

**IBM SPSS Analytic Server  
V2.1**

**管理员指南**

**IBM**

**注意**

在使用本资料及其支持的产品之前，请先阅读第 11 页的『声明』中的信息。

**产品信息**

此版本适用于 IBM SPSS Analytic Server 的 V2.1.1.0 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 概述</b> . . . . .	<b>1</b>	版本信息 . . . . .	9
体系结构 . . . . .	2	日志收集器 . . . . .	9
V2.1 中面向管理员的新增内容 . . . . .	2	常见问题 . . . . .	10
<b>第 2 章 租户管理</b> . . . . .	<b>3</b>	<b>声明</b> . . . . .	<b>11</b>
命名规则 . . . . .	4	商标 . . . . .	12
<b>第 3 章 用户入门</b> . . . . .	<b>5</b>		
<b>第 4 章 Analytic Server 作业名</b> . . . . .	<b>7</b>		
<b>第 5 章 故障诊断</b> . . . . .	<b>9</b>		
日志记录 . . . . .	9		



---

## 第 1 章 概述

IBM® SPSS® Analytic Server 是一套用于大型数据分析的解决方案，融合了带有大型数据系统的 IBM SPSS 技术，允许您使用熟悉的 IBM SPSS 用户界面来解决问题，而这是以前无法做到的。

### 为什么大型数据分析至关重要

各组织所收集的数据量正成倍增长；例如，金融和零售业保留着一年的（或两年、甚至十年）内的所有客户交易数据，电信服务提供商保留着呼叫数据记录（CDR）和设备传感器读数，互联网公司保留着网络检索结果。

在以下情况下都需要大型数据分析：

- 存在大量数据（TB、PB、EB），尤其是结构化和非结构化数据混合时
- 存在迅速变化/累积的数据

在以下情况下同样存在大型数据分析：

- 构建大量（数以千计）模型
- 频繁构建/刷新模型

### 挑战

收集大量数据的相同组织在利用这些数据时实际上经常都会遇到困难，有以下原因：

- 传统分析产品的体系结构不适用于分布式计算，并且
- 现有的统计算法并不是设计来处理大型数据的（这些算法期望数据向它们移动，但是移动大型数据成本过高），因此
- 执行大型数据分析的当前发展状况亟需大型数据系统的新技能和深入的认识。很少有分析人员具有这些技能。
- 内存解决方案能够解决中型问题，但对于真正的大型数据，其扩展性不是很好。

### 解决方案

Analytic Server 提供以下功能：

- 可重用大型数据系统的以数据为中心的体系结构，例如，数据位于 HDFS 的 Hadoop Map/Reduce。
- 定义的接口可合并新的统计算法（设计为流向数据）。
- 熟悉的 IBM SPSS 用户接口隐藏了大型数据环境的详细信息，以便分析人员能够集中于分析数据。
- 可随意扩展以解决任何大小的问题的解决方案。

## 体系结构

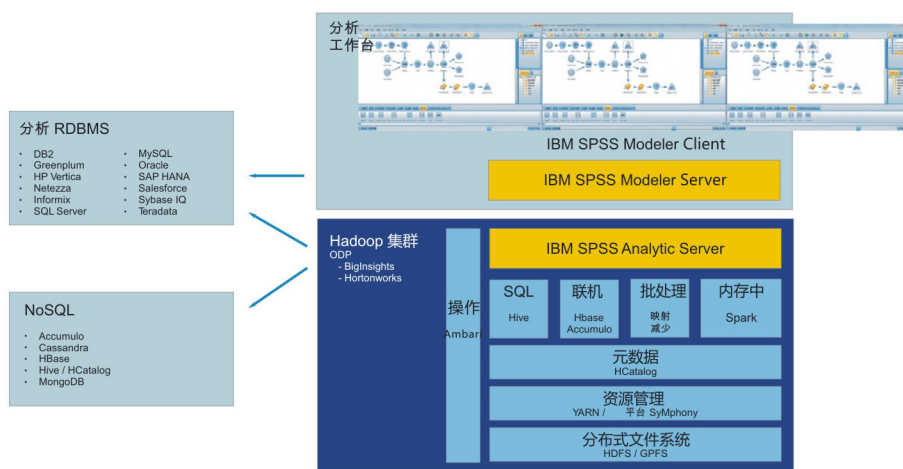


图 1. 体系结构

Analytic Server 位于客户机应用程序和 Hadoop 云之间。假设数据存在于云中，那么使用 Analytic Server 的方法概述如下：

1. 在云中的数据上定义 Analytic Server 数据源。
2. 定义要在客户机应用程序中执行的分析。对于当前发行版，客户机应用程序是 IBM SPSS Modeler。
3. 当您运行分析时，客户机应用程序将提交一个 Analytic Server 执行请求。
4. Analytic Server 会对该作业进行编排以在 Hadoop 云中运行，然后将结果报告给客户机应用程序。
5. 您可以使用这些结果来定义更进一步的分析，如此重复循环。

## V2.1 中面向管理员的新增内容

### Analytic Server 控制台

#### 读者角色

在租户中，您可以对无法登录 Analytic Server 但是可以通过 Modeler 中的 Analytic Server 源节点读取 Analytic Server 数据源的用户和组指定“读者”角色。

---

## 第 2 章 租户管理

租户提供用户、项目和数据源的高层次划分，以使租户之间不能共享对象。每个用户以为其分配的租户身份访问系统。

您可以在 Analytic Server 控制台中管理租户和将用户分配至租户。“租户”页面的视图取决于登录到控制台的用户的角色：

- 在安装过程中设置的“超级用户”管理员是租户管理员。只有这个用户能够创建新租户和编辑任何租户的属性。
- 具有“管理员”角色的用户能够编辑其登录的租户的属性。
- 具有“用户”角色的用户无法编辑租户属性。“租户”页面对于这些用户是隐藏的。
- 具有“读者”角色的用户无法编辑数据源，甚至无法登录 Analytic Server 控制台。

管理员能够访问“项目”和“数据源”页面，并且能够管理所有项目和数据源以进行清除和管理。请参阅《IBM SPSS Analytic Server 用户指南》以获取更多信息。

### 租户列表

“租户”主页面可显示表中现有租户。只有“超级用户”管理员才能在此页面上进行编辑。

- 单击租户名称以显示其详细信息并编辑其属性。
- 单击租户的 URL 以在该租户的上下文中打开控制台。

**注：**您将从控制台注销，并且将需要使用针对该租户有效的凭证来登录。

- 在搜索区域输入以过滤列表，只显示名称中包含搜索字符串的租户。
- 单击**新建**以使用在**添加新租户**对话框中指定的名称创建新租户。请参阅第 4 页的『命名规则』，以了解可为租户提供的名称的限制。
- 单击**删除**以除去所选租户。
- 单击**刷新**以更新列表。

### 个人租户详细信息

内容区域划分为多个可折叠部分。

#### 详细信息

<b>名称</b>	可编辑的文本字段，显示租户名称。
<b>描述</b>	可编辑的文本字段，您可以提供关于该租户的说明性文本。
<b>URL</b>	这是将提供给用户通过 Analytic Server 控制台登录租户的 URL，还可以用于配置 SPSS Modeler 服务器。请参阅《IBM SPSS Analytic Server 安装与配置指南》，以获取有关配置 SPSS Modeler 的详细信息。
<b>状态</b>	<b>活动</b> 租户目前正在使用中。使租户 <b>处于不活动状态</b> 可防止用户登录该租户，但是不会删除任何底层信息。

#### 主体

主体是从安装过程中设置的安全提供程序获取的用户和组。您可以向租户添加“管理员”或“用户”角色的主体。

- 在文本框中输入可按名称中的搜索字符串对用户和组进行过滤。 从下拉列表中选择**管理员**、**用户**或**读者**可以分配用户在该租户中的角色。 单击**添加参与者**按钮可将他们添加到作者列表。
- 要移除参与者，可以选择成员列表中的用户或组，然后单击**移除参与者**按钮。

**度量** 允许您为租户配置资源限制。 报告租户当前使用的磁盘空间。

- 您可以为租户设置最大磁盘空间配额；当达到此限制时，在此租户上无法再将数据写入磁盘，直至清除足够的磁盘空间使租户磁盘空间使用量达到配额以下为止。
- 您可以为租户设置磁盘空间警告级别；当超出此配额时，主体无法在此租户上提交任何分析作业，直至清除足够的磁盘空间使租户磁盘空间使用量达到配额以下为止。
- 您可以为此租户设置可同时运行的并行作业的最大数量；当超出此配额时，主体无法在此租户上提交任何分析作业，直至当前正在运行的作业完成为止。
- 您可以设置数据源能够具有的最大字段数。 每当创建或更新数据源时，都会检查该限制。
- 您可以设置最大文件大小（以兆字节为单位）。 上传文件时，会检查该限制。

### 安全提供程序配置

允许您指定用户认证提供程序。 **缺省值**使用在安装与配置期间设置的缺省租户的提供程序。 **LDAP** 允许您使用外部 LDAP 服务器（例如，Active Directory 或 OpenLDAP）对用户进行认证。 为提供程序指定设置并（可选）指定过滤设置以控制“主体”部分中可用的用户和组。

---

## 命名规则

对于在 Analytic Server 被赋予唯一名称的一切（例如，数据源和项目），以下规则适用于这些名称。

- 在单个租户中，名称必须在同类型的对象中唯一。例如，两个数据源不能同时命名为 insuranceClaims，但是一个数据源和一个项目可分别命名为 insuranceClaims。
- 名称区分大小写。例如，insuranceClaims 和 InsuranceClaims 均被视为唯一名称。
- 名称忽略前置和后置空格。
- 以下字符在名称中无效。

~, #, %, &, \*, {, }, \, :, <, >, ?, /, |, ", \t, \r, \n



---

## 第 3 章 用户入门

告知用户浏览至 `http://<host>:<port>/<context-root>/admin/<tenant>`，并输入其用户名和密码以登录到 Analytic Server 控制台。

### <host>

Analytic Server 主机的地址。

### <port>

Analytic Server 正在侦听的端口号。缺省情况下，此端口号为 8080。

### <context-root>

Analytic Server 的上下文根。缺省情况下为 `analyticserver`。

### <tenant>

在多租户环境中，这是您所属的租户。在单租户环境中，缺省租户为 **ibm**。

例如，如果主机的 IP 地址为 9.86.44.232，您已创建“mycompany”租户并向其添加了用户，并且其他设置已保留为缺省值，那么用户应该导航至 `http://9.86.44.232:8080/analyticserver/admin/mycompany` 以访问 Analytic Server 控制台。



---

## 第 4 章 Analytic Server 作业名

Analytic Server 可生成 map-reduce 作业，可通过 Hadoop 的集群的作业跟踪器来监控这些作业。

map-reduce 作业名采用以下结构。

AS/{tenant name}/{user name}/{algorithm name}

### **{tenant name}**

这是在其中运行作业的租户的名称。

### **{user name}**

这是请求该作业的用户。

### **{algorithm name}**

这是作业中的主要算法。 请注意，单个流可能生成多个 map-reduce 作业；类似地，单个流中的多个操作可包含在单个 map-reduce 作业中。



---

## 第 5 章 故障诊断

Analytic Server 提供了多种实用工具用于问题确定。

---

### 日志记录

Analytic Server 会创建有助于诊断问题的客户日志文件和跟踪文件。通过缺省 Liberty 安装，可以在 {AS\_ROOT}/ae\_wlpserver/usr/servers/aeserver/logs 目录中找到日志文件。

缺省日志记录配置将生成两个每天轮换的日志文件。

**as.log** 此文件包含参考警告和错误消息的高级别摘要。发生服务器错误时如果无法通过使用用户界面中显示的错误消息来解决，请首先检查此文件。

#### **as\_trace.log**

此文件包含来自 ae.log 的所有条目，但添加了额外信息，这些信息主要面向 IBM 支持和开发人员，用于调试目的。

Analytic Server 使用 Apache LOG4J 作为其底层的日志记录工具。通过使用 LOG4J，可通过编辑 {AS\_SERVER\_ROOT}/configuration/log4j.xml 配置文件来动态调整日志记录。支持人员可能要求您执行此项操作以帮助诊断问题，或者您可能希望修改此文件以限制保留的日志文件数量。对文件进行的更改会在数秒钟内自动完成检测，因此无需重新启动 Analytic Server。

有关 log4j 和配置文件的更多信息，请参阅位于以下地址的 Apache 官方 Web 站点上的文档：<http://logging.apache.org/log4j/>。

---

### 版本信息

您可以通过检查 {AS\_ROOT}/properties/version 文件夹确定所安装的 Analytic Server 版本。以下文件包含版本信息。

#### **IBM\_SPSS\_Analytic\_Server-\*.swtag**

包含详细的产品信息。

#### **version.txt**

已安装的产品版本和构建号。

---

### 日志收集器

无法通过直接复查日志文件来解决问题时，可以将所有日志捆绑在一起并发送给 IBM 支持人员。其中提供了一个实用程序来简化所有必要数据的收集。

通过使用命令 shell，运行以下命令：

```
cd {AS_ROOT}/bin
run >sh ./logcollector.sh
```

这些命令会在 {AS\_ROOT}/bin 下创建一个压缩文件。该压缩文件包含所有日志文件和产品版本信息。

---

## 常见问题

本节描述了一些常见管理问题以及它们的解决方法。

### 安全性

#### 尝试访问 HCatalog 数据源时 Kerberos 认证失败

如果您在日志中看到类似于以下内容的错误:

```
cause:javax.security.sasl.SaslException: Failure to initialize security context  
com.spss.analyticframework.api.exceptions.ComponentException: Cannot access HCatalog
```

那么必须确保 HDFS 用户的 Kerberos TGT 已高速缓存并且在 Analytic Server 服务器主机上可用。为此:

1. 停止 Analytic Server 进程。
2. 从 Analytic Server 主机运行 `kinit -f $hdfs.user`, 其中 `$hdfs.user` 是在 `config.properties` 文件中定义的, 并且对分析 root 用户具有写许可权
3. 启动 Analytic Server。

### Analytic Server 控制台

#### 在 iOS 上通过 Safari 访问 Analytic Server 控制台

租户状态下拉列表无法正常工作, 并且您无法更新基于文件的数据源的数据模型。执行操作时, 请使用另一个浏览器。

### 运行中的流

#### PySpark 作业运行失败

1. 在 Ambari 控制台中, 将 `export SPARK_HOME=/usr/iop/current/spark-client` 添加到 YARN 服务的 `yarn-env` 参数中。
2. 重新启动 YARN 服务。

---

## 声明

本信息是为在美国国内供应的产品和服务而编写的。

IBM 可能在其他国家或地区不提供本文中讨论的产品、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暗示只能使用 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  
Armonk, NY 10504-1785  
U.S.A.

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.  
1623-14, Shimotsuruma, Yamato-shi  
Kanagawa 242-8502 Japan

以下段落对于英国和与当地法律有不同规定的其他国家或地区均不适用：INTERNATIONAL BUSINESS MACHINES CORPORATION“按现状”提供本出版物，不附有任何种类的（无论是明示的还是暗含的）保证，包括但不限于暗含的有关非侵权、适销和适用于某特定用途的保证。某些国家或地区在某些交易中不允许免除明示或暗含的保证。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 Software Group  
ATTN: Licensing

200 W. Madison St.  
Chicago, IL; 60606  
U.S.A.

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。

此处包含的任何性能数据都是在受控环境中测得的。因此，在其他操作环境中获得的数据可能会有明显的不同。有些测量可能是在开发级的系统上进行的，因此不保证与一般可用系统上进行的测量结果相同。此外，有些测量是通过推算而估计的，实际结果可能会有差异。本文档的用户应当验证其特定环境的适用数据。

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。

所有关于 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。所有这些名字都是虚构的，若现实生活中实际业务企业使用的名字和地址与此相似，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：

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。所有这些名字都是虚构的，若现实生活中实际业务企业使用的名字和地址与此相似，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.

如果您正在查看本信息的软拷贝，图片和彩色图例可能无法显示。

---

## 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。在 Web 站点 [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml) 上的“Copyright and trademark information”部分中提供了 IBM 商标的最新列表。

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