



TSMC 教育訓練課程

HBase Programming

< V 0.20 >

王耀聰 陳威宇

Jazz@nchc.org.tw

waue@nchc.org.tw

© TemplatesWise.com



財團法人國家實驗研究院

國家高速網路與計算中心

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING



Outline

- HBase 程式編譯方法
- HBase 程式設計
 - ◆ 常用的HBase API 說明
 - ◆ 實做 I/O 操作
 - ◆ 搭配Map Reduce 運算
- 其他用法補充
- 其他專案



HBase

程式編譯方法

此篇介紹兩種編譯與執行HBase程式的方法：

Method 1 – 使用Java JDK 1.6

Method 2 – 使用Eclipse 套件



財團法人國家實驗研究院

國家高速網路與計算中心

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING



1. Java 之編譯與執行

1. 將hbase_home目錄內的 .jar檔全部拷貝至
hadoop_home/lib/ 資料夾內
2. 編譯
 - ◆ javac Δ -classpath Δ **hadoop-*-core.jar:hbase-*.jar** Δ -d
 Δ **MyJava** Δ MyCode.java
3. 封裝
 - ◆ jar Δ -cvf Δ MyJar.jar Δ -C Δ **MyJava** Δ .
4. 執行
 - ◆ bin/hadoop Δ jar Δ MyJar.jar Δ MyCode Δ {Input/ Δ Output/ }

-
- 所在的執行目錄為Hadoop_Home
 - ./MyJava = 編譯後程式碼目錄
 - Myjar.jar = 封裝後的編譯檔

- 先放些文件檔到HDFS上的input目錄
- ./input; ./ouput 不一定為 hdfs的輸入、輸出目錄

2. Eclipse 之編譯與執行



HBase 程式設計

© TemplatesWise.com

此篇介紹如何撰寫HBase程式

常用的HBase API 說明

實做 I/O 操作

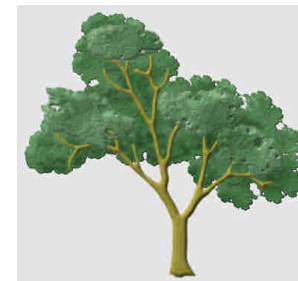
搭配Map Reduce 運算



財團法人國家實驗研究院

國家高速網路與計算中心

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING





HBase 程式設計

常用的HBase API 說明

© TemplatesWise.com



財團法人國家實驗研究院

國家高速網路與計算中心

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING

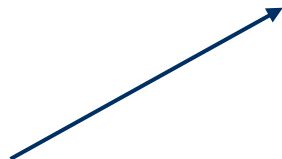


HTable 成員

- Table, Family
- Column, Qualifier
- Row, TimeStamp,
- Cell, Lock

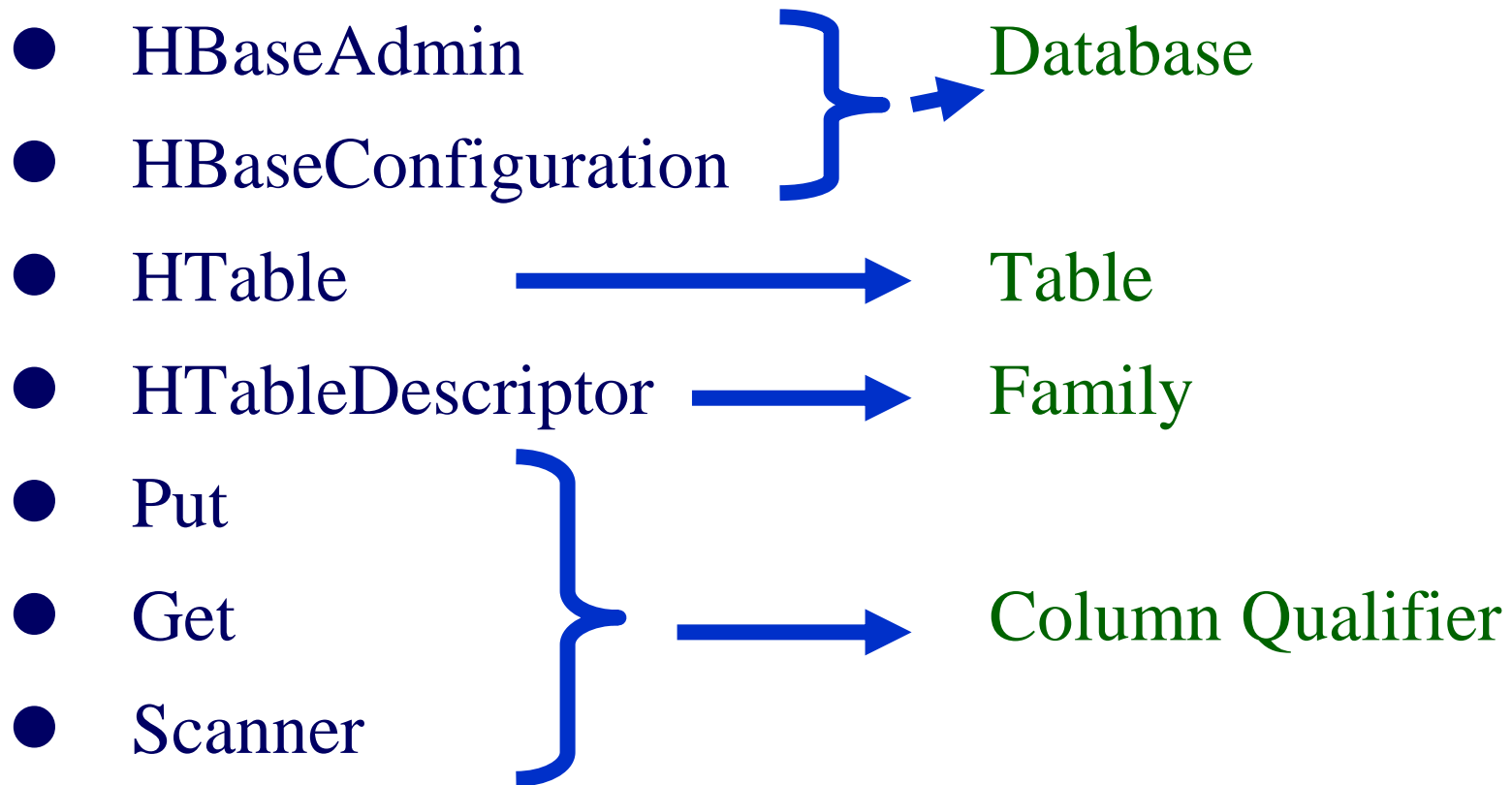
Row Key	Time Stamp	Column (Family) "content:"
com.cnn.www	t9	"<html>..."
	t6	"<html>..."

Row Key	Time Stamp	Column (Family) "anchor:"
com.cnn.www	t9	"anchor:cnnsi.com" "CNN"
	t8	"anchor:cnnsi.com" "CNN"
		"anchor:my.loc" "MyLook"



Row Key	Time Stamp	Column (Family) "content:"	Column (Family) "anchor:"
com.cnn.www	t9	"<html>..."	"anchor:cnnsi.com" "CNN"
	t8		"anchor:cnnsi.com" "CNN"
			"anchor:my.lock.ca" "MyLook"
	t6	"<html>..."	

HBase 常用函式



HBaseConfiguration

- Adds HBase configuration files to a Configuration
 - ◆ = new HBaseConfiguration ()
 - ◆ = new HBaseConfiguration (Configuration c)
- 繼承自
org.apache.hadoop.conf.Configuration

```
<property>  
  <name> name  
</name>  
  <value> value  
</value>  
</property>
```

回傳值	函數	參數
void	addResource	(Path file)
void	clear	()
String	get	(String name)
String	getBoolean	(String name, boolean defaultValue)
void	set	(String name, String value)
void	setBoolean	(String name, boolean value)

HBaseAdmin

- HBase的管理介面
 - ◆ = new HBaseAdmin(HBaseConfiguration conf)
- Ex:

```
HBaseAdmin admin = new HBaseAdmin(config);  
admin.disableTable ("tablename");
```

回傳值	函數	參數
void	addColumn	(String tableName, HColumnDescriptor column)
	checkHBaseAvailable	(HBaseConfiguration conf)
	createTable	(HTableDescriptor desc)
	deleteTable	(byte[] tableName)
	deleteColumn	(String tableName, String columnName)
	enableTable	(byte[] tableName)
	disableTable	(String tableName)
HTableDescriptor[]	listTables	()
void	modifyTable	(byte[] tableName, HTableDescriptor htd)
boolean	tableExists	(String tableName)

HTableDescriptor

- HTableDescriptor contains the name of an HTable, and its column families.
 - ◆ = new HTableDescriptor()
 - ◆ = new HTableDescriptor(String name)
- Constant-values
 - ◆ org.apache.hadoop.hbase.HTableDescriptor.TABLE_DESCRIPTOR_VERSION
- Ex:

```
HTableDescriptor htd = new HTableDescriptor(tablename);  
htd.addFamily ( new HColumnDescriptor (“Family”));
```

回傳值	函數	參數
void	addFamily	(HColumnDescriptor family)
HColumnDescriptor	removeFamily	(byte[] column)
byte[]	getName	() = Table name
byte[]	getValue	(byte[] key) = 對應key的value
void	setValue	(String key, String value)

HColumnDescriptor

- An HColumnDescriptor contains information about a column family
 - ◆ = new HColumnDescriptor(String familyname)
- Constant-values
 - ◆ org.apache.hadoop.hbase.HTableDescriptor.TABLE_DESCRIPTOR_VERSION
- Ex:

```
HTableDescriptor htd = new HTableDescriptor(tablename);  
HColumnDescriptor col = new HColumnDescriptor("content:");  
htd.addFamily(col);
```

回傳值	函數	參數
byte[]	getName	() = Family name
byte[]	getValue	(byte[] key) = 對應key的value
void	setValue	(String key, String value)

HTable

- Used to communicate with a single HBase table.
 - ◆ = new HTable(HBaseConfiguration conf, String tableName)
- Ex:

```
HTable table = new HTable (conf, Bytes.toBytes ( tablename ));  
ResultScanner scanner = table.getScanner ( family );
```

回傳值	函數	參數
void	checkAndPut	(byte[] row, byte[] family, byte[] qualifier, byte[] value, Put put)
void	close	()
boolean	exists	(Get get)
Result	get	(Get get)
byte[][]	getEndKeys	()
ResultScanner	getScanner	(byte[] family)
HTableDescriptor	getTableDescriptor	()
byte[]	getTableName	()
static boolean	isTableEnabled	(HBaseConfiguration conf, String tableName)
void	put	(Put put)

Put

- Used to perform Put operations for a single row.
 - ◆ = new Put(byte[] row)
 - ◆ = new Put(byte[] row, RowLock rowLock)
- Ex:

```
HTable table = new HTable (conf, Bytes.toBytes ( tablename ));  
Put p = new Put ( brow );  
p.add (family, qualifier, value);  
table.put ( p );
```

Put	add	(byte[] family, byte[] qualifier, byte[] value)
Put	add	(byte[] column, long ts, byte[] value)
byte[]	getRow	()
RowLock	getRowLock	()
long	getTimeStamp	()
boolean	isEmpty	()
Put	setTimeStamp	(long timestamp)

Get

- Used to perform Get operations on a single row.
 - ◆ = new Get (byte[] row)
 - ◆ = new Get (byte[] row, RowLock rowLock)
- Ex:

```
HTable table = new HTable(conf, Bytes.toBytes(tablename));  
Get g = new Get(Bytes.toBytes(row));
```

Get	addColumn	(byte[] column)
Get	addColumn	(byte[] family, byte[] qualifier)
Get	addColumnns	(byte[][] columns)
Get	addFamily	(byte[] family)
TimeRange	getTimeRange	()
Get	setTimeRange	(long minStamp, long maxStamp)
Get	setFilter	(Filter filter)

Scanner

- All operations are identical to **Get**
 - ◆ Rather than specifying a single row, an optional startRow and stopRow may be defined.
- If rows are not specified, the Scanner will iterate over all rows.
 - ◆ = new Scan ()
 - ◆ = new Scan (byte[] startRow, byte[] stopRow)
 - ◆ = new Scan (byte[] startRow, Filter filter)

Get	addColumn	(byte[] column)
Get	addColumn	(byte[] family, byte[] qualifier)
Get	addColumnns	(byte[][] columns)
Get	addFamily	(byte[] family)
TimeRange	getTimeRange	()
Get	setTimeRange	(long minStamp, long maxStamp)
Get	setFilter	(Filter filter)

Result

- Single row result of a Get or Scan query.
 - ◆ = new Result()
- Ex:

```
HTable table = new HTable(conf, Bytes.toBytes(tablename));
Get g = new Get(Bytes.toBytes(row));
Result rowResult = table.get(g);
Bytes[] ret = rowResult.getValue( (family + ":" + column) );
```

boolean	containsColumn	(byte[] family, byte[] qualifier)
NavigableMap <byte[],byte[]>	getFamilyMap	(byte[] family)
byte[]	getValue	(byte[] column)
byte[]	getValue	(byte[] family, byte[] qualifier)
int	Size	()

Interface ResultScanner

- Interface for client-side scanning. Go to HTable to obtain instances.
- ◆ `HTable.getScanner (Bytes.toBytes(family));`
- Ex:

```
ResultScanner scanner = table.getScanner (Bytes.toBytes(family));  
for (Result rowResult : scanner) {  
    Bytes[] str = rowResult.getValue ( family , column );  
}
```

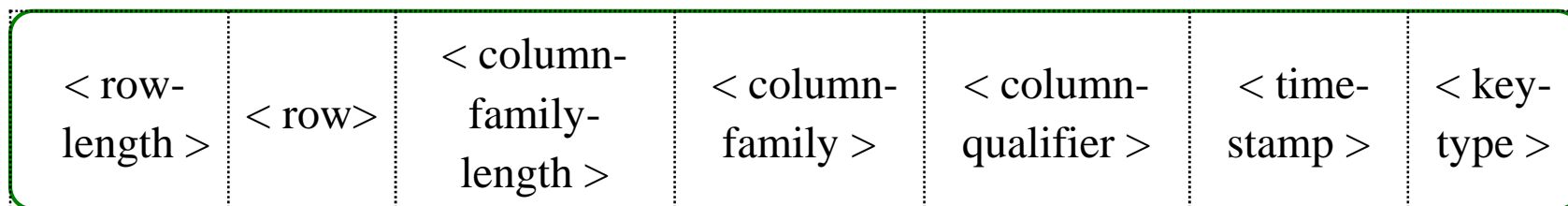
void	close	()
Result	next	()

HBase Key/Value 的格式

- org.apache.hadoop.hbase.KeyValue
- getRow(), getFamily(), getQualifier(), getTimestamp(), and getValue().
- The KeyValue blob format inside the byte array is:

<keylength> <valuelength> <key> <value>

◆ Key 的格式:



- ◆ Rowlength 最大值為 Short.MAX_SIZE,
- ◆ column family length 最大值為 Byte.MAX_SIZE,
- ◆ column qualifier + key length 必須小於 Integer.MAX_SIZE.



HBase 程式設計

實做I/O操作

© TemplatesWise.com



財團法人國家實驗研究院

國家高速網路與計算中心

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING



範例一：新增Table

<指令>

範例一：新增Table

<程式碼>

```
public static void createHBaseTable ( String tablename ) throws IOException
{
    HTableDescriptor htd = new HTableDescriptor(tablename);
    HColumnDescriptor col = new HColumnDescriptor("content:");
    htd.addFamily(col);
    HBaseConfiguration config = new HBaseConfiguration();
    HBaseAdmin admin = new HBaseAdmin(config);
    if(admin.tableExists(tablename))
    {
        admin.disableTable(tablename);
        admin.deleteTable(tablename);
    }
    admin.createTable(htd);
}
```

範例二：Put資料進Column

<指令>

範例二：Put資料進Column

<程式碼>

```
static public void putData(String tablename, String row, String family,
    String column, String value) throws IOException {
    HBaseConfiguration config = new HBaseConfiguration();
    HTable table = new HTable(config, tablename);
    byte[] brow = Bytes.toBytes(row);
    byte[] bfamily = Bytes.toBytes(family);
    byte[] bcolumn = Bytes.toBytes(column);
    byte[] bvalue = Bytes.toBytes(value);
    Put p = new Put(brow);
    p.add(bfamily, bcolumn, bvalue);
    table.put(p);
    table.close();
}
```

範例三： Get Column Value

<指令>

範例三：Get Column Value

〈程式碼〉

```
static String getColumn ( String tablename, String row, String family,
    String column ) {
    HBaseConfiguration conf = new HBaseConfiguration();
    String ret = "";
    HTable table;
    try {
        table = new HTable(conf, Bytes.toBytes(tablename));
        Get g = new Get(Bytes.toBytes(row));
        Result rowResult = table.get(g);
        ret = Bytes.toString(rowResult.getValue(Bytes.toBytes(family) + ":" + column));
        table.close();
    } catch (IOException e) {
        e.printStackTrace();
    }
    return ret;
}
```

範例四：Scan all Column

<指令>

範例四：Scan all Column

<程式碼>

```
static void ScanColumn(String tablename, String family, String column) {
    HBaseConfiguration conf = new HBaseConfiguration();
    HTable table;
    try {
        table = new HTable(conf, Bytes.toBytes(tablename));
        ResultScanner scanner = table.getScanner(Bytes.toBytes(family));
        int i = 1;
        for (Result rowResult : scanner) {
            byte[] by = rowResult.getValue(
                Bytes.toBytes(family), Bytes.toBytes(column) );
            String str = Bytes.toString ( by );
            System.out.println("row " + i + " is \"" + str + "\"");
            i++;
        }
    } catch (IOException e) {
        e.printStackTrace();
    }
}
```

範例五：刪除資料表

<指令>

範例五：刪除資料表

<程式碼>

```
static void drop ( String tablename ) {
    HBaseConfiguration conf = new HBaseConfiguration();
    try {
        HBaseAdmin admin = new HBaseAdmin(conf);
        if (admin.tableExists(tablename))
        {
            admin.disableTable(tablename);
            admin.deleteTable(tablename);
            System.out.println("Dropped the table [" + tablename+ "]);
        }else{
            System.out.println("Table [" + tablename+ "] was not found!");
        }
    } catch (IOException e) {
        e.printStackTrace();
    }
}
```



HBase 程式設計

MapReduce 與 HBase 的搭配

© TemplatesWise.com



財團法人國家實驗研究院

國家高速網路與計算中心

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING



範例六：WordCountHBase

- 程式說明

範例六：WordCountHBase

<1>

```
public class WordCountHBase
{
    public static class Map extends
        Mapper<LongWritable,Text,Text,
            IntWritable>
    {
        private IntWritable i = new
            IntWritable(1);
        public void map(LongWritable key,Text
            value,Context context) throws
            IOException, InterruptedException
        {
            String s[] =
                value.toString().trim().split(" ");
            for( String m : s)
            {
                context.write(new Text(m), i);
            }
        }
    }
}
```

```
public static class Reduce extends
    TableReducer<Text, IntWritable,
        NullWritable>
{
    public void reduce(Text key,
        Iterable<IntWritable> values, Context
        context) throws IOException,
        InterruptedException
    {
        int sum = 0;
        for(IntWritable i : values)
        {
            sum += i.get();
        }
        Put put = new
            Put(Bytes.toBytes(key.toString()));
        put.add(Bytes.toBytes("content"),
            Bytes.toBytes("count"),
            Bytes.toBytes(String.valueOf(sum)));
        context.write(NullWritable.get(), put);
    }
}
```

範例六：WordCountHBase

<2>

```
public static void createHBaseTable(String
    tablename)throws IOException
{
    HTableDescriptor htd = new
        HTableDescriptor(tablename);
    HColumnDescriptor col = new
        HColumnDescriptor("content:");
    htd.addFamily(col);
    HBaseConfiguration config = new
        HBaseConfiguration();
    HBaseAdmin admin = new
        HBaseAdmin(config);
    if(admin.tableExists(tablename))
    {
        admin.disableTable(tablename);
        admin.deleteTable(tablename);
    }
    System.out.println("create new table: " +
        tablename);
    admin.createTable(htd);
}
```

```
public static void main(String args[]) throws Exception
{
    String tablename = "wordcount";
    Configuration conf = new Configuration();
    conf.set(TableOutputFormat.OUTPUT_TABLE,
        tablename);
    createHBaseTable(tablename);
    String input = args[0];
    Job job = new Job(conf, "WordCount table with " + input);
    job.setJarByClass(WordCountHBase.class);
    job.setNumReduceTasks(3);
    job.setMapperClass(Map.class);
    job.setReducerClass(Reduce.class);
    job.setMapOutputKeyClass(Text.class);
    job.setMapOutputValueClass(IntWritable.class);
    job.setInputFormatClass(TextInputFormat.class);
    job.setOutputFormatClass(TableOutputFormat.class);
    FileInputFormat.addInputPath(job, new Path(input));
    System.exit(job.waitForCompletion(true)?0:1);
}
```

範例六：執行結果

範例七：LoadHBaseMapper

說明：

此程式碼將HBase的資料取出來，再將結果塞回hdfs上

運算方法：

將此程式運作在hadoop 0.20 平台上，用(參考2)的方法加入hbase參數後，將此程式碼打包成XX.jar

執行：

```
-----  
hadoop jar XX.jar LoadHBaseMapper <hdfs_output>  
-----
```

結果：

```
$ hadoop fs -cat <hdfs_output>/part-r-00000  
-----
```

```
54 30 31      GunLong  
54 30 32      Esing  
54 30 33      SunDon  
54 30 34      StarBucks  
-----
```

注意：

1. 請注意hbase 上必須要有 table, 並且已經有資料
2. 運算完後，程式將執行結果放在你指定 hdfs的<hdfs_output> 內
請注意 沒有 <hdfs_output> 資料夾

範例七：LoadHBaseMapper

<1>

```
public class LoadHBaseMapper {
public static class HtMap extends
    TableMapper<Text, Text> {
public void
    map(ImmutableBytesWritable
        key, Result value,
        Context context) throws
        IOException,
        InterruptedException {
    String res =
        Bytes.toString(value.getValue(
            Bytes.toBytes("Detail"),

                Bytes.toBytes("Name")));
    context.write(new
        Text(key.toString()), new
        Text(res));
}}
}
```

```
public static class HtReduce extends
    Reducer<Text, Text, Text, Text> {
public void reduce(Text key, Iterable<Text>
    values, Context context)
    throws IOException,
    InterruptedException {
    String str = new String("");
    Text final_key = new Text(key);
    Text final_value = new Text();
    for (Text tmp : values) {
        str += tmp.toString(); }
    final_value.set(str);
    context.write(final_key, final_value);
}}
}
```

範例七：LoadHBaseMapper

<2>

```
public static void main(String args[]) throws
    Exception {
    String input = args[0];
    String tablename = "tsmc";
    Configuration conf = new Configuration();
    Job job = new Job(conf, tablename + " hbase
        data to hdfs");
    job.setJarByClass(LoadHBaseMapper.class);
    TableMapReduceUtil.initTableMapperJob
    (tablename, myScan,
        HtMap.class, Text.class, Text.class,
        job);
    job.setMapperClass(HtMap.class);
```

```
    job.setReducerClass(HtReduce.class);
    job.setMapOutputKeyClass(Text.class);
    job.setMapOutputValueClass(Text.class);
    job.setInputFormatClass(TableInputFormat.clas
        s);
    job.setOutputFormatClass(TextOutputFormat.c
        lass);
    job.setOutputKeyClass(Text.class);
    job.setOutputValueClass(Text.class);
    FileOutputFormat.setOutputPath(job, new
        Path(input));
    System.exit(job.waitForCompletion(true) ? 0 :
        1);
    }
    }
```

範例七：執行結果



其他用法補充

HBase內contrib的項目，如

Trancational

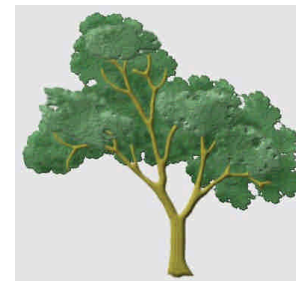
Thrift



財團法人國家實驗研究院

國家高速網路與計算中心

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING



1. Transactional HBase

- Indexed Table = Secondary Index = Transactional HBase
- 內容與原本table相似的另一張table，但key不同，利於排列內容

Primary Table

	name	price	description
1	apple	10	xx
2	orig	5	ooo
3	banana	15	vvvv
4	tomato	8	uu



Indexed Table

	name	price	description
2	orig	5	ooo
4	tomato	8	uu
1	apple	10	xx
3	banana	15	vvvv

1. 環境設定

需在 `$HBASE_INSTALL_DIR/conf/hbase-site.xml` 檔內
增加兩項內容

```
<property>
  <name>hbase.regionserver.class</name>
  <value>org.apache.hadoop.hbase.ipc.IndexedRegionInterface</value>
</property>
<property>
  <name>hbase.regionserver.impl</name>
  <value>
    org.apache.hadoop.hbase.regionserver.tableindexed.IndexedRegionServer
  </value>
</property>
```

1.a Ex : 從一個原有的Table 增加 IndexedTable

```
public void addSecondaryIndexToExistingTable(String
    TableName,
        String IndexID, String IndexColumn) throws
    IOException {
    HBaseConfiguration conf = new HBaseConfiguration();
    conf.addResource(new Path("/opt/hbase/conf/hbase-
    site.xml"));
    IndexedTableAdmin admin = null;
    admin = new IndexedTableAdmin(conf);
    admin.addIndex(Bytes.toBytes(TableName), new
    IndexSpecification(
        IndexID, Bytes.toBytes(IndexColumn)));
}}
```

1.b Ex : 建立一個新的Table 附帶 IndexedTable

```
public void createTableWithSecondaryIndexes(String TableName,
      String IndexColumn) throws IOException {
    HBaseConfiguration conf = new HBaseConfiguration();
    conf.addResource(new Path("/opt/hbase/conf/hbase-site.xml"));
    HTableDescriptor desc = new HTableDescriptor(TableName);
    desc.addFamily(new HColumnDescriptor("Family1"));
    IndexedTableDescriptor Idxdesc = new
    IndexedTableDescriptor(desc);
    Idxdesc.addIndex(new IndexSpecification(IndexColumn, Bytes
      .toBytes(" Family1 :" + IndexColumn)));
    IndexedTableAdmin admin = new IndexedTableAdmin(conf);
    admin.createIndexedTable(Idxdesc);
}
```

2. Thrift



其他專案

© TemplatesWise.com

王耀聰 陳威宇

Jazz@nchc.org.tw

waue@nchc.org.tw



財團法人國家實驗研究院

國家高速網路與計算中心

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING



PIG

Hive



Conclusions

- a

© TemplatesWise.com



財團法人國家實驗研究院

國家高速網路與計算中心

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING





Questions and Thanks

© TemplatesWise.com



財團法人國家實驗研究院

國家高速網路與計算中心

NATIONAL CENTER FOR HIGH-PERFORMANCE COMPUTING

