

Требования:

- sap sdk
- pyRFC
- pySpark
- impala jdbc, версия 4

Общие сведения:

- Место расположения:
/opt/Erp-Hadoop_integration
- Виртуальная среда для запуска:
/opt/Erp-Hadoop_integration/virtual
- Работа с приложением осуществляется через настроечные скрипты. Предполагается, что один скрипт – одна интеграция.
- Запуск скрипта выполняется через настроенную виртуальную среду:
/opt/Erp-Hadoop_integration/virtual/bin/python /opt/Erp-Hadoop_integration/[имя_скрипта]

Пример:

Пример создания новой интеграции (доступен по пути: /opt/Erp-Hadoop_integration/
ZFM_GET_2LIS_17_I3HDR.py)

Вызов SAP ФМ "ZFM_GET_2LIS_17_I3HDR" и запись полученных данных в бд Impala, таблица "dwh_dm.zfm_get_2lis_17_i3hdr_parquet":

```
from Extractors.RFC import SapExtractor
from Loaders.Hadoop import HadoopLoader
from Integration.Integration import DataIntegration
from configs.Configs import Configuration

#
# Options to extract data from sap erp
#

# {Function module name, {FM arguments}}
extract_options = {"fm": "ZFM_GET_2LIS_17_I3HDR",
                  "args":
                    {

                    }

                  }

#
# Options to load data into Impala db
#

#{ "projection": {[erp_table]: [impala_table]}, ...}
load_options = {"projection":
                {
                    "E_TABLE": "dwh_dm.zfm_get_2lis_17_i3hdr_parquet"
                },
                "format": "parquet",
                "compress": "gzip"}

#
# Create extractor and loader for Data Integration
#

# create sap rfc extractor
```

```
extractor_rfc = SapExtractor(Configuration.GetConfig("sapnwrfc.cfg", "connection"),
extract_options)
# create impala loader
loader_hadoop = HadoopLoader(Configuration.GetConfig("impala.cfg", "connection"),
load_options)

#
# Write data
#

# Create integration
data_integration = DataIntegration(extractor_rfc, loader_hadoop)
# Write data into impala
data_integration.Write()
```