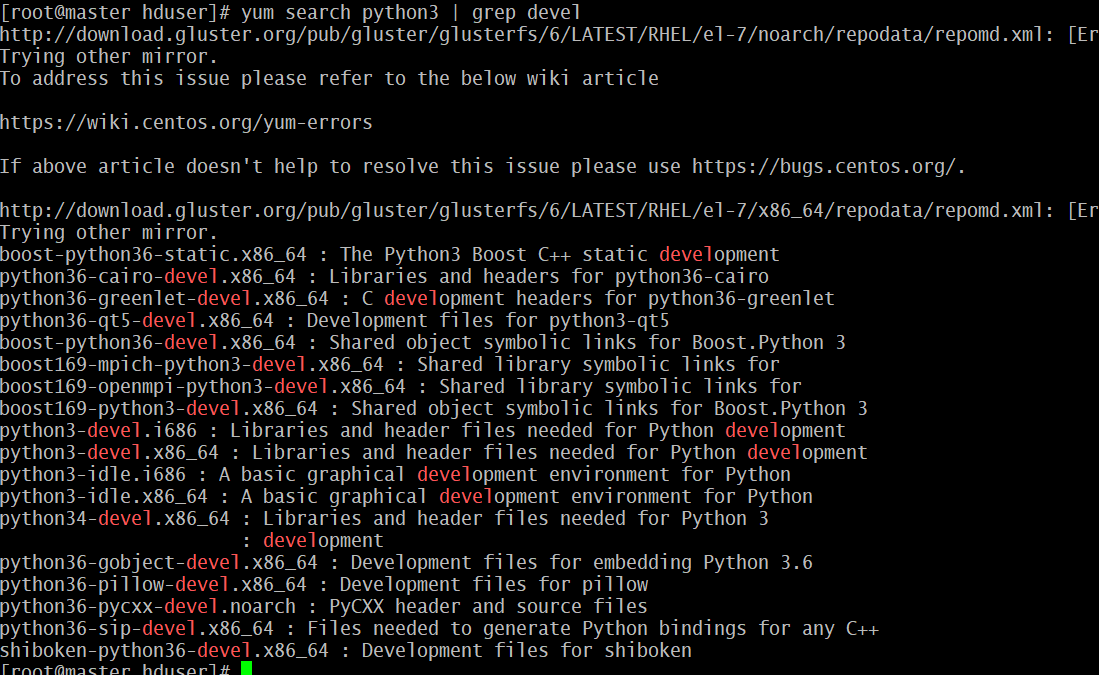
# Setting up the Python environment for IBM database servers on linux

[root@master hduser]# yum search python3 | grep devel



[root@master hduser]# yum install -y python3-devel.x86\_64

[root@master hduser]# pip install ibm\_db

[root@master hduser]# pip install ibm\_db\_sa

Successfully installed ibm\_db and ibm\_db\_sa

Now read and write tables from remote IBM Db2 databases.

Reading

#get the connection to read the database tables from DB2

conn = ibm\_db.connect("AUTHENTICATION=SERVER;DATABASE=database\_name;HOSTNAME=ipaddress;PORT=50000;PROTOCOL=TCPIP;UID=username;PWD=password;", "", "")

#example

#conn = ibm\_db.connect("AUTHENTICATION=SERVER;DATABASE=books;HOSTNAME=12.157.80.120;PORT=50000;PROTOCOL=TCPIP;UID=db2admin;PWD=123456;", "", "")

conn2 = ibm\_db\_dbi.Connection(conn)

#read data as pandas dataframe

df2 = pd.read\_sql("select \* from database.table", conn2)

writing

#select column names

subset = df[['COMM\_ADDR','COMM\_CITY','COMM\_PIN','CUR\_ADDR','CUR\_CITY','CUR\_PIN','data\_availability','data\_missing','DOB','EMAIL\_ID','FATHER\_NAME','FEED\_SYS','GRP\_NO','INSERT\_TIME','LOCATION\_CODE','MM\_YEAR','MOBILE1','MOBILE\_2','Month','MONTH\_YEAR','MONTY\_YEAR','MOTHER\_NAME','NAME','OFF\_ADDR','OFF\_CITY','OFF\_PIN','over\_all\_addr\_available','over\_all\_city\_available','over\_all\_phone\_available','over\_all\_pin\_available','PAN','PHONE1','PHONE2','PHONE3','PSX\_ID\_NEW','Quarter','SL','SOURCE','SOURCE\_NEW','SPOUSE\_NAME','Year','NAME\_VALID\_INVALID','NAME\_TOKENS','Name\_Length','FATHER\_NAME\_VALID\_INVALID','FATHER\_NAME\_TOKENS','FATHER\_NAME\_Length','SPOUSE\_NAMEE\_VALID\_INVALID','SPOUSE\_NAME\_TOKENS','SPOUSE\_NAME\_Length','MOTHER\_NAME\_VALID\_INVALID','MOTHER\_NAME\_TOKENS','MOTHER\_NAME\_Length','AGE','DOB\_VALID\_INVALID','PAN\_VALID\_INVALID','PAN\_Length','CUR\_ADDR\_VALID\_INVALID','CUR\_ADDR\_TOKENS','CURR\_Addr\_Length','COMM\_ADDR\_VALID\_INVALID','COMM\_ADDR\_TOKENS','COMM\_Addr\_Length','OFF\_ADDR\_VALID\_INVALID','OFF\_ADDR\_TOKENS','OFF\_Addr\_Length','CUR\_CITY\_VALID\_INVALID','CUR\_CITY\_TOKENS','cur\_city\_Length','COMM\_CITY\_VALID\_INVALID','COMM\_CITY\_TOKENS','cOMM\_city\_Length','OFF\_CITY\_VALID\_INVALID','OFF\_CITY\_TOKENS','OFF\_city\_Length','CUR\_PIN\_INVALID\_VALID','CUR\_PIN\_TOKENS','COMM\_PIN\_INVALID\_VALID','COMM\_PIN\_TOKENS','OFF\_PIN\_INVALID\_VALID','OFF\_PIN\_TOKENS','PHONE\_1\_VALID\_INVALID','phone1\_tokens','phone1\_Length','PHONE\_2\_VALID\_INVALID','phone2\_tokens','phone2\_Length','PHONE\_3\_VALID\_INVALID','phone3\_tokens','phone3\_Length','MOBILE1\_VALID\_INVALID','MOBILE1\_tokens','MOBILE1\_Length','MOBILE\_2\_VALID\_INVALID','MOBILE\_2\_tokens','MOBILE\_2\_Length','EMAIL\_VALID\_INVALID','EMAIL\_TOKENS','EMAIL\_LENGTH']]

tuple\_of\_tuples = tuple([tuple(x) for x in subset.values])

#query maintain the number question marks same as number of column names

sql = "INSERT INTO database.table VALUES(?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?,?)"

#get the connection

cnn = ibm\_db.connect("AUTHENTICATION=SERVER;DATABASE=database\_name;HOSTNAME=ipaddress;PORT=50000;PROTOCOL=TCPIP;UID=username;PWD=password;", "", "")

# example

# cnn = ibm\_db.connect("AUTHENTICATION=SERVER;DATABASE=books;HOSTNAME=12.157.80.120;PORT=50000;PROTOCOL=TCPIP;UID=db2admin;PWD=123456;", "", "")

stmt = ibm\_db.prepare(cnn, sql)

ibm\_db.execute\_many(stmt, tuple\_of\_tuples)