Notice

BlueData Software, Inc. believes that the information in this publication is accurate as of its publication date. However, the information is subject to change without notice. THE INFORMATION IN THIS PUBLICATION IS PROVIDED “AS IS.” BLUEDATA SOFTWARE, INC. MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, or distribution of any BlueData software described in this publication requires an applicable software license.

For the most up-to-date regulatory document for your product line, please refer to your specific agreements or contact BlueData Technical Support at support@bluedata.com.

The information in this document is subject to change. This manual is believed to be complete and accurate at the time of publication and no responsibility is assumed for any errors that may appear. In no event shall BlueData Software, Inc. be liable for incidental or consequential damages in connection with or arising from the use of this manual and its accompanying related materials.

Contact Information

BlueData Software, Inc.
148 Castro Street, Suite A-1
Mountain View, CA 94041
Email: info@bluedata.com
Website: www.bluedata.com

Copyrights and Trademarks

Copyright 2015 by BlueData Software, Inc. All rights reserved.
This manual describes the known issues and workarounds in the following builds:

- EPIC 1.5.0 GA (Build 2055)
- EPIC Lite 1.5.0 GA (Build 2052)

**CAUTION: THESE RELEASE NOTES ARE NOT VALID FOR ANY BUILD OF EPIC OR EPIC LITE OTHER THAN THE BUILD(S) LISTED HERE.**
1.1 - New Features

The listed builds of EPIC and EPIC Lite include the new features described in this section.

1.1.1 - Installation

Support for RHEL 6.6 and CentOS 6.6
Both EPIC Lite and EPIC Enterprise support installation on Red-Hat Enterprise Linux 6.6 and CentOS 6.6. EPIC Enterprise also continues to support installation on 6.5.

EPIC Lite supports deployment in VirtualBox or Amazon EC2
See the EPIC Lite Installation Guide for details.

Local HDFS system storage uses Kerberos
If local HDFS is chosen for system storage at install time, it will be set up with Kerberos protection. If you are upgrading from a previous EPIC installation, then Kerberos protection will NOT be added to existing local HDFS storage.

1.1.2 - Site Administration

Kerberos support for HDFS system storage
As noted above, local system storage is initially configured to use Kerberos protection. The details of this configuration can be viewed and edited in the system storage settings. System storage can also be configured to use Kerberos-protected remote HDFS. It is also still possible to use non-Kerberos remote HDFS or other remote filesystem types for system storage.

User-authored edge node types in the Catalog
See the Edge Node Authoring Guide for details (EPIC Enterprise) only.

1.1.3 - Tenant Administration

Kerberos support for HDFS DataTaps
DataTaps can now be configured to access Kerberos-protected HDFS storage. It is also still possible to access non-Kerberos HDFS (or other filesystem types) with DataTaps.
1.1.4 - Operation

Catalog changes for EPIC Enterprise
CDH 5.3, HDP 2.2, Spark 1.3.1, and unmodified CentOS 6.4 have been added to the catalog. CDH 4.4 and Spark 0.9.0 have been removed.

EPIC Lite Catalog
The EPIC Lite Catalog is different from both EPIC Enterprise and the previous EPIC One. CDH 5.2, HDP 2.2, and Spark 1.3.1 are currently supported.

New application support
Sqoop2 support is now included for Hadoop distros. Ambari and Cloudera Manager are now included for HDP 2.2 and CDH 5.3 distros (respectively) in the EPIC Enterprise catalog.

Support for Edge node deployment with clusters
If Edge node types have been created (see the Edge Node Authoring Guide for details), then Edge node applications will be available as selections during creation of the relevant cluster configurations.
1.2 - Known Issues & Workarounds

This section lists the known issues that are present in the listed versions of EPIC and EPIC Lite and methods to work around/recover from these issues.

1.2.1 - Installation

**Phase Two of installation may fail with “ImportError: cannot import name Random” message (HAATHI-10752)**

EPIC installation requires that the python-crypto RPM package either be absent, or be installed as version python-crypto-2.6.1. If a different version is present then EPIC installation can fail during the OpenStack portion of Phase Two (the Web-initiated phase).

**Recovery:** This problem can be avoided before executing the EPIC installer by removing python-crypto or ensuring it is the correct version. If however the problem is detected after executing the installer, DO NOT remove python-crypto as that would remove its dependent packages (such as bluedata-con- troller). Instead, upgrade python-crypto to the latest version provided by BlueData by running the following command:

```
yum update python-crypto
```

Re-start Phase Two through your Web browser after running this command.

1.2.2 - Site Administration

**Worker services stay down after Controller node reboot (HAATHI-10088)**

Rebooting the Controller node at any point after adding Worker nodes to the EPIC platform may cause certain services on the Worker nodes (nova compute and openvswitch) to become unresponsive as a result of losing connection with the rabbitmq service on the Controller node. As a result, existing virtual nodes on the affected Worker nodes will not be able to communicate with the Controller. The same problem is also observed if the rabbitmq service is restarted on the Controller node.

**Recovery:** Either restart the following services on the affected worker hosts or, if permissible, reboot the affected Worker nodes.

```
service openstack-nova-compute restart
service neutron-openvswitch-agent restart
```

**Syslog logging consumes root disk space (HAATHI-9846)**

By default, the EPIC services are configured to log messages to the /var/log/messages file. These messages will cause the /var/log/messages files to grow and could fill the root file system of the server.
**Workaround:** Use the `logrotate` service to limit the amount of disk space that the `/var/log/messages` file can occupy.

For example, to limit the `/var/log/messages` usage to 2 files of 50 MB each, edit the `/etc/logrotate.conf` file and add the following lines to the end of the file:

```
/var/log/messages {
    size 50M
    rotate 1
    daily
    postrotate
        service rsyslog restart
    endscript
}
```

After editing the file, run the following script to have the change take effect:

```
/etc/cron.daily/logrotate
```

You must edit the `/etc/logrotate.conf` file on both the Controller node and on each Worker node in the EPIC platform.

---

**Worker services stay down after Controller node reboot (HAATHI-10088)**

Rebooting the Controller node at any point after adding Worker nodes to the EPIC platform may cause certain services on the Worker nodes (`nova compute` and `openvswitch`) to become unresponsive as a result of losing connection with the `rabbitmq` service on the Controller node. As a result, existing virtual nodes on the affected Worker nodes will not be able to communicate with the Controller. The same problem is also observed if the `rabbitmq` service is restarted on the Controller node.

**Recovery:** Either restart the following services on the affected worker hosts or, if permissible, reboot the affected Worker nodes.

```
service openstack-nova-compute restart
service neutron-openvswitch-agent restart
```

---

**DataTap charts show incorrect bytes read/written until first use (HAATHI-10326, HAATHI-10170)**

On the Physical Nodes screen, the DataTap Charts tab includes graphs that show total bytes read and written over the selected time frame. The bytes-read value will be inaccurate until some DataTap has first been used to perform read operations; similarly the bytes-written value will be inaccurate until the first write.

---

**Worker installation can fail because an HDFS datanode's clock is out of sync (HAATHI-10755)**

If local HDFS system storage is used, a Kerberos-protected HDFS service makes use of local disks on every worker host added to the EPIC platform. If a worker host's clock is too far out of sync, this process can fail, as Kerberos is sensitive to clock differences.

**Recovery:** The worker-add process does configure the NTP service on worker hosts, but it may not bring the host's clocks into sync rapidly enough. You can try to manually set the clock of the
worker host to match that of the controller; or force an NTP sync on the worker host by logging in as root and executing these commands:

```
    service ntp stop; ntpd -gq; service ntp start
```

Existing CDH 4.4 or Spark 0.9 clusters prevent upgrade (HAATHI-10750)

CDH 4.4 and Spark 0.9 support has been removed for EPIC Enterprise 1.5. Attempts to upgrade to 1.5 will fail if any persistent clusters of these types exist.

**Workaround:** Manually delete any persistent clusters of these types before performing the upgrade to 1.5.

After platform HA enabled, UI upload of new system storage keytab can cause issues (HAATHI-10683)

UI upload of a keytab will only upload a keytab to the current controller host, even if platform HA is enabled. It will not be automatically uploaded to the shadow controller host. For DataTap keytabs this is not an issue; however, a system storage keytab must be present on the current controller host whenever a new tenant or a new temporary cluster filesystem is created. Therefore the upload of a new system storage keytab followed by a failover to the shadow controller will make such operations vulnerable to failure.

**Workaround:** If uploading a new system storage keytab, and platform HA has previously been enabled, the new keytab should also be manually transferred to the `/srv/bluedata/keytab/site_admin` directory of the current shadow controller host.

UI upload of a new system storage keytab named “datasrvr.headless.keytab” fails (HAATHI-10751)

If local HDFS is used for system storage, it is initially configured to use a keytab file named `datasrvr.headless.keytab`. You can change the system storage settings to use a different keytab file if you wish, but you cannot use the UI to upload a keytab file with the same name `datasrvr.headless.keytab` (i.e. you cannot use the UI to overwrite that specific initial keytab file). An attempt to do so will result in an error dialog about the operation encountering an exception.

**Workaround:** Although the current error message is cryptic, this restriction is intentional. Note that this restriction only applies to the system storage settings, and not to the editing of any tenant DataTaps that might use copies of that keytab file. If for some reason you genuinely need to replace the original system storage keytab with a file of the same name, you can do so manually.

On EC2, the EPIC Lite UI becomes unresponsive during Catalog installation (HAATHI-10702)

If EPIC Lite is deployed into an Amazon EC2 instance, the EPIC web UI can become intermittently unresponsive while catalog images are being downloaded and installed.
Workaround: Wait a few minutes for Catalog installation to complete.

1.2.3 - Operational

HBase client configuration uses VM hostnames (HAATHI-10037)

Spark worker UI is not accessible from master page (HAATHI-10060)

Some URLs in CDH 5.0 ResourceManager use hostname instead of IP address (HAATHI-10155)

Each virtual node created by EPIC is assigned a host name. These host names are only known to other virtual nodes within the EPIC platform. Client computers outside the EPIC platform may wish to follow web links to services within the virtual nodes, but those clients cannot resolve links that are based on the virtual node host names instead of IP addresses.

Workaround: Users must configure the hosts file on the client computer that will be accessing the virtual nodes in a given cluster.

To update the /etc/hosts file on a Linux system:
After creating a cluster, navigate to the Clusters page.

For each cluster, there will be a purple Hosts File Info icon (screen) that will provide the template for setting up the /etc/hosts file.

The template file should look like this:

To update the /etc/hosts file on a Windows 7 system:
1. Click Start>All Programs>Accessories.
2. Right click Notepad and select Run as administrator.
3. If applicable, click Continue in the Windows needs your permission UAC window.
4. In Notepad, click File>Open.
5. In the **File Name** field, type `C:\Windows\System32\Drivers\etc\hosts`, and then click **Open**.

**To update the /etc/hosts file on a Windows NT/2000/XP system:**

1. Click **Start>All Programs>Accessories>Notepad**.
2. In Notepad, click **File>Open**.
3. In the filename field type `C:\Windows\System32\Drivers\etc\hosts`, and then click **Open**.

The `C:\Windows\System32\Drivers\etc\hosts` file is a hidden file, and you will therefore need to enable the hidden files folder option as shown below:

---

**Spark jobs run out of memory on large datasets (HAATHI-10057)**

Spark jobs may need to set certain configuration parameters before execution. The EPIC interface for dispatching Spark jobs does not provide the option to set these configuration parameters while submitting a job.

**Workaround:** To set these parameters, log in into the Spark master virtual node and follow the guidelines specific to that Spark release.

For example:

For Spark 1.0.0, if a job needs to set `spark.executor.memory` to 2048m, and then edit the `spark-defaults.conf` file in the `$SPARK_HOME/conf` directory.

Copy the `spark-defaults.conf.template` file as `spark-defaults.conf`, and then add the line `spark.executor.memory 2048m`. The Spark job reads this file for any updated configuration parameter.

---

**An HDP 2.2 virtual node fails to restart its services if rebooted while its controller is down (HAATHI-10745)**

On EPIC Enterprise (not EPIC Lite), a virtual node deployed with HDP 2.2 + Ambari and then rebooted will fail to restart necessary services if the controller virtual node of that cluster is down. This is true even for an HDP 2.2 cluster with YARN HA enabled; such a cluster can continue to operate if its controller virtual node goes down, but will become nonfunctional if its standby node is rebooted while that controller node is down.
**Workaround:** Bring the virtual cluster's controller node back up if possible, then reboot the affected virtual nodes.

**HDP 2.2 cluster expansion fails to start services in the newly added nodes (HAATHI-10719)**

On EPIC Enterprise (not EPIC Lite), expansion of an HDP 2.2 + Ambari cluster through the EPIC UI will add virtual nodes to the cluster but not start the necessary services on those nodes.

**Workaround:** Start the services using the cluster’s Ambari interface, through the following steps:

1. Wait for the EPIC cluster expansion operation to be successful. Then, in the EPIC UI, go to the cluster node list and take note of the newly added node names and IP addresses.
2. SSH into to each of the new virtual nodes and execute the following command:
   
   ```bash
   echo "export JAVA_HOME=`cat /root/BD_Setup/JAVA_HOME`" >> /etc/profile
   ```
3. From the cluster's Ambari dashboard, click **Hosts** in the top bar and then select **Add New Hosts** from the **Actions** pull-down. This starts the **Add Host** wizard.
4. In the **Target Hosts** box of the **Add Host** wizard, put one hostname per line to identify your new nodes. Each hostname is formed from the node name and the domain `openstacklocal`. For example a node named `bluedata-88` would have the hostname `bluedata-88.openstacklocal`.
5. For the **Host Registration Information**, you may either upload the SSH key for the nodes, or choose manual registration.
6. Click **Register and Confirm**. You can ignore warnings generated at this point.
7. Once the registration is successful, continue to the **Assign Slaves and Clients** step and select **NodeManager, Client**, and all other services needed for each node. Click **Next**.
8. Click **Next** to accept the default configurations.
9. Click **Deploy**.
10. When the deployment is complete, continue with any further service actions recommended by Ambari.

**NFS DataTaps are not accessible when using EPIC Lite in VirtualBox (HAATHI-10637)**

When using EPIC Lite deployed in a VirtualBox VM, DataTaps for NFS storage services may be in an error state (red name) or incorrectly appear to contain no files.

**Recovery:** Add a bridged network in the Settings of the VirtualBox VM that is hosting EPIC Lite.
Hive jobs that use DataTap paths may fail with a SemanticException error (HAATHI-10733)

When Hive creates a table, the location where the table metadata is stored comes from the Hive configuration parameter fs.defaultFS by default (which will point to the cluster filesystem). If a Hive job references DataTap paths outside of the filesystem where the table metadata is stored, then the job will fail with a SemanticException error, because Hive enforces that all data sources come from the same filesystem.

**Workaround:** Explicitly set the table metadata location to a path on the same DataTap that you will use for the job inputs and/or outputs, using the `LOCATION` clause when creating the table. For example, if you intend to use the TenantStorage DataTap, you would set the table metadata location to some path on that DataTap such as:

```
CREATE TABLE docs (c1 INT, c2 STRING) LOCATION 'dtap://TenantStorage/hive-table-docs'
```

Ambari dashboard does not show YARN or Flume service metrics (HAATHI-10667)

The Ambari YARN and Flume summary dashboards of an HDP 2.2 + Ambari cluster will display **No data** for service metrics.

**Workaround:** Some metrics are shown on the top-level Ambari dashboard. Some other metrics can be seen on the cluster’s Ganglia dashboard; in the `<Cluster>` page on the EPIC web UI, select the **Charts** tab, and then click on the **Ganglia Dashboard** link at the bottom of the page; however there is not currently a complete workaround for this issue.

Spark jobs cannot be successfully submitted to a CDH 5.0 cluster from the Bluedata UI (HAATHI-10754)

Submitting a Spark job (Java, Scala, or Python) through the EPIC UI to a CDH 5.0 cluster will fail, even if that cluster has Spark enabled.

**Workaround:** Spark jobs can still be run manually when connected to the cluster’s nodes via SSH; for reference see:

- [https://spark.apache.org/docs/0.9.0/scala-programming-guide.html](https://spark.apache.org/docs/0.9.0/scala-programming-guide.html)

Also, several other kinds of EPIC virtual clusters can accept Spark jobs submitted through the EPIC UI.
This page intentionally left blank.
**EPIC and EPIC Lite Release Notes (see front cover for version)**

This book or parts thereof may not be reproduced in any form without the written permission of the publishers. Printed in the United States of America. Copyright 2015 by BlueData Software, Inc. All rights reserved.

**Contact Information:**

BlueData Software, Inc.
148 Castro Street, Suite A-1
Mountain View, CA 94041
**Email:** info@bluedata.com
**Website:** www.bluedata.com