

# Harvest History

- All jobs
  - Job list links
    - JobID details
    - Harvest definition job history
    - Specific run listing
    - Details on a terminated job
- All job per domain
- Running jobs

## All jobs

**Harvest Status** in the left menu by default shows a chronological list of all jobs ever harvested with status **Started** in **ascending order**. The same does **All jobs**

The screenshot shows the Harvest History interface. On the left is a 'Menu' with links: Definitions, Harvest status, All Jobs, All Jobs per domain, Running Jobs, Bitpreservation, Quality Assurance, and Systemstate. The main area has a search bar with 'Job status' (dropdown: All, New, Submitted, Started), 'Harvest name', and 'Start date' fields. Below the search bar are 'Order' (dropdown: Ascending), 'Display' (input: 100), and 'rows per page.' buttons, along with 'Show' and 'Reset' buttons. A message says 'Search results: 2, displaying results 1 to 2.' and 'previous / next' navigation. A table titled 'Job Status' has columns: Job ID, Harvest name, Run number, Start time, End time, Status, Harvest errors, Upload errors, Number of configurations, and a checkbox. Two rows are shown: Job 1 (My\_Harvest, Run 0, Started, 1 config) and Job 2 (My\_Snapshot\_Harvest, Run 0, Started, 17 configs). A 'Resubmit selected failed jobs' button is at the bottom.

Job ID	Harvest name	Run number	Start time	End time	Status	Harvest errors	Upload errors	Number of configurations	
1	My_Harvest	0	2011/09/05 14:57:49	-	Started	-	-	1	<input type="checkbox"/>
2	My_Snapshot_Harvest	0	2011/09/05 14:57:49	-	Started	-	-	17	<input type="checkbox"/>

If information is needed for jobs with other statuses (or **All** statuses) or other sort order, then this can be specified in the combo-boxes in the top of the page and then activated by clicking the **Show** button.

For each job the page shows information about the job and its status as well as information about errors (harvest errors or upload errors) and number of configurations in the job.

- Chose *JobID* if you want to check details on a specific job. See [Job list links](#).
- Chose Harvest name if you want to check details on the history of a specific harvest definition. See [History of a harvestdefinition](#) for details.
- Chose *Run number* if you want to check details on a specific run of that harvestdefinition – note that a run can consist of multiple jobs. See [Specific run listing](#).

In case of Harvest errors, a **Restart** button will appear and the operator can choose to resubmit that specific job to be harvested again. When resubmitting a failed job, the status will say 'Resubmitted' and refer to the new resubmitted job.

It is possible to go directly to the following

## Job list links

The following pages with job details can be accessed from the job status row

### JobID details



## Menu

- Definitions
- Harvest status
  - All Jobs
  - All Jobs per domain
  - Running Jobs
- Bitpreservation
- Quality Assurance
- Systemstate

## Details for Job 4

Job ID	Type	Harvest name	Run number	Created	Start time	End time	Status	Harvest errors	Upload errors	Object limit	Byte limit
4	Partial harvest	My_selective_harvest00		Sep 5, 2012 11:05:21 AM	Sep 5, 2012 11:05:49 AM		Started			100	500,000,000

## QA job selection

Select this job for QA with viewerproxy

The link above will select this job for the viewerproxy browse index. This will only work if your browser is set up to use the viewerproxy as web proxy.

## Included domains and configurations

Domain	Configuration	Bytes Harvested	Documents Harvested	Stopped due to
netarkivet.dk	defaultconfig	-	-	-

## Seed list

[www.netarkivet.dk](http://www.netarkivet.dk)

## Harvest order template (based on default\_orderxml)

[Show harvest template for job 4](#)

Clicking on a jobID on any of the harvest history pages will give you a very detailed report on the job.

This page gives all the information available about the job itself (e.g. max-bytes limit) and about the single domains included in the job.

Furthermore the page shows the complete seedlist used with the job and the complete **Harvest order template** as well as detailed error information in case of errors. The two latter is mainly for advanced users debugging specific crawls where things didn't go as expected.

### Harvest definition job history



## Menu

- Definitions
- Harvest status
  - All Jobs
  - All Jobs per domain
  - Running Jobs
- Bitpreservation
- Quality Assurance
- Systemstate

## Searching for 'My\_Snapshot\_Harvest' returned 1 hits:

Search results: 1, displaying results 1 to 1.

[previous](#) / [next](#)

## Harvest history for full harvest My\_Snapshot\_Harvest

Run number	Start time	End time	Bytes Harvested	Documents Harvested	Total number of jobs	Number of failed jobs	Number of resubmitted jobs
<a href="#">0</a>	Sep 5, 2011 2:57:49 PM	Sep 5, 2011 3:08:17 PM	82,235,861	2,702	1 <a href="#">Show jobs</a>	0	0

The history page for a harvestdefinition is the same as you can reach from the frontpage with the **History** buttons. This history page gives you further information for each run of the harvestdefinition: Start time, End time, number of bytes harvested and number of documents harvested. The page also show how many jobs each run consists of and how many of these that failed and eventually got resubmitted.

### Specific run listing

If a link in the 'Run number' column is clicked, the jobs for the specific run of the harvests definition is listed. A scheduled harvest definition might generate a number of jobs depending on the configurations and number of domains. See [Harvester design](#) for details on how jobs are generated.



**Menu**

- Definitions
- Harvest status
  - All Jobs
  - All Jobs per domain
  - Running Jobs
- Bitpreservation
- Quality Assurance
- Systemstate

[Dansk](#) [English](#) [Deutsch](#) [Italiano](#) [Français](#)

All  
 New  
 Submitted  
 Started  
 Done  
 Failed  
 Resubmitted

**Job status**

**Harvest name**

**Status for harvest `Snapshot-harvest-1, run 0`**

Search results: 3, displaying results 1 to 3.

[previous](#) / [next](#)

### Job Status

Job ID	Status	Harvest errors	Upload errors	Number of configurations
14	Failed	-	No hosts report found	5
15	Failed	-	No hosts report found	10
16	Done	-	-	17

### QA job selection

[Select these jobs for QA with viewerproxy](#)

The link above will select these jobs for the viewerproxy browse index. This will only work if your browser is set up to use the viewerproxy as web proxy.

### Details on a terminated job

If you terminates a running job in the Heritrix GUI, you can view Job Details and see that the job is stopped due to "Harvesting aborted". If it is a bigger (snapshot) harvest, that includes several jobs, all the finished jobs, will appear as "Done". And only the ones that was actually stopped will appear as stopped due to "Harvesting aborted" for some domains.

### All job per domain

This link is obsolete and will just advise the user to access the `harvest history` for a domain functionality from [the domain details page](#).

### Running jobs

The running jobs page displays details about the jobs currently being run by the harvesters. The information show here is continuously extracted from the running Heritrix instance and returned to the GUI.



**Menu**

- Definitions
- Harvest status
  - All Jobs
  - All Jobs per domain
  - Running Jobs
- Bitpreservation
- Quality Assurance
- Systemstate

[Dansk](#) [English](#) [Deutsch](#) [Italiano](#) [Français](#)

### Running Jobs

There is 1 running Jobs.

**Harvest definition `My_Snapshot_Harvest`**

Job ID	Host	Progress	Elapsed time	Queued files	Queues				Performance			Alerts
					Queues	Active	Retired	Exhausted	URL/s	KB/s	Threads	
11	●	0%	0d 00:00:00	0	0	0	-1	-1	0 (0)	0 (0)	0	0

● - crawl in preparation, ● - crawler is running, ● - crawler is pausing, ● - crawler is paused, ● - crawl finished

Find job harvesting domain



