

Elephant Bird Installation

Elephant Bird benötigt einen aktuellen Compiler für Protocol Buffers (mindestens Version 2.3)

Der Protocol Buffers Compiler ist den Ubuntu Paketquellen enthalten und vorinstalliert, andernfalls könnt ihr ihn je nach Distribution wie folgt nachinstallieren.

Quellcode

1. `sudo apt-get install protobuf-compiler`

Elephant Bird benötigt außerdem, dass Thrift auf eurem Computer installiert ist. Wie ihr Thrift installieren könnt erfahrt ihr in diesem Wiki Beitrag: [wiki][Apache Thrift Installation](#)[/wiki]

Um die JAR Dateien für Elephant Bird zu bauen wird ant verwendet. Neben ant müsst ihr noch die aktuelle JDK installieren.

Statt der Sun JDK könnt ihr natürlich auch openjdk-6-jdk verwenden. Aber gerade wenn ihr die Cloudera Distribution verwendet, dann bringen einige Pakete die Sun JDK bereits als Paketabhängigkeit mit.

Quellcode

1. `sudo apt-get install sun-java6-jdk ant`

Ladet euch nun die aktuellste stabile Version von Elephant Bird herunter: github.com/kevinweil/elephant-bird

Mit wget funktioniert der Download wie folgt. Ihr solltet den SSL Check für github deaktivieren.

Quellcode

1. `wget --no-check-certificate https://github.com/kevinweil/elephant-bird/tarball/master`

Nun entpackt ihr die Dateien und startet anschließend das Bauen der JAR Dateien

Quellcode

1. `tar xvfz kevinweil-elephant-bird*`
2. `cd elephant-bird/`
3. `ant`

Die erfolgreiche Installation sieht bei mir wie folgt aus:

Quellcode

1. Buildfile: /home/tb/elephant-bird/build.xml
2. release:
4. [echo] Building in release mode...
6. init:
8. compile-protobuf:
9. [exec] Result: 1
10. [apply] Applied thrift to 1 file and 0 directories.
11. [apply] Applied protoc to 4 files and 0 directories.
12. [javac] /home/tb/elephant-bird/build.xml:170: warning: 'includeantruntime' was not set, defaulting to build.sysclasspath=last; set to false for repeatable builds
13. [javac] Compiling 9 source files to /home/tb/elephant-bird/build/classes
14. [javac] Note: Some input files use unchecked or unsafe operations.
15. [javac] Note: Recompile with -Xlint:unchecked for details.
16. [javac] /home/tb/elephant-bird/build.xml:116: warning: 'includeantruntime' was not set, defaulting to

- build.sysclasspath=last; set to false for repeatable builds
- 17. [javac] Compiling 45 source files to /home/tb/elephant-bird/build/classes
- 18. [javac] /home/tb/elephant-bird/src/java/com/twitter/elephantbird/pig/util/ProtobufToPig.java:13: warning: com.sun.org.apache.xerces.internal.impl.dv.xs.SchemaDateTimeException is Sun proprietary API and may be removed in a future release
- 19. [javac] import com.sun.org.apache.xerces.internal.impl.dv.xs.SchemaDateTimeException;
- 20. [javac] ^
- 21. [javac] Note: Some input files use or override a deprecated API.
- 22. [javac] Note: Recompile with -Xlint:deprecation for details.
- 23. [javac] Note: Some input files use unchecked or unsafe operations.
- 24. [javac] Note: Recompile with -Xlint:unchecked for details.
- 25. [javac] 1 warning
- ~~26.~~ compile:
- 28. [javac] /home/tb/elephant-bird/build.xml:90: warning: 'includeantruntime' was not set, defaulting to build.sysclasspath=last; set to false for repeatable builds
- 29. [javac] Compiling 52 source files to /home/tb/elephant-bird/build/classes
- 30. [javac] Note: Some input files use or override a deprecated API.
- 31. [javac] Note: Recompile with -Xlint:deprecation for details.
- 32. jar:
- 34. [mkdir] Created dir: /home/tb/elephant-bird/dist
- 35. [jar] Building jar: /home/tb/elephant-bird/dist/elephant-bird-1.2.1.jar
- ~~36.~~ release-jar:
- 38. BUILD SUCCESSFUL
- 40. Total time: 9 seconds

Alles anzeigen

Ihr findet die Elephant Bird JAR Datei von Elephant Bird im dist Ordner.

== Elephant Bird mit Pig 0.8 ==

Elephant Bird in der Version 1 unterstützt kein Pig 0.8, wie es in den Cloudera Quellen zu finden ist. Elephant Bird in Version 2 mit Unterstützung für Pig 0.8 ist zum gegenwärtigen Zeitpunkt noch in der Entwicklung. Um den aktuellen Snapshot zu bauen, müsst ihr die Quellen aus git kompilieren.

Quellcode

1. git clone https://github.com/dvryaboy/elephant-bird.git
2. cd elephant-bird
3. git branch -t pig-08 remotes/origin/pig-08
4. git checkout pig-08
6. # build jar file with ant
7. ant

Der Name der Jar Datei und auch die Paketnamen der Pig Loader haben sich verändert. Hier ein Beispiel Pig Script:

Quellcode

1. register dist/elephant-bird-2.0-SNAPSHOT.jar;
2. raw_data = load '/tmp/thrift/' using com.twitter.elephantbird.pig8.load.LzoThriftBlockPigLoader('com.example.thrift.VectorSequence');