

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 5
- IBM SPSS Modeler Server
- IBM Analytical Decision Management (DM) version 7

Overview

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

The EA repository must be set up on a machine that uses Modeler Server to host the repository. DM 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 15 premium release this is an IBM SolidDB database server) and one instance of the EA service running on each machine.

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

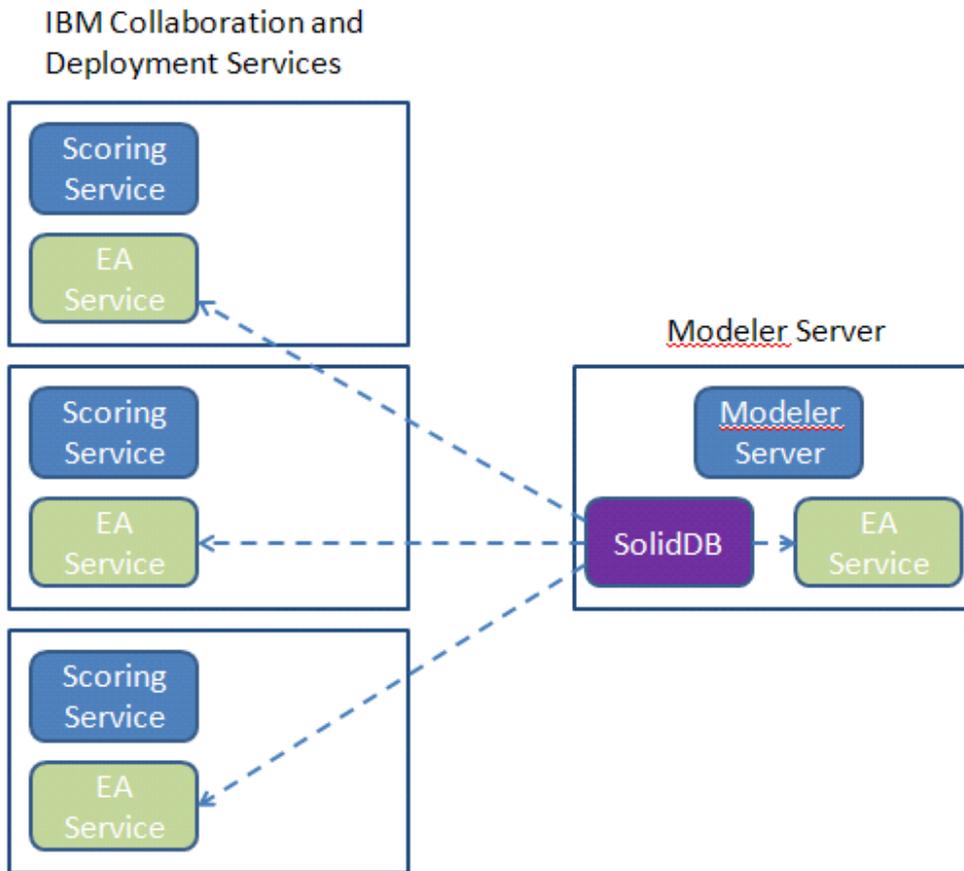


Figure 1 – C&DS – Configuration of Cluster Machines

Configuration

As a first step, make a note of the location of the folder where EA is installed on the Modeler Server and Scoring Service machines; you will need this information during configuration.

The following list shows the default locations for different platforms.

Windows XP, Windows Server 2003:

C:\Documents and Settings\All Users\Application Data\IBM\SPSS\Modeler\15\EA

Windows Vista, Windows 7, Windows Server 2008:

C:\ProgramData\IBM\SPSS\Modeler\15\EA

UNIX:

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

Modeler Server - Repository Host Configuration

As a prerequisite Entity Analytics must be installed onto a Modeler Server machine and an Entity Analytics repository created on this machine. This machine is used to create and test streams that will be later stored in Collaboration and Deployment Services (C&DS) and deployed for scoring in the Scoring Service.

This section describes how to modify the standard configuration on the Modeler Server machine for each repository that is to be shared with the Scoring Service. This change is required to make solidDB listen on an external port rather than a local port.

To modify the configuration, edit the file <EA folder>/repositories/<repository-name>/solid.ini (where <EA folder> is the installed location of EA, and <repository-name> is the name of the repository you are configuring).

In the solid.ini file, change:

```
[Com]
  Listen=tcpip 1320
  Connect=tcpip localhost 1320
```

To:

```
[Com]
  Listen=tcpip <fully-qualified-hostname> 1320
  Connect=tcpip localhost 1320
```

For example:

```
[Com]
  Listen=tcpip jupiter.mycompany.com 1320
  Connect=tcpip localhost 1320
```

In the example above, the port that SolidDB is listening on is 1320. Make a note of your port number as you use it when you configure the scoring service machines.

After editing `solid.ini`, restart the SolidDB database server and the EA service on the Modeler Server machine. For instructions on how to do this using the `manage_repository` script see Managing SolidDB and EA services.

C&DS Configuration

On each C&DS or C&DS Scoring Service machine, for each repository to be shared with the Scoring Service, the following configuration changes are required:

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

(Where `<repository-name>` is the name of the repository on the 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:

```
<?xml version="1.0" encoding="UTF-8" standalone="no" ?>
<config>
  <g2instances>
    <instance name="AAA"
      g2port="1975" g2host="localhost"
      host="jupiter.mycompany.com" port="1320" />
  </g2instances>
</config>
```

The significant information is:

- The **name** attribute (the repository name) must match the name of the repository used on the Modeler Server machine.
 - Set the **g2port** attribute (identifying the port at which the g2 service will listen to a port that is not currently used on the Scoring Service machine).
 - Set the **g2host** attribute to "localhost".
 - Point the **host** attribute to the host name of the machine running Modeler Server / SolidDB.
 - Set the **port** attribute to the port number used by SolidDB.
- (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`.

- If on UNIX, change {Port} to a string with the format: "<host> <port>", indicating where SolidDB is listening on the Modeler Server machine.
- If on Windows, change {Port} to a string with the format: "<port>", indicating where SolidDB is listening on the Modeler Server machine.

Original g2.ini file:

```
#
# G2.INI
#

[GNR]
  SUPPORTPATH={DataFolder}

[ER]
  SUPPORTPATH={DataFolder}

[SQL]
  CONNECTION=solid://G2USER:G2PASSWD@tcp {Port}
  LAST_TOUCH_WAIT=10
  AUTO_UNTOUCH=1
...
```

Example of an edited g2.ini file (UNIX):

```
#
# G2.INI
#

[GNR]
  SUPPORTPATH=/usr/IBM/SPSS/ModelerServer/15.0/ext/bin/pasw.entityanalytics/data

[ER]
  SUPPORTPATH=/usr/IBM/SPSS/ModelerServer/15.0/ext/bin/pasw.entityanalytics/data

[SQL]
  CONNECTION=solid://G2USER:G2PASSWD@tcp jupiter.mycompany.com 1320
  LAST_TOUCH_WAIT=10
  AUTO_UNTOUCH=1
...
```

Example of an edited g2.ini file (Windows):

```
#
# G2.INI
#
```

```
[GNR]
  SUPPORTPATH=C:\Program Files\IBM\SPSS\Modeler\15\ext\bin\pasw.entityanalytics\data

[ER]
  SUPPORTPATH= C:\Program Files\IBM\SPSS\Modeler\15\ext\bin\pasw.entityanalytics\data

[SQL]
  CONNECTION=solid://G2USER:G2PASSWD@tcp 1320
  LAST_TOUCH_WAIT=10
  AUTO_UNTOUCH=1
  ...
```

(4) **Windows Only** - Set up an ODBC connection pointing at the SolidDB server.

- The ODBC driver should be SolidDB 7.0 (Unicode). This driver should be installed as part of the Modeler Adapter install.
- Set the DSN to “tcp <port>”, where <port> is the port on the Modeler Server machine on which SolidDB is listening.
- Set the NetworkName to “tcp <hostname> <port>”, where <hostname> is the host name of the Modeler Server machine and <port> is the port on which the SolidDB server is running.

(5) Edit the <modeler-install-directory>/ext/bin/pasw.entityanalytics/ea.cfg file; add the line:

```
service_management, 0
```

After completing these configuration steps, you will need to start the EA service on each C&DS machine using the manage_repository script. For further information, see Managing SolidDB and EA services.

Configuration for Decision Management to run streams containing Streaming EA nodes

When using Decision Management to run streams containing a streaming EA node against a Modeler Server instance, configure and administer Entity Analytics on the Modeler Server machine as follows:

(1) Edit the file <modeler install folder>/ext/bin/pasw.entityanalytics/ea.cfg

Add the line:

```
service_management, 0
```

and restart Modeler Server.

(2) Use the manage_repository script to start and stop repository services on the modeler server machine. Ensure that a particular repository's services have been started before streams executed from Decision Management can successfully access the Entity Analytics repository.

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

Managing SolidDB and EA services

On the Modeler Server machine, the SolidDB 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 must be used on the Scoring Service machine to start and stop the EA service (on the Scoring Service machines the SolidDB instance is not running locally and is not started or stopped). The EA service is not started automatically.

To start the EA service (and SolidDB 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 SolidDB server...  
SolidDB server started  
Starting EA service...
```

To check if EA service and SolidDB 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  
SolidDB server is running on host localhost, port 1320  
EA service is running on host localhost, port 1321
```

To stop the EA service (and SolidDB 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 SolidDB server  
Shutdown command sent to SolidDB server
```

Configuration for a COP cluster

This configuration is very similar to the configuration for C&DS.

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

Configure the remaining Modeler Server machines according to the instructions in C&DS Configuration.

NOTE: Omit step (5) - reconfiguring ea.cfg.

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

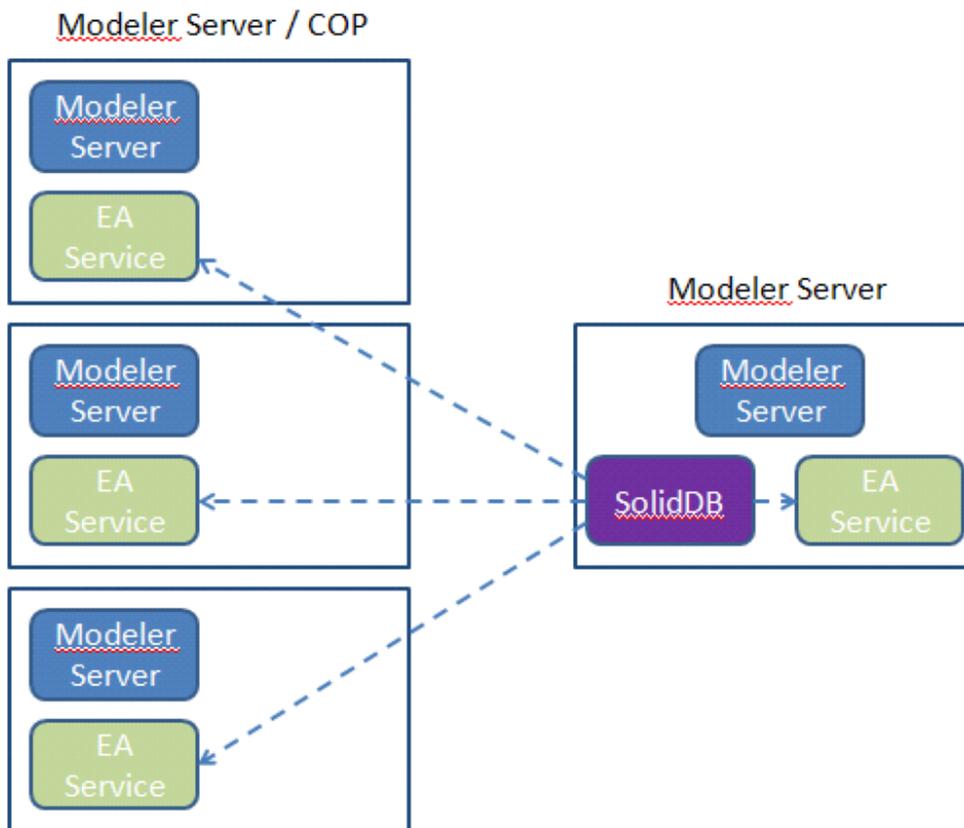


Figure 2 – Modeler Server – Configuration of COP Cluster Machines