

SliderProposal

Slider Proposal

Abstract

Slider is a collection of tools & technologies to package, deploy, and manage long running applications on Apache Hadoop YARN clusters.

Background

Slider is a framework to support deployment and management of arbitrary applications on YARN and leverage YARN's resource management capabilities without having to rewrite the applications. Slider is actively being worked on to expand the ecosystem of applications that can be easily deployed and managed on Apache Hadoop YARN clusters.

The core Slider technologies were initially developed at Hortonworks as part of the *Hoya* project – an effort to support the deployment of HBase and later Accumulo clusters in YARN. This work showed the value in supporting more applications on YARN, that the client should be an API ~~rather than just a command-line~~ and what key issues need to be addressed.

Slider is an evolution of the previous proposal, in that the proposal now includes agent-based deployment, makes packaging applications to be deployable and manageable a core area of work.

Rationale

Hadoop YARN offers the following key capabilities:

Availability (always-on) - YARN works with the application to ensure recovery or restart of running application components.

Flexibility (dynamic scaling) - YARN provides the application with the facilities to allow for scale-up or scale-down

Resource Management - YARN handles allocation of cluster resources --and hence the scheduling of work across a Hadoop cluster.

Today, developers need to design or re-engineer their application to operate in a YARN clusters using the YARN APIs and its application architecture.

Slider's objective is to make it easy for existing distributed application to be deployed on a YARN cluster without changes and with little or no custom code.

Proposal Details

Slider allows users to deploy distributed applications across a Hadoop cluster, leveraging the YARN Resource Manager to allocate and distribute components of an application across the cluster. Key characteristics of Slider:

- No need to change the application code [as long as the application follows developer guidelines]
- No need to develop a custom Application Master or other YARN code
- Slider leverages YARN facilities to manage:
 - Application recovery in cases of container failure
 - Resource allocation and flexing (adding/removing containers)

Initial Goals

1. Donate the Slider source code and documentation to the Apache Software Foundation
2. Set up and standardize the open governance of the Slider project
3. Build a user and developer community
4. Tie in better with Apache HBase, Apache Accumulo, and other projects – both ASF and external – that can be deployed in a YARN cluster without any code changes
5. Improve Slider capabilities to expand on list of apps that can be deployed on YARN using Slider

Longer Term Goals

There are some longer term possibilities that could improve Slider:

1. Implement a reusable management API for managing Slider applications by tools such as Apache Ambari
2. Provide a Java API to ease creation and manipulation of Slider-deployed clusters by other programs.
3. Address the service registration and discovery problem, to aid discovery and binding to YARN applications.
4. Explore load-driven cluster sizing.
5. Collaborate with other YARN applications, libraries, and frameworks to develop better libraries for YARN applications and their clients, monitoring and management, and configuration

Slider is driving YARN service support via YARN-896. We intend to evolve features and get practical experience using them before merging them into the Hadoop codebase.

Current Status

Slider is currently under active development and functions end to end following the Slider specifications.

Meritocracy

The core of Slider was originally driven by Steve Loughran, who has long-standing experience in Apache projects, and is being advanced with significant contributions from Ted Yu, Josh Elser, Billie Rinaldi, Sumit Mohanty, and Jon Maron with deep experience architecting and implementing key parts of key Apache projects including, HBase, Accumulo, Ambari, and other open source projects.

Community

We are happy to report that there are folks in Accumulo, HBase, and some users outside Hortonworks who are closely involved in the project already.

We hope to extend the user and developer base further in the future and build a solid open source community around Slider, growing the community and adding committers following the Apache meritocracy model.

Alignment

The project is completely aligned with Apache, from its build process up. It depends on Apache Hadoop, and it currently deploys HBase and Accumulo.

Slider and Apache Samza are driving the work of supporting long-lived services in YARN. While many of these relate to service longevity, there is also the challenge of having low-latency table lookups co-exist with CPU-and-IO intensive analytics workloads.

Relationship with Apache Twill

Twill is a library that one can use to write YARN applications. Slider aims to provide a general framework using which one can take existing applications (HBase & Accumulo to start with), and make them run well in a YARN cluster, without intruding at all into their internals.

The key differentiators are

- **Long lived static applications:** the application's containers are expected to be relatively stable, with their termination being an unexpected event to which Slider must react.
- **No application code-changes:** The only glue between the App and Slider is a Slider interface that the App needs to implement for it to be deployable/manageable by Slider.

Twill and Slider are therefore very different. The former is a convenience library for new YARN applications, the latter a YARN Framework to adapt existing applications to YARN.

While Slider can be written using Twill libraries (which is something we should pursue as part of long/medium-term collaboration between the two projects), the goals of the two projects are different - Twill will continue to make YARN application developers' lives easier, while Slider is a framework that can deploy distributed-applications easily in a YARN cluster, and perform basic management operations.

Capabilities such as dynamic patching of the application's configuration to run in the YARN cluster, failure detection, reacting to failures, storing application state to facilitate better application restart behavior, etc. are under the purview of Slider.

Management frameworks could use Slider as a tool to start/stop/shrink/expand an instance of an application.

Relationship with Apache Helix

Slider shares some common goals with Apache Helix. Helix is more sophisticated and is designed to work standalone. Slider is designed to work only in the context of a YARN cluster, and focuses on that YARN integration.

We have discussed Slider with the Helix team, and feel that the work we are doing in YARN integration, and driving YARN changes, will be of direct benefit to Helix. We plan to collaborate on features which can be shared across both projects.

Relationship with Apache Accumulo and Apache HBase

We offer Accumulo and HBase the flexible operation in a YARN cluster. As such, it should expand the uses of the applications, and their user base.

There may be some changes that the applications can make to help them live more easily in a YARN cluster, and to be managed by Slider. To date, changes have focused on supporting dynamic port allocations and reporting of the values.

It may be in future that we encounter situations where other changes to the applications can help them work even better in Slider-managed deployments. If these arise we would hope to work with the relevant teams to get the changes adopted - knowing up front that neither of these project teams would countenance any changes that interfered with classic static application deployments.

The initial Slider committer list includes committers for both Accumulo and HBase, who can maintain cross-project collaboration.

Known Risks

The biggest risk is getting the critical mass of use needed to build a broad development team. We don't expect to have or need many full-time developers, but active engagement from the HBase and Accumulo developers would significantly aid adoption and governance.

The other risk is YARN not having the complete feature set needed for long lived services: restarting, security token renewal, log-capture and other issues. We are working with the YARN developers to address these issues, issues shared with other long-lived services on YARN.

Orphaned Products

Steve, Sumit, Jon, and Billie will continue to work on Slider 100% of the time for the foreseeable future with others from Hortonworks and growing community contributing as well.

Inexperience with Open Source

All of the core developers have long-standing experience in open source, Two of them are Accumulo committers and two are HBase committers. Steve Loughran has been a committer on various ASF projects since 2001 (Ant, Axis), a mentor to Incubated projects, a Hadoop committer since 2008, and full-time developer on HP's open -source [SmartFrog](#) project from 2005-2012. Sumit and Billie are committers on Ambari. Jon Maron has worked extensively with Ambari APIs and has contributed to the [OpenStack](#) Savanna (now Sahara) project.

Homogeneous Developers

The current core developers are all from Hortonworks. However, we hope to establish a developer community that includes users of Slider and developers on the applications themselves - HBase, Accumulo, etc.

Reliance on Salaried Developers

Currently, the developers are paid to do work on Slider. A key goal for the incubation process will be to broaden the developer base.

Relationships with Other Apache Products

This is covered in the Alignment section.

An Excessive Fascination with the Apache Brand

While we respect the reputation of the Apache brand and have no doubts that it will attract contributors and users, our interest is primarily to give Slider a solid home as an open source project with a broad developer base -and to encourage adoption by the related ASF projects.

Documentation

All Slider documentation is currently in [markdown-formatted text files in the source repository](#); they will be delivered as part of the initial source donation.

Initial Source

The initial source ~~all ASF licensed~~ can be found at <https://github.com/hortonworks/slider>

Slider is written in Java. Its source tree is entirely self-contained and relies on Apache Maven as its build system. Alongside the application, it contains unit, localhost, and functional tests. The latter for use with remote clusters.

Source and IP Submission Plan

1. All source will be moved to Apache Infrastructure
2. All outstanding issues in our in-house JIRA infrastructure will be replicated into the Apache JIRA system.
3. We have pre-emptively acquired a currently-unused twitter handle @apacheslider which would be passed to the PMC.

External Dependencies

Slider has no external dependencies except for some Java libraries that are considered ASF-compatible (JUnit, SLF4J, jcommander, groovy), BSD-licensed Jinja, and Apache artifacts : Hadoop, Log4J and the transient dependencies of all these artifacts.

Required Resources

Mailing Lists:

1. slider-dev
2. slider-commits
3. slider-private

Infrastructure:

1. Git repository
2. JIRA Slider (Slider)
3. Gerrit for reviewing patches

The existing code includes local host integration tests, so we would like a Jenkins instance to run them whenever a new patch is submitted.

Initial Committers

1. Steve Loughran (stevel at a.o)
2. Jon Maron
3. Sumit Mohanty
4. Billie Rinaldi (billie at a.o)
5. Ted Yu (tedyu at a.o)
6. Josh Elser (elserj at a.o)

Sponsors

Champion: Vinod Kumar Vavilapalli

Nominated Mentors:

1. Jean-Baptiste Onofré
2. Mahadev Konar
3. Arun Murthy
4. Devaraj Das (ddas at a.o)

Sponsoring Entity

Incubator PMC