

 **wibi!data**

# Sqoop: The Early Days

---

Aaron Kimball – CTO

WibiData, Inc.

# It started with a talk...

- In Feb '09 I was asked to give a talk at the Bay Area HUG on DBInputFormat
- The next month, I cleaned this up and made a blog post:

[cloudera.com/blog/2009/03/database-access-with-hadoop/](http://cloudera.com/blog/2009/03/database-access-with-hadoop/)

# Using this InputFormat is hard...

```
class MyRecord implements Writable, DBWritable {
    long id;
    String name;

    public void readFields(DataInput in) throws IOException {
        this.id = in.readLong();
        this.name = Text.readString(in);
    }
    public void readFields(ResultSet resultSet) throws SQLException {
        this.id = resultSet.getLong(1);
        this.name = resultSet.getString(2);
    }
    public void write(DataOutput out) throws IOException {
        out.writeLong(this.id);
        Text.writeString(out, this.name);
    }
    public void write(PreparedStatement stmt) throws SQLException {
        stmt.setLong(1, this.id);
        stmt.setString(2, this.name);
    }
}
```

# So then I got to thinking...

- Writing serializable classes sucks.
- The field types in the classes directly match the types in the database.
- This is the sort of thing we have robots for.

# Hackathons



# The first version

- HADOOP-5815 – Submitted May '09
- Supported MySQL through JDBC...
  - and that's about it
- Didn't have any special connectors
- Named by Aaron Newton
  - (I wanted to call it “SQLImport”)

# “Direct imports” to MySQL

- HADOOP-5844... a couple weeks later
- The first special-purpose connector
  - Referred to as “local” mode – this got changed.
- (I think it barely worked)



# More timeline...

- Oracle support: July '09
  - (Around here was when The Project Split happened; this was MAPREDUCE-654)
- Exports: Oct '09
- Moved to GitHub: March '10
- Third-party connectors
  - Quest, Netezza, Teradata...
- Sqoop chapter in *Hadoop: The Definitive Guide*
- Moved to Apache: August '11
- Apache TLP: April '12

# How do you build community?

“It’ll show up when it’s ready”

- Doug C.

# How do you build community?

- Good documentation
- Good developer documentation
- Pay attention to mailing lists
- Fix the glaring errors:
  - ... suck less each time and don't worry

# Thanks to...

- Arvind Prabhakar
- Kathleen Ting
- The ASF
- All of you!