
HCP
MQ
Tool

**HCP Metadata Query Tool
Documentation**

Release 1.0.11

Thorsten Simons

January 19, 2015

1	Introduction	1
2	Contents	3
2.1	Prerequisites	3
2.2	Tutorial	3
2.3	Technical description	6
2.4	License	7

Introduction

HCP Metadata Query Tool (*HCPmqt*) queries Hitachi Content Platform (*HCP*) for information about object transactions. These are the ingestion and the deletion of objects within HCP. The term *deletion* also includes disposition, purging and pruning of objects.

Warning:

- Using *HCPmqt* can put severe load on *HCP*, especially if there is a huge number of objects stored. You might want to monitor *HCP* performance during a query and tune the load parameters accordingly.
- The output file generated by this tool will be huge, depending on the number of objects (expect round-about 20 GB for 100 million objects found). If you use the 'Dirtree' feature, expect the tool to claim up to 512 MB memory per 100 million objects. The tool will fail if it runs out of memory; to add insult to injury, this may affect other applications running on your system!

Handle with care!

2.1 Prerequisites

To be able to use *HCPmqd*, there are several prerequisites:

- Using an system-level user account, log into the *System Console*:
 - Navigate to *Security / Permissions* and make sure that the **Search** permission is checked within the *Systemwide Permission Mask*.
 - A System-level user account can be used with *HCPmqd* for those Tenants that have delegated administrative rights to *System-level users*. In this case, the user account is required to have the *Search* role: navigate to *Security / Users* (or *Groups*), select the desired user in the list to open its panel. Make sure that **Search** is checked in the *Role* panel.
- Using an Tenant user account, log into the *Tenant Management Console* of the Tenant to be queried:
 - Check if the **Search** permission is set in the *Permissions* panel on the *Overview* page.
 - Check that the **Search** permission is set for each Namespace to be queried.
 - Make sure the account to be used with *HCPmqd* has the *Search* permission: Navigate to *Security / Users* (or *Groups*) and select the user from the list. In the list of Namespaces at the bottom, select the Namespaces of interest and make sure that **Search** is checked.
- To use a data access user (a Tenant Account without administrative rights):
 - use an administrative Tenant user to make sure that the respective user has the *Search* permission.

Tip: Tenants do not need to have the *Search* feature enabled to be queried by *HCPmqd*, nor is there a need to index the content!

2.2 Tutorial

This is a step-by-step guide on how to use HCP Metadata Query Tool.

2.2.1 HCP access Parameters

In this fields, you specify the parameter needed to access HCP and your area of interest within.

Tip: Depending on the access rights you have for HCP, use different names:

- enter `admin.hcp.your-domain.com` if you have a *HCP System Console* account with the Search Role enabled
- enter `tenant.hcp.your-domain.com` if you have a *HCP Tenant Console* account for the specified Tenant with the Search Role enabled
- enter `tenant.hcp.your-domain.com` **and** `namespace.tenant` if you have a *Data Access* account for the specified Namespace with the Search Role enabled

You may further restrict the result by defining folders that should be queried - this will skip any other folders.

2.2.2 HCP load Parameters

Warning: Using *HCPmq* can put severe load on *HCP*, especially if there is a huge number of objects stored. You might want to monitor *HCP* performance during a query and tune the load parameters accordingly.

These values are intended to tune the load generated within *HCP* when running *HCPmq*.

Use the *Records / page* field to specify the number of records that gets fetched from *HCP* with a single call. Larger number speed things up a bit, but need more local memory - where smaller number slow down things a bit, but need less memory. 5,000 to 10,000 is a value known as good.

The *Throttle (sec/page)* field asks the tool to pause for the defined number of seconds between subsequent page requests

Tip: Both values may be changed while a query is running. Please note that changes won't take place until the page in work has been processed.

2.2.3 HCP query Parameters

Select the type of operational records you want to get.

Transaction type	Description
create	existing (!) objects
delete	objects that have been deleted
dispose	objects that have been automatically deleted by HCP after the objects retention had expired ³
prune	object's versions that have been automatically deleted after their lifetime has passed ⁴
purge	object's versions that have been deleted when the <i>head</i> -object (the newest version) was deleted

2.2.4 Time Range

You can specify a time range for the query.

Per default, values are provided for a *full query*, which means anything from Jan. 1st, 1970 until **now** (use the *Reset* button to reset the fields).

Tip: Normally, you need to enter a timestamp exactly in the given format. In addition to this, a number of seconds counted from Jan. 1st 1970 (the Unix-epoch) will be accepted, also.

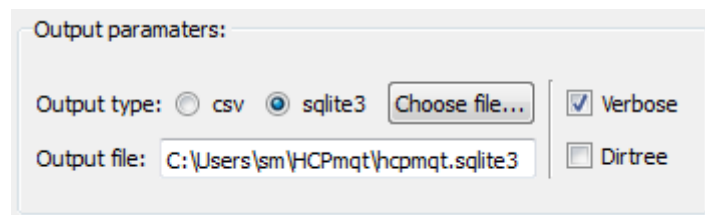
¹Disposition will take place if the *Disposition Service* is enabled in the System Console and for the Namespace, too.

²Namespaces that are enabled for *Versioning* define a periode of time during that versions of objects are kept. After a version is leaving this periode of time, it will be pruned (deleted) automatically.

³Disposition will take place if the *Disposition Service* is enabled in the System Console and for the Namespace, too.

⁴Namespaces that are enabled for *Versioning* define a periode of time during that versions of objects are kept. After a version is leaving this periode of time, it will be pruned (deleted) automatically.

2.2.5 Output



Two different output types are available:

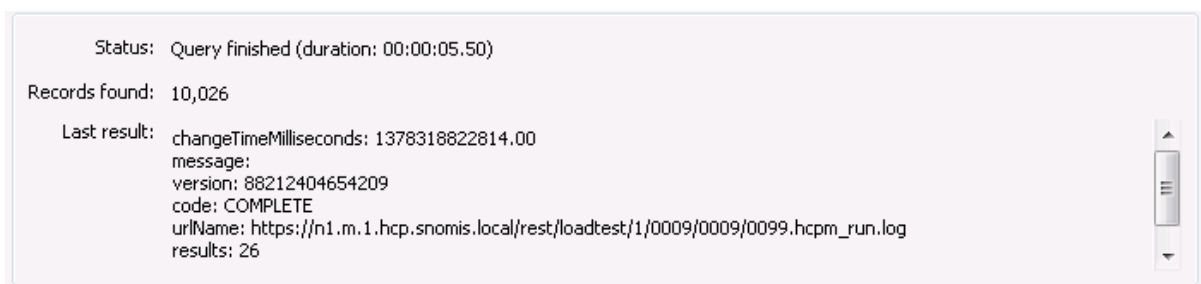
- csv - comma separated values (used to import data into spreadsheet software)
- sqlite3 - a single-file database ⁵

Normally, the output will hold selected information only: `urlName`, `version`, `operation` and `changeTimeMilliseconds`.

If **Verbose** is checked, all information will be provided: `urlName`, `objectPath`, `utf8Name`, `version`, `namespace`, `operation`, `type`, `size`, `retention`, `retentionString`, `retentionClass`, `ingestTimeString`, `ingestTime`, `accessTimeString`, `accessTime`, `changeTimeString`, `changeTimeMilliseconds`, `updateTimeString`, `updateTime`, `hashScheme`, `hash`, `acl`, `dpl`, `customMetadata`, `hold`, `index`, `replicated`, `shred`, `permissions`, `owner`, `uid`, `gid`

Tip: If you need statistical data for a Namespace (or a HCP system at all), check **Dirtree**. This will write an additional file holding a JSON-structure containing the complete directory tree, including the number of files and subfolders per folder.

2.2.6 Status



After pressing the **Run Query** button, the status frame will show information about the progress of a query.

You can pause a query at any time and you can cancel a query, as well; nevertheless you need to wait for the actual page query being ready before pause or cancelation takes place.

2.3 Technical description

This is what happens when you hit the **Run Query** button:

- The Domain Name Server is queried for the IP addresses of *HCP*. The first address received is the one that will be used for all communication with HCP during this query.

⁵ SQLite3 databases can be used by most programming languages. You can also discover them by using the *SQLite Shell* available from sqlite.org if you like to use the commandline; if you prefer a GUI, try the *SQLite Manager Add-on* for the Firefox Webbrowser.

- The parameters given are used to build a query in XML format,
- which is then send to *HCP*.
- HCP runs an internal query against its database and delivers the first page of results back to *HCPmqt*.
- *HCPmqt* processes the page and -if the page isn't flagged as *COMPLETED*- will build the next query-XML.

This runs in a loop until all requested records have been received.

2.4 License

HCP Metadata Query Tool is free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version.

HCP Metadata Query Tool is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

You should have received a copy of the GNU General Public License along with HCP Metadata Query Tool. If not, see [the license page at gnu.org](http://gnu.org).

Copyright 2012-2015 Thorsten Simons

search