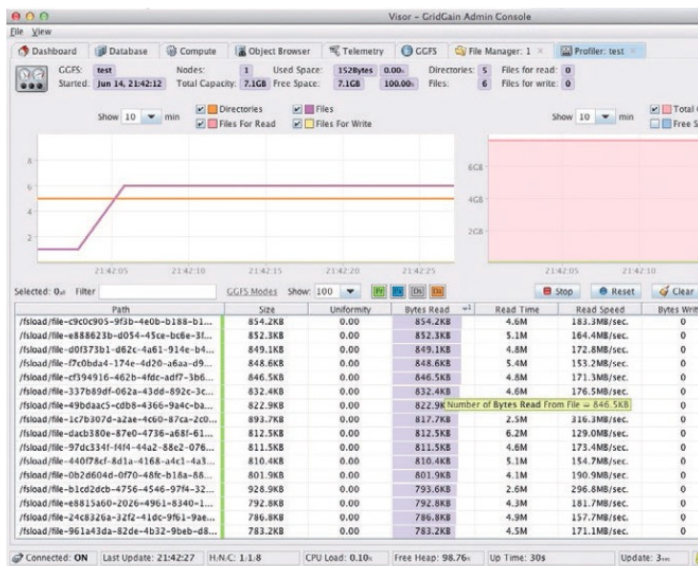
## BOOST HDFS PERFORMANCE

GridGain File System (GGFS) is 100% compatible with HDFS and can act as a drop-in replacement or extension for HDFS. Benchmarks below compare raw GGFS and HDFS performance against the same set of operations:

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%

## READ-THROUGH AND WRITE-THROUGH HDFS CACHING

When GGFS is used as an intelligent caching layer on top of the primary HDFS file system, once data is read from GGFS and it's not currently in memory, it will be read through from HDFS on first access and then remain in the memory cache. When a write occurs, the data will get written into memory and will also get either synchronously or asynchronously written to HDFS.



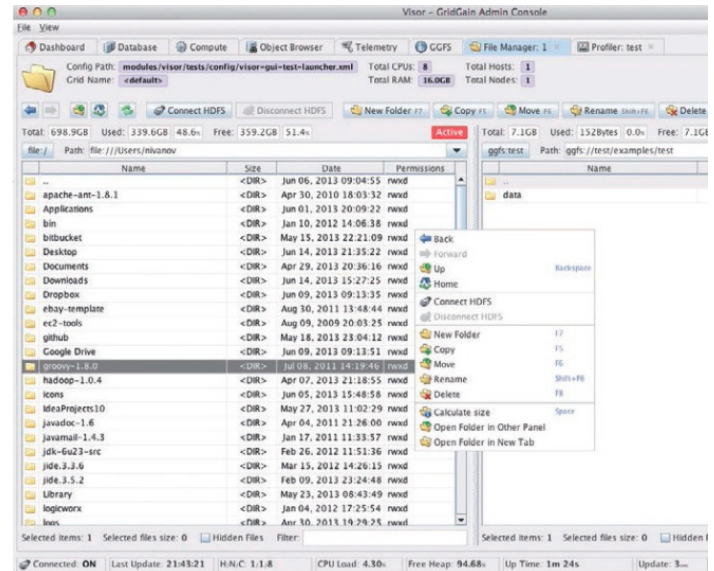
## FILE-BASED AND BLOCK-BASED LRU EVICTION

GGFS supports data sets that are significantly larger than the memory available on all participating cluster nodes. Whenever read-through or write-through is enabled, GGFS utilizes a block-based eviction policy and evicts the oldest unused file blocks from memory. Whenever evicted data is requested in the future, it will be automatically read through from the underlying disk-based HDFS file system.



## GUI-BASED GGFS PROFILER

Keeps track of all operations your GGFS or HDFS file systems make and identifies potential hot spots. It tracks speed and throughput of reads, writes and various directory operations for all files, and displays these metrics in a GUI view, allowing you to sort based on any profiled criteria, e.g. from slowest write to fastest. Profiler also makes suggestions whenever it's possible to gain performance by loading file data into in-memory GGFS.



## GUI-BASED FILE MANAGER

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

## ABOUT GRIDGAIN™

GridGain is revolutionizing real-time data access and processing by offering the first enterprise-grade In-Memory Data Fabric powered by Apache Ignite™. Apache Ignite™ is a top-level project (TLP) by the Apache Software Foundation, which integrates a compute grid, data grid, service grid and streaming grid, as well as Hadoop/Spark integration into a single solution that connects traditional and emerging data stores (SQL, NoSQL, Hadoop) with cloud-scale applications. The GridGain In-Memory Data Fabric enables massive data throughput and ultra-low latencies across any number of clustered commodity servers; it is uniquely designed to conquer today's Fast Data challenges and unleash the competitive advantage of any real-time business, whether on-premises or in the cloud. Offering the most comprehensive, enterprise-grade in-memory computing solution for high-volume transactions, real-time analytics and hybrid data processing, GridGain enables Fortune 500 companies and innovative mobile, web and SaaS companies to anticipate and innovate ahead of market changes. GridGain is headquartered in Foster City, California. Download the GridGain In-Memory Data Fabric [here](http://gridgain.com). For more information, visit [gridgain.com](http://gridgain.com) and follow us [@gridgain](https://twitter.com/gridgain).