

NiFiProposal

NiFi Proposal

Abstract

NiFi is a dataflow system based on the concepts of flow-based programming.

Proposal

NiFi supports powerful and scalable directed graphs of data routing, transformation, and system mediation logic. Some of the high-level capabilities and objectives of NiFi include:

- Web-based user interface for seamless experience between design, control, feedback, and monitoring of data flows
- Highly configurable along several dimensions of quality of service such as loss tolerant versus guaranteed delivery, low latency versus high throughput, and priority based queuing
- Fine-grained data provenance for all data received, forked, joined, cloned, modified, sent, and ultimately dropped as data reaches its configured end-state
- Component-based extension model along well defined interfaces enabling rapid development and effective testing

Background

Reliable and effective dataflow between systems can be difficult whether you're running scripts on a laptop or have a massive distributed computing system operated by numerous teams and organizations. As the volume and rate of data grows and as the number of systems, protocols, and formats increase and evolve so too does the complexity and need for greater insight and agility. These are the dataflow challenges that NiFi was built to tackle.

NiFi is designed in a manner consistent with the core concepts described in flow-based programming as originally documented by J. Paul Morrison in the 1970s. <http://www.jpaulmorrison.com/fbp/> This model lends itself well to visual diagramming, concurrency, componentization, testing, and reuse. In addition to staying close to the fundamentals of flow-based programming, NiFi provides integration system specific features such as: guaranteed delivery; back pressure; ability to gracefully handle backlogs and data surges; and an operator interface that enables on-the-fly data flow generation, modification, and observation.

Rationale

NiFi provides a reliable, scalable, manageable and accountable platform for developers and technical staff to create and evolve powerful data flows. Such a system is useful in many contexts including large-scale enterprise integration, interaction with cloud services and frameworks, business to business, intra-departmental, and inter-departmental flows. NiFi fits well within the Apache Software Foundation (ASF) family as it depends on numerous ASF projects and integrates with several others. We also anticipate developing extensions for several other ASF projects such as Cassandra, Kafka, and Storm in the near future.

Initial Goals

- Ensure all dependencies are compliant with Apache License version 2.0 and all that all code and documentation artifacts have the correct Apache licensing markings and notice.
- Establish a formal release process and schedule, allowing for dependable release cycles in a manner consistent with the Apache development process.
- Establish a process which allows different release cycles for the core framework and extensions.
- Grow the community to establish diversity of background and expertise.

Current Status

Meritocracy

An integration platform is only as good as its ability to integrate systems in a reliable, timely, and repeatable manner. The same can be said of its ability to attract talent and a variety of perspectives as integration systems by their nature are always evolving. We will actively seek help and encourage promotion of influence in the project through meritocracy.

Community

Over the past several years, NiFi has developed a strong community of both developers and operators within the U.S. government. In open sourcing NiFi we plan to grow the community to a broader base of industries and will work to align the interaction of our existing community.

Core Developers

The initial core developers are employed by the National Security Agency and defense contractors. We will work to grow the community among a more diverse set of developers and industries.

Alignment

From its inception, NiFi was developed with an open source philosophy in mind and with the hopes of eventually being truly open sourced. The Apache way is consistent with the approach we have taken to date. The ASF clearly provides a mature and effective environment for successful development as is evident across the spectrum of well-known projects. Further, NiFi depends on numerous ASF libraries and projects including; ActiveMQ, Ant, Commons, Lucene, Hadoop, HttpClient, Jakarta and Maven. We also anticipate extensions and dependencies with several more ASF projects, including Accumulo, Avro, Cassandra, HBase, JClouds, Storm, Kafka, Thrift, Tika, and others. This existing alignment with Apache and the desired community makes the Apache Incubator a good fit for NiFi.

Known Risks

Orphaned Products

Risk of orphaning is limited though it is important to grow the community. The project user and developer base is substantial, growing, and there is already extensive operational use of NiFi.

Inexperience with Open Source

The initial committers to NiFi have limited experience with true open source software development. However, despite the project origins being from closed source development we have modeled our behavior and community development on The Apache Way to the greatest extent possible. This environment includes widely accessible source code repositories, published artifacts, ticket tracking, and extensive documentation. We also encourage contributions and frequent debate and hold regular, collaborative discussions through e-mail, chat rooms, and in-person meet-ups. We are committed to the ideals of open source software and will eagerly seek out mentors and sponsors who can help us quickly come up to speed.

Homogenous Developers

The initial committers of NiFi come from a limited set of entities though we are committed to recruiting and developing additional committers from a broad spectrum of industries and backgrounds.

Reliance on Salaried Developers

We expect NiFi development to continue on salaried time and through volunteer time. The initial committers are paid by their employers to contribute to this project. We are committed to developing and recruiting participation from developers both salaried and non-salaried.

Relationship with other Apache Projects

As described in the alignment section, NiFi is already heavily dependent on other ASF projects and we anticipate further dependence and integration with new and emerging projects in the Apache family.

An Excessive Fascination with the Apache Brand

We respect the laudable Apache brand and that is certainly a factor in the decision to propose NiFi for the Apache Incubator. The ASF is a natural home for NiFi given our existing dependency and alignment with ASF projects. We intend to provide a great deal of energy and capability to the ASF through this project. We will be sensitive to and respectful of any overuse of the Apache brand and ensure our focus remains on how we benefit the Apache community.

Documentation

At this time there is no NiFi documentation on the web. However, we have extensive documentation included within the application that details usage of the many functions. Using incubator INFRA we will be rapidly expanding the available documentation to cover things like installation, developer guide, frequently asked questions, best practices, and more.

Initial Source

NiFi has been in active development since late 2006 with contributions from dozens of developers and feedback from hundreds of users and developers. The core codebase is written in Java and includes detailed Javadocs and feature documentation.

Source and Intellectual Property Submission

Previously referred to as Niagarafiles, the NiFi code and documentation materials will be submitted by the National Security Agency. NiFi has been developed by a mix of government employees and private companies under government contract. Material developed by the government employees is in the public domain and no U.S. copyright exists in works of the federal government. For the contractor developed material in the initial submission, the U.S. Government has sufficient authority to open source per DFARS 252.227-7014. NSA has submitted the Software Grant Agreement and Corporate Contributor License Agreement to the Apache Software Foundation.

External Dependencies

We have at least one dependency on an LGPL library which we will promptly address. Otherwise, we believe all current dependencies are compatible with the ASF guidelines. Our dependency licenses come from the following license styles: Apache v 2.0, BSD, Public Domain, Eclipse Public v1, MIT, CDDL v1.

Cryptography

Consistent with <http://www.apache.org/licenses/exports/> we believe NiFi is classified as ECCN 5D002. NiFi doesn't implement any cryptographic algorithms but is designed to use algorithms provided by Oracle Java Cryptographic Extensions, Bouncy Castle, and JCraft, Inc. These cryptographic algorithm providers are used to support SSL, SSH/SFTP, and the encryption and decryption of sensitive properties. In the event that it becomes necessary we will engage with appropriate Apache members to ensure we file any necessary paperwork or clarified any cryptographic export license concerns.

Required Resources

Mailing Lists

- user@nifi.incubator.apache.org
- dev@nifi.incubator.apache.org
- private@nifi.incubator.apache.org
- commits@nifi.incubator.apache.org

Source Control

NiFi requests use of Git for source control ([git://git.apache.org/nifi.git](https://git.apache.org/nifi.git)). We request a writeable Git repo for NiFi with mirroring to be setup to Github through INFRA. We request sponsor Benson Margulies (bimargulies) to assist with creating the INFRA ticket for this.

Issue Tracking

JIRA NiFi (NIFI)

Initial Committers

- Brandon DeVries <[brandon.devries at gmail dot com](mailto:brandon.devries@gmail.com)>, CLA confirmed
- Jason Carey <[jcarey03 at gmail dot com](mailto:jcarey03@gmail.com)>, CLA submitted
- Matt Gilman <[matt.c.gilman at gmail dot com](mailto:matt.c.gilman@gmail.com)>, CLA confirmed
- Tony Kurc <[trkurc at gmail dot com](mailto:trkurc@gmail.com)>, CLA confirmed
- Mark Payne <[markap14 at hotmail dot com](mailto:markap14@hotmail.com)>, CLA confirmed
- Adam Taft <[adam at adamtaft dot com](mailto:adam@adamtaft.com)>, CLA submitted
- Joseph Witt <[joewitt at gmail dot com](mailto:joewitt@gmail.com)>, CLA confirmed

Affiliations

- Brandon DeVries (Requiest, Inc.)
- Jason Carey (Twitter)
- Matt Gilman (Raytheon)
- Tony Kurc (National Security Agency)
- Mark Payne (Sotera Defense Solutions, Inc.)
- Adam Taft (Requiest, Inc.)
- Joseph Witt (National Security Agency)

Sponsors

Champion

- Benson Margulies (Basis Technology) <[bimargulies at apache dot org](mailto:bimargulies@apache.org)>, IPMC Member

Nominated Mentors

- Drew Farris (Booz Allen Hamilton) <[drew at apache dot org](mailto:drew@apache.org)>, IPMC Member
- Brock Noland (Cloudera) <[brock at apache dot org](mailto:brock@apache.org)>, IPMC Member
- Billie Rinaldi (Hortonworks) <[billie at apache dot org](mailto:billie@apache.org)>, IPMC Member
- Josh Elser (Hortonworks) <[elserj at apache dot org](mailto:elserj@apache.org)>, Needs to request IPMC membership.
- Arvind Prabhakar (StreamSets) <[arvind at apache dot org](mailto:arvind@apache.org)>, IPMC Member
- Sergio Fernandez (Redlink, Salzburg Research) <[wikier at apache dot org](mailto:wikier@apache.org)>, IPMC Member
- Andrew Purtell (Salesforce) <[apurtell at apache dot org](mailto:apurtell@apache.org)>, IPMC Member

Sponsoring Entity

We request the Apache Incubator to sponsor this project.