KNOX-88: Support HDFS HA

Design details and discussion for KNOX-88

Definition

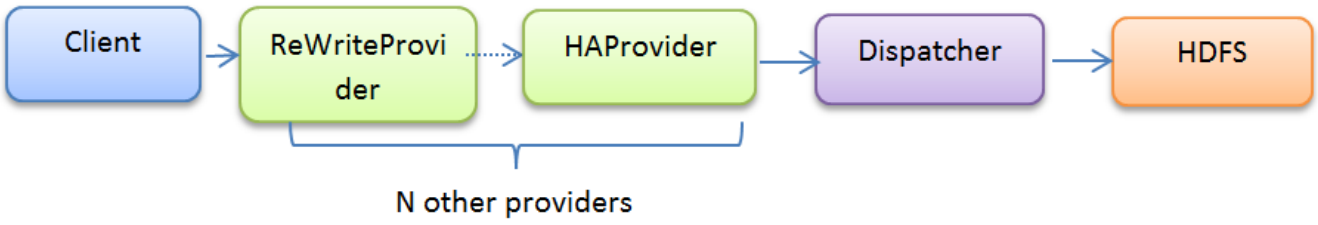
Knox HA is a set of routines for transparent work with Hadoop service that stands in HA mode.

Purpose of Knox HA service

1. Automatic failover. (Example: switch request from not responding name-node to active name-node.)
2. Pluggable support of failover strategies.
3. Daemon-service for regular ping of Hadoop service state (Performance optimization to keeping actual state of service).

Architecture

New provider will be added (descendant ProviderDeploymentContributorBase class) with a set of filters. See Pic.#1 for common architecture.



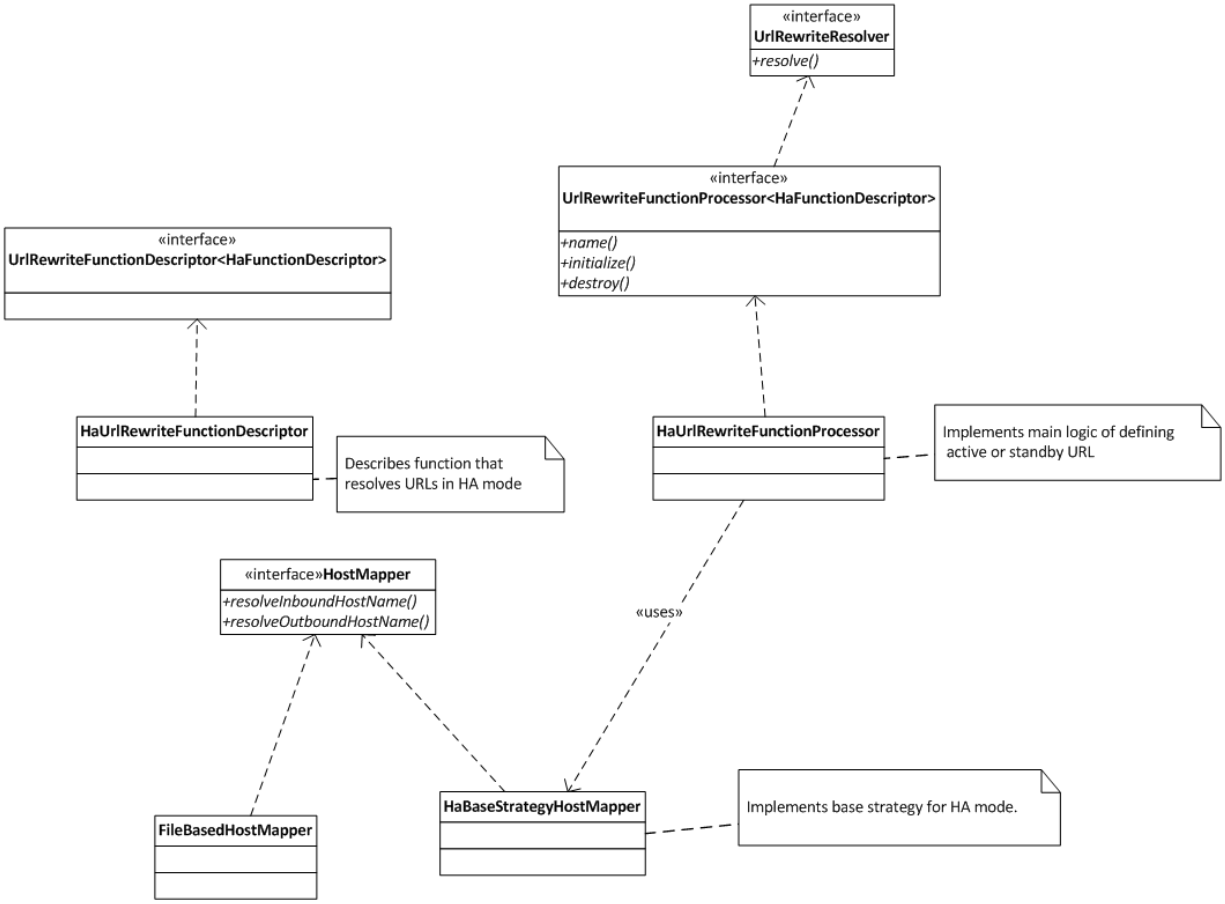
Pic. #1 – Providers architecture

Definition.

Alias – set of Hadoop name-nodes configured for High Availability mode.

Definition.

High Availability Strategy – plan of defining active name-node and switching between active and stand-by name-nodes. Strategy may contain such parameters as retryCount and timeoutInterval. See Pic.#2 for class diagram for HA mode.



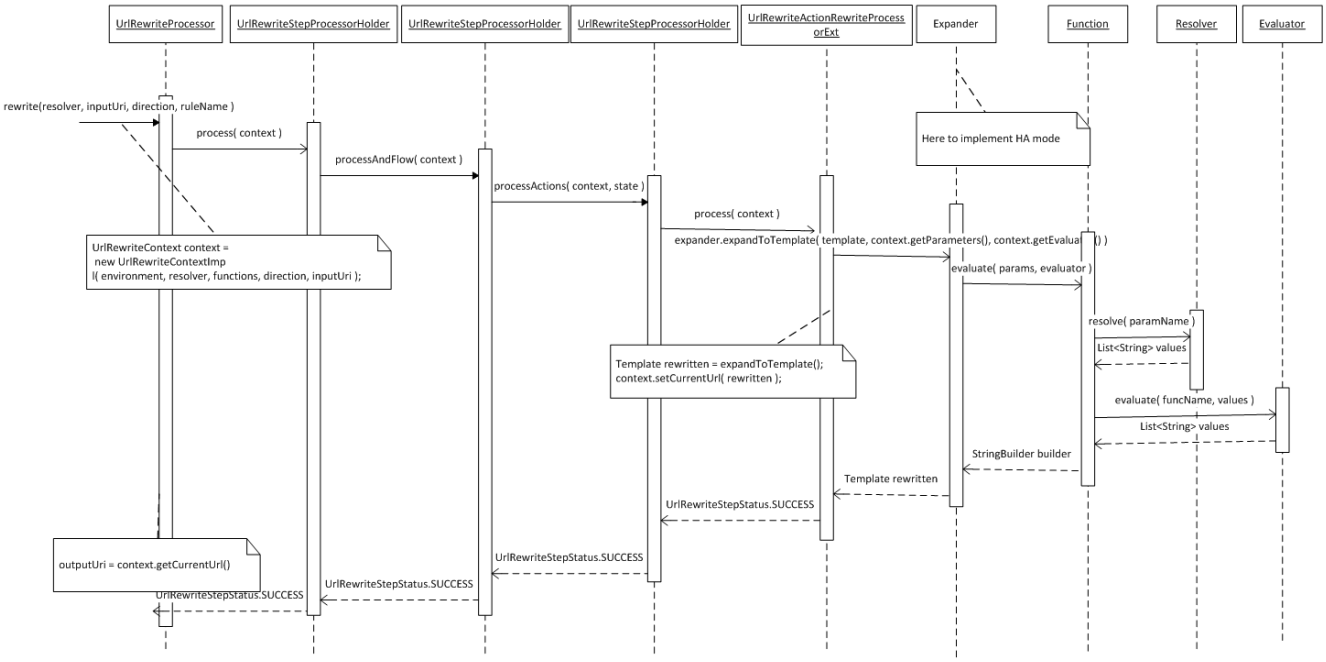
Pic.#2 Class diagram for HA mode.

See Table #1 for class description.

Table #1. – HA mode new classes description.

#	Class name	Description
1	HaUrlRewriteFunctionDescriptor	Describes function that resolves URLs in HA mode
2	HaUrlRewriteFunctionProcessor	Implements main logic of defining active or standby URL
3	HaBaseStrategyHostMapper	Implements base strategy for HA mode. Contains parameters: retryCount, timeoutInterval.

See Pic.#2 for UML sequence diagram for UrlRewriteProcessor.



Pic #3 – UML sequence diagram for UrlRewriteProcessor.

Provider configuration example

Please look at the WebHDFS HA section http://knox.apache.org/books/knox-0-5-0/knox-0-5-0.html#WebHDFS