IBM SPSS Modeler Entity Analytics Advanced Configuration

Introduction

The intended audience for this guide is system administrators who are configuring IBM SPSS Modeler Entity Analytics (EA) to run in a deployed fashion with one of the following:

- IBM SPSS Collaboration and Deployment Services (C&DS) version 8

- IBM SPSS Modeler Server

- IBM Analytical Decision Management (ADM) version 18

- IBM SPSS Modeler Solution Publisher

Overview

This guide shows how to configure the IBM SPSS Collaboration and Deployment Services, Analytical Decision Management, Solution Publisher, or Modeler Servers running in a Coordinator of Processes (COP) cluster configuration to access an Entity Analytics repository.

The EA repository must be set up on a machine that uses Modeler Server to host the repository. ADM, Solution Publisher, and C&DS Jobs and Scoring Service can then use the Streaming EA node within Modeler that is configured to search this repository.

Each EA repository is served by a single database server (for example, in the Modeler 17 release this is an IBM DB2 database server) and one instance of the EA service running on each machine.

Note: If you only use Modeler Server most of this guide does not apply to you; however, you may want to be able to shutdown the EA services. In this case you would use the manage_repository script, see Managing DB2 and EA services for details.

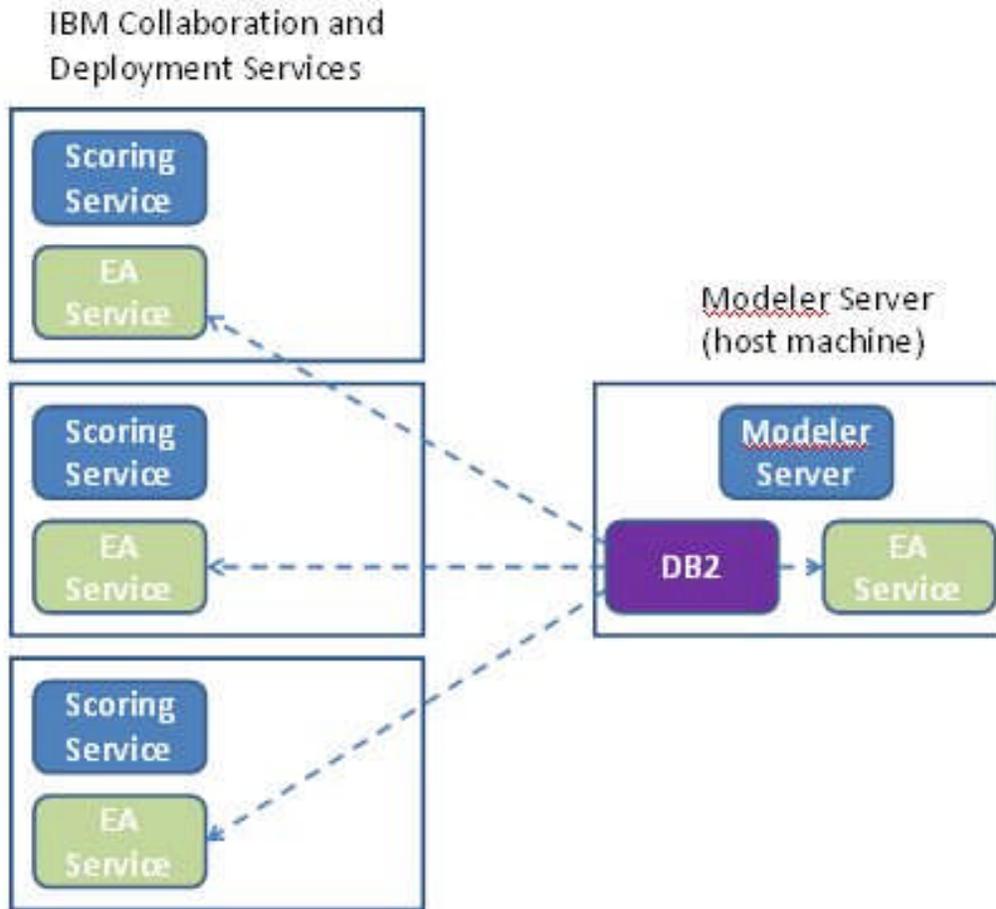The following diagram shows the configuration for a cluster of machines running the C&DS Scoring Service.

Figure 1 – C&DS – Configuration of Cluster Machines

Configuration

As a first step make a note of the location of the folder where EA is installed on the repository host Modeler Server and on each other machine; you will need this information during configuration.

The following list shows the default locations for different platforms.

Windows Vista, Windows 7, Windows Server:

C:\ProgramData\IBM\SPSS\Modeler\18\EA

Note that EA requires the Microsoft Visual C++ 2010 run time library on Windows. If necessary, install the Visual C++ 2010 runtime library (available from the Microsoft Download Center) before running any EA functions.

UNIX:

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

Modeler Server - Repository Host Configuration

As a prerequisite Entity Analytics must be installed onto a repository host Modeler Server machine and an Entity Analytics repository created on this machine (Note that this machine may be either local or remote). This machine hosts the repository that is accessed from other services, such as:

- Collaboration and Deployment Services (C&DS)

- Decision Management

- Solution Publisher

- Other Modeler Servers running in a Coordinator of Processes (COP) cluster.

Use the manage_repository script to start and stop repository services on the repository host Modeler Server machine. Ensure that a particular repository's services have been started before streams executed from C&DS, Analytical Decision Management, or Solution Publisher, or configured for scoring in C&DS, can successfully access the Entity Analytics repository.

For further information on how to use the manage_repository script, see Managing DB2 and EA services.

C&DS Configuration

After configuring the repository, you must use C&DS Deployment Manager client to add an Entity Analytics server definition. This is required both to enable real-time scoring for Entity Analytics and also to enable the use of a Modeler stream that contains an Entity Analytics node in a C&DS job (in other words, to run Entity Analytics streams in C&DS). The server definition must match the repository name in the stream; this definition is used to tell the stream where to find the repository and to give it the connection information it needs.

To create an Entity Analytics server definition:

(1) Edit the C&DS startServer.sh startup script by adding two environment variables, as in the following example.

Note In this example, */opt/ibm/V10.5* is the DB2 installation directory, and */home/g2user* is the home folder of the DB2 instance user that is used during the repository creation.

DB2_INSTALL_PATH=/opt/ibm/V10.5/
export DB2_INSTALL_PATH
DB2_INSTANCE_HOME=/home/g2user
export DB2_INSTANCE_HOME
ICC_PKCS11_ROOT=<modeler>/ext/bin/pasw.entityanalytics/g2
export ICC_PKCS11_ROOT

Note Ensure that you have read/write/edit permissions for the ICC_PKCS11_ROOT folder.

(2) Restart the C&DS server.

(3) In Deployment Manager client, add a new credentials definition. The user name and password must match the credential used on the Modeler Server machine.

(4) In Deployment Manager client, add a new server definition:

   a. Enter the server name. This must match the name of the repository used on the Modeler Server machine.

   b. Select **Entity Analytics Server** for the type and click **Next**.

   c. Enter the directory path to, or URL of, the Entity Analytics server to which the connection is made. This is the "connect" value that is shown in the g2.ini file.

   d. Enter the Salt name and checksum related to the repository that is used for any encrypted data; this is shown in the g2.ini file.

   e. Select the credential to be used. This is the credential you added in Step 1.

   f. Export the salt value from Modeler Server and import it to C&DS. Note that before you export or import the salt value by using the command line, you should set/export the environment variable: ICC_PKCS11_ROOT.

   - To export one special salt value:
     manage_repository -exportsalt password saltname exportfilename

   - To export all salt values:
     manage_repository -exportsalt -all password exportfolder

- To import one special salt value:
  manage_repository -importsalt password saltname saltfilename

- To import all salt values:
  manage_repository -importsalt -all password saltindexfile

Notes:

- You can find the saltname in the g2.ini file of the repository.

- The saltname that you import must be the same as the saltname set in the g2.ini file of the repository.

Note that you must also add the Entity Analytics server information to the job's Entity Analytics tab.

Further details relating to the Entity Analytics Server set up are available in the IBM SPSS Modeler Deployment Guide (*ModelerDeployment.pdf*).

Configuration for Analytical Decision Management to run streams containing Streaming EA nodes

After you have used C&DS Deployment Manager client to add an Entity Analytics server definition as explained in C&DS Configuration, no further steps are necessary to enable Analytical Decision Management to run streams.

Configuration for Solution Publisher

When using Solution Publisher to run images containing any EA node against a repository, configure and administer Entity Analytics on each Solution Publisher machine as follows:

(1) Edit the modelerrun script, as in the following example.

   Note In this example, */opt/ibm/V10.5* is the DB2 installation directory, and */home/g2user* is the home folder of the DB2 instance user that is used during the repository creation.

   DB2_INSTALL_PATH=/opt/ibm/V10.5/
   export DB2_INSTALL_PATH
   DB2_INSTANCE_HOME=/home/g2user
   export DB2_INSTANCE_HOME
    ICC_PKCS11_ROOT=<modeler>/ext/bin/pasw.entityanalytics/g2
   export ICC_PKCS11_ROOT

   Note Ensure that you have read/write/edit permissions for the ICC_PKCS11_ROOT folder.

(2) Configure the Solution Publisher installation according to the instructions for the satellite machines in the Configuration for a COP cluster section.

Note that the example in the instructions shows Modeler Server paths – change these to use the equivalent paths for Solution Publisher.

(3) Edit the file *<solution publisher install folder>*/ext/bin/pasw.entityanalytics/ea.cfg

Add the line:

 service_management, 0

Managing DB2 and EA services

On the repository host Modeler Server machine, the DB2 and EA service processes are managed separately from Modeler Server. They are started on demand, if not already running when required to execute a Modeler stream, and left running independently of whether Modeler Server is running or stopped. They can also be started, checked, and stopped by an administrator using the manage_repository script.

The manage_repository script can be used on other (satellite) machines to start and stop the EA service (on these machines the DB2 instance is not running locally and is not started or stopped). The EA service is not started automatically.

To start the EA service (and DB2 server if applicable):

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

Example:
pasw.entityanalytics >./manage_repository.sh -start AAA myuser mypass
Starting DB2 server...
DB2 server started
Starting EA service...

To check if EA service and DB2 are running:

Invoke the manage_repository script with the –check option:

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

Example:
pasw.entityanalytics > ./manage_repository.sh -check AAA myuser mypass
DB2 server is running on host localhost, port 1320
EA service is running on host localhost, port 1321

To stop the EA service (and DB2 if applicable):

Invoke the manage_repository script with the –stop option:

cd *<modeler-install-directory>*\ext\bin\pasw.entityanalytics

manage_repository.bat|.sh –stop *<repository-name> <username> <password>*

Example:
pasw.entityanalytics> ./manage_repository.sh -stop AAA myuser mypass
Stopping local EA service
Shutdown command sent to EA service
Stopping local DB2 server
Shutdown command sent to DB2 server

Multiple DB2 instances

If you have multiple DB2 instances, to specify which instance to use for the repository, set the
environment variable *DB2INSTANCE* in *modelersrv.sh* to the name of the required instance.

Configuration for a COP cluster

Chose one Modeler Server machine to host the EA repository and reconfigure it according to the
instructions in Modeler Server - Repository Host Configuration. This machine is referred to as the host
Modeler Server machine.

The following configuration changes are required for the remaining non-host Modeler Server machines,
referred to as satellite machines:

(1) Create a folder : *<EA folder>*/repositories/*<repository-name>*.

(Where *<repository-name>* is the name of the EA repository on the host Modeler Server
machine.)

(2) Edit the file *<EA folder>*/g2_config.xml and add a new instance for each repository.

This file is used by the EA components to locate the g2 web service for a named repository. For
example:

**<instance external="false" g2host="localhost" g2port="1321" host="9.30.214.79"
name="AAA"
path="C:\ProgramData\IBM\SPSS\Modeler\18\EA\repositories\AAA\g2.ini"
port="1320"/>**

The significant information is:

- The **name** attribute (the repository name) must match the name of the repository used on the
  host Modeler Server machine.

- The **path** attribute (the repository location):

  - Windows:
    C:\ProgramData\IBM\SPSS\Modeler\18\EA\repositories\<repository-name>\g2.ini

- UNIX:
  <modeler-install-directory>/ext/bin/pasw.entityanalytics/EA/repositories/<repository-name>/g2.ini

- Set the **g2port** attribute (identifying the port at which the g2 service will listen) to a port that is not currently used on the satellite machine.

- Set the **g2host** attribute to "localhost".

- Point the **host** attribute to the name of the host Modeler Server machine.

- Set the **port** attribute to the port number used by the EA repository on the host Modeler Server machine.

(3) Copy the *<modeler-install-directory>*/ext/bin/pasw.entityanalytics/templates/g2.ini file to the new folder created in step (1) and edit the new copy of g2.ini:

- Change the two occurrences of {DataFolder} to the full path to the folder: *<modeler-install-directory>*/ext/bin/pasw.entityanalytics/data.

- Change the {Database} to the name of the repository used on the host Modeler Server machine.

- Enter the Salt name and checksum that relate to the repository that is used for any encrypted data; this is shown in the g2.ini file. Note that before you export or import the salt value by using the command line, you should set/export the environment variable: ICC_PKCS11_ROOT. The following commands are available to export the salt value from the host Modeler Server and import it:

    - To export one special salt value:
      manage_repository -exportsalt password saltname exportfilename

    - To export all salt values:
      manage_repository -exportsalt -all password exportfolder

    - To import one special salt value:
      manage_repository -importsalt password saltname saltfilename

    - To import all salt values:
      manage_repository -importsalt -all password saltindexfile

    Notes:

    - You can find the saltname in the g2.ini file of the repository on the host Modeler Server.

    - The saltname that you import must be the same as the saltname set in the g2.ini file of the repository.

(4) Change user to the one who created the repository (for example: g2user) then use the following commands to catalog the repository used on the host Modeler Server machine:

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) Restart Modeler Server on the satellite machine.

(6) Connect to the satellite Modeler Server machine from Modeler Client and connect to the configured repository.

The following diagram shows the configuration for a COP cluster of machines running Modeler Server.
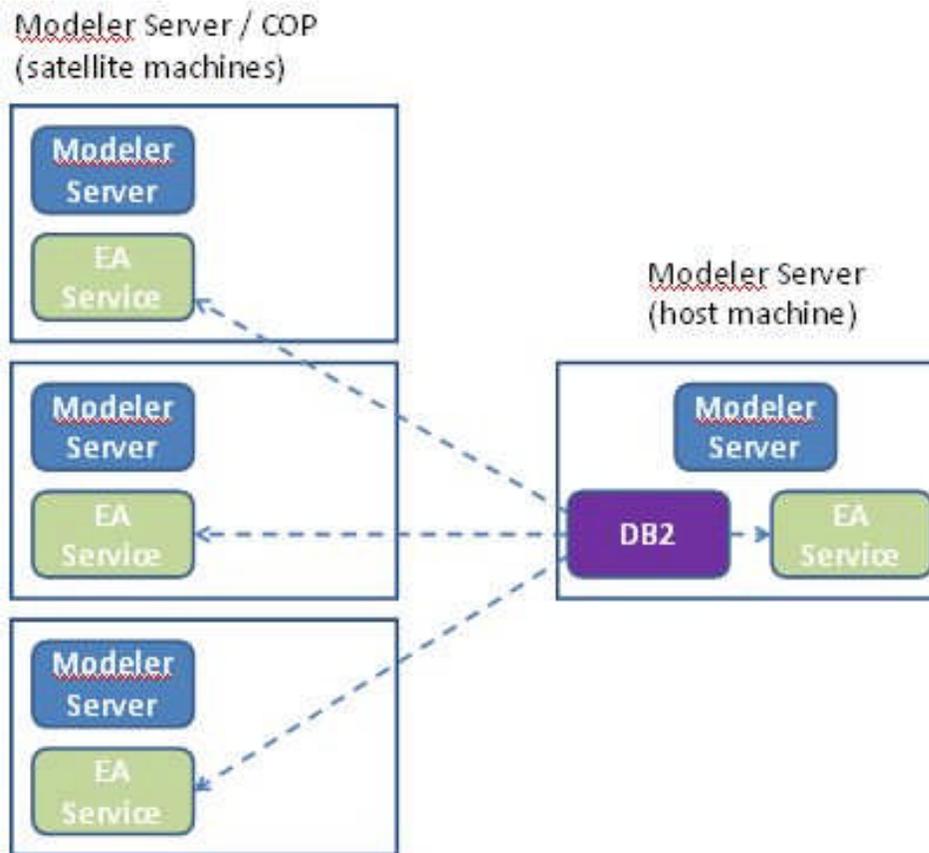

Figure 2 – Modeler Server – Configuration of COP Cluster Machines