

Controller Flow: Add Host(s)

Overview:

This document describes a high level flow for a add hosts invocation in the Savanna controller. It is assumed that a mechanism will exist to plug cloud and Hadoop providers into the Savanna controller. Described is the interaction of the controller with the cloud and Hadoop providers with the intent of identifying the interaction points and possible interfaces. The focus here is primarily the Hadoop provider functionality and the services that it requires. The add hosts invocation results in a number of VM's being provisioned by the cloud provider which are then added to an existing Hadoop cluster by the Hadoop provider.

- controller receives add host(s) request
 - received arguments
 - cloud_spec
 - specifies cloud provider (openstack ...)
 - provider specific details used to provision vm(s)
 - vm_image(s), flavor(s), node_count, etc.
 - cluster_spec
 - specifies Hadoop provider
 - provider specific cluster details
 - cluster, host, host_components, etc.
 - controller calls provision_nodes(cloud_spec) on cloud provider
 - cloud provider provisions vm(s) based on cloud_spec
 - blocks until network info (ip addr) are available for node(s)
 - returns cloud_ctx
 - general cluster information
 - servers[]
 - server specific information such as hostname, public/private ip addr, vm_image, flavor, etc.
 - functionality related to interacting with provisioned cluster/vm's
 - install_package(package)
 - yum -y, zypper --non-interactive, ...
 - execute(command)
 - execute a command on a server
 - execute_interactive(command, prompts)
 - execute a command that requires interactive responses
 - open_file(file, mode)
 - returns a file pointer to a remote file
- Controller calls add_hosts(cloud_ctx, cluster_spec) on Hadoop provider
 - provider adds nodes to Hadoop cluster and starts all specified services
 - HDP specific details
 - install/confirm various epel packages on node(s)
 - cloud_ctx.server[0].install('epel-release')
 - ...
 - install/setup/start Ambari agent
 - install ambari-agent package
 - cloud_ctx.server[0].install('ambari-agent')
 - update Ambari agent configuration

- set Ambari server address in ambari-agent.ini configuration
 - lookup master host information for cluster
 - obtain via savanna_ctx
 - provided to Hadoop provider during init
 - server = savanna_ctx.lookup(cluster_name)
 - f =
cloud_ctx.servers[0].open_file('ambari-agent.ini',
'r+')
 - f.write(...)
 - start ambari-agent
 - cloud_ctx.server[0].execute('ambari-agent start')
- update various Hadoop configurations on hosts
- interact with Ambari via REST api
 - invoked via savanna_ctx
 - savanna_ctx.invoke_rest(request)
 - install Hadoop services
 - start Hadoop services