HiTune

Visual Reporter Utility Manual

Contacts

Jason, Dai (jason.dai@intel.com)

Jie, Huang (jie.huang@intel.com)

Contents

Visual Reporter

Visual reporter is the additional utility for HiTune, which renders the figures and charts using the report sheets under Windows. It provides an insightful and visual view of the analysis results. We offer this utility as a method to use the raw data on the report sheets (generated by the *Analysis Engine*).

This document contains the information about how to configure and use this visual report utility based on the *Analysis Engine*'s output. There is no need to install any program other than Microsoft Excel 2007 on your Windows machine.

Currently the visual report engine supports two types of *Report Folder*, namely, the SINGLE-JOB and MULTI-JOB folders. In the SINGLE-JOB folder, there is only one report folder (belonging to a single job). While in the MUTI-JOB folder, there are multiple report sub-folders. The MUTI-JOB type allows you to load several jobs' data into the visual report template at the same time, and output the comparison report for these jobs. The raw data of the *Report Folder* can be retrieved from the *Analysis Engine*'s output originally.

Below is the "***3-step***" usage guide, and the further details can be found in the following sections:

 ***MUST do item***  ***Click Action***

1. ***Copy the Report Folder*** generated by *Analysis Engine* from a Linux machine to a Windows machine
2. ***Create and edit a TYPE file*** (the template is under the HiTune package) as the configuration step and copy the AnalysisReport.xlsm to the *Report Folder* on that Window machine, more details see ***2. Configure report folder*** .
3. ***Open the AnalysisReport.xlsm*** (contained in the HiTune package, ***click 1.configure***) file on that Windows machine, and import the data into the template(***click 2.load data***), more details see ***2. Configure report folder*** and ***3. Import report folder***.

There is an example of visual report (i.e., VisualReportExample) with some data results in the HiTune package for your reference, from which you can intuitively obtain the brief information about the major functionalities and usage offered by the visual utility.

1. Prepare report folder

This section describes how to retrieve and compose your *Report Folder*.

Under each generated *Report Folder*, there are several sheets with pre-defined name. ***Do not change the name of these sheets.*** You can copy these report sheets from the *Report Folder* generated by the *Analysis Engine* into a SINGLE-JOB folder, or as a sub-folder in the MULTI-JOB.

2. Configure report folder

This section describes how to configure your report folder before you run the visual report utility.

Before importing the data into the visual report template, you need to prepare the configurations for the source report folder. The sample file can be found under HiTune distribution package as well, which is named as "TYPE".

-  For SINGLE-JOB folder, you can just edit the TYPE file under the report folder as follows:

```
TYPE,NAME,AUTHOR,TIME,DESCRIPTION

SINGLE-JOB, your_test_name, your_name, your_test_time, your_test_description
```

The TYPE file should be put under the *Report Folder* along with other raw data files. **The first line is the header which should not be changed** and you need to prepare from the next line with comma-separated items. The items prefixed with "your_" in the above template needs to be modified based on your environment.

- For MULTI-JOBS folder, you need to put all the SINGLE-JOB *Report Folders* into one common MULTI-JOB folder. Then, create a new TYPE file under that MULTI-JOB folder as follows:

TYPE,NAME,AUTHOR,TIME,DESCRIPTION

MULTI-JOB, your_test_name, your_name, your_test_time, your_test_description

In addition, you need to copy the **AnalysisReport.xlsm** file in the HiTune package into your *Report Folder* (in addition to the TYPE file and the raw data files).

Now ***open*** the AnalysisReport.xlsm file and configure the location of the report folder for your visual report template. Please click the ***1.Configure*** button 

, it pops up a dialog box as follows in Figure 1.

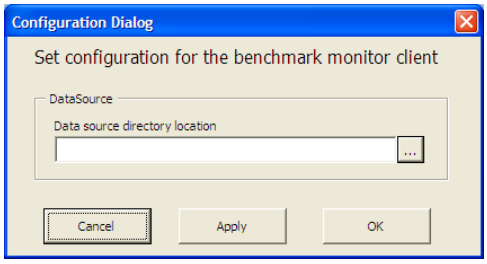


Figure 1. Configuration dialog

- Select the *Report Folder* (either *MULTI-JOB* or *SINGLE-JOB)* folder as the import data folder type
- Click ***Cancel*** to quit and without any changes
- Click ***Apply*** to import the configuration file without closing the dialog box
- Click ***OK*** to import the configuration file and close the dialog box

3. Import report folder

This section shows the steps to import a job data into the visual report template.

- To import the data, you can select the items from the Benchmark(s) Drop-down list in Figure 2.

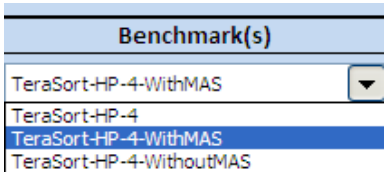
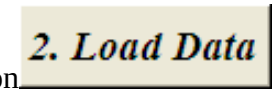


Figure 2. The Drop-down benchmark list

- Then click the ***2.Load Data*** button 

to import the data. It will cleanup all the dirty sheets and import the new data from the source report folder which you selected.

It pops up 2 dialogs waiting for you to input the Hadoop metrics sampling period in seconds, the system statistics' interval in second (sampling period for sysstat), you can choose the default values by clicking ***OK***. The default values are shown in the input text box.

4. View reports

This section describes the visual report's features in 3 perspectives.

- To view the details in each sub-category.
- To view the aggregation content in each sub-category for the whole cluster.
- To view the comparison content among several jobs.

4.1 Details report

The details report shows the certain metrics for each sub-category in both a compact way and raw data sheet. To see the details about each report sheet, you can choose the Metric Selection Drop-down list as follows in Figure 3.

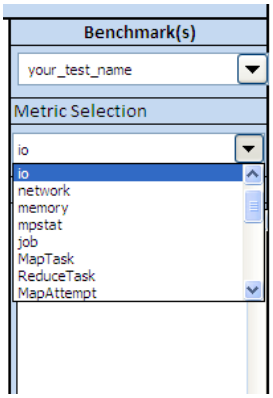


Figure 3. Metric selection Drop-down list

Then click ***Details*** button 

to see the detailed report sheet. Here is a snapshot for Reduce tasks analysis report in Figure 4.

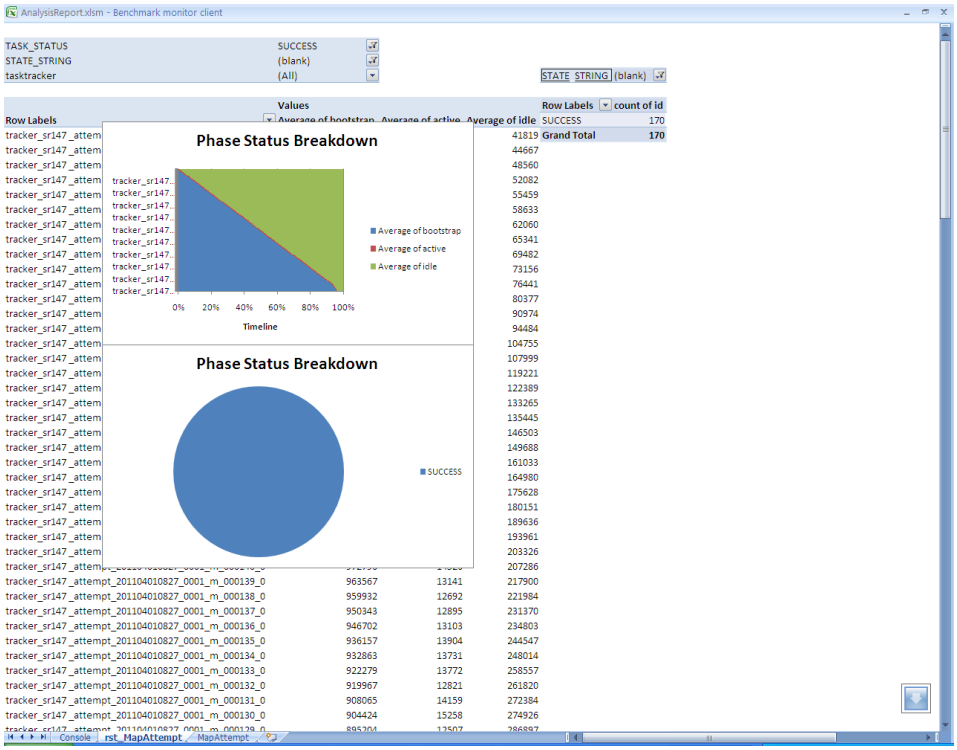


Figure 4. The snapshot for Reduce tasks analysis report

4.2 Aggregation report

The aggregation report prints out all the statistics figures for each node in the target cluster in a summary page. To see the aggregation report, you can choose the Metric Selection Drop-down

list first. Then click ***Aggregation*** button Aggregation to see the aggregation report. Generally, the aggregation report will show the visual charts for each node in the *print-out* style.

Here is a snapshot for network utilization aggregation report for the whole cluster in Figure 5.

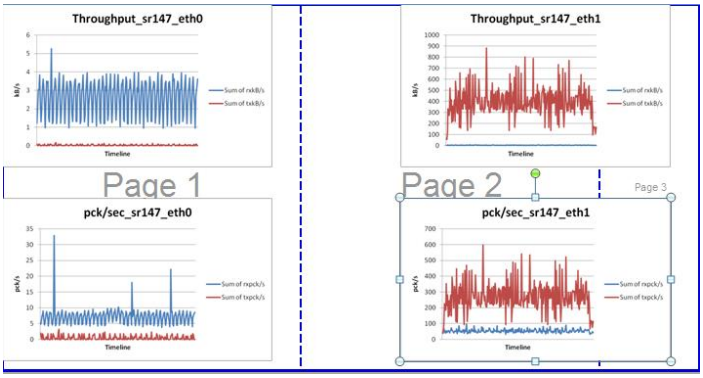


Figure 5. The snapshot of network utilization for the whole cluster

4.3 Comparison report

The comparison feature offers a summary figure to compare the running time with further break down among target jobs. You can get comparison report with the normalized running times, and the denominator is decided by the job running time of "***Ref benchmark***".

- First, you can choose one or more tests in the multi-selection list in Figure 6.

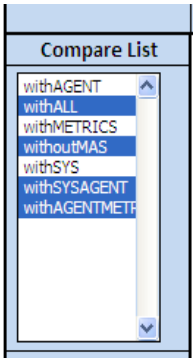


Figure 6. Multi-selection list for comparison benchmarks

- Select the reference test in Figure 7. The job running time of the selected Ref benchmark will be normalized as 100%.

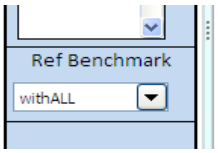


Figure 7. Reference benchmark for the comparison benchmark list

- Select the Radio button "***Normalized Data***" in Figure 8.

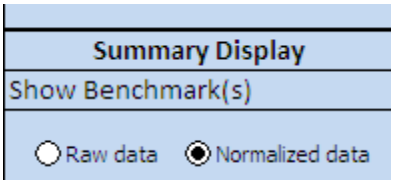


Figure 8. Comparing ratio button

The result will be shown as follows in Figure 9:

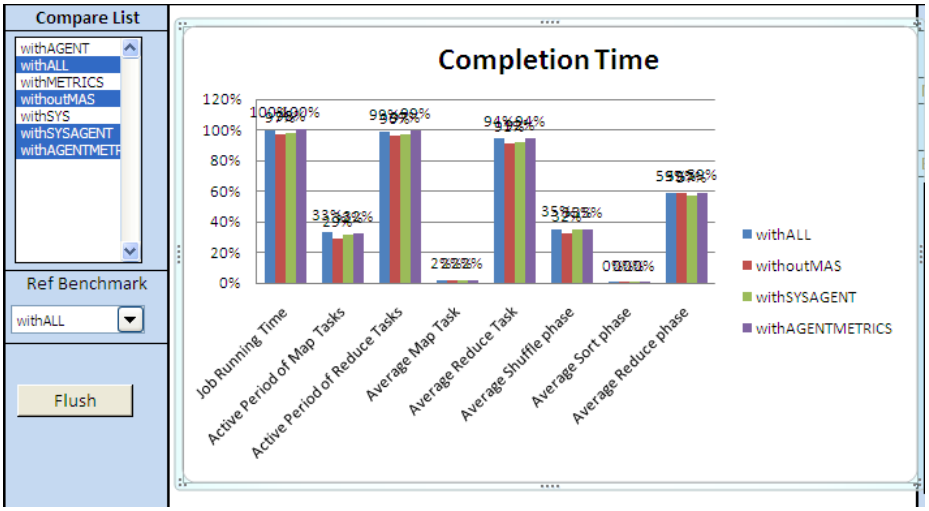


Figure 9. Comparison report among several benchmarks

Appendix

1. Visual Report component list

In this part, we list the supporting components in this visual report utility currently in the following list (Table 1).

Table 1. HiTune visual reporter's supporting components list

Category	Sub-category	Metrics
System statistics	CPU	Utilization
	IO	IO throughput, latency, utilization, request size
	Network	Throughput
	Memory	Utilization
	mpstat	Utilization for each core
History Log	Map Tasks	Active period, bootstrap, idle of Map Stage
	Reduce Tasks	Active period, bootstrap, idle of Reduce Stage
	Map Attempt	Active period, bootstrap, idle of each map task
	Reduce Attempt	Active period(shuffle, sort, reduce), bootstrap, idle for each reduce task
	Job	N/A
Dataflow based statistics	Stage for Map Tasks	Active period, bootstrap, idle for each map task; Further breakdowns in Map Task Pool;
	Stage for Reduce Tasks	Active period, bootstrap, idle for each reduce task; Shuffle period breakdown for total reduce tasks; Further breakdowns in Reduce Task Pool;
Hotspot analysis statistics	Stage for Map Tasks	Top N called functions in Map Task Pool
	Stage for Reduce Tasks	Top N called functions in Reduce Task Pool
Hadoop Metrics	JVM	Thread status distribution, Thread heap utilization
	Datanode	N/A
	TaskTracker	Map and Reduce Tasks execution time line
	Namenode	NameNode operations
	JobTracker	Tasks throughput for the cluster
	Map Shuffle out	Busy status percentage for Shuffle out
	Reduce Shuffle in	Busy status percentage for Shuffle in

2. Terminology in Visual Report

Following table shows the decryption and its detailed formula for all those HiTune's terms in this visual report.

Table 2. Explanations and detailed formula for all terms in HiTune visual report

Terminology	Description	Formula
Job Running Time	The total elapsed time for a single job	= FINISH_TIME - SUMBMIT_TIME
Active Period of Map Tasks	The period of time when any map task is running	= LAST_MAP_FINISH_TIME - FIRST_MAP_START_TIME
Active Period of Reduce Tasks	The period of time when any reduce task is running	= LAST_REDUCE_FINISH_TIME -FIRST_ REDUCE_START_TIME
Average Map Task	Average completion time for all map tasks	= SUM(FINISH_TIMEmap_i - START_TIMEmap_i) / COUNT(MAP)
Average Reduce Task	Average completion time for all reduce tasks	= SUM(FINISH_TIMEreduce_i - START_TIMEreduce_i) / COUNT(REDUCE)
Average Shuffle phase	Average completion time of the shuffle phase for all reduce tasks	= SUM(SHUFFLE_FINISH_TIMEreduce_i - START_TIMEreduce_i) / COUNT(REDUCE)
Average Sort phase	Average completion time of the sort phase for all reduce tasks	= SUM(SORT_FINISH_TIMEreduce_i – SHUFFLE_FINISH_TIMEreduce_i) / COUNT(REDUCE)
Average Reduce phase	Average completion time of the reduce phase for all reduce tasks.	= SUM(REDUCE_FINISH_TIMEreduce_i – SORT_FINISH_TIMEreduce_i) / COUNT(REDUCE)
BOOTSTRAP	The period of time when a task is in the pending status.	= TASK_START_TIME – JOB_START_TIME
ACTIVE	The period of time when a task is running on certain Tracker	= TASK_FINISH_TIME – TASK_START_TIME
IDLE	The period of time between each task's finishing time and the whole job finishing time.	= JOB_ FINISH _TIME – TASK_FINISH_TIME
SHUFFLE	The period of time when the reduce task copies those map output data into its local node.	= SHUFFLE_ FINISH _TIME – REDUCE_START_TIME
SORT	The period of time when the reduce task sort all fetched keys	= SORT_FINISH_TIME - SHUFFLE_ FINISH _TIME
REDUCE	The period of time the reduce task does the real user defined reduce(aggregation) function.	= REDUCE_FINISH_TIME - SORT_FINISH_TIME