Development Guide for Kylin 4

Source code

```
git clone https://github.com/apache/kylin.git
# Compile
mvn clean install -DskipTests
```

The environment on dev machine

Install Maven

The latest maven can be found at http://maven.apache.org/download.cgi, we create a symbol link so that `mvn` can be run anywhere.

```
cd ~
wget http://xenia.sote.hu/ftp/mirrors/www.apache.org/maven/maven-3/3.2.5/binaries/apache-maven-3.2.5-bin.tar.gz
tar -xzvf apache-maven-3.2.5-bin.tar.gz
ln -s /root/apache-maven-3.2.5/bin/mvn /usr/bin/mvn
```

Install Spark

Manually install the Spark binary in a local folder like /usr/local/spark. Kylin supports the community version of Spark. You can go to apache spark's official website and download spark 2.4.6.

How to Debug

There are two modes to debug source code: Debug with local metadata(recommended), or debug with Hadoop sandbox.

Configuration

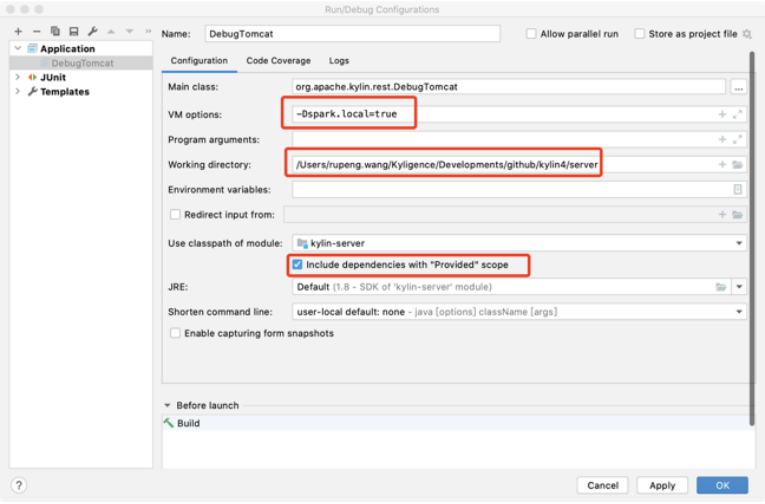
Debug with local metadata

1. Edit the properties of $KYLIN_SOURCE_DIR/examples/test_case_data/sandbox/kylin.properties

```
# Need to use absolute path
kylin.metadata.url=${KYLIN_SOURCE_DIR}/examples/test_case_data/sample_local
kylin.storage.url=${KYLIN_SOURCE_DIR}/examples/test_case_data/sample_local
kylin.env.zookeeper-is-local=true
kylin.env.hdfs-working-dir=file://$KYLIN_SOURCE_DIR/examples/test_case_data/sample_local
kylin.engine.spark-conf.spark.master=local
kylin.engine.spark-conf.spark.eventLog.dir=/path/to/local/dir
kylin.engine.spark-conf.spark.sql.shuffle.partitions=1
kylin.env=LOCAL
```

2. Open DebugTomcat.java, start to debug.

3. Edit Configuration



VM options "-Dspark.local=true" is for query engine.

Debug with Hadoop sandbox

Local configuration must be modified to point to your Hadoop sandbox (or CLI) machine.

- In **examples/test_case_data/sandbox/kylin.properties**
  - Find `sandbox` and replace with your Hadoop hosts (if you're using HDP sandbox, this can be skipped)
  - Find `kylin.job.use-remote-cli` and change it to "true" (in the code repository the default is false, which assumes running it on Hadoop CLI)
  - Find `kylin.job.remote.cli.username` and `kylin.job.remote.cli.password`, fill in the user name and password used to login Hadoop cluster for Hadoop command execution; If you're using HDP sandbox, the default username is `root` and password is `hadoop`.
- In **examples/test_case_data/sandbox**
  - For each configuration XML file, find all occurrences of `sandbox` and `sandbox.hortonworks.com`, replace with your Hadoop hosts; (if you're using HDP sandbox, this can be skipped)

An alternative to the host replacement is updating your `hosts` file to resolve `sandbox` and `sandbox.hortonworks.com` to the IP of your sandbox machine.

Launch Kylin Web Server

Copy server/src/main/webapp/WEB-INF to webapp/app/WEB-INF

```
cp -r server/src/main/webapp/WEB-INF webapp/app/WEB-INF
```

Download JS for Kylin web GUI. `npm` is part of `Node.js`, please search about how to install it on your OS.

```
cd webapp
npm install -g bower
bower --allow-root install
```

If you encounter a network problem when run "bower install", you may try:

```
git config --global url."git://".insteadOf https://
```

Note, if on Windows, after installing bower, need to add the path of "bower.cmd" to system environment variable 'PATH', and then run:

```
bower.cmd --allow-root install
```

In IDE, launch `org.apache.kylin.rest.DebugTomcat`. Please set the path of the “server” module as the “Working directory”, set “kylin-server” for “Use classpath of module”, and check the “Include dependencies with ‘Provided’ scope” option in IntelliJ IDEA 2018. If you’re using IntelliJ IDEA 2017 and older, you need modify “server/kylin-server.iml” file, replace all “PROVIDED” to “COMPILE”, otherwise a “java.lang.NoClassDefFoundError: org/apache/catalina/LifecycleListener” error may be thrown.

You may also need to tune the VM options:

```
-Dhdp.version=2.4.0.0-169 -DSPARK_HOME=/usr/local/spark -Dkylin.hadoop.conf.dir=/workspace/kylin/examples/test_case_data/sandbox -Xms800m -Xmx800m -XX:PermSize=64M -XX:MaxNewSize=256m -XX:MaxPermSize=128m
```

Also remember that if you debug with local mode, you should add a VM option for the query engine:

```
-Dspark.local=true
```

If you worked with Kerberized Hadoop Cluster, the additional VM options should be set:

```
-Djava.security.krb5.conf=/etc/krb5.conf -Djava.security.krb5.principal=kylin -Djava.security.krb5.keytab=/path/to/kylin/keytab
```

And Hadoop environment variable:

```
HADOOP_USER_NAME=root
```

By default Kylin server will listen on the 7070 port; If you want to use another port, please specify it as a parameter when run `DebugTomcat`.

Check Kylin Web at `http://localhost:7070/kylin` (user:ADMIN, password:KYLIN)

How to Package and Deploy

```
cd ${KYLIN_SOURCE_CODE}
# For HDP2.x
./build/script/package.sh

# For CDH5.7
./build/script/package.sh -P cdh5.7
# After finished, the package will be avaliable in the directory ${KYLIN_SOURCE_CODE}/dist/

# If running on HDP, you need to uncomment the following properties in kylin.properties
kylin.engine.spark-conf.spark.driver.extraJavaOptions=-Dhdp.version=current
kylin.engine.spark-conf.spark.yarn.am.extraJavaOptions=-Dhdp.version=current
kylin.engine.spark-conf.spark.executor.extraJavaOptions=-Dhdp.version=current
```