HBase and Hive integration use case

HBase implementation

Start HBase

start-hbase.sh

hbase shell

create namespace and table

create_namespace 'bigdata'

create 'bigdata:AnalysisLog_ID', 'LogInfo'

```
hbase(main):001:0>
hbase(main):001:0> create_namespace 'bigdata'
Took 1.6625 seconds
hbase(main):002:0> create 'bigdata:AnalysisLog_ID','LogInfo'
Created table bigdata:AnalysisLog_ID
Took 0.7793 seconds
=> Hbase::Table - bigdata:AnalysisLog_ID
```

put data

put 'bigdata:AnalysisLog_ID', '001', 'LogInfo:hostname', 'FI01Miner01'

put 'bigdata:AnalysisLog_ID', '001', 'LogInfo:service', 'miner'

put 'bigdata:AnalysisLog_ID', '001', 'LogInfo:type', 'info'

put 'bigdata:AnalysisLog_ID', '002', 'LogInfo:hostname', 'FI01Miner02'

put 'bigdata:AnalysisLog_ID', '002', 'LogInfo:service', 'zookeeper'

put 'bigdata:AnalysisLog_ID', '002', 'LogInfo:type', 'info'

put 'bigdata:AnalysisLog_ID', '003', 'LogInfo:hostname', 'FI01Miner01'

put 'bigdata:AnalysisLog_ID', '003', 'LogInfo:service', 'miner'

```
put 'bigdata:AnalysisLog_ID', '003', 'LogInfo:type', 'error'


put 'bigdata:AnalysisLog_ID', '004', 'LogInfo:hostname', 'FI01Miner02'
put 'bigdata:AnalysisLog_ID', '004', 'LogInfo:service', 'miner'
put 'bigdata:AnalysisLog_ID', '004', 'LogInfo:type', 'error'


put 'bigdata:AnalysisLog_ID', '005', 'LogInfo:hostname', 'FI01Miner02'
put 'bigdata:AnalysisLog_ID', '005', 'LogInfo:service', 'spark'
put 'bigdata:AnalysisLog_ID', '005', 'LogInfo:type', 'info'


put 'bigdata:AnalysisLog_ID', '006', 'LogInfo:hostname', 'FI01Miner01'
put 'bigdata:AnalysisLog_ID', '006', 'LogInfo:service', 'zookeeper'
put 'bigdata:AnalysisLog_ID', '006', 'LogInfo:type', 'debug'


put 'bigdata:AnalysisLog_ID', '007', 'LogInfo:hostname', 'FI01Miner03'
put 'bigdata:AnalysisLog_ID', '007', 'LogInfo:service', 'zookeeper'
put 'bigdata:AnalysisLog_ID', '007', 'LogInfo:type', 'debug'


put 'bigdata:AnalysisLog_ID', '008', 'LogInfo:hostname', 'FI01Miner02'
put 'bigdata:AnalysisLog_ID', '008', 'LogInfo:service', 'miner'
put 'bigdata:AnalysisLog_ID', '008', 'LogInfo:type', 'info'


put 'bigdata:AnalysisLog_ID', '009', 'LogInfo:hostname', 'FI01Miner01'
put 'bigdata:AnalysisLog_ID', '009', 'LogInfo:service', 'miner'
put 'bigdata:AnalysisLog_ID', '009', 'LogInfo:type', 'info'
```

```
Took 0.0067 seconds
hbase(main):040:0> put 'bigdata:AnalysisLog_ID', '009', 'LogInfo:type', 'info'
Took 0.0166 seconds
hbase(main):041:0> scan 'bigdata:AnalysisLog_ID'
ROW                              COLUMN+CELL
 001                             column=LogInfo:hostname, timestamp=2025-05-09T14:14:25.661, value=FI01Miner01
 001                             column=LogInfo:service, timestamp=2025-05-09T14:16:09.295, value=miner
 001                             column=LogInfo:type, timestamp=2025-05-09T14:17:43.836, value=info
 002                             column=LogInfo:hostname, timestamp=2025-05-09T14:17:58.798, value=FI01Miner02
 002                             column=LogInfo:service, timestamp=2025-05-09T14:17:58.836, value=zookeeper
 002                             column=LogInfo:type, timestamp=2025-05-09T14:18:03.465, value=info
 003                             column=LogInfo:hostname, timestamp=2025-05-09T14:18:18.735, value=FI01Miner01
 003                             column=LogInfo:service, timestamp=2025-05-09T14:18:18.769, value=miner
 003                             column=LogInfo:type, timestamp=2025-05-09T14:18:19.800, value=error
 004                             column=LogInfo:hostname, timestamp=2025-05-09T14:18:32.894, value=FI01Miner02
 004                             column=LogInfo:service, timestamp=2025-05-09T14:18:32.933, value=miner
 004                             column=LogInfo:type, timestamp=2025-05-09T14:18:34.083, value=error
 005                             column=LogInfo:hostname, timestamp=2025-05-09T14:18:49.693, value=FI01Miner02
 005                             column=LogInfo:service, timestamp=2025-05-09T14:18:49.773, value=spark
 005                             column=LogInfo:type, timestamp=2025-05-09T14:18:50.721, value=info
 006                             column=LogInfo:hostname, timestamp=2025-05-09T14:19:01.944, value=FI01Miner01
 006                             column=LogInfo:service, timestamp=2025-05-09T14:19:02.122, value=zookeeper
 006                             column=LogInfo:type, timestamp=2025-05-09T14:19:03.184, value=debug
 007                             column=LogInfo:hostname, timestamp=2025-05-09T14:19:12.919, value=FI01Miner03
 007                             column=LogInfo:service, timestamp=2025-05-09T14:19:12.993, value=zookeeper
 007                             column=LogInfo:type, timestamp=2025-05-09T14:19:13.945, value=debug
 008                             column=LogInfo:hostname, timestamp=2025-05-09T14:19:23.717, value=FI01Miner02
 008                             column=LogInfo:service, timestamp=2025-05-09T14:19:23.764, value=miner
 008                             column=LogInfo:type, timestamp=2025-05-09T14:19:24.753, value=info
 009                             column=LogInfo:hostname, timestamp=2025-05-09T14:19:36.484, value=FI01Miner01
 009                             column=LogInfo:service, timestamp=2025-05-09T14:19:36.519, value=miner
 009                             column=LogInfo:type, timestamp=2025-05-09T14:19:37.867, value=info
9 row(s)
Took 0.2585 seconds
hbase(main):042:0> █
```

Hive integration

- **Create table in Hive**
-

```
CREATE EXTERNAL TABLE Log_hive_ID(
row_key STRING,
hostname STRING,
service STRING,
type STRING
)
STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
WITH SERDEPROPERTIES (
"hbase.columns.mapping" = ":key,LogInfo:hostname,LogInfo:service,LogInfo:type"
)
TBLPROPERTIES (
"hbase.table.name" = "bigdata:AnalysisLog_ID"
);
```

- **Query using Hive**
```
SELECT * FROM Log_hive_ID WHERE service = 'miner';
```

```
hive> CREATE EXTERNAL TABLE Log_hive_ID(
    > row_key STRING,
    > hostname STRING,
    > service STRING,
    > type STRING
    > )
    > STORED BY 'org.apache.hadoop.hive.hbase.HBaseStorageHandler'
    > WITH SERDEPROPERTIES (
    > "hbase.columns.mapping" = ":key,LogInfo:hostname,LogInfo:service,LogInfo:type"
    > )
    > TBLPROPERTIES (
    > "hbase.table.name" = "bigdata:AnalysisLog_ID"
    > );
OK
Time taken: 1.062 seconds
hive> SELECT * FROM Log_hive_ID WHERE service = 'miner';
OK
001     FI01Miner01     miner   info
003     FI01Miner01     miner   error
004     FI01Miner02     miner   error
008     FI01Miner02     miner   info
009     FI01Miner01     miner   info
Time taken: 2.659 seconds, Fetched: 5 row(s)
hive>
```