

IBM SPSS Analytic Server
Version 2.1

Installation and Configuration Guide

IBM

Note

Before using this information and the product it supports, read the information in "Notices" on page 25.

Product Information

This edition applies to version 2.1, release 1, modification 0 of IBM SPSS Analytic Server and to all subsequent releases and modifications until otherwise indicated in new editions.

Contents

Chapter 1. Overview	1	Enabling relational database sources	15
Architecture	2	Enabling HCatalog data sources	16
What is new for installers in version 2.1	2	Changing ports used by Analytic Server.	16
Chapter 2. Installation.	5	High availability Analytic Server	17
Prerequisites	5	Optimizing JVM options for small data	17
Installation	6	Updating client dependencies	17
Chapter 3. Configuration	9	Chapter 4. Migration	19
Security	9	Chapter 5. Uninstalling	21
Making changes to the basic registry	9	Chapter 6. Troubleshooting	23
Configure an LDAP registry.	10	Notices	25
Configuring Kerberos	11	Trademarks	27
Enabling Secure Socket Layer (SSL) connections to the Analytic Server console	12		
Enabling Support for Essentials for R.	13		
Configuring IBM SPSS Modeler for use with IBM SPSS Analytic Server	14		

Chapter 1. Overview

IBM® SPSS® Analytic Server is a solution for big data analytics that combines IBM SPSS technology with big data systems and allows you to work with familiar IBM SPSS user interfaces to solve problems on a previously unattainable scale.

Why big data analytics matters

Data volumes collected by organizations are growing exponentially; for example, financial and retail businesses have all customer transactions for a year (or two years, or ten years), telco providers have call data records (CDR) and device sensor readings, and internet companies have the results of web crawls.

Big data analytics is needed where there exists:

- A large volume of data (terabytes, petabytes, exabytes), especially when it is a mixture of structured & unstructured data
- Rapidly changing/accumulating data

Big data analytics also assists when:

- A large number (thousands) of models are being built
- Models are frequently built/refreshed

Challenges

The same organizations that collect large volumes of data often have difficulty actually making use of it, for a variety of reasons:

- The architecture of traditional analytic products are not suited to distributed computation, and
- Existing statistical algorithms are not designed to work with big data (these algorithms expect the data to come to them, but big data is too costly to move), thus
- Performing state of the art analytics on big data requires new skills and intimate knowledge of big data systems. Very few analysts have these skills.
- In-memory solutions work for medium-size problems, but do not scale well to truly big data.

Solution

Analytic Server provides:

- A data-centric architecture that leverages big data systems, such as Hadoop Map/Reduce with data in HDFS.
- A defined interface to incorporate new statistical algorithms designed to go to the data.
- Familiar IBM SPSS user interfaces that hide the details of big data environments so that analysts can focus on analyzing the data.
- A solution that is scalable to any size problem.

Architecture

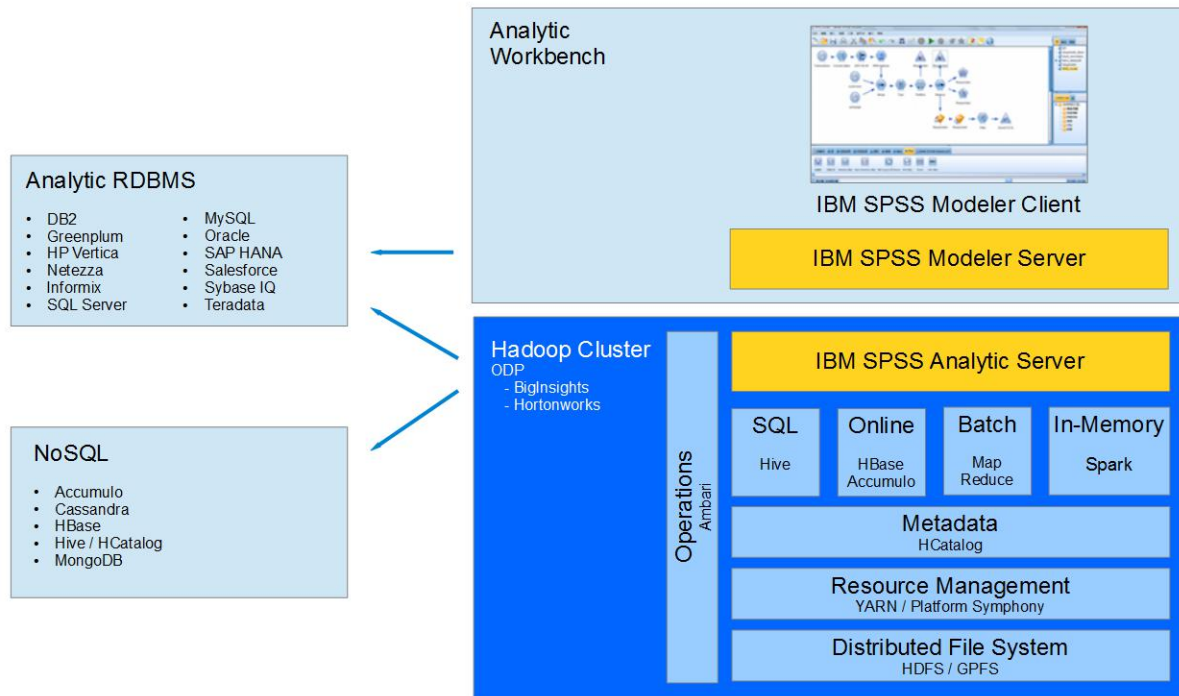


Figure 1. Architecture

Analytic Server sits between a client application and Hadoop cloud. Assuming that the data resides in the cloud, the general outline for working with Analytic Server is to:

1. Define Analytic Server data sources over the data in the cloud.
2. Define the analysis you want to perform in the client application. For the current release, the client application is IBM SPSS Modeler.
3. When you run the analysis, the client application submits an Analytic Server execution request.
4. Analytic Server orchestrates the job to run in the Hadoop cloud and reports the results to the client application.
5. You can use the results to define further analyses, and the cycle repeats.

What is new for installers in version 2.1

Installation and configuration

Analytic Server is now installed and runs as an Apache Ambari service.

Platform

Support for operating systems and Hadoop distributions is changed from version 2.

Operating systems

Analytic Server now runs on Red Hat Enterprise Linux (Power LE) in addition to existing operating system support.

Hadoop distributions

Analytic Server runs with Big Insights and Hortonworks.

Metadata repository

Analytic Server no longer supports Derby as the default metadata repository, and instead uses MySQL. DB2 is still supported as an alternative repository.

Database data sources

Data sources can be defined for Amazon Redshift, in addition to existing database support.

For the most up-to-date system requirements information, use the Detailed system requirements reports at the IBM Technical Support site: <http://publib.boulder.ibm.com/infocenter/prodguid/v1r0/clarity/softwareReqsForProduct.html>. On this page:

1. Type SPSS Analytic Server as the product name and click **Search**.
2. Select the wanted version and scope of report, then click **Submit**.

Chapter 2. Installation

Prerequisites

Before installing Analytic Server, review the following information.

System requirements

For the most up-to-date system requirements information, use the Detailed system requirements reports at the IBM Technical Support site: <http://publib.boulder.ibm.com/infocenter/prodguid/v1r0/clarity/softwareReqsForProduct.html>. On this page:

1. Type SPSS Analytic Server as the product name and click **Search**.
2. Select the wanted version and scope of report, then click **Submit**.

Power systems

Ensure that the IBM XLC and XLF compilers are installed and included in the PATH on all hosts in the cluster.

You can find more information about getting a license for these compilers at the following web sites:

- XL C for AIX: <http://www-03.ibm.com/software/products/en/xlcaix>
- XL Fortran for AIX: <http://www-03.ibm.com/software/products/en/xlfortran-aix>

Ambari

Analytic Server is installed as an Ambari service. Prior to installing Analytic Server, you should ensure that HDFS, YARN, MapReduce2, Hive, and Zookeeper are added as Ambari services.

Password-less SSH

Set up password-less SSH for the root user between the Analytic Metastore host and all hosts in the cluster.

Hive/HCatalog

If you plan to use NoSQL data sources, then configure Hive and HCatalog for remote access. Also ensure that `hive-site.xml` contains a `hive.metastore.uris` property in the form `thrift://<host_name>:<port>` that points to the active Thrift Hive Metastore server. Refer to your Hadoop distribution documentation for details.

Metadata repository

By default, Analytic Server installs and uses a MySQL database. Alternatively, you can configure Analytic Server to use an existing DB2 installation. Regardless of which type of database you choose, it must have an encoding of UTF-8.

MySQL

The default character set for MySQL is dependent upon the version and operating system. Use the following steps to determine whether your installation of MySQL is set to UTF-8.

1. Determine the version of MySQL.

```
mysql -V
```
2. Determine the default character set for MySQL by running the following query from the MySQL command line interface.

```
mysql>show variables like 'char%';
```

If the character sets already set to UTF-8 no further changes are needed.

3. Determine the default collation for MySQL by running the following query from the MySQL command line interface.

```
mysql>show variables like 'coll%';
```

If the collation is already set to UTF-8 no further changes are needed.

4. If the default character set or collation is not UTF-8 refer to the MySQL documentation for details on how to edit `/etc/my.cnf` and restart the MySQL daemont to change the character set to UTF-8.

DB2 For more information on configuring DB2, see the Knowledge Center http://www-01.ibm.com/support/knowledgecenter/SSEPGG_10.5.0/com.ibm.db2.luw.kc.doc/welcome.html.

High-availability clusters

Load balancer

Your high availability cluster should have a load balancer that supports session affinity, sometimes also known as sticky sessions. Analytic Server identifies sessions with the cookie "request-token". This identifies a session for the duration of a user login for use in application-controlled session affinity. Please consult the documentation of your particular load balancer for the details of how it supports session affinity.

Installation

A self-extracting binary file specific to your stack, stack version, and hardware architecture can be downloaded from the [IBM Passport Advantage® Web Site](#).

1. Execute the self-extracting binary file and follow the instructions to (optionally) view the license, accept the license, and choose online or offline installation.

Online installation

Choose online installation if your Ambari server host, and all nodes in the cluster, are able to access <http://ibm-open-platform.ibm.com>.

Offline installation

Choose offline if your Ambari server host does not have internet access. Offline installation will download the necessary RPM files, and should be run on a machine that can access <http://ibm-open-platform.ibm.com>. The RPM files can then be copied to the Ambari server host.

- a. Install the tool that allows you to create a local Yum repository.

```
yum install createrepo
```
- b. Create a new directory that will serve as the repository for the Analytic Server RPM files. See the following example.

```
mkdir /home/root/repos/IBM-SPSS-AnalyticServer/x86_64
```
- c. Copy the necessary Analytic Server RPM files to this directory. The RPM files you need depend on your distribution, version, and architecture, shown below.

BigInsights 4.1 (x86_64)

```
IBM-SPSS-AnalyticServer-ambari-2.1-BI-4.1-2.1.0.0-1.x86_64.rpm
```

```
IBM-SPSS-AnalyticServer-2.1.0.0-1.x86_64.rpm
```

BigInsights 4.1 (PPC64LE)

```
IBM-SPSS-AnalyticServer-ambari-2.1-BI-4.1-2.1.0.0-1.ppc64le.rpm
```

```
IBM-SPSS-AnalyticServer-2.1.0.0-1.ppc64le.rpm
```

HDP 2.3 (x86_64)

```
IBM-SPSS-AnalyticServer-2.1.0.0-1.x86_64.rpm
```

```
IBM-SPSS-AnalyticServer-ambari-2.1-HDP-2.3-2.1.0.0-1.x86_64.rpm
```

- d. Create the local repository definition. For example, create a file called `analyticserver.repo` in `/etc/yum.repos.d/` (for RHEL, CentOS) or `/etc/zypp/repos.d/` (for SLES) with the following contents.

```
[IBM-SPSS-AnalyticServer]
name=IBM-SPSS-AnalyticServer-2.1.0.0
baseurl=file:///path to local repository
enabled=1
gpgcheck=0
protect=1
```

- e. Create the local Yum repository. See the following example.


```
createrepo /home/root/repos/IBM-SPSS-AnalyticServer/x86_64
```
- f. Install the Ambari metadata RPM from your local repository; this is the file of form IBM-SPSS-AnalyticServer-ambari-xxx.rpm. For example, to install the RPM for Ambari 2.1 on BigInsights 4.1 on RHEL or CentOS, run the following command.


```
sudo yum install IBM-SPSS-AnalyticServer-ambari-2.1-BI-4.1
```

On SLES, the command changes as follows.

```
sudo zypper install IBM-SPSS-AnalyticServer-ambari-2.1-BI-4.1
```

- g. Update your Ambari repository file `repoinfo.xml`, typically located in `/var/lib/ambari-server/resources/stacks/$stackName/$stackVersion/repos/`, to use the local Yum repository, by adding the following lines.


```
<os type="host_os">
  <repo>
    <baseurl>file:///path to local repository/</baseurl>
    <repopid>IBM-SPSS-AnalyticServer</repopid>
    <reponame>IBM-SPSS-AnalyticServer-2.1.0.0</reponame>
  </repo>
</os>
```

2. Restart your Ambari server.

```
ambari-server restart
```

3. Log on to your Ambari server and install Analytic Server as a service via the Ambari UI.

Metadata repository

Analytic Server uses MySQL by default to track information about data sources, projects, and tenants. During installation you need to provide a username (**metadata.repository.user.name**) and password **metadata.repository.password** used in the JDBC connection between Analytic Server and MySQL. The installer creates the user in the MySQL database but that user is specific to the MySQL database and does not need to be an existing Linux or Hadoop user.

To change the metadata repository to DB2, follow these steps.

Note: You cannot change the metadata repository after installation is complete.

- a. Ensure that DB2 is installed on another machine. For more information, see the metadata repository section of the topic “Prerequisites” on page 5.
- b. In the Ambari Services tab, navigate to the Configs tab of the Analytic Server service.
- c. Open the **Advanced analytics-env** section.
- d. Change the value of **as.database.type** from `mysql` to `db2`.
- e. Open the **Advanced analytics-meta** section.
- f. Change the value of **metadata.repository.driver** from `com.mysql.jdbc.Driver` to `com.ibm.db2.jcc.DB2Driver`.
- g. Change the value of **metadata.repository.url** to `jdbc:db2://{DB2_HOST}:{PORT}/{DBName}:currentSchema={SchemaName};`, where
 - {DB2_HOST} is the hostname of the server where DB2 is installed
 - {PORT} is the port on which DB2 is listening
 - {SchemaName} is an available, unused schema.

If you are unsure of what values to enter, work with your DB2 administrator.

- h. Supply valid DB2 credentials in **metadata.repository.user.name** and **metadata.repository.password**.
- i. Click **Save**.

Configuration settings that should not be changed after installation

Do not change the following settings after installation, or Analytic Server will fail to work.

- Analytic_Server_User
 - Analytic_Server_UserID
 - as.database.type
 - metadata.repository.driver
 - distrib.fs.root
4. You now have a functioning instance of Analytic Server. Further configuration is optional. For more information on configuring and administrating Analytic Server, see the topic: Chapter 3, "Configuration," on page 9. For information on migrating an existing configuration to a new installation, see the topic: Chapter 4, "Migration," on page 19.
 5. Open a web browser and enter the address `http://<host>:<port>/analyticserver/admin/ibm`, where `<host>` is the address of the Analytic Server host, and `<port>` is the port that Analytic Server is listening on. By default this is 9080. This URL opens the login dialog for the Analytic Server console. Log in as the Analytic Server administrator. By default this userid is admin and has password admin.

Chapter 3. Configuration

After installation, you can optionally configure and administer Analytic Server through the Ambari UI.

Note: The following conventions are used for Analytic Server file paths.

- {AS_ROOT} refers to the location where Analytic Server is deployed; for example, /opt/IBM/SPSS/AnalyticServer/{version}.
- {AS_SERVER_ROOT} refers to the location of the configuration, log, and server files; for example, /opt/IBM/SPSS/AnalyticServer/{version}/ae_wlpserver/usr/servers/aeserver.
- {AS_HOME} refers to the location on HDFS that is used by Analytic Server as a root folder.

Security

The **security.config** parameter defines the registry of users and groups that can be added as principals to the Analytic Server system.

By default, a basic registry is defined with a single user, admin, with password admin. You can change the registry by editing **security.config** or configuring Kerberos.

Note: If you edit the **security.config** parameter to alter the registry, you then need to add any new users as principals to the Analytic Server system. See the *IBM SPSS Analytic Server Administrator's Guide* for details on tenant management.

Making changes to the basic registry

The basic registry allows you to define a database of users and groups within the **security.config** parameter.

The default basic registry looks like the following.

```
<basicRegistry id="basic" realm="ibm">
  <user name="admin" password="admin"/>
</basicRegistry>
```

An example of an altered basic registry follows.

```
<basicRegistry id="basic" realm="ibm">
  <user name="user1" password="{xor}Dz4sLG5tbGs="/>
  <user name="user2" password="Pass"/>
  <user name="user3" password="Pass"/>
  <user name="user4" password="Pass"/>
  <user name="admin" password="{xor}KzosKw=="/>
  <group name="Development">
    <member name="user1"/>
    <member name="user2"/>
  </group>
  <group name="QA">
    <member name="user3"/>
    <member name="user4"/>
  </group>
  <group name="ADMIN">
    <member name="user1"/>
    <member name="admin"/>
  </group>
</basicRegistry>
```

Passwords can be encoded to obfuscate their values with the securityUtility tool, which is in {AS_ROOT}/ae_wlpserver/bin.

```
securityUtility encode changeit  
{xor}Pdc+MTg6Nis=
```

Note: See http://www-01.ibm.com/support/knowledgecenter/SSD28V_8.5.5/com.ibm.websphere.wlp.core.doc/ae/rwlp_command_securityutil.html for details of the securityUtility tool.

Note: The basic registry is useful in a sandbox environment, but is not recommended for a production environment.

Configure an LDAP registry

The LDAP registry allows you to authenticate users with an external LDAP server such as Active Directory or OpenLDAP.

Here is an example of an ldapRegistry for OpenLDAP.

```
<ldapRegistry  
  baseDN="ou=people,dc=aeldap,dc=org"  
  ldapType="Custom"  
  port="389"  
  host="server"  
  id="OpenLDAP"  
  bindDN="cn=admin,dc=aeldap,dc=org"  
  bindPassword="{xor}Dz4sLG5tbGs="  
  searchTimeout="300000m"  
  recursiveSearch="true">  
  <customFilters  
    id="customFilters"  
    userFilter="(&(uid=%v)(objectClass=inetOrgPerson))"  
    groupFilter="(&(cn=%v)(|(objectclass=organizationalUnit)))"  
    groupMemberIdMap="posixGroup:memberUid"/>  
</ldapRegistry>
```

For more examples of configurations, see the templates folder {AS_ROOT}/ae_wlpserver/templates/config.

Configure a secure socket layer (SSL) connection from Analytic Server to LDAP

1. Login to each of the Analytic Server machines as the Analytic Server user and create a common directory for SSL certificates.

Note: By default, as_user is the Analytic Server user; see **Service accounts** under the Admin tab in the Ambari console.

2. Copy the key store and trust store files to some common directory on all Analytic Server machines. Also add the LDAP client CA certificate to the trust store. Below are some sample instructions.

```
mkdir /home/as_user/security  
cd /home/as_user/security  
openssl s_client -connect <ldap-hostname>:636 -showcerts > client.cert  
$JAVA_HOME/bin/keytool -import -file ./client.cert -alias ldapCA -keystore mytrust.jks  
password : changeit
```

Note: JAVA_HOME is the same JRE used for Analytic Server startup.

3. Passwords can be encoded to obfuscate their values with the securityUtility tool, which is in {AS_ROOT}/ae_wlpserver/bin. An example follows.

```
securityUtility encode changeit  
{xor}Pdc+MTg6Nis=
```

4. Login to the Ambari console and update the Analytic Server configuration setting **ssl.keystore.config** with the correct SSL configuration settings. An example follows.

```

<ssl id="defaultSSLConfig" keyStoreRef="defaultKeyStore" trustStoreRef="defaultTrustStore"
  clientAuthenticationSupported="true"/>
  <keyStore id="defaultKeyStore" location="/home/as_user/security/mykey.jks" type="JKS"
    password="{xor}0zo5PiozKxYdEgwPDaWeDG1uDz4sLCg7"/>
  <keyStore id="defaultTrustStore" location="/home/as_user/security/mytrust.jks" type="JKS"
    password="{xor}PDC+MTg6Nis="/>

```

Note: Use the absolute path for key and trust store files.

- Update the Analytic Server configuration setting **security.config** with the correct LDAP configuration settings. For example, in the **ldapRegistry** element, set the **sslEnabled** attribute to true and the **sslRef** attribute to defaultSSLConfig.

Configuring Kerberos

Analytic Server supports Kerberos using Ambari.

- Create accounts in the Kerberos user repository for all users you plan to give access to Analytic Server.

Note: If the Analytic Server installation uses a basic registry, it must include the Kerberos user accounts, using "-" as the password. An example follows.

```

<basicRegistry id="basic" realm="ibm">
  <user name="admin" password="-"/>
  <user name="user1" password="-"/>
  <user name="user2" password="-"/>
  <group name="group1">
    <member name="admin"/>
    <member name="user1"/>
    <member name="user2"/>
  </group>
  <group name="group2">
    <member name="admin"/>
    <member name="user1"/>
  </group>
</basicRegistry>

```

- Create an OS user account for each of the users created in the previous step on on each and every Analytic Server node and Hadoop node.
 - Make sure that the UID for these users matches on all machines. You can test this using the `kinit` command to log in to each of the accounts.
 - Make sure that the UID adheres to the "Minimum user ID for submitting job" Yarn setting. This is the **min.user.id** parameter in `container-executor.cfg`. For example, if **min.user.id** is 1000, then each user account created must have a UID greater than or equal to 1000.
- Create a user home folder on HDFS for all principals in Analytic Server. For example, if you add `testuser1` to the Analytic Server system, then create a home folder like `/user/testuser1` on HDFS and ensure that `testuser1` has read and write permissions to this folder.
- [Optional] If you plan to use HCatalog data sources and Analytic Server is installed on a different machine from the Hive metastore, you need to impersonate the Hive client on HDFS.
 - Navigate to the Configs tab of the HDFS service in the Ambari console.
 - Edit the **hadoop.proxyuser.hive.groups** parameter to have the value `*`, or a group that contains all of the users allowed to log in to Analytic Server.
 - Edit the **hadoop.proxyuser.hive.hosts** parameter to have the value `*`, or the list of hosts on which the Hive metastore and every instance of Analytic Server are installed as services.
 - Restart the HDFS service.

After these steps have been performed and Analytic Server is installed, Analytic Server silently and automatically configures Kerberos.

Configuring HAProxy for Single Sign On (SSO) using Kerberos

1. Configure and start HAProxy per the HAProxy documentation guide: <http://www.haproxy.org/#docs>
2. Create the Kerberos principle (HTTP/<proxyHostname>@<realm>) and keytab file for the HAProxy host, where <proxyHostname> is the full name of the HAProxy host, and <realm> is the Kerberos realm.
3. Copy the keytab file to each of the Analytic Server hosts as /etc/security/keytabs/spnego_proxy.service.keytab
4. Update permissions to this file on each of the Analytic Server hosts. An example follows.

```
chown root:hadoop /etc/security/keytabs/spnego_proxy.service.keytab  
chmod 440 /etc/security/keytabs/spnego_proxy.service.keytab
```
5. Open the Amabri console and update the following properties in the Analytic Server 'Custom analytics.cfg' section.

```
web.authentication.kerberos.keytab=/etc/security/keytabs/spnego_proxy.service.keytab  
web.authentication.kerberos.principal=HTTP/<proxy machine full name>@<realm>
```
6. Save the configuration and restart all Analytic Server services from the Amabri console.

Now users are able to log in to Analytic Server using Kerberos SSO.

Disabling Kerberos

1. Disable Kerberos in the Ambari console.
2. Stop the Analytic Server service.
3. Remove the following parameters from Custom analytics.cfg.

```
default.security.provider  
hdfs.keytab  
hdfs.user  
java.security.krb5.conf  
jdbc.db.connect.method.kerberos  
web.authentication.kerberos.keytab  
web.authentication.kerberos.principal
```

4. Click **Save** and restart the Analytic Server service.

Enabling Secure Socket Layer (SSL) connections to the Analytic Server console

By default, Analytic Server generates self-signed certificates to enable Secure Socket Layer (SSL), so you can access the Analytic Server console through the secure port by accepting self signed certificates. In order to make HTTPS access more secure, you need to install 3rd party vendor certificates.

To install 3rd party vendor certificates, follow these steps.

1. Copy the 3rd party vendor key store and trust store certificates to the same directory in all Analytic Server nodes; for example, /home/as_user/security.

Note: The Analytic Server User must have read access to this directory.

2. In the Ambari Services tab, navigate to the Configs tab of the Analytic Server service.
3. Edit the **ssl.keystore.config** parameter.

```
<ssl id="defaultSSLConfig"  
  keyStoreRef="defaultKeyStore"  
  trustStoreRef="defaultTrustStore"  
  clientAuthenticationSupported="true"/>  
<keyStore id="defaultKeyStore"  
  location="<KEYSTORE-LOCATION>"  
  type="<TYPE>"  
  password="<PASSWORD>"/>
```



```
<keyStore id="defaultTrustStore"
  location="<TRUSTSTORE-LOCATION>"
  type="<TYPE>"
  password="<PASSWORD>"/>
```

Replace

- <KEYSTORE-LOCATION> with the absolute location of the key store; for example:
/home/as_user/security/mykey.jks
- <TRUSTSTORE-LOCATION> with the absolute location of the trust store; for example:
/home/as_user/security/mytrust.jks
- <TYPE> with the type of the certificate; for example: JKS, PKCS12 etc.
- <PASSWORD> with the encrypted password in Base64 encryption format. For encoding you can use the securityUtility; for example: /opt/ibm/spss/analyticsserver/2.1/ae_wlpserver/bin/securityUtility encode <password>

If you want to generate a self-signed certificate, you can use securityUtility; for example:
/opt/ibm/spss/analyticsserver/2.1/ae_wlpserver/bin/securityUtility createSSLCertificate
--server=myserver --password=mypassword --validity=365
--subject=CN=mycompany,O=myOrg,C=myCountry. For more information on securityUtility and other SSL settings, refer to the WebSphere Liberty Profile documentation.

4. Click **Save** and restart the Analytic Server service.

Enabling Support for Essentials for R

Analytic Server supports scoring R models and running R scripts.

To configure support for R after a successful Analytic Server installation:

1. Download the self-extracting archive (BIN) for IBM SPSS Modeler Essentials for R RPM. Essentials for R is available for download (<https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=swg-tspssp>). Choose the file specific to your stack, stack version, and hardware architecture.
2. Execute the self-extracting binary file and follow the instructions to (optionally) view the license, accept the license, and choose online or offline installation.

Online installation

Choose online installation if your Ambari server host, and all nodes in the cluster, are able to access <http://ibm-open-platform.ibm.com>.

Offline installation

Choose offline if your Ambari server host does not have internet access. Offline installation will download the necessary RPM files, and should be run on a machine that can access <http://ibm-open-platform.ibm.com>. The RPM files can then be copied to the Ambari server host.

- a. Copy the necessary Essentials for R RPM files to any location on your Ambari server host. The RPM files you need depend on your distribution, version, and architecture, shown below.

BigInsights 4.1 (x86_64)

IBM-SPSS-ModelerEssentialsR-ambari-2.1-BI-4.1-8.2.0.0-1.x86_64.rpm

BigInsights 4.1 (PPC64LE)

IBM-SPSS-ModelerEssentialsR-ambari-2.1-BI-4.1-8.2.0.0-1.ppc64le.rpm

HDP 2.3 (x86_64)

IBM-SPSS-ModelerEssentialsR-ambari-2.1-HDP-2.3-8.2.0.0-1.x86_64.rpm

- b. Install the RPM. For example, the following command installs Essentials for R on Hortonworks 2.3.

```
rpm -i IBM-SPSS-ModelerEssentialsR-ambari-2.1-HDP-2.3-8.2.0.0-1.x86_64.rpm
```

- Restart your Ambari server.
`ambari-server restart`
- Log on to your Ambari server and install SPSS Essentials for R as a service via the Ambari console. SPSS Essentials for R should be installed on every host where Analytic Server and the Analytic Metastore is installed.

Note: Ambari will install `gcc-c++` and `gcc-gfortran` (RHEL) and `gcc-fortran` (SUSE) prior to installing R. Ensure that the servers where R is to be installed and executed are configured to download `gcc-c++` and `gcc-[g]fortran` RPMs or have GCC and FORTRAN compilers installed.

- Refresh the Analytic Server service.
- Run the `update_clientdeps` script using the instructions in “Updating client dependencies” on page 17.
- You must also install Essentials for R on the machine that hosts SPSS Modeler Server. See the SPSS Modeler documentation for details.

Configuring IBM SPSS Modeler for use with IBM SPSS Analytic Server

In order to enable SPSS Modeler for use with Analytic Server, you need to make some updates to the SPSS Modeler Server installation.

- Configure SPSS Modeler Server to associate it with an Analytic Server installation.
 - Edit the `options.cfg` file in the `config` subdirectory of the main server installation directory, and add or edit the following lines:

```
as_ssl_enabled, {Y|N}
as_host, "{AS_SERVER}"
as_port, PORT
as_context_root, "{CONTEXT-ROOT}"
as_tenant, "{TENANT}"
as_prompt_for_password, {Y|N}
as_kerberos_auth_mode, {Y|N}
as_kerberos_krb5_conf, {CONF-PATH}
as_kerberos_krb5_spn, {AS-SPN}
```

as_ssl_enabled

Specify Y if secure communication is configured on Analytic Server; otherwise, N.

as_host

The IP address of the server that hosts Analytic Server.

as_port

The port on which Analytic Server is listening (by default this is 8080).

as_context_root

The Analytic Server context root (by default this is `analyticserver`).

as_tenant

The tenant the SPSS Modeler Server installation is a member of (the default tenant is `ibm`).

as_prompt_for_password

Specify N if the SPSS Modeler Server is configured with the same authentication system for users and passwords as that used on Analytic Server; for example, when using Kerberos authentication. Otherwise, specify Y.

When running SPSS Modeler in batch mode, you add `-analytic_server_username {ASusername} -analytic_server_password {ASpassword}` as arguments to the `clemb` command.

as_kerberos_auth_mode

Specify Y to enable Kerberos SSO from SPSS Modeler.

as_kerberos_krb5_conf

Specify the path to the Kerberos configuration file that Analytic Server should use; for example, `\etc\krb5.conf`.

as_kerberos_krb5_spn

Specify the Analytic Server Kerberos SPN; for example, `HTTP/ashost.mydomain.com@MYDOMAIN.COM`.

- b. Restart the SPSS Modeler Server service.

In order to connect to an Analytic Server installation that has SSL/TLS enabled, there are some further steps to configuring your SPSS Modeler Server and client installations.

- a. Navigate to `http{s}://{HOST}:{PORT}/{CONTEXT-ROOT}/admin/{TENANT}` and log on to the Analytic Server console.
- b. Download the certification file from the browser and save it to your file system.
- c. Add the certification file to the JRE of both your SPSS Modeler Server and SPSS Modeler Client installations. The location to update can be found under the `/jre/lib/security/cacerts` subdirectory of the SPSS Modeler installation path.
 - 1) Make sure the `cacerts` file is not read-only.
 - 2) Use the `keytool` program Modeler ships with – this can be found in the `/jre/bin/keytool` subdirectory of the SPSS Modeler installation path.

Run the following command

```
keytool -import -alias <as-alias> -file <cert-file> -keystore "<cacerts-file>"
```

Note that `<as-alias>` is an alias for the `cacerts` file. You can use any name you like as long as it is unique to the `cacerts` file.

So an example command would look like the following.

```
keytool -import -alias MySSLCertAlias -file C:\Download\as.cer  
-keystore "c:\Program Files\IBM\SPSS\Modeler\{ModelerVersion}\jre\lib\security\cacerts"
```

- d. Restart your SPSS Modeler Server and SPSS Modeler Client .

2. [optional] Install IBM SPSS Modeler - Essentials for R , if you plan to score R models in streams with Analytic Server data sources. IBM SPSS Modeler - Essentials for R is available for download (<https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=swg-tspssp>).

Enabling relational database sources

Analytic Server can use relational database sources if you supply the JDBC drivers in a shared directory on each Analytic Server host. By default, this directory is `/usr/share/jdbc`.

To change the shared directory, follow these steps.

1. In the Ambari Services tab, navigate to the Configs tab of the Analytic Server service.
2. Open the **Advanced analytics.cfg** section.
3. Specify the path of the shared directory of JDBC drivers in **jdbc.drivers.location**.
4. Click **Save**.
5. Stop the Analytic Server service.
6. Click **Refresh**.
7. Start the Analytic Server service.

Table 1. Supported databases

Database	Supported versions	JDBC driver jars	Vendor
Amazon Redshift	8.0.2 or later	postgresql-8.0.2-xxxx.jdbc41.jar or later	PostgreSQL
DB2 for Linux, UNIX, and Windows	10.5, 10.1, 9.7	db2jcc.jar	IBM

Table 1. Supported databases (continued)

Database	Supported versions	JDBC driver jars	Vendor
DB2 z/OS	11, 10	db2jcc.jar, db2_license_cisuz.jar	IBM
Greenplum	5, 4.2.x	postgresql.jar	Greenplum
Netezza	7, 6.x	nzjdbc.jar	IBM
Oracle	12c, 11g R2 (11.2)	ojdbc6.jar, orai18n.jar	Oracle
SQL Server	2014, 2012, 2008 R2	sqljdbc4.jar	Microsoft
Sybase IQ	16.x, 15.4, 15.2	jconnect70.jar	Sybase
Teradata	14, 14.1, 15	tdgssconfig.jar, terajdbc4.jar	Teradata

Enabling HCatalog data sources

Analytic Server provides support for a number of data sources through Hive/HCatalog. Some sources require manual configuration steps.

1. Collect the necessary JAR files to enable the data source. See the sections below for details.
2. Add these JAR files to the {HIVE_HOME}/auxlib directory and to the /usr/share/hive directory on each Analytic Server node.
3. Restart the Hive Metastore service.
4. Refresh the Analytic Metastore service.
5. Restart each and every instance of the Analytic Server service.

NoSQL databases

Analytic Server supports any NoSQL database for which a Hive storage handler is available from the vendor.

No additional steps are necessary to enable support for Apache HBase and Apache Accumulo.

For other NoSQL databases, contact the database vendor and obtain the storage handler and related jars.

File-based Hive tables

Analytic Server supports any file-based Hive tables for which a built-in or custom Hive SerDe (serializer-deserializer) is available.

The Hive XML SerDe for processing XML files is located in the Maven Central Repository at <http://search.maven.org/#search%7Cga%7C1%7Chivexmlserde>.

Changing ports used by Analytic Server

Analytic Server uses the 9080 port for HTTP and the 9443 port for HTTPS by default. To change the port settings, follow these steps.

1. In the Ambari Services tab, navigate to the Configs tab of the Analytic Server service.
2. Open the **Advanced analytics.cfg** section.
3. Specify the desired HTTP and HTTPS ports in **http.port** and **https.port**, respectively.
4. Click **Save**.
5. Restart the Analytic Server service.

High availability Analytic Server

You can make Analytic Server highly available by adding it as a service to multiple nodes in your cluster.

1. In the Ambari console, navigate to the Hosts tab.
2. Select a host that is not already running Analytic Server as a service.
3. On the Summary tab, click **Add** and select Analytic Server.
4. Click **Confirm Add**.

Optimizing JVM options for small data

You can edit JVM properties in order to optimize your system when running small (M3R) jobs.

In the Ambari console, see the Advanced analytics-jvm-options section of the Configs tab in the Analytic Server service. Modifying the following parameters sets the heap size for jobs run on the server that hosts Analytic Server; that is, not Hadoop. This is important if running small (M3R) jobs, and you may need to experiment with these values to optimize your system.

```
-Xms512M  
-Xmx2048M
```

Updating client dependencies

This section describes how to update the Analytic Server service's dependencies using the `update_clientdeps` script.

1. Login to Ambari server host as root.
2. Change directory to `/var/lib/ambari-server/resources/stacks/<stack-name>/<stack-version>/services/ANALYTICSERVER/package/scripts`; see the following example.

```
cd "/var/lib/ambari-server/resources/stacks/HDP/2.3/services/ANALYTICSERVER/package/scripts"
```
3. Run the `update_clientdeps` script with the following arguments.

```
-u <ambari-user>  
    The Ambari account username  
  
-p <ambari-password>  
    The password for the Ambari account user.  
  
-h <ambari-host>  
    The hostname of the Ambari server.  
  
-x <ambari-port>  
    The port on which Ambari is listening.
```

See the following example.

```
./update_clientdeps.sh -u admin -p admin -h host.domain -x 8080
```

4. Restart the Ambari server using the following command.

```
ambari-server restart
```

Chapter 4. Migration

Analytic Server allows you to migrate data and configuration settings from an existing Analytic Server installation to a new installation.

Upgrade to a new version of Analytic Server

If you have an existing installation of Analytic Server 2.0 and have purchased 2.1, then you can migrate your 2.0 configuration settings to your 2.1 installation.

Restriction: If you have a version earlier than 2.0 installed, you must first migrate from the earlier version to 2.0 and then from version 2.0 to 2.1.

Restriction: Your 2.0 and 2.1 installations cannot coexist on the same Hadoop cluster. If you configure your 2.1 installation to use the same Hadoop cluster as your 2.0 installation, the 2.0 installation will no longer function.

Migration steps, 2.0 to 2.1

1. Install the new installation of Analytic Server according to the instructions in “Installation” on page 6.
2. Copy the analytic root from your old installation to your new one.

- a. If you are unsure of the location of the analytic root, run `hadoop -fs ls`. The path to the analytic root will be of form `/user/aeuser/analytic-root`, where `aeuser` is the userid that owns the analytic root.

- b. Change ownership of the analytic root from `aeuser` to `as_user`

```
hadoop dfs -chown -R {as_user:{group}} {path to 2.0 analytic-root}
```

Note: If you plan to use the existing Analytic Server installation after migration, make a copy of the `analytic-root` directory in HDFS and then change the ownership on the copy of the directory.

- c. Log in to the host of the new Analytic Server installation as `as_user`. Delete the `/user/as_user/analytic-root` directory, if it exists.

- d. Run the following copy script.

```
hadoop distcp http://{host of 2.0 namenode}:50070/{path to 2.0 analytic-root} hdfs://{host of 2.1 namenode}/user/as_user/analytic-root
```

3. In the Ambari console, stop the Analytic Server service.
4. Ensure the Analytic Metastore service is running.
5. Collect the configuration settings from the old installation.
 - a. Copy the `configcollector.zip` archive in your new installation to `{AS_ROOT}\tools` in your old installation.
 - b. Extract the copy of `configcollector.zip`. This creates a new `configcollector` subdirectory in your old installation.
 - c. Run the configuration collector tool in your old installation by executing the **configcollector** script in `{AS_ROOT}\tools\configcollector`. Copy the resulting compressed (ZIP) file to the server that hosts your new installation.

6. Run the migration tool by executing the **migrationtool** script and passing the path of the compressed file created by the configuration collector as an argument. An example follows.

```
migrationtool.sh /opt/ibm/spss/analyticserver/2.1/ASConfiguration_2.0.0.0.xxx.zip
```

7. In the Ambari console, start the Analytic Server service.

Note: If you configured R for use with the existing Analytic Server installation, you will need to follow the steps to configure it with the new Analytic Server installation.

Chapter 5. Uninstalling

1. On the Analytic Metastore host, run the `remove_as.sh` script in the `{AS_ROOT}/bin` directory with the following parameters.

- u** Required. The Ambari Server administrator's user ID.
- p** Required. The Ambari Server administrator's password.
- h** Required. The Ambari Server host name.
- x** Required. The Ambari Server port.
- l** Optional. Enables secure mode.

Examples follow.

```
remove_as.sh -u admin -p admin -h one.cluster -x 8081
```

Removes Analytic Server from a cluster with Ambari host `one.cluster`.

```
remove_as.sh -u admin -p admin -h one.cluster -x 8081 -l
```

Removes Analytic Server from a cluster with Ambari host `one.cluster`, in secure mode.

Note: This operation removes the Analytic Server folder on HDFS.

Note: This operation does not remove any DB2 schemas associated with Analytic Server. Consult the DB2 documentation for information on manually removing schemas

Chapter 6. Troubleshooting

This section describes some common installation and configuration issues and how you can fix them.

General issues

Installation succeeds with warnings, but users are unable to create data sources with error "Unable to complete the request. Reason: Permission denied"

Setting the **distrib.fs.root** parameter to a directory that the Analytic Server user (by default, `as_user`) doesn't have access to will result in errors. Make certain that the Analytic Server user is authorized to read, write, and execute the **distrib.fs.root** directory.

Issues with specific Hadoop distributions

Refresh action for Analytic Server service is disabled on Hortonworks 2.3

To manually refresh Analytic Server libraries on Hortonworks 2.3 use the following steps.

1. Log on to the host running the Analytic Metastore as the Analytic Server user (by default `as_user`).

Note: You can find this host name from the Ambari console.

2. Run the **refresh** script in the directory `{AS_ROOT}/bin`; for example:

```
cd /opt/ibm/spss/analyticserver/2.1/bin
./refresh
```

3. Restart the Analytic Server service in the Ambari console.

High availability clusters

Analytic Server cannot be added to more hosts due to changes in dependencies

Run the `update_clientdeps` script using the instructions in "Updating client dependencies" on page 17.

java.net.SocketTimeoutException: Read timed out

Change the Liberty ND timeout environment variable as follows:

```
export LIBERTYND_READ_TIMEOUT=<milliseconds>
```

where `<milliseconds>` is the number of seconds to use for the JMX read timeout.

java.io.IOException: CWWKX7202E: The timeout value 60 (seconds) for command ./server start expired

Add the following to the Controller Server `server.xml`

```
<!-- Increase start and stop server timeout to accommodate slow hardware -->
<serverCommands startServerTimeout="120" stopServerTimeout="120"/>
```

java.lang.OutOfMemoryError: Java heap space

Add the following lines to `jvm.options` on every member of the HA cluster.

```
-Xms512M
-Xmx2048M
```

"The Analytic Cluster Service has unexpectedly lost contact with Zookeeper, this JVM is being terminated to maintain cluster integrity."

One thing that may cause this is if the amount of data being written to Zookeeper is too large. If, in the Zookeeper logs are exceptions like:

```
java.io.IOException: Unreasonable length = 2054758
```

or in the Analytic Server logs are messages like:

Caused by: java.io.UTFDataFormatException: encoded string too long: 2054758 bytes
at java.io.DataOutputStream.writeUTF(DataOutputStream.java:375)

1. In the Ambari console, navigate to the Zookeeper service Configs tab and add the following line to the env-template, then restart the Zookeeper service.

```
export JVMFLAGS="-Xmx2048m -Djute.maxbuffer=2097152"
```

2. In the Ambari console, navigate to the Analytic Server service Configs tab and add the following in the Advanced analytics-jvm-options, then restart the Analytic Cluster service.

```
-Djute.maxbuffer=2097152
```

The number to specify for the jute.maxbuffer setting should be higher than the number indicated in the exception messages.

Zookeeper transaction data becomes unmanageable

Set the **autopurge.purgeInterval** parameter in `zoo.cfg` to 1 to enable automatic purges of the Zookeeper transaction log.

Analytic cluster service loses contact with Zookeeper

Review and modify the **tickTime**, **initLimit**, and **syncLimit** parameters in `zoo.cfg`. For example:

```
# The number of milliseconds of each tick
tickTime=2000
# The number of ticks that the initial
# synchronization phase can take
initLimit=30
# The number of ticks that can pass between
# sending a request and getting an acknowledgement
syncLimit=15
```

See the Zookeeper documentation for details: <https://zookeeper.apache.org/doc/r3.3.3/zookeeperAdmin.html>

Analytic Server jobs do not resume

There are two common situations in which Analytic Server jobs do not resume.

1. When an Analytic Server job fails because a cluster member fails, the job is normally restarted automatically on another cluster member. If the job does not resume, check to ensure there are at least 4 cluster members in the High Availability cluster.
2. When you quiesce a cluster member, then all of the Analytic Server jobs on that server are normally resumed on another cluster member. In order to ensure that the jobs resume, set `-Dcom.spss.ae.remoteclient.failover.threshold=100` and use remote mode.

Analytic Server servers hang occasionally upon server shutdown

Kill the server manually.

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
U.S.A.

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Licensees of this program who want to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Software Group
ATTN: Licensing
200 W. Madison St.
Chicago, IL; 60606
U.S.A.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

All IBM prices shown are IBM's suggested retail prices, are current and are subject to change without notice. Dealer prices may vary.

This information is for planning purposes only. The information herein is subject to change before the products described become available.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Each copy or any portion of these sample programs or any derivative work, must include a copyright notice as follows:

© your company name) (year). Portions of this code are derived from IBM Corp. Sample Programs.

© Copyright IBM Corp. _enter the year or years_. All rights reserved.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at www.ibm.com/legal/copytrade.shtml.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

ITIL is a registered trademark, and a registered community trademark of The Minister for the Cabinet Office, and is registered in the U.S. Patent and Trademark Office.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.



Printed in USA