Apache Bigtop on openEuler

徐国栋

Linaro 资深技术主管

开放原子开源基金会 OPENATOM FOUNDATION | OpenEuler

自我介绍

徐国栋

Linaro, 高级技术主管

个人主页:
guodong.xu@linaro.org
https://github.com/docularxu

技术方向:

大数据，Apache Bigtop
Arm SVE/SVE2优化，进行软件库优化
OpenSSL 3.0+
Linux内核设备驱动程序开发

关于 linaro.org :
Linaro works with businesses and open source communities to develop software on Arm-based technology. We create solutions that drive forward the Arm software ecosystem, enhance standardization, promote collaboration across industries and contribute to real-world applications.

Agenda

- Apache Bigtop 介绍
- openEuler Bigdata SIG
- 创新项目：Apache Bigtop on openEuler
- Apache Bigtop 未来计划

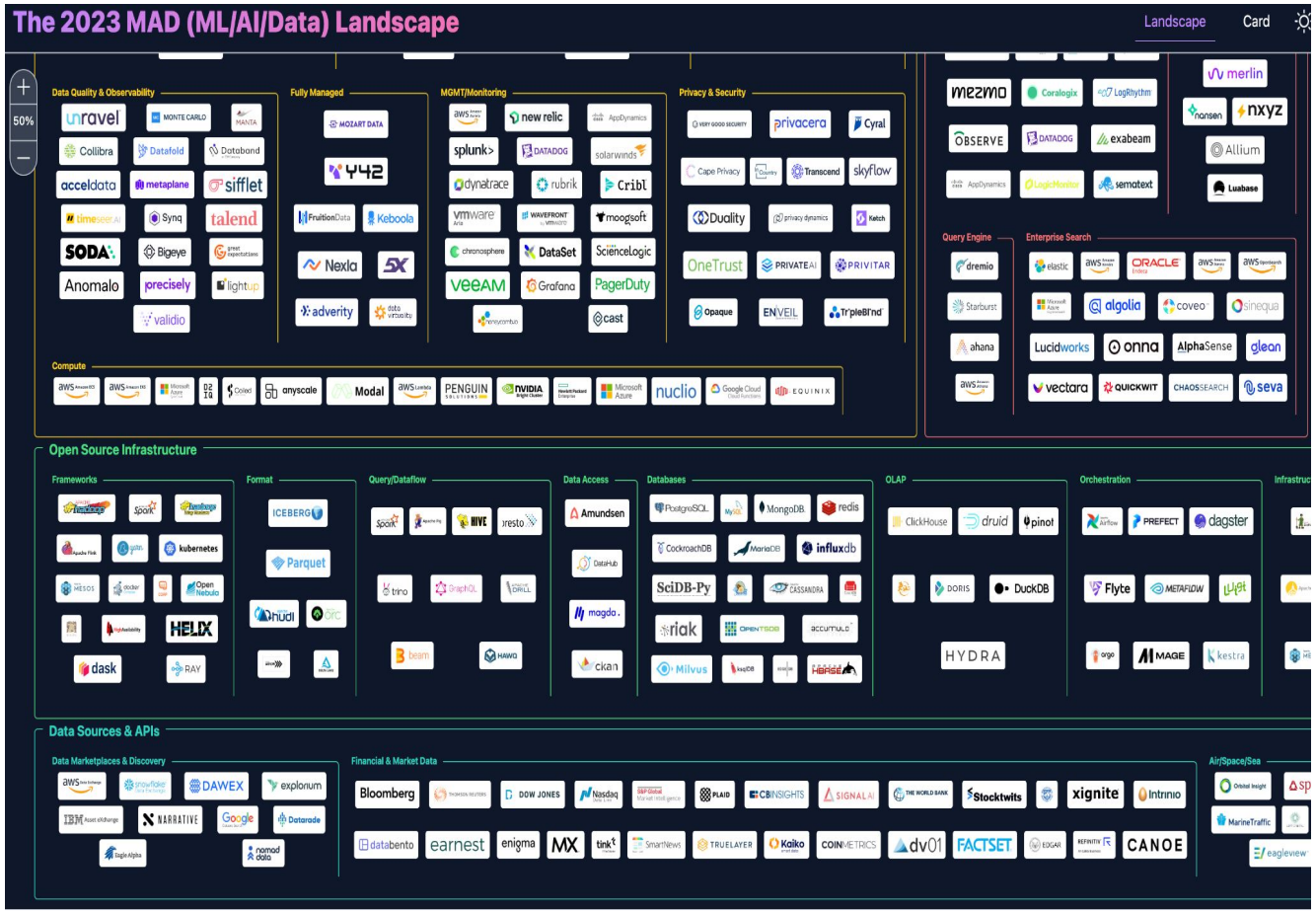
Apache Bigtop 介绍

Big Data and Data Science Ecosystem

Big Data in itself is a huge ecosystem. It is just too large, complex and redundant. It has too many standards, too many engines, too many vendors.

Categorizing Big Data Components

- Core Components
- Operational Components
- Data Ingestion
- Streaming / In-Memory Processing
- Data Warehousing
- Messaging System
- NoSQL
- File formats
- Dashboards and Management
- Security/Governance
- Data Science Tools / Machine Learning Components
- Notebooks



https://mad.firstmark.com/

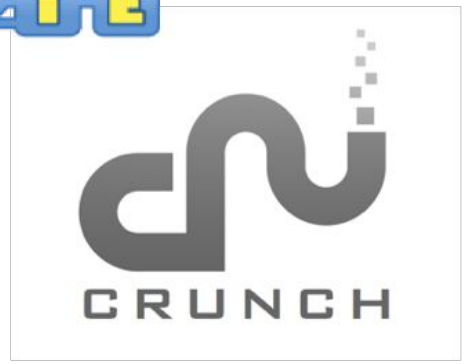
What is Apache Bigtop ?

Apache Bigtop is the only open source Big Data distro in the market today. It provides full functionality of packaging, testing, and deployment for users to build their own Hadoop Distro from ground up.

Bigtop Benefits:

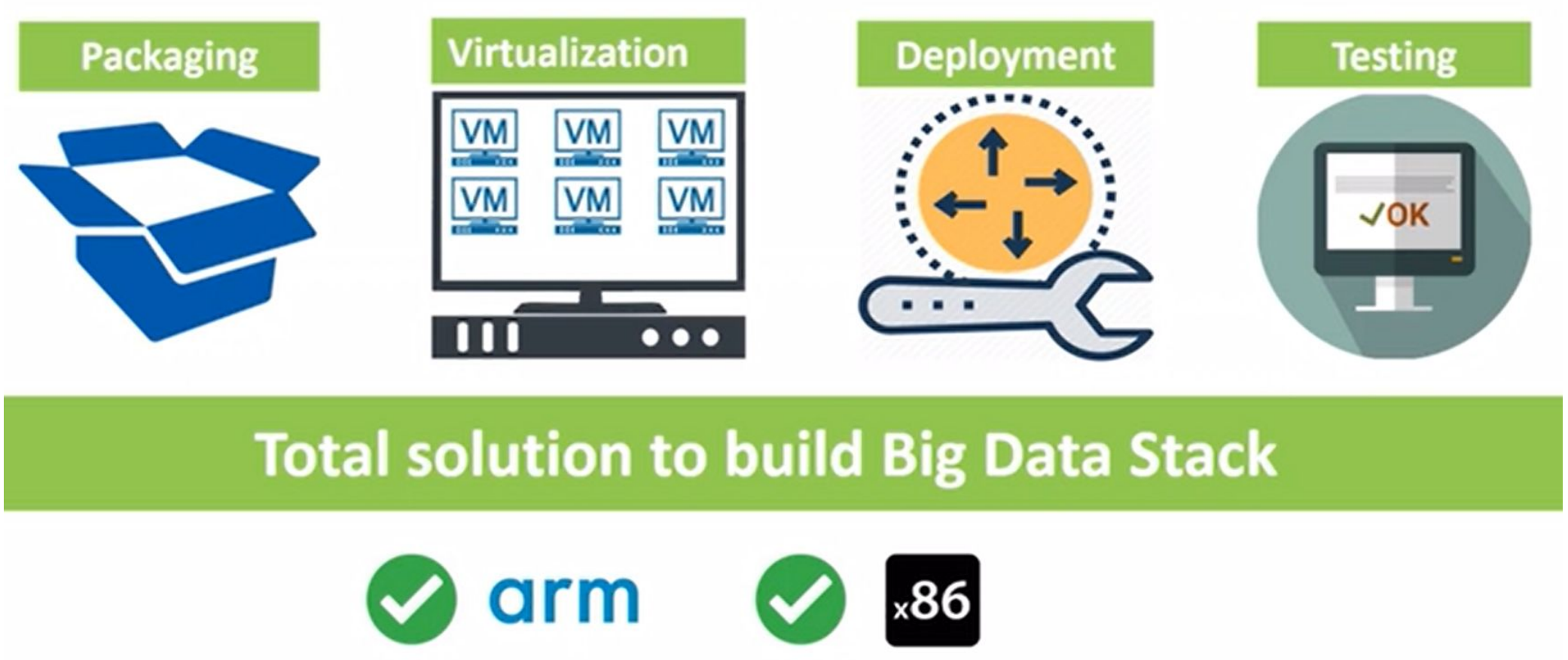
- Comprehensive list of component support compared to any other closed source
- Support for Multiple Architectures
- Only distro with **ARM64** Support
- Build and **packaging** code
- **Deployment** code
- VM and bundle creation code
- Easy **provisioning** and incremental updates
- Configuration for integrated ecosystem
- Blueprints
- Integration tests for all of the various Hadoop ecosystem components and package artifacts themselves
- Smoke tests for all the various components of Hadoop ecosystem
- Packaged distribution of the Hadoop big data management stack for various **Linux-based Distros**
- Seamless Integration into CI

Supports 25+ Hadoop Ecosystem Projects



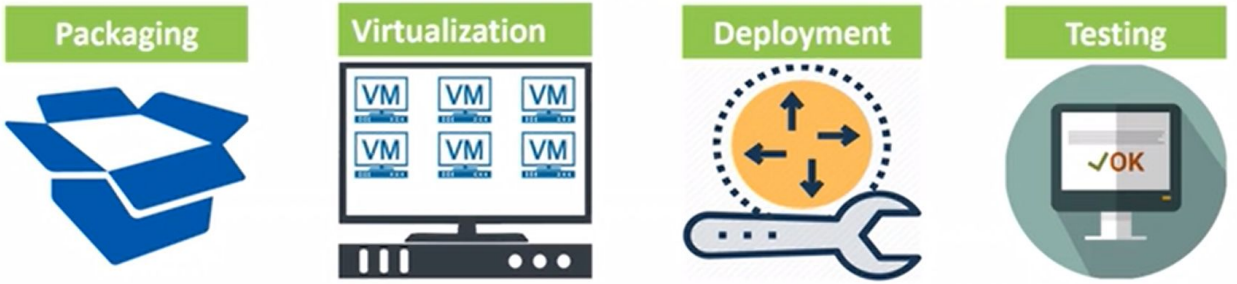
X

Bigtop Feature Set



Bigtop Feature Set

- Package Hadoop ecosystem to RPM/DEB artifacts
- Support 25+ Hadoop Ecosystem Components

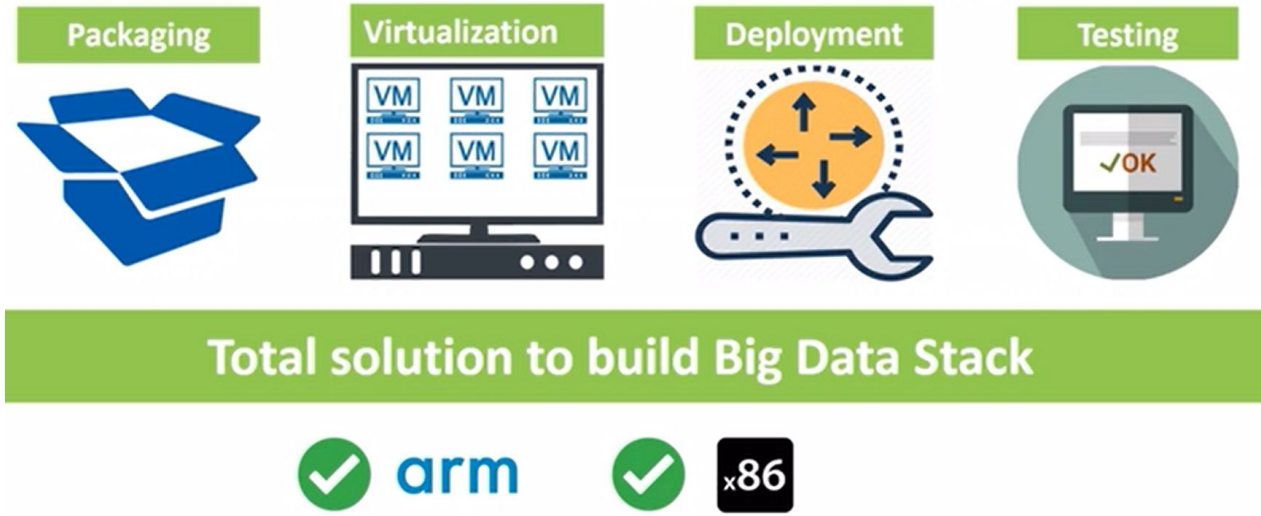


Total solution to build Big Data Stack

arm x86

Bigtop Feature Set

- Package Hadoop ecosystem to RPM/DEB artifacts
- Support 25+ Hadoop Ecosystem Components

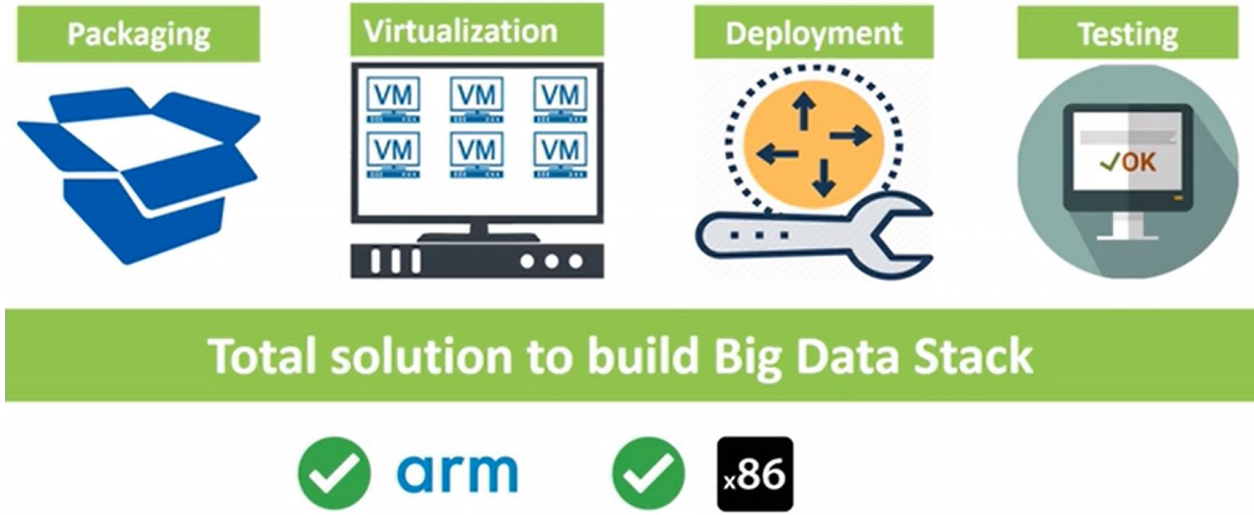


Starting from Bigtop 3.2: One-Click deploying Bigtop stack by Ambari

关于使用 Ambari 部署Bigtop的更多内容，欢迎大家移步 Linaro 网站继续学习：
https://resources.linaro.org/en/search?q=LVC21F-202
“BIGTOP 3.0 with the upgraded Mpack: New era of BigData Distribution ”，By Gu Yuqi

Bigtop Feature Set

- Package Hadoop ecosystem to RPM/DEB artifacts
- Support 25+ Hadoop Ecosystem Components



Starting from Bigtop 3.2: One-Click deploying Bigtop stack by Ambari

Supported OS : **Debian, Ubuntu, CentOS, Fedora, openSUSE …**

\+ **openEuler OS**

Commercial Apps / Services using Bigtop

cz.nic
TREND MICRO
Qubole
sas
WIKIMEDIA FOUNDATION
VISA
SBERBANK
DB BAHN
LINAGORA
NAVER Cloud
wanDISCO
STRATOSCALE
TERRADATUM
Persistent
VERTICA
GridGain
XIIIab
altiscale
ampool IN-MEMORY COMPUTING
Capgemini
DATATORRENT
splunk>
TERADATA
PLDT
zData
Zettaset
CANONICAL

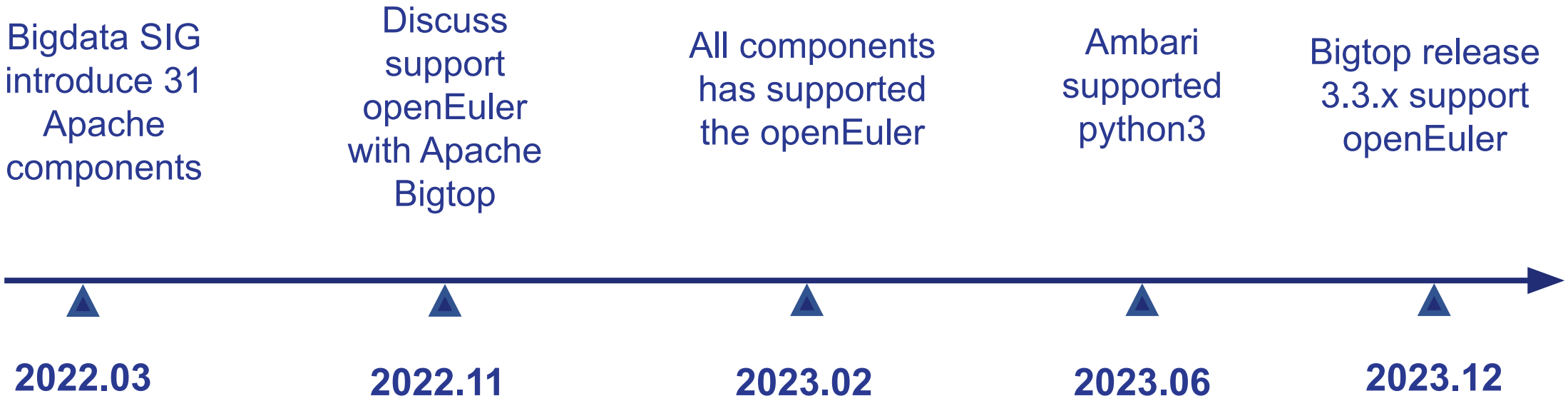
openEuler Bigdata SIG

openEuler Bigdata SIG 愿景

- Build and improve the bigdata ecosystem of openEuler community
- Create the technology exchange and innovation platform of bigdata
- Unleash the ultimate performance of bigdata components by mining software and hardware capabilities

为社区开发者，为大数据用户，提供交流创新的平台。

与上游社区 Bigtop 的互动



Linaro
Developer Services

Support openEuler on Bigtop

- **Support openEuler OS steps:**
  1. Create the openEuler docker for puppet and slaves
  2. Modify the component's code for support openEuler
  3. Compile all components on openEuler
  4. Smoke test all components

- **The code support openEuler OS has merged into the master branch.**

OpenEuler 22.03 software version	
Name	Version
maven	3.6.3-1.oe2203
cmake	3.22.0-4.oe2203
rpm-build	4.17.0-9.oe2203
fuse-devel	2.9.9-9.oe2203
lzo-devel	2.10-1.oe2203
openssl-devel	1.1.1m-8.oe2203
openEuler-rpm-config	30-23.oe2203
unzip	6.0-48.oe2203
jdk	1.8.0.342.b07-0.oe2203
git	2.33.0-3.oe2203
autoconf	2.71-2.oe2203
curl	7.79.1-9.oe2203
ruby	3.0.3-124.oe2203
net-tools	2.10-1.oe2203
pkgconfig	1.5.1-2.oe2203
gcc	10.3.1-11.oe2203
g++	10.3.1-11.oe2203
protobuf	3.14.0-4.oe2203
python3-crypto	2.6.1-28.oe2203
doxygen	1.9.2-4.oe2203
cyrus-sasl	2.1.27-14.oe2203
gyp	0.1-1.oe2203
python3	**3.9.9-8.oe2203**
node-gyp	3.6.0-4.oe2203

Bigtop 3.2+

在 openEuler 平台上的

部署+测试

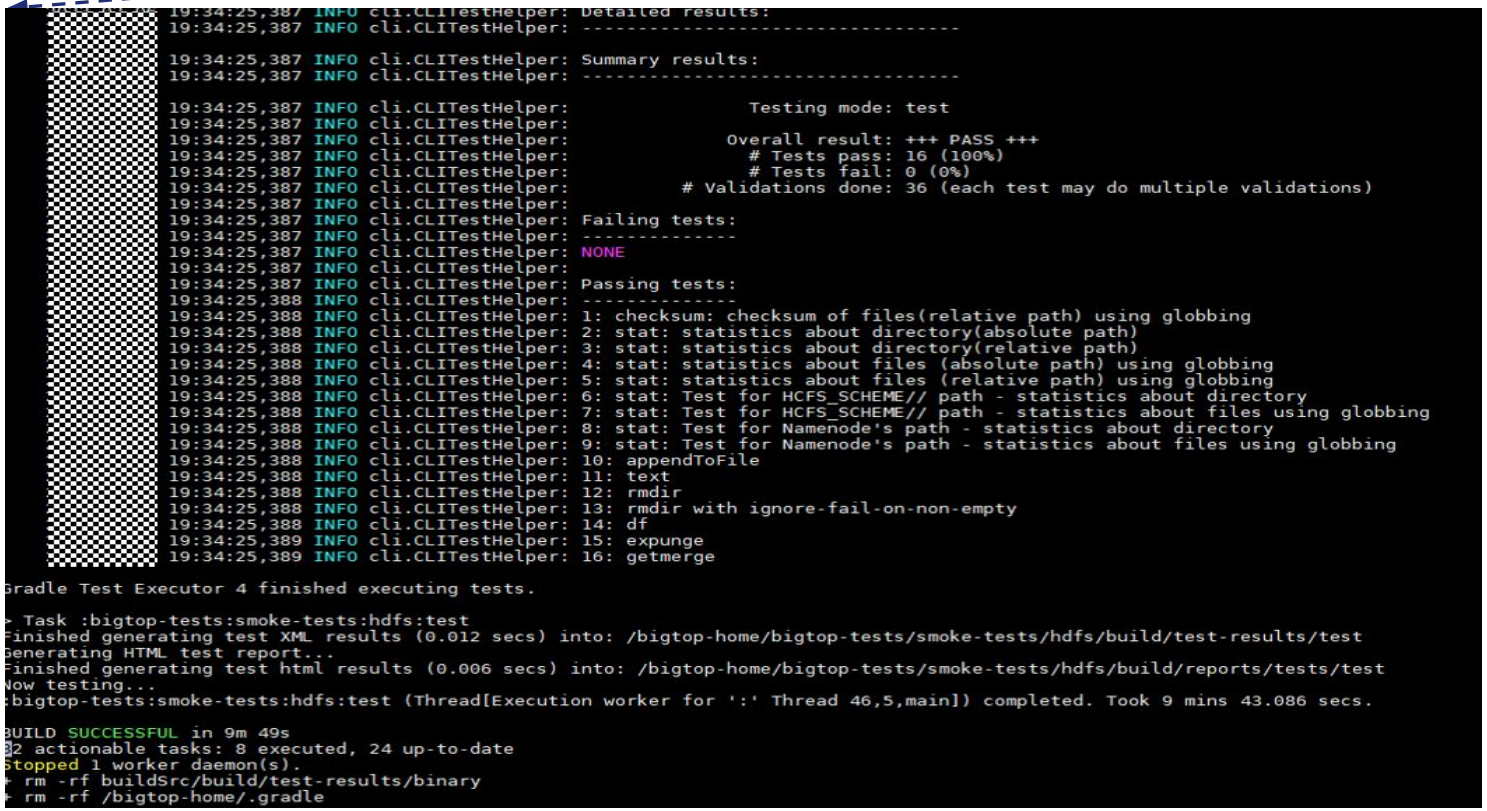
Docker-based Deployment on openEuler

./docker-hadoop.sh --create 3

./docker-hadoop.sh -s hdfs

Docker config → Create docker instance → Smoke Test → Result

```
docker:
    image: "bigtop/puppet:trunk-openEuler-22.03"
    repo: "file:///bigtop-home/output/"
distro: openEuler
components: [hdfs, yarn, mapreduce]
enable_local_repo: true
smoke_test_components: [hdfs]
```



Bigtop deploys Hadoop on openEuler

```
19:34:25,387 INFO cli.CLITestHelper: Detailed results:
19:34:25,387 INFO cli.CLITestHelper: ---------------------------------

19:34:25,387 INFO cli.CLITestHelper: Summary results:
19:34:25,387 INFO cli.CLITestHelper: ---------------------------------

19:34:25,387 INFO cli.CLITestHelper:                 Testing mode: test
19:34:25,387 INFO cli.CLITestHelper:
19:34:25,387 INFO cli.CLITestHelper:               Overall result: +++ PASS +++
19:34:25,387 INFO cli.CLITestHelper:                 # Tests pass: 16 (100%)
19:34:25,387 INFO cli.CLITestHelper:                 # Tests fail: 0 (0%)
19:34:25,387 INFO cli.CLITestHelper:           # Validations done: 36 (each test may do multiple validations)
19:34:25,387 INFO cli.CLITestHelper:
19:34:25,387 INFO cli.CLITestHelper: Failing tests:
19:34:25,387 INFO cli.CLITestHelper: --------------
19:34:25,387 INFO cli.CLITestHelper: NONE
19:34:25,387 INFO cli.CLITestHelper:
19:34:25,387 INFO cli.CLITestHelper: Passing tests:
19:34:25,388 INFO cli.CLITestHelper: --------------
19:34:25,388 INFO cli.CLITestHelper: 1: checksum: checksum of files(relative path) using globbing
19:34:25,388 INFO cli.CLITestHelper: 2: stat: statistics about directory(absolute path)
19:34:25,388 INFO cli.CLITestHelper: 3: stat: statistics about directory(relative path)
19:34:25,388 INFO cli.CLITestHelper: 4: stat: statistics about files (absolute path) using globbing
19:34:25,388 INFO cli.CLITestHelper: 5: stat: statistics about files (relative path) using globbing
19:34:25,388 INFO cli.CLITestHelper: 6: stat: Test for HCFS_SCHEME// path - statistics about directory
19:34:25,388 INFO cli.CLITestHelper: 7: stat: Test for HCFS_SCHEME// path - statistics about files using globbing
19:34:25,388 INFO cli.CLITestHelper: 8: stat: Test for Namenode's path - statistics about directory
19:34:25,388 INFO cli.CLITestHelper: 9: stat: Test for Namenode's path - statistics about files using globbing
19:34:25,388 INFO cli.CLITestHelper: 10: appendToFile
19:34:25,388 INFO cli.CLITestHelper: 11: text
19:34:25,388 INFO cli.CLITestHelper: 12: rmdir
19:34:25,388 INFO cli.CLITestHelper: 13: rmdir with ignore-fail-on-non-empty
19:34:25,388 INFO cli.CLITestHelper: 14: df
19:34:25,389 INFO cli.CLITestHelper: 15: expunge
19:34:25,389 INFO cli.CLITestHelper: 16: getmerge

Gradle Test Executor 4 finished executing tests.

> Task :bigtop-tests:smoke-tests:hdfs:test
Finished generating test XML results (0.012 secs) into: /bigtop-home/bigtop-tests/smoke-tests/hdfs/build/test-results/test
Generating HTML test report...
Finished generating test html results (0.006 secs) into: /bigtop-home/bigtop-tests/smoke-tests/hdfs/build/reports/tests/test
Now testing...
:bigtop-tests:smoke-tests:hdfs:test (Thread[Execution worker for ':' Thread 46,5,main]) completed. Took 9 mins 43.086 secs.

BUILD SUCCESSFUL in 9m 49s
12 actionable tasks: 8 executed, 24 up-to-date
Stopped 1 worker daemon(s).
+ rm -rf buildSrc/build/test-results/binary
+ rm -rf /bigtop-home/.gradle
```

Deploy success

```
[root@a6597325db6b /]# jps
6368 ResourceManager
8177 DataNode
7798 NameNode
7223 WebAppProxyServer
10840 Jps
6907 JobHistoryServer
9039 NodeManager
```

Bigtop

接下来的计划

Bigtop Immediate Roadmap - 2024

Right Now Bigtop is working v3.3 release, with **openEuler** support.

, as soon as it is done the next release will be planned out. For now the below is in consideration:

- Remove Ambari dependency and HDP dependency
- **Ambari** mpack
- **Parallel** compilation support

- Add Apache **Beam**
- Enable **Kerberos** support
- Enable **Ranger** and **Security** Components and support for various components

- Upgrade various components
- Upgrade **JDK** support

THANKS