

Exporting Probability Distributions using sierra_stats

ClustrixDB provides the ability to export Probability Distributions (PDs), which can be useful along with schema (DDL) to investigate the performance of a query. Typically, this procedure is performed upon request by Clustrix Support. The following section outlines how to export PDs using a python script, sierra_stats.

Usage of sierra_stats

Sierra_stats is included by default as part of installation in /opt/clustrix/bin/ directory. The required inputs are --host,--user, and -p (prompt for password), and --file, which is the name of the .tar.gz file to produce.

```
./sierra_stats -u username -H hostname_or_ip -p --db database -f filename [space_separated_list_of_tables]
```

Running sierra_stats requires read only access to SYSTEM tables. If no list of tables is specified, sierra_stats can export all the tables in one database. To export tables from multiple database, include a fully qualified list of tables.

sierra_stats requires Python version 2.7

The output of sierra_stats is a .tar.gz file consisting:

1. a SQL text file for each database with CREATE TABLE statements.
2. A single pds.bin file is created containing PD data for all tables included in the export.

Options

option flag(s)	description
-h/--help	Show help message and exit
--host/-H	Clustrix node hostname to connect to. In system.users if host for the user is 127.0.0.1, this must be used for hostname.
--user/-u	Username to connect as
--port/-P	Database port (default:3306)
--passwd/-PASSWD	Provide password on CLI (not recommended)
--prompt-passwd/-p	Prompt for password (recommended)
--db	Database name to dump tables from
--file/-f	filename for export output

Caveats for sierra_stats

- sierra_stats cannot be used with users that use sha256 authentication