

**IBM SPSS Analytic Server
V1**

用户指南

IBM

注释

在使用本信息及其支持的产品之前，请先阅读第 17 页的『声明』中的信息。

产品信息

此版本适用于 IBM SPSS Analytic Server V1.0.0 及所有后续发行版和修订版，直至在新版本中另有说明为止。

目录

第 1 章 概述	1	数据源	9
体系结构	2	设置 (文件数据源)	12
第 2 章 SPSS Modeler 集成	3	预览和元数据 (数据源)	13
Analytic Server 源	3	项目	14
Analytic Server 选择数据源	3	声明	17
Analytic Server 导出	3	商标	18
Analytic Server 流属性	4	索引	21
受支持的节点	4		
第 3 章 Analytic Server 控制台	9		

第 1 章 概述

IBM® SPSS® Analytic Server 是大型数据分析的解决方案，它将 IBM SPSS 技术与大型数据系统结合，使您能够使用熟悉的 IBM SPSS 用户界面来解决先前不能解决的问题。

为什么大型数据分析非常重要

组织收集的数据量正在迅速增长；例如，金融和零售业务拥有一年（或两年，或十年）的所有客户交易，电信提供商拥有呼叫数据记录（CDR）和设备传感器读数，因特网公司拥有网络搜寻的结果。

当存在以下情况时，需要大型数据分析：

- 大量数据（太字节、拍字节、艾字节），特别是当结构化数据和非结构化数据同时存在时
- 迅速变化/累积的数据

当存在以下情况时，大型数据分析也非常有帮助：

- 正在构建大量（成千上万的）模型
- 频繁构建/刷新模型

挑战

由于以下各种原因，收集大量数据的相同组织在使用数据方面实际上经常遇到困难：

- 传统分析产品的体系结构不适合分布式计算
- 现有统计算法未设计用于使用大型数据（这些算法需要移入数据，但移动大型数据的成本太大）
- 因此，对大型数据执行最先进的分析需要大型数据系统的新技能和详细知识。而具有这些技能的分析人员却非常少。
- 内存解决方案用于解决中型问题，不能很好地解决真正大型的数据。

解决方案

Analytic Server 提供：

- 利用大型数据系统的以数据为中心的体系结构，如 Hadoop Map/减少 HDFS 中的数据。
- 已定义的界面，以将设计用于转到数据的新统计算法合并在一起。
- 熟悉的 IBM SPSS 用户界面，隐藏大型数据环境的详细信息以便分析人员能够将注意力放在分析数据上。
- 可解决任何规模问题的解决方案。

体系结构

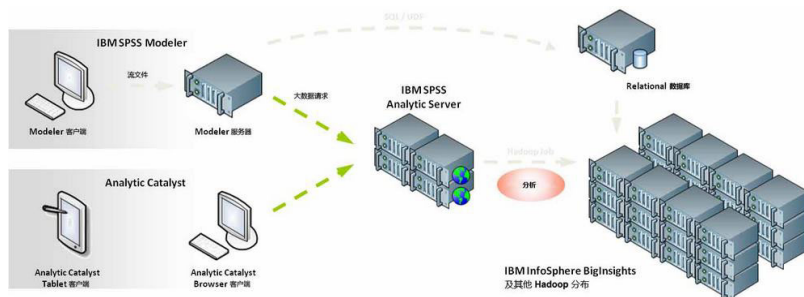


图 1. 体系结构

Analytic Server 位于客户机应用程序和 Hadoop 云之间。假设数据驻留在云中，使用 Analytic Server 的大纲将执行以下操作：

1. 为云中的数据定义 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，客户机应用程序为 IBM SPSS Modeler 和 IBM SPSS Analytic Catalyst。
3. 运行分析时，客户机应用程序提交 Analytic Server 执行请求。
4. Analytic Server 把将在 Hadoop 云中运行的作业协调地结合起来并向客户机应用程序报告结果。
5. 您可以使用这些结果来定义进一步的分析，且循环重复。

第 2 章 SPSS Modeler 集成

SPSS Modeler 是一种具有可视化分析方法的数据挖掘工作台。用画布上的节点表示作业中的每个不同操作（从访问数据源到合并记录、以及写入新文件或构建模型）。将这些操作链接起来形成一个分析流。

要构造一个可以在 HDFS 中运行的 SPSS Modeler 流，请以『Analytic Server 源』节点开始。SPSS Modeler 会将尽可能多的流推送回 Analytic Server，然后，如果需要，将 HDFS 中的记录子集降级，从而在 SPSS Modeler 服务器上完成执行“本地”流。您可以在第 4 页的『Analytic Server 流属性』中设置 SPSS Modeler 将下载的最大记录数。

如果分析以写回 HDFS 的记录结束，请以『Analytic Server 导出』节点完成流。

Analytic Server 源

Analytic Server 源允许您运行 Hadoop 分布式文件系统 (HDFS) 上的流。Analytic Server 数据源中的信息可来自各种地方，包括：

- HDFS 上的文本文件
- 数据库
- HCatalog

通常，会在 HDFS 上执行具有“Analytic Server 源”的流；但是，如果流含有不支持在 HDFS 上执行的节点，那么会将尽可能多的流“推送回”Analytic Server，然后 SPSS Modeler Server 会尝试处理该流的剩余部分。您需要对非常大的数据集做两次采样；例如，通过在流中放置“样本”节点的方式。

数据源。假设 SPSS Modeler Server 管理员已经建立了连接，那么您可以选择含有要使用的数据的数据源。数据源包含与该源关联的文件和元数据。单击**选择**可显示一个可用数据源的列表。请参阅『Analytic Server 选择数据源』主题以获取更多信息。

如果需要创建新的数据源或编辑现有数据源，请单击**启动数据源编辑器...**。请参阅第 9 页的『数据源』，以获取创建和编辑数据源的详细信息。

Analytic Server 选择数据源

“数据源”表显示了一个可用数据源的列表。选择要使用的源，然后单击**确定**。

单击**显示所有者**可显示数据源所有者。

过滤依据允许您对含有**关键字**的数据源列表进行过滤，可根据数据源名称和数据源描述或**所有者**检查过滤条件。您可以输入字符串组合、数字或通配符 (%) 作为过滤条件。搜索字符串是区分大小写的。单击**刷新**可更新“数据源”表。

Analytic Server 导出

“Analytic Server 导出”允许您将数据从分析写入现有 Analytic Server 数据源。这些数据可以是 Hadoop 分布式文件系统 (HDFS) 或数据库中的文本文件。

通常，具有“Analytic Server 导出”节点的流也可以“Analytic Server 源”节点开始，提交到 Analytic Server，并在 HDFS 上执行。或者，为了上载相对较小的数据集（不超过 100,000 个记录）以与 Analytic Server 配合使用，具有“本地”数据源的流可以“Analytic Server 导出”节点结束。

数据源。选择含有要使用的数据的数据源。数据源包含与该源关联的文件和元数据。单击**选择**可显示一个可用数据源的列表。请参阅第 3 页的『Analytic Server 选择数据源』主题以获取更多信息。

如果需要创建新的数据源或编辑现有数据源，请单击**启动数据源编辑器...**。请参阅第 9 页的『数据源』，以获取创建和编辑数据源的详细信息。

方式。选择**附加**可添加到现有数据源，或选择**覆写**可替换数据源的内容。

为此数据生成导入节点。在导出到指定数据源时，选择为此数据生成源节点。将此节点添加到数据流区域。

Analytic Server 流属性

这些设置提供了使用 Analytic Server 的若干选项。

在 Analytic Server 之外进行处理的最大记录数

指定要从 Analytic Server 数据源导入到 SPSS Modeler 服务器的最大记录数。

当无法在 Analytic Server 中处理节点时进行通知

此设置确定了将要提交给 Analytic Server 的流中包含无法在 Analytic Server 中进行处理的节点时将发生什么。指定是发出警告并继续处理流，还是抛出错误并停止处理。

拆分模型存储设置

当超过模型大小 (MB) 时，通过引用在 Analytic Server 上存储拆分模型

模型块通常存储在流中。含许多拆分的拆分模型会生成大型块，在流和 Analytic Server 之间来回移动块会影响性能。解决方案是，如果拆分模型超过了指定大小，那么将其存储在 Analytic Server 上，SPSS Modeler 中的块含有对该模型的引用。

执行完成后，通过引用在 Analytic Server 上存储模型的缺省文件夹

指定要在 Analytic Server 上存储拆分模型的缺省路径。该路径必须以有效的 Analytic Server 项目名称开头。

存储升级模型的文件夹

指定要存储“升级”模型的缺省路径。SPSS Modeler 会话结束后，不会清除升级模型。

受支持的节点

HDFS 上支持执行许多 SPSS Modeler 节点，但是某些节点的执行可能有所不同，另有一些节点目前尚不支持。本主题详细描述了目前的支持级别。

概述

- Analytic Server 将不接受某些在带引号的建模器字段名称中通常可接受的字符。
- 对于要在 Analytic Server 中运行的“建模器”流，它必须以一个或多个 Analytic Server 源节点开始，并以单个“建模”节点或 Analytic Server “导出”节点结束；允许合并，但不允许派生。
- 建议将连续目标的存储设置为实数而非整数。评分模型始终将实数值写入连续目标的输出数据文件，而评分的输出数据模型会跟在目标存储后面。因此，如果连续目标是整数存储，那么写入的值和评分的数据模型将不匹配，这种情况会在尝试读取评分数据时造成错误。

源

- 将在本地运行以 Analytic Server 源节点之外的任何节点开始的流。

记录操作

支持所有的“记录”操作。以下是有关这些节点功能的进一步注释。

选择

- 支持 Derive 节点所支持的一组函数（见下文）。

样本

- 不支持块级采样。
- 不支持“复合抽样”方法。

聚合

- 不支持相连的键。
- 不支持顺序统计（中间值、第一个四分位值和第三个四分位值）。

排序

- 不支持“优化”选项卡。

合并

- 不支持按顺序合并。
- 不支持按条件合并。
- 不支持“优化”选项卡。
- 当前不支持在 Analytic Server “源”节点和“合并”节点之间放置“样本”节点或模型块。通常可以指定“选择”节点来取代“样本”节点的功能。
- Analytic Server 不会连接空字符串键；即，如果正在合并的某个键包含空字符串，那么包含该空字符串的任何记录将会从已合并的输出中删除。
- 合并操作相对较慢。如果 HDFS 中有闲置的空间，相比在各个流中合并数据源，先合并一次数据源，然后在后面的流中使用已合并的源，后者的操作速度将快得多。

字段操作

支持自动数据准备、类型、过滤、衍生、整体、填料、重新分类、装箱、RFM 分析、分区、SetToFlag、重构和字段重新排序节点。以下是有关这些节点功能的进一步注释。

自动数据准备

- 不支持培训节点。支持将受培训的自动数据准备节点中的转换应用于新的数据。

类型

- 不支持检查列。
- 不支持“格式”选项卡。

衍生

- 支持所有的衍生函数，序列函数除外。
- 无法在将 Split 字段用作拆分的同一流中派生 Split 字段；需要创建两个流；一个用于派生 Split 字段，另一个用于将该字段用作拆分。
- 在进行对比时，标志字段不能单独使用；即，if (flagField) then ... endif 这种用法会导致错误；变通方法是使用 if (flagField=trueValue) then ... endifendif will cause an error; the work-around is to use if (flagField=trueValue) then ... endif
- 为了与“建模器”中的结果相匹配，建议在使用 ** 运算符时，将指数指定为实数（如 x**2.0，而非 x**2）。

填料

- 支持衍生节点所支持的一组函数（见上文）。

RFM 分析

- 不支持用于处理间距的“Keep in current”选项。RFM 近度 (recency)、频度 (frequency)和值度 (monetary) 评分与“建模器”用同一数据计算出的评分不一定始终匹配。评分范围相同，但是评分赋值 (bin 号) 可能不同。

图形 支持所有的“图形”节点。

建模 支持数量有限的“建模”节点，包括 Linear、Neural Net、C&RT、Chaid 和 Quest。以下是有关这些节点功能的进一步注释。

Linear

- 不支持对现有 PSM 模型进行连续性培训。
- 仅当定义了 split 字段时才建议使用“标准”模型构建目标，从而每个拆分中的记录数量不会太大。其中“太大”的定义取决于 Hadoop 集群中个别节点的能力。相比之下，您也需要注意确保勿将拆分定义得过于精确而使用于构建模型的记录过少。
- 不支持 Boosting 目标。
- 不支持 Bagging 目标。
- 如果记录很少，那么不建议使用“非常”大的数据集目标；它通常不会构建模型或不会构建降级的模型。如果输入记录以某种违反了随机性假定（位于所用的整体模型构建算法后面）的系统方法进行排序，可能也会出现问题。
- 不支持 Automatic Data Preparation。尝试用含有许多缺失值的数据构建模型时，可能会出现这个问题；通常，缺失值会插补到 Automatic Data Preparation 中。变通方法是使用树模型或具有“高级”设置的神经网络来插补所选的缺失值。
- 未计算拆分模型的准确性统计。

Neural Net

- 不支持对现有的标准模型或 PSM 模型进行连续性培训。
- 仅当定义了 split 字段时才建议使用“标准”模型构建目标，从而每个拆分中的记录数量不会太大。其中“太大”的定义取决于 Hadoop 集群中个别节点的能力。相比之下，您也需要注意确保勿将拆分定义得过于精确而使用于构建模型的记录过少。
- 不支持 Boosting 目标。
- 不支持 Bagging 目标。
- 如果记录很少，那么不建议使用“非常”大的数据集目标；它通常不会构建模型或不会构建降级的模型。如果输入记录以某种违反了随机性假定（位于所用的整体模型构建算法后面）的系统方法进行排序，可能也会出现问题。
- 如果数据中存在许多缺失值，请使用“高级”设置来插补这些缺失值。
- 未计算拆分模型的准确性统计。

C&R Tree、CHAID 和 Quest

- 不支持对现有 PSM 模型进行连续性培训。
- 仅当定义了 split 字段时才建议使用“标准”模型构建目标，从而每个拆分中的记录数量不会太大。其中“太大”的定义取决于 Hadoop 集群中个别节点的能力。相比之下，您也需要注意确保勿将拆分定义得过于精确而使用于构建模型的记录过少。
- 不支持 Boosting 目标。
- 不支持 Bagging 目标。

- 如果记录很少，那么不建议使用“非常”大的数据集目标；它通常不会构建模型或不会构建降级的模型。如果输入记录以某种违反了随机性假定（位于所用的整体模型构建算法后面）的系统方法进行排序，可能也会出现问题。
- 不支持交互式会话。
- 未计算拆分模型的准确性统计。

模型评分

以下模型块支持评分：Linear、Neural Net、C&RT、Chaid 和 Quest、Logistic、Regression、Genlin、TwoStep Cluster、C5、Bayesian Networks、SVM、R 和 Text Mining。

- 不会对原始倾向或已调整的倾向进行评分。变通方法是使用含以下表达式的 Derive 节点，手动计算原始倾向以获取相同效应：if 'predicted-value' == 'value-of-interest' then 'prob-of-that-value' else 1-'prob-of-that-value' endif
- 对模型进行评分时，Analytic Server 不会检查模型中所用的所有字段是否都位于数据集中，因此模型在 Analytic Server 中运行之前，请先确保模型中所用的所有字段都位于数据集中。

输出 支持“矩阵”节点、“分析”节点、“数据审计”节点、“转换”节点、“统计”节点和“平均值”节点。

导出 流可以 Analytic Server 源节点开始，并以导出节点而非 Analytic Server 导出节点结束，但是数据会从 HDFS 移动到 SPSS Modeler Server，并最终移动到导出位置。

第 3 章 Analytic Server 控制台

Analytic Server 提供了用于管理数据源和项目的瘦客户机接口，可通过以下两种方式进行访问：

- 通过 Analytic Server “源”或“导出”节点，这些节点会在浏览器中打开控制台
- 直接通过浏览器
 1. 在浏览器的地址栏中输入 Analytic Server 的 URL。该 URL 可从服务器管理员处获取。
 2. 输入登录该服务器所使用的用户名。
 3. 输入与指定用户名关联的密码。

登录后，缺省折页为数据源折页。

导航控制台

Analytic Server 控制台有以下四个组件：

- 标题显示产品名和当前登录的用户名。
- 左列显示可用折页或功能分组。所选折页决定内容区域中显示的内容。
- 内容区域。显示与当前所选折页关联的控件。每个折页内容的详细信息显示在以下部分中。
- 页脚包含注销链接并显示 Analytic Server 的已安装版本。

数据源

数据源为记录的集合以及数据模型，它定义用于分析的数据集。记录源可以是 HDFS 上的文件（定界文本、固定宽度文本 和 Excel）、数据库或 HCatalog。数据模型定义了分析数据必需的所有元数据（字段名称、存储、测量级别等）。数据源所有者可以授予或限制对数据源的访问权。

左列

左列显示折页标题下的现有数据源。

- 选择数据源时，会在内容区域中显示其详细信息，并可以编辑其属性。在搜索区域输入时，会对列表进行过滤，从而仅显示其名称中包含搜索字符串的数据源。
- 单击**新建数据源**按钮时，会使用**添加新数据源**对话框中指定的名称和内容类型来创建新的数据源。
 - 数据源名称必须区分大小写。将忽略首尾空格。会拒绝某些特定的名称以防止 SQL 注入。
 - 可用的内容类型是“文件”、“数据库”和 HCatalog。

注：选择后，将无法编辑内容类型。

- 单击**删除数据源**按钮可移除数据源。此操作将保持与数据源关联的所有文件完整无缺。

内容区域

内容区域分为几个部分，这些部分取决于数据源的内容类型。为数据源指定了设置之后，单击预览和元数据可最终完成数据源。

数据源属性

对所有内容类型通用的设置。

名称	这是一个可编辑的文本字段，用于显示数据源名称。数据源名称必须区分大小写。将忽略首尾空格。会拒绝某些特定的名称以防止 SQL 注入。
描述	这是一个可编辑的文本字段，用于提供数据源解释性文本。
公开	这是一个复选框，用于指示是任何人都可以查看项目（选中情况下）还是必须将用户和组明确添加到所有者列表中才可以查看项目（未选中情况下）。
共享	<p>通过将用户和组添加为作者，您可以共享数据源的所有权。</p> <ul style="list-style-type: none"> 在文本框中输入时，会对其名称中包含搜索字符串的用户和组进行过滤。单击添加参与者按钮可将参与者添加到作者列表。 要移除作者，请在“作者”列表中选择用户或组，然后单击移除参与者按钮。 <p>注： 无论是否将管理员列为作者，他们对每个数据源都具有读写访问权。</p>

文件输入

专门用于定义具有文本内容类型的数据源的设置。

文件查看器

显示了可包含在数据源中的文件。选择**项目**方式可查看 Analytic Server 项目结构中的文件，或者选择 **HDFS** 可查看其余的 Hadoop 分布式文件系统。或者，您可以浏览文件夹结构，但是 HDFS 完全不可编辑，而 Analytic Server 文件夹结构仅在已定义的项目中可编辑。即，您无法在**项目**方式的 root 级别上添加文件、创建文件夹或删除项。要创建、编辑或删除项目，请使用项目折页。

- 单击**将文件上传到 HDFS**按钮时，会将文件上传到当前项目/子文件夹。
- 单击**创建新文件夹**按钮时，会在当前文件夹下创建一个新文件夹，并会以“新文件夹名称”对话框中指定的名称命名。
- 单击**将文件下载到本地文件系统**按钮时，会将选中的文件下载到本地文件系统。
- 单击**删除选中的文件**按钮时，会移除选中的文件/文件夹。

数据源定义中所包含的文件

使用**移动**按钮可将选中的文件添加到数据源或从数据源移除。对于数据源中每个选中的文件，单击**设置**可定义读取文件的规范。

文件输出

通过从 Analytic Server 上运行的流中输出，可附加具有文件内容类型的数据源。选择**可写**以启用附加功能，并选择将要写入新文件的输出文件夹。

数据库选择

为包含记录内容的数据库指定连接参数。

数据库 选择要连接到的数据库的类型。从以下类型中选择：DB2、Oracle、SQL Server、TeraData 或 Netezza。

服务器地址

输入托管数据库的服务器的 URL。

服务器端口

数据库将侦听的端口号。

数据库名称

要连接到的数据库的名称。

用户名 如果数据库受密码保护，请输入您的用户名。

密码 如果数据库受密码保护，请输入您的密码。

表名称 从要使用的数据库输入表的名称。

最大并行读取数

数据库输出

通过从 Analytic Server 上运行的流中输出，可附加具有数据库内容类型的数据源。选择**可写**以启用附加功能，并选择将要写入输出数据的输出数据库表。

HCatalog 选择

指定用于访问受管于 Apache HCatalog 的数据的参数。

数据库 HCatalog 数据库名称。

表名称 从要使用的数据库输入表的名称。

过滤器 这是表的分区过滤器（将表创建为分区表的情况下）。

HCatalog 模式

显示指定表的结构。HCatalog 可以支持高度结构化的数据。要在此类数据上定义 Analytic Server 数据源，必须对结构进行序列化，从而转换成简单行和简单列。选择模式中的元素，然后单击移动按钮，可以将元素映射到字段以进行分析。不是所有节点都可以进行映射的。例如，将复杂类型的阵列或映射视为“父级”，并且无法进行映射。可以通过树中以 `....array:struct` 或 `....map:struct` 结尾的标签来确定这些节点。

HCatalog 字段映射

向数据源中的字段显示 HCatalog 中的元素的映射。单击**预览原始数据**可查看存储在 HCatalog 中的记录；这可以帮助您确定如何将 HCatalog 模式映射到字段。

HCatalog 元素

双击单元以进行编辑。HCatalog 元素为阵列或映射时，必须对单元进行编辑。对于阵列，根据您想要将其映射到某个字段的阵列的成员来指定整数。对于映射，根据您想要将其映射到某个字段的关键字来指定加引号的字符串。请参阅第 12 页的图 2 以获取关于如何使用原始数据预览来确定对应于映射索引的字符串的示例。

映射字段

Analytic Server 数据源中出现的字段。双击单元以进行编辑。不允许复制“映射字段”列中的值，否则将引起错误。

存储 字段的存储。存储来自 HCatalog，并且无法进行编辑。

注：单击编辑和元数据来最终完成 HCatalog 数据源时，没有编辑选项。

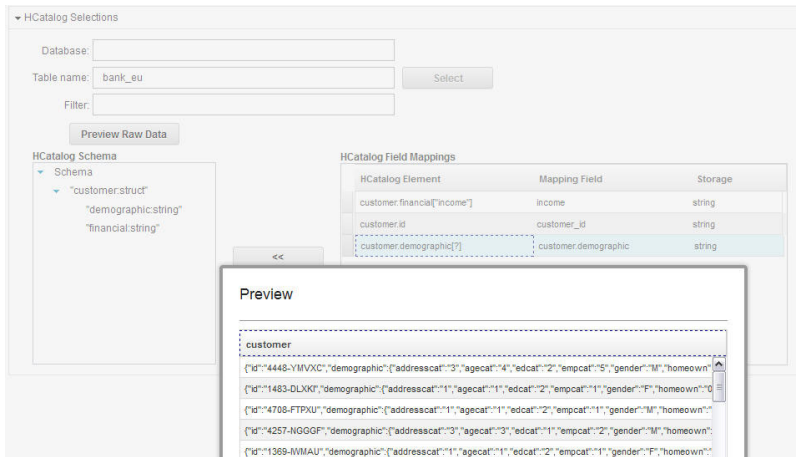


图 2. 数据源折页，用于定义 HCatalog 数据源

设置（文件数据源）

字符集编码

文件的字符编码。选择或指定 Java 字符集名称，如“UTF-8”、“ISO-8859-2”和“GB18030”。缺省值为“UTF-8”。

语言环境

定义语言环境。可选。缺省为服务器语言环境。必须将语言环境字符串指定为：
<language>[_country[_variant]]，其中：

language

是 ISO-639 所定义的有效的小写字母代码。必需。

country

是 ISO-3166 所定义的有效的大写字母代码。可选。

variant

是专门用于供应商或浏览器的代码。可选。

删除空格

将字符串字段开头和/或结尾的空格字符移除。缺省为“无”。支持以下值：

- 无 不必移除空格字符。
- 左侧 将字符串字段开头的空格字符移除。
- 右侧 将字符串字段结尾的空格字符移除。
- 两侧 将字符串字段开头和结尾的空格字符都移除。

分组符号

设置是否必须使用用于千位分隔符且特定于语言环境的字符。

字段分隔符

一个或多个标记字段边界的字符。每个字符都被视为独立的分隔符。例如，“\t”表示逗号或制表符标记字段边界。如果控制字符对字段进行了定界，那么除了控制字符之外，也将此处指定的字符视为定界符。如果控制字符未对字段和空字符串进行定界，那么缺省为“;”。

控制字符对字段进行定界

设置是否将 ASCII 控制字符（LF 和 CR 除外）视为字段分隔符。缺省为否。

首行包含字段名称

设置是否使用首行来确定字段名称。缺省为否。

要跳过的起始字符数

文件开头将被跳过的字符数。非负整数。缺省值为零。

合并空格

设置是否将相邻出现的多个空格和/或制表符作为一个字段分隔符。如果空格或制表符都不是字段分隔符，那么没有作用。缺省值为 true。

行尾注释字符

一个或多个用于标记行尾注释的字符。将忽略记录上该注释后的字符和所有内容。每个字符都被视为独立的注释标记。例如，“/*”表示注释开头的斜杠或星号。无法定义诸如“/”之类的多字符注释标记。空字符串表示未定义注释字符。如果已定义，那么在处理引号或跳过起始字符之前，先检查注释字符。缺省值为空字符串。

无效字符

确定无效字符（与编码中的字符不对应的字节序列）的处理方式。空字符串表示将抛弃这些字符。非空字符串（通常为单个字符）表示这些字符将替换为该字符串的内容。缺省值为空字符串。

单引号 指定单引号（撇号）的处理。缺省为保留。

保留 单引号没有特殊含义，可将其视为任何其他字符。

删除 除非是引用内容，否则删除单引号

成对 将单引号视为引用字符，成对单引号之间的字符将失去任何特殊含义（将其视为引用内容）。单引号本身是否可以出现在带有单引号的字符串中，这取决于设置**可以用双重引号括起引号**。

双引号 指定双引号的处理。缺省为成对。

保留 双引号没有特殊含义，可将其视为任何其他字符。

删除 除非是引用内容，否则删除双引号

成对 将双引号视为引用字符，成对双引号之间的字符将失去任何特殊含义（将其视为引用内容）。双引号本身是否可以出现在带有双引号的字符串中，这取决于设置**可以用双重引号括起引号**。

可以用双重引号括起引号

当设置为**成对**时，指示双引号是否可以显示在带有双引号的字符串中，以及单引号是否可以显示在带有单引号的字符串中。如果是，那么双引号可以在使用双重双引号的字符串中进行转义，以及单引号可以在使用双重单引号的字符串中进行转义。如果否，那么不能在带有双引号的字符串中使用双引号，并且不能在带有单引号的字符串中使用单引号。缺省为是。

预览和元数据（数据源）

单击**预览和元数据**时，会根据记录的样本生成数据源的数据模型。此时您可以查看和编辑基本元数据信息。

预览 “预览”选项卡显示了记录的小样本及其字段值。

编辑

“编辑”选项卡允许您修改基本字段元数据。

字段 双击字段名称可对其进行编辑。

- 度量** 这是用于描述给定字段中数据特征的测量级别。
- 角色** 用于告知建模节点，机器学习过程的字段将是“输入”（预测变量字段）还是“目标”（预测字段）。“两者皆是”和“无”也是可与“分区”一起使用的角色，指示了用于将记录划分到单独样本以进行培训、测试和验证的字段。**Split** 值指定为字段的每个可能的值构建单独的模型。
- 存储** 存储描述了将数据存储于字段中的方式。例如，值为 1 和 0 的字段用于存储整数数据。这与测量级别不同，测量级别描述了数据的使用情况，而不会影响存储。例如，要设置整数字段的测量级别，可将值标记为 1 和 0。通常 1 = True, 0 = False。

项目

项目是用于存储输入和访问作业输出的工作空间。项目提供了用于包含文件和文件夹的顶级组织结构。项目可以与单个用户和组分享。

左列

左列显示折页标题下的现有项目。

- 选择项目时，会在内容区域中显示其详细信息，并可以编辑其属性。在搜索区域输入时，会对列表进行过滤，从而仅显示其名称中包含搜索字符串的项目。
- 单击**新建项目**按钮时，会创建新项目，并以“添加新项目”对话框中指定的名称命名。名称区分大小写，忽略首尾空格，并且防止 SQL 注入。
- 单击**删除项目**按钮时，会移除该项目。此操作将保持与数据源关联的所有文件完整无缺。

内容区域

内容区域分为**设置**、**所有者**和**版本**选项卡。

设置

项目描述

这是一个可编辑的文本字段，用于提供项目解释性文本。

公开 这是一个复选框，用于指示是任何人都可以查看项目（选中情况下）还是必须将用户和组明确添加到所有者列表中才可以查看项目（未选中情况下）。

单击**保存**将保存设置的当前状态。

项目数据源

这是一个不可编辑的区域，用于列示与该项目相关联的所有数据源。

项目结构窗格

右侧窗格显示当前所选项目的项目/文件夹结构。您可以浏览文件夹结构，但是不能对其进行编辑，除非通过这些按钮。

- 单击**将文件下载到本地文件系统**按钮时，会将选中的文件下载到本地文件系统。
- 单击**删除选中的文件**按钮时，会移除选中的文件/文件夹。

所有者

所有者是项目的正式成员，可修改该项目以及该项目中的文件夹和文件。

“可用用户和组”列表显示了当前与此项目不关联的活动租户中的用户和组。

- 输入搜索窗格将对其名称中包含搜索字符串的用户和组进行过滤。

- 选择列表上方的用户图标将显示可用用户。取消选择该图标将隐藏这些用户。缺省情况下，选择此图标。
- 选择列表上方的组图标将显示可用组。取消选择该图标将隐藏这些组。缺省情况下，选择此图标。

使用移动按钮可将用户和组移至项目用户和组。通过 IBM SPSS Modeler 连接到 Analytic Server 时，这些组的这些用户和成员将对此项目具有读（Analytic Server 源节点）和写（Analytic Server 导出节点）访问权。

注：会自动立即应用“所有者”选项卡上所作的更改。

注：无论是否将管理员明确列为所有者，他们对每个项目都具有读写访问权。

版本

项目的版本号基于对文件和文件夹内容的更改。对项目属性的更改（如描述、是否公共的以及与谁共享）不需要新版本。添加、修改或删除文件或文件夹不需要新版本。

项目版本控制表

该表显示现有项目版本、其创建和提交日期、负责每个版本的用户以及父版本。父版本是所选版本基于的版本。

- 单击**锁定**可以对所选项目版本内容做出更改。
- 单击**提交**将保存对项目做出的所有更改并使此版本成为项目的当前可视状态。
- 单击**放弃**将放弃对锁定项目做出的所有更改并将项目的可视状态返回到最近提交的版本。
- 单击**删除**按钮将除去所选版本。

超过版本数量时，自动清除

版本数量超过指定数量时，自动删除最早提交的项目版本。

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 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。当前的 IBM 商标列表，可从 Web 站点 www.ibm.com/legal/copytrade.shtml 上『版权和商标信息』部分获取。

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。

Java 和所有基于 Java 的商标和徽标是 Oracle 和/或其子公司的商标或注册商标。

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。

索引

[D]

导出节点

Analytic Server 导出 3

[L]

流属性

分析服务器 4

[S]

数据源 9

[X]

项目 14

[Y]

源节点

Analytic Server 源 3

A

Analytic Server 导出 3

Analytic Server 源 3



Printed in China