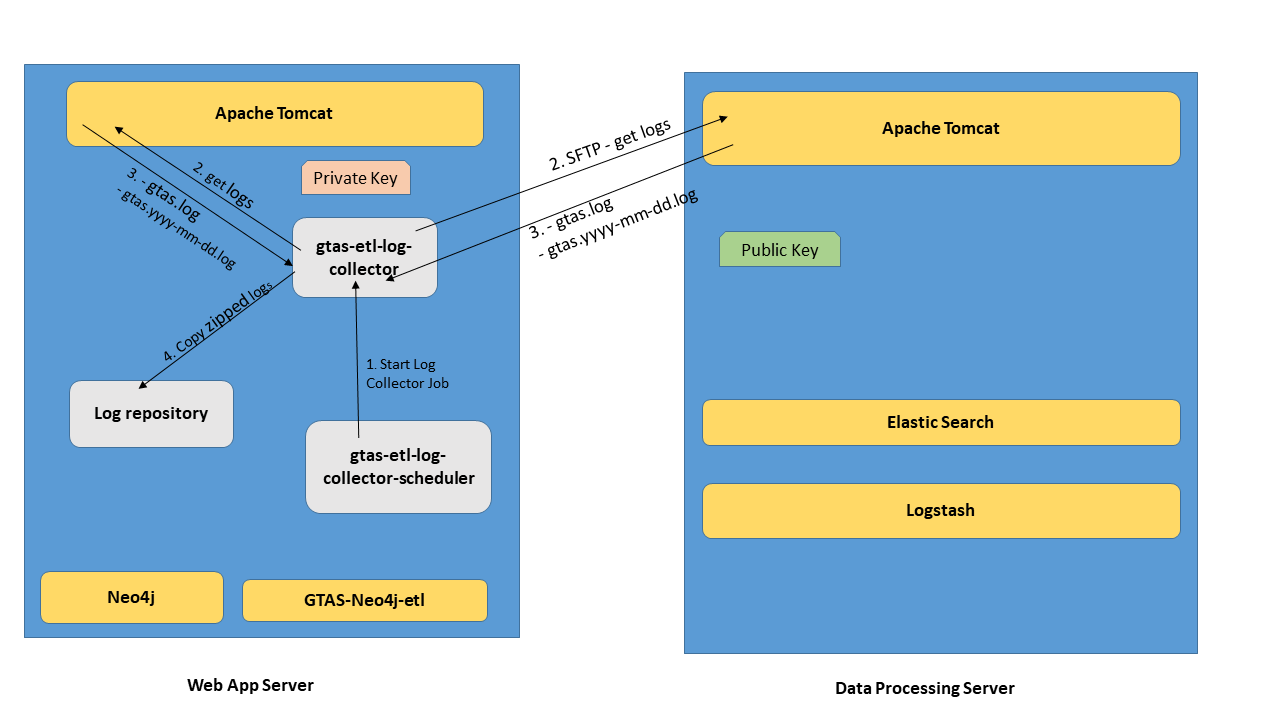
**Deployment Guide for GTAS Log Collector Job**



**Required tools on the host server (Web App Server):**

* JRE 1.8
* Pentaho Kettle Data Integration Remix (kettle-neo4j-remix-8.2.0.3-519-REMIX)

**Features:**

* Zip and copy Apache Tomcat logs from the remote server to a repository on the host server. Other than the current log, all previous logs will be deleted from the remote server once they are copied to the local repository on the host server.
* Zip and copy Apache Tomcat Web logs from the host (local) server to the local repository. Other than the current log, all previous logs will be deleted once they are copied to the local repository.
* Enable/Disable a flag to delete or keep collected logs in the repository that are older than N days
* Enable/Disable a flag to use or not to use RSA key for SFTP communication between servers

1. **Create a user with an appropriate scope on the Data Processing server**

Although any current active user can be used for this purpose, it is strongly recommended to have a separate user for this task.

* NOTE: This user should have read/write access on the folder in which the log exists on the Data Processing server

1. **(Optional) Create Public and Private keys for SFTP**

* NOTE 1: Setting a passphrase for the private key is strongly recommended.
* NOTE 2. The private key should be in *.pem* format

1. **(Optional) Install the Public Key on the Data Processing Server for the user created in step 1.**

* NOTE 1. Make sure the SFTP port (22) is open for incoming and outgoing traffic on both the Data Processor and Web Application servers.
* NOTE 2. Before proceeding to the next step, test step 1, 2, and 3 by making an SFTP connection to the Data Processing Server using the above user, public key, and private key.

1. **Build the scheduler for the GTAS Log Collector ETL**

4.1 Pull the gtas-log-collector-scheduler from https://github.com/US-CBP/GTAS.git

4.2 Build the project to create the ***gtas-log-collector-sch-1.jar***

1. **Deploy the GTAS Log Collector ETL Job**

Create the following directory structure and copy the scheduler.jar, all configuration files, ETL files, and the private keyas follows.

* /gtas-log-etl

*→ gtas-log-collector-sch-1.jar*

* + /config

→ *application.properties*

→ *gtas-log-etl-config.properties*

→ *run-record. properties*

* + /job

*→ All ETL files (.kjb and .ktr files)*

* + /key

→ *Private Key file (should be in .pem format)*

* + /log
* /var
  + /gtas-service-logs
    - /delete
    - /temp
      * + /apache-tomcat
        + /apache-tomcat-web
        + /apache-tomcat-catalina
    - /repository
      * + /apache-tomcat
        + /apache-tomcat-web
* NOTE. After creating the above directory structure and copy all the configuration files, ETL files, the scheduler jar file, and the private key, make sure that the user on this server (Web Application Server) has the right privileges on the directory structure and its contents
* *sudo chmod –R 755 /gtas-log-etl*
* *sudo chmod –R 755 /var/gtas-service-logs*
* *sudo chown –R <username> /gtas-log-etl*
* *sudo chown –R <username> /var/gtas-service-logs*
* *sudo chmod +x /gtas-log-etl/gtas-log-collector-sch.jar*

1. **Configure the ETL Job**

**5.1 application.properties**

This file is used by the GTAS Log Collector Scheduler application (gtas-log-collector-sch.jar).

**Note:** No change is required on this file if the default directory structure is used.

|  |  |  |
| --- | --- | --- |
| **Property** | **Value** | **Description** |
| opSystem | linux | The operating system on which the job runs on |
| pdiDir | /opt/pentaho/data-integration/./kitchen.sh | The Pentaho Data Integration launcher file |
| jobDir | /gtas-log-etl/job/gtas-log-collector.kjb | The gtas-log-collector job starter file |
| logLevel | Basic | The log level for the ETL job |
| logDir | /gtas-log-etl/log/gtas-log-collector-job | The path to the folder in which ETL logs are written |
| configFilePropertyName | EXT\_ETL\_CONFIG\_FILE | The property name for the ETL Configuration file |
| configFile | /gtas-log-etl/config/gtas-log-etl-config.properties | The ETL Configuration file. |

* 1. **run-record.properties**

The ETL job updates this file once per day if the option to delete old log files is enabled. If the option to delete old log files is enabled, the ETL file will update the below value to the current date on the first successful run.

**Note:** No change is required on this configuration file.

|  |  |  |
| --- | --- | --- |
| **Property** | **Value** | **Description** |
| DELETE\_JOB\_LAST\_RUN | 2019-01-01 | The last execution date of the Delete Job in YYYY-MM-DD format. The default value for first time deployment is 2019-01-01 or can be any dates. |

* 1. **gtas-log-etl-config.properties**

|  |  |  |
| --- | --- | --- |
| **Property** | **Value** | **Description** |
| **#GENERAL** |  |  |
| EXT\_VAR\_DATA\_LOADER\_NAME | GTAS-LOG-COLLECTOR-ETL | The name of the GTAS Log Collector ETL |
| **#SFTP** |  |  |
| EXT\_VAR\_REMOTE\_DATA\_PROC\_SERVER |  | The Private IP address of the Data Processing Server |
| EXT\_VAR\_REMOTE\_DATA\_PROC\_SERVER\_PORT | 22 | The SFTP port number |
| EXT\_VAR\_REMOTE\_DATA\_PROC\_SERVER\_USER\_NAME |  | The username of the user of the Data Processor Server |
| EXT\_VAR\_REMOTE\_DATA\_PROC\_SERVER\_USER\_PASSWORD |  | The user password for the username. |
| EXT\_VAR\_REMOTE\_DATA\_PROC\_SERVER\_PRIVATE\_KEY |  | The full file name of the Private Key |
| EXT\_VAR\_REMOTE\_DATA\_PROC\_SERVER\_PASSPHRASE |  | The passphrase of the private key |
| EXT\_VAR\_REMOTE\_USE\_KEY | Y or N | Y for Yes, N for No to enable or disable use of SFTP key |
| **#APACHE TOMCAT** |  |  |
| EXT\_VAR\_TARGET\_TOMCAT\_LOG\_DIR |  | The final repository for the Apache Tomcat zipped log files |
| EXT\_VAR\_TARGET\_TOMCAT\_LOG\_FILE\_NAME |  | The final full log file name of the Apache Tomcat current log |
| EXT\_VAR\_TEMP\_TOMCAT\_DIR |  | The temporary working directory where Apache Tomcat Files are processed |
| EXT\_VAR\_REMOTE\_TOMCAT\_LOG\_DIR |  | The location of the Apache Tomcat Log Files on the remote server |
| EXT\_VAR\_TOMCAT\_CURRENT\_LOG\_WC | gtas.log | The wildcard string that the ETL uses to read current Apache Tomcat log files |
| EXT\_VAR\_TOMCAT\_ROTATED\_LOG\_WC | ^gtas[.][0-9.-]\*.log | The wildcard string that the ETL uses to read rotated Apache Tomcat log files |
| EXT\_VAR\_TOMCAT\_TEMP\_FILE\_NAME\_WC | gtas.\* | The wildcard string that the ETL uses to read Apache Tomcat Log files from the temporary working directory |
| EXT\_VAR\_APACHE\_TOMCAT\_TEMP\_EXCLD\_FILE\_NAME\_WC | .\*\.zip | The wildcard string that ETL uses to read Apache Tomcat zipped log files |
| **#APACHE TOMCAT WEB** |  |  |
| EXT\_VAR\_TARGET\_TOMCAT\_WEB\_LOG\_DIR |  | The final repository for the Apache Tomcat web zipped log files |
| EXT\_VAR\_TARGET\_TOMCAT\_WEB\_LOG\_FILE\_NAME |  | The final full log file name of the Apache Tomcat web current log |
| EXT\_VAR\_TEMP\_TOMCAT\_WEB\_LOG\_DIR |  | The temporary working directory where Apache Tomcat web Files are processed |
| EXT\_VAR\_LOCAL\_TOMCAT\_WEB\_LOG\_DIR |  | The location of the Apache Tomcat web Log Files on the local server |
| EXT\_VAR\_TOMCAT\_WEB\_CURRENT\_LOG\_WC | gtas.log | The wildcard string that the ETL uses to read current Apache Tomcat web log files |
| EXT\_VAR\_TOMCAT\_WEB\_ROTATED\_LOG\_WC | ^gtas[.][0-9.-]\*.log | The wildcard string that the ETL uses to read rotated Apache Tomcat web log files |
| EXT\_VAR\_TOMCAT\_WEB\_TEMP\_FILE\_NAME\_WC | gtas.\* | The wildcard string that the ETL uses to read Apache Tomcat web Log files from the temporary working directory |
| EXT\_VAR\_TOMCAT\_WEB\_TEMP\_EXCLD\_FILE\_NAME\_WC | .\*\.zip | The wildcard string that ETL uses to read Apache Tomcat web zipped log files |
| **#DELETE OLD LOGS** |  |  |
| EXT\_VAR\_DELETE\_DIR | /var/gtas-service-logs/delete | The folder in which files to be delete will be stored |
| EXT\_VAR\_DELETE\_OLD\_LOGS | Y | Delete old logs?  Y=yes, N= no |
| EXT\_VAR\_DELETE\_BEFORE\_DAYS | 10 | Delete logs older than X number of days |
| EXT\_VAR\_DELETE\_WC | .\*\.zip | The wildcard string that ETL uses to delete all logs in the delete folder |
| EXT\_VAR\_DELETE\_JOB\_RECORD | /gtas-log-etl/config/run-record.properties | The properties file that the ETL uses to track the last execution of the delete job. |

1. **Configure the Scheduler as a service**
2. Execute the following commands:

* *touch /etc/systemd/system/log-collector-sch.service*
* *chmod 664 /etc/systemd/system/ log-collector-sch.service*
* *sudo systemctl enable /etc/systemd/system/ log-collector-sch.service*
* *sudo systemctl daemon-reload*

1. **Run the ETL job**

* To start the service: *sudo systemctl start log-collector-sch.service*
* Check the status of the service*: sudo systemctl status log-collector-sch.service*
* To stop the service: *sudo systemctl stop log-collector-sch.service*

The service should start and stop without errors