

Operational Database powered by Apache Accumulo 2.1.2

OpDB powered by Apache Accumulo Overview

Date published: 2021-03-03

Date modified: 2024-10-30

CLOUDERA

<https://docs.cloudera.com/>

Legal Notice

© Cloudera Inc. 2024. All rights reserved.

The documentation is and contains Cloudera proprietary information protected by copyright and other intellectual property rights. No license under copyright or any other intellectual property right is granted herein.

Unless otherwise noted, scripts and sample code are licensed under the Apache License, Version 2.0.

Copyright information for Cloudera software may be found within the documentation accompanying each component in a particular release.

Cloudera software includes software from various open source or other third party projects, and may be released under the Apache Software License 2.0 (“ASLv2”), the Affero General Public License version 3 (AGPLv3), or other license terms. Other software included may be released under the terms of alternative open source licenses. Please review the license and notice files accompanying the software for additional licensing information.

Please visit the Cloudera software product page for more information on Cloudera software. For more information on Cloudera support services, please visit either the Support or Sales page. Feel free to contact us directly to discuss your specific needs.

Cloudera reserves the right to change any products at any time, and without notice. Cloudera assumes no responsibility nor liability arising from the use of products, except as expressly agreed to in writing by Cloudera.

Cloudera, Cloudera Altus, HUE, Impala, Cloudera Impala, and other Cloudera marks are registered or unregistered trademarks in the United States and other countries. All other trademarks are the property of their respective owners.

Disclaimer: EXCEPT AS EXPRESSLY PROVIDED IN A WRITTEN AGREEMENT WITH CLOUDERA, CLOUDERA DOES NOT MAKE NOR GIVE ANY REPRESENTATION, WARRANTY, NOR COVENANT OF ANY KIND, WHETHER EXPRESS OR IMPLIED, IN CONNECTION WITH CLOUDERA TECHNOLOGY OR RELATED SUPPORT PROVIDED IN CONNECTION THEREWITH. CLOUDERA DOES NOT WARRANT THAT CLOUDERA PRODUCTS NOR SOFTWARE WILL OPERATE UNINTERRUPTED NOR THAT IT WILL BE FREE FROM DEFECTS NOR ERRORS, THAT IT WILL PROTECT YOUR DATA FROM LOSS, CORRUPTION NOR UNAVAILABILITY, NOR THAT IT WILL MEET ALL OF CUSTOMER’S BUSINESS REQUIREMENTS. WITHOUT LIMITING THE FOREGOING, AND TO THE MAXIMUM EXTENT PERMITTED BY APPLICABLE LAW, CLOUDERA EXPRESSLY DISCLAIMS ANY AND ALL IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO IMPLIED WARRANTIES OF MERCHANTABILITY, QUALITY, NON-INFRINGEMENT, TITLE, AND FITNESS FOR A PARTICULAR PURPOSE AND ANY REPRESENTATION, WARRANTY, OR COVENANT BASED ON COURSE OF DEALING OR USAGE IN TRADE.

Contents

Release notes.....	4
OpDB powered by Apache Accumulo overview.....	4
OpDB CLI tool support.....	4
OpDB system requirements.....	5

Release notes

You can go through the release notes of the respective Accumulo parcel releases.

- For Accumulo-on-cdp 1.1.0 parcel release notes, see [Apache Accumulo 2.0.0](#) and [Apache Accumulo 2.0.1](#) release notes.
- For Accumulo-on-cdp 1.10.3.7.1.7.2000.0 parcel release notes, see [Apache Accumulo 1.10.0](#) and [Apache Accumulo 1.10.3](#) release notes.
- For Accumulo-on-cdp 2.1.2.7.1.9.0.0 parcel release notes, see [Apache Accumulo 2.1.2](#) release notes.

OpDB powered by Apache Accumulo overview

Apache Accumulo™ is an ideal solution for government agencies looking for a secure, distributed NoSQL data store to serve their most performance-intensive Big Data applications. The Accumulo service is the Cloudera package of Apache Accumulo.

Apache Accumulo™ is an open source project integrated with Hadoop that provides the ability to store data in massive tables (billions of rows / millions of columns) for fast and random access. Accumulo was created and contributed to the Apache Software Foundation by the National Security Agency (NSA). It has quickly gained adoption as a Hadoop-based key or value store for applications that have unique and stringent information security requirements.

For more information about Apache Accumulo™, see the [Apache Accumulo](#) project.

OpDB CLI tool support

The Accumulo CLI toolset is different in Apache Accumulo 2.0 and later versions than it was in Accumulo 1.x. As the initial Operational Database powered by Apache Accumulo (OpDB) release is based on the Apache Accumulo 2.0.1 version, it is important to learn about these differences.

- For Accumulo-on-cdp 1.1.0 (Based on upstream version: Apache Accumulo 2.0.1) parcels:

Before Apache Accumulo™ 2.0 the following toolsets were available:

- accumulo
- accumulo-tool

Since Apache Accumulo™ 2.0 the following toolsets are available:

- accumulo
- accumulo-cluster
- accumulo-service
- accumulo-util



Attention: Do not use accumulo-cluster and accumulo-service with the OpDB service. These cluster and service manager tools are not compatible with Cloudera Manager. If you use them, your OpDB service can become unavailable.

- For Accumulo-on-cdp 1.10.0 (Based on upstream version: Apache Accumulo 1.10.3) parcel:

The Accumulo CLI toolset is different in Apache Accumulo 2.0 and later versions than it was in Accumulo 1.x. As the initial Accumulo parcel release is based on the Apache Accumulo 1.10.3 version, it is important to learn about these differences.

With Apache Accumulo™ 1.10.0 the following toolsets were available:

- accumulo
- accumulo-tool
- For Accumulo-on-cdp 2.1.2 (Based on upstream version: Apache Accumulo 2.0.1) parcel:

The Accumulo CLI toolset is different in Apache Accumulo 2.1.2 and later versions than it was in Accumulo 1.x. As the initial OpDB Accumulo release is based on the Apache Accumulo 2.0.1 version, it is important to learn about these differences.

Before Apache Accumulo™ 2.0 the following toolsets were available:

- accumulo
- accumulo-tool

With Apache Accumulo™ 2.0 the following toolsets were available:

- accumulo
- accumulo-cluster
- accumulo-service
- accumulo-util



Note: Do not use accumulo-cluster and accumulo-service with the OpDB service. These cluster and service manager tools are not compatible with Cloudera Manager. If you use them, your OpDB service can become unavailable.

OpDB system requirements

Before you start using the Accumulo parcel, review the prerequisites and system requirements to ensure all requirements are fulfilled.

- Accumulo depends on HDFS and ZooKeeper libraries and configuration information. Therefore, HDFS and ZooKeeper are mandatory services for OpDB.
- Accumulo TabletServers must be collocated with DataNodes.
- Optionally, you can use Accumulo with YARN (MapReduce 2).

Table 1: OpDB system requirements for Accumulo-on-cdp 2.1.2 parcel

	Supported versions	Note
CDP	CDP Private Cloud Base from 7.1.9	CDP Public Cloud is not supported. CDP Private Cloud Data Services are not supported.
Operating System	RHEL 7, RHEL 8, and RHEL 9 Ubuntu 20.04 Focal Fossa SUSE Linux Enterprise Server 12 or 15	Supported on the operating system with the same minor versions that are supported for the CDP installation with which OpDB is installed.
JDK	Java 11 and Java 17	Supported for the CDP install with which OpDB is installed.

Table 2: OpDB system requirements for Accumulo-on-cdp 1.1.0 parcel

	Supported versions	Note
CDP	CDP Private Cloud Base 7.1.7 SP1 and higher	CDP Public Cloud is not supported. CDP Private Cloud Experiences are not supported.
Operating System	RHEL 7 / RHEL 8 Ubuntu 18.04 Bionic Beaver Ubuntu 20.04 Focal Fossa SUSE Linux Enterprise Server 12	Supported on the operating system with the same minor versions that are supported for the CDP installation with which OpDB is installed.
JDK	Java 8 and Java 11	Supported for the CDP install with which OpDB is installed.



Note: *OpDB Powered by Apache Accumulo 1.1.0* does not run on JDK 17, this scenario is not supported.

Table 3: OpDB system requirements for Accumulo-on-cdp 1.10.0 parcel

	Supported versions	Note
CDP	CDP Private Cloud Base 7.1.7 SP2 and higher	CDP Public Cloud is not supported. CDP Private Cloud Experiences are not supported.
Operating System	RHEL 7 / RHEL 8 Ubuntu 18.04 Bionic Beaver Ubuntu 20.04 Focal Fossa SUSE Linux Enterprise Server 12	Supported on the operating system with the same minor versions that are supported for the CDP installation with which OpDB is installed.
JDK	Java 8 and Java 11	Supported for the CDP install with which OpDB is installed.



Note: *Accumulo 1.10.0* does not run on JDK 17, this scenario is not supported.

Table 4: Hardware requirements for Accumulo parcels

Component	Java Heap	CPU	Disk
Master	Minimum: 1 GB <ul style="list-style-type: none"> 100-1000 tablets: 4 GB 10,000 or more tablets with 200 or more Tablet Servers: 8 GB 10,000 or more tablets with 300 or more Tablet Servers: 12 GB Set this value using the Tablet Server Max Heapsize Accumulo configuration property. See Flume Memory Consumption .	2 Cores. Add more for large clusters or bulk load.	1 disk for local logs

Component	Java Heap	CPU	Disk
Tablet Server	<ul style="list-style-type: none"> Minimum: 512 MB Java Heap Recommended: 8 GB Medium scale production (200 or more tablets per Tablet Server): 16 GB <p>Set this value using the Tablet Server Max Heapsize Accumulo configuration property.</p>	Minimum 4 dedicated cores. You can add more cores for larger clusters, when using replication, or for bulk loads.	<ul style="list-style-type: none"> 4 or more spindles for each HDFS DataNode 1 disk for local logs (this disk can be shared with the operating system and/or other Hadoop logs)
Tracer	<p>1 - 2 GB, depending on cluster workloads.</p> <p>Set this value using the Tracer Max Heapsize Accumulo configuration property.</p>	2 or more dedicated cores, depending on cluster size and workloads	1 disk for local logs, which can be shared with the operating system and/or other Hadoop logs
GC Role	<p>1 - 2 GB</p> <p>Set this value using the Garbage Collector Max Heapsize Accumulo configuration property.</p>	2 or more dedicated cores, depending on cluster size	1 disk for local logs, which can be shared with the operating system and/or other Hadoop logs
Monitor Role	<p>1 - 2 GB</p> <p>Set this value using the Monitor Max Heapsize Accumulo configuration property.</p>	2 or more dedicated cores, depending on cluster size and workloads	1 disk for local logs, which can be shared with the operating system and/or other Hadoop logs