

# HBase 101

Ryan Rawson  
Systems Architect, [Stumbleupon.com](http://Stumbleupon.com)

# Introduction

- Quickly get started with HBase
- Use standalone mode on your laptop
- Write “Hello HBase World!”

# Requirements

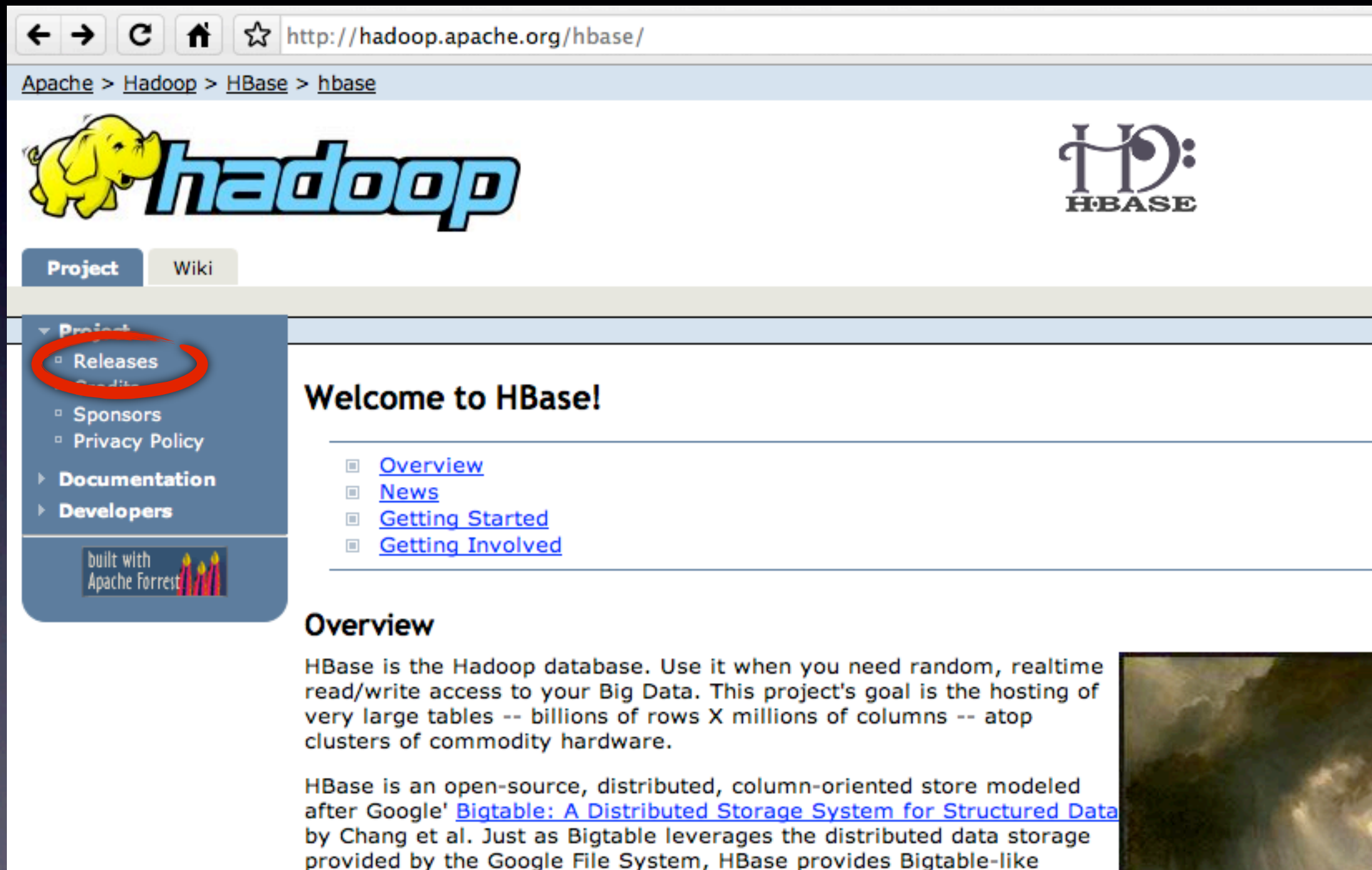
- Java 6
- Windows? Cygwin maybe



# Get Lab on Github

- Handy download:
- <http://github.com/ryanobjc/hbase-lab>



# Download HBase



The screenshot shows a web browser window with the address bar displaying <http://hadoop.apache.org/hbase/>. The breadcrumb navigation path is [Apache](#) > [Hadoop](#) > [HBase](#) > [hbase](#). The page features the Hadoop logo (a yellow elephant) and the HBase logo (a stylized 'H' with a colon). Below the logos are tabs for 'Project' and 'Wiki'. A left sidebar contains a 'Project' dropdown menu with the following items: 'Releases' (highlighted with a red circle), 'Sponsors', 'Privacy Policy', 'Documentation', and 'Developers'. At the bottom of the sidebar is a 'built with Apache Forrest' logo. The main content area has a heading 'Welcome to HBase!' followed by a list of links: 'Overview', 'News', 'Getting Started', and 'Getting Involved'. Below this is an 'Overview' section with two paragraphs of text. The first paragraph describes HBase as a Hadoop database for random, realtime read/write access to Big Data. The second paragraph describes it as an open-source, distributed, column-oriented store modeled after Google's Bigtable. A small image of a cloudy sky is visible in the bottom right corner of the page.

← → ↻ 🏠 ☆ <http://hadoop.apache.org/hbase/>

[Apache](#) > [Hadoop](#) > [HBase](#) > [hbase](#)

**Project** Wiki

▼ Project

- ▣ Releases
- ▣ Sponsors
- ▣ Privacy Policy
- ▶ Documentation
- ▶ Developers

built with Apache Forrest

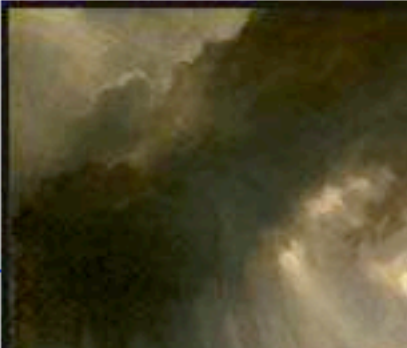
## Welcome to HBase!

- ▣ [Overview](#)
- ▣ [News](#)
- ▣ [Getting Started](#)
- ▣ [Getting Involved](#)

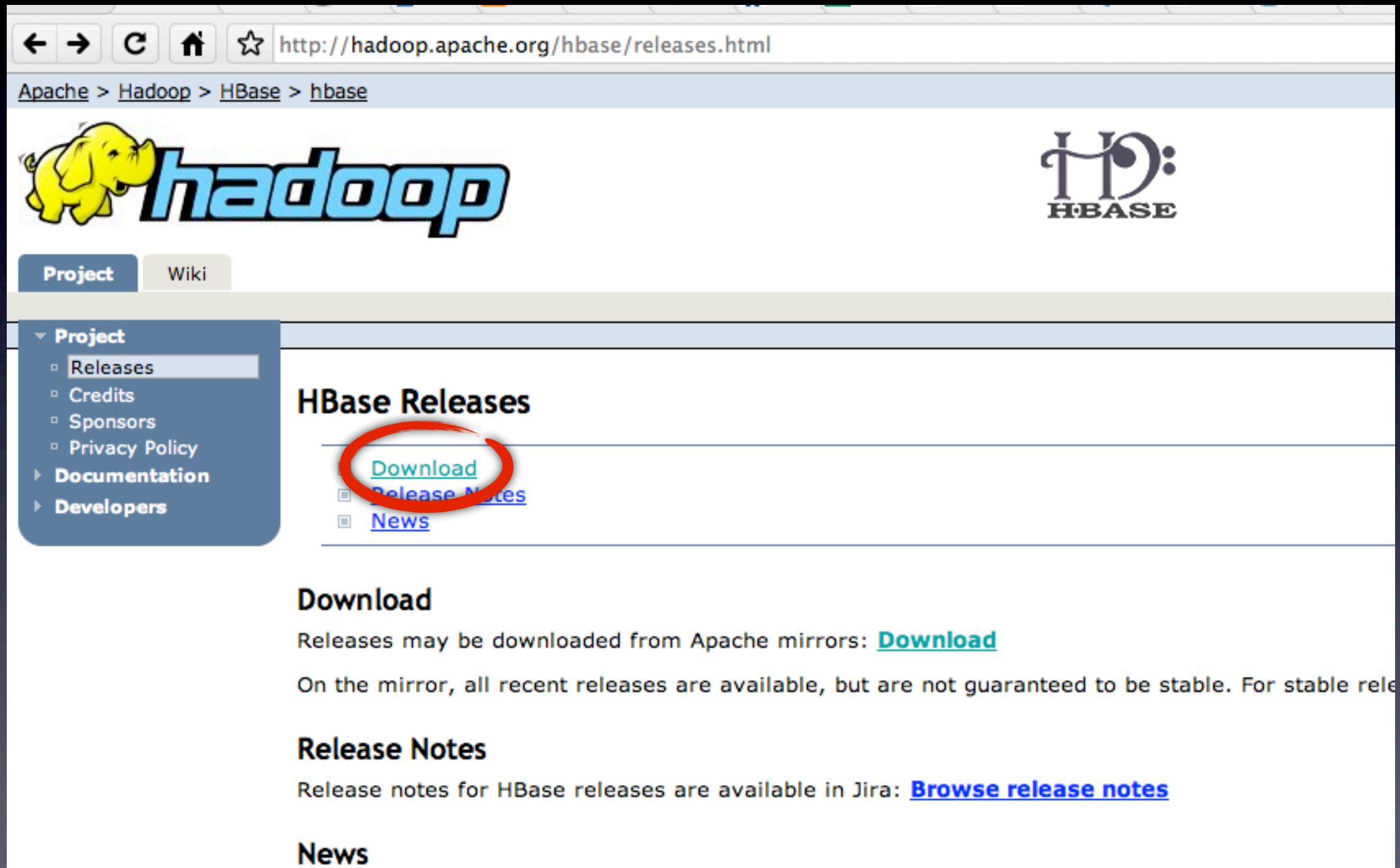
### Overview

HBase is the Hadoop database. Use it when you need random, realtime read/write access to your Big Data. This project's goal is the hosting of very large tables -- billions of rows X millions of columns -- atop clusters of commodity hardware.

HBase is an open-source, distributed, column-oriented store modeled after Google's [Bigtable: A Distributed Storage System for Structured Data](#) by Chang et al. Just as Bigtable leverages the distributed data storage provided by the Google File System, HBase provides Bigtable-like




# Download HBase



The screenshot shows a web browser window with the address bar displaying `http://hadoop.apache.org/hbase/releases.html`. The breadcrumb navigation path is `Apache > Hadoop > HBase > hbase`. The page features the Hadoop logo (a yellow elephant) and the HBase logo (a stylized 'H' with a colon). Below the logos are tabs for 'Project' and 'Wiki'. A left sidebar menu is expanded, showing options like 'Releases', 'Credits', 'Sponsors', 'Privacy Policy', 'Documentation', and 'Developers'. The main content area is titled 'HBase Releases' and contains a red circle around a 'Download' link. Below this, there are sections for 'Download', 'Release Notes', and 'News', each with descriptive text and additional links.

← → ↻ 🏠 ☆ `http://hadoop.apache.org/hbase/releases.html`

Apache > Hadoop > HBase > hbase

Project Wiki

▼ Project

- Releases
- Credits
- Sponsors
- Privacy Policy
- ▶ Documentation
- ▶ Developers

## HBase Releases

- [Download](#)
- [Release Notes](#)
- [News](#)

### Download

Releases may be downloaded from Apache mirrors: [Download](#)

On the mirror, all recent releases are available, but are not guaranteed to be stable. For stable releases, see the [Release Notes](#).

### Release Notes

Release notes for HBase releases are available in Jira: [Browse release notes](#)

### News



# Untar package

- `tar -zxvf hbase-0.20.3.tar.gz`

# Optional

- Use my 0.20.3-r1 for easier standalone on windows
- Patch will be rolled into 0.20.4
- It's on github
- If you do this, replace hbase-0.20.3 with hbase-0.20.3-r1 in the following examples



# Start HBase standalone

- `cd hbase-0.20.3`
- `bin/start-hbase.sh`
- `bin/hbase shell`

# Create table

- In shell:
- bin/hbase shell
- > create 'test\_table', {NAME => 'test'}

# Setup project

- Using IntelliJ or Eclipse, new project
- Add hbase-0.20.3/\*.jar hbase-0.20.3/lib/\*.jar and hbase-0.20.3/conf to classpath



# Ant?

- To build:
- \$ ant
- To get classpath:
- \$ ant classpath

# Client I: Put

```
HBaseConfiguration conf = new HBaseConfiguration();  
  
HTable table = new HTable(conf, "test_table");  
  
Put put = new Put(Bytes.toBytes("hello_row"));  
put.add(Bytes.toBytes("test"), Bytes.toBytes("qualifier"),  
        Bytes.toBytes("Hello HBase World!"));  
table.put(put);
```

# Examine results

- In Shell:
- > get 'test\_table', 'hello\_world'
- > scan 'test\_table'



# Client 2: Get

```
HBaseConfiguration conf = new HBaseConfiguration();
```

```
HTable table = new HTable(conf, "test_table");
```

```
Get get = new Get(Bytes.toBytes("hello_row"));
```

```
Result result = table.get(get);
```

```
byte [] value = result.getValue(Bytes.toBytes("test"),
```

```
    Bytes.toBytes("qualifier"));
```

```
System.out.println(Bytes.toString(value));
```

# Client 3: Scan

- Too much for presentation!

# What's next?

- Running HBase on a cluster
- Client configuration requirements:
  - quorum, ZK path
  - Adjusting `hadoop-env.sh` to use HBase