

## Table of Content

1. Requirement.....	2
2. Input .....	2
2.1 Schema file .....	2
2.2 Scale factor .....	2
2.3 Table files .....	2
2.4 Number reducers .....	3
3. Installation.....	3
4. How to use.....	3
4.1 Command line version .....	3
4.2 GUI version .....	3

## 1. Requirement

The source code was compiled on Ubuntu 10.04 32-bit edition with Java<sup>TM</sup> 1.6 and hadoop 0.20.2 installed.

## 2. Input

### 2.1 Schema file

Schema file is formatted in the following way:

Line 1: database name

Line 2: the number of tables in the database

This is then followed by the description of each table.

Line 1: table name, followed by table status.

Line 2: the number of attributes, followed by the number of foreign keys.

From line 3 onwards are the description of each attribute.

The attributes are arranged by primary key, then foreign keys, then other attributes.

All attributes descriptions start from attribute name and end with attribute type.

However, if it is foreign key attribute, the foreign table name is between the attribute name and attribute type.

There are 3 kinds of table status.

S: Static table

D: Dynamic table

Please read the sample schema information configuration file in the executable directory.

### 2.2 Scale factor

A positive real number

### 2.3 Table files

Table files records the content of the original dataset. Each file stores the tuples in one table. The file must meet the following requirements:

The file name should be exactly the same as the table name in the schema file, with “tbl” as name extension.

Each line in the file is a tuple.

The attribute values in a tuple are separated by “|”.

Please read the sample table files in the table folder.

## 2.4 Number reducers

Number reducers is the number of reducers you want to use in the MapReduce task.

You can set this parameter according to the number of working nodes in your cluster.

## 3. Installation

Before install UpSizeR, please install Java<sup>TM</sup> 1.6 and hadoop 0.20.2. For java installation please refer to [www.java.com](http://www.java.com). For hadoop installation, please refer to <http://hadoop.apache.org/>.

To install UpSizeR, simply decompress the file UpSizeR.zip.

## 4. How to use

### 4.1 Command line version

The command line version can run in a single node or a cluster with hadoop deployed.

Before running UpSizeR, please make sure you have the table files and the schema file in hands. Please load the following inputs into HDFS:

- Tables: Create a folder naming “tables” in HDFS. The command could be “hadoop dfs –mkdir tables”. Then copy the tables into this folder: “hadoop dfs –copyFromLocal tableName.tbl tables”.
- NumberReduc: Create a folder named numReducer in HDFS and a file which records the number of reducers. Put this file into the folder.

After loading these input into HDFS, please run the command to run UpSizeR: “hadoop jar UpSizeR\_Command.jar sg.edu.nus.UpSizeR schemaFileName scaleFactor numReducer”, in which schemaFileName is the path of the schema file, scaleFactor is the scale factor and numReducer is the number of reducers.

### 4.2 GUI version

For GUI version please refer to “UpSizeR GUI version.pdf”