

Heritrix Prerequisites

If you are to crawl a web site you will need to make sure you have the following points covered:

- The Aspire Server running the crawl must have access to the seed(s) URL (configured in the content source configuration).
- Check for any credentials needed to access the sites to be crawled. (Basic, Digest, HTTP forms and NTLM are supported)