

```
$HARVARD_DATA_TOOLS_BASE/  
java/data_client  
java/data_tool_generator  
java/aws_data_tools  
java/canvas_data_hadoop  
python/canvas_generate_tools.py  
schema/phase_1_schema_additions  
schema/phase_1_schema_additions
```

Generate
Code

```
$GENERATED_CODE_DIR/  
aws_data_tools.jar  
create_redshift_tables.sql  
hadoop_jobs.jar  
phase_1_copy_unmodified_files.sh  
phase_1_create_tables.sh  
phase_1_hive.sh  
phase_2_copy_unmodified_files.sh  
phase_2_create_tables.sh  
phase_2_hive.sh  
s3_to_redshift_loader.sql
```

```
$HARVARD_DATA_TOOLS_BASE/  
hive/phase_1/*.q  
$GENERATED_CODE_DIR/  
hadoop_jobs.jar  
phase_1_copy_unmodified_files.sh  
phase_1_create_tables.sh  
phase_1_hive.sh
```

```
$HDFS/PHASE_0  
account_dim/  
assignment_dim/  
...  
wiki_page_dim/  
wiki_page_fact/
```

Phase 1

```
$HDFS/PHASE_1  
account_dim/  
assignment_dim/  
...  
wiki_page_dim/  
wiki_page_fact/
```

```
$HARVARD_DATA_TOOLS_BASE/  
python/canvas_download_and_verify.py  
$GENERATED_CODE_DIR/  
aws_data_tools.jar
```

Download
and Verify

```
$CANVAS_S3_LOCATION/  
account_dim/  
assignment_dim/  
...  
wiki_page_dim/  
wiki_page_fact/  
DynamoDB Tables:  
dumps
```

```
$HARVARD_DATA_TOOLS_BASE/  
hive/phase_2/*.q  
$GENERATED_CODE_DIR/  
hadoop_jobs.jar  
phase_2_copy_unmodified_files.sh  
phase_2_create_tables.sh  
phase_2_hive.sh
```

```
$HDFS/PHASE_1  
account_dim/  
assignment_dim/  
...  
wiki_page_dim/  
wiki_page_fact/
```

Phase 2

```
$HDFS/PHASE_2  
account_dim/  
assignment_dim/  
...  
wiki_page_dim/  
wiki_page_fact/
```

```
$CANVAS_S3_LOCATION/  
account_dim/  
assignment_dim/  
...  
wiki_page_dim/  
wiki_page_fact/
```

Copy S3
to HDFS

```
$HDFS/PHASE_0  
account_dim/  
assignment_dim/  
...  
wiki_page_dim/  
wiki_page_fact/
```

```
$GENERATED_CODE_DIR/  
create_redshift_tables.sql  
s3_to_redshift_loader.sql  
$HDFS/PHASE_2  
account_dim/  
assignment_dim/  
...  
wiki_page_dim/  
wiki_page_fact/
```

Copy to
Redshift

```
Redshift Tables:  
account_dim  
assignment_dim  
...  
wiki_page_dim  
wiki_page_fact
```

```
$GENERATED_CODE_DIR/  
phase_n_create_tables.sh  
$HDFS/PHASE_n-1  
account_dim/  
...  
wiki_page_fact/
```

Phase *n*
Create Tables

```
HDFS Tables:  
in_account_dim  
out_account_dim  
...  
in_wiki_page_fact  
out_wiki_page_fact  
$HDFS/PHASE_n  
account_dim/  
...  
wiki_page_fact/
```

```
$GENERATED_CODE_DIR/  
hadoop_jobs.jar  
$HDFS/PHASE_n-1  
account_dim/  
...  
wiki_page_fact/
```

Phase *n*
Hadoop

```
$HDFS/PHASE_n  
account_dim/  
...  
wiki_page_fact/
```

```
$HARVARD_DATA_TOOLS_BASE/  
hive/phase_n/*  
$GENERATED_CODE_DIR/  
phase_n_hive.sh  
HDFS Tables:  
in_account_dim  
...  
in_wiki_page_fact
```

Phase *n* Hive

```
HDFS Tables:  
out_account_dim  
...  
out_wiki_page_fact
```

```
$GENERATED_CODE_DIR/  
phase_n_copy_unmodified_files.sh  
$HDFS/PHASE_n-1  
account_dim/  
...  
wiki_page_fact/
```

Phase *n* Copy
Tables

```
$HDFS/PHASE_n  
account_dim/  
...  
wiki_page_fact/
```

```
$GENERATED_CODE_DIR/  
aws_data_tools.jar  
REDSHIFT TABLES:  
account_dim  
...  
wiki_page_fact  
HDFS TABLES:  
phase_1_verify_account_dim  
phase_1_verify_table_1  
...  
verify_wiki_page_fact  
$HDFS/PHASE_n  
account_dim/  
...  
wiki_page_fact/
```

Phase *n*
Verify

```
DynamoDB Tables:  
phase_n_verify
```