

Apache HBase - Installing and Configuring a Crawl Status Database

HBase can run in two modes:

- Standalone. This is the default mode. In standalone mode, HBase does not use HDFS—it uses the local filesystem instead
- Distributed. For a production environment, distributed mode is advised. In distributed mode, multiple instances of HBase daemons run on multiple servers in the cluster. HDFS is used. Please see [Aspire Parcel and Service for Cloudera](#) for an example of Aspire running with a distributed HBase setup on the Cloudera environment.

Installation and configuration instructions is part of this book: <http://hbase.apache.org/book.html>.

How to Configure Aspire with HBase is described here: [HBase Settings](#).

HBase version used for development: server 1.2.6, hbase-client 1.2.4, zookeeper 3.4.3