

GTAS ETL Jobs and ETL Scheduler Deployment Guide (Linux)

1. Required tools

- Java 1.8
- Pentaho Data Integration tool (See separate installation guide)
- Neo4j Graph Database (see separate installation guide) – required for Neo4j ETL
- Elasticsearch 7.5.2 and Kibana 7.5.2 - required for the report ETL

2. Create Local ETL Environment

- a. Create your local ETL parent directory *gtas-etl* in the root directory (/)

Note: Although the ETL jobs and scheduler can be installed under any directory, the following instruction show the installations under the root directory (/). If you choose to create this parent directory at a different location, make sure that you review and update the properties file listed in the next section.

- b. Pull the latest from the GTAS GitHub repository. Go to the *gtas-etl-main* directory and copy ETL and configuration files with their directories to your local ETL parent directory, *gtas-etl*

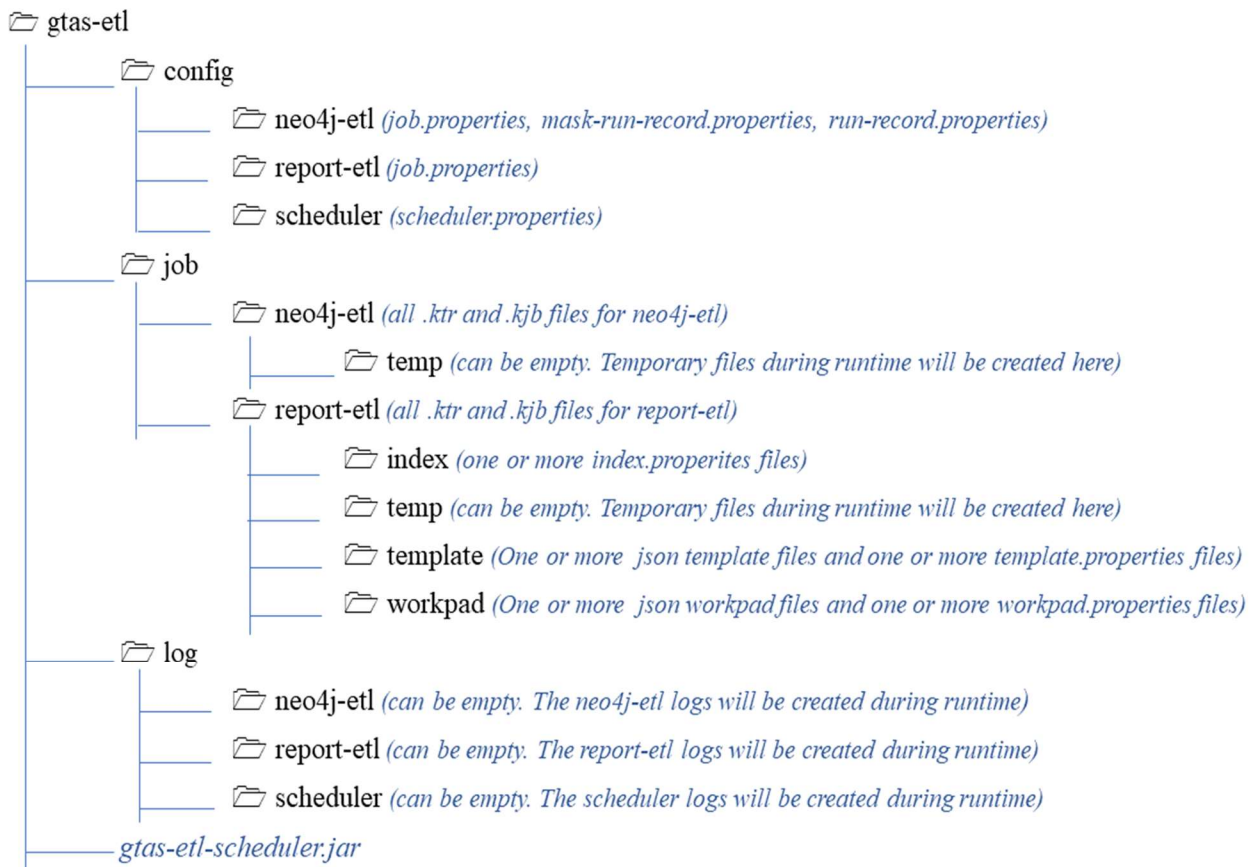
Copy from	Copy to
GTAS/gtas-etl-main/ (all files and directories)	/gtas-etl/

- c. Go to the GTAS/gtas-etl-scheduler project and build it as follows:

- *mvn clean install*

After a successful build, copy the *gtas-etl-scheduler.jar* file from *GTAS/gtas-etl-scheduler/target* to your ETL parent directory */gtas-etl*

- d. Verify that your local ETL environment looks as follows:



3. Update Property Files

Note: If you did not create `gtas-etl` in the root directory (`/`), please review each property file carefully. In addition to the following properties, you will have to update properties with values pointing to a file or a directory structure.

a. **`/gtas-etl/config/scheduler/scheduler.properties`**

- No update is required

b. **`/gtas-etl/config/neo4j-etl/job.properties`**

- `EXT_VAR_GTAS_DB_USER_NAME`
- `EXT_VAR_GTAS_DB_PASSWORD`
- `EXT_VAR_NEO4J_DB_USER_NAME`
- `EXT_VAR_NEO4J_DB_PASSWORD`

c. **`/gtas-etl/config/report-etl/job.properties`**

- `EXT_VAR_GTAS_DB_USER_NAME`
- `EXT_VAR_GTAS_DB_PASSWORD`
- `EXT_VAR_ELSEARCH_PROTOCOL`
- `EXT_VAR_ELSEARCH_HOST_NAME`
- `EXT_VAR_ELSEARCH_USERNAME`
- `EXT_VAR_ELSEARCH_PASSWORD`
- `EXT_VAR_KIBANA_PROTOCOL`
- `EXT_VAR_KIBANA_HOST_NAME`
- `EXT_VAR_KIBANA_USERNAME`
- `EXT_VAR_KIBANA_PASSWORD`

Note: If you have configured your Kibana and Elasticsearch environments to use the HTTPS protocol with SSL certificate, then import the public certificate to your local java keystore and set the following properties:

- EXT_VAR_JAVA_TRUST_STORE
- EXT_VAR_JAVA_TRUST_STORE_PASSWORD

4. Launch the gtas-etl-scheduler

- a. Start the scheduler as follows
- `java -jar gtas-etl-scheduler.jar`

5. Important features

- a. Job Scheduler properties

Common ETL Properties		
Property	Description	Allowed Values
opSystem	The operating system on which the ETL is running	windows/linux
pdiDir	Points to the Pentaho Kitchen executable file. The full file path to kitchen.sh (for Linux) or kitchen.bat (for Windows)	Path to .../data-integration/./kitchen.sh (on Linux) or .../data-integration/kitchen.bat (on Windows)
Neo4j ETL properties		
Property	Description	Allowed Value
neo4j.execInterval	The time interval to launch the Neo4j ETL job after its first execution.	The recommended minimum is 60
neo4j.enableEtl	Enables/Disables the Neo4j ETL job	true/false
neo4j.logLevel	The log level for the Neo4j ETL job	
neo4j.etlName	The name given to the Neo4j ETL job	Any name. The default name is NEO4J_ETL
neo4j.configFilePropertyName	The global parameter name for the config property file name	EXT_ETL_CONFIG_FILE
neo4j.jobDir	The main entry point of the Neo4j ETL job	Path to /gtas-etl/job/neo4j-etl/gtas-to-neo-job.kjb
neo4j.logDir	The log file of the Neo4j ETL job	Path to /gtas-etl/log/neo4j-etl/neo4j-etl-log
neo4j.configFile	The main configuration file of the Neo4j ETL job	Path to /gtas-etl/config/neo4j-etl/job.properties
report.execInterval	The time interval to launch the Report ETL job after its first execution.	The recommended minimum is 90
report.enableEtl	Enables/Disables the Report ETL job	true/false
report.logLevel	The log level for the Neo4j ETL job	

report.etlName	The name given to the Report ETL job	Any name. The default name is REPORT_ETL
report.configFilePropertyName	The global parameter name for the config property file name	EXT_ETL_CONFIG_FILE
report.jobDir	The main entry point of the Report ETL job	Path to /gtas-etl/job/report-etl/gtas-report-job.kjb
report.logDir	The log file of the Report ETL job	Path to /gtas-etl/log/report-etl/report-etl-log
report.configFile	The main configuration file of the Report ETL job	Path to /gtas-etl/config/report-etl/job.properties