

# Linux Data Warehouse Documentation 01

Dear All

Steps and Links needed to maintain already existing ETL structure:

## URL Links (Bookmark and use Chrome while in ROWAD Domain or FortiClient)

- Airflow : <http://10.10.11.242:8080/> - user: Nada - password: Nada123
- Spark : <http://10.10.11.242:8081/>

## Remote Desktop (while in ROWAD Domain or FortiClient)

- Windows Server : 10.10.11.241 – user: Rowad\Mohamed GHassan – password : Rowad1234 (use normal remote desktop connection)
- Linux Server : 10.10.11.242 – user: PMO – password: PMO@1234 (use attached file “linux.rdp”)

## Re Initiate Airflow/Spark in case they’re down

- Run attached file “05.startup.airflow.spark.rev05.ipynb” – (need pip install paramiko)

## Add New Query as DAG

- Use GUI Xrdp to access linux machine (use attached file “linux.rdp”)
- Go to path home\airflow\dags – copy paste an existing py file, right click text editor, rename all previous report names then replace existing query with the new one

- Note: the schedule for the new query will be written in crone , for reference <https://crontab.guru/>
- Open Airflow to confirm

Also attached are the current dags in rar “00.main.py.dags.rar”

Also attached a word file “00.if.needed.docx” that will be helpful later which contains more data about linux terminal commands and solutions to some possible issues

Thank you