

Spark build workaround steps:

Github:

Parquet:

<https://github.com/intel-innersource/libraries.databases.thirdparty.parquet-mr/tree/xf-parquet-1.12.2-2023-03-30>

Spark:

<https://github.com/intel-innersource/libraries.databases.thirdparty.parquet-mr/tree/xf-spark3.3.2-2023-03-30>

1. Source env variable:

```
export JAVA_HOME=/home/fxie/openjdk17/jdk-17.0.2/
export PATH=/home/fxie/openjdk17/jdk-17.0.2/bin:$PATH
export PATH=/home/fxie/bbb/libraries.databases.thirdparty.parquet-mr/build/scala-2.12.17/bin:$PATH
```

2. Build parquet-mr package

git clone -b xf-parquet-1.12.2-2023-03-30 <https://github.com/intel-innersource/libraries.databases.thirdparty.parquet-mr.git>

mvn clean install -DskipTests -P vector-plugins -Pci-test -Djapicmp.skip=true

3. Build Spark

git clone -b xf-spark3.3.2-2023-03-30 <https://github.com/intel-innersource/libraries.databases.thirdparty.parquet-mr.git>

Cd \$SPARK_SRC_HOME

./build/mvn -DskipTests clean package

Maybe meet some error during build.

Cd \$SPARK_SRC_HOME/sql/catalyst, build catalyst as below:

.././build/mvn -DskipTests clean package

Add a jar link spark-catalyst_2.12-3.3.2.jar to your maven repository

Cd ~/.m2/repository/org/apache/spark/spark-catalyst_2.12/3.3.2,

Remove spark-catalyst_2.12-3.3.2.jar

Add a link to spark-catalyst_2.12-3.3.2.jar -> /home/fxie/spark3.3.2-

src/libraries.databases.thirdparty.parquet-mr/sql/catalyst/target/spark-catalyst_2.12-3.3.2.jar

vi \${SPARK_SRC_HOME}/pom: do below change comment out "incremental" add "all" :

```
<!--recompileMode>incremental</recompileMode-->
```

```
<recompileMode>all</recompileMode>
```

Cd \${SPARK_SRC_HOME}/sql/core

.././build/mvn -DskipTests clean package

The jar spark-sql_2.12-3.3.2.jar

under /home/fxie/spark3.3.2-src/libraries.databases.thirdparty.parquet-mr/sql/core/target is new jar

include vector api implementation

Copy new built jar into your Spark environment

parquet-column-1.12.2.jar
parquet-encoding-1.12.2.jar
parquet-encoding-vector-1.13.0-SNAPSHOT.jar
parquet-hadoop-1.12.2.jar
spark-catalyst_2.12-3.3.2.jar
spark-sql_2.12-3.3.2.jar