INTRODUCING APACHE IGNITE
An Apache Incubator Project

Table of Contents

Overview

In-Memory computing is characterized by using high-performance, integrated and distributed memory systems to compute and transact on large-scale data sets in real-time, orders of magnitude faster than possible with traditional disk-based or flash technologies.
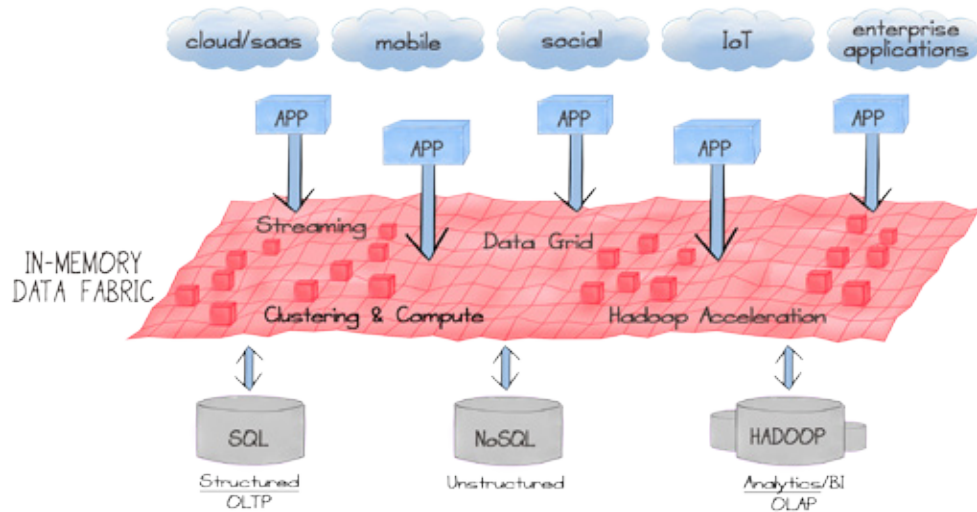
Apache Ignite (incubating) is an In-Memory Data Fabric that delivers unprecedented speed and unlimited scale to modern data processing. It enables high-performance transactions, real-time streaming and fast analytics in a single, comprehensive data access and processing layer. Apache Ignite is designed to easily power both existing and new applications in a distributed, massively parallel architecture on affordable, industry-standard hardware.

Apache Ignite provides a unified API that connects cloud-scale applications with multiple data stores containing structured, semi-structured and unstructured data (SQL, NoSQL, Hadoop). It offers a high-performance data environment that allows companies to process full ACID transactions and generate valuable insights from real-time, interactive and batch queries.

In-memory data fabrics offer a strategic approach to in-memory computing that delivers performance, scale and comprehensive capabilities far above and beyond what traditional in-memory databases (IMDBs), in-memory data grids (IMDGs) or other in-memory-based point solutions can offer by themselves.

Unlike IMDBs, Apache Ignite works on top of existing databases and hence requires no rip-and-replace of an existing RDBMS, or any changes to it. Users can keep their existing RDBMSs in place and deploy Apache Ignite as a layer above it. In fact, one of the best features of Apache Ignite is its ability to automatically integrate with different RDBMS systems, such as Oracle, MySql, Postgres, DB2, Microsoft SQL, and others; this feature automatically generates the application domain model based on the schema definition of the underlying database, and then loads the data. Moreover, IMDBs typically only provide a SQL interface, while Apache Ignite provides a much wider eco-system of supported access and processing paradigms, in addition to ANSI SQL. Apache Ignite supports key/value stores, SQL access, MapReduce, HPC/MPP processing, streaming/CEP processing and Hadoop acceleration – all in one well-integrated in-memory data fabric.

When comparing Apache Ignite with IMDGs, it should be noted that an in-memory data grid is just one of the capabilities that Apache Ignite provides; besides the data grid function, Apache Ignite also supports HPC/MPP processing, streaming, clustering, and Hadoop acceleration, allowing for a much broader set of use cases than a typical IMDG.

Architecture

Apache Ignite is a JVM-based distributed middleware software based on a homogeneous cluster topology implementation that does not require separate server and client nodes – all nodes in Apache Ignite cluster are equal and can play any logical role per runtime application requirement.

At the core of the Apache Ignite implementation is a Service Provider Interface (SPI) design. SPI-based design makes every internal component of Apache Ignite fully customizable and pluggable by the developer – enabling tremendous configurability of the system and its adaptability to any existing or future server infrastructure.

Another core tenet of Apache Ignite is the direct support for parallelization of distributed computations based on Fork/Join, MapReduce or MPP-style processing – the largest implementation eco-system of distributed processing algorithms. Distributed parallel computations are not only extensively used internally by Apache Ignite, but they are also fully exposed at the API level for user-defined functionality.

Key Features

IN-MEMORY DATA GRID

One of Apache Ignite's core capabilities is an in-memory data grid, which handles distributed in-memory data management including ACID transactions, failover and advanced load balancing, extensive SQL support and many other features. The data grid feature incorporated into the Apache Ignite In-Memory Data Fabric is a distributed, object-based, ACID transactional, in-memory key-value store. As such, Apache Ignite stores its data in memory as opposed to traditional Database Management Systems, which utilize disk as their primary storage mechanism. By utilizing system memory rather than disk, Apache Ignite is orders of magnitude faster than traditional DBMS systems.
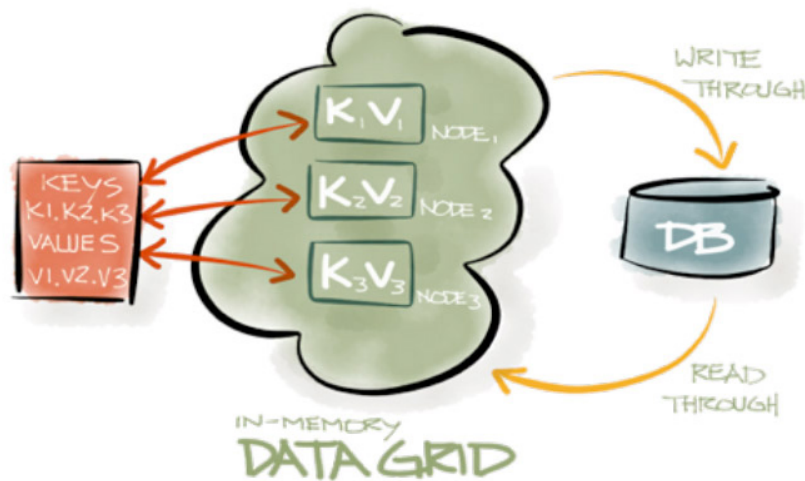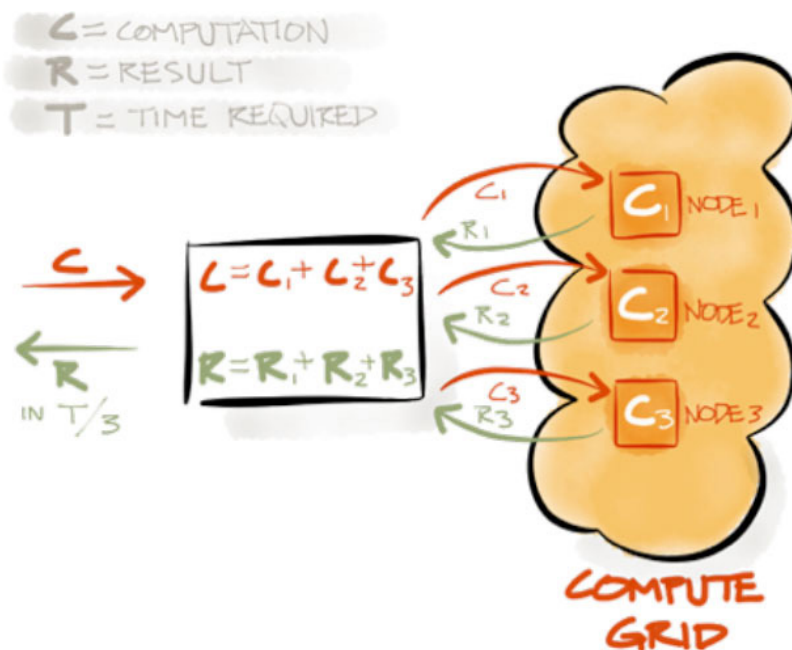

Main features of the **In-Memory Data Grid** in Apache Ignite include:

- Distributed in-memory caching
- Lightning-fast performance
- Elastic scalability
- Distributed in-memory ACID transactions
- Distributed in-memory queue and other data structures

- Web session clustering
- Hibernate L2 cache integration
- Tiered off-heap storage
- Distributed SQL queries with distributed joins

IN-MEMORY COMPUTE GRID

Besides the data grid capability, Apache Ignite also includes a Compute Grid, which provides the means for parallel, in-memory processing of CPU-intensive or other resource-intensive tasks, including traditional High Performance Computing (HPC) and Massively Parallel Processing (MPP).

Main features of the **In-Memory Compute Grid** in Apache Ignite include:

- Dynamic clustering
- Fork-Join & MapReduce processing
- Distributed closure execution
- Load balancing and fault tolerance
- Distributed messaging and events
- Linear scalability
- Standard Java ExecutorService support

IN-MEMORY SERVICE GRID

The Apache Ignite Service Grid allows you to have complete control over services being deployed on the cluster. It allows you to control how many instances of your service should be deployed on each cluster node, ensuring proper deployment and fault tolerance. The Service Grid guarantees continuous availability of all deployed services in case of node failures.

Main features of the **In-Memory Service Grid** in Apache Ignite include:

- Automatic deployment of multiple instances of a service
- Automatic deployment of a service as singleton
- Automatic deployment of services on node start-up
- Fault tolerant deployment
- Removal of deployed services
- Retrieval of service deployment topology information
- Remote access to deployed services via service proxy

IN-MEMORY STREAMING

In-memory streaming processing addresses a large family of applications for which traditional processing methods and disk-based storages, such as disk-based databases or file systems, fall short. Such applications are pushing the limits of traditional data processing infrastructures.

Streaming support allows for the ability to query into so-called rolling windows of incoming data, and to enable users to answer questions such as "What are the 10 most popular products over the last 2 hours?", or "What is the average product price in a certain category for the past day?".

Another most common use case for stream processing is the ability to control and properly pipeline a distributed events workflow. As events are coming into the system at high rates, the processing of events is split into multiple stages and each stage has to be properly routed within a cluster for processing.

Main features of the **In-Memory Streaming** in Apache Ignite include:

- Programmatic window-based querying
- Customizable event workflow / Complex Event Processing (CEP)
- At-least-once guarantee
- Built-in, user-defined sliding windows
- Streaming data indexing
- Distributed streaming queries
- Co-location with in-memory data grid

IN-MEMORY HADOOP ACCELERATION

Hadoop Acceleration is another capability of Apache Ignite which enhances existing Hadoop environments by enabling fast data processing using the tools and technology your organization is already using today.
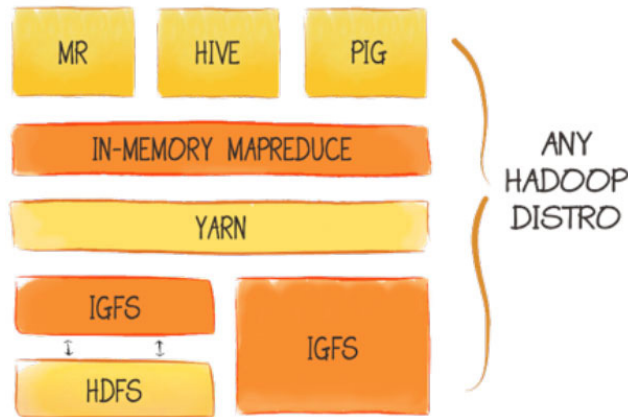
In-Memory Hadoop Acceleration in Apache Ignite is based on the industry's first dual-mode, high-performance in-memory file system that is 100% compatible with Hadoop HDFS, and an in-memory optimized MapReduce implementation. In-memory HDFS and in-memory MapReduce provide easy to use extensions to disk-based HDFS and traditional MapReduce, delivering up to 100x faster performance.

This feature requires minimal to no integration and works with open source Hadoop as well as any commercial version of Hadoop, including Cloudera, HortonWorks, MapR, Apache, Intel, AWS, as well as any other Hadoop 1.x and Hadoop 2.x distributions.

Main features of the **In-Memory Hadoop Acceleration** in Apache Ignite include:

- Up to 100x faster performance for MapReduce and HIVE jobs
- In-memory MapReduce
- Highly optimized in-memory processing
- Dual mode – standalone IGFS file system & primary caching layer for HDFS
- Highly tunable read-through and write-through behavior

DISTRIBUTED IN-MEMORY FILE SYSTEM

One of the unique capabilities of Apache Ignite is a file system interface to its in-memory data called Ignite File System (IGFS). IGFS delivers similar functionality to Hadoop HDFS, including the ability to create a fully functional file system in memory. In fact, IGFS is at the core of the Apache Ignite In-Memory Hadoop Acceleration feature.

The data from each file is split on separate data blocks and stored in cache. Developers can access each file's data with a standard Java streaming API. Moreover, for each part of the file a developer can calculate an affinity and process the file's content on corresponding nodes to avoid unnecessary networking.

Main features of the **Distributed In-Memory File System** in Apache Ignite include:

- Standard file system "view" on in-memory data
- Listing of directories or information for a single path
- Create/move/delete of files or directories
- Write/read of data streams into/from files

ADVANCED CLUSTERING

The Apache Ignite In-Memory Data Fabric provides one of the most sophisticated clustering technologies on Java Virtual Machines (JVM). With Apache Ignite, nodes can automatically discover each other. This helps to scale the cluster when needed, without having to restart the entire cluster. Developers can also take advantage of the hybrid cloud support in Apache Ignite that allows establishing connections between private clouds and public clouds such as Amazon Web Services.

Main features of the **Advanced Clustering** in Apache Ignite include:

- Dynamic topology management
- Automatic discovery on LAN, WAN, and AWS
- Automatic "split-brain" (i.e. network segmentation) resolution
- Unicast, broadcast, and group-based message exchange
- On-demand and direct deployment
- Support for virtual clusters and node groupings

DISTRIBUTED MESSAGING

Apache Ignite provides high-performance cluster-wide messaging functionality to exchange data via publish-subscribe and direct point-to-point communication models.

Main features of the **Distributed Messaging** in Apache Ignite include:

- Support for topic-based publish-subscribe model
- Support for direct point-to-point communication
- Pluggable communication transport layer
- Support for message ordering
- Cluster-aware message listener auto-deployment

DISTRIBUTED EVENTS

The distributed events functionality in Apache Ignite allows applications to receive notifications about cache events occurring in a distributed grid environment. Developers can use this functionality to get notified about the execution of remote tasks or any cache data changes within the cluster.

In Apache Ignite, event notifications can be grouped together and sent in batches and/or timely intervals. Batching notifications help attain high cache performance and low latency.

Main features of the **Distributed Events** in Apache Ignite include:

- Subscribing of local and remote listeners
- Ability to enable and disable any event
- Local and remote filters for fine-grained control over notifications
- Automatic batching of notifications for enhanced performance

DISTRIBUTED DATA STRUCTURES

Apache Ignite allows for most of the data structures from the java.util.concurrent framework to be used in a distributed fashion. For example, you can take java.util.concurrent.BlockingDeque and add to it on one node and poll it from another node. Or you could have a distributed Primary Key generator, which would guarantee uniqueness on all nodes.

Distributed Data Structures in Apache Ignite include support for these standard Java APIs:

- Concurrent map
- Distributed queues and sets
- AtomicLong
- AtomicSequence
- AtomicReference
- CountDownLatch

Deploying Apache Ignite in the Enterprise

GridGain is the commercial company behind Apache Ignite, offering the most comprehensive, enterprise-grade in-memory data fabric for high-volume transactions, real-time analytics and hybrid data processing. For enterprises who want to use Ignite in production, GridGain supplies the **GridGain In-Memory Data Fabric Enterprise Edition**, which includes:

- A certified version of Ignite that has passed GridGain's rigorous internal testing, quality assurance, and performance benchmarking process
- Portable Objects to seamlessly interoperate between Java, .NET and C++ applications
- Datacenter Replication to replicate remote data across two or more regions connected via wide area networks
- GUI-based Management and Monitoring
- Enterprise-grade Security to support the highest levels of protection for enterprise customers
- Network Segmentation Protection to prevent 'split brain' scenarios in the network
- Rolling Production Updates to prevent downtime when performing software updates
- Expert 9x5 and 24x7 Support
- Ongoing Maintenance Releases

A free, 30-day trial version of the GridGain In-Memory Data Fabric Enterprise Edition is available on www.gridgain.com. Production deployments require that a software subscription be purchased.

Summary

Apache Ignite is an incubating Apache project, originally developed by GridGain Systems, and the leading open source In-Memory Data Fabric. It provides a high-performance, integrated and distributed in-memory platform for computing and transacting on large-scale data sets in real-time, and orders of magnitude faster than possible with traditional disk-based or flash technologies. As an in-memory data management software layer, it sits between applications and various data sources, and does not require any rip-and-replace of existing databases.

Apache Ignite comprises a set of key in-memory data fabric capabilities, including a full-fledged in-memory data grid, an in-memory compute grid, an in-memory service grid, in-memory streaming processing as well as in-memory Hadoop acceleration – all in one, well-integrated framework. Despite the breadth of its feature set, Apache Ignite is very easy to use and deploy: There are no custom installers. The code base comes as one ZIP file, and it has only 1 mandatory dependency – *ignite-core.jar*; all other dependencies, such as integration with Spring for configuration, can be added to the process a la carte. The project is fully mavenized, and is composed of over a dozen of maven artifacts that can be imported and used in any combination. Apache Ignite is based on standard Java APIs, and for distributed caches and data grid functionality Ignite implements the *JCache (JSR107)* standard.

Apache Ignite's large scale, distributed in-memory framework offers transactional and analytical applications performance gains of 100-1,000 times faster throughput and/or lower latencies. Apache Ignite is an important open source foundation that holds the key to the world of Fast Data across high-volume transactions, real-time analytics and the emerging class of hybrid transactional-analytical workloads.

Next Steps

Community contributions are always welcome. If you would like to contribute, provide your own ideas or just like to find out where the project is headed, send an email to the Ignite dev list on the Apache Ignite Incubator program page. Another way, of course, to get familiar with the capabilities of Apache Ignite is to download the binaries on the Apache Ignite homepage.

ABOUT GRIDGAIN™

GridGain is revolutionizing enterprise data access and processing by offering the first enterprise-grade In-Memory Data Fabric built on Apache Ignite (incubating). The GridGain In-Memory Data Fabric is 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 a 30-day free trial of the GridGain In-Memory Data Fabric here. For more information, visit gridgain.com and follow us @gridgain.

ABOUT APACHE IGNITE

Apache Ignite (incubating) is the leading open source In-Memory Data Fabric integrating a compute grid, data grid, service grid, as well as streaming, messaging and Hadoop acceleration into a single, highly scalable real-time access and processing layer for Fast Data. It connects both traditional and emerging data stores (SQL, NoSQL, Hadoop) with hyper-scale applications, enabling massive data throughput and ultra-low latencies across any number of clustered commodity servers. For more information, visit the Apache Ignite website and follow the project @apacheignite.

For more information contact:
info@gridgain.com
650.241.2281