DataPlatform cluster configuration for EDG with Apache Zookeeper

The following supplements the failover guidance for EDG with Data Platform. Please read the Data Platform guidance and familiarize yourself with the technical details of the Data Coordinator. TopQuadrant created this detailed playbook for customer's on premise IT department to get up and running with a cluster of Data Coordinator servers for EDG. The Data Platform documentation contained in the distribution has its own guidelines but this is specific for EDG.

for each DataPlatform instance:

unzip rdf-delta-dist-1.1.0.zip
cd rdf-delta-1.1.0
cp -a Tutorial/zk-example zk

#define members in zoo.dynamic:
zk/zoo.dynamic
server.1=dp1:2181:2184:participant;2281
server.2=dp2:2182:2185:participant;2282
server.3=dp3:2183:2186:participant;2283
#where dp1,dp2,dp3 are short hostnames. arbitrary.
#ports that matter here are 2281,2282,2283
#make sure each node can connect to other nodes of the defined ports

#define zoo.cfg
zk/zoo.cfg
admin.enableServer=false
dataDir=/opt/zkData/
initLimit=5
tickTime=2000
syncLimit=2

#define id's
Node1:
echo 1 > zk/myid
Node2:
echo 2 > zk/myid
Node3:
echo 3 > zk/myid

start each instance of DataPlatform *ensure java11 is 'java' binary running

cd zk

```
java -jar ../delta-server.jar --port 1071 --zk=dp1:2281,dp2:2282,dp3:2283 --zkCfg=zoo.cfg
[2022-07-05 19:38:08] PatchStoreZk INFO  Connecting to zookeeper
[2022-07-05 19:38:09] Delta     INFO  Provider: zk
[2022-07-05 19:38:09] Config    INFO  Delta Server port=1071
[2022-07-05 19:38:09] Learner   WARN  Got zxid 0x7000002e1 expected 0x1
[2022-07-05 19:38:09] Delta     INFO  RDF Delta 1.1.0 2021-12-19T17:55:49+0000
[2022-07-05 19:38:09] Delta     INFO   Data source: [id:259956, dp1, <urn:x-e…
```

Configure EDG to use a DataPlatform instance from the cluster

Preferably to use a LoadBalanced IP of the DP Cluster, but can use a single DP instance URL

edg-setup.properties: (this can be set using UI setup wizard or file directly)

```
workspacePath=/var/lib/topbraid/workspace
securityRoles=,admin
vaultPath=/var/lib/topbraid/vault
isPrimaryNode=true
dpZone=Zone
setupComplete=true
dpServerURL=http\://dp1\:1071/
databaseType=DataPlatform
vaultPassword=Password
sessionTimeout=90
authMethod=form
```

Now changes you do on EDG will propagate to all the members of the DP/ZK Cluster and all EDG instances connected to the cluster as above will also be made aware of all changes from the cluster.

Can verify this by pointing a second EDG to another DP node directly and see changes reflected.

*Note: DP stdout console logging will not reflect sync changes. This is because the console can only capture what directly goes into the DP instance directly, and does not display the sync messages behind the zookeeper cluster.