Overview

Table of contents

1. About Zebra

Zebra is an access path library for reading and writing data in a column-oriented fashion. Zebra functions as an abstraction layer between your client application and data on the Hadoop Distributed File System (HDFS). Data is written to HDFS using Zebra's TableStore class. Data is read from HDFS using Zebra's TableLoad class. Zebra supports client applications written as Pig, MapReduce, or streaming jobs. Keep in mind that Zebra works with Zebra tables; you cannot use Zebra to process text or sequence files.

2. Zebra Setup

2.1. Prerequisites

Zebra requires:
- Pig 0.7.0 or later
- Hadoop 0.20.2 or later

Also, make sure the following software is installed on your system:
- JDK 1.6
- Ant 1.7.1

Note: Zebra requires Pig.jar in its classpath to compile and run.

2.2. Download Zebra

Zebra is a Pig contrib project and is available at:
http://svn.apache.org/viewvc/pig/trunk/contrib/zebra/

To work with Zebra you need to check out the Pig trunk:
http://svn.apache.org/repos/asf/pig/trunk/

2.3. Compile Zebra

To compile Zebra follow these steps.

Step 1:
- Move to the top level of your Pig installation
- Run 'ant jar' (this builds the Pig classes and creates the Pig JAR files)
- (optional) Run 'ant -Dtestcase=none test-core' (this builds the Pig test classes which are

needed by the Zebra tests)

Step 2:
- cd ./contrib/zebra
- Run 'ant jar' (this builds the Zebra classes and creates the Zebra JAR file)
- (optional) Run 'ant test' (this verifies that Zebra is working correctly)