

Docker to SquashFs Conversion Tool

Problem Statement

YARN-9564 defines a docker-to-squashfs image conversion tool that relies on python2, multiple libraries, squashfs-tools and root access in order to convert Docker images to squashfs images for use with the runc container runtime in YARN.

YARN-9943 was created to investigate alternatives, as the response to merging YARN-9564 has not been very positive. This proposal outlines the design for a CLI conversion tool in 100% pure Java that will work out of the box.

Design

We propose a new `hadoop-runc` Maven module (under `hadoop-tools`) which will provide a CLI tool to perform the Docker image conversion. This tool will connect to a Docker repository, download image layers and metadata, convert the `tar.gz` layers to SquashFs, and upload the results to HDFS in the format defined by **YARN-10943**.

Example Usage

The CLI tool can be invoked with the standard `hadoop jar` command as follows:

```
hadoop jar \  
  $HADOOP_HOME/share/hadoop/tools/lib/hadoop-runc-*.jar \  
  org.apache.hadoop.runc.tools.ImportDockerImage \  
  repository.docker.com/library/debian:latest \  
  os/debian-latest
```

This will contact the Docker repository at `repository.docker.com`, download the Docker image `library/debian:latest`, convert the image layers to squashfs, and upload to the default repository location within HDFS and tag the image as `os/debian-latest` within the runc repository.