

IBM SPSS Modeler Entity Analytics 高级配置

简介

本指南的目标读者为系统管理员，他们配置 IBM SPSS Modeler Entity Analytics (EA) 以部署在下列系统之一中：

- IBM SPSS Collaboration and Deployment Services (C&DS) V7
- IBM SPSS Modeler Server
- IBM Analytical Decision Management (ADM) V17
- IBM SPSS Modeler Solution Publisher

概述

本指南介绍如何配置使用 Coordinator of Processes (COP) 聚类配置运行的 IBM SPSS Collaboration and Deployment Services、Analytical Decision Management、Solution Publisher 或 Modeler Server 以访问 Entity Analytics 存储库。

EA 存储库必须设置在使用 Modeler Server 管理存储库的计算机上。然后，ADM、Solution Publisher 及 C&DS Jobs and Scoring Service 可以使用在 Modeler 中配置为搜索此存储库的流 EA 节点。

每个 EA 存储库由单一的数据库服务器（例如，在 Modeler 17 发行版中为 IBM DB2 数据库服务器）以及每台机器上运行的某个 EA 服务实例提供服务。

注：如果您仅使用 Modeler Server，那么本指南的大部分内容不适用于您；但是，您可能希望能够关闭 EA 服务。在这种情况下，您可以使用 `manage_repository` 脚本，请参阅管理 DB2 和 EA 服务以获取详细信息。

下图显示一组运行 C&DS 评分服务的机器的配置。

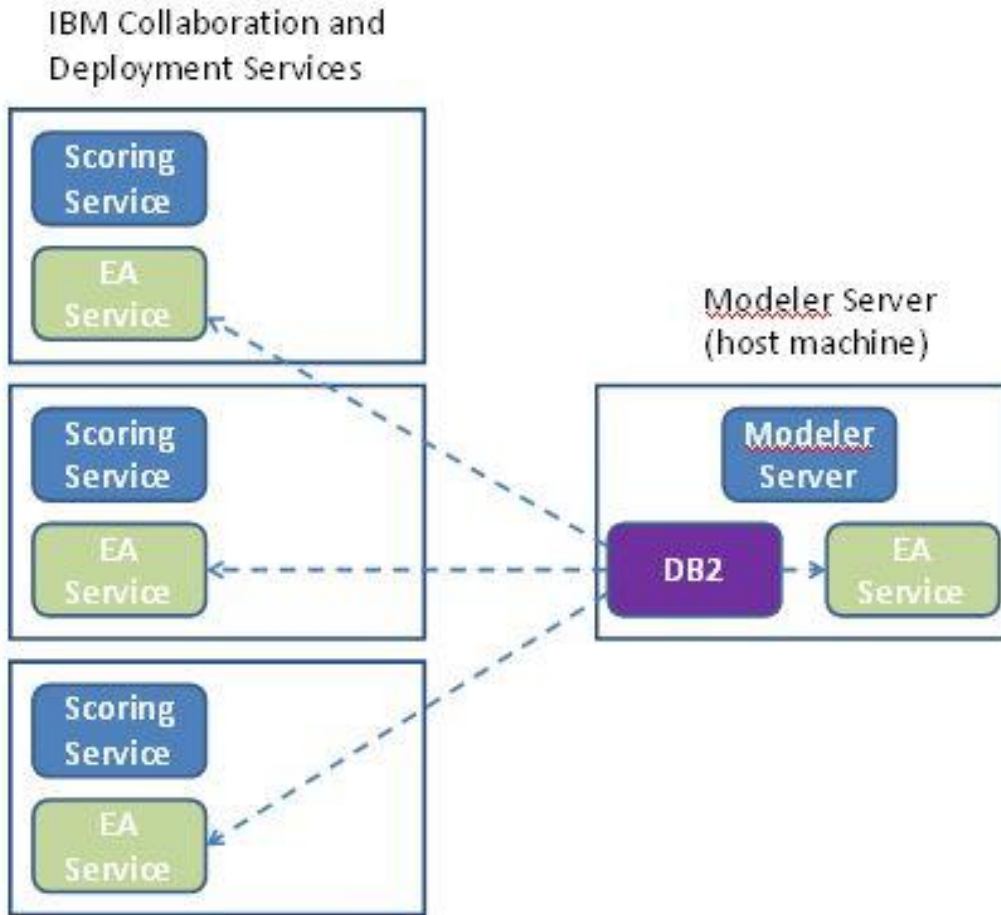


图 1 – C&DS – 聚类机器的配置

配置

第一步，记下存储库主机 Modeler Server 和每台其他机器上的 EA 安装文件夹的位置；您在进行配置时需要此信息。

以下列表显示了不同平台的默认位置。

Windows Vista、Windows 7 和 Windows Server:

C:\ProgramData\IBM\SPSS\Modeler\17.1\EA

请注意，在 Windows 上，EA 需要 Microsoft Visual C++ 2010 运行时库。必要时，请在运行任何 EA 功能之前安装 Visual C++ 2010 运行时库（可以从 Microsoft 下载中心获取）。

UNIX:

<modeler-install-directory>/ext/bin/pasw.entityanalytics/EA

Modeler Server - 存储库主机配置

作为必备软件，Entity Analytics 必须安装到存储库主机 Modeler Server 和在此机器上创建的 Entity Analytics 存储库上。此机器托管从其他服务访问的存储库，例如：

- Collaboration and Deployment Services (C&DS)
- Decision Management
- Solution Publisher
- 在 Coordinator of Processes (COP) 聚类中运行的其他 Modeler Server。

使用 `manage_repository` 脚本在存储库主机 Modeler Server 上启动和停止存储库服务。请确保已启动特定存储库的服务，这样从 C&DS/Analytical Decision Management 或 Solution Publisher 执行的流或为 C&DS 中的评分进行配置的流才能成功访问 Entity Analytics 存储库。

有关如何使用 `manage_repository` 脚本的更多信息，请参阅管理 DB2 和 EA 服务。

C&DS 配置

配置存储库后，您必须使用 C&DS Deployment Manager 客户机来添加 Entity Analytics 服务器定义。如果要对 Entity Analytics 启用实时评分，以及要在 C&DS 作业中使用包含 Entity Analytics 节点的 Modeler 流（换言之，要在 C&DS 中运行 Entity Analytics 流），那么此定义是必需的。此服务器定义必须与流中的存储库名称相匹配；此定义用于告知流可以在何处找到存储库并向其提供所需的连接信息。

要创建 Entity Analytics 服务器定义：

- (1) 通过添加两个环境变量来编辑 C&DS startServer.sh 启动脚本，如以下示例中所示。

注意：在此示例中，`/opt/ibm/V10.5` 是 DB2 安装目录，`/home/g2user` 是创建存储库期间使用的 DB2 实例用户的主文件夹。

```
DB2_INSTALL_PATH=/opt/ibm/V10.5/  
export DB2_INSTALL_PATH  
DB2_INSTANCE_HOME=/home/g2user  
export DB2_INSTANCE_HOME
```

- (2) 重新启动 C&DS 服务器。
- (3) 在 Deployment Manager 客户机中，添加新的凭证定义。用户名和密码必须与 Modeler Server 机器上使用的凭证相匹配。
- (4) 在 Deployment Manager 客户机中，添加新的服务器定义：
 - a. 输入服务器名称。此名称必须与 Modeler Server 机器上使用的存储库的名称相匹配。
 - b. 对类型选择 **Entity Analytics 服务器**，然后单击下一步。
 - c. 输入与之建立连接的 Entity Analytics 服务器的目录路径或 URL。这是 g2.ini 文件中显示的“连接”值。
 - d. 输入与用于任何加密数据的存储库相关的 salt 值；该值显示在 g2.ini 文件中。
 - e. 选择要使用的凭证。此凭证是您在步骤 1 中添加的凭证。

请注意，您还必须将 Entity Analytics 服务器信息添加到作业的 Entity Analytics 选项卡中。

IBM SPSS Modeler Deployment Guide (*ModelerDeployment.pdf*) 中提供了与 Entity Analytics 服务器设置相关的更多详细信息。

Analytical Decision Management 用于运行包含流 EA 节点的流的配置

按照 C&DS 配置中的说明，使用 C&DS Deployment Manager 客户机添加 Entity Analytics 服务器定义之后，无需执行其他步骤即可使 Analytical Decision Management 运行流。

Solution Publisher 的配置

使用 Solution Publisher 对存储库运行包含任何 EA 节点的映像时，请配置和管理每台 Solution Publisher 机器上的 Entity Analytics，如下所示：

- (1) 编辑 modelerrun 脚本，如以下示例所示。

注意：在此示例中，`/opt/ibm/V10.5` 是 DB2 安装目录，`/home/g2user` 是创建存储库期间使用的 DB2 实例用户的主文件夹。

```
DB2_INSTALL_PATH=/opt/ibm/V10.5/
export DB2_INSTALL_PATH
DB2_INSTANCE_HOME=/home/g2user
export DB2_INSTANCE_HOME
```

- (2) 根据 配置 COP 聚类一节中针对附属机器的指示信息，配置 Solution Publisher 安装。

请注意，指示信息中的示例显示了 Modeler Server 路径 - 请将这些路径更改为使用 Solution Publisher 的等效路径。

- (3) 编辑文件 `<solution_publisher_install_folder>/ext/bin/pasw.entityanalytics/ea.cfg`

添加以下行：

```
service_management, 0
```

管理 DB2 和 EA 服务

在存储库主机 Modeler Server 上，DB2 与 EA 服务进程是从 Modeler Server 单独管理的。在收到请求来执行一个 Modeler 流时，若它们未在运行中，则会按需启动，之后无论 Modeler Server 运行与否，它们都将保持独立运行。它们也可由管理员使用 `manage_repository` 脚本来启动、检查或停止。

`manage_repository` 脚本可用于其他（附属）机器上以启动和停止 EA 服务（在这些机器上，DB2 实例不在本地运行，因而无法启动或停止）。EA 服务不会自动启动。

要启动 EA 服务和 DB2 服务器（如果适用）：

```
cd <modeler-install-directory>\ext\bin\pasw.entityanalytics
manage_repository.bat|.sh -start <repository-name> <username> <password>
```

示例：

```
pasw.entityanalytics >./manage_repository.sh -start AAA myuser mypass
```

正在启动 DB2 服务器...

DB2 服务器已启动

正在启动 EA 服务...

要检查 EA 服务器与 DB2 是否处于运行状态：

调用 `manage_repository` 脚本，并使用 `-check` 选项：

```
cd <modeler-install-directory>\ext\bin\pasw.entityanalytics  
manage_repository.bat|.sh -check<repository-name> <username> <password>
```

示例：

```
pasw.entityanalytics > ./manage_repository.sh -check AAA myuser mypass  
DB2 服务器正在主机 localhost (端口 1320) 上运行  
EA 服务正在主机 localhost (端口 1321) 上运行
```

要停止 EA 服务和 DB2（如果适用）：

调用 `manage_repository` 脚本，并使用 `-stop` 选项：

```
cd <modeler-install-directory>\ext\bin\pasw.entityanalytics  
manage_repository.bat|.sh -stop <repository-name> <username> <password>
```

示例：

```
pasw.entityanalytics> ./manage_repository.sh -stop AAA myuser mypass  
正在停止本地 EA 服务  
已向 EA 服务发送关闭命令  
正在停止本地 DB2 服务器  
已向 DB2 服务器发送关闭命令
```

配置 COP 聚类

选择某个 Modeler Server 机器以托管 EA 存储库并根据 Modeler Server - 存储库主机配置中的指示信息对其进行重新配置。此机器称为 Modeler Server 主机。

对于称为附属机器的其余非主机 Modeler Server 机器，需要以下配置更改：

(1) 创建文件夹： `<EA folder>/repositories/<repository-name>`。

（其中 `<repository-name>` 是 Modeler Server 主机上 EA 存储库的名称。）

(2) 编辑文件 `<EA folder>/g2_config.xml` 并为每个存储库添加一个新实例。

EA 组件使用此文件来查找指定存储库的 g2 Web 服务。例如：

```
<instance external="false" g2host="localhost" g2port="1321" host="9.30.214.79"  
name="AAA"  
path="C:\ProgramData\IBM\SPSS\Modeler\17\EA\repositories\AAA\g2.ini"  
port="1320"/>
```

其中的重要信息为：

- **name** 属性（存储库名称）必须与 Modeler Server 主机上使用的存储库的名称相匹配。
- **path** 属性（存储库位置）：
 - Windows:
C:\ProgramData\IBM\SPSS\Modeler\17.1\EA\repositories*<repository-name>*\g2.ini
 - UNIX:
<modeler-install-directory>/ext/bin/pasw.entityanalytics/EA/repositories/*<repository-name>*/g2.ini
- 将 **g2port** 属性（用于标识 g2 服务将侦听的端口）设置为附属机器上当前未使用的端口。
- 将 **g2host** 属性设置为“localhost”。
- 将 **host** 属性指向 Modeler Server 主机的名称。
- 将 **port** 属性设置为 Modeler Server 主机上 EA 存储库使用的端口号。

(3) 将 *<modeler-install-directory>*/ext/bin/pasw.entityanalytics/templates/g2.ini 文件复制到步骤 (1) 中创建的新文件夹，然后编辑 g2.ini 的新副本：

- 将两处 {DataFolder} 更改为该文件夹的完整路径：*<modeler-install-directory>*/ext/bin/pasw.entityanalytics/data。
- 将 {Database} 更改为 Modeler Server 主机上使用的存储库的名称。
- 将 {Salt} 更改为 Modeler Server 主机上使用的存储库的 g2.ini 文件中找到的 Salt 值

(4) 将用户更改为创建存储库的用户（例如：g2user），然后使用下列命令对 Modeler Server 主机上使用的存储库进行编目：

```
su - g2user
db2 catalog tcpip node <Your Node Name> remote <host machine IP> server 50000
db2 catalog database <Repository on Host Modeler Server Machine> at node <Your Node Name>
```

(5) 重新启动附属机器上的 Modeler Server。

(6) 从 Modeler Client 连接到 Modeler Server 附属机器并连接到已配置的存储库。

下图显示运行 Modeler Server 的机器的 COP 聚类配置。

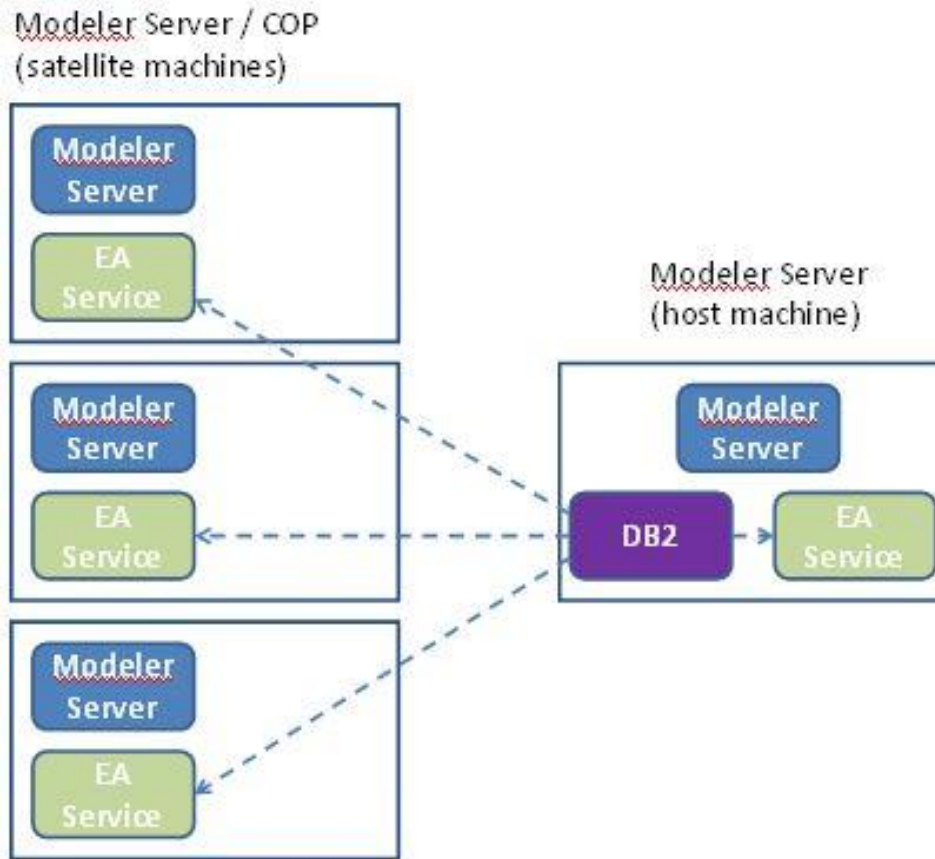


图 2 – Modeler Server – COP 聚类机器配置