

WebHCat Configuration

Table of contents

1 Variables.....	2
------------------	---

The configuration for WebHCat (Templeton) merges the normal Hadoop configuration with the WebHCat-specific variables. Because WebHCat is designed to connect services that are not normally connected, the configuration is more complex than might be desirable.

The WebHCat-specific configuration is split into two layers:

1. **webhcat-default.xml** - All the configuration variables that WebHCat needs. This file sets the defaults that ship with WebHCat and should only be changed by WebHCat developers. Do not copy this file and/or change it to maintain local installation settings. Because webhcat-default.xml is present in the WebHCat war file, editing a local copy of it will not change the configuration.
2. **webhcat-site.xml** - The (possibly empty) configuration file in which the system administrator can set variables for their Hadoop cluster. Create this file and maintain entries in it for configuration variables that require you to override default values based on your local installation.

The configuration files are loaded in this order with later files overriding earlier ones.

Note: the WebHCat server will require restart after any change to the configuration.

To find the configuration files, WebHCat first attempts to load a file from the CLASSPATH and then looks in the directory specified in the TEMPLETON_HOME environment variable.

Configuration files may access the special environment variable `env` for all environment variables. For example, the Pig executable could be specified using:

```
${env.PIG_HOME}/bin/pig
```

Configuration variables that use a filesystem path try to have reasonable defaults. However, it's always safe to specify the full and complete path if there is any uncertainty.

Note: The location of the log files created by WebHCat and some other properties of the logging system are set in the webhcat-log4j.properties file.

1 Variables

Name	Default	Description
templeton.port	50111	The HTTP port for the main server.
templeton.hadoop.config.dir	<code>\$(env.HADOOP_CONFIG_DIR)</code>	The path to the Hadoop configuration.
templeton.jar	<code>\${env.TEMPLETON_HOME}/share/webhcat/ivr/webhcat-0.11.0.jar</code>	The path to the WebHCat jar file.

Name	Default	Description
templeton.libjars	<code>\${env.TEMPLETON_HOME}/share/webhcat/ivr/lib/zookeeper-3.4.5.jar</code>	Jars to add to the classpath.
templeton.override.jars	<code>hdfs:///user/templeton/ugi.jar</code>	Jars to add to the HADOOP_CLASSPATH for all Map Reduce jobs. These jars must exist on HDFS.
templeton.override.enabled	<code>false</code>	Enable the override path in templeton.override.jars
templeton.streaming.jar	<code>hdfs:///user/templeton/hadoop-streaming.jar</code>	The hdfs path to the Hadoop streaming jar file.
templeton.hadoop	<code>\${env.HADOOP_PREFIX}/bin/hadoop</code>	The path to the Hadoop executable.
templeton.pig.archive	<code>hdfs:///user/templeton/pig-0.11.1.tar.gz</code>	The path to the Pig archive.
templeton.pig.path	<code>pig-0.11.1.tar.gz/pig-0.11.1/bin/pig</code>	The path to the Pig executable.
templeton.hcat	<code>\${env.HCAT_PREFIX}/bin/hcat</code>	The path to the HCatalog executable.
templeton.hive.archive	<code>hdfs:///user/templeton/hive-0.11.0.tar.gz</code>	The path to the Hive archive.
templeton.hive.path	<code>hive-0.11.0.tar.gz/hive-0.11.0/bin/hive</code>	The path to the Hive executable.
templeton.hive.properties	<code>hive.metastore.local=false hive.metastore.uris=thri localhost:9933, hive.metastore.sasl.enab</code>	Properties to set when running Hive. To use it in a cluster with Kerberos security enabled set <code>hive.metastore.sasl.enabled=false</code> and add <code>hive.metastore.execute.setugi=true</code> . Using localhost in metastore uri does not work with Kerberos security.
templeton.exec.encoding	<code>UTF-8</code>	The encoding of the stdout and stderr data.

Name	Default	Description
templeton.exec.timeout	10000	How long in milliseconds a program is allowed to run on the WebHCat box.
templeton.exec.max-procs	16	The maximum number of processes allowed to run at once.
templeton.exec.max-output-bytes	1048576	The maximum number of bytes from stdout or stderr stored in ram.
templeton.controller.mr.child.opts	-server -Xmx256m -Djava.net.preferIPv4Stack	Java options to be passed to WebHCat controller map task.
templeton.exec.envs	HADOOP_PREFIX, HADOOP_HOME	The environment variables passed through to exec.
templeton.zookeeper.hosts	127.0.0.1:2181	ZooKeeper servers, as comma separated host:port pairs.
templeton.zookeeper.session-timeout	30000	ZooKeeper session timeout in milliseconds.
templeton.callback.retry.interval	10000	How long to wait between callback retry attempts in milliseconds.
templeton.callback.retry.attempts	5	How many times to retry the callback.
templeton.storage.class	org.apache.hcatalog.templeton	The class to use as storage.
templeton.storage.root	/templeton-hadoop	The path to the directory to use for storage.
templeton.hdfs.cleanup.interval	43200000	The maximum delay between a thread's cleanup checks.
templeton.hdfs.cleanup.maxage	604800000	The maximum age of a WebHCat job.
templeton.zookeeper.cleanup.interval	43200000	The maximum delay between a thread's cleanup checks.
templeton.zookeeper.cleanup.maxage	604800000	The maximum age of a WebHCat job.
templeton.kerberos.secret	A random value	The secret used to sign the HTTP cookie value. The default value is a random value. Unless multiple

Name	Default	Description
		WebHCat instances need to share the secret the random value is adequate.
templeton.kerberos.principal	None	The Kerberos principal to used by the server. As stated by the Kerberos SPNEGO specification, it should be USER/\${HOSTNAME}@{REALM}. It does not have a default value.
templeton.kerberos.keytab	None	The keytab file containing the credentials for the Kerberos principal.