

IgniteProposal

Ignite Apache Incubator Proposal

Abstract

Apache Ignite will be a unified In-Memory Data Fabric providing high-performance, distributed in-memory data management software layer between various data sources and user applications.

Proposal

Apache Ignite is written mostly in Java and Scala with small amount of C++ code and will initially combine the following technologies under one unified umbrella:

- In-Memory Data Grid
- In-Memory Compute Grid
- In-Memory Streaming Processing

This unified in-memory fabric will provide high-performance, distributed in-memory software layer that sits in between various data sources and user applications. Data sources can include SQL RDBMS, NoSQL, or HDFS. Applications APIs will be available for Java (and Java-based scripting languages), Scala, C++ and .NET (C#).

GridGain Systems, Inc. submits this proposal to donate its Apache 2.0-licensed open source project generally known as "GridGain In-Memory Computing Platform", its source code, documentation, and websites to the Apache Software Foundation ("ASF") with the goal of extending the vibrant open source community around this technology ultimately governed by "Apache Way". Proposed Naming

We have been advised by the ASF mentors that the name "Ignite" may not be ideal because the name may be too generic and may not pass ASF legal check. Here are the alternatives that we have come up with and any of those will be acceptable for the project pending the ASF legal green light:

- Apache Silk (preferable name)
- Apache Sylk
- Apache Memstor
- Apache Ignite

Background & Rationale

In-Memory Data Fabric is a natural and evolutionary consolidation of various "in-memory technologies" from the last decade. From simple local caching (JSR-107), to distributed caching, to data grids and databases, to streaming and plug-n-play acceleration - the in-memory space has grown quite dramatically.

With rapid advances in NVRAM and significant price reduction of traditional DRAM on one hand, and growing sophistication and demand for faster data processing on another - many users of these silo-ed technologies and products started to look for a "strategic approach" to in-memory - an in-memory data fabric - that would provide suitable APIs for different types of payloads: from data caching, to data grids, to in-memory SQL data stores, to HPC, to streaming processing.

With expensive and proprietary in-memory computing products from companies like Oracle, SAP, Microsoft, and IBM - the developers worldwide need an unhindered access to advanced open source in-memory software technology, the technology they can trust to develop with and deploy for critical applications. Current Status

Apache Ignite will be based on the technology that is currently developed by [GridGain Systems](http://www.gridgain.org) and available under Apache 2.0 license (<http://www.gridgain.org>). The software has been in development since 2007 and in production since 2009. It is currently used by over 500 production deployments with over 1,000,000 downloads to date, and with over 20,000,000 [GridGain](http://www.gridgain.org) nodes started in the last 5 years.

Comparative analysis to relevant projects

Ignite vs. Spark

Apache Spark is a data-analytic and ML centric system that ingest data from HDFS or another distributed file system and performs in-memory processing of this data. Ignite is an In-Memory Data Fabric that is data source agnostic and provides both Hadoop-like computation engine (MapReduce) as well as many other computing paradigms like MPP, MPI, Streaming processing. Ignite also includes first-class level support for cluster management and operations, cluster-aware messaging and zero-deployment technologies. Ignite also provides support for full ACID transactions spanning memory and optional data sources. Ignite is a broader in-memory system that is less focused on Hadoop. Apache Spark is more inclined towards analytics and ML, and focused on MR-specific payloads.

Ignite vs. Storm, Samza

Apache Storm is streaming processing framework. Apache Samza is a distributed stream processing engine. Ignite is a multi-purpose In-Memory Data Fabric that also includes streaming processing capabilities (and we can argue better capabilities when it comes to streaming and CEP).

Generally, Apache Storm and Apache Samza provide a very different implementation for one of the functional areas of Ignite.

Ignite vs. Hadoop

Apache Hadoop is a batch oriented data warehouse system. Ignite is a real-time, transactional In-Memory Data Fabric focused on real-time processing of operational data. In most cases, Hadoop acts as one of the “slower” datasources on top of which Ignite is typically deployed.

Initial Goals

The number one goal during ASF incubation will coalesce around building a true active and vibrant community governed by the “Apache Way”. The initial development goals for Ignite primarily revolve around migrating the existing code base, documentation, and refactoring of the existing internal build, test & release processes. We believe these initial goals are sufficiently difficult to be considered early milestones.

Some of the specific initial goals include:

- Migrate the existing Ignite code base to the ASF.
- Refactor development, testing, build and release processes to work in ASF.
- Attract developer and user interest in the new Apache Ignite project.
- Road map the integration efforts with “sister” projects in ASF eco-system like Storm and Spark.
- Incorporate externally developed features into the core Apache Ignite project.

Known Risks

This proposal is not without its risks, some of which are outlined below.

The current list of committers are primarily from [GridGain Systems](#). One of the key purposes of proposing Ignite for incubation is to attract new committers and spur the adoption of Ignite. The ASF has a well-deserved reputation of fostering and building open source communities, which makes it the ideal location to attempt this community bootstrap.

Most of the initial committers are supported by their employers to work on Ignite, and may be assigned to work on other priorities. However, the employers of these salaried individuals - [GridGain Systems](#) or current customers and users - have a vested interest in seeing Ignite thrive as a long-term, growing project.

[GridGain Systems](#) understands that their employees are acting as individuals when contributing to Apache projects. As a major initial contributor [GridGain Systems](#) is prepared to bring additional staff on board to assist with Ignite development to ensure its active growth.

One of the key motivators in creating the Ignite project as part of the Incubator is to leverage the vendor-neutral nature of the ASF. The ASF has a strong and recognized brand as being a leader in open source, and by hosting Ignite at the ASF, we hope to attract developers to build a viable community for the project.

Meritocracy

Apache Ignite plans to adopt the policy that encourages an environment that supports a meritocracy. We intend to actively ask the community for help, listing/specifying the work that needs to be done, and keeping track of and encouraging members of the community who make any contributions. Community & Core Developers

[GridGain](#) project has been actively building community of users in the last couple of years with an active [StackOverflow](#) group, support groups, and Meetups (<http://www.meetup.com/Bay-Area-In-Memory-Computing>). This group includes active members of Apache community as well. We strongly believe that this community will grow and develop substantially as part of Apache family and that's our commitment.

Existing Documentation

Current documentation for [GridGain](#) project can be found here: <http://www.gridgain.org/documentation/> We intend to migrate it into ASF podling.

Initial Source

Initial Apache 2.0 licensed source code can be found here: <http://www.gridgain.org/download/>

External Dependencies

Here's the list of 3rd party JAR-only dependencies:

- Apache Hadoop
- Apache Commons
- H2
- JTS
- Apache Lucene
- Spring

Here's the list of the all licenses for 3rd party libraries currently used:

- Apache 2.0

Required Resources

Mailing lists

- private@ignite.incubator.apache.org (with moderated subscriptions)
- dev@ignite.incubator.apache.org
- committs@ignite.incubator.apache.org

Git & JIRA

- Git: <https://git-wip-us.apache.org/repos/asf/incubator-ignite.git>
- JIRA: JIRA Ignite (IGNITE)

Initial Committers & Affiliation

- Dmitry Setrakyan (GridGain Systems, [dsetrakyan at gridgain dot com](mailto:dsetrakyan@gridgain.com))
- Yakov Zhdanov (GridGain Systems, [yzhdanov at gridgain dot com](mailto:yzhdanov@gridgain.com))
- Alexey Goncharuk (GridGain Systems, [agoncharuk at gridgain dot com](mailto:agoncharuk@gridgain.com))
- Sergey Vladykin (GridGain Systems, [svladykin at gridgain dot com](mailto:svladykin@gridgain.com))
- Valentin Kulichenko (GridGain Systems, [vkulichenko at gridgain dot com](mailto:vkulichenko@gridgain.com))
- Semen Boikov (GridGain Systems, [sboikov at gridgain dot com](mailto:sboikov@gridgain.com))
- Vladimir Ozerov (GridGain Systems, [vvozerov at gridgain dot com](mailto:vvozerov@gridgain.com))
- Nikita Ivanov (GridGain Systems, [nivanov30 at gmail dot com](mailto:nivanov30@gmail.com))
- Sergey Khisamov (FitechSource, [skh at gmail dot com](mailto:skh@gmail.com))
- Ilya Sterin (ChronoTrack, [isterin at gmail dot com](mailto:isterin@gmail.com))
- Ryan Rawson (WANdisco, [rwrason at apache dot org](mailto:rwrason@apache.org))
- Konstantin Boudnik (WANdisco, [kboudnik at apache dot org](mailto:kboudnik@apache.org))
- Roman Shaposhnik (Pivotal, [rshaposhnik at pivotal dot io](mailto:rshaposhnik@pivotal.io))
- Branko Cibej (WANdisco, [bcibej at apache dot org](mailto:bcibej@apache.org))

Sponsors

Apache Champion

- Konstantin Boudnik ([cos at apache dot org](mailto:kboudnik@apache.org))

Nominated Mentors

- Michael Stack ([mstack at apache dot org](mailto:mstack@apache.org))
- Roman Shaposhnik ([rshaposhnik at apache dot org](mailto:rshaposhnik@pivotal.io))
- Konstantin Boudnik ([cos at apache dot org](mailto:kboudnik@apache.org))
- Henry Saputra ([hsaputra at apache dot org](mailto:hsaputra@apache.org))
- Branko Cibej ([brane at apache dot org](mailto:bcibej@apache.org))

Sponsoring Entity

- Apache Incubator PMC