Big Data Team Space

About Us

Engineers

<table>
<thead>
<tr>
<th>Assignee, ARM</th>
<th>Assignee, Cavium</th>
</tr>
</thead>
<tbody>
<tr>
<td>The content of this macro can only be viewed by users who have logged in.</td>
<td>The content of this macro can only be viewed by users who have logged in.</td>
</tr>
</tbody>
</table>

Assignee, ARM
The content of this macro can only be viewed by users who have logged in.

Assignee, Cavium
The content of this macro can only be viewed by users who have logged in.

Member Engineer, ARM
The content of this macro can only be viewed by users who have logged in.

IRC Channel: #linaro-bigdata
Mailing List: leg-bigdata@linaro.org

Portal

Documentations

ERP

- ERP 16.12: Installing Hadoop 2.7.2, Spark 2.0 and Hive 2.0.1
- ERP 17.08 Building ELK (ElasticSearch, LogStash and Kibana) on Aarch64
- ERP 17.08 - Build, Configure, Install and Test Hadoop, Spark, Hive, Hbase in containerized environment
- Resources
  - ERP 17.08 deb packages

Big Data Components

1. Apache Bigtop
   - Building Hadoop 2.7.2, Spark 2.0, Hive 2.0.1 using Apache BigTop
   - Building BigTop using Docker container
   - Bigtop Sandbox with Hadoop, Spark, Hive, HBase for Aarch64
   - Bigtop Smoke Tests
   - Resources
     - Bigtop 1.2 Components
     - Bigtop Docker images
     - Upstream github
     - CI Loop
     - v1.2 Release
     - Bigtop AArch64 patches
     - Bigtop Project Ideas and Roadmap

2. ODPI
   - Building ODPI BigTop packages
   - Setup, Configure and Install ODPI Hadoop
   - ODPI Hadoop Cluster Setup Guide
   - ODPI BigTop Hadoop Installation, setup and running

3. Big Data Core Components
   - Apache Hadoop
4. **Big Data Operations**
   - **Apache Ambari**
     - Build Apache Ambari on AArch64
     - Build Apache Ambari on AArch64 (Copy)
     - Install, Setup and Configure Apache Ambari
     - JVM Null pointer error in javac when building Apache Ambari
   - **Apache Zookeeper**
     - Enablement of Apache Zookeeper on AArch64
   - **Apache Oozie**
   - Apache Falcon
   - Ganglia

5. **Big Data Streaming Tools**
   - **Apache Spark**
     - Building Hadoop 2.7.2, Spark 2.0, Hive 2.0.1 using Apache Bigtop
     - Spark 2.0 - Build, configure and installation steps using Apache Bigtop
     - Profiling Spark on AArch64
     - Building Apache Spark 1.6 on AArch64
     - Building Apache Spark 2.0
     - Spark Dependency Libraries
   - **Apache Flink**
   - Apache Beam
   - Apache Tez
   - Apache Flume
   - Apache Storm
   - Apache Alluxio
   - **Apache Kafka**
     - Apache Kafka Streams
   - **Apache NiFi**
     - Difference between NiFi and MiNiFi
   - **Apache MiNiFi**

6. **Big Data Warehousing and Database Tools**
   - **Apache Hive**
     - Building Apache Hive
   - Apache HBase
     - Enablement of Apache HBase on AArch64
   - **Apache Cassandra**
   - Postgres
   - Memcached
   - MySQL
   - Redis
   - Apache Drill
   - Apache Sqoop

7. **Big Data Data Governance and Security**
   - **Apache Ranger**
   - Apache Knox
   - Apache Atlas
   - Apache Sentry

8. **Big Data Data Science Notebooks**
   - **Apache Jupyter**
   - Apache Zeppelin

9. **Big Data Analytics**
   - **ELK - ElasticSearch, Logstash and Kibana**
     - ERP 17.08 Building ELK (ElasticSearch, LogStash and Kibana) on AArch64
     - ELK Setup and Run on AArch64
   - Apache Arrow
   - Apache Kudu
11. **Big Data ML - Machine Learning**
   - H20
     - H20 Install and Run Guide
     - H20 Scaling Study
     - Sparkling Water - Spark + H20

12. **Big Data component dependencies**
    - NETLIB-JAVA AArch64 Natives Support

**Tests**

- **Smoke Tests**
  - BigTop smoke tests
- **Integration Tests**
- **ODPi Spec Tests**

**Benchmarking**

- TeraSort
- Building, Running, Configuring and Profiling Apache Hadoop
- Spark Bench
- TPC-H
- TPCxHS
- Apache Bench
- BigBench
- Building and Running HiBench on AArch64 Platform
- HiveTestBench
- **HiBench**
  - Spark profiling on Arm64

**Build and Port**

- Build Apache Ambari on AArch64
- HBase Enablement on AArch64
- Apache Flink on AArch64
- Zookeeper Enablement on AArch64
- NETLIB-JAVA AArch64 Natives Support
- Apache Ambari Install, Setup and Configuration

**Machine Learning**

- Building and Installing Tensorflow on AArch64

**Misc**

- Onboarding info - Welcome to Team
- Big Data team work summary v1.0
- CRC32 vs Non-CRC32 Study
- Configuring archiva with Tomcat

**Bigtop**

**Blogs/Presentations**

State of Big Data on Aarch64 - Apache Bigtop

Big Data benchmarking

Himalayan Odyssey
Big Data Roadmap

Strategic Engineering
Big Data and OpenJDK Strategic Engineering - 2018
Big Data and OpenJDK Strategic Engineering - 2017

Big Data Epics

<table>
<thead>
<tr>
<th>key</th>
<th>summary</th>
<th>type</th>
<th>created</th>
<th>updated</th>
<th>due</th>
<th>assignee</th>
<th>reporter</th>
<th>priority</th>
<th>status</th>
<th>resolution</th>
</tr>
</thead>
</table>

⚠️ Jira project doesn't exist or you don't have permission to view it.

View these issues in Jira

JIRA Inprogress

<table>
<thead>
<tr>
<th>key</th>
<th>summary</th>
<th>type</th>
<th>created</th>
<th>updated</th>
<th>due</th>
<th>assignee</th>
<th>reporter</th>
<th>priority</th>
<th>status</th>
<th>resolution</th>
</tr>
</thead>
</table>

⚠️ Jira project doesn't exist or you don't have permission to view it.

View these issues in Jira

Recently updated

Big Data JIRA status 03 2020
Mar 31, 2020 • updated by Ganesh Raju • view change

Apache Knox v1.0
Mar 18, 2020 • updated by Naresh Bhat • view change

Apache Knox
Mar 18, 2020 • created by Naresh Bhat