

HBase, the Open-Source BigTable

Henggang Cui

15-799 lightning talk

What is HBase

- HBase
 - an open-source implementation of BigTable
 - a distributed, scalable, big data store
 - for hosting billions of rows X millions of columns
 - on top of clusters of commodity hardware

Apache Open-source Systems VS Google Systems

- HDFS \leftrightarrow GFS
- HBase \leftrightarrow BigTable
- Hadoop \leftrightarrow MapReduce
- ...

Use HBase

- Via Hbase shell

```
hbase> create 'test', 'cf'
```

```
hbase> put 'test', 'row1', 'cf:a', 'v1'
```

```
hbase> scan 'test'
```

```
ROW    COLUMN+CELL
```

```
row1   column=cf:a, timestamp=1288380, value=v1
```

```
hbase> get 'test', 'row1'
```

```
COLUMN    CELL
```

```
cf:a      timestamp=1288380, value=v1
```

Use Hbase (Cont.)

- In Java code

```
Configuration conf = HBaseConfiguration.create();
HTable hTable = new HTable(conf, "test");
Put put1 = new Put(toBytes("row1"));
put1.add(toBytes("cf"),
        toBytes("a"),
        toBytes("val1"));
hTable.put(put1);
```

Use Hbase (Cont.)

- In Hadoop Job
 - Each map job usually processes one row

```
@Override
protected void map(ImmutableBytesWritable rowKey,
                   Result result,
                   Context context) {
    ...
}
```