

YARN sysfs interface

Provide cluster information to application

Design Document

Version 1.0
August 30, 2018

1 Overview	3
1.1 Requirement Description	3
1.2 Solution Summary	3
2 External	4
2.1 Target Audience	4
2.2 Usage	4
3 Internal	4
3.1 Design	4
4 Known Issues	5
4.1 Unstable application to increase network traffic	5
4.2 Cluster information is not available in real time	6

1 Overview

This document describe the introduction of a new YARN API called YARN sysfs. The pseudo file system abstraction provides environment information to application. This information can be consumed by application developers to write distribute application without depend on heavy client libraries to discover the running environment.

1.1 Requirement Description

Container framework usually requires external system administrator to write a configuration file to describe the running environment, and the cluster hostname and port information are presented to application through various command line parameters. Some of the well known software like Docker swarm, or Kubernetes pod info are exported via environment variables. There are usually a set of boilerplate code to obtain cluster information then translate to application parameterization. However, this is implemented, this information is required to be delivered to container for application to consume the essential information.

YARN provides ability to forward environment variables to containers, but late binding environment change can not be exposed properly. For example, Docker container can not change launching environment variables after the container is already running. Alternative method of exporting cluster information needs to be explored.

1.2 Solution Summary

In Linux, sysfs is used to provide abstraction between kernel information and user land utilities. YARN sysfs is inspired by sysfs to provide cluster information via file abstraction to expose the information more securely and flexible than environment variables offers. We choose to use pseudo file abstraction for exporting cluster information in YARN.

The motivation of introducing YARN sysfs API is to enable non-Hadoop native framework, and container friendly applications to have ability to adapt to Hadoop resource management framework with minimum amount of code changes. This helps third party developers to write distributed clustering software with ease.

2 External

2.1 Target Audience

This API is targeted for third party software developer to write clustering software without depending on language specific client libraries.

2.2 Usage

Define a environment variable (YARN_CONTAINER_RUNTIME_YARN_SYSFS) during YARN application launch. Docker container that has this flag enabled, will have an extra path /hadoop/yarn/sysfs/service.json available in the container.

3 Internal

3.1 Design

The entire flow of the code execution is illustrated in Figure 1.

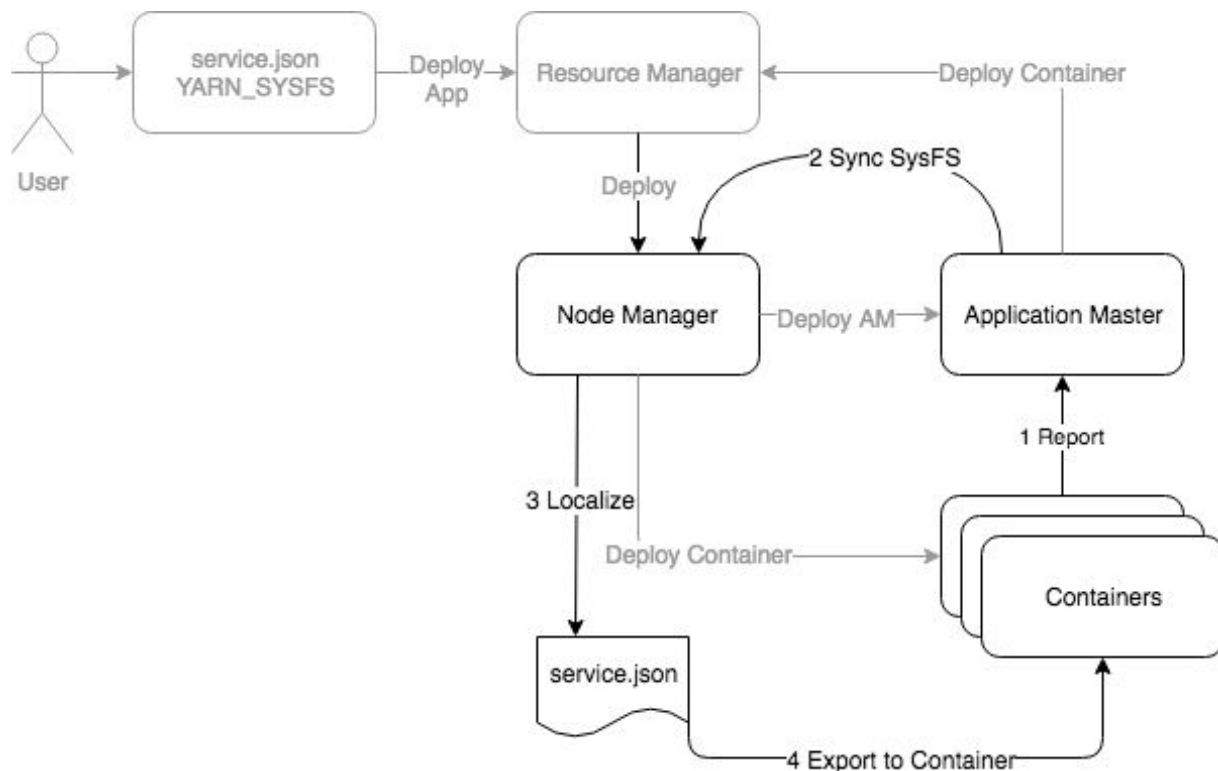


Figure 1

1. Application Master tracks the creation of containers, and obtain hostname, IP address and useful information.
2. When service reaches stable state, application master will issue REST API call with cluster information to node manager.
3. Node manager localize the content of cluster information as a file in node manager private directory.
4. Node manager invokes container-executor to copy the cluster information from node manager private directory to container working directory.

4 Known Issues

4.1 Unstable application increases network traffic

If an application is rapidly changing, and the container live time is only a fraction of seconds. It is possible that rapid creation and deletion of containers increase network load for synchronizing YARN sysfs information to all running containers. The recommendation is to not use YARN sysfs feature for containers that runs less than a minute.

4.2 Cluster information is not available in real time

When cluster is constructing or changing, the information is not immediately available to all containers. This is to conserve network bandwidth required for exporting cluster configuration. If there are delays during container construction, it could be minutes before service.json shows up in /hadoop/yarn/sysfs directory. It is recommend to use file creation notification event to monitor changes in /hadoop/yarn/sysfs directory to be notified of cluster changes.