UnomiProposal

Apache Unomi incubation proposal

Abstract

Unomi is a reference implementation of the OASIS Context Server specification currently being worked on by the OASIS Context Server Technical Committee. It provides a high-performance user profile and event tracking server.

Proposal

The server is a data management platform that provides real-time user profile and event storage as well as a rule engine and fully featured REST+JSON interfaces which make integration with very different systems easier.

One of the main objectives of this implementation is to be modular and flexible enough to relatively easily build integration into systems such as CMS, WCM, CRM or even ERP. Built using Apache Karaf and ElasticSearch, it is designed to also be easily clusterable to address high-traffic scenarios where lots of profile data will need to be updated at a high frequency.

Background

Unomi was built from scratch to serve as a reference implementation for the OASIS Context Server proposal. It has been developed by Jahia Software.

It initially served as a test-bed while working on the specifics of the new proposal, to make sure that the concepts being discussed and decided would make sense in an actual implementation. At the same time, Jahia, one of the companies involved in the standardization work, needed a basis to build their new Marketing Factory product, so it made a lot of sense to start this way. Thanks to a lot of work done in the standard, the implementation was produced relatively quickly and this made it possible to build a commercial product on top of this implementation that will be released in Winter 2015. This implies that the core of the commercial product being proposed here will undergo a full set of QA tests, including and scalability tests.

Another important objective was to provide a platform that deals with user data in an "ethical" way, in order to make the usage of personal data more transparent and to improve privacy. Through the work in the standard and this open source implementation, it is the hope of the original designers that it will be easier for all users of this technology to fully understand how the data is being used and what type of control is exercised over it. A transparent usage of customer data could become, in the near future, a key tool to keep customer loyalty. Companies would then be able to leverage digital marketing frameworks capabilities such as Unomi, but, as a fair - and quite possibly legally mandatory - counterpart, should allow end users to understand the data being used, as well as the way it is used, in order to give them more control over which data to share or to anonymise.

The purpose is to bring Jahia Marketing Factory codebase into the Apache Software Foundation (ASF) in order to build a vibrant, diverse and selfgoverned open source community around the technology. Jahia will continue to market and sell Jahia Marketing Factory based on Unomi. Unomi and Jahia Marketing Factory will be managed separately. This proposal covers the Unomi source code (mainly written in Java), Unomi documentation and other materials currently available on GitHub. Unomi is our primary choice for a name of the project.

Rationale

Building highly scalable, performant and flexible personalization architectures requires major efforts. The complexities of requirements, integrations and use cases is directly related to the number of users, and Apache projects usually get a lot of visibility and usage. Therefore it was a goal from the start to share this project with the ASF, as it is believed this is the best way to achieve the true potential of this codebase, to make it highly visible and continually improve on its design and implementation to make it the best tool for the job for a wide variety of use cases.

The project, as it sits a bit higher than usual BigData frameworks, could be used a front-end to a lot of existing Apache project that could be integrated to help deliver more power into the platform. For example Unomi doesn't currently include machine learning but it could very well be expanded to include a module that would use Apache Mahout. In the same way, for the persistence layer, an already existing Service Provider Interface could be used to provide other backends such as HDFS, Cassandra, Solr or any other.

Initial Goals

Our initial goals are to bring Unomi into the ASF, transition internal engineering processes into the open, and foster a collaborative development model according to the "Apache Way." Jahia plans to develop new functionality in an open, community-driven way. To get there, the existing internal build, test and release processes will be refactored to support open development.

Current Status

Currently, the project code base is licensed under the Apache License and is available on Github (https://github.com/Jahia/unomi).

The Jahia Marketing Factory, as soon as it is released, plans on building a robust end user community of paying and non-paying customers, and hopefully this will trickle down to the community on the Apache Unomi project once it is established.

Documentation is currently being worked on, but as it is the implementation of a public standard, most of the conceptual and API documentation will be covered by the specification. Any specifics of the implementation will of course be documented in the incubator project

Meritocracy

Our proposed list of initial committers include the current Marketing Factory R&D team, Jahia Engineers. This group will form a base for the broader community we will invite to collaborate on the codebase. We intend to radically expand the initial developer and user community by running the project in accordance with the "Apache Way". Users and new contributors will be treated with respect and welcomed. By participating in the community and providing quality patches/support that move the project forward, they will earn merit. They also will be encouraged to provide non-code contributions (documentation, events, community management, etc.) and will gain merit for doing so. Those with a proven support and quality track record will be encouraged to become committers.

Community

If Unomi is accepted for incubation, the primary initial goal will be transitioning the core community towards embracing the Apache Way of project governance. We would solicit major existing contributors to become committers on the project from the start.

Core Developers

While a few core developers are skilled in working in openly governed Apache communities. Most of the core developers are currently NOT affiliated with the ASF and would require new ICLAs before committing to the project.

Alignment

The following existing ASF projects can be considered when reviewing Unomi proposal:

- Apache Karaf : Apache Karaf is a small OSGi based runtime which provides a lightweight container onto which various components and applications can be deployed. Unomi is actually built on top of Karaf, taking full benefit of its modularity and powerful features as an OSGi runtime. More subprojects could of the Karaf project could be integrated such as Karaf Cave, Cellar or more interestingly Decanter which could offer real-time insight into the Unomi runtime.
- Apache Mahout : The Apache Mahout™ project's goal is to build an environment for quickly creating scalable performant machine learning
 applications. Unomi could integrate with Mahout to use it to perform recommendations or other types of machine learning based on what events
 are being fed to it.
- Apache Hadoop is a distributed storage and processing framework for very large datasets focusing primarily on batch processing for analytic purposes. Unomi could serve as a data set for Hadoop, and batch processing could be done on top of this data.
- Apache CXF : Apache CXF is an open source services framework. CXF helps you build and develop services using frontend programming APIs like JAX-WS and JAX-RS. These services can speak a variety of protocols such as SOAP, XML/HTTP, RESTful HTTP, or CORBA and work over a variety of transports such as HTTP, JMS or JBI. Unomi actually uses CXF to expose it's services as pure HTTP REST+JSON services. But this could be expanded further than what the specs supports to support other bindings supported by CXF out of the box.
- Incubator Apache Pistachio (http://wiki.apache.org/incubator/PistachioProposal): Pistachio is a fault-tolerant low latency distributed storage system which enables simple embedding the computation to the storage layer to achieve best data locality. It evolves from Yahoo's global user profile storage system. Unomi could possibly also use Pistachio as a backend system (if compatible). This might provide an interesting alternative to ElasticSearch for some users.
- Apache Geode : Geode is a data management platform that provides real-time, consistent access to data-intensive applications throughout widely
 distributed cloud architectures. Geode pools memory (along with CPU, network and optionally local disk) across multiple processes to manage
 application objects and behavior. It uses dynamic replication and data partitioning techniques for high availability, improved performance,
 scalability, and fault tolerance. Geode is both a distributed data container and an in-memory data management system providing reliable
 asynchronous event notifications and guaranteed message delivery. Unomi could leverage the power of Geode to help it with the processing of
 the incoming data from user events.
- Apache Ignite : Apache Ignite In-Memory Data Fabric is a high-performance, integrated and distributed in-memory platform for computing and transacting on large-scale data sets in real-time, orders of magnitude faster than possible with traditional disk-based or flash technologies. Unomi could leverage this project to perform any input processing.

Known Risks

Development has been sponsored mostly by a single company (or its predecessors) thus far and coordinated mainly by the core Jahia Unomi team.

For the project to fully transition to the Apache Way governance model, development must shift towards the meritocracy-centric model of growing a community of contributors balanced with the needs for extreme stability and core implementation coherency.

The tools and development practices in place for the Jahia Marketing Factory product are compatible with the ASF infrastructure and thus we do not anticipate any on-boarding pains. Migration from the current GitHub repository is also expected to be straightforward.

Orphaned products

Jahia is fully committed to Jahia Marketing Factory and the product will continue to be based on the Unomi project. Moreover, Jahia has a vested interest in making Unomi succeed by driving its close integration with sister ASF projects. We expect this to further reduces the risk of orphaning the product. Unomi is the core server at the heart of the Jahia Marketing Factory product, and at the time of this proposal both the commercial offering server and Unomi source code are 100% the same. It is expected over time that Jahia will develop additional modules on top of Unomi that may not be open sourced but other companies may of course do the same thing.

Inexperience with Open Source

Jahia has been developing and using open source software since its inception in 2001. Jahias CTO and Chief Software Architects are Apache Committers and other employees have contributed to various ASF projects over the years. Additionally, several ASF veterans agreed to mentor the project and are listed in this proposal. The project will rely on their guidance and collective wisdom to quickly transition the entire team of initial committers towards practicing the Apache Way.

Homogeneous Developers

While most of the initial committers are employed by Jahia, we have already seen a healthy level of interest from our existing customers and partners. We intend to convert that interest directly into participation and will be investing in activities to recruit additional committers from other companies.

Reliance on Salaried Developers

Most of the contributors are paid to work in web marketing/web content management space. While they might wander from their current employers, they are unlikely to venture far from their core expertises and thus will continue to be engaged with the project regardless of their current employers.

Relationships with Other Apache Products

As mentioned in the Alignment section, Unomi may consider various degrees of integration and code exchange with Apache Karaf, Apache Mahout, Apache Pistachio (incubating), and many others. We look forward to collaborating with these communities as well as other communities under the Apache umbrella.

An Excessive Fascination with the Apache Brand

While we intend to leverage the Apache 'branding' when talking to other projects as testament of our project's 'neutrality', we have no plans for making use of Apache brand in press releases nor posting billboards advertising acceptance of Unomi into Apache Incubator.

Interaction with the OASIS standard

In order to properly implement the standard and possibly use the feedback from the implementation as a way to improve the specifications, currently the cochair of the technical committee is also the proposer on the Apache proposal. This relationship makes it easy for communication between the two works, and will hopefully grow as the community grows around both the standard and the implementation. This is quite similar to the situation with the Apache Chemistry project, that is an open source implementation of the CMIS specification that is also an OASIS standard.

Documentation

See documentation for the current state of the project documentation available as part of the GitHub repository at https://github.com/Jahia/unomi

Initial Source

Jahia is releasing the source code for Unomi under the ASF license at https://github.com/Jahia/unomi . We encourage ASF community members interested in this proposal to download the source code, review and try out the software.

Source and Intellectual Property Submission Plan

As soon as Unomi is approved to join Apache Incubator, the source code will be transitioned via the Software Grant Agreement onto ASF infrastructure and in turn made available under the Apache License, version 2.0. We know of no legal encumberments that would inhibit the transfer of source code to the ASF.

External Dependencies

Embedded dependencies (relocated):

None

Runtime dependencies:

- apache-karaf
- pax-web
- apache-cxf
- slf4j
- org.apache.httpcomponents:httpclient-osgi
- org.apache.httpcomponents:httpcore-osgi
- org.apache.servicemix.bundles.jakarta-regexp
- pax-cdi
- javax.servlet-api
- jackson
- org.osgi.core

- commons-lang3
- commons-collections
- org.mvel:mvel2
- commons-beanutils
- commons-ognl
- commons-email
- commons-logging
- commons-codec
- commons-io
- geronimo-jcdi
- geronimo-atinject •
- geronimo-annotations string-template (BSD)
- javax.json
- ٠ net.sf.uadetector
- net.sf.quality-check
- com.maxmind.geoip2:geoip2 (Commons Commons Attribution-ShareAlike 3.0 Unported, http://dev.maxmind.com/geoip/geoip2/geolite2/#License)
- · elasticsearch
- spatial4j (optional)
- org.ow2.asm:asm
- groovy (optional)
- com.vividsolutions:jts (lgpl, optional)
- net.java.dev.jna (optional)
 org.fusesource:sigar (optional)

Module or optional dependencies:

None

Build only dependencies:

None

Test only dependencies:

- org.ops4j.pax.exam:pax-exam-container-karaf
- org.ops4j.pax.exam:pax-exam-junit4
- org.ops4j.pax.exam:pax-exam
- org.ops4j.pax.url:pax-url-aether
- javax.inject
- junit
- com.carrotsearch:junit-benchmarks
- com.h2database:h2
- gatling
- scala library

Cryptography N/A

Required Resources

Mailing lists

- private@unomi.incubator.apache.org (moderated subscriptions)
- commits@unomi.incubator.apache.org ٠
- . dev@unomi.incubator.apache.org
- issues@unomi.incubator.apache.org

Git Repository

https://git-wip-us.apache.org/repos/asf/incubator-unomi.git

Issue Tracking

• JIRA Project Unomi (UNOMI)

Other Resources

Means of setting up regular builds for Unomi on builds.apache.org

Initial Committers

- Serge Huber
- Thomas Draier
- Sergiy Shyrkov

- Kevan Jahanshahi
- Cédric Mailleux
- Régis Mora
- Quentin Lamerand
- Chris Laprun
- David Griffon
 Jean-Baptiste Onofré
- Raul Kripalani

Affiliations

- Jahia Solutions: everyone else on this proposal
 NASA JPL: Chris Mattmann
- Talend : Jean-Baptiste Onofré
- Pivotal : Roman Shaposhnik • Independent: Raul Kripalani

Sponsors

Champion

• Jean-Baptiste Onofré - Apache Member, Talend

Mentors

The initial mentors are listed below:

- Bertrand Delacretaz Apache Member, Adobe
- Roman Shaposhnik Apache Member, Pivotal
 Chris Mattmann Apache Member, NASA JPL

Sponsoring Entity

We would like to propose Apache incubator to sponsor this project.