

Choosing a Database to Hold Crawl States

What is the crawl state?

The crawl state is the representation of the progress made by an in-progress crawl. By using the crawl state, Aspire is able to distribute processing tasks between multiple Aspire nodes, as well as pause and resume crawls. In order to synchronize the crawl state between multiple Aspire nodes, as well as preserve the state on restarts, it must be stored in an external database.

Related pages

- [HBase Settings](#)
- [HBase Connector](#)

Which databases does Aspire use?

As of Aspire 3.2, we support two different databases:

1. [MongoDB 3.4.1](#) - Open Source, document-oriented database.
 - For more information, see [Installing and Configuring a 3.4.1 Crawl Status Database](#)
2. [Apache HBase](#) - Open Source, distributed, scalable, big data store.
 - For more information, see [Configuring Aspire to use HBase](#)