



Using Informix for your Data Warehouse: OVERVIEW

IBM **Information Management** software

Informix.
software

TABLE OF CONTENTS

1. Introduction	2
2. Informix Warehouse Components.....	3
2.1 Informix Dynamic Server 11.5.....	3
2.2 Design Studio	3
2.2.1 Data Flows.....	5
2.2.2 Control Flows	5
2.3 SQL Warehousing Tool	6
2.4 Warehouse Administration Console.....	8
3. Conclusion.....	13

1. Introduction

IBM® Informix® Dynamic Server (IDS) is now available with tools to build a powerful warehouse infrastructure platform to lower costs and increase profits by doing more with existing operational data and infrastructure. IDS already offers many of the key characteristics needed in a mission-critical warehousing environment. The Informix Warehouse Feature simplifies design and deployment of an Informix warehouse. It allows customers to more easily enable these business applications, supplying a state of the art Extract-Load-Transform (ELT) tool, in an easy to use GUI/Eclipse environment. This comprehensive platform provides the foundation you need to cost effectively deploy and build next-generation analytic solutions using the IBM Informix Dynamic Server.

Using Informix for a data warehouse is an ideal solution for Informix users who want to build end-to-end business intelligence and reporting solutions using data from various sources, including IDS. They can utilize more effectively front-end analysis and reporting tools, like IBM Cognos®, or develop mashups and other dashboards. Informix customers can simplify operational complexity and reduce costs by using a single database server for both operational and warehouse data.

The Informix Warehouse Feature includes the SQL Warehouse (SQW) tool that has been integrated with IDS V11.50. SQW includes the following components:

- ✓ Design Studio
- ✓ SQL Warehousing Tool
- ✓ Warehouse Administration Console

The SQW Design Studio, is a client tool available on Windows and Linux:

Platform	Operating System
Intel x86-32	Microsoft Windows® XP SP2 Professional, Windows Vista Business/Enterprise/Ultimate Microsoft Windows® XP and Vista with FDCC support Linux® SLES 10 SP2, RHEL 5.2

The SQW Administration Console may run on a WebSphere Application Server, which is included with the Informix Warehouse Feature. The SQW Administration Console is available on the following server platforms:

Platform	Operating System
Intel x86-32	Microsoft Windows® 2003 SP2 (32-bit)
Intel/AMD x86-64	Microsoft Windows® 2003 SP2 (64-bit) Linux® SLES 10 SP2, RHEL 5.2
Sun Sparc	Sun Solaris® 9, 10 (64-bits)
IBM Power PC	IBM AIX® 5.3 TL8, 6.1 SP4

For platform availability of Informix Dynamic Server 11.5, consult the [Informix system requirements site](#).

2. Components for an Informix warehouse platform

2.1 Informix Dynamic Server 11.50

The legendary performance, scalability, resilience and ease-of-use combined with low administration requirements of IDS makes it an ideal database server of choice to support integrated mission-critical OLTP applications. Customers who want to do more with their operational data in real-time for just-in-time dynamic business decisions can continue to rely on IDS to support data warehouse workloads.

The following key features in IDS will help you support the specific needs of a demanding data warehouse or data mart repository workload:

- ✓ **Multi-threading IBM Informix Dynamic Scalable Architecture™ (DSA)** - which makes IDS one of the best and most powerful DBMS technologies available, delivering mainframe-caliber scalability, performance, manageability and flexibility, with efficient use of hardware and operating system resources for both OLTP and DSS environments
- ✓ **Decision Support Systems (DSS) parameters** - to set up the engine to optimize memory for DSS queries as well as efficient hash joins
- ✓ **Parallel Data Query (PDQ)** - for parallel operations such as light scans, extensive calculations, sorts and multiple joins; ideal for decision support queries and batch operations
- ✓ **Time-cyclic Data Management** - fragment elimination, fragment attach and detach operations, and data and index distribution schemas, to ultimately improve manageability of large volumes of data, and also increase performance by maximizing I/O throughput
- ✓ **Configurable page sizes** - in disk and memory, for additional performance advantages
- ✓ **Ability to handle large chunks** - allowing an IDS instance to support large volumes of data of up to 128 PB.
- ✓ **Quick sequential scans** – essential for table scans common to DSS environments

Starting with IDS 11, the legendary availability and reliability of IDS includes a full active-active cluster solution for high availability and low cost scalability. You can use Informix to manage workload distribution across multiple read-only or full-transaction nodes, and dynamically add different types of nodes into your cluster environment to scale out or increase the availability in the most demanding environments. Warehouse workloads have the flexibility to on the same database with operational data, running real-time on a separate node in the cluster. Data can also be replicated in real-time using Enterprise Replication, or copied to a separate data warehouse server. With Informix you have the flexibility to design the system to meet your needs and to make the most of your existing infrastructure.

2.2 SQW Design Studio

The SQW Design Studio provides an Eclipse-based common design environment for connecting to source and target databases, creating and reverse-engineering physical data models, and building SQL-based data flows and control flows. You can use Design Studio to physically model tables that will be used as sources and targets for the SQL Warehousing Tool, to design data transformation and data flows, to create control flows to put control logic in the execution of the data flows, and to model and consolidate your data and from different sources into your Informix data warehouse.

Each window in the Design Studio contains one or more perspectives. Perspectives contain views and editors and control what the Design Studio displays in certain menus and tool bars. The Design Studio opens the Data Warehousing perspective by default. You create projects in the Design Studio where you can work on different parts of your data warehousing solution.

A project is a set of objects that you create in the Design Studio as part of the data transformation or warehouse building processes. You can build projects in the Design Studio, and test their validity without impacting the database. Each project that you build is represented by an icon in the Data Project Explorer, where you can expand it, explore its contents, and access editors that enable you to work with it.

The following figure shows an example of using the Design Studio to create a data flow.

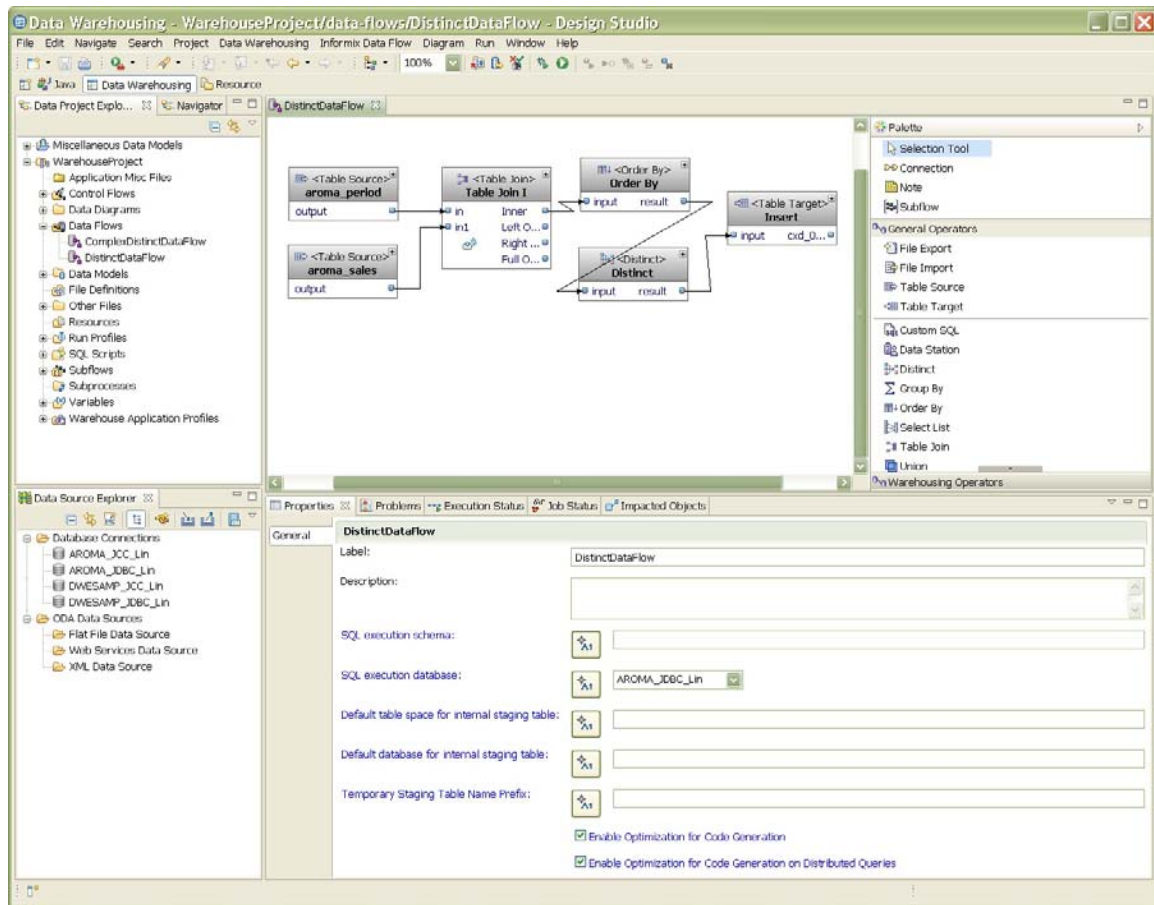


Figure 1: Design Studio

2.2.1 Data Flows

Data flows define the SQL-based data movement and transformation activities that run in an Informix database. Data flows define data transformation activities that extract data from flat files or relational tables, transform the data, and load it into an Informix data warehouse.

The Design Studio provides an intuitive way to visualize and design data flows. Database metadata, imported to or created in the Data Project Explorer, describes the schemas of database objects. Graphical operators model the various steps of a data flow activity. By arranging source, transform, and target operators in a canvas work area, connecting them, and defining their properties, you can create models that meet your business requirements.

Users visually build a graph of operators to describe their SQL data movement and transformation requirements in an Informix server. This data flow is processed by the SQL Warehousing Tool to generate Informix-optimized SQL for execution. The Design studio offers various useful operators including “sources” such as file imports and table sources and “targets” such as inserts, updates, merges, and file exports.

The following figure shows an example of a data flow where rows from two source tables are first joined, arranged in a specific order, and then made distinct. Finally, the resulting rows are inserted into a target table.

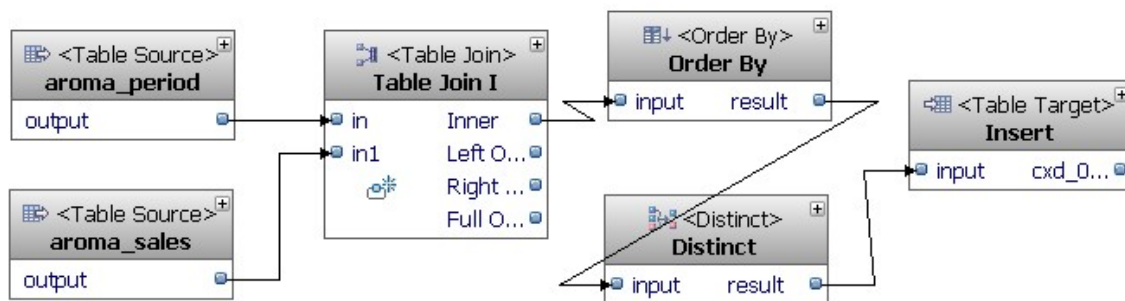


Figure 2: Example of a data flow

2.2.2 Control Flows

Control flows are used to coordinate (sequence in order or run in parallel) several activities, including data flows and other control flows. One or more control flows can be packaged into “warehouse application” profiles and deployed to a runtime environment via the Warehouse Administration Console, where such processes can be scheduled and monitored.

Sub-processes are control flow modules that can be reused when they are embedded inside main control flows (or other sub-processes).

Variables are convenient in control flows especially when used as parameters for data flow operators. For example, if you have a directory of data files that need to be loaded through a data flow, an Iterator

operator can loop through the directory and assigns a variable with each file name. This variable can be used as the file name in a File Import operator in the data flow.

The following figure shows the usage of an iterator value as part of IN parameter for a stored procedure in the loop. At the same time, the Variable Comparison operator checks if the OUT value from the stored procedure execution meets the condition or not. If it does, it continues the loop. If not, it breaks the loop. While in the entire loop, a data flow is also run.

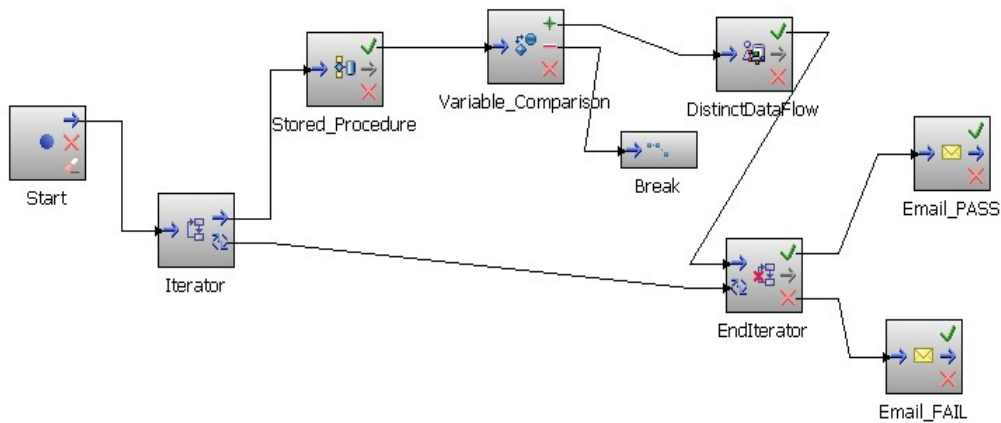


Figure 3: Example of a control flow

2.3 SQL Warehousing Tool

The SQL Warehousing Tool is a graphical environment that works with the Design Studio to enable you to quickly and easily build in-database data movements and transformations into your warehouse. It solves data movement and integration problems in an Informix warehouse environment. Users can model logical flows of higher-level operations, which generate units of code that are organized inside execution plans. The tool provides a metadata system and an IDE to create, edit, and manage these flows, and a code generation system that translates the flow models into optimized SQL code to test-run, schedule and execute.

The following figures show the SQL Warehousing operator palettes for data flows and control flows. Note that there are Informix specific data flow operators available.

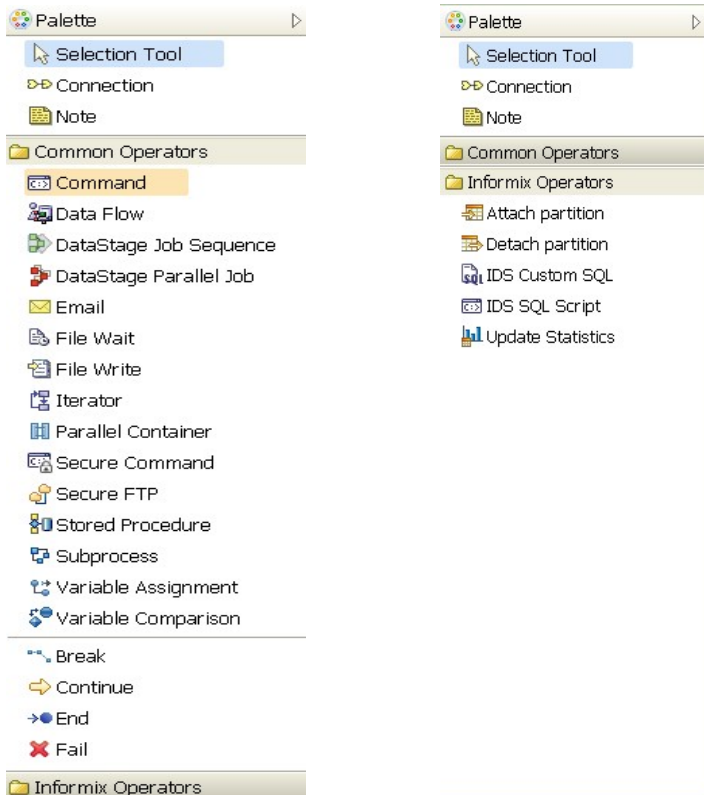


Figure 4: SQL Warehouse Data Flow Operators

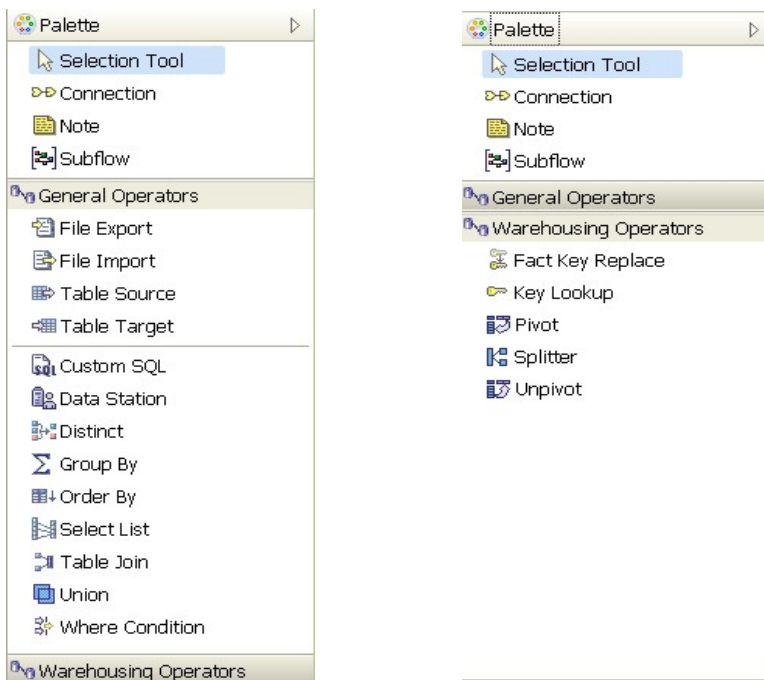


Figure 5: SQL Warehouse Control Flow Operators

When flow development is complete, the second part of the system comes into play: the packaging of the generated code and associated artifacts into a data warehouse application that can be deployed into various target runtime systems, via the Warehouse Administration Console.

2.4 SQW Administration Console

The SQW Administration Console is the single web-based application for managing and monitoring the data flows that you design using the Design Studio. The Administration Console allows you to:

- ✓ Manage common resources such as database connections and machine resources
- ✓ Schedule when the execution of control flows
- ✓ Monitor the execution status

The following figure illustrates the Warehouse Administration Console where you can manage your warehouse applications and their processes.

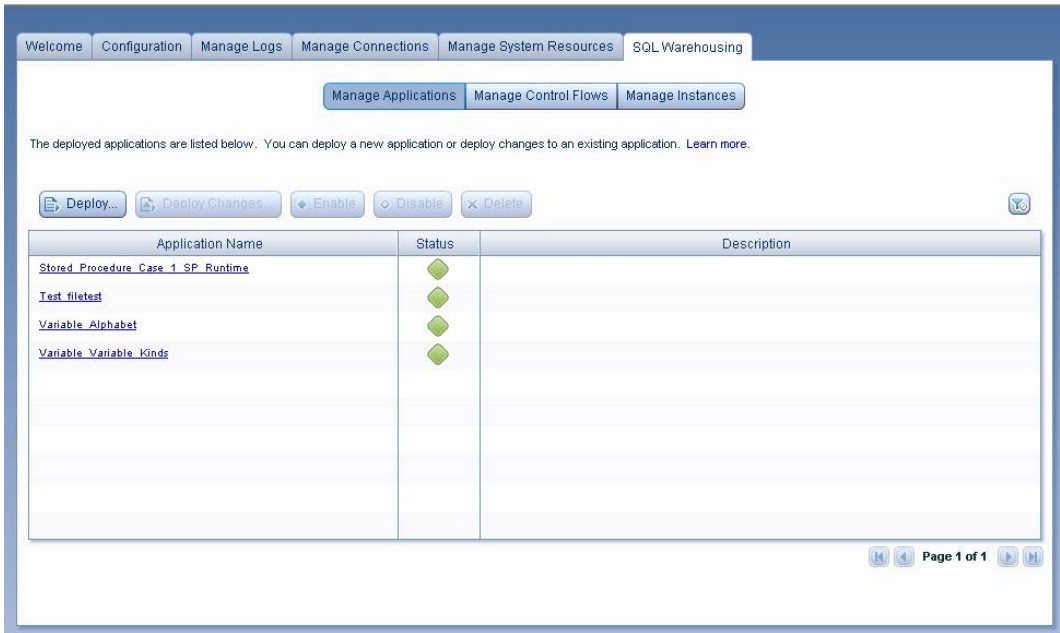


Figure 6: Warehouse Administration Console – Manage Applications

In Figure 7, the Manage Schedules tab under the Manage Control Flows tab allow the creation, modification, and deletion of schedules for when control flows run. The other tabs also manage applications and instances.

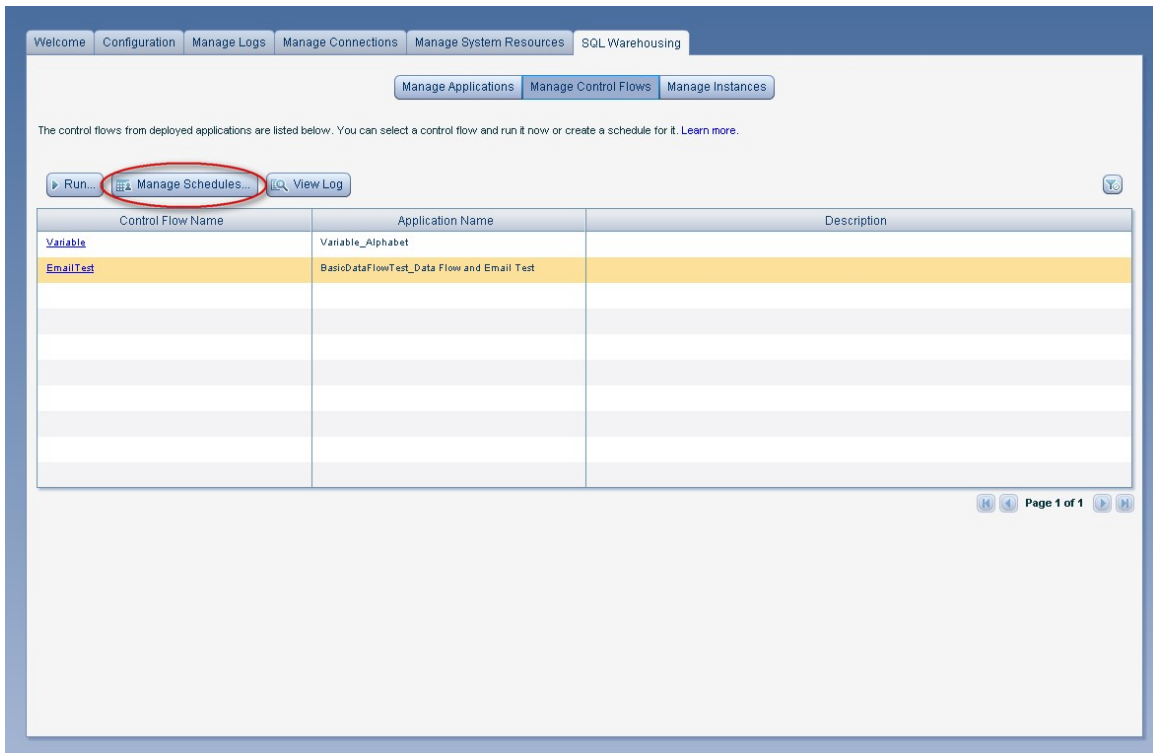


Figure 7: Warehouse Administration Console – Manage Schedules

To explore the Manage Schedules further, a user goes through a series of steps in scheduling a control flow. See the following figure for an example of scheduling a control flow to start on February 3, 2009 and repeat every day for nine days.

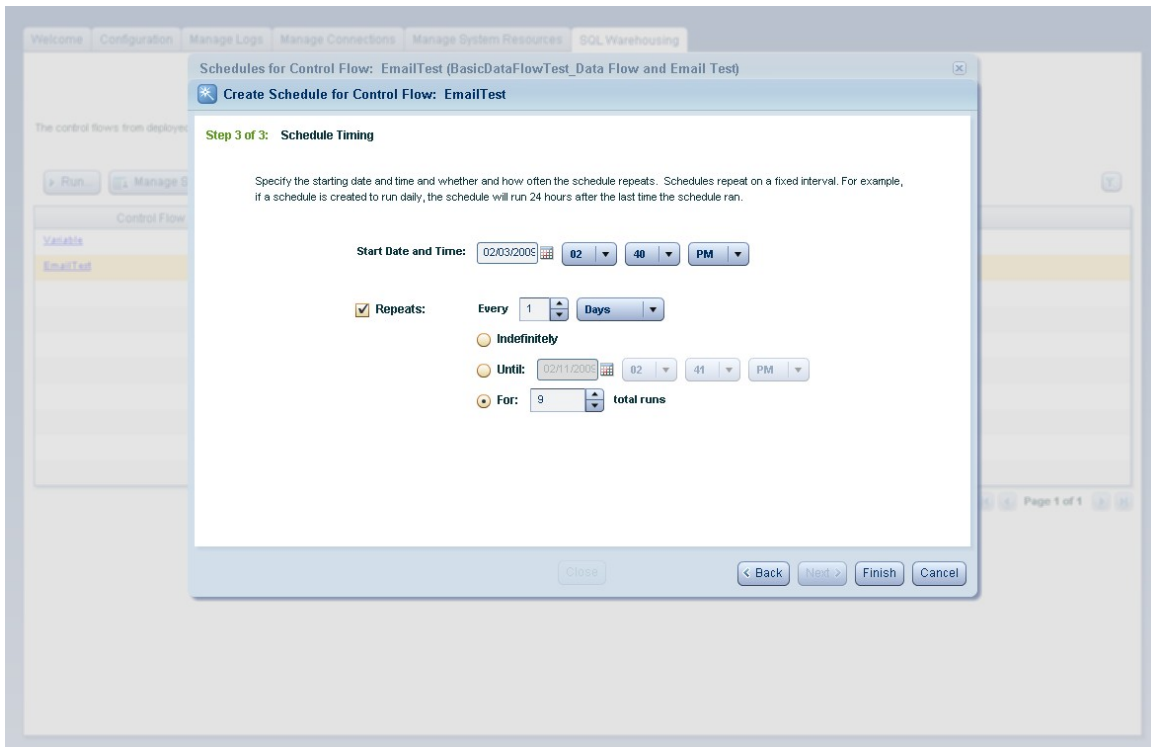


Figure 8: Warehouse Administration Console – Create Schedule

After the control flows are scheduled, you would want to monitor the status of the instance. See Figure 9 for an example. The application has two activities: run a data flow and send an email. We can see that both activities succeeded.

