hashdb Quick Reference

http://github.com/simsong/hashdb/wiki

General Usage

`hashdb <command> <options> <parameters>`	Run hashdb command, `-q` for quiet mode, `-f <flags>` to control B-Tree flags

New Database

`create [-p <hash block size>] [-m <maximum duplicates>] [<bloom settings>] <hashdb.hdb>`	Create a new hash database

Import/Export

`import [-r <repository name>] <hashdb.hdb> <dfxml.xml>`	Import from DFXML file into hash database
`export <hashdb.hdb> <dfxml.xml>`	Export hash database to DFXML file

Database Manipulation

`add <A.hdb> <B.hdb>`	$A + B \rightarrow B$ Add A into B
`add_multiple <A.hdb> <B.hdb> <C.hdb>`	$A + B \rightarrow C$ add A and B into C
`add_repository <A.hdb> <B.hdb> <repository name>`	$A + B \rightarrow B$ Add A into B but only when the repository name matches
`intersect <A.hdb> <B.hdb> <C.hdb>`	$A \cap B \rightarrow C$ add intersection of A and B into C
`intersect_hash <A.hdb> <B.hdb> <C.hdb>`	$A \cap B \rightarrow C$ intersect into C when hashes match
`subtract <A.hdb> <B.hdb> <C.hdb>`	$A - B \rightarrow C$ add A but not B into C
`subtract_hash <A.hdb> <B.hdb> <C.hdb>`	$A - B \rightarrow C$ add A but not hashes in B into C
`deduplicate <A.hdb> <B.hdb>`	Copy $A \rightarrow B$ except for hashes with duplicates

Scan Services

`scan <path or socket> <dfxml.xml>`	Scan DFXML file for matching hashes
`scan_hash <path or socket> <hash value>`	Scan for hash match
`scan_expanded <hashdb.hdb> <dfxml.xml>`	Scan DFXML file for matches showing all sources
`scan_expanded_hash <hashdb.hdb> <hash value>`	Scan for hash match showing all sources
`server <hashdb.hdb> <port number>`	Start scan service at port

Statistics

`size <hashdb.hdb>`	Print sizes of internal database tables
`sources <hashdb.hdb>`	Print source metadata
`histogram <hashdb.hdb>`	Print hash distribution
`duplicates <hashdb.hdb> <number>`	Print hashes sourced the given number of times
`hash_table <hashdb.hdb> <repository name> <filename>`	Print the hashes associated with this source
`expand_identified_blocks <hashdb.hdb> <identified_blocks.txt>`	Expand to include source information for each source
`explain_identified_blocks [-m <number>] <hashdb.hdb> <identified_blocks.txt>`	Print information about less frequently observed hashes

Tuning

`rebuild_bloom [<bloom settings>] <hashdb.hdb>`	Rebuild Bloom filter
`upgrade <hashdb.hdb>`	Make database from v1.0.0 compatible with v1.1.0

Performance Analysis

`add_random [-r <repository name>] <hashdb.hdb> <count>`	Add random hashes, log performance in `log.xml`
`scan_random <hashdb.hdb> <copy.hdb>`	Scan random hashes, log performance in `log.xml`

***bulk_extractor* Scanner**

`bulk_extractor -E hashdb -S hashdb_mode=import -o outdir1 -R my_import_dir`	Import directory
`bulk_extractor -E hashdb -S hashdb_mode=scan -S hashdb_scan_path_or_socket= outdir1/hashdb.hdb -o outdir2 my_image2`	Scan image

October 31, 2014