

Project Trident: An Introduction

Microsoft Project Trident: A Scientific Workflow Workbench

Version 1.0 – November 6, 2009

Abstract

Microsoft Project Trident: A Scientific Workflow Workbench simplifies the process of creating, running, and sharing scientific data analysis workflows. The paper provides a general discussion of Project Trident, including overviews of workflows, the Trident architecture, the Trident Workbench applications, runtime services, and the Trident Registry.

**Note:**

* Many of the resources discussed in this paper are provided with the Project Trident package.

For a complete list of documents and software discussed in this paper, see “Resources” at the end of this document.

* For Project Trident updates and news, see:   
  <http://research.microsoft.com/en-us/collaboration/tools/trident.aspx>

Disclaimer: The information contained in this document represents the current view of Microsoft Corporation on the issues discussed as of the date of publication. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information presented after the date of publication.

This White Paper is for informational purposes only. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS DOCUMENT.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

Unless otherwise noted, the example companies, organizations, products, domain names, e-mail addresses, logos, people, places and events depicted herein are fictitious, and no association with any real company, organization, product, domain name, email address, logo, person, place or event is intended or should be inferred.

© 2009 Microsoft Corporation. All rights reserved.

Microsoft, MSDN, Silverlight, SQL Server, Visual Studio, Windows, Windows Server, and Windows Vista are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

The names of actual companies and products mentioned herein may be the trademarks of their respective owners.

Document History

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Date | Change |  |  |  |
| April 14, 2009 | Version 0.9 preview draft | | | |
| November 6, 2009 | Version 1.0 updates to Software and Hardware Requirements sections. | | | |

Contents

[Introduction 4](#_Toc245211496)

[Terminology 5](#_Toc245211497)

[Getting Started 6](#_Toc245211498)

[Prerequisites 6](#_Toc245211499)

[Software and Hardware Requirements 6](#_Toc245211500)

[How to Install Trident 7](#_Toc245211501)

[Workflow Overview 7](#_Toc245211502)

[Flow Charts 7](#_Toc245211503)

[Windows Workflow Foundation 8](#_Toc245211504)

[Project Trident Workflow Platform 9](#_Toc245211505)

[Trident Architecture 10](#_Toc245211506)

[Trident Workbench 11](#_Toc245211507)

[Trident Workflow Components 11](#_Toc245211508)

[Standard Workflows 11](#_Toc245211509)

[Standard Activities 12](#_Toc245211510)

[Standard Type Initializers 12](#_Toc245211511)

[Custom Activities and Type Initializers 12](#_Toc245211512)

[Composer Overview 13](#_Toc245211513)

[Management Studio Overview 14](#_Toc245211514)

[Workflow Manager 15](#_Toc245211515)

[Registry Manager 16](#_Toc245211516)

[Workflow Applications 16](#_Toc245211517)

[Client Workflow Application 17](#_Toc245211518)

[Silverlight Workflow Applications 18](#_Toc245211519)

[Runtime Services and Applications 18](#_Toc245211520)

[Executor 18](#_Toc245211521)

[Scheduling Service 18](#_Toc245211522)

[Blackboard Service 19](#_Toc245211523)

[Provenance Service 19](#_Toc245211524)

[Monitoring Service 19](#_Toc245211525)

[Trident Registry 19](#_Toc245211526)

[External Data Sets 20](#_Toc245211527)

[Registry UI 20](#_Toc245211528)

[Data Storage 21](#_Toc245211529)

[Execution Nodes 22](#_Toc245211530)

[Resources 24](#_Toc245211531)

# Introduction

As a scientist, you know that many important scientific questions must be addressed as a whole, by multiple disciplines. For example, some important oceanography problems require large-scale multidisciplinary experiments, such as the Neptune project. This integrated approach to science has considerable promise, but it also presents considerable challenges.

After the project begins generating data, you will probably face two key issues:

* How to analyze a wide variety of data from a large number of sources.

The research issues are often not well understood at this point, so you need a tool that lets you easily experiment with different analytical approaches.

* How to analyze terabytes or even petabytes of data.

That’s often too much for a single computer; you need the power of a high-performance cluster.

Microsoft® Project Trident: A Scientific Workflow Workbench effectively addresses both these issues, and includes some useful related features. To summarize:

* Trident is based on the Windows® Workflow Foundation (WF), which is a flexible and powerful platform for implementing workflows.

A workflow is a pipeline of individual data processing components—called activities—that is basically a programmatic realization of a flow chart. Trident provides a simple and flexible way to create and run WF data analysis workflows.

* Trident includes two applications—Trident Composer and Trident Management Studio—that are referred to collectively as Trident Workbench.

Workbench provides the tools to create, run, and manage Trident workflows.

* You don’t have to write a single line of code to create a Trident workflow.

You create workflows by using Composer’s visual design tool to “snap” together and configure the appropriate activities. Trident includes flow control activities, such as an **If Else** activity, that allow you to construct arbitrarily complex workflows.

* You don’t have to implement an application to host and run the workflow.

Workbench hosts and runs Trident workflows. You can even schedule workflows to run automatically, on a daily or weekly basis. A Microsoft Silverlight® workflow application can be run remotely by using Silverlight-enabled browsers, which are available for a variety of operating systems.

* Trident is extensible.

Many scientific workflows will require activities that are not included in the Trident package. Developers can extend the standard set of Trident activities by implementing custom activities to handle tasks such as specialized data processing procedures.

* Trident is powerful.

Trident is designed to work seamlessly with Windows® HPC Server 2008, which allows you to run multiple workflows concurrently on different cluster nodes. For large data sets, you can implement Dryad-based activities to perform data analysis as a distributed application on the cluster.

* You can easily share Trident workflows with colleagues.

You can package a workflow as an Open Packaging Convention (OPC) file and upload it directly to My Experiment, or send it to colleagues. With the Project Trident: Word Add-in for Reproducible Research, you can embed workflow links in Microsoft Office Word 2007 documents. You can then provide the document to colleagues, who can run the workflow from the document.

This paper is an introduction to Project Trident, including a brief description of Workbench and a discussion of the underlying Trident architecture and functionality. For more detailed information, see the following documents.

* “Trident Composer User’s Guide” describes how to use Composer to create and run workflows.
* “Trident Management Studio User’s Guide” describes how to use Management Studio to manage workflows and activities.
* “Trident Programming Guide” describes how to implement custom Trident activities and type initializers.
* “Trident Activity Reference” describes the activities that are included with the Trident package.
* “Project Trident: Word Add-in for Reproducible Research: User Guide” describes how to embed Trident workflow links in a Word 2007 document.

Links to these documents are in “Resources” at the end of this paper.

# Terminology

This section defines the key terms for Trident and Windows Workflow Foundation (WF) that are used in this paper.

activity

A component of a workflow.

basic activity

An activity that performs a particular task.

composite activity

An activity that serves as a container for other activities.

job

A workflow instance that is either running or scheduled to run at a specified time.

root activity

The composite activity that hosts the workflow proper. For Trident Workbench, the root activity is always a **SequentialActivity** activity.

Trident Registry

A database that contains Trident workflows, activities, type initializers, and so on.

type initializer

A module that provides a user interface (UI) for initializing properties with complex data types.

workflow

A structured collection of activities that forms a data analysis pipeline.

workflow runtime

A Windows Workflow Foundation component that manages running workflows and provides an interface between workflows and their host applications. Workbench is the host application for Trident workflows.

# Getting Started

This section discusses how to get started with Project Trident.

## Prerequisites

You don’t need to be a programmer to construct workflows. Trident Workbench is a largely self-contained tool that does not require you to implement any code. However, programming experience—Microsoft .NET Framework in particular—is helpful.

Extending Trident by implementing components such as custom activities requires significant programming experience. In particular, you should be familiar with the .NET Framework and the WF programming model and API.

## Software and Hardware Requirements

This section describes the system requirements for running Trident Workbench.

Software

Project Trident package

Windows Vista®, Windows 7, or Windows Server® 2008 operating system (or later version)

Microsoft SQL Server® 2008 or SQL Server® 2005

SQL Server Express is adequate for learning how to use Trident Workbench, but a full version of SQL Server is recommended for production use.

Microsoft .NET Framework version 3.5, SP1

Microsoft Visual Studio® 2008, if you want to implement custom activities

Microsoft Silverlight with Internet Information Services (IIS) 7.0 enabled, if you want to use or create Silverlight workflow applications

Microsoft Office 2007, if you want to use the Trident Document Add-in

Hardware

A Windows Vista or Windows 7capable computer

A Windows HPC Server 2008 cluster, if you need more performance that can be achieved with a single computer

## How to Install Trident

The Project Trident package consists of Trident Workbench, client and Silverlight workflow applications, and supporting components. The Trident Document Add-in is provided in a separate package.

For details on how to install Trident and the Trident Document Add-in, see “Trident Installation Guide.” The procedure installs the applications and supporting components, and places icons for Composer and Management Studio on your desktop.

# Workflow Overview

A workflow is basically a mechanism for programmatically realizing a procedure defined by a flow chart. This section provides a brief overview of workflows and the Windows Workflow Foundation.

## Flow Charts

Flow charts are useful tools for describing a wide variety of processes—including data analysis, process control, and decision making. A basic understanding of flow charts is essential to understanding workflows. Figure 1 shows a simple data analysis flow chart.

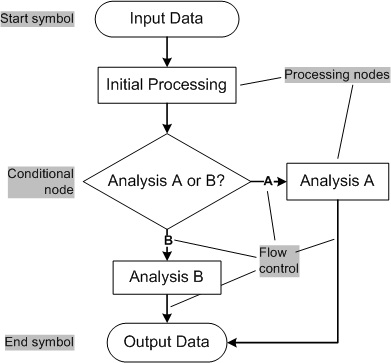


Figure 1. A simple flow chart

Figure 1 includes the key components of a typical data processing flow chart:

Input data

Data processing typically starts with stored input data, but it could also be real-time data.

Processing nodes

These nodes modify the data in some way, such as applying a filter to a time series.

Flow control

The arrows indicate how control passes from node to node.

Conditional nodes

These nodes determine control flow, based on a user-defined condition.

Figure 1 includes an if-else conditional node that passes control to either Analysis A or Analysis B.

Output data

The final output, which could be a data file, a graphical display, and so on.

Working flow charts are similar to Figure 1. However, they are usually larger and more complex, and include additional specialized symbols. For an introduction to flow-charts, see “Flowchart,” which is listed in the “Resources” section at the end of this document.

## Windows Workflow Foundation

Flow charts such as Figure 1 are visual representations and must be translated into working code. Windows Workflow Foundation provides a straightforward way to represent a flow chart as a structured sequence of components—called activities—which form a data analysis pipeline—called a workflow. An activity is a .NET object that implements part of the flow chart’s functionality. The WF workflow runtime manages running workflows and provides an interface between the workflow and the host application.

Windows Workflow Foundation includes a large collection of activities that can handle many data processing and flow control tasks. For specialized processing requirements, you can implement custom activities. There are two categories of activity:

* Basic activities correspond to flow-chart processing nodes.

They perform a particular task, such as filtering a time series or displaying processed data.

* Composite activities host collections of activities.

They manage how control flows through the child activities.

Composite activities can host both composite and basic activities, which allows you to construct arbitrarily complex workflows.

Figure 2 shows a how the flow chart from Figure 3 would be implemented as a WF workflow.



Figure 2. A simple Windows Workflow Foundation workflow

The workflow’s root activity is a **Sequence** activity—a composite activity that hosts the workflow proper and manages the overall control flow. Several basic activities correspond to the data processing nodes in Figure 2. **IfElse** is a composite activity that corresponds to the if-else conditional in Figure 1. It determines whether control passes to Analysis A or Analysis B.

For information on how to implement applications with Windows Workflow Foundation, see books such as *Windows Workflow Foundation: Step by Step* and *Windows Workflow Foundation Essentials*. For MSDN® documentation, including the API reference, see “Windows Workflow Foundation” in the “Resources” section.

## Project Trident Workflow Platform

You could implement your data analysis procedure by using Windows Workflow Foundation directly, but it’s not an ideal solution:

* You don’t have a lot of time for routine programming.

Implementing a WF workflow requires a fair amount of .NET programming experience, and you must also implement an application to host the workflow.

* You need to experiment with different analytical schemes.

You don’t want to rewrite WF workflow and host-application code each time you want to try a somewhat different analytical approach.

* You are primarily concerned with how data flows from one processing step to the next.

Windows Workflow Foundation is primarily concerned with how control flows from one activity to the next.

Trident addresses these issues by providing a workflow platform that simplifies the process of creating and running WF workflows for scientific data analysis. It is tailored for the needs of working scientists, and allows you to focus on understanding your data, not the mechanics of implementing workflows and host applications.

# Trident Architecture

Figure 3 shows the basic Trident architecture.



Figure 3. Trident Architecture

The following list is a brief summary of the components of Figure 3. The following sections describe them in more detail.

Trident Workbench

Workbench consists of two applications:

Composer is a tool for creating or modifying workflows.

Management Studio is a tool for running workflows, and managing the collection of workflows, activities, and related components.

Workflow Applications

Workflow applications provide a simple way to run workflows. There are two types of workflow applications.

The client Workflow Application runs workflows locally on any Windows computer with Trident installed.

The Microsoft Silverlight Workflow Application is similar to the client Workflow Application, but runs on a server as a Web application.

You don’t have to install Trident to run a Silverlight Workflow Application. You can run it from any browser that supports Silverlight, which includes computers running operating systems other than Windows.

Windows Workflow Foundation and .NET Framework V.3.5 SP1

Trident is based on the Windows Workflow Foundation, which is part of the .NET Framework Version 3.5 SP1. This version of the .NET Framework is included with Visual Studio 2008, or you can download it directly from Microsoft. For a download link, see “Resources.”

Runtime Services

Trident depends on several runtime services and applications, including:

Execution service runs Trident workflows as a service.

Tray application runs Trident workflows as user applications.

Blackboard service handles internal messaging.

Scheduling service runs scheduled Trident jobs at the appropriate time.

Provenance service records executing workflows for later examination.

Monitoring service tracks workflow execution.

Trident Registry

The Trident Registry is a data store that represents the collection of workflows, activities, type initializers, and so on that make up a Trident workflow. The Trident Registry also represents related components such as stored data products. The Trident Registry contains only metadata.

Data Storage

The actual data that are represented by the Trident Registry are in the associated data store—an instance of either SQL Server 2008, SQL Server 2005, or SQL Server Express.

Execution Nodes

The computer that the workflow runs on is called the execution node. Trident workflows can run on a variety of execution nodes, including:

Locally, on your computer

On a remote computer

On Windows HPC Server 2008

# Trident Workbench

The two Workbench applications—Composer and Management Studio—are easy-to-use tools for creating, managing, and running Trident workflows. Some tasks—such as importing custom activities—can be performed with either application. From a user perspective, just use the application that best suits your needs. Both applications run on top of the Trident Registry, so any changes that you make with Composer are immediately reflected in Management Studio, and vice versa.

## Trident Workflow Components

Composer and Management Studio both work with the collection of basic components stored in the Trident Registry. The Project Trident package includes a set of standard components, and you can create custom components as required. The components fall into three primary categories: workflows, activities, and type initializers.

### Standard Workflows

Trident includes a set of workflows, mostly for oceanographic and bioinformatics data analysis. Depending on your project, you might be able to use them directly or with some modifications. They are also useful learning tools. More will be available in the future, and you can also share workflows on public sites such as MyExperiment.

### Standard Activities

Trident includes a standard set of activities, which fall into two general categories:

Infrastructure

Trident includes a set of composite activities—such as **SequenceActivity** and **If Else**—that provide workflow structure. You can implement additional composite activities if necessary, but the default set is sufficient for most scientific workflows. Trident also includes some general-purpose basic activities—such as activities that write data products to or read data products from the Trident Registry—that are useful to many workflows.

Specialized data processing

Trident includes a collection of specialized data-processing activities—mostly related to oceanography and bioinformatics—that can be incorporated into any suitable workflow.

Additional activities will be posted on Trident’s Microsoft Connect Web site, as they become available. See “Resources” for a link.

### Standard Type Initializers

By default, you configure an activity in Composer by typing strings in a text box to initialize some or all of the input property values. This approach is adequate for integer or string properties, but not for complex data types such as structures. In addition, some data are inconvenient to enter manually. For example, fully-qualified file names are strings, but typing a long file path is tedious and error-prone.

A type initializer is a .NET component that is associated with a particular data type. It provides a UI—similar to a dialog box—that is customized for the particular data type, and gives users a straightforward way to initialize the associated properties.

### Custom Activities and Type Initializers

The standard composite activities should be sufficient for most scientific workflows. However, many projects will require basic activities that are not included in the Trident package to handle tasks such as specialized data processing. In that case, you must obtain appropriate custom activities and import them into Workbench.

There are two primary ways to obtain custom activities:

* Implement them.

Trident activities are basically lightly modified WF activities, and are straightforward to implement. For details, see “Trident Programming Guide.”

* Obtain them from colleagues.

Trident activities are packaged as DLLs, which can be easily shared. In particular, you can share Trident activities through the Trident page on Microsoft Connect.

If your custom activity uses complex data types, you can implement a custom type initializer for the data type. Typically, an activity DLL includes any required type initializers, but they can also be packaged as stand-alone DLLs.

For details on how to implement custom activities and type initializers, see “Trident Programming Guide.”

## Composer Overview

Composer is an easy-to-use tool for creating Trident workflows. This section provides a brief description of Composer. For a detailed description, including a tutorial on how to create Trident workflows, see “Trident Composer User’s Guide.”

Figure 4 shows Composer’s primary UI.



Figure 4. Composer primary UI

The Composer UI has four components:

* The Toolbar allows you to run a variety of commands, such as starting a new workflow, or saving a completed workflow to the Trident Registry.

There are actually four toolbars, which you select by clicking the appropriate tab. Figure 4 shows the **File** toolbar.

* The Workflow Catalog shows all the available workflows and activities in the Trident Registry.

You use these items to create or modify workflows.

* The working pane is where most of the work takes place.

Composer has three working panes, which you select by clicking the appropriate tab. The most important is the Composer pane, which hosts a visual designer that simplifies the process of creating or modifying workflows.

* The status bar displays the status of executing workflows.

Figure 5 is a clip from the Composer working pane, showing an example of a workflow under construction. The workflow consists of three activities: Demux Data, ParallelActivity, and Display Data.

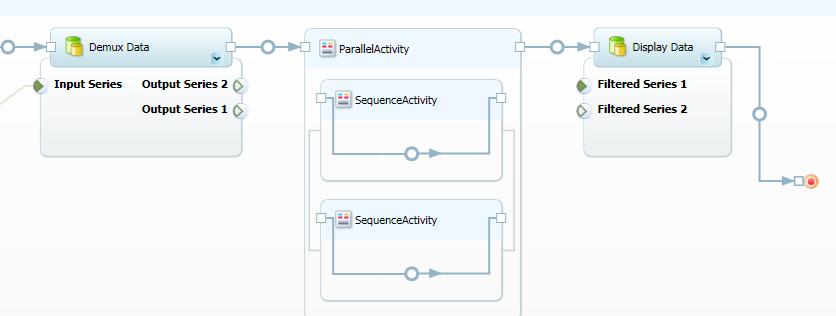


Figure 5. The Composer working pane

To create a workflow, you drag activities—or even complete workflows—from the Workflow Catalog and drop them in the appropriate location on the Composer working pane. Workbench handles the mechanics of connecting the activities. Each activity has a set of input and output properties—displayed below the activity box—that you set to configure the activity. You specify data flow by connecting an activity’s output properties to the appropriate input properties on a downstream activity.

You can run the completed workflow by clicking a button. You can also use Management Studio, which provides a more sophisticated set of tools for running workflows.

## Management Studio Overview

Composer allows you to run workflows and manage the Trident Registry, but that’s not its primary purpose. Management Studio provides a much more powerful and flexible set of tools for such tasks, including some features that are not available in Composer. This section provides a brief description of Management Studio. For a detailed description, see “Trident Management Studio User’s Guide.”

Figure 6 shows Management Studio’s primary UI.



Figure 6. Management Studio primary UI

The Management Studio UI is divided into four regions:

* The Manager pane allows you to select the Trident component that you want to work on.

The Manager pane has two primary displays—Workflow Manager and Registry Manager—which you select by clicking the bars at the bottom of the pane. When Workflow Manager is displayed, you can also click the Search tab to search for workflows.

* The toolbar contains a limited set of tools.

Most Management Studio tasks are handled by the working pane toolbars.

* The working pane is where you do the actual work, such as importing an activity or scheduling a job.

The working pane details—including the tasks supported by its toolbar—depends on the particular item that you have selected.

* The status bar displays the status of executing workflows.

### Workflow Manager

Workflow Manager handles a variety of workflow-related tasks, including:

* Run workflows.

You can run the selected workflow immediately by clicking a button.

* Schedule jobs.

A job is an instance of a workflow. You can schedule a job to run once—at a specified date and time—or you can create recurring jobs, which run daily or weekly.

* Manage running jobs.

You can pause, resume, or stop a running job, or examine its details.

* Examine completed jobs.

You can, for example, examine the input or output data of the different activities, which can be useful for debugging.

* Restore deleted workflows.

Deleted workflows are moved to a separate list, but remain in the Trident Registry. If you change your mind, you can restore a deleted workflow to the list of available workflows.

### Registry Manager

You use the Registry Manager to manage the contents of the Trident Registry, including workflows, activities, and type initializers. The Registry Manager allows you to perform tasks such as:

* Share Trident workflows.

You can package a workflow as an OPC-formatted file and upload it to your MyExperiment account, or send it directly to colleagues.

* Import Trident workflows.

You can import a workflow created elsewhere and run it, or use it to create new workflows.

* Import custom Trident activities.

You can use Management Studio to import the activities into the Trident Registry, and then incorporate them into workflows.

* Import custom type initializers.

You might need to import custom type initializers for custom activities.

* Manage the registry’s structure.

You can manage how workflows and activities are displayed by creating new folders, moving activities or workflows from one folder to another, and so on.

* Manage access

Workbench allows you to create user accounts and set access limits for those accounts. For example, you can grant administrative access to users who should have complete access to your Trident installation. You can also set access limits on individual workflows. For example, a user might be allowed to run a particular workflow but not modify it. You can also manage access rights by setting up groups of users and specifying access rights for the group.

# Workflow Applications

You typically use Workbench to run the workflows that you create. However, you might want to provide the workflow and the accompanying data to your colleagues, so that they can run it themselves. They might not have Trident installed on their computers and might not even be using computers that are running the Windows operating system.

The solution is to use a workflow application. There are two types of workflow application: client and Silverlight-based.

## Client Workflow Application

The client workflow application runs locally, on any computer with Trident installed. Figure 7 shows an example of a Workflow Application window for the Ocean Current workflow.

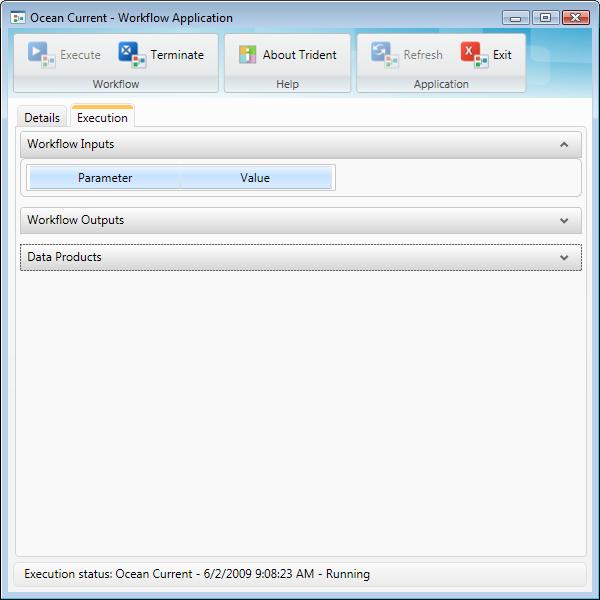


Figure 7. Windows Workflow application window

If you save a workflow as a WF Link, it automatically runs the client workflow application. An easy way to send WF links to colleagues is to install the Trident Document Add-in, which allows you to embed a WF link in a Microsoft Office Word document, along with any other relevant information. When they run the link, they connect to your Trident Registry, and then run the workflow. You can also use the add-in as a research tool, to perform tasks such as documenting the process of constructing a workflow.

When you install the add-in, the installer adds a Trident tab to the Ribbon—shown in Figure 8—that allows you to connect to a Trident Registry and embed a selected workflow in your document as a WF link. For more information, see “Trident Document Add-in User’s Guide.”

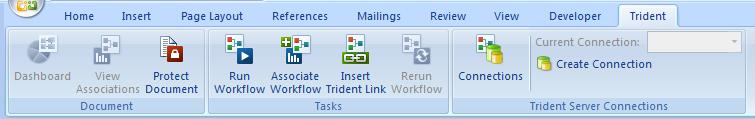


Figure 8. Trident Document Add-in

## Silverlight Workflow Applications

The Microsoft Silverlight workflow application is a Web application and is hosted on a server. It provides access to all the workflows in the associated Trident Registry. You don’t have to install Trident to run a Silverlight workflow application. You can run the application from any browser that supports Silverlight, which includes computers running operating systems other than Windows.

The Silverlight workflow application UI is very similar to a client workflow application.

# Runtime Services and Applications

Trident depends on a variety of services and applications, which run in the background to support the process of running Trident workflows. This section discusses the key services and applications.

## Executor

When you run a workflow, an executor handles the process of running the workflow. Trident has two executors: the execution service and the tray application. The executor that runs the workflow is determined by the execution mode:

* Non-interactive workflows are handled by the Trident execution service.

Services are typically more reliable than user applications. However, because they run in Session 0, non-interactive workflows cannot display a UI or save files locally.

* Interactive workflows are handled by the Trident tray application.

The tray application runs workflows in the user’s session. This allows the workflow to display a UI—such as the graph displayed by Ocean Current—or save files locally.

When you install Trident, it installs the two executors on your system and adds the tray application to your list of Startup applications. Remote execution nodes have their own executors, which run the workflow on the remote system.

## Scheduling Service

You can use Management Studio to schedule recurring jobs to run daily or weekly. The scheduling service is responsible for managing the list of scheduled jobs and ensuring that they run at the appropriate time.

The schedule itself is stored in the Trident Registry, and the scheduling service polls the Registry periodically for changes. When a new job appears, the scheduling service uses the Windows Task API to schedule the job. When the task runs, it calls a Trident API to notify the executor to run the workflow on the specified execution node.

## Blackboard Service

The blackboard service supports communication between the various Trident components, primarily for event notifications. The service is based on the Windows Communication Foundation and uses a publish-subscribe architecture; publishers send messages to the blackboard service, which distributes them to subscribers. For example, the execution service publishes data to the blackboard, and the provenance service subscribes to data from the blackboard.

## Provenance Service

The provenance service records a detailed history of each Trident job, including:

* Who ran the job, when it started, and how long it ran.
* The data associated with properties specified as workflow inputs or outputs.
* Any data products that activities store in the Trident Registry.
* A detailed diagram of the workflow, showing the activities, properties, property bindings, and so on.

The Provenance data are displayed in the Provenance working pane, or in a separate window. For details, see “Trident Composer User’s Guide” or “Trident Management Studio User’s Guide.”

## Monitoring Service

The monitoring service tracks jobs as they execute and records information such as:

* The execution status, as the workflow progresses.
* A variety of event notifications, such as when the various activities begin executing.
* CPU and memory usage.
* Inputs and outputs.

The monitoring data are displayed in real-time in the Monitor working pane or in a separate window. For details, see “Trident Composer User’s Guide” or “Trident Management Studio User’s Guide.”

# Trident Registry

The Trident Registry is a data store that represents all Trident data. It is a metadata repository for the various components that are used by Trident, including:

* Workflows
* Activities
* Workflow inputs or outputs

If you mark an activity’s property as a workflow input or output, Trident stores the value assigned to the property when the workflow runs.

* Type initializers
* Execution nodes
* Data products from workflows

You can explicitly store data in the registry as a data product.

* Scheduled jobs
* User data

For simplicity, the documentation often refers to data as being stored in the Trident Registry. Strictly, the Trident Registry acts as an abstraction layer between user applications such as Composer and a SQL Server or SQL Express database which contains the actual data. Applications add data to, or request data from, the Trident Registry, and the Registry resolves the reference and performs the required operation on the data store.

For example, when you import a custom activity into Workbench, the activity’s metadata is added to the Trident Registry, and a copy of associated DLL is placed in the database. When you run a workflow that uses the activity, the Trident Registry recovers the DLL.

Each data store has its own Trident Registry. For example, if you create a workflow and run it on your local system, you connect to the local Trident Registry. If a colleague sends you a workflow application, you must connect to your colleague’s Trident Registry on a remote system. It’s even possible to have multiple Trident Registries using the same database server. Each Registry represents a separate database of Trident components.

## External Data Sets

You can save large data sets in external storage, such as a SQL Server database on a server. You store links to the data in the Trident Registry, so workflows can access the data as required.

## Registry UI

From the user’s perspective, the contents of the Trident Registry are exposed through Composer’s Workflow Catalog and Management Studio’s Registry Manager as a set of tree-structured lists. For example, Figure 9 shows the collection of workflows as they are displayed in Registry Manager.



Figure 9. Workflows in Registry Manager

Both Composer and Management Studio connect to the same Trident Registry and can modify its contents. The UI for both applications includes “refresh” controls to synchronize the application with the Trident Registry and ensure that the view is current.

# Data Storage

Each registry is associated with an underlying database, which contains the data that is managed by the Trident Registry. Trident can use:

* A SQL Server or SQL Express database server running locally.
* A SQL Server database server running on a server or other remote computer.

A Trident Registry is simply a database, so it is possible for a server to host multiple Trident Registries. For this reason, you connect to a particular Trident Registry, not the database server itself. The converse is generally not true—a Trident Registry is associated with a single database server. The exception to that rule is that Trident allows you to store large data sets separately from the primary data store. The Trident Registry resolves the references to those data sets.

The Trident installation package creates a single Trident Registry on the SQL Server or SQL Server Express database on your system. When you run Workbench for the first time, it automatically connects to this Trident Registry. You can switch to a different Trident Registry—perhaps on a remote server—by using the Connection Manager, shown in Figure 10.

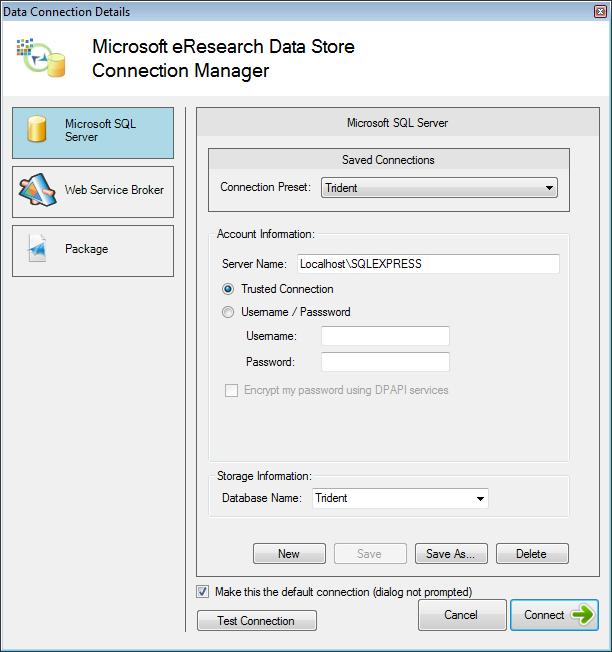


Figure 10. Data store Connection Manager

A Trident application can connect to only one Trident Registry at a time, so the contents of the Trident Registry that you see in Composer or Management Studio reflects the contents of the Trident Registry that is currently connected.

The Connection Manager is also displayed when you run a workflow application on a workflow that is located on a remote system, to allow you to connect to the associated Trident Registry.

# Execution Nodes

You can run Trident workflows locally or on remote systems. Before you run a workflow—including a client workflow application—you must specify an execution node, which is the machine name of the computer on which the workflow will run.

There are three types of execution nodes:

* Your local system
* Remote computers
* A Windows HPC Server 2008 cluster

Figure 11 shows the basic architectures.



Figure 11. Trident nodes

In the basic architectures:

Local Computer

The simplest way to run workflows is on the computer on which you have installed Trident. In addition to the applications, the installation process installs all the associated infrastructure, including a Trident Registry and an executor. When you first run Workbench, it automatically adds your computer name to the Registry’s node list, which will have only that node.

To run a workflow locally, select your machine name from the node list.

Remote Computers

Remote computers can also serve as Trident execution nodes. Each remote node is associated with a Trident Registry, so you obtain access to remote execution nodes by connecting to the appropriate Trident Registry. For example, you could send colleagues a link to your workflow and have them connect to the Trident Registry on your local system, or perhaps on a dedicated server. When they connect, that machine name automatically appears on the node list, and they can run workflows on that node.

A Trident Registry can be associated with multiple execution nodes. In that case, when you connect, you will see multiple machine names in the node list. Multiple nodes allow you, for example, to run multiple workflows simultaneously by assigning each workflow to a different node.

Windows HPC Server 2008

A more robust approach to distributing workflows across multiple nodes is to install Trident on a Windows HPC Server 2008 cluster. Instead of manually assigning each workflow to a different node, you simply assign your workflows to the cluster’s head node, which then assigns each workflow to an appropriate cluster computer.

Trident cannot distribute an individual workflow across multiple nodes. If you want to process large amounts of data in a single workflow, consider using DryadLINQ to implement your processing activities as distributed applications on the HPC cluster. For more information, see “An Introduction to Dryad and DryadLINQ,” listed in “Resources” at the end of this paper.

# Resources

This section provides links to additional information about Project Trident and related topics.

#### Project Trident references and resources

Microsoft Research Web site

<http://research.microsoft.com>

Project Trident Web site

<http://research.microsoft.com/en-us/collaboration/tools/trident.aspx>   
Provides the Project Trident package and related documentation, including:   
“Trident Composer User’s Guide”   
“Trident Management Studio User’s Guide”  
“Trident Programming Guide”  
“Trident Activity Reference”  
“Project Trident: Word Add-in for Reproducible Research: User Guide”

Project Trident on Microsoft Connect Web Page

<https://connect.microsoft.com/Trident>

#### Other references and resources

An Introduction to Dryad and DryadLINQ

Download at [http://research.microsoft.com/en-us/collaboration/tools/dryad.aspx](http://research.microsoft.com/en-us/collaboration/tools/dryad.aspx%20)

Microsoft Silverlight

<http://www.microsoft.com/silverlight/default.aspx>

My Experiment

<http://www.myexperiment.org/>

.NET Framework version 3.5

<http://www.microsoft.com/downloads/details.aspx?FamilyID=333325fd-ae52-4e35-b531-508d977d32a6&DisplayLang=en>