

Anarion Technologies – Apache CarbonData

Note: Search product on Azure marketplace and click on "Get it now"

 Microsoft | Marketplace

Products ▾ 🔍 Search Marketplace

≡ All Products ▾ Categories ▾ Industries ▾ Partners ▾

Products > Apache CarbonData on Ubuntu26



Apache CarbonData on Ubuntu26

by Anarion Technologies

[Virtual Machines](#)

Starting at \$0.03 [Get it now](#) ☆ Save to my list

[Overview](#) [Plans + Pricing](#) [Ratings + reviews](#) [Details + support](#)

Production Ready VM on Ubuntu26.04 LTS + Free Support

Apache CarbonData on Ubuntu 26.04 is a high-performance open-source columnar storage format designed for big data analytics and large-scale data processing. Built to integrate seamlessly with Apache Spark, CarbonData enhances query performance through advanced indexing, data compression, and efficient storage management. Running on Ubuntu 26.04 LTS, this virtual machine provides a stable and production-ready environment for deploying, testing, and evaluating Apache CarbonData.

Apache CarbonData enables organizations to store massive volumes of structured and semi-structured data while significantly reducing storage requirements through optimized compression techniques. Its intelligent indexing mechanisms, including blocklet indexes and data maps, help accelerate analytical queries and improve overall processing efficiency compared to traditional storage formats.

Designed for modern data lake and data warehouse architectures, CarbonData works natively with Apache Spark to

Click on Continue

Confirm your details to continue



Apache CarbonData on Ubuntu26

By Anarion Technologies

Plan

Apache CarbonData on Ubuntu26



Pricing

Starting at \$0.03/hour

Description

Apache CarbonData on Ubuntu26

You're signed in as **Market Place Products** (mp@anariontech.com).

[Edit your details](#)

 By clicking **Get it now**, I give Microsoft permission to use or share my account information so that the provider can contact me regarding this product. I agree to the provider's [terms of use](#) and [privacy policy](#) and understand that the rights to use this product do not come from Microsoft, unless Microsoft is the provider. Use of Marketplace is governed by separate [terms](#) and [privacy](#).

Get it now

Click on Create

Home

Apache CarbonData on Ubuntu26

Anarion Technologies



Apache CarbonData on Ubuntu26

[Add to Favorites](#)

Anarion Technologies | Virtual Machine

Subscription	Plan		
Azure subscription 1	Apache CarbonData on Ubuntu26	Create	Start with a pre-set configuration

Want to deploy programmatically? [Get started](#)

[Overview](#) [Plans + Pricing](#) [Usage Information + Support](#) [Ratings + Reviews](#)

Apache CarbonData on Ubuntu 26.04 is a high-performance open-source columnar storage format designed for big data analytics and large-scale data processing. Built to integrate seamlessly with Apache Spark, CarbonData enhances query performance through advanced indexing, data compression, and efficient storage management. Running on Ubuntu 26.04 LTS, this virtual machine provides a stable and production-ready environment for deploying, testing, and evaluating Apache CarbonData.

Apache CarbonData enables organizations to store massive volumes of structured and semi-structured data while significantly reducing storage requirements through optimized compression techniques. Its intelligent indexing mechanisms, including blocklet indexes and data maps, help accelerate analytical queries and improve overall processing efficiency compared to traditional storage formats.

Designed for modern data lake and data warehouse architectures, CarbonData works natively with Apache Spark to provide high-speed data ingestion, faster query execution, and scalable analytics capabilities. Organizations can leverage CarbonData to process large datasets efficiently while reducing infrastructure costs and improving business intelligence performance.

One of CarbonData's key advantages is its ability to optimize read and write operations through columnar storage, predicate pushdown, and indexing technologies. These features help data engineers and analysts achieve lower query latency and better resource utilization across enterprise-scale analytics environments.

Creating a virtual machine, enter or select appropriate values for zone, machine type, resource group and so on as per your choice.

» **Create a virtual machine** ...

[Basics](#) [Disks](#) [Networking](#) [Management](#) [Advanced](#) [Tags](#) [Review + create](#)

Create a virtual machine that runs Linux or Windows. Select an image from Azure marketplace or use your own customized image. Complete the Basics tab then Review + create to provision a virtual machine with default parameters or review each tab for full customization. [Learn more](#)

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ Azure subscription 1

Resource group * ⓘ Demo [Create new](#)

Instance details

Virtual machine name * ⓘ Demo ✓

Region * ⓘ (US) East US

[Review + create](#) [< Previous](#) [Next : Disks >](#)

Create a virtual machine ...

Basics Disks Networking Management Advanced Tags Review + create

Azure VMs have one operating system disk and a temporary disk for short-term storage. You can attach additional data disks. The size of the VM determines the type of storage you can use and the number of data disks allowed. [Learn more](#)

Disk options

OS disk type * ⓘ Premium SSD (locally-redundant storage) ▼

Encryption type * (Default) Encryption at-rest with a platform-managed key ▼

Enable Ultra Disk compatibility ⓘ ☐
Ultra disk is available only for Availability Zones in eastus.

Data disks

You can add and configure additional data disks for your virtual machine or attach existing disks. This VM also comes with a temporary disk.

i Adding unmanaged data disks is currently not supported at the time of VM creation. You can add them after the VM is

Review + create

< Previous

Next : Networking >

Create a virtual machine ...

Basics Disks Networking Management Advanced Tags Review + create

Define network connectivity for your virtual machine by configuring network interface card (NIC) settings. You can control ports, inbound and outbound connectivity with security group rules, or place behind an existing load balancing solution. [Learn more](#)

Network interface

When creating a virtual machine, a network interface will be created for you.

Virtual network * ⓘ (new) Demo-vnet ▼
[Create new](#)

Subnet * ⓘ (new) default (10.1.0.0/24) ▼

Public IP ⓘ (new) Demo-ip ▼
[Create new](#)

NIC network security group ⓘ ☐ None
☒ Basic

Review + create

< Previous

Next : Management >

Create a virtual machine ...

Basics Disks Networking Management Advanced Tags Review + create

Configure monitoring and management options for your VM.

Azure Security Center

Azure Security Center provides unified security management and advanced threat protection across hybrid cloud workloads. [Learn more](#)

✓ Your subscription is protected by Azure Security Center basic plan.

Monitoring

Boot diagnostics ⓘ ☐ On
☒ Off

Enable OS guest diagnostics ⓘ ☐

Identity

Review + create

< Previous

Next : Advanced >

Create a virtual machine ...

Basics Disks Networking Management Advanced Tags Review + create

Tags are name/value pairs that enable you to categorize resources and view consolidated billing by applying the same tag to multiple resources and resource groups. [Learn more about tags](#)

Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.

Name ⓘ	Value ⓘ	Resource
	:	12 selected ▼

Review + create

< Previous

Next : Review + create >

[Home](#) >

Create a virtual machine ...

✓ Validation passed

Basics Disks Networking Management Monitoring Advanced Tags **[Review + create](#)**

i Cost given below is an estimate and not the final price. For all your pricing needs, please use the pricing calculator. [↗](#)

Price

1 X Standard B1s
by Microsoft
[Terms of use](#) | [Privacy policy](#)

Subscription credits apply ⓘ

0.8652 INR/hr
[Pricing for other VM sizes](#)

TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

[< Previous](#)

[Next >](#)

Create

After Process of Create Virtual Machine. You have got an Option Go to Resource Group

Click Go to Resource Group

Copy the Public IP Address

SSH into Terminal and Run these Commands:

```
$ sudo su
```

```
$ cd ../../
```

```
$ sudo apt update && sudo apt upgrade -y
```



```

testing@testing:~$ sudo su
root@testing:/home/testing# cd ../../
root@testing:/# sudo apt update && sudo apt upgrade -y
Hit:1 http://azure.archive.ubuntu.com/ubuntu resolute InRelease
Get:2 http://azure.archive.ubuntu.com/ubuntu resolute-updates InRelease [137
kB]
Hit:3 http://azure.archive.ubuntu.com/ubuntu resolute-backports InRelease
Get:4 http://azure.archive.ubuntu.com/ubuntu resolute-security InRelease [13
7 kB]
Get:5 https://aquasecurity.github.io/trivy-repo/deb resolute InRelease [3238
B]
Get:6 http://azure.archive.ubuntu.com/ubuntu resolute-updates/main amd64v3 P
ackages [282 kB]
Get:7 http://azure.archive.ubuntu.com/ubuntu resolute-updates/main Translati
on-en [74.0 kB]
Get:8 http://azure.archive.ubuntu.com/ubuntu resolute-updates/main amd64 Com
ponents [49.4 kB]
Get:9 http://azure.archive.ubuntu.com/ubuntu resolute-updates/main amd64 c-n
-f Metadata [5244 B]
Get:10 http://azure.archive.ubuntu.com/ubuntu resolute-updates/universe amd6
4v3 Packages [159 kB]
Get:11 http://azure.archive.ubuntu.com/ubuntu resolute-updates/universe Tran
slation-en [50.1 kB]
Get:12 http://azure.archive.ubuntu.com/ubuntu resolute-updates/universe amd6
4v3 Packages [159 kB]

```

Verify CarbonData JAR

\$ find /opt/carbondata-parent-2.3.2 -name "carbondata-spark*.jar"

```

root@testing:/# find /opt/carbondata-parent-2.3.2 -name "carbondata-spark*.j
ar"
/opt/carbondata-parent-2.3.2/integration/spark/target/carbondata-spark_3.1-2
.3.2-tests.jar
/opt/carbondata-parent-2.3.2/integration/spark/target/carbondata-spark_3.1-2
.3.2.jar
root@testing:/#

```

Verify Spark Installation

\$ spark-shell --version

Verify CarbonData Provider

```
> spark.sql("""
```

```
CREATE TABLE carbon_test(
```

```
id INT,
```

```
name STRING
```

```
)
```

```
USING carbon
```

```
""")
```

```
scala> spark.sql("""
  CREATE TABLE carbon_test(
  id INT,
  name STRING
  )
  USING carbon
  """)
26/06/26 16:24:10 WARN HiveConf: HiveConf of name hive.stats.jdbc.timeout does not exist
26/06/26 16:24:10 WARN HiveConf: HiveConf of name hive.stats.retries.wait does not exist
26/06/26 16:24:20 WARN ObjectStore: Version information not found in metastore. hive.metastore.schema.validation is not enabled so recording the schema version 2.3.0
26/06/26 16:24:20 WARN ObjectStore: setMetaStoreSchemaVersion called but recording version is disabled: version = 2.3.0, comment = Set by MetaStore root@10.2.0.4
26/06/26 16:24:20 WARN ObjectStore: Failed to get database default, returning NoSuchObjectException
26/06/26 16:24:21 WARN HiveExternalCatalog: Couldn't find corresponding Hive SerDe for data source provider carbon. Persisting data source table `spark_catalog`.`default`.`carbon_test` into Hive metastore in Spark SQL specific format, which is NOT compatible with Hive.
26/06/26 16:24:22 WARN SessionState: METASTORE_FILTER_HOOK will be ignored, since hive.security.authorization.manager is set to instance of HiveAuthorizerFactory.
26/06/26 16:24:23 WARN HiveConf: HiveConf of name hive.internal.ss.authz.settings.applied.marker does not exist
26/06/26 16:24:23 WARN HiveConf: HiveConf of name hive.stats.jdbc.timeout does not exist
26/06/26 16:24:23 WARN HiveConf: HiveConf of name hive.stats.retries.wait does not exist
res1: org.apache.spark.sql.DataFrame = []
```

Verify Table Creation

```
$ spark.sql("SHOW TABLES").show(false)
```

```
scala> spark.sql("SHOW TABLES").show(false)
26/06/26 16:24:57 WARN ObjectStore: Failed to get database global_temp, returning NoSuchObjectException
+-----+-----+-----+
|namespace|tableName  |isTemporary|
+-----+-----+-----+
|default  |carbon_test|false      |
+-----+-----+-----+
```

Verify Table Metadata

```
$ spark.sql("DESCRIBE FORMATTED carbon_test").show(false)
```

```
scala> spark.sql("DESCRIBE FORMATTED carbon_test").show(false)
+-----+-----+-----+
|col_name      |data_type      |comment|
+-----+-----+-----+
|id             |int            |NULL   |
|name          |string         |NULL   |
|# Detailed Table Information|
|Catalog       |spark_catalog  |
|Database      |default        |
|Table         |carbon_test    |
|Owner         |root           |
|Created Time   |Fri Jun 26 16:24:23 UTC 2026|
|Last Access    |UNKNOWN        |
|Created By     |Spark 3.5.1    |
|Type          |MANAGED        |
|Provider       |carbon         |
|Location       |file:/spark-warehouse/carbon_test|
|Serde Library  |org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe|
|InputFormat    |org.apache.hadoop.mapred.SequenceFileInputFormat|
|OutputFormat   |org.apache.hadoop.hive.ql.io.HiveSequenceFileOutputFormat|
+-----+-----+-----+
```

```
$ spark.sql("SHOW CREATE TABLE carbon_test").show(false)
```

```
scala> spark.sql("SHOW CREATE TABLE carbon_test").show(false)
+-----+
|createtab_stmt|
+-----+
|CREATE TABLE spark_catalog.default.carbon_test (\n  id INT,\n  name STRING)\nUSING carbon\n|
+-----+
```

```
$ spark.catalog.listTables().show(false)
```

```
scala> spark.catalog.listTables().show(false)
```

name	catalog	namespace	description	tableType	isTemporary
carbon_test	spark_catalog	[default]	NULL	MANAGED	false

```
$ spark.catalog.tableExists("carbon_test")
```

Expected:

```
resX: Boolean = true
```

```
scala> spark.catalog.tableExists("carbon_test")  
res15: Boolean = true
```

Thanks!!!