

**IBM SPSS Analytic Server  
V3.1.2**

## **安装和配置指南**

**IBM**

**注释**

在使用本信息及其支持的产品之前，请先阅读第 67 页的『声明』中的信息。

**产品信息**

此版本是用于 IBM SPSS Analytic Server 的 V3.1.2 以及后续发行版和修订版，直至在新版本中另有说明为止。

# 目录

<b>第 1 章 先决条件</b> . . . . .	<b>1</b>
<b>第 2 章 Ambari 安装和配置</b> . . . . .	<b>3</b>
特定于 Ambari 的先决条件 . . . . .	3
安装预检查和后检查工具 - Ambari . . . . .	3
在 Ambari 上安装 . . . . .	6
联机安装. . . . .	6
脱机安装. . . . .	9
在由外部管理的 MySQL 环境中安装 Analytic Server . . . . .	14
配置. . . . .	14
安全性 . . . . .	15
启用对 Essentials for R 的支持 . . . . .	19
启用关系数据库源 . . . . .	21
启用 HCatalog 数据源. . . . .	23
更改 Analytic Server 所使用的端口 . . . . .	24
高可用性 Analytic Server . . . . .	24
针对小型数据优化 JVM 选项 . . . . .	25
更新客户机依赖关系 . . . . .	25
配置 Apache Knox. . . . .	26
针对每个 IBM SPSS Analytic Server 租户配置单独的 YARN 请求 - HDP. . . . .	30
在 Ambari 上迁移 IBM SPSS Analytic Server . . . . .	31
卸载. . . . .	33
卸载 Essentials for R . . . . .	34
<b>第 3 章 Cloudera 安装和配置</b> . . . . .	<b>35</b>
Cloudera 概述 . . . . .	35
特定于 Cloudera 的先决条件 . . . . .	35
启用 Kerberos 的 Cloudera 环境 . . . . .	35
为 Analytic Server 配置 MySQL . . . . .	37
安装预检查和后检查工具 - Cloudera . . . . .	37
在 Cloudera 上安装 . . . . .	39
配置 Cloudera . . . . .	44
安全性 . . . . .	45
启用对 Essentials for R 的支持 . . . . .	49
启用关系数据库源 . . . . .	50
启用 HCatalog 数据源. . . . .	51
配置 Apache Impala . . . . .	53
更改 Analytic Server 所使用的端口 . . . . .	54
高可用性 Analytic Server . . . . .	54
针对小型数据优化 JVM 选项 . . . . .	55
针对每个 IBM SPSS Analytic Server 租户配置单独的 YARN 请求 - Cloudera . . . . .	55
迁移. . . . .	56
在 Cloudera 上卸载 Analytic Server. . . . .	57
<b>第 4 章 配置 IBM SPSS Modeler 以用于 IBM SPSS Analytic Server</b> . . . . .	<b>59</b>
<b>第 5 章 使用 SLM 标记来跟踪许可证发放</b> . . . . .	<b>61</b>
<b>第 6 章 故障诊断</b> . . . . .	<b>63</b>
声明 . . . . .	67
商标. . . . .	68



---

## 第 1 章 先决条件

在安装 Analytic Server 之前，请复审以下信息。

### 系统需求

有关最新的系统需求信息，请使用 IBM 技术支持站点上的详细系统需求报告：<http://publib.boulder.ibm.com/infocenter/prodguid/v1r0/clarity/softwareReqsForProduct.html>。在此页面上：

1. 输入 SPSS Analytic Server 作为产品名称并单击**搜索**。
2. 选择想要的版本和报告范围，然后单击**提交**。

### WebSocket 流量

您必须确保防火墙、VPN 或其他端口阻止方法未阻止客户机和 Analytic Server 之间的 WebSocket 流量。WebSocket 端口与常规 Analytic Server 端口相同。

### SuSE Linux (SLES) 12

在 SuSE Linux 12 上安装 Analytic Server 之前，执行以下任务：

1. 从以下 URL 将公用密钥下载到主机：

```
https://ibm-open-platform.ibm.com/repos/IBM-SPSS-AnalyticServer/3.1.2.0/IBM-SPSS-ANALYTICSERVER-GPG-KEY.public
```

2. 通过在主机上运行以下命令，导入公用密钥：

```
rpm --import IBM-SPSS-ANALYTICSERVER-GPG-KEY.public
```

### Power 系统

确保在集群的所有主机上的 PATH 中安装并包含 IBM XLC 和 XLF 编译器。

您可以在以下 Web 站点找到有关获取这些编译器的许可证的更多信息：

- XL C for Linux: <http://www-03.ibm.com/software/products/en/xlcpp-linux>
- XL Fortran for Linux: <http://www-03.ibm.com/software/products/en/xlfortran-linux>

### Hortonworks 数据平台 (HDP)

在安装 Analytic Server 之前，必须确保至少在集群环境中部署了一个 HDP 客户机。由于托管 Ambari 管理器的节点预期的是 /usr/hdp 目录，因此如果缺少 HDP 客户机，Analytic Server 将会失败。

### Hive/HCatalog

如果您计划使用 NoSQL 数据源，请针对远程访问配置 Hive 和 HCatalog。另外，请确保 hive-site.xml 包含 `hive.metastore.uris` 属性，此属性具有指向活动 Thrift Hive Metastore 服务器的格式 `thrift://<host_name>:<port>`。请参阅 Hadoop 发行版文档以获取详细信息。

**注：** Analytic Server Metastore 不能安装在与 Hive Metastore 相同的机器上。

如果想要使用 Hive 2.1，必须在 Ambari 控制台中启用 Hive 2.1 来启用**交互查询**设置，然后在 Analytic Server 安装期间输入 2.x 作为 `hive.version` 属性。

1. 打开 Ambari 控制台，并在**分析服务器高级 analytics.cfg** 部分中添加以下属性。
  - 键: `hive.version`
  - 值: 输入相应的 Hive 版本 (例如 2.x)
2. 保存配置。

**注：** Hive 2.1 在具有 Spark 2.x 的 HDP 2.5 和 2.6 上受支持。

## 元数据存储库

缺省情况下, Analytic Server 安装并使用 MySQL 数据库。另外, 您可以配置 Analytic Server 以使用现有 Db2 安装。无论选择何种数据库类型, 都必须采用 UTF-8 编码。

### MySQL

MySQL 的缺省字符集取决于版本和操作系统。使用以下步骤以确定您的 MySQL 安装是否设置为 UTF-8。

1. 确定 MySQL 的版本。

```
mysql -V
```

2. 通过从 MySQL 命令行界面运行以下查询来确定 MySQL 的缺省字符集。

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

如果字符集已设置为 UTF-8, 那么无需进一步更改。

3. 通过从 MySQL 命令行界面运行以下查询来确定 MySQL 的缺省排序规则。

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

如果排序规则已设置为 UTF-8, 那么无需进一步更改。

4. 如果缺省字符集或排序规则不是 UTF-8, 请参阅 MySQL 文档以获取有关如何编辑 `/etc/my.cnf` 的详细信息, 然后重新启动 MySQL 守护程序以将字符集更改为 UTF-8。

**Db2** 有关配置 Db2 的更多信息, 请参阅 Knowledge Center [http://www-01.ibm.com/support/knowledgecenter/SSEPGG\\_10.5.0/com.ibm.db2.luw.kc.doc/welcome.html](http://www-01.ibm.com/support/knowledgecenter/SSEPGG_10.5.0/com.ibm.db2.luw.kc.doc/welcome.html)。

## 高可用性集群

### 负载均衡器

高可用性集群应具有支持会话亲缘关系 (有时候也称为粘滞会话) 的负载均衡器。Analytic Server 标识包含 `cookie="request-token"` 的会话。这标识要在应用程序控制的会话亲缘关系中使用的用户登录持续时间的会话。请参阅特定负载均衡器的文档以获取如何支持会话亲缘关系的详细信息。

### Analytic Server 作业失败

Analytic Server 作业因集群成员失败而失败时, 作业通常将在另一集群成员上自动重新启动。如果该作业未恢复, 请检查以确保高可用性集群中至少有 4 个集群成员。

---

## 第 2 章 Ambari 安装和配置

---

### 特定于 Ambari 的先决条件

除了一般的先决条件，还需查看以下信息。

**服务** Analytic Server 作为 Ambari 服务进行安装。在安装 Analytic Server 之前，您必须确保已将下列客户机作为 Ambari 服务安装：

- HDFS/HDFS\_CLIENT
- MAPREDUCE2/MAPREDUCE2\_CLIENT
- HIVE/HIVE\_CLIENT
- SPARK/SPARK\_CLIENT (如果使用 Spark 1.x)
- SPARK2/SPARK2\_CLIENT (如果使用 Spark 2.x)
- HBASE/HBASE\_CLIENT (如果使用 HBASE)
- YARN
- Zookeeper

### 无密码 SSH

在 Analytic Metastore 主机和集群中的所有主机之间为 root 用户设置无密码 SSH。

---

## 安装预检查和后检查工具 - Ambari

### 预检查工具概述

Analytic Server 安装预检查工具有助于在 Analytic Server 安装之前通过识别潜在的环境问题来减少安装问题和运行时错误。

预检查工具验证：

- 本地系统上的 OS 和 Ambari 版本
- 本地系统上的 OS ulimit 设置
- 本地系统上的可用磁盘空间
- Hadoop 版本
- Ambari 服务可用性 (HDFS、HCatalog、Spark、Hive、MapReduce、Yarn 和 Zookeeper, 等等)
- Analytic Server 指定 Ambari 设置

注：可在运行自解压 Analytic Server 二进制文件后使用预检查工具。

### 后检查工具概述

Analytic Server 安装后检查工具通过提交 REST API 处理请求，在 Analytic Server 安装之后识别配置问题：

- HDFS 中的数据
- Hive/HCatalog 中的数据
- 压缩的数据 (包括 deflate、bz2 和 snappy)
- PySpark 数据

- 使用本机 SPSS 组件（包括 alm、tree、neuralnet、scoring 和 tascoreing）的数据
- MapReduce 数据
- 内存 MapReduce 数据

## 工具位置和先决条件

在安装 Analytic Server 服务之前，在将属于 Analytic Server 服务的所有节点上运行预检查工具以验证 Linux 环境已准备好安装 Analytic Server。

在安装过程中自动调用预检查工具。工具在每个节点上运行安装之前检查 Analytic Metastore 和每个 Analytic Server 节点。您还可以手动在 Ambari 服务器节点上调用预检查工具，此工具将在安装服务之前验证机器。

运行自解压 Analytic Server 二进制文件后，预检查工具位于以下目录中：

- **HDP**

```
/opt/ibm/spss/analyticserver-ambari/3.1/ANALYTICSERVER/package/chktool/precheck.py  
  
[root@servername chktool]# cd /opt/ibm/spss/analyticserver-ambari/3.1/ANALYTICSERVER/package/chktool  
[root@servername chktool]# ls  
checkers data lib logs postcheck.py precheck.py readme.txt
```

安装 Analytic Server 后，后检查工具位于以下目录中：

- **HDP**

```
/opt/ibm/spss/analyticserver/3.1/tools/com.spss.ibm.checker.zip
```

该工具必须作为根目录运行，并需要 Python 2.6.X（或更高版本）。

如果预检查工具报告了任何故障，那么这些故障必须在您继续 Analytic Server 安装之前解决。

运行 Analytic Server 自抽取二进制文件（第 6 页的『在 Ambari 上安装』部分中的步骤 2）之后，chktool 目录将可用。如果您选择运行第 9 页的『脱机安装』，那么 chktool 目录将在安装元数据 RPM 之后可用。

## 运行预检查工具

### 自动

在通过 Ambari 控制台作为服务的一部分安装 Analytic Server 时，可作为 Analytic Server 安装的一部分自动调用预检查工具。您必须手动输入 Ambari 服务器用户名和密码：



▼ Advanced analytics-env

Analytic_Server_UserID	<input style="width: 90%;" type="text" value="3124"/>	<span style="color: green;">+</span> <span style="color: blue;">C</span>
ambari.user.name	<input style="width: 90%;" type="text" value="admin"/>	
ambari.user.password	<input style="width: 45%; height: 20px;" type="password" value="•••••"/> <input style="width: 45%; height: 20px;" type="password" value="•••••"/>	
as.database.type	<input style="width: 90%;" type="text" value="mysql"/>	<span style="color: green;">+</span> <span style="color: blue;">C</span>

图 1. *Advanced analytics-env* 设置

## 手动

您可以手动在 Ambari 服务器节点上调用预检查工具。

以下预检查示例检查了在 `myambarihost.ibm.com:8080` 上运行的启用了 SSL 的 Ambari 集群 MyCluster，并使用了登录凭证 `admin:admin:`

```
python ./precheck.py --target B --cluster MyCluster --username admin
--password admin --host myambarihost.ibm.com --port 8080 --as_host myashost.ibm.com --ssl
```

### 注释：

- `as_host` 值必须由任一 IP 地址或标准域名提供。
- 工具将在省略密码自变量时提示密码。
- `precheck.py` 命令包括使用情况帮助，它与 `--h` 自变量一起显示 (`python ./precheck.py --help`)。
- `--cluster` 自变量是可选的（未使用 `--cluster` 时标识当前集群）。

当预检查工具运行其检查时，每个检查的状态将显示在命令窗口中。发生故障时，在日志文件中提供详细信息（命令窗口中提供了确切的日志文件位置）。当需要更多支持时，日志文件可以提供给 IBM 技术支持。

## 运行后检查工具

后检查工具验证 Analytic Server 是否正常运行，并能够处理简单的作业。以下后检查示例检查了在 `myanalyticserverhost.ibm.com:9443` 上运行的启用了 SSL 的 Analytic Server 实例，并使用了登录凭证 `admin:ibmspss:`

```
python ./postcheck.py --host myanalyticserverhost.ibm.com --port 9443
--username admin --password ibmspss --ssl
```

当 Knox 与 Analytic Server 一起使用时，命令如下所示：

```
python ./postcheck.py --host myknoxserverhost.ibm.com --port 8443
--username admin --password ibmspss --ssl --gateway_url /gateway/default
```

要执行单个检查，请使用以下命令：

```
python ./postcheck.py --host myknoxserverhost.ibm.com --port 8443
--username admin --password ibmspss --ssl --gateway_url /gateway/default --check AS_PYSPARK_BUILDMODEL
```

### 注释：

- 工具将在省略密码自变量时提示密码。
- `postcheck.py` 命令包括用法帮助，它与 `--h` 自变量一起显示 (`python ./postcheck.py --help`)。

当后检查工具运行其检查时，每个检查的状态将显示在命令窗口中。发生故障时，在日志文件中提供详细信息（命令窗口中提供了确切的日志文件位置）。当需要更多支持时，日志文件可以提供给 IBM 技术支持。

## 在 Ambari 上安装

基本过程是在 Ambari 集群中的主机上安装 Analytic Server 文件，然后将 Analytic Server 添加为 Ambari 服务。

### 『联机安装』

如果 Ambari 服务器主机以及集群中的所有节点都能够访问 <https://ibm-open-platform.ibm.com>，那么选择联机安装。

### 第 9 页的『脱机安装』

如果 Ambari 服务器主机无法访问因特网，请选择脱机安装。

## 联机安装

如果 Ambari 服务器主机以及集群中的所有节点都能够访问 <https://ibm-open-platform.ibm.com>，那么选择联机安装。

1. 浏览至 [IBM Passport Advantage® Web 站点](https://ibm-passport-advantage.com)，并将特定于您的堆栈、堆栈版本和硬件体系结构的自解压二进制文件下载到 Ambari 管理器节点。可用的 Ambari 二进制文件为：

表 1. Analytic Server 自抽取二进制文件

描述	二进制文件名
IBM® SPSS® Analytic Server 3.1.2 for Hortonworks Data Platform 2.5 and 2.6 Linux x86-64 英文版	spss_as-3.1.2.0-hdp2.5-2.6-1x86.bin
IBM SPSS Analytic Server 3.1.2 for Hortonworks Data Platform 2.6 Linux on System p LE 英文版	spss_as-3.1.2.0-hdp2.6-1ppc64.bin

2. 执行自解压二进制文件并遵循指示信息以查看许可证，接受许可证，选择联机安装，然后选择 Analytic Server 使用的数据库类型的安装过程。为您提供以下数据库类型选项：
  - 新的 MySQL 实例
  - 预先存在的 MySQL 或 Db2 实例
3. 从 `/var/lib/ambari-server/resources/stacks/<stack-name>/<stack-version>/services/ANALYTICSERVER/package/scripts` 目录，使用相应的自变量运行 `update_clientdeps.sh` 脚本（对于示例使用 `--help` 自变量）。
4. 重新启动 Ambari 服务器。

```
ambari-server restart
```
5. 登录 Ambari 服务器并通过 Ambari UI 将 Analytic Server 作为服务进行安装。

### 元数据存储库

缺省情况下，Analytic Server 使用 MySQL 来跟踪数据源、项目和租户的相关信息。在安装期间，您需要提供 Analytic Server 和 MySQL 之间的 JDBC 连接中使用的用户名 (`metadata.repository.user.name`) 和密码 `metadata.repository.password`。安装程序会在 MySQL 数据库中创建用户，但该用户特定于 MySQL 数据库并且无需是现有 Linux 或 Hadoop 用户。

要将元数据存储库更改为 Db2，请执行下列步骤。

注：您无法在安装完成后更改元数据存储库。

- a. 确保 Db2 安装在另一台机器上。有关更多信息，请参阅主题第 1 页的第 1 章，『先决条件』的元数据存储库部分。
- b. 在"Ambari 服务"选项卡中，浏览至 Analytic Server 服务的"配置"选项卡。
- c. 打开 **Advanced analytics-env** 部分。
- d. 将 **as.database.type** 的值从 `mysql` 更改为 `db2`。
- e. 打开 **Advanced analytics-meta** 部分。
- f. 将 **metadata.repository.driver** 的值从 `com.mysql.jdbc.Driver` 更改为 `com.ibm.db2.jcc.DB2Driver`。
- g. 将 **metadata.repository.url** 的值更改为 `jdbc:db2://{Db2_HOST}:{PORT}/{DBName};currentSchema={SchemaName};`，其中
  - {Db2\_HOST} 是安装了 Db2 的服务器的主机名
  - {PORT} 是 Db2 正在侦听的端口
  - {SchemaName} 是未使用的可用模式。
 如果不确定要输入什么值，请咨询 Db2 管理员。
- h. 在 **metadata.repository.user.name** 和 **metadata.repository.password** 中提供有效的 Db2 凭证。
- i. 单击**保存**。

## LDAP 配置

Analytic Server 使用 LDAP 服务器来存储和认证用户和组。在 Analytic Server 安装期间提供必需的 LDAP 配置信息。

表 2. LDAP 配置设置

LDAP 设置	描述
<code>as.ldap.type</code>	LDAP 类型。值可以是 <code>ads</code> 、 <code>ad</code> 或 <code>openldap</code> 。 <ul style="list-style-type: none"> <li>• <code>ads</code> - Apache Directory Server (缺省设置)</li> <li>• <code>ad</code> - Microsoft Active Directory</li> <li>• <code>openldap</code> - OpenLDAP</li> </ul>
<code>as.ldap.host</code>	LDAP 主机
<code>as.ldap.port</code>	LDAP 端口号
<code>as.ldap.binddn</code>	LDAP 绑定 DN
<code>as.ldap.bindpassword</code>	LDAP 绑定 DN 密码
<code>as.ldap.basedn</code>	LDAP 基本 DN
<code>as.ldap.filter</code>	LDAP 用户和组过滤器规则
<code>as.ldap.ssl.enabled</code>	指定是否使用 SSL 在 Analytic Server 和 LDAP 之间进行通信。该值可以是 <code>true</code> 或 <code>false</code> 。
<code>as.ldap.ssl.reference</code>	LDAP SSL 引用标识
<code>as.ldap.ssl.content</code>	LDAP SSL 配置

- 缺省情况下，`as.ldap.type` 设置为 `ads`，其他相关设置包含缺省设置。例外是，必须针对 `as.ldap.bindpassword` 设置提供密码。Analytic Server 使用配置设置以安装 Apache Directory Server (ADS) 和运行服务器初始化。缺省 ADS 概要文件包含用户 `admin` 以及 `admin` 的密码。您可以通过 Analytic Server 控制台执行用户管理，或者通过 `<Analytic Root>/bin` 文件夹中的 `importUser.sh` 脚本，从 XML 文件导入用户和组信息。

- 如果计划使用外部 LDAP 服务器，例如，Microsoft Active Directory 或 OpenLDAP，那么必须根据实际 LDAP 值定义配置设置。有关更多信息，请参阅在 Liberty 中配置 LDAP 用户注册表。
- 安装 Analytic Server 后，您可以更改 LDAP 配置（例如，从 Apache Directory Server 更改为 OpenLDAP）。但是，如果初始从 Microsoft Active Directory 或 OpenLDAP 开始，并且决定以后切换到 Apache Directory Server，那么 Analytic Server 不会在安装期间安装 ApacheDirectory Server。仅在初始 Analytic Server 安装期间选中时，才会安装 Apache Directory Server。

The screenshot displays the configuration interface for 'Advanced analytics-ldap'. The fields and their values are as follows:

- as.ldap.basedn:** dc=ibm,dc=com
- as.ldap.binddn:** uid=admin,ou=system
- as.ldap.bindpassword:** Two masked password input fields.
- as.ldap.filter:**

```
<customFilters id="customFilters"
userFilter="(&cn=%v)(objectClass=organizationalPerson)"
groupFilter="(&cn=%v)(objectclass=groupOfNames)"
useridMap="":cn"
groupidMap="":cn"
```
- as.ldap.host:** {analytic\_metastore\_host}
- as.ldap.port:** 10636
- as.ldap.ssl.content:**

```
<ssl id="LDAPSSLSettings" keyStoreRef="LDAPTrustStore"
trustStoreRef="LDAPTrustStore" />
<keyStore id="LDAPTrustStore" location="/opt/ibm/spss/analyticserver/{as_version}
/ads/public/trustads.jks" type="JKS" password="changeit" />
```
- as.ldap.ssl.enabled:** true
- as.ldap.ssl.reference:** LDAPSSLSettings
- as.ldap.type:** ads

图 2. 示例 LDAP 配置设置

#### 安装后不应进行更改的配置设置

在安装后请勿更改以下设置，否则 Analytic Server 将无法工作。

- Analytic\_Server\_User
- Analytic\_Server\_UserID
- as.database.type
- metadata.repository.driver

- distrib.fs.root
6. 现在，您已有 Analytic Server 的正常运行示例。可以选择进行进一步配置。有关配置和管理 Analytic Server 的更多信息，请参阅主题：第 14 页的『配置』。有关将现有配置迁移到新安装的信息，请参阅主题：第 31 页的『在 Ambari 上迁移 IBM SPSS Analytic Server』。
  7. 打开 Web 浏览器并输入地址 `http://<host>:<port>/analyticserver/admin/ibm`，其中 `<host>` 是 Analytic Server 主机的地址，而 `<port>` 是 Analytic Server 正在侦听的端口。缺省情况下，此端口为 9080。此 URL 将打开 Analytic Server 控制台的登录对话框。以 Analytic Server 管理员身份登录。缺省情况下，此用户标识是 `admin` 并且密码为 `admin`。

## 脱机安装

可自动或手动完成 IBM SPSS Analytic Server 脱机安装。

### 『HDP 上的自动安装』

自动安装过程利用 Ambari REST API，是首选安装方法。

### 第 10 页的『HDP (RHEL 和 SLES) 上的手动安装』

用于手动在 Hortonworks Data Platform 上安装 Analytic Server

### 第 12 页的『HDP (Ubuntu) 上的手动安装』

用于手动在 Ubuntu Linux 上安装 Analytic Server。

## HDP 上的自动安装

自动安装过程利用 Ambari REST API，是首选安装方法。

### 要点：

- 脱机自动安装过程将安装嵌入式 Apache Directory Server (ADS)。如果想要使用第三方 LDAP 服务器，那么可在完成 IBM SPSS Analytic Server 安装后配置 LDAP 设置。
- 脱机自动安装过程只能安装单个 Analytic Server 服务实例。在完成初始安装后，可添加更多实例。
- 脱机自动安装过程不支持在启用 Kerberos 的集群上安装 Analytic Server。

这些限制不适用于手动 HDP 或 Ubuntu 安装。

1. 浏览至 [IBM Passport Advantage® Web 站点](https://ibm-open-platform.ibm.com)，将自解压二进制文件下载到可访问 `https://ibm-open-platform.ibm.com` 的计算机。

表 3. *Analytic Server* 自抽取二进制文件

描述	二进制文件名
IBM SPSS Analytic Server 3.1.2 for Hortonworks Data Platform 2.5 and 2.6 Linux x86-64 英文版	<code>spss_as-3.1.2.0-hdp2.5-2.6-1x86.bin</code>
IBM SPSS Analytic Server 3.1.2 for Hortonworks Data Platform 2.6 Linux on System p LE 英文版	<code>spss_as-3.1.2.0-hdp2.6-1ppc64.bin</code>

2. 运行在步骤 1 中下载的可执行二进制文件并指定脱机安装。脱机安装将下载安装过程中后面所需的 RPM 或 DEB 文件，并且应在可访问 `https://ibm-open-platform.ibm.com` 的计算机上运行。已下载的文件位于当前可执行二进制文件目录 `./IBM-SPSS-AnalyticServer` 中。
3. 将可执行二进制文件文件目录 `./IBM-SPSS-AnalyticServer` 的全部内容从可访问因特网的机器复制到 Ambari 管理器节点（防火墙后面）。
4. 在 Ambari 管理器节点上，使用以下命令检查 Ambari 服务器是否正在运行：  
`ambari-server status`

- 在 Ambari 管理器节点和想要部署 Analytic Server 的所有其他节点上，安装创建本地 yum 存储库的工具。

```
yum install createrepo (RHEL, CentOS)
```

or

```
apt-get install dpkg-dev (Ubuntu)
```

- 在 Ambari 管理器节点上，运行可执行二进制文件 `./IBM-SPSS-AnalyticServer/packages/spss_as-ambari-offlineinstall.bin`。在安装期间，可执行二进制文件验证必需的 Analytic Server RPM/DEB 文件位于软件包目录中。您所需的 RPM 文件取决于分发版、版本和体系结构。

#### HDP 2.5 和 2.6 (x86\_64)

```
IBM-SPSS-AnalyticServer-ambari-2.x-3.1.2.0-1.noarch.rpm
```

```
IBM-SPSS-AnalyticServer-3.1.2.0-1.x86_64.rpm
```

#### HDP 2.6 (PPC64LE)

```
IBM-SPSS-AnalyticServer-ambari-2.x-3.1.2.0-1.noarch.rpm
```

```
IBM-SPSS-AnalyticServer-3.1.2.0-1.ppc64le.rpm
```

#### HDP 2.5 和 2.6 (Ubuntu)

```
IBM-SPSS-AnalyticServer-ambari-2.x_3.1.2.0_amd64.deb
```

```
IBM-SPSS-AnalyticServer_1_amd64.deb
```

在安装期间，将提示您输入 Analytic Server 版本、JDBC 驱动程序、Spark 版本、Hive 版本等等。

## HDP (RHEL 和 SLES) 上的手动安装

HDP (RHEL 和 SLES) 上的手动脱机安装的常规工作流程如下所示：

- 浏览至 [IBM Passport Advantage® Web 站点](https://ibm-open-platform.ibm.com)，将自解压二进制文件下载到可访问 <https://ibm-open-platform.ibm.com> 的计算机。

表 4. Analytic Server 自抽取二进制文件

描述	二进制文件名
IBM SPSS Analytic Server 3.1.2 for Hortonworks Data Platform 2.5 and 2.6 Linux x86-64 英文版	spss_as-3.1.2.0-hdp2.5-2.6-1x86.bin
IBM SPSS Analytic Server 3.1.2 for Hortonworks Data Platform 2.6 Linux on System p LE 英文版	spss_as-3.1.2.0-hdp2.6-1ppc64.bin

- 运行在步骤 1 中下载的可执行二进制文件并指定脱机安装。脱机安装将下载安装过程中后面所需的 RPM 文件，并且应在可访问 <https://ibm-open-platform.ibm.com> 的计算机上运行。已下载的文件位于当前可执行二进制文件目录 `./IBM-SPSS-AnalyticServer` 中。
- 将可执行二进制文件目录 `./IBM-SPSS-AnalyticServer` 的全部内容从可访问因特网的机器复制到 Ambari 管理节点的 `<AS_INSTALLABLE_HOME>` 目录（Ambari 管理器节点位于防火墙后面）。
- 在 Ambari 管理器节点上，使用以下命令检查 Ambari 服务器是否正在运行：

```
ambari-server status
```

- 安装创建本地 yum 存储库的工具。

```
yum install createrepo (RHEL, CentOS)
```

or

```
zypper install createrepo (SLES)
```

6. 创建用作 Analytic Server RPM 文件的存储库的目录。 请参阅以下示例。

```
mkdir /home/root/repos/IBM-SPSS-AnalyticServer/x86_64
```

7. 将必需的 Analytic Server RPM 文件复制到新目录。 您所需的 RPM 文件取决于分发版、版本和体系结构。

#### **HDP 2.5 和 2.6 (x86\_64)**

```
IBM-SPSS-AnalyticServer-ambari-2.x-3.1.2.0-1.noarch.rpm
```

```
IBM-SPSS-AnalyticServer-3.1.2.0-1.x86_64.rpm
```

#### **HDP 2.6 (PPC64LE)**

```
IBM-SPSS-AnalyticServer-ambari-2.x-3.1.2.0-1.noarch.rpm
```

```
IBM-SPSS-AnalyticServer-3.1.2.0-1.ppc64le.rpm
```

8. 创建本地存储库定义。 例如，在 `/etc/yum.repos.d/` 中创建名为 `IBM-SPSS-AnalyticServer-3.1.2.0.repo` 的文件（对于 RHEL、CentOS）或 `/etc/zypp/repos.d/`（对于 SLES），并包含下列内容。

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

9. 创建本地 yum 存储库。

```
createrepo /home/root/repos/IBM-SPSS-AnalyticServer/x86_64 (RHEL, CentOS, SLES)
```

10. 在 root 用户命令窗口中，`cd` 至 `<AS_INSTALLABLE_HOME>/IBM-SPSS-AnalyticServer` 目录，然后运行 `run ./offLineInstall.sh`。 脚本读取先前执行的二进制可执行安装命令的持久响应，并发出相应的平台命令（用于安装 rpm）。

注：仅当使用外部管理的 MySQL 环境时，步骤 11 才适用。

11. 在安装了将用作 `AS_MetaStore` 的节点/主机上运行 `add_mysql_user.sh` 脚本。

- a. 将 `add_mysql_user.sh` 脚本从 `<AS_INSTALLABLE_HOME>/IBM-SPSS-AnalyticServer` 复制到 MySQL 实例将用作 `AS_MetaStore` 的节点/主机。

- 在 MySQL node/host 上运行 `add_mysql_user.sh`。 例如 `./add_mysql_user.sh -u as_user -p spss -d aedb`

#### **注释：**

- 用户名和密码必须与在 Ambari 配置屏幕上输入的 `AS_Metastore` 的数据库用户名和密码相匹配。
- 可以手动更新 `add_mysql_user.sh` 脚本，以发出命令（如有需要）。
- 针对受保护的（root 用户访问权）MySQL 数据库运行 `add_mysql_user.sh` 脚本，使用 `-r` 和 `-t` 参数传入 `dbuserid` 和 `dbuserid_password`。 该脚本使用 `dbuserid` 和 `dbuserid_password` 来执行 MySQL 操作。

注：必须在 **AS\_Configuration** 屏幕 (**Advanced analytics-meta**) 上更改 `metadata.repository.url` 设置，以指向 MySQL 数据库主机。 例如，将 JDBC 设置 `mysql://{{analytic_metastore_host}}/aedb?createDatabaseIfNotExist=true` 更改为 `mysql://{MySQL_DB}/aedb?createDatabaseIfNotExist=true`

12. 通过添加下列行，更新通常位于 `/var/lib/ambari-server/resources/stacks/${stackName}/${stackVersion}/repos/` 中的 Ambari 存储库文件 `repoinfo.xml`，以使用本地 yum 存储库。

```
<os type="host_os">
  <repo>
    <baseurl>file:///{{path to local repository}}/</baseurl>
```

```

    <repoid>IBM-SPSS-AnalyticServer</repoid>
    <reponame>IBM-SPSS-AnalyticServer-3.1.2.0</reponame>
  </repo>
</os>

```

示例 {path to local repository} 将类似以下:

```
/home/root/repos/IBM-SPSS-AnalyticServer/x86_64/
```

13. 重复以下每个 Ambari 非服务器集群节点的步骤。

a. 将相应的 <AS\_INSTALLABLE\_HOME> 目录的全部内容从可以访问因特网的机器复制到 Ambari 非服务器集群节点。

b. 安装创建本地 yum 存储库的工具。

```
yum install createrepo (RHEL, CentOS)
```

or

```
zypper install createrepo (SLES)
```

c. 创建用作 Analytic Server RPM 文件的存储库的目录。 请参阅以下示例。

```
mkdir /home/root/repos/IBM-SPSS-AnalyticServer/x86_64
```

d. 将必需的 Analytic Server RPM 文件复制到新目录。 您所需的 RPM 文件取决于分发版、版本和体系结构。

#### **HDP 2.5 和 2.6 (x86\_64)**

```
IBM-SPSS-AnalyticServer-ambari-2.x-3.1.2.0-1.noarch.rpm
```

```
IBM-SPSS-AnalyticServer-3.1.2.0-1.x86_64.rpm
```

#### **HDP 2.6 (PPC64LE)**

```
IBM-SPSS-AnalyticServer-ambari-2.x-3.1.2.0-1.noarch.rpm
```

```
IBM-SPSS-AnalyticServer-3.1.2.0-1.ppc64le.rpm
```

e. 创建本地存储库定义。 例如，在 /etc/yum.repos.d/ 中创建名为 IBM-SPSS-AnalyticServer-3.1.2.0.repo 的文件（对于 RHEL、CentOS）或 /etc/zypp/repos.d/（对于 SLES），并包含下列内容。

```

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

```

f. 创建本地 yum 存储库。

```
createrepo /home/root/repos/IBM-SPSS-AnalyticServer/x86_64 (RHEL, CentOS, SLES)
```

14. 继续执行第 6 页的『联机安装』部分中的步骤 3。

## **HDP (Ubuntu) 上的手动安装**

HDP (Ubuntu) 上的手动脱机安装的常规工作流程如下所示:

1. 浏览至 [IBM Passport Advantage® Web 站点](https://ibm-open-platform.ibm.com)，将相应的 Ubuntu 自解压二进制文件下载到可访问 <https://ibm-open-platform.ibm.com> 的计算机。

表 5. Analytic Server 自抽取二进制文件

描述	二进制文件名
IBM SPSS Analytic Server 3.1.2 for Hortonworks Data Platform 2.5 and 2.6 Linux x86-64 英文版	spss_as-3.1.2.0-hdp2.5-2.6-1x86.bin



2. 运行在步骤 1 中下载的可执行二进制文件并指定脱机安装。脱机安装将下载安装过程中后面所需的 DEB 文件，并且应在可访问 <https://ibm-open-platform.ibm.com> 的计算机上运行。已下载的文件位于当前可执行二进制文件目录 `./IBM-SPSS-AnalyticServer` 中。
3. 将可执行二进制文件目录 `./IBM-SPSS-AnalyticServer` 的全部内容从可访问因特网的机器复制到 Ambari 管理节点的 `<AS_INSTALLABLE_HOME>` 目录（Ambari 管理器节点位于防火墙后面）。
4. 在 Ambari 管理器节点上，使用以下命令检查 Ambari 服务器是否正在运行：

```
ambari-server status
```

5. 创建用作 Analytic Server DEB 文件的存储库的 `<local_repo>` 目录。例如：

```
mkdir -p /usr/local/mydebs
```

6. 将必需的 Analytic Server DEB 文件复制到 `<local_repo>` 目录。

- `IBM-SPSS-AnalyticServer-ambari-2.x_3.1.2.0_amd64.deb`
- `IBM-SPSS-AnalyticServer_1_amd64.deb`

7. 创建本地存储库。

- a. 安装创建本地存储库的工具：

```
apt-get install dpkg-dev
```

- b. 生成源软件包文件：

```
cd <local_repo>
dpkg-scanpackages . /dev/null | gzip -9c > Packages.gz
```

- c. 创建本地存储库的组件（主）和体系结构（例如，`binary-i386` 和 `binary-amd64`）：

```
mkdir -p <local_repo>/dists/IBM-SPSS-AnalyticServer/main/binary-amd64/
mkdir -p <local_repo>/dists/IBM-SPSS-AnalyticServer/main/binary-i386/
```

- d. 复制源软件包：

```
cp -fr <local_repo>/Packages.gz <local_repo>/dists/IBM-SPSS-AnalyticServer/main/binary-amd64/Packages
cp -fr <local_repo>/Packages.gz <local_repo>/dists/IBM-SPSS-AnalyticServer/main/binary-i386/Packages
```

8. 创建本地存储库定义。例如，使用以下内容在 `/etc/apt/sources.list.d` 中创建名为 `IBM-SPSS-AnalyticServer-3.1.2.0.list` 的文件。

```
deb file:/usr/local/mydebs ./
```

9. 运行以下命令以更新存储库列表：

```
apt-get update
```

10. 运行以下命令以安装 IBM SPSS Analytic Server 3.1.2：

```
apt-get install ./IBM-SPSS-AnalyticServer-ambari-2.x
```

注：要验证本地存储库设置是否正确，请勿在 `<local_repo>` 目录中运行先前的命令。如果安装找不到软件包，那么意味着本地存储库设置不正确（在此情况下，必须验证所有先前步骤）。

11. 重复以下每个 Ambari 非服务器集群节点的步骤。

- a. 创建用作 Analytic Server DEB 文件的存储库的 `<local_repo>` 目录。例如：

```
mkdir -p /usr/local/mydebs
```

- b. 将 `<local_repo>` 目录的全部内容从 Ambari 管理器节点机器复制到 Ambari 非服务器集群节点的 `<local_repo>` 目录。该目录应包含以下文件：

- `<local_repo>/IBM-SPSS-AnalyticServer-ambari-2.x_3.1.2.0_amd64.deb`
- `<local_repo>/IBM-SPSS-AnalyticServer_1_amd64.deb`
- `<local_repo>/Packages.gz`
- `<local_repo>/dists/IBM-SPSS-AnalyticServer/main/binary-amd64/Packages`
- `<local_repo>/dists/IBM-SPSS-AnalyticServer/main/binary-i386/Packages`

- c. 创建本地存储库定义。例如，使用以下内容在 `/etc/apt/sources.list.d` 中创建名为 `IBM-SPSS-AnalyticServer-3.1.2.0.list` 的文件。

```
deb file:/usr/local/mydebs ./
```

12. 继续执行第 6 页的『联机安装』部分中的步骤 3。

## 在由外部管理的 MySQL 环境中安装 Analytic Server

在由外部管理的 MySQL 环境中安装 Analytic Server 的过程与正常安装不同。

以下步骤说明了在由外部管理的 MySQL 环境中安装 Analytic Server 的过程。

1. 浏览到 [IBM Passport Advantage® Web 站点](#)，并将特定于您的堆栈、堆栈版本和硬件体系结构的自抽取二进制文件下载到 Ambari 集群中的主机。
2. 执行自解压缩二进制文件并遵循指示信息以（可选）查看许可证，接受许可证。
  - a. 选择联机选项。
  - b. 在提示时选择外部 **MySQL 数据库** 选项。
3. 将 `add_mysql_user.sh` 脚本从 `<AS_INSTALLABLE_HOME>/IBM-SPSS-AnalyticServer` 复制到 MySQL 实例将用作 `AS_MetaStore` 的节点/主机。
  - 在 MySQL node/host 上运行 `add_mysql_user.sh`。例如 `./add_mysql_user.sh -u as_user -p spss -d aedb`

### 注释：

- 用户名和密码必须与在 Ambari 配置屏幕上输入的 `AS_Metastore` 的数据库用户名和密码相匹配。
  - 可以手动更新 `add_mysql_user.sh` 脚本，以发出命令（如有需要）。
  - 针对受保护的（root 用户访问权）MySQL 数据库运行 `add_mysql_user.sh` 脚本，使用 `-r` 和 `-t` 参数传入 `dbuserid` 和 `dbuserid_password`。该脚本使用 `dbuserid` 和 `dbuserid_password` 来执行 MySQL 操作。
4. 重新启动 Ambari 服务器。

```
ambari-server restart
```
  5. 通常在 Ambari 控制台上添加 AnalyticServer 服务（输入步骤 3 中输入的数据库用户名和密码）。

注：必须在 **AS\_Configuration** 屏幕 (**Advanced analytics-meta**) 上更改 `metadata.repository.url` 设置，以指向 MySQL 数据库主机。例如，将 JDBC 设置 `mysql://{analytic_metastore_host}/aedb?createDatabaseIfNotExist=true` 更改为 `mysql://{MySQL_DB}/aedb?createDatabaseIfNotExist=true`

---

## 配置

安装后，您可以通过 Ambari UI 有选择地配置和管理 Analytic Server。

注：以下约定用于 Analytic Server 文件路径。

- `{AS_ROOT}` 指的是部署 Analytic Server 的位置；例如 `/opt/IBM/SPSS/AnalyticServer/3.1`。
- `{AS_SERVER_ROOT}` 指的是配置文件、日志文件和服务器文件的位置；例如，`/opt/IBM/SPSS/AnalyticServer/3.1/ae_wlpserver/usr/servers/aeserver`。
- `{AS_HOME}` 指的是 HDFS 上 Analytic Server 用作根文件夹的位置。

## 安全性

### 配置 LDAP 注册表

LDAP 注册表允许您使用外部 LDAP 服务器（例如，Active Directory 或 OpenLDAP）对用户进行认证。

**要点：**在 Ambari 中，必须将 LDAP 用户指定为 Analytic Server 管理员。

以下是针对 OpenLDAP 的 ldapRegistry 示例。

```
<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>
```

以下示例提供使用 Active Directory 进行 Analytic Server 认证的方法：

```
<ldapRegistry id="Microsoft Active Directory" realm="ibm"
  host="host"
  port="389"
  baseDN="cn=users,dc=adtest,dc=mycompany,dc=com"
  bindDN="cn=administrator,cn=users,dc=adtest,dc=mycompany,dc=com"
  bindPassword="adminpassword"
  ldapType="Custom"
  <customFilters
    userFilter="(&(sAMAccountName=%v)(objectcategory=user))"
    groupFilter="(&(cn=%v)(objectcategory=group))"
    userIdMap="user:sAMAccountName"
    groupIdMap="*:cn"
    groupMemberIdMap="memberOf:member" />
</ldapRegistry>
```

**注：**使用 LDAP 查看器第三方工具通常有助于验证 LDAP 配置。

以下示例提供使用 Active Directory 进行 WebSphere Liberty 概要文件认证的方法：

```
<ldapRegistry id="ldap" realm="SampleLdapADRealm"
  host="ldapservers.mycity.mycompany.com" port="389" ignoreCase="true"
  baseDN="cn=users,dc=adtest,dc=mycity,dc=mycompany,dc=com"
  bindDN="cn=testuser,cn=users,dc=adtest,dc=mycity,dc=mycompany,dc=com"
  bindPassword="testuserpwd"
  ldapType="Microsoft Active Directory"
  sslEnabled="true"
  sslRef="LDAPSSLSettings">
  <activatedFilters
    userFilter="(&(sAMAccountName=%v)(objectcategory=user))"
    groupFilter="(&(cn=%v)(objectcategory=group))"
    userIdMap="user:sAMAccountName"
    groupIdMap="*:cn"
    groupMemberIdMap="memberOf:member" >
  </activatedFilters>
</ldapRegistry>

<ssl id="LDAPSSLSettings" keyStoreRef="LDAPKeyStore" trustStoreRef="LDAPTrustStore" />
```

```
<keyStore id="LDAPKeyStore" location="\${server.config.dir}/LdapSSLKeyStore.jks"
  type="JKS" password="{xor}CDo9Hgw=" />

<keyStore id="LDAPTrustStore" location="\${server.config.dir}/LdapSSLTrustStore.jks"
  type="JKS" password="{xor}CDo9Hgw=" />
```

#### 注释:

- 在 Analytic Server 中对 LDAP 的支持由 WebSphere Liberty 进行控制。有关更多信息，请参阅在 Liberty 中配置 LDAP 用户注册表。
- 使用 SSL 保护 LDAP 时，请遵循以下"配置从 Analytic Server 到 LDAP 的安全套接字层 (SSL) 连接"部分中的指示信息。

### 配置从 Analytic Server 到 LDAP 的安全套接字层 (SSL) 连接

如果在 Analytic Server 安装期间选择 Apache Directory Server (ads) LDAP 选项并且使用缺省配置，那么将在配置和启用 SSL 的情况下安装 Apache Directory Server (Analytic Server 将自动使用 SSL 以与 Apache Directory Server 通信)。

在 Analytic Server 安装期间选择其他一个 LDAP 选项时 (例如，在使用外部 LDAP 服务器时)，使用以下步骤配置 SSL。

1. 以 Analytic Server 用户身份登录每台 Analytic Server 机器，然后为 SSL 证书创建公共目录。

**注:** 缺省情况下，as\_user 是 Analytic Server 用户；请在 Ambari 控制台中查看"管理"选项卡下的服务帐户。

2. 将密钥库和信任库文件复制到所有 Analytic Server 机器上的某一公共目录。另外，将 LDAP 客户机 CA 证书添加至信任库。以下是一些指示信息示例。

```
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
```

**注:** JAVA\_HOME 是用于 Analytic Server 启动的同一 JRE。

3. 可以使用 {AS\_ROOT}/ae\_wlpserver/bin 的 securityUtility 工具对密码进行编码，从而对它们的值进行加密。示例如下所示。

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

4. 登录 Ambari 控制台并使用正确的 SSL 配置设置更新 Analytic Server 配置设置 **ssl.keystore.config**。示例如下所示。

```
<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="/>
```

**注:** 对密钥库文件和信任库文件使用绝对路径。

5. 使用正确的 LDAP 配置设置更新 Analytic Server 配置设置 **security.config**。例如，在 **ldapRegistry** 元素中，将 **sslEnabled** 属性设置为 true 并将 **sslRef** 属性设置为 defaultSSLConfig。

## 配置 Kerberos

Analytic Server 支持使用 Ambari 的 Kerberos。

注：IBM SPSS Analytic Server 与 Apache Knox 结合使用时，不支持 Kerberos 单点登录 (SSO)。

1. 在 Kerberos 用户存储库中，为您计划向其提供对 Analytic Server 的访问权的所有用户创建帐户。
2. 在 LDAP 服务器上创建相同帐户（从先前步骤）。
3. 在每个 Analytic Server 节点和 Hadoop 节点上，为前一步骤中创建的每个用户创建一个操作系统用户帐户。
  - 确保这些用户的 UID 在所有机器上都匹配。您可以使用 `kinit` 命令登录每个帐户来测试此内容。
  - 确保 UID 符合"用于提交作业的最小用户标识"Yarn 设置。这是 `container-executor.cfg` 中的 `min.user.id` 参数。例如，如果 `min.user.id` 为 1000，那么创建的每个用户帐户都必须具有大于或等于 1000 的 UID。
4. 在 Analytic Server 中，为所有主体在 HDFS 上创建用户主文件夹。例如，如果您将 `testuser1` 添加至 Analytic Server 系统，请在 HDFS 上创建类似于 `/user/testuser1` 的主文件夹，并确保 `testuser1` 具有对此文件夹的读写许可权。
5. [可选] 如果您计划使用 HCatalog 数据源并且 Analytic Server 与 Hive metastore 安装在不同的机器上，那么必须在 HDFS 上模拟 Hive 客户机。
  - a. 在 Ambari 控制台中，浏览至 HDFS 服务的"配置"选项卡。
  - b. 编辑 `hadoop.proxyuser.hive.groups` 参数以使其具有值 `*`，或者编辑包含所有能够登录 Analytic Server 的用户的组。
  - c. 编辑 `hadoop.proxyuser.hive.hosts` 参数以使其具有值 `*`，或者编辑 Hive metastore 和每个 Analytic Server 实例都作为服务安装在其上主机的列表。
  - d. 重新启动 HDFS 服务。

执行这些步骤并且安装了 Analytic Server 步骤后，Analytic Server 将以静默方式自动配置 Kerberos。

## 使用 Kerberos 针对单点登录 (SSO) 配置 HAProxy

1. 按照 HAProxy 文档指南配置并启动 HAProxy：<http://www.haproxy.org/#docs>
2. 针对 HAProxy 主机创建 Kerberos 原则 (`HTTP/<proxyHostname>@<realm>`) 和密钥表文件，其中，`<proxyHostname>` 是 HAProxy 主机的完整名称，而 `<realm>` 是 Kerberos 域。
3. 将密钥表文件作为 `/etc/security/keytabs/spnego_proxy.service.keytab` 复制到每个 Analytic Server 主机
4. 在每个 Analytic Server 主机上更新此文件的许可权。示例如下所示。

```
chown root:hadoop /etc/security/keytabs/spnego_proxy.service.keytab
chmod 440 /etc/security/keytabs/spnego_proxy.service.keytab
```
5. 打开 Amabri 控制台并更新 Analytic Server "Custom analytics.cfg" 部分中的以下属性。

```
web.authentication.kerberos.keytab=/etc/security/keytabs/spnego_proxy.service.keytab
web.authentication.kerberos.principal=HTTP/<proxy machine full name>@<realm>
```
6. 从 Amabri 控制台保存配置并重新启动所有 Analytic Server 服务。

现在，用户能够使用 Kerberos SSO 登录到 Analytic Server。

## 启用 Kerberos 模拟

模拟允许在安全上下文中执行线程，该安全上下文不同于拥有线程的进程的安全上下文。例如，模拟为 Hadoop 作业提供以除标准 Analytic Server 用户 (as\_user) 以外的其他用户身份运行的方式。要启用 Kerberos 模拟：

1. 在启用集群的 Kerberos 中运行时，将模拟配置属性添加到 HDFS (或 Hive 服务配置)。在 HDFS 的情况下，必须将以下属性添加到 HDFS core-site.xml 文件：

```
hadoop.proxyuser.<analytic_server_service_principal_name>.hosts = *
hadoop.proxyuser.<analytic_server_service_principal_name>.groups = *
```

其中，<analytic\_server\_service\_principal\_name> 是 Analytic Server 配置的 Analytic\_Server\_User 字段中指定的缺省 as\_user 值。

在通过 Hive/HCatalog 从 HDFS 访问数据的情况下，还必须将以下属性添加到 HDFS core-site.xml 文件：

```
hadoop.proxyuser.hive.hosts = *
hadoop.proxyuser.hive.groups = *
```

2. 如果将 Analytic Server 配置为使用 as\_user 以外的用户名，那么您必须修改属性名称以显示其他用户名 (例如，hadoop.proxyuser.xxxxx.hosts，其中 xxxxx 为 Analytic Server 配置中指定的配置的用户名)。
3. 在 Analytic Server 节点上从命令 shell 运行以下命令：

```
hadoop fs -chmod -R 755 /user/as_user/analytic-root/analytic-workspace
```

## 禁用 Kerberos

1. 在 Ambari 控制台中禁用 Kerberos。
2. 停止 Analytic Server 服务。
3. 从定制的 analytics.cfg 中除去下列参数。

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

4. 单击保存，然后重新启动 Analytic Server 服务。

## 启用与 Analytic Server 控制台的安全套接字层 (SSL) 连接

缺省情况下，Analytic Server 会生成自签名证书以启用安全套接字层 (SSL)，因此您可以接受自签名证书，通过安全端口来访问 Analytic Server 控制台。要使 HTTPS 访问更安全，您需要安装第三方供应商证书。

要安装第三方供应商证书，请执行下列步骤。

1. 将第三方供应商密钥库和信任库复制到所有 Analytic Server 节点中的同一目录；例如，/home/as\_user/security。

**注：**Analytic Server 用户必须具有对此目录的读访问权。

2. 在"Ambari 服务"选项卡中，浏览至 Analytic Server 服务的"配置"选项卡。
3. 编辑 **ssl.keystore.config** 参数。

```
<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>"/>

```

将

- <KEYSTORE-LOCATION> 替换为密钥库的绝对位置；例如， /home/as\_user/security/mykey.jks
- <TRUSTSTORE-LOCATION> 替换信任库的绝对位置；例如： /home/as\_user/security/mytrust.jks
- <TYPE> 替换为证书的类型；例如： JKS、PKCS12 等。
- <PASSWORD> 替换为使用 Base64 加密格式的加密密码。对于编码，您可以使用 securityUtility；例如： /opt/ibm/spss/analyticserver/3.1/ae\_wlpserver/bin/securityUtility encode <password>

如果想要生成自签名证书，您可以使用 securityUtility；例如： /opt/ibm/spss/analyticserver/3.1/ae\_wlpserver/bin/securityUtility createSSLCertificate --server=myserver --password=myspassword --validity=365 --subject=CN=mycompany,O=myOrg,C=myCountry。

注：您必须针对 CN 值提供相应的主机域名。

有关 securityUtility 和其他 SSL 设置的更多信息，请参阅 WebSphere Liberty 概要文件文档。

4. 单击保存，然后重新启动 Analytic Server 服务。

## 通过 SSL 与 Apache Hive 通信

您必须更新 hive.properties 文件以通过 SSL 连接与 Apache Hive 通信。或者，如果已针对高可用性启用 Apache Hive 环境，那么可选择主 Analytic Server 数据源页面上的高可用性参数。

### 更新 hive.properties 文件

1. 打开 hive.properties 文件。该文件位于： /opt/ibm/spss/AnalyticServer-3.1.2.0/ae\_wlpserver/usr/servers/aeserver/configuration/database
2. 找到以下行：

```
jdbcur1 = jdbc:hive2://{db.servername}:{db.serverport}/{db.databasesname};user={db.username};password={db.password}
```

3. 通过添加以下**粗体**信息更新行：

```
jdbcur1 = jdbc:hive2://{db.servername}:{db.serverport}/{db.databasesname};user={db.username};password={db.password};
ssl=true;sslTrustStore=pathtotheirtruststorefile;trustStorePassword=xxxtheirTrustStorePassword
```

4. 保存 hive.properties 文件。

## 启用对 Essentials for R 的支持

Analytic Server 支持对 R 模型进行评分，并支持运行 R 脚本。

为了配置对 R 的支持，成功安装 Analytic Server 后，请执行以下步骤：

1. 供应 Essentials for R 的服务器环境。

### RedHat Linux x86\_64

运行以下命令：

```

yum update
yum install -y zlib zlib-devel
yum install -y bzip2 bzip2-devel
yum install -y xz xz-devel
yum install -y pcre pcre-devel
yum install -y libcurl libcurl-devel

```

## Ubuntu Linux

运行以下命令：

```

apt-get update
apt-get install -y zlib1g-dev
apt-get install -y libreadline-dev
apt-get install -y libx32-dev
apt-get install -y bzip2
apt-get install -y libbz2-dev
apt-get install -y liblzma-dev
apt-get install -y libpcre3 libpcre3-dev
apt-get install -y libcurl4-openssl-dev
apt-get install -y liblzma-dev
apt-get install -y libpcre3 libpcre3-dev
apt-get install -y libcurl4-openssl-dev

```

## SUSE Linux

SUSE 上的 Essentials for R 安装需要兼容的 FORTRAN，它通常在配置的 Zypper 存储库中是不提供的（只能从 SUSE SDK 媒体获取）。因此，在 SUSE 服务器上运行 Essentials for R 的 Ambari 安装将会失败，因为它将无法安装 FORTRAN。在 SUSE 上请使用以下步骤来供应：

a. 安装 GCC C++。

```
zypper install gcc-c++
```

b. 安装 GCC FORTRAN。可以从 SUSE SDK 媒体复制所需的 RPM 文件，而且必须按下列顺序安装。

```

zypper install libquadmath0-4.7.2_20130108-0.19.3.x86_64.rpm
zypper install libgfortran3-4.7.2_20130108-0.19.3.x86_64.rpm
zypper install gcc43-fortran-4.3.4_20091019-0.37.30.x86_64.rpm
zypper install gcc-fortran-4.3-62.200.2.x86_64.rpm

```

c. 运行下列命令以安装 Essentials for R 库。

```

R_PREFIX=/opt/ibm/spss/R
cd $R_PREFIX
rm -fr $R_PREFIX/r_libs
mkdir -p $R_PREFIX/r_libs
cd $R_PREFIX/r_libs
wget https://zlib.net/fossils/zlib-1.2.11.tar.gz --no-check-certificate
tar xzvf zlib-1.2.11.tar.gz
cd zlib-1.2.11/
./configure
make && make install
cd $R_PREFIX/r_libs
wget http://www.bzip.org/1.0.6/bzip2-1.0.6.tar.gz
tar xzvf bzip2-1.0.6.tar.gz
cd bzip2-1.0.6
sed "s|^CC=gcc|CC=gcc -fPIC|" -i ./Makefile
make -f Makefile-libbz2_so
make clean
make
make install
cd $R_PREFIX/r_libs
wget https://tukaani.org/xz/xz-5.2.3.tar.gz
tar xzvf xz-5.2.3.tar.gz
cd xz-5.2.3
./configure
make -j3
make install
cd $R_PREFIX/r_libs
wget http://ftp.pcre.org/pub/pcre/pcre-8.38.tar.gz
tar xzvf pcre-8.38.tar.gz
cd pcre-8.38
./configure --enable-utf8
make
make install
cd $R_PREFIX/r_libs
wget https://www.openssl.org/source/openssl-1.0.2l.tar.gz --no-check-certificate
tar xzvf openssl-1.0.2l.tar.gz
cd openssl-1.0.2l/
./config shared
make
make install
echo '/usr/local/ssl/lib' >> /etc/ld.so.conf
ldconfig
cd $R_PREFIX/r_libs
wget --no-check-certificate https://curl.haxx.se/download/curl-7.50.1.tar.gz
tar xzvf curl-7.50.1.tar.gz
cd curl-7.50.1
./configure --with-ssl
make -j3

```



```
make install
cd $R_PREFIX/r_libs
wget ftp://rpmfind.net/linux/opensuse/distribution/12.3/repo/oss/suse/x86_64/libgomp1-4.7.2_20130108-2.1.6.x86_64.rpm --no-check-certificate
rpm -ivh libgomp1-4.7.2_20130108-2.1.6.x86_64.rpm
```

2. 下载用于 IBM SPSS Modeler Essentials for R RPM 或 DEB 的自抽取归档 (BIN)。Essentials for R 可供下载 (<https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=swg-tspssp>)。选择特定于堆栈、堆栈版本及硬件体系结构的文件。
3. 执行自解压缩二进制文件并遵循指示信息以（可选）查看许可证，接受许可证，然后选择联机或脱机安装。

### 联机安装

如果 Ambari 服务器主机以及集群中的所有节点都能够访问 <https://ibm-open-platform.ibm.com>，那么选择联机安装。

### 脱机安装

如果 Ambari 服务器主机无法访问因特网，请选择脱机安装。脱机安装将下载必需的 RPM 文件，并且应该在可以访问 <https://ibm-open-platform.ibm.com> 的机器上运行。然后，可以将 RPM 文件复制到 Ambari 服务器主机。

- a. 将必需的 Essentials for R RPM 或 DEB 文件复制到 Ambari 服务器主机上的任何位置。您需要的 RPM/DEB 文件取决于您的发行版、版本和体系结构，如下所示。

#### HDP 2.5 和 2.6 (x86\_64)

```
IBM-SPSS-ModelerEssentialsR-ambari-2.1-HDP-2.6-9.0.0.0-1.x86_64.rpm
```

#### HDP 2.6 (PPC64LE)

```
IBM-SPSS-ModelerEssentialsR-ambari-2.1-HDP-2.6-9.0.0.0-1.ppc64le.rpm
```

#### HDP 2.5 和 2.6 (Ubuntu)

```
IBM-SPSS-ModelerEssentialsR-ambari-3.1.2.0_3.1.2.0_amd64.deb
```

- b. 安装 RPM 或 DEB。在以下示例中，命令在 HDP 2.6 (x86\_64) 上安装 Essentials for R。  

```
rpm -i IBM-SPSS-ModelerEssentialsR-ambari-2.1-HDP-2.6-9.0.0.0-1.x86_64.rpm
```

在以下示例中，命令在 HDP 2.5 (Ubuntu) 上安装 Essentials for R。

```
dpkg -i IBM-SPSS-ModelerEssentialsR-ambari-3.1.2.0_3.1.2.0_amd64.deb
```

4. 重新启动 Ambari 服务器。

```
ambari-server restart
```

5. 登录 Ambari 服务器并通过 Ambari 控制台将 SPSS Essentials for R 作为服务进行安装。应该在安装了 Analytic Server 和 Analytic Metastore 的每个主机上安装 SPSS Essentials for R。

**注：**Ambari 将尝试在安装 R 之前安装 gcc-c++ 和 gcc-gfortran (RHEL) 以及 gcc-fortran (SUSE)。这些软件包声明为对 R 的 Ambari 服务定义的从属项。请确保要安装和执行 R 的服务器配置为下载 gcc-c++ 和 gcc-[g]fortran RPM，或者已安装 GCC 和 FORTRAN 编译器。如果 Essentials for R 安装失败，请在安装 Essentials for R 之前手动安装这些包。

6. 刷新 Analytic Server 服务。
7. 使用第 25 页的『更新客户机依赖关系』中的指示信息运行 `update_clientdeps` 脚本。
8. 还必须在托管 SPSS Modeler 服务器的机器上安装 Essentials for R。请参阅 SPSS Modeler 文档以获取详细信息。

## 启用关系数据库源

如果您在每个 Analytic Server 主机的共享目录中提供 JDBC 驱动程序，那么 Analytic Server 可以使用关系数据库源。缺省情况下，此目录为 `/usr/share/jdbc`。

要更改共享目录，请执行下列步骤。

1. 在"Ambari 服务"选项卡中，浏览至 Analytic Server 服务的"配置"选项卡。
2. 打开 **Advanced analytics.cfg** 部分。
3. 在 **jdbc.drivers.location** 中，指定 JDBC 驱动程序的共享目录的路径。
4. 单击**保存**。
5. 停止 Analytic Server 服务。
6. 单击**刷新**。
7. 启动 Analytic Server 服务。

表 6. 受支持的数据库

数据库	受支持的版本	JDBC 驱动程序 JAR	供应商
Amazon Redshift	8.0.2 或更高版本	RedshiftJDBC41-1.1.6.1006.jar 或更高版本	Amazon
BigSQL	4.1.0.0 或更高版本	db2jcc.jar	IBM
DashDB	Bluemix 服务	db2jcc.jar	IBM
Db2 for Linux、UNIX 和 Windows	11.1、10.5、10.1 和 9.7	db2jcc.jar	IBM
Db2 z/OS	11 和 10	db2jcc.jar 和 db2_license_cisuz.jar	IBM
Greenplum	5	postgresql.jar	Greenplum
Hive	1.2 和 2.1	hive-jdbc-*.jar	Apache
MySQL	5.6 和 5.7	mysql-connector-java-commercial-5.1.25-bin.jar	MySQL
Netezza	7 和 6.x	nzjdbc.jar	IBM
Oracle	12c 和 11g R2 (11.2)	ojdbc6.jar 和 orai18n.jar	Oracle
SQL Server	2014、2012 和 2008 R2	sqljdbc4.jar	Microsoft
Teradata	15 和 15.1	tdgssconfig.jar 和 terajdbc4.jar	Teradata

## 注释

- 如果在安装 Analytic Server 之前创建了 Redshift 数据源，那么需要执行以下步骤，才能使用该 Redshift 数据源。
  1. 在 Analytic Server 控制台中，打开该 Redshift 数据源。
  2. 选择 Redshift 数据库数据源。
  3. 输入 Redshift 服务器地址。
  4. 输入数据库名称和用户名。密码应自动填充。
  5. 选择数据库表。
- BigSQL 是 Apache Hadoop 环境的 IBM SQL 接口。BigSQL 不是关系数据库，但是 Analytic Server 支持通过 JDBC 来访问它（JDBC jar 文件与用于 Db2 的文件相同）。

BigSQL 与 Analytic Server 的常见用法是通过 HCatalog 数据源访问 BigSQL Hadoop/HBase 表。

## 启用 HCatalog 数据源

Analytic Server 通过 Hive/HCatalog 提供对许多数据源的支持。一些源需要手动配置步骤。

1. 收集必需的 JAR 文件以启用数据源。不需要任何额外步骤即可启用对 Apache HBase 及 Apache Accumulo 的支持。对于其他 NoSQL 数据源，请联系联系人供应商并获取存储处理程序及相关 JAR。有关受支持的 HCatalog 数据源的信息，请参阅 IBM SPSS Analytic Server 3.1.2 用户指南中的“使用 HCatalog 数据源”部分。
2. 将这些 JAR 文件添加至 {HIVE\_HOME}/auxlib 目录及每个 Analytic Server 节点的 /usr/share/hive 目录。
3. 重新启动 Hive Metastore 服务。
4. 刷新 Analytic Metastore 服务。
5. 重新启动 Analytic Server 服务的每个实例。

### 注释：

- Analytic Server Metastore 不能安装在与 Hive Metastore 相同的机器上。
- 在通过 Analytic Server HCatalog 数据源访问 HBase 数据时，访问用户必须具有 HBase 表的读许可权。
  - 在非 Kerberos 环境中，Analytic Server 使用 as\_user 访问 HBase (as\_user 必须具有 HBase 的读许可权)。
  - 在 Kerberos 环境中，as\_user 和登录用户必须具有 HBase 表的读许可权。

## NoSQL 数据库

Analytic Server 支持任何 NoSQL 数据库，可以从供应商处获取用于此数据库的 Hive 存储处理程序。

不需要任何额外步骤即可启用对 Apache HBase 及 Apache Accumulo 的支持。

对于其他 NoSQL 数据库，请联系联系人供应商并获取存储处理程序及相关 JAR。

## 基于文件的 Hive 表

Analytic Server 支持任何基于文件的 Hive 表，有一个内置或定制 Hive SerDe (序列化器-反序列化器) 可用于此表。

用于处理 XML 文件的 Hive XML SerDe 位于 Maven Central Repository 中，网址为 <http://search.maven.org/#search%7Cga%7C1%7Cchivexmlserde>。

## MapReduce v2 作业

使用 Analytic Server **Custom analytic.cfg** 部分中的 **preferred.mapreduce** 设置来控制 MapReduce 作业的处理方式：

表 7. 定制 *analytics.cfg* 属性

属性	描述
preferred.mapreduce	控制运行 MapReduce 作业的方法。有效值包括： <ul style="list-style-type: none"><li>• spark</li><li>• m3r</li><li>• hadoop</li></ul> 例如：preferred.mapreduce=spark

## Apache Spark

如果想要使用 Spark (V1.5 或更高版本)，必须在 Analytic Server 安装期间手动添加 spark.version 属性。

1. 打开 Amabri 控制台，并在 Analytic Server 的高级 **analytics.cfg** 部分中添加以下属性。
  - **Key:** spark.version
  - **Value:** 输入相应的 Spark 版本号 (例如, 1.x、2.x 或 None)。
2. 保存配置。

注：您可以通过定制的 analytics.cfg 设置强制 HCatalog 不要使用 Spark。

1. 打开 Amabri 控制台，并在 Analytic Server 的定制 **analytic.cfg** 部分中添加以下属性。
  - **Key:** spark.hive.compatible
  - **Value:** false

## 更改 Analytic Server 所使用的端口

缺省情况下，Analytic Server 对 HTTP 和 HTTPS 分别使用 9080 和 9443 端口。要更改端口设置，请执行下列步骤。

1. 在"Ambari 服务"选项卡中，浏览至 Analytic Server 服务的"配置"选项卡。
2. 打开 **Advanced analytics.cfg** 部分。
3. 在 **http.port** 和 **https.port** 中分别指定所需的 HTTP 和 HTTPS 端口。
4. 单击保存。
5. 重新启动 Analytic Server 服务。

## 高可用性 Analytic Server

通过将 Analytic Server 作为服务添加至集群中的多个节点，您可以实现它的高可用性。

1. 在 Ambari 控制台中，浏览至"主机"选项卡。
2. 选择尚未将 Analytic Server 作为服务运行的主机。
3. 在"摘要"选项卡上，单击添加并选择 Analytic Server。
4. 单击确认添加

## 多集群支持

多集群功能是 IBM SPSS Analytic Server 的高可用性功能的增强功能，它可在多租户环境中提供更强的隔离。缺省情况下，安装 Analytic Server 服务 (在 Ambari 或 ClouderaManager 中) 会导致系统定义一个分析服务器集群。

该集群规范定义 Analytic Server 集群成员信息。修改集群规范可通过 XML 内容来完成 (在 Ambari Analytic Server 配置的 analytics-cluster 字段中或通过手动编辑 Cloudera Manager 的 configuration/analytics-cluster.xml 文件)。当配置多个 Analytic Server 集群时，必须使用每个 Analytic Server 集群自身的负载均衡器将请求馈送到这些集群。

使用多集群功能可确保一个租户的工作不可能对在其他租户的集群上执行的工作产生负面影响。对于高度可用的作业，作业故障转移仅发生在启动工作的 Analytic Server 集群的作用域中。下面的示例提供了一个多集群 XML 规范。

注：可以通过将 Analytic Server 作为服务添加到集群中的多个节点来使它具有高可用性。

```

<analyticServerClusterSpec>
  <cardinality>1+</cardinality>
  <cluster name="cluster1">
    <memberName>one.cluster</memberName>
    <memberName>two.cluster</memberName>
  </cluster>
  <cluster name="cluster2">
    <memberName>three.cluster</memberName>
    <memberName>four.cluster</memberName>
  </cluster>
</analyticServerClusterSpec>

```

在上面的示例中，需要两个负载均衡器。一个负载均衡器将请求发送至 cluster1 成员（one.cluster 和 two.cluster），另一个负载均衡器将请求发送至 cluster2 成员（three.cluster 和 four.cluster）。

以下示例提供了一个单集群 XML 规范（缺省配置）。

```

<analyticServerClusterSpec>
  <cardinality>1</cardinality>
  <cluster name="cluster1">
    <memberName>*</memberName>
  </cluster>
</analyticServerClusterSpec>

```

在上面的示例中，需要一个负载均衡器来处理配置了多个集群成员的情况。

## 注释

- 只有单集群支持在 **memberName** 元素中使用通配符（例如，集群基数 ="1"）。cardinality 元素的有效值是 1 和 1+。
- 必须以与分配了 Analytic Server 角色的主机名相同的方式指定 **memberName**。
- 应用了集群配置更改后，必须重新启动所有集群中的所有服务器。
- 在 Cloudera Manager 中，必须修改并维护所有 Analytic Server 节点上的 analytics-cluster.xml 文件。对所有节点都必须进行维护，以确保它们包含相同的内容。

## 针对小型数据优化 JVM 选项

您可以编辑 JVM 属性，以便在运行小型 (M3R) 作业时优化系统。

在 Ambari 控制台中，查看 Analytic Server 服务中"配置"选项卡的 Advanced analytics-jvm-options 部分。修改下列参数会设置托管 Analytic Server 的服务器上运行的作业的堆大小；即，并非 Hadoop。运行较小的 (M3R) 作业时，这一点很重要，并且您可能需要对这些值进行试验以优化系统。

```

-Xms512M
-Xmx2048M

```

## 更新客户机依赖关系

本节描述如何使用 update\_clientdeps 脚本更新 Analytic Server 服务的依赖关系。

1. 以 root 用户身份登录到 Ambari 服务器主机。
2. 将目录更改为 /var/lib/ambari-server/resources/stacks/<stack-name>/<stack-version>/services/ANALYTICSERVER/package/scripts；请参阅以下示例。

```
cd "/var/lib/ambari-server/resources/stacks/HDP/2.6/services/ANALYTICSERVER/package/scripts"
```

3. 使用以下参数运行 update\_clientdeps 脚本。

```

-u <ambari-user>
    Ambari 帐户用户名
-p <ambari-password>
    Ambari 帐户用户的密码。

```

**-h <ambari-host>**

Ambari 服务器的主机名。

**-x <ambari-port>**

Ambari 侦听的端口。

请参阅以下示例。

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

4. 使用以下命令重新启动 Ambari 服务器。

```
ambari-server restart
```

## 配置 Apache Knox

Apache Knox Gateway 是针对 Apache Hadoop 服务提供单点安全访问的系统。系统简化用户（访问集群数据和运行作业）和操作员（控制访问权和管理集群）的 Hadoop 安全性。Gateway 作为服务器（或服务器集群）运行，为一个或多个 Hadoop 集群提供服务。

注：IBM SPSS Analytic Server 与 Kerberos 单点登录 (SSO) 结合使用时，不支持 Apache Knox。

Apache Knox Gateway 有效地隐藏 Hadoop 集群拓扑详细信息并与 Enterprise LDAP 和 Kerberos 相集成。以下部分提供有关必需的 Apache Knox 和 Analytic Server 配置任务的信息。

### 先决条件

- 已知的 Apache Knox 问题不传播在 HTTP cookie 和头中包含的安全信息（有关更多信息，请参阅 <https://issues.apache.org/jira/browse/KNOX-895>）。在 Knox 0.14.0（或更高版本）中解决了此问题。您必须获取更新的 Hortonworks 分发版（包含 Knox 0.14.0 或更高版本），然后再将 Knox 用于 Analytic Server。请联系 Hortonworks 提供商以获取更多信息。
- Analytic Server 节点必须以无密码的 SSH 连接方式连接到 Knox 服务器。无密码的 SSH 连接从 Analytic Server 移动到 Knox (**Analytic Server > Knox**)。
- 要安装 Analytic Server，必须首先安装 Knox 服务。

在某些情况下，会发生意外而导致无法自动复制配置文件。在这些情况下，您必须手动复制以下配置文件：

- com.ibm.spss.knox\_0.6-3.1.2.0.jar：必须将该文件从 Analytic Server 位置：

```
<Analytic_Server_Installation_Path>/ae_wlpserver/usr/servers/aeserver/apps/AE_BOOT.war/WEB-INF/lib
```

到 Knox 服务器节点：

```
/KnoxServicePath/ext
```

例如：/usr/iop/4.1.0.0/knox/ext

- rewrite.xml 和 service.xml：必须从以下 Analytic Server 位置复制这些文件：

```
<Analytic_Server_Installation_Path>/ae_wlpserver/usr/servers/aeserver/configuration/knox
```

到 Knox 服务器节点：

```
/KnoxServicePath/data/services
```

例如：/usr/iop/4.1.0.0/knox/data/services

注：存在两组 rewrite.xml 和 service.xml 文件（一组用于 http://rest 流量，一组用于 ws://websocket 流量）。将两个 analyticserver 和 analyticserver\_ws 的所有 rewrite.xml 和 service.xml 文件复制到 Knox 服务器节点。

## 配置 Ambari

必须在 Ambari 用户界面中配置 Analytic Server 服务：

1. 在 Ambari 用户界面中，浏览至 **Knox > 配置 > 高级拓扑**。当前 Knox 配置设置在内容窗口中显示。
2. 将以下两个服务添加到 Knox 配置中的高级拓扑部分：

```
<service>
  <role>ANALYTICSERVER</role>
  <url>http://{analyticserver-host}:{analyticserver-port}/analyticserver</url>
</service>
<service>
  <role>ANALYTICSERVER_WS</role>
  <url>ws://{analyticserver-host}:{analyticserver-port}/analyticserver</url>
</service>
```

{analyticserver-host} 和 {analyticserver-port} 必须替换为相应的 Analytic Server 服务器名称和端口号：

- 可以在 Ambari 用户界面 (**SPSS Analytic Server > 摘要 > Analytic Server**) 中找到 {analyticserver-host} URL。
- 可以在 Ambari 用户界面 (**SPSS Analytic Server > 配置 > Advanced analytics.cfg > http.port**) 中找到 {analyticserver-port} 号。

注：在将 Analytic Server 部署到多个节点并且使用 LoadBalancer时，{analyticserver-host} 和 {analyticserver-port} 必须对应于 LoadBalancer URL 和端口号。

3. 重新启动 Knox 服务。

在使用 LDAP 时，Knox 缺省为提供的“演示”LDAP。您可以更改为企业 LDAP 服务器（例如，Microsoft LDAP 或 OpenLDAP）。

## 配置 Analytic Server

要将 LDAP 用于 Analytic Server，必须将 Analytic Server 配置为使用 Apache Knox 所使用的相同 LDAP 服务器。必须更新以下 Ambari 设置的 <value> 条目以反映相应的 Knox LDAP 服务器设置：

- main.ldapRealm.userDnTemplate
- main.ldapRealm.contextFactory.url

Ambari 用户界面中提供的值位于：**Knox > 配置 > 高级拓扑**。例如：

```
<param>
  <name>main.ldapRealm.userDnTemplate</name>
  <value>uid={0},ou=people,dc=hadoop,dc=apache,dc=org</value>
</param>
<param>
  <name>main.ldapRealm.contextFactory.url</name>
  <value>ldap://{knox_host_name}:33389</value>
</param>
```

更新 Knox LDAP 设置后，重新启动 Knox 服务。

要点：Analytic Server 管理员密码必须与 Knox 管理员密码相同。

## 配置 Apache Knox

1. 刷新 Knox gateway.jks 文件：
  - a. 在 Knox 服务器上，停止 Knox 服务。
  - b. 从 /var/lib/knox/data-2.6.2.0-205/security/keystores 删除 gateway.jks。
  - c. 重新启动 Knox 服务。
2. 在 Knox 服务器上，创建子目录 <knox\_server>/data/service/analyticserver/3.1.2.0，然后将 service.xml 和 rewrite.xml 文件上载到此新目录。这两个文件位于 Analytic Server 的 <analytic\_server>/configuration/knox/analyticserver/3.1.2.0 上（例如，/opt/ibm/spss/analyticserver/3.1/ae\_wlpserver/usr/servers/aeserver/configuration/knox/analyticserver/3.1.2.0/\*.xml）
3. 在 <knox\_server>/bin 中，运行脚本 ./knoxcli.sh redeploy--cluster default
4. 将 com.ibm.spss.knoxservice\_0.6-\*.jar 文件上载到 <knox\_server>/ext。该文件位于 Analytic Server 的 <analytic\_server>/apps/AE\_BOOT.war/WEB-INF/lib/com.ibm.spss.knox\_0.6-3.1.2.0.jar 上（例如，/opt/ibm/spss/analyticserver/3.1/ae\_wlpserver/usr/servers/aeserver/apps/AE\_BOOT.war/WEB-INF/lib/com.ibm.spss.knox\_0.6-3.1.2.0.jar）。
5. 在 Ambari 用户界面中，浏览至 **Knox** > **配置** > **高级拓扑**，以添加下列元素。

```
<service>
  <role>ANALYTICSERVER</role>
  <url>http://{analyticserver-host}:{analyticserver-port}/analyticserver</url>
  <role>ANALYTICSERVER_WS</role>
  <url>ws://{analyticserver-host}:{analyticserver-port}/analyticserver</url>
</service>
```

注：缺省情况下，禁用 WebSocket 功能。可通过将 /conf/gateway-site.xml 文件中的 gateway.websocket.feature.enabled 属性更改为 true 来启用此功能。

6. 在 Ambari 用户界面中，浏览至 **Knox** > **配置** > **高级用户-ldif**（例如 admin、qauser1、qauser2），以添加或更新用户。
7. 浏览至 **Knox** > **服务操作** > **启动演示 LDAP** 以重新启动 LDAP。
8. 重新启动 Knox 服务。

## 在 Hortonworks 数据平台 (HDP) 上安装 Apache Knox

下列步骤概述了在 HDP 集群中安装 Apache Knox 的过程。

1. 验证 Knox 用户是否存在于 HDP 集群。 如果不存在，必须创建一个 Knox 用户。
2. 下载 Apache Knox 并将其解压缩到 /home/knox 目录下的文件夹中。
3. 在 HDP 中，切换到 Knox 用户，然后转至 knox 文件夹。 Knox 用户必须对所有 knox 子文件夹具有 permission(RWX) 许可权。
4. 为 Analytic Server 配置 Apache Knox。 有关更多信息，请参阅配置 **Apache Knox** 部分。
  - a. 在 {knox}/data/services 下创建 analyticserver/3.1.2.0 文件夹层次结构。
  - b. 从以下 Analytic Server 位置复制 rewrite.xml 和 service.xml 文件：

```
/opt/ibm/spss/analyticserver/3.1/ae_wlpserver/usr/servers/aeserver/configuration/knox/
analyticserver/3.1.2.0
```

到 Knox 服务器节点：

```
{knox}/data/services/analyticserver/3.1.2.0
```

- c. 从 Analytic Server 主机复制 Knox \*.jar 文件：



/opt/ibm/spss/analyticserver/3.1/ae\_wlpserver/usr/servers/aeserver/apps/AE\_BOOT.war/WEB-INF/lib/com.ibm.spss.knox\_0.6-\*.jar

到 Knox ext 目录:

{knox}/ext

- d. 在 {knox}/conf/topologies 目录下更新 default.xml 文件, 以符合以下示例:

注: 如果不存在, 必须创建此文件。

```
<topology>
  <gateway>
    <provider>
      <role>authentication</role>
      <name>ShiroProvider</name>
      <enabled>true</enabled>
      <param>
        <name>sessionTimeout</name>
        <value>30</value>
      </param>
      <param>
        <name>main.ldapRealm</name>
        <value>org.apache.hadoop.gateway.shiro.realm.KnoxLdapRealm</value>
      </param>
      <param>
        <name>main.ldapRealm.userDnTemplate</name>
        <value>uid={0},ou=people,dc=hadoop,dc=apache,dc=org</value>
      </param>
      <param>
        <name>main.ldapRealm.contextFactory.url</name>
        <value>ldap://localhost:33389</value>
      </param>
      <param>
        <name>main.ldapRealm.contextFactory.authenticationMechanism</name>
        <value>simple</value>
      </param>
      <param>
        <name>urls./**</name>
        <value>authcBasic</value>
      </param>
    </provider>
    <provider>
      <role>identity-assertion</role>
      <name>Default</name>
      <enabled>true</enabled>
    </provider>
    <provider>
      <role>authorization</role>
      <name>AclsAuthz</name>
      <enabled>true</enabled>
    </provider>
  </gateway>

  <!--other service-->
  <service>
    <role>ANALYTICSERVER</role>
    <!--replace the {analyticserver-host} and {analyticserver-port} with real value-->
    <url>http://{analyticserver-host}:{analyticserver-port}/analyticserver</url>
    <role>ANALYTICSERVER_WS</role>
    <url>ws://{analyticserver-host}:{analyticserver-port}/analyticserver</url>
  </service>
</topology>
```

注: 缺省情况下, 禁用 WebSocket 功能。可通过将 /conf/gateway-site.xml 文件中的 gateway.websocket.feature.enabled 属性更改为 true 来启用此功能。

5. 运行 {knox}/bin/knoxcli.sh。
6. 运行 {knox}/bin/ldap.sh start。

注: 此脚本使用端口 33389。请确保该端口当前不在使用中。

7. 运行 {knox}/bin/gateway.sh start。

注: 此脚本使用端口 8443。请确保该端口当前不在使用中。

8. 验证安装。

- a. 在 Knox URL 上针对 Analytic Server 运行 curl 命令：

```
curl -ikvu {username}:{password} https://{knox-host}:8443/gateway/default/analyticserver/admin
```

## 故障诊断

**问题：**安装后，Analytic Server 无法在 Knox 中运行。

**解决方案：**停止 Knox，除去 {knox}/data/deployments/\* 目录下的所有文件，然后重新启动 Knox。

**问题：**无法通过 Knox 登录 Analytic Server。

**解决方案：**验证 {knox}/conf/users.ldif 中的用户。更新现有用户，或添加新的 Analytic Server 用户。Knox 用户主体和凭证必须与 Analytic Server 用户匹配。

## 启用 Apache Knox 的 Analytic Server 的 URL 结构

启用 Knox 的 Analytic Server 用户界面 URL 为 `https://{knox-host}:{knox-port}/gateway/default/analyticserver/admin`

- https 协议 - 用户必须接受证书才能在 Web 浏览器中继续。
- knox-host 是 Knox 主机。
- knox-port 是 Knox 端口号。
- URI 为 gateway/default/analyticserver。

## 针对每个 IBM SPSS Analytic Server 租户配置单独的 YARN 请求 - HDP

通过使用 Spark 动态资源分配技术完成 Yarn 队列配置。

### Hortonworks Data Platform 2.x

1. 在 Ambari 用户界面中，浏览至 **SPSS Analytic Server 服务 > 配置 > 高级 analytics.cfg** 选项卡。
2. 将 **resource.pool.enabled** 值更改为 true。
3. 在定制 **analytics.cfg** 选项卡上添加以下属性：

```
config.folder.path=/etc/spark2/conf
resource.pool.mapping=tenant1:test,tenant2:production
resource.pool.default=default
spark.scheduler.mode=FAIR
spark.yarn.queue=default
```

表 8. 定制 analytics.cfg 属性

属性	描述
config.folder.path	该目录包含 fairscheduler.xml 文件，此文件包含 Spark 池属性信息。该文件是必需的并且必须手动创建。有关更多信息，请参阅 <b>fairscheduler.xml</b> 示例部分。
resource.pool.mapping	<p><b>Spark：</b>将租户映射到在 fairscheduler.xml 文件中定义的池。必须用逗号分隔租户对（例如，tenant1:test,tenant2:production）。在指定池之前，确保在 fairscheduler.xml 文件中配置池。</p> <p><b>MapReduce：</b>将租户映射到在 YARN 队列管理器中定义的队列。必须用逗号分隔租户对（例如，tenant1:test,tenant2:production）。指定队列之前，确保使用队列配置系统，并且允许访问以向队列提交作业。</p> <p>注：如果想要一起运行 Spark 和 MapReduce 作业，那么租户映射值必须是 fairscheduler.xml 文件和 YARN 队列管理器中的相同名称。</p>

表 8. 定制 *analytics.cfg* 属性 (续)

属性	描述
resource.pool.default	<p><b>Spark:</b> 定义缺省资源池。该值可以是 default 或在 fairscheduler.xml 文件中定义的池名称。如果未配置租户 (或者配置不正确), 那么使用 default 设置。</p> <p><b>MapReduce:</b> 定义向其提交作业的缺省队列。</p>
spark.scheduler.mode=FAIR	<p><b>Spark:</b> 启用公平调度程序。不应更改该属性。</p>
spark.yarn.queue	<p><b>Spark:</b> 将应用程序提交至的 YARN 队列的名称。您可以在 YARN 队列管理器中指定定制的 YARN 队列名称。</p>

4. 保存配置并重新启动 Analytic Server 服务。

### fairscheduler.xml 示例

fairscheduler.xml 文件包含 Spark 池属性信息。该文件是必需的并且必须手动创建。

```
<?xml version="1.0"?>
<allocations>
  <pool name="production">
    <schedulingMode>FAIR</schedulingMode>
    <weight>1</weight>
    <minShare>2</minShare>
  </pool>
  <pool name="test">
    <schedulingMode>FIFO</schedulingMode>
    <weight>2</weight>
    <minShare>3</minShare>
  </pool>
</allocations>
```

### 参考

请参阅以下站点以获取更多信息：

- <https://spark.apache.org/docs/latest/job-scheduling.html#dynamic-resource-allocation>
- <https://spark.apache.org/docs/latest/running-on-yarn.html>

## 在 Ambari 上迁移 IBM SPSS Analytic Server

Analytic Server 可以将数据和配置设置从现有 Analytic Server 迁移到新安装。迁移可在相同集群环境或新集群环境中发生。

### 从 Analytic Server 3.1.1.1 迁移至相同服务器集群上的 3.1.2

如果您有 Analytic Server 3.1.1.1 的现有安装, 那么可将 3.1.1.1 配置设置迁移至相同服务器集群上的 3.1.2 安装。

- 从旧的 Analytic Server 版本 (Analytic Server 3.1.1.1) 中收集配置设置。
  - 展开 {AS\_ROOT}\tools\unzip configcollector.zip 归档 (这将创建新的名为 configcollector 的文件夹)。
  - 运行 configcollector 文件夹中的 configcollector.sh 脚本。将生成的压缩 (ZIP) ASConfiguration\_3.1.1.1.xxx.zip 文件复制到不同的文件夹位置 (作为备份)。
- 将分析根目录从旧的 Analytic Server 3.1.1.1 版本安装被分到新位置。
  - 如果不确定分析根目录的位置, 请运行 **hadoop fs -ls** 命令。分析根目录的路径类似于 /user/as\_user/analytic-root/analytic-workspace, 其中, as\_user 是拥有分析根目录的用户标识。

- b. 使用 `hadoop fs -copyToLocal` 和 `hadoop fs -copyFromLocal` 命令以将旧的 Analytic Server 版本 `analytic-workspace` 文件夹复制到新位置（例如，`/user/as_user/analytic-root/AS31Location`）。
3. 如果使用嵌入式 Apache Directory Server，那么使用第三方 LDAP 客户机工具备份当前用户/组配置。安装 Analytic Server 3.1.2 后，将备份用户/组配置导入到 Apache Directory Server。

注：如果使用外部 LDAP 服务器，那么可跳过此步骤。

4. 打开 Ambari 控制台并停止 **Analytic Server** 服务。
5. 卸载旧的 Analytic Server 版本 (Analytic Server 3.1.1.1)，然后安装 Analytic Server 3.1.2。有关安装指示信息，请参阅第 3 页的第 2 章，『Ambari 安装和配置』。
6. 打开 Ambari 控制台并停止 **Analytic Server** 服务（在 Ambari 中，确保 **Analytic Metastore** 服务正在运行）。
7. 将来自步骤 2 的备份 Analytic Server 3.1.1.1 分析根目录复制到新的 Analytic Server 版本位置。
  - a. 从新安装的 Analytic Server 版本除去 `analytic-workspace`。
  - b. 将备份 Analytic Server 3.1.1.1 分析工作空间文件夹 (`/user/as_user/analytic-root/AS31Location`) 复制到新版本位置（例如，`/user/as_user/analytic-root/analytic-workspace`）。必须确保将分析工作空间所有者定义为 `as_user`。
8. 清除 Zookeeper 状态。在 Zookeeper bin 目录（例如，Hortonworks 上的 `/usr/hdp/current/zookeeper-client`）中，运行以下命令：

```
./zkCli.sh rmr /AnalyticServer
```
9. 将来自步骤 1 的备份归档 `ASConfiguration_3.1.1.1.xxx.zip` 复制到新的 Analytic Server 版本位置（例如，`/opt/ibm/spss/analyticserver/3.1/`）。
10. 通过运行 `migrationtool.sh` 脚本并将 `ASConfiguration_3.1.1.1.xxx.zip` 归档文件（配置收集器创建）的路径作为自变量进行传递，运行迁移工具。例如：

```
migrationtool.sh /opt/ibm/spss/analyticserver/3.1/ASConfiguration_3.1.1.1.xxx.zip
```
11. 更新每个 Analytic Server 集群节点上的 `ae_wlpserver/usr/servers/aeserver/configuration/config.properties` 文件。

- 在该文件中添加一个 `as_user` 的条目。例如：

```
hdfs.user=as_user/host@REALM
```

`host` 必须与 `config.properties` 文件驻留的 Analytic Server 节点主机名匹配。

每个节点都有不同的 `hdfs.user` 值；每个 `host` 值必须与其驻留的 Analytic Server 主机匹配。

12. 在 Analytic Server 节点上从命令 shell 运行以下命令：

```
hadoop fs -chmod -R 755 /user/as_user/analytic-root/analytic-workspace
```
13. 在 Ambari 控制台中，启动 **Analytic Server** 服务。

## 从 Analytic Server 3.1.1.1 迁移至新服务器集群上的 3.1.2

如果您有 Analytic Server 3.1.1.1 的现有安装，那么可将 3.1.1.1 配置设置迁移至新服务器集群上的 3.1.2 安装。

1. 根据第 6 页的『在 Ambari 上安装』中的指示信息，安装新的 Analytic Server 版本。
2. 将分析工作空间从旧安装复制到新安装。
  - a. 如果您不确定分析工作空间的位置，请运行 `hadoop fs -ls`。分析工作空间的路径类似于 `/user/as_user/analytic-root/analytic-workspace`，其中，`as_user` 是拥有分析工作空间的标识。
  - b. 从新服务器上除去 `analytic-workspace`。

- c. 使用 `hadoop fs -copyToLocal` 和 `hadoop fs -copyFromLocal` 以将旧服务器的分析工作空间复制到新服务器的 `/user/as_user/analytic-root/analytic-workspace` 文件夹（确保将所有者设置为 `as_user`）。
3. 如果使用嵌入式 Apache Directory Server，那么使用第三方 LDAP 客户机工具备份当前用户/组配置。安装 Analytic Server 3.1.2 后，将备份用户/组配置导入到 Apache Directory Server。

注：如果使用外部 LDAP 服务器，那么可跳过此步骤。

4. 在新服务器上，打开 Ambari 控制台，然后停止 Analytic Server 服务（在 Ambari 上，确保 Analytic Metastore 服务正在运行）。
5. 从旧安装中收集配置设置。
  - a. 将新安装版本中的 `configcollector.zip` 归档复制到旧安装版本中的 `{AS_ROOT}\tools`。
  - b. 解压 `configcollector.zip` 的副本，它会在旧安装中创建新的 `configcollector` 子目录。
  - c. 通过运行 `{AS_ROOT}\tools\configcollector` 中的 **configcollector** 脚本，在旧安装版本中运行配置收集器工具。将生成的压缩 (ZIP) 文件复制到托管新安装的服务器。

**要点：**所提供的 **configcollector** 脚本可能与最新的 Analytic Server 版本不兼容。如果遇到有关 **configcollector** 脚本的问题，请联系 IBM 技术支持代表。

6. 清除 Zookeeper 状态。在 Zookeeper bin 目录（例如，Hortonworks 上的 `/usr/hdp/current/zookeeper-client`）中，运行以下命令。

```
./zkCli.sh rmr /AnalyticServer
```

7. 运行 **migrationtool** 脚本并将配置收集器所创建的压缩文件的路径作为参数进行传递。示例如下所示。

```
migrationtool.sh /opt/ibm/spss/analyticserver/3.1/ASConfiguration_3.1.1.1.xxx.zip
```

8. 更新每个 Analytic Server 节点上的 `ae_wlpserver/usr/servers/aeserver/configuration/config.properties` 文件。在该文件中添加一个 `as_user` 的条目。例如：

```
hdfs.user=as_user/host@REALM
```

`host` 必须与 `config.properties` 文件驻留的 Analytic Server 节点主机名匹配。每个节点都有不同的 `hdfs.user` 值；每个 `host` 值必须与其驻留的 Analytic Server 主机匹配。

9. 在 Analytic Server 节点上从命令 shell 运行以下命令：

```
hadoop fs -chmod -R 755 /user/as_user/analytic-root/analytic-workspace
```

10. 在 Ambari 控制台中，启动 Analytic Server 服务。

注：如果已将 R 配置为与现有 Analytic Server 安装配合使用，那么请遵循相应步骤，将其配置为用于新的 Analytic Server 安装。

---

## 卸载

**要点：**在安装 Essentials for R 时，首先必须运行 `remove_R.sh` 脚本。在 Analytic Server 之前未能卸载 Essentials for R 将导致以后无法卸载 Essentials for R。在卸载 Analytic Server 时，将除去 `remove_R.sh` 脚本。有关卸载 Essentials for R 的信息，请参阅第 34 页的『卸载 Essentials for R』。

1. 在 Analytic Metastore 主机上，使用以下参数运行 `{AS_ROOT}/bin` 目录中的 `remove_as.sh` 脚本。

**u** 必需。Ambari 服务器管理员的用户标识。

**p** 必需。Ambari 服务器管理员的密码。

**h** 必需。Ambari 服务器主机名。

**x** 必需。 Ambari 服务器端口。

**l** 可选。 启用安全方式。

示例如下所示：

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

从具有 Ambari 主机 one.cluster 的集群中除去 Analytic Server。

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

以安全方式从具有 Ambari 主机 one.cluster 的集群中除去 Analytic Server。

注：此操作除去 HDFS 上的 Analytic Server 文件夹。

注：该操作不会移除与 Analytic Server 关联的任何 Db2 模式。请参阅 Db2 文档以了解有关手动移除模式的信息

## 卸载 Essentials for R

1. 在 Essentials for R 主机上，在 {AS\_ROOT}/bin 目录中运行具有以下参数的 remove\_R.sh 脚本。

**u** 必需。 Ambari 服务器管理员的用户标识。

**p** 必需。 Ambari 服务器管理员的密码。

**h** 必需。 Ambari 服务器主机名。

**x** 必需。 Ambari 服务器端口。

**l** 可选。 启用安全方式。

示例如下所示：

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

从包含 Ambari 主机 one.cluster 的集群中移除 Essentials for R。

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

从包含 Ambari 主机 one.cluster 的集群中以安全方式移除 Essentials for R。

2. 从 Ambari 服务器服务目录中移除 R 服务目录。例如，在 HDP 2.6 中，ESSENTIALR 目录位于 /var/lib/ambari-server/resources/stacks/HDP/2.6/services 中。
3. 在 Ambari 控制台中，验证 Essentials for R 服务是否不再存在。

---

## 第 3 章 Cloudera 安装和配置

---

### Cloudera 概述

Cloudera 是一个开放式源代码 Apache Hadoop 分发版。包含 Apache Hadoop (CDH) 的 Cloudera 分发版针对使用该技术的企业级部署。

Analytic Server 可以在 CDH 平台上运行。CDH 包含 Hadoop 的主要核心元素以及其他面向企业的组件，这些元素提供对大数据集（主要是 MapReduce 和 HDFS）的可靠可伸缩的分布式数据处理，而这些组件提供安全、高可用性以及与硬件和其他软件的集成。

---

### 特定于 Cloudera 的先决条件

除了一般的先决条件，还需查看以下信息。

**服务** 请确保将以下实例安装在每个 Analytic Server 主机上。

- HDFS: Gateway、DataNode 或 NameNode
- Hive: Gateway、Hive Metastore Server 或 HiveServer2
- Yarn: Gateway、ResourceManager 或 NodeManager

以下实例仅在使用其功能时是必需的。

- Accumulo: Gateway
- HBase: Gateway、Master 或 RegionServer
- Spark: Gateway
- Spark 2: Gateway

#### 元数据存储库

您可以使用 Db2 和 MySQL 作为 Analytic Server 元数据存储库。如果计划将 MySQL 用作 Analytic Server 元数据存储库，请根据第 37 页的『为 Analytic Server 配置 MySQL』的指示信息进行操作。

### 启用 Kerberos 的 Cloudera 环境

如果计划在启用 Kerberos 的 Cloudera 环境中安装 Analytic Server，那么必须验证以兼容 Analytic Server 的方式正确配置了 Kerberos。

以下部分适用于已安装 Kerberos 的 Cloudera 环境。在 Cloudera 中安装 Analytic Server 前必须遵循以下部分。假定您具有基本 Kerberos 认证知识，因为这些部分包含特定于 Kerberos 的术语（例如，**kinit** 和 **kadmin** 等）。

**注：**Analytic Server 检查 HDFS 配置以获取 Kerberos 相关值从而用于认证。

### Kerberos 认证

验证在安装 Analytic Server 之前在每个 Cloudera 集群节点上配置了 Kerberos 认证。有关更多信息，请参阅 Cloudera 产品文档中的在 Cloudera Manager 中配置认证。

注：在每个 Cloudera 集群节点上配置 Kerberos 认证后，必须先重新启动 **cloudera-scm-server** 和 **cloudera-scm-agent** 服务，然后才能安装 Analytic Server。必须重新启动所有集群节点上的 **cloudera-scm-agent** 服务。

## 在 Kerberos 中创建必需的帐户

1. 在 Kerberos 用户存储库中，为您计划向其提供对 Analytic Server 的访问权的所有用户创建帐户。
2. 在 LDAP 服务器上创建相同帐户（从先前步骤）。
3. 在每个 Analytic Server 节点和 Hadoop 节点上，为前一步骤中创建的每个用户创建一个操作系统用户帐户。
  - 确保这些用户的 UID 在所有机器上都匹配。您可以使用 `kinit` 命令登录每个帐户来测试此内容。
  - 确保 UID 符合用于提交作业的最小用户标识 Yarn 设置。这是 `container-executor.cfg` 中的 `min.user.id` 设置。例如，如果 `min.user.id` 为 1000，那么创建的每个用户帐户都必须具有大于或等于 1000 的 UID。
4. 在 HDFS 上为 Analytic Server 管理员用户创建用户主文件夹。文件夹许可权必须设置为 777，所有者必须定义为 `admin`，并且用户组必须设置为 `hdfs`。请参阅以下粗体示例：

```
[root@xxxx configuration]# hadoop fs -ls /user
Found 9 items
drwxrwxrwx - hdfs supergroup 0 2017-07-26 03:41 /user/AE
drwxrwxrwx - admin hdfs 0 2017-06-08 01:33 /user/admin
drwxr-x--x - as_user hdfs 0 2017-06-06 01:00 /user/as_user
drwx----- - hdfs supergroup 0 2017-07-31 00:17 /user/hdfs
drwxrwxrwx - mapred hadoop 0 2017-06-05 00:28 /user/history
drwxrwxr-t - hive hive 0 2017-06-05 00:30 /user/hive
drwxrwxr-x - hue hue 0 2017-06-05 00:30 /user/hue
drwxrwxr-x - impala impala 0 2017-07-19 00:52 /user/impala
drwxr-x--x - spark spark 0 2017-06-05 01:34 /user/spark
```

5. 如果您计划使用 HCatalog 数据源，并且 Analytic Server 与 Hive metastore 安装在不同的机器上，那么需要在 HDFS 上模拟 Hive 客户机。
  - a. 在 Cloudera Manager 中，浏览至 HDFS 服务的"配置"选项卡。

注：如果尚未设置以下设置，那么可能不会显示在配置选项卡上。在此情况下，请运行搜索来查找这些参数。
  - b. 编辑 `hadoop.proxyuser.hive.groups` 设置以使其具有值 `*`，或者编辑包含所有能够登录 Analytic Server 的用户的组。
  - c. 编辑 `hadoop.proxyuser.hive.hosts` 设置以使其具有值 `*`，或者编辑 Hive metastore 和每个 Analytic Server 实例都作为服务安装在其上主机的列表。
  - d. 重新启动 HDFS 服务。

执行这些步骤并且安装了 Analytic Server 步骤后，Analytic Server 将以静默方式自动配置 Kerberos。

## 启用 Kerberos 模拟

模拟允许在安全上下文中执行线程，该安全上下文不同于拥有线程的进程的安全上下文。例如，模拟为 Hadoop 作业提供以除标准 Analytic Server 用户 (`as_user`) 以外的其他用户身份运行的方式。要启用 Kerberos 模拟：

1. 打开 Cloudera Manager 并在 `core-site.xml` 的集群范围高级配置片段（安全值）（位于 **HDFS**（服务范围） > 配置选项卡上）区域中添加或更新以下属性。
  - 名称：`hadoop.proxyuser.as_user.hosts`
  - 值：`*`
  - 名称：`hadoop.proxyuser.as_user.groups`



- 值: \*

注: **core-site.xml** 设置应用于 Hadoop 配置 (而不是 Analytic Server)。

2. 在 Analytic Server 节点上从命令 shell 运行以下命令:

```
hadoop fs -chmod -R 755 /user/as_user/analytic-root/analytic-workspace
```

## 为 Analytic Server 配置 MySQL

在 Cloudera Manager 中配置 IBM SPSS Analytic Server 要求安装和配置 MySQL 服务器数据库。

1. 从存储 MySQL 数据库的节点上的命令窗口运行以下命令:

```
yum install mysql-server
```

注: 对于 SuSE Linux 使用 `zypper install mysql`。

2. 在每个 Cloudera 集群节点上的命令窗口运行以下命令:

```
yum install mysql-connector-java
```

注: 对于 SuSE Linux, 使用 `sudo zypper install mysql-connector-java`。

3. 决定并记录 Analytic Server 在访问 MySQL 数据库时使用的 Analytic Server 数据库名称、数据库用户名和数据库密码。
4. 根据第 39 页的『在 Cloudera 上安装』中的指示信息安装 Analytic Server。
5. 将 `/opt/cloudera/parcels/AnalyticServer/bin/add_mysql_user.sh` 脚本从 Cloudera 管理的其中一个服务器复制到安装了 MySQL 数据库的节点。使用适合您特定配置参数运行该脚本。例如:

```
./add_mysql_user.sh -u <database_user_name> -p <database_password> -d  
<database_name>
```

注释: 当数据库以安全方式 (即设置了 root 用户密码) 运行时需要 `a -r <dbRootPassword>` 参数。

当数据库使用非 root 用户名以安全方式运行时需要 `-r <dbUserPassword>` 和 `-t <dbUserName>` 参数。

---

## 安装预检查和后检查工具 - Cloudera

### 工具位置和先决条件

在安装 Analytic Server 服务之前, 在将属于 Analytic Server 服务的所有节点上运行预检查工具以验证 Linux 环境已准备好安装 Analytic Server。

在安装过程中自动调用预检查工具。工具在每个节点上运行安装之前检查每个 Analytic Server 节点。您还可以手动在每个节点上调用预检查工具, 此工具可在安装服务之前验证机器。

运行自解压 Analytic Server 二进制文件后, 预检查工具位于以下目录中:

- **Cloudera**

```
/opt/cloudera/parcels/AnalyticServer-3.1.2.0/tools/com.spss.ibm.checker.zip  
  
[root@servername ~]# cd /opt/cloudera/parcels/AnalyticServer-3.1.2.0/tools/  
[root@servername tools]# ls  
com.spss.ibm.checker.zip configcollector.zip regex-files
```

注: 在运行可执行二进制文件, 并随后在 Cloudera Manager 的“包”页面中分发 (下载 > 分发) 和激活 Analytic Server 之前, tools 目录中不存在预检查工具。

安装 Analytic Server 后, 后检查工具位于以下目录中:

- **Cloudera**

/opt/cloudera/parcels/AnalyticServer-3.1.2.0/tools/com.spss.ibm.checker.zip

该工具必须作为根目录运行，并需要 Python 2.6.X（或更高版本）。

如果预检查工具报告了任何故障，那么这些故障必须在您继续 Analytic Server 安装之前解决。

## 运行预检查工具

### 自动

在通过 Cloudera Manager 控制台作为服务的一部分安装 Analytic Server 时，可作为 Analytic Server 安装的一部分自动调用预检查工具。您必须手动输入 Cloudera Manager 服务器用户名和密码：

## Add SPSS Analytic Server Service to Cluster 1

### Review Changes

The screenshot shows the 'Review Changes' step in Cloudera Manager. It is divided into two main sections:

- Cloudera Manager Administrator account username:** The label is 'Cloudera Manager Administrator account username' with the variable name 'cm.admin.username' below it. The input field contains the text 'admin'. Below the field is a red error message: 'Missing required value: Cloudera Manager Administrator account username'.
- Cloudera Manager Administrator account password:** The label is 'Cloudera Manager Administrator account password' with the variable name 'cm.admin.password' below it. The input field contains six dots '••••••'. Below the field is a red error message: 'Missing required value: Cloudera Manager Administrator account password'.

图 3. Cloudera Manager 管理员设置

### 手动

您可以手动在每个集群节点上调用预检查工具。

以下预检查示例检查在 myclouderahost.ibm.com:7180 上运行的 Cloudera 集群 MyCluster，并使用登录凭证 admin:admin：

```
python ./precheck.py --target C --cluster MyCluster --username admin  
--password admin --host myclouderahost.ibm.com --port 7180 --as_host myashost.ibm.com
```

### 注释：

- as\_host 值必须由任一 IP 地址或标准域名提供。
- 工具将在省略密码自变量时提示密码。
- precheck.py 命令包括使用情况帮助，它与 --h 自变量一起显示 (python ./precheck.py --help)。
- --cluster 自变量是可选的（未使用 --cluster 时标识当前集群）。

当预检查工具运行其检查时，每个检查的状态将显示在命令窗口中。发生故障时，在日志文件中提供详细信息（命令窗口中提供了确切的日志文件位置）。当需要更多支持时，日志文件可以提供给 IBM 技术支持。

## 运行后检查工具

后检查工具验证 Analytic Server 是否正常运行，并能够处理简单的作业。以下后检查示例检查了在 myanalyticserverhost.ibm.com:9443 上运行的启用了 SSL 的 Analytic Server 实例，并使用了登录凭证 admin:ibmspss:

```
python ./postcheck.py --host myanalyticserverhost.ibm.com --port 9443
--username admin --password ibmspss --ssl
```

当 Knox 与 Analytic Server 一起使用时，命令如下所示：

```
python ./postcheck.py --host myknoxserverhost.ibm.com --port 8443
--username admin --password ibmspss --ssl --gateway_url /gateway/default
```

要执行单个检查，请使用以下命令：

```
python ./postcheck.py --host myknoxserverhost.ibm.com --port 8443
--username admin --password ibmspss --ssl --gateway_url /gateway/default --check AS_PYSPARK_BUILDMODEL
```

注释：

- 工具将在省略密码自变量时提示密码。
- postcheck.py 命令包括用法帮助，它与 --h 自变量一起显示 (python ./postcheck.py --help)。

当后检查工具运行其检查时，每个检查的状态将显示在命令窗口中。发生故障时，在日志文件中提供详细信息（命令窗口中提供了确切的日志文件位置）。当需要更多支持时，日志文件可以提供给 IBM 技术支持。

---

## 在 Cloudera 上安装

以下步骤说明了在 Cloudera Manager 中手动安装 IBM SPSS Analytic Server 的过程。

### Analytic Server 3.1.2

#### 联机安装

1. 浏览到 [IBM Passport Advantage® Web 站点](#)，并将特定于您的堆栈、堆栈版本和硬件体系结构的自抽取二进制文件下载到 Cloudera 集群中的主机。可用的 Cloudera 二进制文件为：

表 9. Analytic Server 自抽取二进制文件

描述	二进制文件名
IBM SPSS Analytic Server 3.1.2 for Cloudera 5.11、5.12、5.13、5.14 和 5.15 Ubuntu 英文版	spss_as-3.1.2.0-cdh5.11-5.15-ubun.bin
IBM SPSS Analytic Server 3.1.2 for Cloudera 5.11、5.12、5.13、5.14 和 5.15 Linux x86-64 英文版	spss_as-3.1.2.0-cdh5.11-5.15-lx86.bin

2. 在 Cloudera Manager 主集群节点上，运行 Cloudera 自抽取 \*.bin 安装程序。接受许可协议并保持缺省 CSD 安装目录不变，以按照安装提示进行操作。

注：如果从缺省位置变更了 CSD 目录，必须指定其他 CSD 目录。

3. 安装完成后使用以下命令以重新启动 Cloudera Manager：

```
service cloudera-scm-server restart
```

4. 打开 Cloudera Manager 界面（例如，使用缺省登录凭证 admin/admin 登录的 [http://\\${CM\\_HOST}:7180/cmfd/login](http://${CM_HOST}:7180/cmfd/login)），刷新远程包存储库 URL（位于主机 > 包 > 单击配置中），并验证 URL 是否正确。例如：

```
https://ibm-open-platform.ibm.com
```

注：可以更新包更新频率和远程包存储库 URL 以满足您的特定需要。

5. 在 Cloudera Manager 刷新包文件（可以通过单击**检查新包**来手动刷新包文件）之后，您会看到 **AnalyticServer** 包状态设置为**远程可用**。
6. 选择**下载 > 分发 > 激活**。**AnalyticServer** 包状态更新为**已分发和已激活**。
7. 在 Cloudera Manager 中，将 Analytic Server 添加为服务，并决定要放置 Analytic Server 的位置。您需要在"添加服务向导"中提供以下信息：

注："添加服务向导"显示服务创建过程的每个阶段中的总体进度，并在服务成功安装和配置到集群上时提供最终确认消息。

- Analytic Server metastore 主机名
- Analytic Server metastore 数据库名称
- Analytic Server metastore 用户名
- Analytic Server metastore 密码

#### MySQL 作为 Analytic Server 元数据存储库

- Analytic Server metastore 驱动程序类：com.mysql.jdbc.Driver
- Analytic Server metastore 存储库 URL：jdbc:mysql://{MySQL\_DB}/{DBName}?createDatabaseIfNotExist=true  
  
{MySQL\_DB} 是安装了 MySQL 的服务器的主机名

#### Db2 作为 Analytic Server 元数据存储库

- Analytic Server metastore 驱动程序类：com.ibm.db2.jcc.DB2Driver
- Analytic Server metastore 存储库 URL：jdbc:db2://{Db2\_HOST}:{PORT}/{DBName}:currentSchema={SchemaName};  
  
{Db2\_HOST} 是安装了 Db2 的服务器的主机名。  
  
{PORT} 是 Db2 正在侦听的端口。  
  
{SchemaName} 是未使用的可用模式。

如果不确定要输入的值，请咨询 Db2 管理员。

#### LDAP 配置

Analytic Server 使用 LDAP 服务器来存储和认证用户和组。在 Analytic Server 安装期间提供必需的 LDAP 配置信息。

表 10. LDAP 配置设置

LDAP 设置	描述
as.ldap.type	LDAP 类型。值可以是 ads、ad 或 openldap。 <ul style="list-style-type: none"> <li>• ads - Apache Directory Server (缺省设置)</li> <li>• ad - Microsoft Active Directory</li> <li>• openldap - OpenLDAP</li> </ul>
as.ldap.host	LDAP 主机
as.ldap.port	LDAP 端口号
as.ldap.binddn	LDAP 绑定 DN
as.ldap.bindpassword	LDAP 绑定 DN 密码
as.ldap.basedn	LDAP 基本 DN

表 10. LDAP 配置设置 (续)

LDAP 设置	描述
as.ldap.filter	LDAP 用户和组过滤器规则 注：如果该值包含竖线   字符，那么必须用反斜杠字符来转义该字符 (例如, \ )。
as.ldap.ssl.enabled	指定是否使用 SSL 在 Analytic Server 和 LDAP 之间进行通信。该值可以是 true 或 false。
as.ldap.ssl.reference	LDAP SSL 引用标识
as.ldap.ssl.content	LDAP SSL 配置

- 缺省情况下，as.ldap.type 设置为 ads，其他相关设置包含缺省设置。例外是，必须针对 as.ldap.bindpassword 设置提供密码。Analytic Server 使用配置设置以安装 Apache Directory Server (ADS) 和运行服务器初始化。缺省 ADS 概要文件包含用户 admin 以及 admin 的密码。您可以通过 Analytic Server 控制台执行用户管理，或者通过 <Analytic Root>/bin 文件夹中的 importUser.sh 脚本，从 XML 文件导入用户和组信息。
- 如果计划使用外部 LDAP 服务器，例如，Microsoft Active Directory 或 OpenLDAP，那么必须根据实际 LDAP 值定义配置设置。有关更多信息，请参阅在 Liberty 中配置 LDAP 用户注册表。
- 安装 Analytic Server 后，您可以更改 LDAP 配置（例如，从 Apache Directory Server 更改为 OpenLDAP）。但是，如果初始从 Microsoft Active Directory 或 OpenLDAP 开始，并且决定以后切换到 Apache Directory Server，那么 Analytic Server 不会在安装期间安装 ApacheDirectory Server。仅在初始 Analytic Server 安装期间选中时，才会安装 Apache Directory Server。

<b>LDAP type</b> as ldap.type	Analytic Server Default Group <input type="radio"/> openldap <input type="radio"/> ad <input checked="" type="radio"/> ads	?
<b>LDAP host</b> as ldap.host	Analytic Server Default Group <input type="text"/> Missing required value: LDAP host	?
<b>Bind DN</b> as ldap.binddn	Analytic Server Default Group <input type="text" value="uid=admin,ou=system"/>	?
<b>Bind password</b> as ldap.bindpassword	Analytic Server Default Group <input type="text"/> Missing required value: Bind password	?
<b>Base DN</b> as ldap.basedn	Analytic Server Default Group <input type="text" value="dc=ibm,dc=com"/>	?
<b>Enable SSL</b> as ldap.ssl.enabled	<input checked="" type="checkbox"/> Analytic Server Default Group	?
<b>SSL settings id</b> as ldap.ssl.reference	Analytic Server Default Group <input type="text" value="LDAPSSLSettings"/>	?
<b>SSL configuration</b> as ldap.ssl.content	Analytic Server Default Group <input type="text" value="&lt;ssl id='LDAPSSLSettings' keyStoreRef='LDAPTrustStore' trustStoreRef='LDAPTrustStore' /&gt; &lt;keyStore id='LDAPTrustStore' location='/opt"/>	?
<b>LDAP user and group filter</b> as ldap.filter	Analytic Server Default Group <input "="" type="text" value="&lt;customFilters id='customFilters' userFilter='(&amp;(cn=%v)(objectClass=organizationalPerson))' groupFilter='(&amp;(cn=%v)(objectclass="/>	?
<b>LDAP Port</b> as ldap.port	Analytic Server Default Group <input type="text" value="10636"/>	?

图 4. 示例 LDAP 配置设置

8. 在启用 Kerberos 的 Cloudera 环境中安装 Analytic Server 时，还必须在“添加服务向导”中配置以下设置：

注：Analytic Server 检查 HDFS 配置以获取 Kerberos 相关值从而用于认证。

- 如果想要在登录到 Analytic Server 控制台时启用 Kerberos 认证，请选择 Kerberos 作为 **Analytic Server 安全性** 设置。在选择 **Kerberos** 作为 **Analytic Server 安全性** 设置时，Analytic Server 控制台缺省为 Kerberos 登录方式。
- 想要连接到启用 Kerberos 的数据库时，选择 Kerberos 作为 **Analytic Server 数据库数据源连接方法** 设置。如果选择 **Kerberos** 作为 **Analytic Server 数据库数据源连接方法** 设置，那么在连接到数据库时 Analytic Server 控制台使用 Kerberos 方式。
- **Kerberos 域名和 KDC 主机** 设置是必需的。**Kerberos 域名和 KDC 主机** 值位于 Kerberos 密钥分发中心 (KDC) 服务器上的 krb5.conf 文件中。

<b>Analytic Server security</b> default.security.provider	Analytic Server Default Group ← <input checked="" type="radio"/> Kerberos <input type="radio"/> WebSphere
<b>Analytic Server database datasource connection method</b> as.db.connect.method	Analytic Server Default Group ← <input type="radio"/> Basic <input checked="" type="radio"/> Kerberos
<b>Kerberos Realm Name</b> kdcrealm	Analytic Server Default Group ← <input type="text" value="IBM.COM"/>
<b>KDC host</b> kdcserver	Analytic Server Default Group ← <input type="text" value="alcoa1.fyre.ibm.com"/>

图 5. 示例 Kerberos 设置

注释:

- **Analytic Server** 安全性和 **Analytic Server** 数据库数据源连接方法设置适用于 IBM SPSS Modeler 客户机和 Analytic Server 控制台认证。
- 在将 **Analytic Server** 数据库数据源连接方法设置为 Kerberos 时，必须确保目标数据库也启用 Kerberos。
- **Analytic Server** 安全性和 **Analytic Server** 数据库数据源连接方法设置不在 Hadoop 集群上配置 Kerberos 认证。有关更多信息，请参阅“启用 Kerberos 模拟”部分。
- 如果想要在登录时启用 Kerberos 认证，必须部署 IBM SPSS Modeler 客户机作为有效的 Kerberos 客户机。通过在 Kerberos 密钥分发中心 (KDC) 服务器中使用 **addprinc** 命令来完成。有关更多信息，请参阅 IBM SPSS Modeler 文档。

在启用 Kerberos 的 Cloudera 环境中安装 Analytic Server 时，还必须在 Kerberos 中创建必需的帐户并启用 Kerberos 模拟。有关更多信息，请参阅第 45 页的『配置 Kerberos』。

注：在成功安装 Analytic Server 之后，请勿单击 Cloudera Manager 中 Analytic Server 服务页面的“操作”列表上的创建 **Analytic Server Metastore**。创建 metastore 以覆盖现有元数据存储库。

脱机安装

脱机安装步骤与联机步骤相同，唯一的差别是您必须手动下载适用于您的特定操作系统的包文件和元数据。

RedHat Linux 需要以下文件：

- AnalyticServer-3.1.2.0-el7.parcel
- AnalyticServer-3.1.2.0-el7.parcel.sha
- manifest.json

SuSE Linux 需要以下文件：

- AnalyticServer-3.1.2.0-sles11.parcel
- AnalyticServer-3.1.2.0-sles11.parcel.sha
- manifest.json

or

- AnalyticServer-3.1.2.0-sles12.parcel
- AnalyticServer-3.1.2.0-sles12.parcel.sha

Ubuntu Linux 14.04 需要以下文件：

- AnalyticServer-3.1.2.0-trusty.parcel
- AnalyticServer-3.1.2.0-trusty.parcel.sha

Ubuntu Linux 16.04 需要以下文件：

- AnalyticServer-3.1.2.0-xenial.parcel
- AnalyticServer-3.1.2.0-xenial.parcel.sha

1. 在 Cloudera Manager 主集群节点上，下载并运行 Cloudera 自抽取 \*.bin 安装程序。接受许可协议并保持缺省 CSD 安装目录不变，以按照安装提示进行操作。

注：如果与缺省位置不同，必须指定其他 CSD 目录。

2. 将所需的包和元数据文件复制到 Cloudera Manager 主集群节点上的本地 Cloudera repo 路径中。缺省路径是 /opt/cloudera/parcel-repo（该路径可以在 Cloudera Manager 用户界面中配置）。
3. 使用以下命令以重新启动 Cloudera Manager：

```
service cloudera-scm-server restart
```

在 Cloudera Manager 刷新包之后，**AnalyticServer** 包显示为已下载。您可以单击检查新包以强制刷新。

4. 单击分发 > 激活。

此时 **AnalyticServer** 包会显示为已分发和已激活。

5. 在 Cloudera Manager 中，添加 Analytic Server 作为服务。请参阅“联机安装”部分中的步骤 7 和 8 以获取更多信息。

---

## 配置 Cloudera

安装后，您可以通过 Cloudera Manager 有选择地配置和管理 Analytic Server。

注：以下约定用于 Analytic Server 文件路径。

- {AS\_ROOT} 指的是部署 Analytic Server 的位置；例如 /opt/cloudera/parcels/AnalyticServer。
- {AS\_SERVER\_ROOT} 指的是配置文件、日志文件和服务器文件的位置；例如，/opt/cloudera/parcels/AnalyticServer/ae\_wlpserver/usr/servers/aeserver。



- {AS\_HOME} 指的是 HDFS 上 Analytic Server 用作根文件夹的位置；例如，/user/as\_user/analytic-root。

## 安全性

IBM SPSS Modeler options.cfg 文件中的缺省 **tenant\_id** 值为 **ibm**。您可以在 Analytic Server 控制台中查看租户。请参阅《IBM SPSS Analytic Server 管理员指南》，以获取有关租户管理的详细信息。

## 配置 LDAP 注册表

在 Analytic Server 安装期间配置 LDAP。您可以在 Analytic Server 安装后更改为其他 LDAP 服务器。

注：在 Analytic Server 中对 LDAP 的支持由 WebSphere Liberty 进行控制。有关更多信息，请参阅在 Liberty 中配置 LDAP 用户注册表。

## 配置从 Analytic Server 到 LDAP 的安全套接字层 (SSL) 连接

1. 以 Analytic Server 用户身份登录每台 Analytic Server 机器，然后为 SSL 证书创建公共目录。

注：在 Cloudera 上，Analytic Server 用户始终是 as\_user，这一点无法更改。

2. 将密钥库和信任库文件复制到所有 Analytic Server 机器上的某一公共目录。另外，将 LDAP 客户机 CA 证书添加至信任库。以下是一些指示信息示例。

```
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
```

注：JAVA\_HOME 是用于 Analytic Server 启动的同一 JRE。

3. 可以使用 {AS\_ROOT}/ae\_wlpserver/bin 的 securityUtility 工具对密码进行编码，从而对它们的值进行加密。示例如下所示。

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

4. 登录 Cloudera Manager 并使用正确的 SSL 配置设置更新 Analytic Server 配置设置 **ssl\_cfg**。示例如下所示。

```
<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="/>
```

注：对密钥库文件和信任库文件使用绝对路径。

5. 使用正确的 LDAP 配置设置更新 Analytic Server 配置设置 **security\_cfg**。例如，在 **ldapRegistry** 元素中，将 **sslEnabled** 属性设置为 true 并将 **sslRef** 属性设置为 defaultSSLConfig。

## 配置 Kerberos

Analytic Server 支持在 Cloudera 中使用 Kerberos。以下部分提供配置设置以确保以兼容 Analytic Server 的方式正确配置 Kerberos。

注：Analytic Server 检查 HDFS 配置以获取 Kerberos 相关值从而用于认证。

## Analytic Server 和 Kerberos 设置

在启用 Kerberos 的 Cloudera 环境中安装 Analytic Server 时，请记住以下设置。

- 如果想要在登录到 Analytic Server 控制台时启用 Kerberos 认证，请选择 Kerberos 作为 **Analytic Server 安全性** 设置。在选择 **Kerberos** 作为 **Analytic Server 安全性** 设置时，Analytic Server 控制台缺省为 Kerberos 登录方式。
- 想要连接到启用 Kerberos 的数据库时，选择 Kerberos 作为 **Analytic Server 数据库数据源连接方法** 设置。如果选择 **Kerberos** 作为 **Analytic Server 数据库数据源连接方法** 设置，那么在连接到数据库时 Analytic Server 控制台使用 Kerberos 方式。
- **Kerberos 域名**和 **KDC 主机**设置是必需的。**Kerberos 域名**和 **KDC 主机**值位于 Kerberos 密钥分发中心 (KDC) 服务器上的 krb5.conf 文件中。

The image shows a configuration interface for Analytic Server with four sections:

- Analytic Server security** (default.security.provider): Set to Analytic Server Default Group, with Kerberos selected (radio button checked) and WebSphere unselected.
- Analytic Server database datasource connection method** (as.db.connect.method): Set to Analytic Server Default Group, with Kerberos selected (radio button checked) and Basic unselected.
- Kerberos Realm Name** (kdcrealm): Set to Analytic Server Default Group, with the value IBM.COM entered in the text field.
- KDC host** (kdcserver): Set to Analytic Server Default Group, with the value alcoa1.fyre.ibm.com entered in the text field.

图 6. 示例 Kerberos 设置

注释：

- **Analytic Server 安全性**和 **Analytic Server 数据库数据源连接方法**设置适用于 IBM SPSS Modeler 客户机和 Analytic Server 控制台认证。
- 在将 **Analytic Server 数据库数据源连接方法**设置为 Kerberos 时，必须确保目标数据库也启用 Kerberos。

- **Analytic Server 安全性和 Analytic Server 数据库数据源连接方法** 设置不在 Hadoop 集群上配置 Kerberos 认证。有关更多信息，请参阅"启用 Kerberos 模拟"部分。
- 如果想要在登录时启用 Kerberos 认证，必须部署 IBM SPSS Modeler 客户机作为有效的 Kerberos 客户机。通过在 Kerberos 密钥分发中心 (KDC) 服务器中使用 **addprinc** 命令来完成。有关更多信息，请参阅 IBM SPSS Modeler 文档。

## 在 Kerberos 中创建必需的帐户

1. 在 Kerberos 用户存储库中，为您计划向其提供对 Analytic Server 的访问权的所有用户创建帐户。
2. 在 LDAP 服务器上创建相同帐户（从先前步骤）。
3. 在每个 Analytic Server 节点和 Hadoop 节点上，为前一步骤中创建的每个用户创建一个操作系统用户帐户。
  - 确保这些用户的 UID 在所有机器上都匹配。您可以使用 **kinit** 命令登录每个帐户来测试此内容。
  - 确保 UID 符合用于提交作业的最小用户标识 Yarn 设置。这是 **container-executor.cfg** 中的 **min.user.id** 设置。例如，如果 **min.user.id** 为 1000，那么创建的每个用户帐户都必须具有大于或等于 1000 的 UID。
4. 在 HDFS 上为 Analytic Server 管理员用户创建用户主文件夹。文件夹许可权必须设置为 777，所有者必须定义为 **admin**，并且用户组必须设置为 **hdfs**。请参阅以下粗体示例：

```
[root@xxxxx configuration]# hadoop fs -ls /user
Found 9 items
drwxrwxrwx - hdfs supergroup 0 2017-07-26 03:41 /user/AE
drwxrwxrwx - admin hdfs 0 2017-06-08 01:33 /user/admin
drwxr-x--x - as_user hdfs 0 2017-06-06 01:00 /user/as_user
drwx----- - hdfs supergroup 0 2017-07-31 00:17 /user/hdfs
drwxrwxrwx - mapred hadoop 0 2017-06-05 00:28 /user/history
drwxrwxr-t - hive hive 0 2017-06-05 00:30 /user/hive
drwxrwxr-x - hue hue 0 2017-06-05 00:30 /user/hue
drwxrwxr-x - impala impala 0 2017-07-19 00:52 /user/impala
drwxr-x--x - spark spark 0 2017-06-05 01:34 /user/spark
```

5. 如果您计划使用 HCatalog 数据源，并且 Analytic Server 与 Hive metastore 安装在不同的机器上，那么需要在 HDFS 上模拟 Hive 客户机。
  - a. 在 Cloudera Manager 中，浏览至 HDFS 服务的"配置"选项卡。
 

注：如果尚未设置以下设置，那么可能不会显示在配置选项卡上。在此情况下，请运行搜索来查找这些参数。
  - b. 编辑 **hadoop.proxyuser.hive.groups** 设置以使其具有值 **\***，或者编辑包含所有能够登录 Analytic Server 的用户的组。
  - c. 编辑 **hadoop.proxyuser.hive.hosts** 设置以使其具有值 **\***，或者编辑 Hive metastore 和每个 Analytic Server 实例都作为服务安装在其上主机的列表。
  - d. 重新启动 HDFS 服务。

执行这些步骤并且安装了 Analytic Server 步骤后，Analytic Server 将以静默方式自动配置 Kerberos。

## 启用 Kerberos 模拟

模拟允许在安全上下文中执行线程，该安全上下文不同于拥有线程的进程的安全上下文。例如，模拟为 Hadoop 作业提供以除标准 Analytic Server 用户 (**as\_user**) 以外的其他用户身份运行的方式。要启用 Kerberos 模拟：

1. 打开 Cloudera Manager 并在 **core-site.xml** 的集群范围高级配置片段（安全值）（位于 **HDFS**（服务范围） > 配置选项卡上）区域中添加或更新以下属性。
  - 名称：**hadoop.proxyuser.as\_user.hosts**

- 值: \*
- 名称: `hadoop.proxyuser.as_user.groups`
- 值: \*

注: **core-site.xml** 设置应用于 Hadoop 配置 (而不是 Analytic Server)。

2. 在 Analytic Server 节点上从命令 `shell` 运行以下命令:

```
hadoop fs -chmod -R 755 /user/as_user/analytic-root/analytic-workspace
```

## 使用 Kerberos 针对单点登录 (SSO) 配置 HAProxy

1. 按照 HAProxy 文档指南配置并启动 HAProxy: <http://www.haproxy.org/#docs>
2. 针对 HAProxy 主机创建 Kerberos 原则 (`HTTP/<proxyHostname>@<realm>`) 和密钥表文件, 其中, `<proxyHostname>` 是 HAProxy 主机的完整名称, 而 `<realm>` 是 Kerberos 域。
3. 将密钥表文件作为 `/etc/security/keytabs/spnego_proxy.service.keytab` 复制到每个 Analytic Server 主机
4. 在每个 Analytic Server 主机上更新此文件的许可权。 示例如下所示。
 

```
chown root:hadoop /etc/security/keytabs/spnego_proxy.service.keytab
chmod 440 /etc/security/keytabs/spnego_proxy.service.keytab
```
5. 打开 Cloudera Manager, 并在 Analytic Server 的 **analyticserver-conf/config.properties** 的 **Analytic Server 高级配置片段 (安全值)** 区域中添加或更新以下属性。
 

```
web.authentication.kerberos.keytab=/etc/security/keytabs/spnego_proxy.service.keytab
web.authentication.kerberos.principal=HTTP/<proxy machine full name>@<realm>
```
6. 从 Cloudera Manager 保存配置并重新启动所有 Analytic Server 服务。
7. 指示用户将其浏览器配置为使用 Kerberos。

现在, 用户能够使用 Kerberos SSO 登录到 Analytic Server。

## 禁用 Kerberos

1. 在 Cloudera Manager 控制台中禁用 Kerberos。
2. 停止 Analytic Server 服务。
3. 从 **analyticserver-conf/config.properties** 的 **Analytic Server 高级配置片段 (安全值)** 区域中移除以下设置。

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

4. 单击**保存更改**, 然后重新启动 Analytic Server 服务。

## 启用与 Analytic Server 控制台的安全套接字层 (SSL) 连接

缺省情况下, Analytic Server 会生成自签名证书以启用安全套接字层 (SSL), 因此您可以接受自签名证书, 通过安全端口来访问 Analytic Server 控制台。 要使 HTTPS 访问更安全, 您需要安装第三方供应商证书。

要安装第三方供应商证书, 请执行下列步骤。

1. 将第三方供应商密钥库和信任库复制到所有 Analytic Server 节点中的同一目录; 例如, `/home/as_user/security`。

注：Analytic Server 用户必须具有对此目录的读访问权。

2. 在 Cloudera Manager 中，导航到 Analytic Server 服务的"配置"选项卡。
3. 编辑 `ssl_cfg` 参数。

```
<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>"/>
```

将

- `<KEYSTORE-LOCATION>` 替换为密钥库的绝对位置；例如，`/home/as_user/security/mykey.jks`
- `<TRUSTSTORE-LOCATION>` 替换信任库的绝对位置；例如：`/home/as_user/security/mytrust.jks`
- `<TYPE>` 替换为证书的类型；例如：JKS、PKCS12 等。
- `<PASSWORD>` 替换为使用 Base64 加密格式的加密密码。对于编码，您可以使用 `securityUtility`；例如：`{AS_ROOT}/ae_wlpserver/bin/securityUtility encode <password>`

如果要生成自签名证书，您可以使用 `securityUtility`；例如：`{AS_ROOT}/ae_wlpserver/bin/securityUtility createSSLCertificate --server=myserver --password=mypassword --validity=365 --subject=CN=mycompany,O=myOrg,C=myCountry`。有关 `securityUtility` 和其他 SSL 设置的更多信息，请参阅 `WebSphere Liberty` 概要文件文档。

注：您必须针对 CN 值提供相应的主机域名。

4. 单击保存更改，然后重新启动 Analytic Server 服务。

## 通过 SSL 与 Apache Hive 通信

您必须更新 `hive.properties` 文件以通过 SSL 连接与 Apache Hive 通信。或者，如果已针对高可用性启用 Apache Hive 环境，那么可选择主 Analytic Server 数据源页面上的高可用性参数。

### 更新 `hive.properties` 文件

1. 打开 `hive.properties` 文件。该文件位于：`/opt/cloudera/parcels/AnalyticServer-3.1.2.0/ae_wlpserver/usr/servers/aeserver/configuration/database`
2. 找到以下行：

```
jdbcurl = jdbc:hive2://{db.servername}:{db.serverport}/{db.databasesname};user={db.username};password={db.password}
```

3. 通过添加以下**粗体**信息更新行：

```
jdbcurl = jdbc:hive2://{db.servername}:{db.serverport}/{db.databasesname};user={db.username};password={db.password};
ssl=true;sslTrustStore=pathtotheirtruststorefile;trustStorePassword=xxxtheirTrustStorePassword
```

4. 保存 `hive.properties` 文件。

## 启用对 Essentials for R 的支持

Analytic Server 支持对 R 模型进行评分，并支持运行 R 脚本。

要在 Cloudera Manager 中成功安装 Analytic Server 之后安装 Essentials for R：

1. 供应 Essentials for R 的服务器环境。有关更多信息，请参见第 19 页的『启用对 Essentials for R 的支持』中的步骤 1。
2. 下载用于 IBM SPSS Modeler Essentials for R RPM 的自解压缩归档 (BIN)。Essentials for R 可从以下地址下载：<https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=swg-tspssp>。选择特定于堆栈、堆栈版本及硬件体系结构的文件。
3. 在 Cloudera Manager 服务器主机上以 root 或 sudo 用户身份运行自抽取归档。以下包必须安装到已配置的存储库或可从已配置的存储库中使用：
  - Red Hat Linux: gcc-gfortran、zip 和 gcc-c++
  - SUSE Linux: gcc-fortran、zip 和 gcc-c++
  - Ubuntu Linux: gcc-fortran、zip 和 gcc-c++
4. 自抽取安装程序执行以下任务：
  - a. 显示必需许可证并提示安装程序接受这些许可证。
  - b. 提示安装程序输入 R 源位置，或者使用缺省位置继续操作。安装的缺省 R 版本是 3.3.2。要安装其他版本：
    - 联机安装：提供指向必需 R 版本归档的 URL。例如，<https://cran.r-project.org/src/base/R-2/R-2.15.3.tar.gz> 表示 R 2.15.3。
    - 脱机安装：下载必需 R 版本归档，然后将该归档复制到 Cloudera Manager 服务器主机。请勿重命名该归档（缺省情况下，其名称为 R-x.x.x.tar.gz）。提供指向所复制的 R 归档的 URL，如下所示：`file://<R_archive_directory>/R-x.x.x.tar.gz`。如果下载了 R-2.15.3.tar.gz 归档并将其复制到 /root，URL 为 `file:///root/R-2.15.3.tar.gz`。
  - c. 安装 R 所需的软件包。
  - d. 下载并安装 R 以及 Essentials for R 插件。
  - e. 创建包和 parcel.sha 文件，并将它们复制到 /opt/cloudera/parcel-repo。如果位置已发生更改，请输入正确的位置。
5. 安装完成后，在 Cloudera Manager 中分发并激活 **Essentials for R** 包（单击检查新包以刷新包列表）。
6. 如果 Analytic Server 服务已安装：
  - a. 停止服务。
  - b. 刷新 Analytic Server 二进制文件。
  - c. 启动该服务以完成 Essentials for R 安装。
7. 如果未安装 Analytic Server 服务，那么继续其安装。

注：所有 Analytic Server 主机都必须安装相应的归档（zip 和 unzip）软件包。

## 启用关系数据库源

如果您在每个 Analytic Server 主机的共享目录中提供 JDBC 驱动程序，那么 Analytic Server 可以使用关系数据库源。缺省情况下，此目录为 /usr/share/jdbc。

要更改共享目录，请执行下列步骤。

1. 在 Cloudera Manager 中，导航到 Analytic Server 服务的“配置”选项卡。
2. 在 **jdbc.drivers.location** 中，指定 JDBC 驱动程序的共享目录的路径。
3. 单击保存更改。
4. 从操作下拉菜单中选择停止以停止 Analytic Server 服务。

5. 从操作下拉菜单中选择刷新 **Analytic Server** 二进制文件。
6. 从操作下拉菜单中选择启动以启动 **Analytic Server** 服务。

表 11. 受支持的数据库

数据库	受支持的版本	JDBC 驱动程序 JAR	供应商
Amazon Redshift	8.0.2 或更高版本	RedshiftJDBC41-1.1.6.1006.jar 或更高版本	Amazon
Apache Impala	具有 2.5.5 或更高版本的 JDBC 4	ImpalaJDBC4.jar、commons-codec-*.jar、commons-logging-*.jar、httpclient-*.jar、httpcore-*.jar、log4j-*.jar、libthrift-*.jar、libfb303-*.jar、slf4j-api-*.jar、ql.jar、zookeeper-*.jar 和 TCLIServiceClient.jar	Apache
DashDB	Bluemix 服务	db2jcc.jar	IBM
Db2 for Linux、UNIX 和 Windows	11.1、10.5、10.1 和 9.7	db2jcc.jar	IBM
Db2 z/OS	11 和 10	db2jcc.jar 和 db2_license_cisuz.jar	IBM
Greenplum	5.x	postgresql.jar	Greenplum
Hive	1.1	hive-jdbc-*.jar	Apache
MySQL	5.6 和 5.7	mysql-connector-java-commercial-5.1.25-bin.jar	MySQL
Netezza	7 和 6.x	nzjdbc.jar	IBM
Oracle	12c 和 11g R2 (11.2)	ojdbc6.jar 和 orai18n.jar	Oracle
SQL Server	2014、2012 和 2008 R2	sqljdbc4.jar	Microsoft
Teradata	15 和 15.1	tdgssconfig.jar 和 terajdbc4.jar	Teradata

## 注释

- 如果在安装 **Analytic Server** 之前创建了 **Redshift** 数据源，那么需要执行以下步骤，才能使用该 **Redshift** 数据源。
  1. 在 **Analytic Server** 控制台中，打开该 **Redshift** 数据源。
  2. 选择 **Redshift** 数据库数据源。
  3. 输入 **Redshift** 服务器地址。
  4. 输入数据库名称和用户名。密码应自动填充。
  5. 选择数据库表。

## 启用 HCatalog 数据源

**Analytic Server** 通过 **Hive/HCatalog** 提供对许多数据源的支持。一些源需要手动配置步骤。

1. 收集必需的 JAR 文件以启用数据源。请参阅下列各节以获取详细信息。
2. 将这些 JAR 文件添加至 {HIVE\_HOME}/auxlib 目录及每个 **Analytic Server** 节点的 /usr/share/hive 目录。

3. 重新启动 Hive Metastore 服务。
4. 重新启动 Analytic Server 服务的每个实例。

注：

在通过 Analytic Server HCatalog 数据源访问 HBase 数据时，访问用户必须具有 HBase 表的读许可权。

- 在非 Kerberos 环境中，Analytic Server 使用 as\_user 访问 HBase (as\_user 必须具有 HBase 的读许可权)。
- 在 Kerberos 环境中，as\_user 和登录用户必须具有 HBase 表的读许可权。

## NoSQL 数据库

Analytic Server 支持任何 NoSQL 数据库，可以从供应商处获取用于此数据库的 Hive 存储处理程序。

不需要任何额外步骤即可启用对 Apache HBase 及 Apache Accumulo 的支持。

对于其他 NoSQL 数据库，请联系联系人供应商并获取存储处理程序及相关 JAR。

## 基于文件的 Hive 表

Analytic Server 支持任何基于文件的 Hive 表，有一个内置或定制 Hive SerDe (序列化器-反序列化器) 可用于此表。

用于处理 XML 文件的 Hive XML SerDe 位于 Maven Central Repository 中，网址为 <http://search.maven.org/#search%7Cga%7C1%7Chivexmlserde>。

## MapReduce v2 作业

使用 **analyticserver-conf/config.properties** 的 **Analytic Server** 高级配置片段 (安全值) 区域中的 **preferred.mapreduce** 设置来控制 MapReduce 作业的处理方式：

表 12. *analyticserver-conf/config.properties* 的 *Analytic Server* 高级配置片段 (安全值)

属性	描述
preferred.mapreduce	控制运行 MapReduce 作业的方法。有效值包括： <ul style="list-style-type: none"> <li>• spark</li> <li>• m3r</li> <li>• hadoop</li> </ul> 例如：preferred.mapreduce=spark

## Apache Spark

如果想要使用 Spark (V1.5 或更高版本)，必须在 Analytic Server 安装期间选择 spark.version。

1. 打开 Cloudera Manager 并在分析服务器 **Spark** 版本区域中选择相应的 spark.version (例如，None、1.x 或 2.x)。

注：在使用 Spark 1.x 时，还必须在 **analyticserver-conf/config.properties** 的 **Analytic Server** 高级配置片段 (安全值) 区域中添加以下行。

```
spark.extraListeners=org.apache.spark.JavaSparkListener
```

2. 保存配置。



## 配置 Apache Impala

Apache Impala 在 Cloudera 上针对 Analytic Server 数据库数据源或 HCatalog 数据源运行时受支持（无论 Impala 是否启用 SSL）。

### 为 Apache Impala 数据创建数据库数据源

1. 在 Analytic Server 数据源主页上，单击**新建**以创建新数据源。系统将显示“新建数据源”对话框。
2. 在新建数据源字段中输入相应的名称，选择 Database 作为内容类型值，然后单击**确定**。
3. 打开数据库选择部分并输入以下信息。

**数据库：**

从下拉菜单中选择 **Impala**。

**服务器地址：**

输入托管 Impala 守护程序的服务器的 URL。在为 Analytic Server 启用了 Kerberos 时需要标准域名。

**服务器端口：**

输入 Impala 数据库侦听的端口号。

**数据库名称：**

输入要连接到的数据库的名称。

**用户名：**

输入有权登录到 Impala 数据库的用户名。

**密码：**

输入相应的用户名密码。

**表名称：**

输入要使用的数据库中的表的名称。单击**选择**以手动选择文件。

**最大并发读取数：**

输入可以从 Analytic Server 发送到数据库来从数据源中指定的表进行读取的并发查询的数量的限制。

4. 在输入所需信息后，单击**保存**。

### 为 Apache Impala 数据创建 HCatalog 数据源

1. 在 Analytic Server 数据源主页上，单击**新建**以创建新数据源。系统将显示“新建数据源”对话框。
2. 在新建数据源字段中输入相应的名称，选择 HCatalog 作为内容类型值，然后单击**确定**。
3. 打开数据库选择部分并输入以下信息。

**数据库：**

从下拉菜单中选择**缺省**。

**表名称：**

输入要使用的数据库中的表的名称。

**HCatalog 模式**

选择 **HCatalog** 元素选项，然后选择相应的 **HCatalog** 字段映射选项。

4. 在输入所需信息后，单击**保存**。

### 连接到 Apache Impala 启用的数据

1. 在 Cloudera Manager 控制台中定义以下 Impala SSL 设置。

### 为 Impala 启用 TLS/SSL (client\_services\_ssl\_enabled)

选择 **Impala (服务范围)** 选项。

### Impala TLS/SSL 服务器证书文件 (PEM 格式) (ssl\_server\_certificate)

输入自签名 PEM 格式证书位置和文件名 (例如: /tmp/<user\_name>/ssl/114200v21.crt)。

### Impala TLS/SSL 服务器专用密钥文件 (PEM 格式) (ssl\_private\_key)

输入专用密钥 (PEM 格式)、位置和文件名 (例如: /tmp/<user\_name>/ssl/114200v21.key)。

2. 在 Analytic Server 主机上, 将 \*.crf 文件 (用于启用 Impala SSL) 导入到 \*.jks 文件中。该文件可以是 cacerts 文件 (例如 /etc/pki/java/cacerts) 或任何其他 \*.jks 文件。
3. 在 Analytic Server 主机上, 通过附加以下 jdbcurl 键值来更新 Impala 配置文件 (impala.properties):  
SSL=1;AllowSelfSignedCerts=1;CAIssuedCertNamesMismatch=1;

注: 在使用 \*.jks 文件 (不同于 cacerts) 时, 还需要指定以下内容:

```
SSLTrustStore=<your_pks_file>;SSLTrustStorePwd=<password_for_pks_file>;
```

4. 在 Cloudera Manager 控制台中重新启动 Analytic Server。

## 更改 Analytic Server 所使用的端口

缺省情况下, Analytic Server 对 HTTP 和 HTTPS 分别使用 9080 和 9443 端口。要更改端口设置, 请执行下列步骤。

1. 在 Cloudera Manager 中, 导航到 Analytic Server 服务的"配置"选项卡。
2. 在 **http.port** 和 **https.port** 中分别指定所需的 HTTP 和 HTTPS 端口。

注: 您可能需要选择"过滤"部分中的端口和地址类别才能看到这两个参数。

3. 单击**保存更改**。
4. 重新启动 Analytic Server 服务。

## 高可用性 Analytic Server

通过将 Analytic Server 作为服务添加至集群中的多个节点, 您可以实现它的高可用性。

1. 在 Cloudera Manager 中, 导航到 Analytic Server 服务的"实例"选项卡。
2. 单击**添加角色实例**并选择要将 Analytic Server 添加为服务的主机。

## 多集群支持

多集群功能是 IBM SPSS Analytic Server 的高可用性功能的增强功能, 它可在多租户环境中提供更强的隔离。缺省情况下, 安装 Analytic Server 服务 (在 Ambari 或 ClouderaManager 中) 会导致系统定义一个分析服务器集群。

该集群规范定义 Analytic Server 集群成员信息。修改集群规范可通过 XML 内容来完成 (在 Ambari Analytic Server 配置的 analytics-cluster 字段中或通过手动编辑 Cloudera Manager 的 configuration/analytics-cluster.xml 文件)。当配置多个 Analytic Server 集群时, 必须使用每个 Analytic Server 集群自身的负载均衡器将请求馈送到这些集群。

使用多集群功能可确保一个租户的工作不可能对在其他租户的集群上执行的工作产生负面影响。对于高度可用的作业, 作业故障转移仅发生在启动工作的 Analytic Server 集群的作用域中。下面的示例提供了一个多集群 XML 规范。

注: 可以通过将 Analytic Server 作为服务添加到集群中的多个节点来使它具有高可用性。

```

<analyticServerClusterSpec>
  <cardinality>1+</cardinality>
  <cluster name="cluster1">
    <memberName>one.cluster</memberName>
    <memberName>two.cluster</memberName>
  </cluster>
  <cluster name="cluster2">
    <memberName>three.cluster</memberName>
    <memberName>four.cluster</memberName>
  </cluster>
</analyticServerClusterSpec>

```

在上面的示例中，需要两个负载均衡器。一个负载均衡器将请求发送至 cluster1 成员 (one.cluster 和 two.cluster)，另一个负载均衡器将请求发送至 cluster2 成员 (three.cluster 和 four.cluster)。

以下示例提供了一个单集群 XML 规范 (缺省配置)。

```

<analyticServerClusterSpec>
  <cardinality>1</cardinality>
  <cluster name="cluster1">
    <memberName>*</memberName>
  </cluster>
</analyticServerClusterSpec>

```

在上面的示例中，需要一个负载均衡器来处理配置了多个集群成员的情况。

## 注释

- 只有单集群支持在 **memberName** 元素中使用通配符 (例如，集群基数 = "1")。cardinality 元素的有效值是 1 和 1+。
- 必须以与分配了 Analytic Server 角色的主机名相同的方式指定 **memberName**。
- 应用了集群配置更改后，必须重新启动所有集群中的所有服务器。
- 在 Cloudera Manager 中，必须修改并维护所有 Analytic Server 节点上的 analytics-cluster.xml 文件。对所有节点都必须进行维护，以确保它们包含相同的内容。

## 针对小型数据优化 JVM 选项

您可以编辑 JVM 属性，以便在运行小型 (M3R) 作业时优化系统。

在 Cloudera Manager 中，请查看 Analytic Server 服务中"配置"选项卡上的 **JVM 选项 (jvm.options)** 控件。修改下列参数会设置托管 Analytic Server 的服务器上运行的作业的堆大小；即，并非 Hadoop。运行较小的 (M3R) 作业时，这一点很重要，并且您可能需要对这些值进行试验以优化系统。

```

-Xms512M
-Xmx2048M

```

## 针对每个 IBM SPSS Analytic Server 租户配置单独的 YARN 请求 - Cloudera

通过使用 Spark 动态资源分配技术完成 Yarn 队列配置。

### Cloudera 5.x

在向现有集群添加 SPSS Analytic Server 服务时遵循以下步骤。

1. 在 Cloudera Manager 中，浏览至 **SPSS Analytic Server 服务 > 配置**。
2. 将资源池启用：**resource.pool.enabled** 值更改为 true。
3. 将以下属性添加到 **Analytic Server 高级配置片段 (安全值) > analyticserver-conf.config.properties**：

```

config.folder.path=/etc/spark2/conf
resource.pool.mapping=tenant1:test,tenant2:production
resource.pool.default=default
spark.scheduler.mode=FAIR
spark.yarn.queue=default

```

表 13. *analyticserver-conf.config.properties* 设置

属性	描述
config.folder.path	该目录包含 <code>fairscheduler.xml</code> 文件，此文件包含 Spark 池属性信息。该文件是必需的并且必须手动创建。有关更多信息，请参阅 <b>fairscheduler.xml</b> 示例部分。
resource.pool.mapping	<p><b>Spark:</b> 将租户映射到在 <code>fairscheduler.xml</code> 文件中定义的池。必须用逗号分隔租户对（例如，<code>tenant1:test,tenant2:production</code>）。在指定池之前，确保在 <code>fairscheduler.xml</code> 文件中配置池。</p> <p><b>MapReduce:</b> 将租户映射到在动态资源池配置中定义的队列。必须用逗号分隔租户对（例如，<code>tenant1:test,tenant2:production</code>）。指定队列之前，确保使用队列配置系统，并且允许访问以向队列提交作业。</p> <p>注：如果想要一起运行 Spark 和 MapReduce 作业，那么租户映射值必须是 <code>fairscheduler.xml</code> 文件和动态资源池配置中的相同名称。</p>
resource.pool.default	<p><b>Spark:</b> 定义缺省资源池。该值可以是 <code>default</code> 或在 <code>fairscheduler.xml</code> 文件中定义的池名称。如果未配置租户（或者配置不正确），那么使用 <code>default</code> 设置。</p> <p><b>MapReduce:</b> 定义向其提交作业的缺省队列。</p>
spark.scheduler.mode=FAIR	<b>Spark:</b> 启用公平调度程序。不应更改该属性。
spark.yarn.queue	<b>Spark:</b> 将应用程序提交至的 YARN 队列的名称。您可以在动态资源池配置中指定定制的 YARN 队列名称。

#### 4. 保存配置并重新启动 Analytic Server 服务。

### fairscheduler.xml 示例

`fairscheduler.xml` 文件包含 Spark 池属性信息。该文件是必需的并且必须手动创建。

```

<?xml version="1.0"?>
<allocations>
  <pool name="production">
    <schedulingMode>FAIR</schedulingMode>
    <weight>1</weight>
    <minShare>2</minShare>
  </pool>
  <pool name="test">
    <schedulingMode>FIFO</schedulingMode>
    <weight>2</weight>
    <minShare>3</minShare>
  </pool>
</allocations>

```

### 参考

请参阅以下站点以获取更多信息：

- <https://spark.apache.org/docs/latest/job-scheduling.html#dynamic-resource-allocation>
- <https://spark.apache.org/docs/latest/running-on-yarn.html>

## 迁移

通过 Analytic Server，您可以将数据和配置设置从现有 Analytic Server 迁移到新安装。

### 升级到新版本的 Analytic Server

如果您已安装 Analytic Server 3.1 并已购买更新的版本，那么可以将 3.1 配置设置迁移到新安装。

**限制：**3.1 和新安装无法在同一 Hadoop 集群中共存。如果您将新安装配置为使用 3.1 安装所用的 Hadoop 集群，那么 3.1 安装将不再起作用。

### 3.1.1.1 到更新版本的迁移步骤

1. 根据第 39 页的『在 Cloudera 上安装』中的指示信息，安装新的 Analytic Server。
2. 将分析工作空间从旧安装复制到新安装。
  - a. 如果您不确定分析工作空间的位置，请运行 `hadoop -fs ls`。分析工作空间的路径的格式为 `/user/as_user/analytic-root/analytic-workspace`，其中 `as_user` 是拥有分析工作空间的用户标识。
  - b. 以 `as_user` 身份登录新 Analytic Server 安装版本的主机。删除 `/user/as_user/analytic-root/analytic-workspace` 目录（如果存在）。
  - c. 运行下列复制脚本。

```
hadoop distcp hftp://{host of 3.1.1.1 namenode}:50070/{path to 3.1.1.1 analytic-workspace}
hdfs://{host of 3.1.2 namenode}/user/as_user/analytic-root/analytic-workspace
```

3. 如果使用嵌入式 Apache Directory Server，那么使用第三方 LDAP 客户机工具备份当前用户/组配置。在安装 Analytic Server 3.1.2 后，将备份用户/组配置导入到 Apache Directory Server。

**注：**如果使用外部 LDAP 服务器，那么可跳过此步骤。

4. 在 Cloudera Manager 中，停止 Analytic Server 服务。
5. 从旧安装中收集配置设置。
  - a. 将新安装版本中的 `configcollector.zip` 归档复制到旧安装版本中的 `{AS_ROOT}\tools`。
  - b. 解压缩 `configcollector.zip` 的副本。这将在旧安装版本中创建新的 `configcollector` 子目录。
  - c. 通过执行 `{AS_ROOT}\tools\configcollector` 中的 **configcollector** 脚本，在旧安装版本中运行配置收集器工具。将生成的压缩 (ZIP) 文件复制到托管新安装的服务器。

**要点：**所提供的 **configcollector** 脚本可能与最新的 Analytic Server 版本不兼容。如果遇到有关 **configcollector** 脚本的问题，请联系 IBM 技术支持代表。

6. 清除 Zookeeper 状态。在 Zookeeper bin 目录中（例如 Cloudera 上的 `/opt/cloudera/parcels/CDH-5.4...../lib/zookeeper/bin`），运行以下命令。

```
./zkCli.sh rmr /AnalyticServer
```

7. 执行 **migrationtool** 脚本并将配置收集器所创建的压缩文件的路径作为参数进行传递。示例如下所示。

```
migrationtool.sh /opt/ibm/spss/analyticserver/3.1/ASConfiguration_3.1.1.1.xxx.zip
```

8. 在 Analytic Server 节点上从命令 `shell` 运行以下命令：

```
hadoop fs -chmod -R 755 /user/as_user/analytic-root/analytic-workspace
```

9. 在 Cloudera Manager 中，启动 Analytic Server 服务。

**注：**如果已将 R 配置为与现有 Analytic Server 安装配合使用，那么您将需要遵循相应步骤，将其配置为用于新的 Analytic Server 安装。

---

## 在 Cloudera 上卸载 Analytic Server

Cloudera 自动处理卸载 Analytic Server 服务和包所需的大部分步骤。

以下是从 Cloudera 环境中清除 Analytic Server 所必需的步骤：

1. 停止并删除 Analytic Server 服务。
2. 取消激活、从主机除去和删除 Analytic Server 包。
3. 删除 HDFS 中的 Analytic Server 用户目录。缺省位置是 `/user/as_user/analytic-root`。

4. 删除 Analytic Server 使用的数据库或模式。
5. 清除任何其余 Analytic Server 安装软件包。通过删除以下项完成此操作：
  - csd 文件夹
  - 任何现有 3.1.2 文件位于 parcels、parcel-cache 和 parcel-repo 文件夹中。

---

## 第 4 章 配置 IBM SPSS Modeler 以用于 IBM SPSS Analytic Server

为了使 SPSS Modeler 能够与 Analytic Server 配套使用，您需要更新 SPSS Modeler 服务器安装。

1. 配置 SPSS Modeler 服务器以将其与 Analytic Server 安装相关联。
  - a. 编辑主服务器安装目录的 config 子目录中的 options.cfg 文件，并添加或编辑以下行：

```
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**

如果在 Analytic Server 上配置安全通信，那么指定 Y；否则，指定 N。

### **as\_host**

托管 Analytic Server 的服务器的 IP 地址/主机名。

注：在针对 Analytic Server 启用 SSL 时，必须提供相应的 IP 地址/主机域名。

### **as\_port**

Analytic Server 侦听的端口（缺省情况下为 8080）。

### **as\_context\_root**

Analytic Server 上下文根（缺省情况下为 analyticserver）。

### **as\_tenant**

SPSS Modeler 服务器安装所属的租户（缺省租户为 ibm）。

### **as\_prompt\_for\_password**

如果使用与 Analytic Server 上使用的相同用户和密码认证系统配置 SPSS Modeler 服务器，那么指定 N；例如，使用 Kerberos 认证时。否则，请指定 Y。

以批处理方式运行 SPSS Modeler 时，将 `-analytic_server_username {ASusername}`  
`-analytic_server_password {ASpassword}` 作为参数添加到 `clem` 命令。

### **as\_kerberos\_auth\_mode**

指定 Y 可以从 SPSS Modeler 启用 Kerberos SSO。

### **as\_kerberos\_krb5\_conf**

指定 Analytic Server 应该使用的 Kerberos 配置文件的路径；例如，`\etc\krb5.conf`。

### **as\_kerberos\_krb5\_spn**

指定 Analytic Server Kerberos SPN；例如，`HTTP/ashost.mydomain.com@MYDOMAIN.COM`。

- b. 重新启动 SPSS Modeler 服务器服务。

要连接到已启用 SSL/TLS 的 Analytic Server 安装，需要一些进一步的步骤来配置 SPSS Modeler 服务器和客户机安装。

- a. 浏览至 `http{s}://{HOST}:{PORT}/{CONTEXT-ROOT}/admin/{TENANT}` 并登录到 Analytic Server 控制台。
- b. 从浏览器下载证书文件并将其保存到您的文件系统。
- c. 将证书文件添加到 SPSS Modeler 服务器和 SPSS Modeler 客户机安装的 JRE。可以在 SPSS Modeler 安装路径的 `/jre/lib/security/cacerts` 子目录中找到要更新的位置。
  - 1) 请确保 `cacerts` 文件不是只读的。
  - 2) 使用 Modeler 随附的密钥工具程序 - 可以在 SPSS Modeler 安装路径的 `/jre/bin/keytool` 子目录中找到。

运行以下命令

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

请注意，`<as-alias>` 是 `cacerts` 文件的别名。您可以使用任何喜欢的名称，只要名称对于 `cacerts` 文件唯一。

因此，示例命令将如下所示。

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

- d. 重新启动 SPSS Modeler 服务器和 SPSS Modeler 客户机。
2. [可选] 如果您打算对含有 Analytic Server 数据源的流中的 R 模型进行评分，那么请安装 IBM SPSS Modeler - Essentials for R。IBM SPSS Modeler - Essentials for R 可供下载 (<https://www14.software.ibm.com/webapp/iwm/web/preLogin.do?source=swg-tspssp>)。



## 第 5 章 使用 SLM 标记来跟踪许可证发放

SLM 标记基于用于资源利用率测量的 ISO/IEC 19770-4 标准草案。SLM 标记为产品提供了报告其许可证耗用度量（与软件资产的使用有关的资源）的标准化能力。在产品中启用 SLM 后，会生成运行时 XML 文件来自我报告其许可证使用情况。

启动 Analytic Server 时，在 `<as_installation_path>/logs/slmtag` 文件夹中会创建 `slmtag` 文件。

由于存在两种许可证类型，将定期记录两种不同的度量：

- 对于当前的 Analytic Server 版本，许可证基于 Hadoop 集群中的数据节点总数（基于虚拟服务器）发放。节点数量记录在下列 `slmtag` 文件段中。

```
<Type>VIRTUAL_SERVER</Type>
  <SubType>Number of Data Nodes in Hadoop</SubType>
  <Value>2</Value>
  ...
```

- 对于 3.1 之前的 Analytic Server 版本，许可证基于 Hadoop 集群中的 HDFS 存储器大小（基于 RVU）发放。例如，存储器大小（单位为 TB）记录在下列 `slmtag` 文件段中。

```
<Type>RESOURCE_VALUE_UNIT</Type>
  <SubType>HDFS storage (Unit: Tega byte)</SubType>
  <Value>0.21</Value>
```

SLM 标记输出在线程中启动，它受 `SlmTagOutput.properties` 文件中定义的属性影响。该文件位于 `<as_installation_path>/configuration` 文件夹。

表 14. SLM 标记属性

属性	描述
<code>license.metric.logger.output.enabled</code>	控制 SLM 日志文件生成。缺省值为 <code>False</code> 。
<code>license.metric.logger.output.dir</code>	到存储 SLM 标记文件的目录的相对路径。缺省目录为 <code>&lt;as_installation_path&gt;/logs</code> 。
<code>license.metric.logger.output.SLMLogFrequency</code>	用于收集 SLM 日志的时间间隔（单位：毫秒）。
<code>license.metric.logger.file.size</code>	SML 标记文件最大大小，单位为字节。
<code>license.metric.logger.file.number</code>	一个软件身份实例的 SLM 标记文件最大数量。



---

## 第 6 章 故障诊断

本部分描述一些常见的安装和配置问题以及如何解决这些问题。

### 常见问题

**安装成功但带有警告，而用户无法创建数据源，错误为“无法完成请求。原因：权限被拒绝”**

将 `distrib.fs.root` 参数设置为 Analytic Server 用户（缺省情况下，`as_user`）无访问权的目录将导致错误。确定授权 Analytic Server 用户读写以及执行 `distrib.fs.root` 目录。

**Analytic Server 性能逐渐下降。**

Analytic Server 性能不符合预期时，请除去所有 `*.war` 文件，这些文件位于 Knox 服务的部署路径：`<KnoxServicePath>/data/deployments`。例如：`/usr/iop/4.1.0.0/knox/data/deployments`。

**卸载 Ambari 上的 Analytic Server 或 Essentials for R**

在某些情况下，卸载 Ambari 上的 Analytic Server 或 Essentials for R 时，卸载过程将挂起。当发生该问题时，必须手动停止 Ambari 服务器的进程标识。

**当 Analytic Server 安装在使用 OpenJDK 的 POWER System 上时的问题**

如果 Analytic Server 在使用 OpenJDK 的 POWER System 上运行，那么您必须手动执行下列配置步骤来确保坐标系 API 按期望工作

注：如果不使用坐标系 API，可以忽略该配置需求。

1. 在 Ambari 控制台中，导航至 **Analytic Server 服务 > 配置选项卡 > 高级 analytics-jvm-options**，然后在内容区域中添加下列行：

```
-XX:CompileCommand=exclude,com/esri/sde/sdk/pe/engine/PeHznTwoPointEquidistant$GCShorizon.*
```

2. 在 Ambari 控制台中，导航至定制 **analytics.cfg** 部分，然后添加下列 3 项配置：

**spark.executor.extraJavaOptions**

将值设置为：`-XX:CompileCommand=exclude,com/esri/sde/sdk/pe/engine/PeHznTwoPointEquidistant$GCShorizon.*`

**spark.driver.extraJavaOptions**

将值设置为：`-XX:CompileCommand=exclude,com/esri/sde/sdk/pe/engine/PeHznTwoPointEquidistant$GCShorizon.*`

**mapred.child.java.opts**

将值设置为：`-XX:CompileCommand=exclude,com/esri/sde/sdk/pe/engine/PeHznTwoPointEquidistant$GCShorizon.*`

**在 SuSE Linux 12 上安装 Analytic Server 时出错**

在 SuSE Linux 12 上安装 Analytic Server 时可能遇到以下错误：

```
Signature verification failed [4-Signatures public key is not available]
```

在 SuSE Linux 12 上安装 Analytic Server 之前，执行以下任务可解决此问题：

1. 从以下 URL 将公用密钥下载到主机：

```
https://ibm-open-platform.ibm.com/repos/IBM-SPSS-AnalyticServer/3.1.2.0/IBM-SPSS-ANALYTICSERVER-GPG-KEY.public
```

2. 通过在主机上运行以下命令，导入公用密钥：

```
rpm --import IBM-SPSS-ANALYTICSERVER-GPG-KEY.public
```

## 特定 Hadoop 发行版的问题

对于在 **Hortonworks 2.3-2.6** 上禁用的 **Analytic Server** 服务的刷新操作

要在 Hortonworks 2.3-2.6 上手动刷新 Analytic Server 库，请使用下列步骤。

1. 以 Analytic Server 用户（缺省情况下，as\_user）身份登录到运行 Analytic Metastore 的主机。

注：您可以从 Ambari 控制台找到此主机名。

2. 在 {AS\_ROOT}/bin 目录中运行 **refresh** 脚本；例如：

```
cd /opt/ibm/spss/analyticserver/3.1/bin
./refresh
```

3. 在 Ambari 控制台中重新启动 Analytic Server 服务。

从外部站点下载的包在 **Cloudera Manager** 中进行散列检查时失败。

散列验证错误将显示在包列表中。完成下载过程，然后通过 cloudera-scm-server 服务重新启动 Cloudera，便可以解决此问题。重新启动服务后，将不会发生此错误。

## HDFS 超组属性

如果 as\_user 不是以下 HDFS 组属性的成员，那么 Analytic Server 将记录启动期间的异常：**dfs.permissions.supergroup/dfs.permissions.superusergroup**。例如：

```
[11/15/17 7:32:35:510 PST] 000000bf SystemOut
0 2017-11-15 07:32:35,510 | : | | | | ERROR | slmtagoutput.SlmOutputAgent | SLM Logger => Error in performing callback function when calculating number
of nodes in kerberos environment: org.apache.hadoop.ipc.RemoteException(org.apache.hadoop.security.AccessControlException): Access denied for user as_user.
Superuser privilege is required
at org.apache.hadoop.hdfs.server.namenode.FSPermissionChecker.checkSuperuserPrivilege(FSPermissionChecker.java:93)
at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.checkSuperuserPrivilege(FSNamesystem.java:6606)
at org.apache.hadoop.hdfs.server.namenode.FSNamesystem.datanodeReport(FSNamesystem.java:5595)
at org.apache.hadoop.hdfs.server.namenode.NameNodeRpcServer.getDatanodeReport(NameNodeRpcServer.java:928)
at org.apache.hadoop.hdfs.server.namenode.AuthorizationProviderProxyClientProtocol.getDatanodeReport(AuthorizationProviderProxyClientProtocol.java:390)
at org.apache.hadoop.hdfs.protocolPB.ClientNameNodeProtocolServerSideTranslatorPB.getDatanodeReport(ClientNameNodeProtocolServerSideTranslatorPB.java:694)
at org.apache.hadoop.hdfs.protocol.proto.ClientNameNodeProtocolProtos$ClientNameNodeProtocol$2.callBlockingMethod(ClientNameNodeProtocolProtos.java)
at org.apache.hadoop.ipc.ProtobufRpcEngine$Server$ProtoBufRpcInvoker.call(ProtobufRpcEngine.java:617)
at org.apache.hadoop.ipc.RPC$Server.call(RPC.java:1073)
at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2141)
at org.apache.hadoop.ipc.Server$Handler$1.run(Server.java:2137)
at java.security.AccessController.doPrivileged(Native Method)
at javax.security.auth.Subject.doAs(Subject.java:415)
at org.apache.hadoop.security.UserGroupInformation.doAs(UserGroupInformation.java:1912)
at org.apache.hadoop.ipc.Server$Handler.run(Server.java:2135)
```

您必须手动将 as\_user 添加到在 hdfs-site 配置属性中定义的操作系统组：**dfs.permissions.supergroup/dfs.permissions.superusergroup**。

- 对于 Cloudera，缺省属性值是 **supergroup** 并且必须更改为实际存在的操作系统组。有关 Cloudera 中超组设置的信息，请参阅 Cloudera 文档。
- 对于 Ambari，缺省属性值为 **hdfs**。缺省情况下，在 Ambari 安装期间，Analytic Server 将 as\_user 添加到 HDFS 和 Hadoop 组。

在 Linux 上，使用 **usermod** 命令以将 as\_user 添加到 HDFS **superusergroup**（如果尚不存在）。

有关 HDFS 许可权的常规信息，请参阅 HDFS 许可权指南。

## 元数据存储库的问题

在运行 **add\_mysql\_user** 脚本时，操作 **CREATE USER** 失败

在运行 **add\_mysql\_user** 脚本之前，您首先需要从 MySQL 数据库中手动移除尝试添加的用户。您可以通过 MySQL Workbench UI 或 MySQL 命令移除这些用户。例如：

```
mysql -u root -e "DROP USER '$AEDB_USERNAME_VALUE'@'localhost';"
mysql -u root -e "DROP USER '$AEDB_USERNAME_VALUE'@'$METASTORE_HOST';"
mysql -u root -e "DROP USER '$AEDB_USERNAME_VALUE'@'%';"
```

在上述命令中，将 \$AEDB\_USERNAME\_VALUE 替换为您移除的用户名，将 \$METASTORE\_HOST 替换为您在其中安装数据库的主机名。

## Apache Spark 问题

### 在 Spark 进程中运行的流的问题

在 Spark 进程中强制运行 SPSS Modeler 流时失败。运行失败的 SPSS Modeler 流与 Analytic Server 源节点 (HDFS 文件) 一同构建, 且已将此源节点链接到 Sort 节点, 并设置为导出到其他 Analytic Server 数据源。运行该流后, 资源管理器用户界面将指示新应用程序正在运行, 但该流不会结束且保持 Running 状态。在 Analytic Server 日志、YARN 日志或 Spark 日志中没有指示流失败原因的消息。

在 Analytic Server 配置中, 将 spark.executor.memory 设置添加到定制的 analytics.cfg 文件中, 便可以解决此问题。将内存值设置为 4GB, 这样可以使先前失败的 SPSS Modeler 流在 2 分钟内完成 (在单节点集群环境中)。

### 无法使用 Cloudera 5.x 和 Spark 1.x 运行 Spark 作业

在将 Cloudera 5.x 与 Spark 1.x 一起使用时, 可能遇到以下异常:

```
org.apache.spark.SparkException: Exception when registering SparkListener
```

异常是 java.lang.ClassCastException: com.cloudera.spark.lineage.ClouderaNavigatorListener 无法强制类型转换 org.apache.spark.scheduler.SparkListener 所导致的。

为避免此异常, 必须在 **analyticserver-conf/config.properties** 的 **Analytic Server** 高级配置片段 (安全值) 区域中添加以下行。

```
spark.extraListeners=org.apache.spark.JavaSparkListener
```

在运行 **SparkML** 案例时遇到错误 **Exception during HdfsAuthcom.spss.utilities.i18n.LocException: Execution failed. Reason: com.spss.ae.filesystem.exception.FileSystemException: Unable to initialize the file system access.**

在 Spark 无法找到系列日志目录时生成此错误。此问题的变通方法是将 spark.lineage.log.dir 重定向到 /ae\_wlpserver/usr/servers/aeserver/logs/spark。

## 高可用性集群

由于依赖关系更改, 无法将 **Analytic Server** 添加到更多主机

使用第 25 页的『更新客户机依赖关系』中的指示信息运行 update\_clientdeps 脚本。

"分析集群服务与 **Zookeeper** 意外失去联系, 将终止此 **JVM** 以保证集群完整性。"

导致此问题的一个原因是要写入到 Zookeeper 的数据量是否过大。如果, 在 Zookeeper 日志中存在类似以下异常:

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

或者在 Analytic Server 日志中存在类似以下消息:

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

1. 在 Ambari 控制台中, 浏览至 Zookeeper 服务"配置"选项, 并将以下行添加到 env-template, 然后重新启动 Zookeeper 服务。

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

2. 在 Ambari 控制台中, 导航到 Analytic Server 服务的"配置"选项卡, 并在高级分析 JVM 选项中添加以下内容, 然后重新启动分析集群服务。

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

要为 jute.maxbuffer 设置指定的数字应大于异常消息中指示的数字。

## Zookeeper 事务数据无法进行管理

将 zoo.cfg 中的参数 **autopurge.purgeInterval** 设置为 1，以实现自动清除 Zookeeper 事务日志。

### 分析集群服务与 Zookeeper 失去联系

查看并修改 zoo.cfg 中的参数 **tickTime**、**initLimit** 和 **syncLimit**。例如：

```
# 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
```

请参阅 Zookeeper 文档以获取详细信息：<https://zookeeper.apache.org/doc/r3.3.3/zookeeperAdmin.html>

### Analytic Server 作业未恢复

常见情况下，Analytic Server 作业未恢复。

- Analytic Server 作业因集群成员失败而失败时，作业通常将在另一集群成员上自动重新启动。如果该作业未恢复，请检查以确保高可用性集群中至少有 4 个集群成员。

---

## 声明

本信息是为在美国国内供应的产品和服务而编写的。可以从 IBM 获取本资料的其他语言版本。但是，您必须拥有该语言的产品或产品版本副本，才能访问对应语言的资料。

IBM 可能在其他国家或地区不提供本文档中讨论的产品、服务或功能特性。有关您所在区域当前可获得的产品和服务的信息，请向您当地的 IBM 代表咨询。任何对 IBM 产品、程序或服务的引用并非意在明示或默示只能使用 IBM 的产品、程序或服务。只要不侵犯 IBM 的知识产权，任何同等功能的产品、程序或服务，都可以代替 IBM 产品、程序或服务。但是，评估和验证任何非 IBM 产品、程序或服务的操作，由用户自行负责。

IBM 可能已拥有或正在申请与本文档内容有关的各项专利。提供本文档并不意味着授予用户使用这些专利的任何许可。您可以用书面形式将许可查询寄往：

*IBM Director of Licensing*  
*IBM Corporation*  
*North Castle Drive, MD-NC119*  
*Armonk, NY 10504-1785*  
*US*

有关双字节 (DBCS) 信息的许可查询，请与您所在国家或地区的 IBM 知识产权部门联系，或用书面方式将查询寄往：

*Intellectual Property Licensing*  
*Legal and Intellectual Property Law*  
*IBM Japan Ltd.*  
*19-21, Nihonbashi-Hakozakicho, Chuo-ku*  
*Tokyo 103-8510, Japan*

International Business Machines Corporation"按现状"提供本出版物，不附有任何种类的（无论是明示的还是默示的）保证，包括但不限于默示的有关不侵权、适销和适用于某种特定用途的保证。某些管辖区域在某些交易中不允许免除明示或默示的保证。因此本条款可能不适用于您。

本信息可能包含技术方面不够准确的地方或印刷错误。本信息将定期更改；这些更改将编入本信息的新版本中。IBM 可以随时对本出版物中描述的产品和/或程序进行改进和/或更改，而不另行通知。

本信息中对任何非 IBM Web 站点的引用都只是为了方便起见才提供的，不以任何方式充当对那些 Web 站点的保证。那些 Web 站点中的资料不是 IBM 产品资料的一部分，使用那些 Web 站点带来的风险将由您自行承担。

IBM 可以按它认为适当的任何方式使用或分发您所提供的任何信息而无须对您承担任何责任。

本程序的被许可方如果要了解有关程序的信息以达到如下目的：(i) 使其能够在独立创建的程序和其它程序（包括本程序）之间进行信息交换，以及 (ii) 使其能够对已经交换的信息进行相互使用，请与下列地址联系：

*IBM Director of Licensing*  
*IBM Corporation*

North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
US

只要遵守适当的条件和条款，包括某些情形下的一定数量的付费，都可获得这方面的信息。

本文档中描述的许可程序及其所有可用的许可资料均由 IBM 依据 IBM 客户协议、IBM 国际程序许可协议或任何同等协议中的条款提供。

此处引用的性能数据和客户示例仅用于描述目的。实际性能可能因特定配置和操作条件而异。

涉及非 IBM 产品的信息可从这些产品的供应商、其出版说明或其他可公开获得的资料中获取。IBM 没有对这些产品进行测试，也无法确认其性能的精确性、兼容性或任何其他关于非 IBM 产品的声明。有关非 IBM 产品性能的问题应当向这些产品的供应商提出。

关于 IBM 未来方向或意向的声明都可随时更改或收回，而不另行通知，它们仅仅表示了目标和意愿而已。

所有 IBM 的价格均是 IBM 当前的建议零售价，可随时更改而不另行通知。经销商的价格可与此不同。

本信息仅用于规划的目的。在所描述的产品上市之前，此处的信息会有更改。

本信息包含在日常业务操作中使用的数据和报告的示例。为了尽可能完整地说明这些示例，示例中可能会包括个人、公司、品牌和产品的名称。所有这些名称都是虚构的，若实际人员或企业与此相似，纯属巧合。

版权许可证：

本信息包含在日常业务操作中使用的数据和报告的示例。为了尽可能完整地说明这些示例，示例中可能会包括个人、公司、品牌和产品的名称。所有这些名称都是虚构的，若实际人员或企业与此相似，纯属巧合。

凡这些实例程序的每份拷贝或其任何部分或任何衍生产品，都必须包括如下版权声明：

© (贵公司的名称) (年)。此部分代码是根据 IBM Corp. 公司的样本程序衍生出来的。

© Copyright IBM Corp. (输入年份)。All rights reserved.

---

## 商标

IBM、IBM 徽标和 [ibm.com](http://ibm.com) 是 International Business Machines Corp.，在全球许多管辖区域的商标或注册商标。其他产品和服务名称可能是 IBM 或其他公司的商标。最新的 IBM 商标列表可以在 Web 上的 "Copyright and trademark information" 中获取，地址为：[www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml)。

Adobe、Adobe 徽标、PostScript 以及 PostScript 徽标是 Adobe Systems Incorporated 在美国和/或其他国家或地区的注册商标或商标。

IT Infrastructure Library 是 Central Computer and Telecommunications Agency 的注册商标，该企业现已成为 Office of Government Commerce 的一部分。

Intel、Intel 徽标、Intel Inside、Intel Inside 徽标、Intel Centrino、Intel Centrino 徽标、Celeron、Intel Xeon、Intel SpeedStep、Itanium 和 Pentium 是 Intel Corporation 或其子公司在美国和其他国家或地区的商标或注册商标。

Linux 是 Linus Torvalds 在美国和@3B72其他国家或地区的注册商标。



Microsoft、Windows、Windows NT 以及 Windows 徽标是 Microsoft Corporation 在美国和/或其他国家或地区的商标。

ITIL 是一个注册商标，是 Minister for the Cabinet Office 的共同体注册商标，并且已在 U.S. Patent and Trademark Office 进行注册。

UNIX 是 The Open Group 在美国和/或其他国家或地区的注册商标。

Cell Broadband Engine 是 of Sony Computer Entertainment, Inc. 在美国和/或其他国家或地区的商标并且在当地许可证下使用。

Linear Tape-Open、LTO、LTO 徽标、Ultrium 和 Ultrium 徽标是 HP、IBM Corp 和 Quantum 在美国和其他国家或地区的商标。



Printed in China