Welcome to HDFS command line interface.

In this presentation, I will cover the general usage of the HDFS command line interface and commands specific to HDFS. Other commands should be familiar to anyone with UNIX experience and will not be covered.

The HDFS can be manipulated through a Java API or through a command line interface. All commands for manipulating HDFS through Hadoop's command line interface begin with "hadoop", a space, and "fs". This is the file system shell. This is followed by the command name as an argument to "hadoop fs". These commands start with a dash. For example, the "ls" command for listing a directory is a common UNIX command and is preceded with a dash. As on UNIX systems, ls can take a path as an argument. In this example, the path is the current directory, represented by a single dot.

So I have already started the various Hadoop components and I just want to demonstrate quickly using the Hadoop file system shell. So we will execute Hadoop fs – and we will do a list of the current directory. And this is obviously the current directory in HDFS. Or if we wanted to pass in a different path… This gives you a little bit of the concept of executing the Hadoop file shell.

As we saw for the "ls" command, the file system shell commands can take paths as arguments. These paths can be expressed in the form of uniform resource identifiers or URIs. The URI format consists of a scheme, an authority, and path. There are multiple schemes supported. The local file system has a scheme of "file". HDFS has a scheme called "hdfs". For example, let us say you wish to copy a file called

"myfile.txt" from your local filesystem to an HDFS file system on the localhost. You can do this by issuing the command shown. The cp command takes a
URI for the source and a URI for the destination. The scheme and the authority do not always need to be specified. Instead you may rely on their default values. These defaults can be overridden by specifying them in a file named core-site.xml in the conf directory of your Hadoop installation.

So now let’s do another example. This time we will us the copy command and the various schemes that we just talked about. So hadoop fs - cp and I am going to copy from my biadmin home directory to HDFS. And let’s see the results. And we can now see myfile.txt as having been uploaded to HDFS.

HDFS supports many POSIX-like commands. HDFS is not a fully POSIX compliant file system, but it supports many of the commands. The HDFS commands are mostly easily-recognized UNIX commands like cat and chmod. There are also a few commands that are specific to HDFS such as copyFromLocal. We'll examine a few of these.

copyFromLocal and put are two HDFS-specific commands that do the same thing - copy files from the local filesystem to a location on another filesystem.

Their opposite is the copyToLocal command which can also be referred to as get. This command copies files out of the filesystem you specify and into the local filesystem.

getMerge is an enhanced form of get that can merge the files from multiple locations into a single local file.

setRep lets you override the default level of replication. You can do this for one file or, with the -R option, to an entire tree. This command
returns immediately after requesting the new replication level. If you want the command to block until the job is done, pass the -w option.

IBM, with BigInsights, provides the BigInsights Console as graphical way to work with HDFS.

The Cluster tab provides a simiple way to view that status of the Hadoop components.

The Files tab gives you a graphical mechanism to create directories in HDFS, move data into and out of HDFS, delete files and directories, and so on. It is also possible to edit data directly from within the console.

You have control over file permissions as well. You can change ownership of files and how others are able to access those files.

Even within the console, you have access to the Hadoop Files System Shell. You can execute commands in the same way that you can from the command line. Note that you still have to code the commands in the same way as we described doing so from the command line.

And now I would like to quickly demonstrate the BigInsights Console. Here we are on the Files tab. I am going to expand out users. We can see biadmin as a user there. Here is myfile.txt that was uploaded in a previous demonstration. If I click on myfile.txt, I can see the data over here. If I want, I can even edit that data. Let's create a new directory. So we have now created a new directory. Let's upload myfile.txt to this new directory. Browse to locate it. Under biadmin, we will select myfile.txt and that gets added to the

list. If I wanted to upload multiple files to this directory at one time, I could then just browse for those files and add them to my list here. When I have all of the files that I want added to the list, I click ok. And now we can see that myfile.txt has been added to the new directory. Let's show that we can modify and change permissions and if I want to, I can come in and execute my on Hadoop file shell command. And we have now deleted that file. If I do a refresh, you can see that the file now longer exists.

This concludes the presentation. Thank you for watching.

Here is a list of trademarks that might have been referenced in this presenetation.