Alibaba Cloud

ApsaraDB for HBase

ApsaraDB for HBase Product

Document Version: 20201207

⟨-⟩ Alibaba Cloud

Legal disclaimer

Alibaba Cloud reminds you to carefully read and fully understand the terms and conditions of this legal disclaimer before you read or use this document. If you have read or used this document, it shall be deemed as your total acceptance of this legal disclaimer.

1. You shall download and obtain this document from the Alibaba Cloud website or other Alibaba Cloud-authorized channels, and use this document for your own legal business activities only. The content of this document is considered confidential information of Alibaba Cloud. You shall strictly abide by the confidentiality obligations. No part of this document shall be disclosed or provided to any third party for use without the prior written consent of Alibaba Cloud.

2. No part of this document shall be excerpted, translated, reproduced, transmitted, or disseminated by any organization, company or individual in any form or by any means without the prior written consent of Alibaba Cloud.

3. The content of this document may be changed because of product version upgrade, adjustment, or other reasons. Alibaba Cloud reserves the right to modify the content of this document without notice and an updated version of this document will be released through Alibaba Cloud-authorized channels from time to time. You should pay attention to the version changes of this document as they occur and download and obtain the most up-to-date version of this document from Alibaba Cloud-authorized channels.

4. This document serves only as a reference guide for your use of Alibaba Cloud products and services. Alibaba Cloud provides this document based on the "status quo", "being defective", and "existing functions" of its products and services. Alibaba Cloud makes every effort to provide relevant operational guidance based on existing technologies. However, Alibaba Cloud hereby makes a clear statement that it in no way guarantees the accuracy, integrity, applicability, and reliability of the content of this document, either explicitly or implicitly. Alibaba Cloud shall not take legal responsibility for any errors or lost profits incurred by any organization, company, or individual arising from download, use, or trust in this document. Alibaba Cloud shall not, under any circumstances, take responsibility for any indirect, consequential, punitive, contingent, special, or punitive damages, including lost profits arising from the use or trust in this document (even if Alibaba Cloud has been notified of the possibility of such a loss).

5. By law, all the contents in Alibaba Cloud documents, including but not limited to pictures, architecture design, page layout, and text description, are intellectual property of Alibaba Cloud and/or its affiliates. This intellectual property includes, but is not limited to, trademark rights, patent rights, copyrights, and trade secrets. No part of this document shall be used, modified, reproduced, publicly transmitted, changed, disseminated, distributed, or published without the prior written consent of Alibaba Cloud and/or its affiliates. The names owned by Alibaba Cloud shall not be used, published, or reproduced for marketing, advertising, promotion, or other purposes without the prior written consent of Alibaba Cloud. The names owned by Alibaba Cloud include, but are not limited to, "Alibaba Cloud", "Aliyun", "HiChina", and other brands of Alibaba Cloud and/or its affiliates, which appear separately or in combination, as well as the auxiliary signs and patterns of the preceding brands, or anything similar to the company names, trade names, trademarks, product or service names, domain names, patterns, logos, marks, signs, or special descriptions that third parties identify as Alibaba Cloud and/or its affiliates.

6. Please directly contact Alibaba Cloud for any errors of this document.

Document conventions

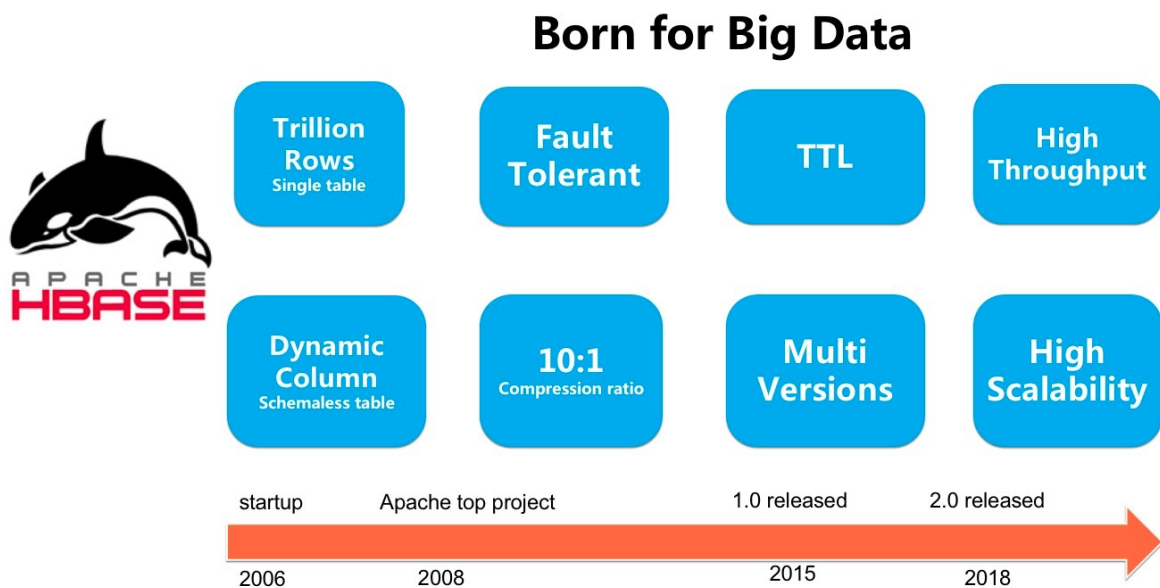| Style	Description	Example
⚠ Danger	A danger notice indicates a situation that will cause major system changes, faults, physical injuries, and other adverse results.	⚠ **Danger:** Resetting will result in the loss of user configuration data.
🔔 Warning	A warning notice indicates a situation that may cause major system changes, faults, physical injuries, and other adverse results.	🔔 **Warning:** Restarting will cause business interruption. About 10 minutes are required to restart an instance.
🔊 Notice	A caution notice indicates warning information, supplementary instructions, and other content that the user must understand.	🔊 **Notice:** If the weight is set to 0, the server no longer receives new requests.
? Note	A note indicates supplemental instructions, best practices, tips, and other content.	? **Note:** You can use Ctrl + A to select all files.
>	Closing angle brackets are used to indicate a multi-level menu cascade.	Click **Settings**> **Network**> **Set network type**.
Bold	Bold formatting is used for buttons , menus, page names, and other UI elements.	Click **OK**.
Courier font	Courier font is used for commands	Run the `cd /d C:/window` command to enter the Windows system folder.
Italic	Italic formatting is used for parameters and variables.	`bae log list --instanceid` *Instance_ID*
[] or [a|b]	This format is used for an optional value, where only one item can be selected.	`ipconfig [-all|-t]`
{} or {a|b}	This format is used for a required value, where only one item can be selected.	`switch {active|stand}`

Table of Contents

1.Overview

Alibaba ApsaraDB for HBase is an end-to-end NoSQL service, which is 100% compatible with open source HBase and deeply optimized and extended. It supports gigabytes (GB) to petabytes (PB) level of data, and is a one-stop NoSQL service that is suitable for processing large-scale throughput, retrieval, and workload analysis. ApsaraDB for HBase is a database service that provides support for many core services of Alibaba such as Taobao recommendation, risk control for Ant Credit Pay, advertising, data dashboard, Cainiao logistics track, Alipay transaction records, and mobile Taobao messages. It provides some enterprise capabilities such as low latency, low cost, full hosting, security, open source standards, global distribution.

ApsaraDB for HBase comprehensively provides capabilities such as real-time storage, high concurrency throughput, light SQL analysis (integrated with Phoenix), full-text index (integrated with Solr), time series query (integrated with OpenTSDB), and spatio temporal query (integrated with Geomesa) for massive semi-structured or structured data. Combined with comprehensive tools and services and rich ecosystem integration, ApsaraDB for HBase provides an end-to-end and efficient solution to meet the storage, retrieval, and analysis needs of enterprises in big data scenarios. ApsaraDB for HBase is the preferred database for scenarios such as risk control, recommendation, advertising, Internet of Things (IoT), Internet of Vehicles, feeds streaming, and data dashboard.

HBase: one of the most popular open source big data storage databases

Overview of ApsaraDB for HBase

Alibaba Cloud HBase

Low Cost, High Scalability, Intelligent NoSQL Services

Compatible with standard HBase protocol, having the core advantages of **low cost storage, high throughput, and intelligent data processing** . It is the preferred database for scenarios such as **recommendation, advertising, risk control, IoT, feed streams**. And also the same database that powers many core Alibaba services, including Tmall, Alipay, Cainiao.

Low Cost

With built-in high-performance engine, the throughput is 7x of open source HBase, data compression ratio up to 10. Provides super cheap storage media.

Elasticity

Allow you elastically scale storage to 10PB and throughput to 10 million ops

Low Latency

Single-digit millisecond latency, The P99 latency is only 1/10 of open source HBase

High Availability

Master-follower dual clusters, 99.99% SLA, strong consistency across data centers

Security

Support VPC, whitelist, authentication, ACL, audit.

Search

With built-in search engine, allow you do the full-text search, multi-dimensional query.

SQL Support

SQL interfaces is provided, allow you use it for the lightweight analysis

Easy-using

Support multi data model , such as time-series, spatio-temporal , queue.
Easy to migrate data with our provided service BDS

Fully Managed

No database management operation is required by yourself. And support pay-as-you-go with Serverless model
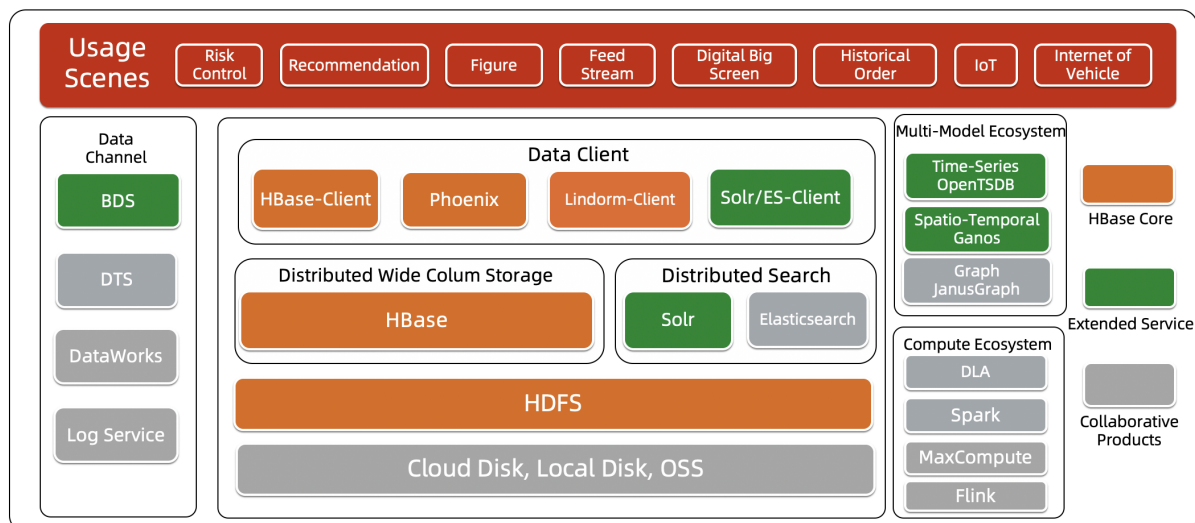
Proven

10-years of use in Alibaba core service, have a scale of more than 10,000 nodes

Architecture of ApsaraDB for HBase

Low cost, low latency, open source compatible
Big data NoSQL service

| Usage Scenes | Risk Control | Recommendation | Figure | Feed Stream | Digital Big Screen | Historical Order | IoT | Internet of Vehicle |

Data Channel
- BDS
- DTS
- DataWorks
- Log Service

Data Client
- HBase-Client
- Phoenix
- Lindorm-Client
- Solr/ES-Client

Distributed Wide Colum Storage
- HBase

Distributed Search
- Solr
- Elasticsearch

HDFS

Cloud Disk, Local Disk, OSS

Multi-Model Ecosystem
- Time-Series OpenTSDB
- Spatio-Temporal Ganos
- Graph JanusGraph

HBase Core

Compute Ecosystem
- DLA
- Spark
- MaxCompute
- Flink

Extended Service

Collaborative Products

Massive data storage, real-time storage, high concurrency, full-text search, lightweight SQL analysis

Scenarios

Type	Storage object	Engine API	Typical scenario	Benefit
KV	key and value	HBase API	Social distribution and chat records	Low latency and high concurrency

Type	Storage object	Engine API	Typical scenario	Benefit
SQL	Relational model	Phoenix SQL	Upgrades of traditional relational databases	Large storage capacity and distributed SQL database
Table storage	Sparse table	HBase API	Risk control and user profiling	Dynamic schema and large storage capacity
Document storage	JSON, XML, and HTML files	HBase API	News	Large storage capacity and high concurrency
Object storage	Images and videos	HBase API	Images and videos	Large storage capacity
Time series data	Continuous temporal data in time dimension	OpenTSDB/HBase API	Sensors and Internet of Things (IoT)	Large storage capacity
Spatio temporal data	Spatial dimension or spatial trajectory	GeoMesa/HBase API	Internet of Vehicles and sports apps	Large storage capacity and high-performance write capacity
OLAP (On-Line Analytical Processing)	Volumetric data	Spark/HBase API	Reports	Large storage capacity, high concurrency, high load throughput, and HFile load
Full text	Full text	Solr/HBase API	Full-text search	High throughput

Benefits

For more information, see Benefits.

2.Benefits

Comparison with self-build HBase

	Alibaba Cloud ApsaraDB for HBase
Latency	The P99 latency of ApsraDB for HBase is one tenth of the open source edition. For more information, see Performance Report.
Performance	I/O performance: three to seven times of the open source edition. For more information, see Performance Report.
Costs	Exclusively supports for ZSTD compression (double compression rate), OSS cold storage, hot and cold separation, elastic scaling, and other capabilities. The total cost can be as low as 50% of fthe costs on user-created HBase databases.
Service mode	Fully managed
Feature extensions	Provides SQL light-weight analysisfull-text search, hot and cold separation, vertical data model, and many other capabilities. This can comprehensively meet the storage, search, and analysis needs of enterprises in volumetric data scenarios.
Reliability	Provides an SLA for features such as multiple disaster recovery capabilities including primary/secondary failover, active-active, and backup.
Elasticity	Supports scale-up: Upsizing and disk expansion; Supports scale out: Productization scaling nodes; Supports Serverless: Automatically scales without the need to sense resource usage.
Ecosystem capability	Open source ecosystem and Alibaba Cloud ecosystem, which made deep production integration with MySQL, Spark, OSS, and MaxCompute.
Data migration	Supports hot migration from any edition of user-created HBase instances and environments to ApsaraDB for HBase. The entire process is automated, efficient, has no impacts on services, and does not require any code changes.
Security	Supports simple account and password authentication and access control lists, as well as network isolation and whitelist capabilities.

	Alibaba Cloud ApsaraDB for HBase
Maturity	ApsaraDB for HBase has been tested inside Alibaba Group for over 10 years, and provides services to tens of thousands of nodes. ApsaraDB for HBase also has also provided support for the core applications Tmall Double 11 for nine years.

Comparison with traditional databases

		Alibaba Cloud HBase	Traditional DB（MySQL/PG）
Scale	Storage scale	0~10PB	0~1TB (suitable)
	Throughput scale	0~100,000,000 ops	0~100,000 ops(suitable)
	Row count per table	0~1 Trillion	0~1 Billion
	Column count per table	0~100,000	0~100
	Row size	0~10MB	0~8KB
Availability	Response Time	Single-digit millisecond	Single-digit millisecond
	SLA	99.9%(1 datacenter) , 99.99%(2 datacenters)	99.99%(2 datacenters)
	MTTR	2~5 mins	<1 min
Cost	Storage media	Local HDD , OSS , Efficiency cloud disk , Local SSD , SSD Cloud Disk	Local SSD , SSD Cloud Disk
	Data compression	On by default, the ration up to 10	Off by default, low compression ratio
	Throughput per node	Up to hundreds of thousands QPS	Tens of thousands QPS
Features	transaction	Weak	Strong
	Standard SQL	Support	Support
	Index	Support	Support
	Free Schema	Strong , dynamic column、TTL、multi-family	Weak
Ecosystem	Data visualization	Weak	Good
	Big data ecosystem	Good at connecting with Spark、Hadoop	Weak

3.Professional service

If you want to learn more about ApsaraDB for HBase or have any questions about it, use the following
methods:Please contact customer service