Juju Big Data Documentation

Release 2.0.0

Juju Big Data Developers

November 20, 2015

The `jujubigdata` Python library is an collection of functions and classes for simplifying the development of Juju Charms for Big Data applications. It includes utilities for:

- Interacting with Apache Hadoop

- Connecting to the core Apache Hadoop platform bundle

- Reading and writing configuration in various formats

- Managing distribution-specific configuration in a generalized, maintainable fashion

Apache Hadoop Core Platform Bundle

The platform bundle deploys the core Apache Hadoop platform, providing a basic Hadoop deployment to use directly, as well as endpoints to which to connect additional components, such as Apache Hive, Apache Pig, Hue, etc. It also serves as a reference implementation and starting point for creating charms for vendor-specific Hadoop platform distributions, such as Cloudera or Hortonworks.

1.1 Deploying the Bundle

Deploying the core platform bundle is as easy as:

```
juju quickstart apache-core-batch-processing
```

1.2 Connecting Components

Once the core platform bundle is deployed, you can add additional components, such as Apache Hive:

```
juju deploy cs:trusty/apache-hive
juju add-relation apache-hive plugin
```

Currently available components include:

- Apache Hive
- Apache Pig
- Hue
- Apache Spark
- Apache Zeppelin

1.3 Charming New Components

New components can be added to the ecosystem using one of the following two relations on the apache-hadoop-plugin endpoint charm:

- **hadoop-rest**: This interface is intended for components that interact with Hadoop only via the REST API, such as Hue. Charms using this interface are provided with the REST API endpoint information for both the

NameNode and the ResourceManager. The details of the protocol used by this interface are documented in the `helper class`, which is the recommended way to use this interface.

- **hadoop-plugin**: This interface is intended for components that interact with Hadoop via either the Java API libraries, or the command-line interface (CLI). Charms using this interface will have a JRE installed, the Hadoop API Java libraries installed, the Hadoop configuration managed in `/etc/hadoop/conf`, and the environment configured in `/etc/environment`. The endpoint will ensure that the distribution, version, Java, etc. are all compatible to ensure a properly functioning Hadoop ecosystem. The details of the protocol used by this interface are documented in the `helper class`, which is the recommended way to use this interface.

1.4 Replacing the Core

As long as it supports the same interfaces described above, the core platform can be replaced with a different distrib-tution. The recommended way to create charms for another distribution is to use the core platform charms as the base and modify the `dist.yaml` and `resources.yaml`.

API Documentation

2.1 jujubigdata.relations

2.2 jujubigdata.handlers

2.3 jujubigdata.utils

Indices and tables

- genindex
- modindex