The IBM SPSS Collaboration and Deployment Services product (a.k.a. C&DS) is a JEE based application that can be leveraged by other IBM SPSS products (Modeler, Statistics, Analytical Decision Management, etc.) to provide additional capabilities. This article will describe five ways that C&DS can be used to enhance your IBM predictive analytic solution.

View more content in this series

Introduction

The IBM SPSS product portfolio consists of several different predictive analytic related products. These include products such as IBM SPSS Modeler, IBM SPSS Statistics, IBM SPSS Analytical Decision Management and IBM SPSS Data Collection. However, you may not be aware that the IBM SPSS Collaboration and Deployment Services (C&DS) product can be used to extend and enhance the base capabilities provided by these other IBM SPSS applications. The integration between the various IBM SPSS products and C&DS is enabled by installing a C&DS “Adapter” that is packaged with the various IBM SPSS base products (such as Modeler Server Adapters for C&DS).

Figure 1: Various IBM SPSS products can be enhanced with IBM SPSS C&DS

This article will introduce five ways that C&DS can be used to enhance your predictive analytic solution by discussing five of the most valuable and commonly used C&DS features.
1. Analytic Repository
2. Automation/Process Management
3. Scoring
4. Asset Lifecycle Management/Object Promotion
5. Model Management

There are additional features that provide value to customers which you can read more about in the IBM SPSS C&DS product documentation linked in the Resources section of this article.

#1 – Analytic Repository

The C&DS product provides you with an analytic repository that can be used to manage various analytical assets. These assets are commonly files produced by the other IBM SPSS products. However, you can also store and manage other types of files such as PDF and XML. The C&DS repository includes version/label capability to help manage the life cycle of the various assets. The repository assets are stored in an enterprise relational database such as IBM DB2, which allows C&DS to leverage database capabilities such as security, backup, and restore.

Let’s walk through a common IBM SPSS Modeler user scenario to see how the C&DS repository can help. In the beginning you may have a single analyst working with the Modeler client product. At this stage the analyst is simply saving and loading all files locally on their client machine. While this level of functionally is fine, it does have some limitations.

- What if I have more than one analyst and I want them to be able to easily share assets?
- What if my analyst develops a predictive model and wants to save and share a specific version, yet continue to work on additional enhancements?
- How do I back up the analyst’s work so that if their machine fails I don’t lose the predictive model that was under development?
- What if I want to restrict access to some artifacts while allowing access to others?

Here is where the C&DS repository can come to the rescue. One or more analysts can be creating various artifacts that can now be easily managed (e.g. shared, versioned, labeled, secured, backed up) from a central location as shown in Figure 2.

**Figure 2: Multiple users storing files in C&DS analytic repository**

In addition to convenient sharing, versioning, and labeling of the stored assets, the C&DS repository supports the notion of users, roles, and permissions for managing access to stored
assets. There are also inherent locking capabilities to avoid collisions between users working with the same assets. Now all users can be assured they are working with the appropriate artifacts in a controlled manner.

#2 – Automation/Process Management

Automation is another major capability provided by the C&DS product. Using C&DS automation you can define a work flow for reoccurring tasks, schedule when the work is initiated, and receive notifications about significant events that occur.

Let’s continue with our analyst example. Suppose the analyst has developed some kind of predictive model or stream that they now intend to run to determine the predictive outcome. Once again, an analyst could manually load and run the desired stream file locally on their client. While this may be sufficient for development purposes, it is not likely desirable in production where efficiency, reliability, and predictability are crucial. One enhancement could be to bring a Modeler Server into the picture and the analyst could load and run the stream file remotely on the Modeler Server (Figure 3).

Figure 3: Analyst runs Modeler stream on remote Modeler server

This has several advantages including,

- Run the stream on more powerful server-class hardware.
- Improved data access by running the stream closer to the input data instead of pulling data out to each client.
- Improved efficiency and availability. Multiple clients can run the stream on shared servers and if one server fails, another is available to continue the work.

Things are getting better, but instead of having the analyst manually run the stream, what if I wanted to schedule the run to happen every week on Sunday night and generate a report of the result ready for everyone on Monday morning? This is where the C&DS automation functionality can come to the rescue. If you are storing your analytic artifacts in the C&DS repository, then you can use the C&DS automation capabilities to configure a job to run automatically on a designated schedule.

You now have a fully automated solution. C&DS automation gives you the ability to design and create custom jobs that contain various job steps of your choosing. You can then schedule these jobs to be run based on various criteria. In addition, the C&DS Remote Process Server can be installed and used as an agent to allow remote custom operations on any generic server machine.
The C&DS automation can therefore be configured to send work to various IBM SPSS servers such as Modeler, Statistics or generic servers as shown in Figure 4.

**Figure 4: C&DS Automation can submit work across various server sets**

An automated job can create artifacts that are stored back into the repository. To facilitate communication about this activity, C&DS also includes a notification service that can be used to send an e-mail, for example, when a meaningful event such as a job completion or report creation occurs.

**#3 – Scoring**

Scoring is about using a predictive model with real input data to produce a prediction. These predictions can be everything from credit worthiness of a potential customer to fraud detection of an insurance claim. Using C&DS scoring you can integrate model predictions into your business process to produce real customer value.

Generally speaking the scoring of a predictive model will fall into one of two broad data categories.

1. Scoring data at rest
2. Scoring data in motion

**Figure 5: Data categories and associated scoring technologies**

The C&DS product contains scoring technologies that are useful when working with either data at rest or data in motion (or a combination of the two). Common technologies for analyzing data at rest include the C&DS automation service jobs, SPSS Modeler Batch, and Modeler Scoring...
UDF’s. For data in motion, the IBM SPSS Solution Publisher, the IBM SPSS Analytics Toolkit for InfoSphere Streams, and the C&DS Scoring Service are leveraged.

In the previous scenario I described how C&DS capabilities can be used to automatically run a Modeler stream on a Modeler server. This is an example of using C&DS for scoring data at rest. Using a C&DS job to automate the score of a Modeler stream is commonly referred to as **batch scoring**.

In some cases, using data at rest is not sufficient. Suppose you have a call center where operators are dealing with customers who have called in with a complaint. You want to predict if the customer is likely to leave and if so, make them a special offer. In this scenario one needs to be able to score the data in motion, based on live customer input while they are still on the phone. Here you could modify your call center application to invoke the C&DS scoring service web service API to produce and return a score.

**Figure 6: Call center application invokes C&DS Scoring Service API in real-time**

Using the C&DS scoring service to initiate and return a score is commonly referred to as **real-time scoring**. The dynamic input data can be passed in with the score request, and any required static customer data can be retrieved by the server from an associated data source as shown in Figure 6.

In these examples I demonstrate two of the most common scenarios.

1. Batch scoring using the C&DS automation service and a Modeler server.
2. Real-time scoring using the C&DS scoring service.

For more information on the various scoring technologies, consult the IBM SPSS C&DS product documentation link in the Resources section of this article.

### #4 – Asset Lifecycle Management/Object Promotion

A common predictive analytic deployment will involve the following activities,

1. Development of a new predictive model.
2. Testing of the model.
3. Deployment of the model for production use.

This **Dev > Test > Prod** scenario is a common work flow. The question becomes how do you manage your predictive analytic assets across these areas? Once again, C&DS comes to the
rescue. For smaller single C&DS server topologies, you can manage these different deployment areas using the C&D repository facilities such as folders, filenames, labels, and access control as shown in Figure 7.

**Figure 7: Single C&DS server with different repository areas for Dev and Prod**

For larger topologies with multiple C&DS servers, C&DS provides object promotion functionality which allows you to promote assets from the analytic repository on one C&DS server to the analytic repository of another C&DS server as shown in Figure 8. This supports having a separate C&DS server for each deployment area, while providing a controlled way of managing the movement of analytic assets between C&DS servers.

**Figure 8: Separate C&DS server for each deployment area**

The C&DS object promotion support contains several features that allow you to manage the process. Some examples include,

- Reusable promotion policies to define the business logic for promotion.
- Version label notification that can be used to initiate the promotion process.
- Limit scope so that only a single file version and associated versions of dependents get promoted.
- Control of promotion activity via a **Promote Objects** security action privilege.
- Support for both immediate and delayed promotion.
The C&DS object promotion support will allow you to properly manage and control the flow of assets across your analytic deployment area.

#5 – Model Management

So far in the example scenario, I have described how C&DS can be used to enhance your productivity all the way from the initial model development to the use of real-time scoring in a call center application. Now we have a model in production and are using it to produce actionable insight, such as a call center response with a special offer. To some it might appear that we are done. However, those that work in the field of predictive analytics understand that the production model and data will drift over time. This is due to the ever present nature of change - factors such as the behavior of people, the process in your company, types of special offers and actions by competitors all change over time.

In order to address these changes, one should have a strategy for the on-going comparison and analysis of older models and data with newer models and data. Once again, C&DS comes to the rescue with the included model management support. C&DS model management is about using, evaluating, and refreshing models. C&DS model management can be used to control the life cycle of a deployed model and contains support for the following functions,

- Model Refresh
- Model Evaluation
- Champion Challenger
- Score

**Figure 9: C&DS model management options for evaluation/refresh of a deployed model**

Using C&DS model management one can perform tasks such as train models, rank models, find the best model to score operational data. By automatically comparing models and selecting the best one (the Champion Challenger) over time you can continue to achieve valid results. This is demonstrated in Figure 9.

Model evaluation and comparison can focus on accuracy, gains, or accreditation.

- **Accuracy** - The accuracy of a model reflects the percentage of target responses that are predicted correctly. Models having a high percentage of correct predictions are preferred to those having a low percentage.
- **Gains** - The gains statistic is an indicator of the performance of a model. This measure compares the results from a model to the results obtained without using a model. The improvement in the results when using the model is referred to as the gains. When comparing two models, the model having the higher gains value at a specified percentile is preferred.
• **Accreditation** - Model accreditation reflects the credibility of a model. This approach examines the similarity between new data and the training data on which a model is based. Accreditation values vary from 0 to 1, with high values indicating greater similarity between the predictors in the two data sets. When comparing two models, the model having the higher accreditation value is based on training data that is more similar to the new data, making it more credible and preferred.

With C&DS model management capabilities you now have control over the entire predictive model life cycle. For more information on C&DS model management, consult the IBM SPSS C&DS product documentation linked to in the Resources section of this article.

**Conclusion**

The IBM SPSS Collaboration and Deployment Services product can be used to extend and enhance the base capabilities of other IBM SPSS applications. While the C&DS product is diverse and provides many capabilities, this article introduced you to five common ways that C&DS can be used to enhance your predictive analytic solution.

1. Analytic Repository
2. Automation/Process Management
3. Scoring
4. Asset Lifecycle Management/Object Promotion
5. Model Management

I hope after reading this article you have a better understanding of the benefits provided by the IBM SPSS Collaboration and Deployment Services product.

**Acknowledgments**

The author is grateful to the following persons who have contributed towards preparing or reviewing the paper for its technical content and accuracy: Alex Jones C&DS product management, Tom Kochie C&DS development manager, and Duane Wiebe C&DS quality assurance.
Resources

- For overall C&DS product information (description, pricing, education, demo downloads), IBM SPSS Collaboration and Deployment Services.
- For C&DS product documentation, IBM SPSS Collaboration and Deployment Services v6.0 - Information Center
- For a deeper understanding of C&DS capabilities from an administration point of view
  - IBM SPSS Collaboration and Deployment Services v6.0 – Deployment Manager User’s Guide
  - IBM SPSS Collaboration and Deployment Services v6.0 – Deployment Manager Server Administration Guide
- For more information on the IBM SPSS Modeler Solution Publisher, IBM SPSS Modeler v16.0 Solution Publisher
About the authors

Kevin Kepros

Kevin Kepros is a senior software engineer at the IBM Software Development Lab in Rochester, Minnesota. Kevin is a lead developer on the IBM SPSS Collaboration and Deployment Services product. Prior to joining the IBM Business Analytics team, Kevin was a lead developer on the WebSphere High Availability Manager component and a member of the WebSphere Clustering development team.

Keith Swanson

Keith Swanson is a software engineer at the IBM Software Development Lab in Rochester, Minnesota. Keith is a developer on the IBM SPSS Collaboration and Deployment Services product specializing in installation and configuration. In addition, Keith maintains the back-end build infrastructure for the product and is involved with release support.

© Copyright IBM Corporation 2014
Trademarks
(www.ibm.com/developerworks/ibm/trademarks/)