

Milestone 2 Management of Data

WQD7005 – DATA MINING

ZULKANAIN BIN HASAN

WQD180031

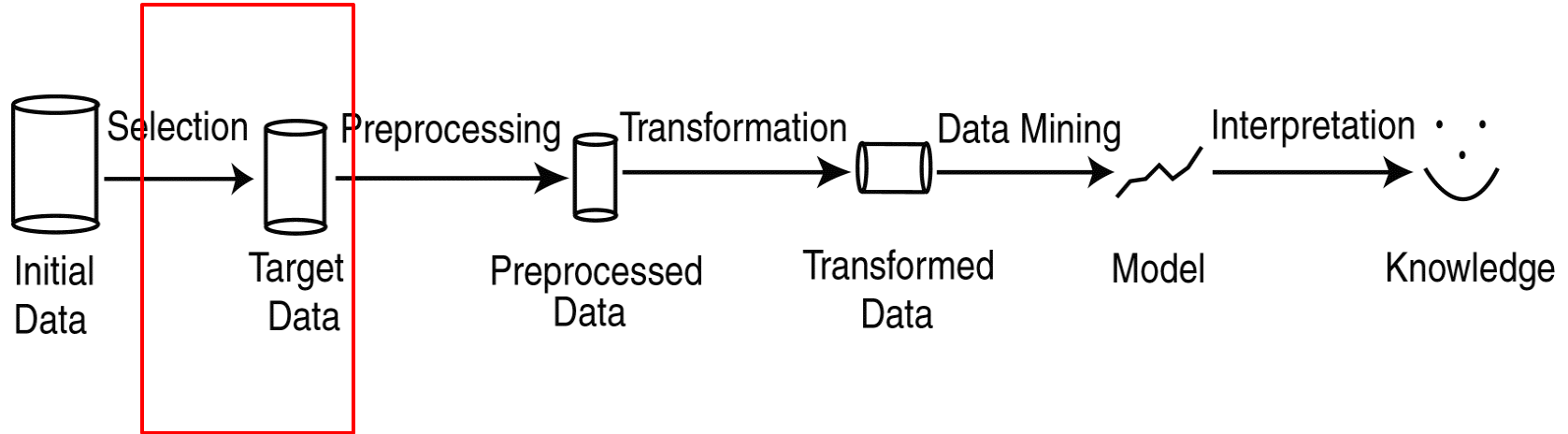
NUR ATHIRAH BT. RAZAK

WQD170072

Knowledge Discovery in Databases (KDD): process of finding useful information and patterns in data:

1. Selection (Pre-Mining 1): Obtain data from various sources

Milestone 2



```
File Edit View Search Terminal Help
(base) zulkanh@elitebook:~$ start-all.sh
This script is Deprecated. Instead use start-dfs.sh and start-yarn.sh
Starting namenodes on [localhost]
zulkanh@localhost's password:
localhost: starting namenode, logging to /home/zulkanh/bigdata/hadoop/logs/hadoop-p-zulkanh-namenode-elitebook.out
zulkanh@localhost's password:
localhost: starting datanode, logging to /home/zulkanh/bigdata/hadoop/logs/hadoop-p-zulkanh-datanode-elitebook.out
Starting secondary namenodes [0.0.0.0]
zulkanh@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /home/zulkanh/bigdata/hadoop/logs/hadoop-s-hadoop-zulkanh-secondarynamenode-elitebook.out
starting yarn daemons
starting resourcemanager, logging to /home/zulkanh/bigdata/hadoop/logs/yarn-zulkanh-resourcemanager-elitebook.out
zulkanh@localhost's password:
localhost: starting nodemanager, logging to /home/zulkanh/bigdata/hadoop/logs/yarn-zulkanh-nodemanager-elitebook.out
(base) zulkanh@elitebook:~$ jps
15889 DataNode
15729 NameNode
16083 SecondaryNameNode
16246 ResourceManager
```

```
File Edit View Search Terminal Help
localhost: starting datanode, logging to /home/zulkanh/bigdata/hadoop/logs/hadoop-zulkanh-datanode-elitebook.out
Starting secondary namenodes [0.0.0.0]
zulkanh@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /home/zulkanh/bigdata/hadoop/logs/hadoop-zulkanh-secondarynamenode-elitebook.out
starting yarn daemons
starting resourcemanager, logging to /home/zulkanh/bigdata/hadoop/logs/yarn-zulkanh-resourcemanager-elitebook.out
zulkanh@localhost's password:
localhost: starting nodemanager, logging to /home/zulkanh/bigdata/hadoop/logs/yarn-zulkanh-nodemanager-elitebook.out
(base) zulkanh@elitebook:~$ jps
15889 DataNode
15729 NameNode
16083 SecondaryNameNode
16246 ResourceManager
16872 Jps
16413 NodeManager
(base) zulkanh@elitebook:~$ hive
Logging initialized using configuration in jar:file:/home/zulkanh/bigdata/hive/lib/hive-common-1.2.2.jar!/hive-log4j.properties
hive>
```

```
File Edit View Search Terminal Help
localhost: starting datanode, logging to /home/zulkanh/bigdata/hadoop/logs/hadoop-zulkanh-datanode-elitebook.out
Starting secondary namenodes [0.0.0.0]
zulkanh@0.0.0.0's password:
0.0.0.0: starting secondarynamenode, logging to /home/zulkanh/bigdata/hadoop/logs/hadoop-zulkanh-secondarynamenode-elitebook.out
starting yarn daemons
starting resourcemanager, logging to /home/zulkanh/bigdata/hadoop/logs/yarn-zulkanh-resourcemanager-elitebook.out
zulkanh@localhost's password:
localhost: starting nodemanager, logging to /home/zulkanh/bigdata/hadoop/logs/yarn-zulkanh-nodemanager-elitebook.out
(base) zulkanh@elitebook:~$ jps
15889 DataNode
15729 NameNode
16083 SecondaryNameNode
16246 ResourceManager
16872 Jps
16413 NodeManager
(base) zulkanh@elitebook:~$ hive

Logging initialized using configuration in jar:file:/home/zulkanh/bigdata/hive/lib/hive-common-1.2.2.jar!/hive-log4j.properties
hive> 
```

```
File Edit View Search Terminal Help
starting yarn daemons
starting resourcemanager, logging to /home/zulkanh/bigdata/hadoop/logs/yarn-zulka
anh-resourcemanager-elitebook.out
zulkanh@localhost's password:
localhost: starting nodemanager, logging to /home/zulkanh/bigdata/hadoop/logs/ya
rn-zulkanh-nodemanager-elitebook.out
(base) zulkanh@elitebook:~$ jps
15889 DataNode
15729 NameNode
16083 SecondaryNameNode
16246 ResourceManager
16872 Jps
16413 NodeManager
(base) zulkanh@elitebook:~$ hive

Logging initialized using configuration in jar:file:/home/zulkanh/bigdata/hive/l
ib/hive-common-1.2.2.jar!/hive-log4j.properties
hive> show databases;
OK
default
wqd7005
wqd7007
Time taken: 0.836 seconds, Fetched: 3 row(s)
hive> 
```

```
File Edit View Search Terminal Help
zulkanh@localhost's password:
localhost: starting nodemanager, logging to /home/zulkanh/bigdata/hadoop/logs/yarn-zulkanh-nodemanager-elitebook.out
(base) zulkanh@elitebook:~$ jps
4212 ResourceManager
4565 NodeManager
3609 NameNode
4666 Jps
4043 SecondaryNameNode
3804 DataNode
(base) zulkanh@elitebook:~$ hive

Logging initialized using configuration in jar:file:/home/zulkanh/bigdata/hive/lib/hive-common-1.2.2.jar!/hive-log4j.properties
hive> show databases;
OK
default
wqd7005
wqd7007
Time taken: 1.172 seconds, Fetched: 3 row(s)
hive> use wqd7005;
OK
Time taken: 0.048 seconds
hive> 
```

```
File Edit View Search Terminal Help
```

```
4565 NodeManager
```

```
3609 NameNode
```

```
4666 Jps
```

```
4043 SecondaryNameNode
```

```
3804 DataNode
```

```
(base) zulkanh@elitebook:~$ hive
```

```
Logging initialized using configuration in jar:file:/home/zulkanh/bigdata/hive/lib/hive-common-1.2.2.jar!/hive-log4j.properties
```

```
hive> show databases;
```

```
OK
```

```
default
```

```
wqd7005
```

```
wqd7007
```

```
Time taken: 1.172 seconds, Fetched: 3 row(s)
```

```
hive> use wqd7005;
```

```
OK
```

```
Time taken: 0.048 seconds
```

```
hive> CREATE EXTERNAL TABLE IF NOT EXISTS 0056kl (c1 DATE, high DOUBLE, low DOUBLE, open DOUBLE, close DOUBLE, volume INT, adjclose DOUBLE) ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ',';
```

```
OK
```

```
Time taken: 0.688 seconds
```

```
hive> █
```



```
File Edit View Search Terminal Help
Logging initialized using configuration in jar:file:/home/zulkanh/bigdata/hive/lib/hive-common-1.2.2.jar!/hive-log4j.properties
hive> show databases;
OK
default
wqd7005
wqd7007
Time taken: 1.172 seconds, Fetched: 3 row(s)
hive> use wqd7005;
OK
Time taken: 0.048 seconds
hive> CREATE EXTERNAL TABLE IF NOT EXISTS 0056kl (c1 DATE, high DOUBLE, low DOUBLE, open DOUBLE, close DOUBLE, volume INT, adjclose DOUBLE) ROW FORMAT DELIMITED
  FIELDS TERMINATED BY ',';
OK
Time taken: 0.688 seconds
hive> LOAD DATA LOCAL INPATH '/home/zulkanh/Desktop/0056KL.csv' OVERWRITE INTO TABLE 0056kl;
Loading data to table wqd7005.0056kl
Table wqd7005.0056kl stats: [numFiles=1, numRows=0, totalSize=223983, rawDataSize=0]
OK
Time taken: 1.643 seconds
hive> 
```

File Edit View Search Terminal Help

hive> SELECT * FROM 0056kl LIMIT 10;

OK

2010-01-04	0.119319997727871	0.075000002980232	0.11931999772787
1	0.075000002980232	0	0.061856783926487
2010-01-05	0.119319997727871	0.075000002980232	0.11591000109911
0.11591000109911	61600	0.095597594976425	
2010-01-06	0.119319997727871	0.109090000391006	0.11249999701976
8	0.119319997727871	237601	0.09841001033783
2010-01-07	0.121019996702671	0.112499997019768	0.11249999701976
8	0.121019996702671	202400	0.099812090396881
2010-01-08	0.122730001807213	0.112499997019768	0.11249999701976
8	0.122730001807213	143440	0.101222433149815
2010-01-11	0.126139998435974	0.119319997727871	0.11931999772787
1	0.126139998435974	108534	0.1040348559618
2010-01-12	0.136360004544258	0.127839997410774	0.12783999741077
4	0.129549995064735	161334	0.106847271323204
2010-01-13	0.136360004544258	0.122730001807213	0.12443000078201
3	0.136360004544258	308001	0.112463869154453
2010-01-14	0.136360004544258	0.131249994039536	0.13636000454425
8	0.136360004544258	776455	0.112463869154453
2010-01-15	0.136360004544258	0.129549995064735	0.13636000454425
8	0.129549995064735	32267	0.106847271323204

Time taken: 0.413 seconds, Fetched: 10 row(s)

hive> █