

Virtual Engineering Solutions, Inc.

Creating SSIS Packages

SQL Server 2005

Ground Water Protection Council's Spring Training
Meeting, Lido Beach, Florida



2008

Creating SSIS Packages

Table of Contents

Begin Developing the Control Flow.	1
Develop the Entity Data Flow	2
Develop the Address Data Flow	4
Develop the EntityType Data Flow	7
Develop the Entity Communications Data Flow	11
Test and Debug the ImportEntity.dtsx	15
PSSST, Look Here!.....	15

Creating SSIS Packages

In this session, participants will develop an SQL Server Integration Services (SSIS) package to convert data from MSOGB's classic RBDMS structure for operator data to the RBDMS.NET Entity module. To participate in this session, you must have SQL Server 2005 with SSIS and the client tools including BIDS already installed on your laptop. You will need to restore the source SQL Server (MS_RBDMS) and the destination SQL Server (now called MSRBDMS; you might wish to re-name the destination MSRBDMS_Test) to your laptop from the .bak files in <http://www.rbdmsonline.org/downloads/ms/MSRBDMSbakForSSISTraining.zip>.

SQL Server 2005 includes a graphical user interface for developing the next generation of data translation services. The application, called the Business Intelligence Development Studio (BIDS), is a special version of Visual Studio 2005 that is Microsoft's replacement for SQL Server 2000 Analysis Manager. For SSIS packages, BIDS allows users to define the steps in a data translation process and shows the defined process as an interactive flow chart each time the SSIS package is run.

Begin Developing the Control Flow

For this workshop, you will be working with the process that was used to populate the **Entity** module in MS RBDMS.NET from the classic RBDMS application. The summary "control flow" is defined by the interactive flow chart diagram screen-captured as Exhibit 1 on the right, and we will reproduce this SSIS package in this session.

1. Open BIDS from the **Start | SQL Server 2005** menu.
2. Open a new Integration Services project.
3. In the Solution Explorer, right-click Data Sources and create two data sources through the Data Source Wizard.
4. Right click in Solution Explorer to create a New SSIS Package named ImportEntity. Set the property Security | Protection level to **EncryptSensitiveWithPassword** so it can be moved to another machine at a later time. Set the password to something you can remember.
5. In the Connection Manager, right click **New Connection From Data Source**, select both connection strings, and click **OK**.

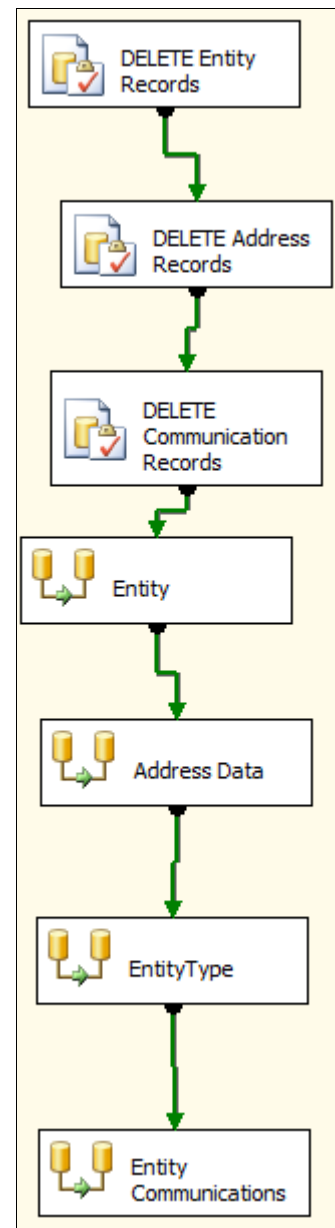



EXHIBIT 1.
Summary Control SSIS Package to Populate MS RBDMS Entity Module.

In the next few steps, we will create Execute SQL Tasks that will clear the records in the three destination tables in MSRBDMMS each time the SSIS package is run:

6. From the toolbox, drag an Execute SQL task to the workspace and re-name it **Delete Entity Records**.
7. Double click the  icon in the task you just created to open the Properties box. Under General, set the Name to **DELETE Entity Records**. Under SQL Statement, Connection, choose **MSRBDMMS**, and for SQL Statement, enter **DELETE from Entity**.
8. Repeat steps 7 and 8 for an Execute SQL task named **DELETE Address Records** from **MSRBDMMS** and enter the SQL Statement **DELETE from Address**.
9. Drag the green arrow from the DELETE Entity Records onto the DELETE Address Records to connect them in the data flow.
10. Repeat steps 7 and 8 for an Execute SQL task named **DELETE Communication Records** from **MSRBDMMS** and enter a SQL Statement **DELETE from Communication**.
11. Drag the green arrow from the DELETE Address Records onto the DELETE Communication Records to connect them in the data flow.
12. Save your work.

Develop the Entity Data Flow

The individual data flows to populate each of the tables in the RBDMS.NET **Entity** module also can be viewed separately. We will demonstrate how to translate information from tblRefCompany to Entity as a defined data flow in the next few steps. The properties of each step can be viewed or edited with a point-and-right click in the SQL Server BIDS. Our product will look like the flow shown in Exhibit 2.

1. From the toolbox, drag a Data Flow Task onto the workspace and rename it **Entity**. Drag the green arrow from the DELETE Communications Records onto the Entity Data Flow so that they are connected in the flow.

2. Double-click the new  icon on the Entity data flow task, which will open in a new tab.
3. Drag an OLE DB Source from the toolbox and rename it **tblRefCompany**.

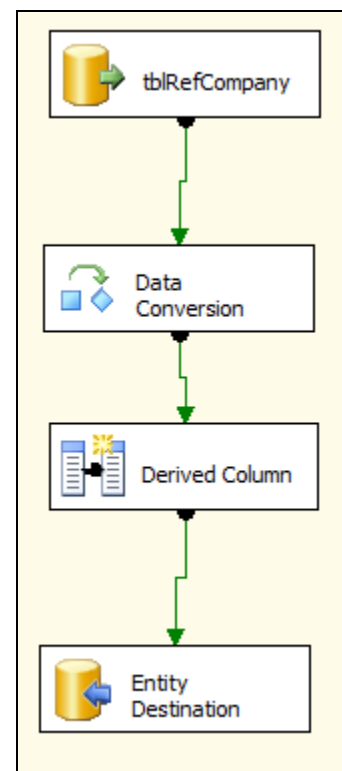







EXHIBIT 2.
Data Flow Task for Entity.

4. Double-click the  icon to open the OLE DB Source Editor box and set the OLE DB connection manager to your source **MS_RBDMS** and set the name of the table to **[dbo].[tblRefCompany]**.
5. Still in the OLE DB Source Editor, under Columns, uncheck everything except *CoNo*, *Stat*, and *CoName*. Click **OK**.
6. Next, create a data conversion task by dragging one from the toolbox to the workspace. Drag the green arrow from *tblRefCodes* onto the Data Conversion task to connect them in the flow.
7. Double-click the  icon to open the Data Conversion Transformation Editor. In the Available Input Columns box mark the *Stat* and *CoName* fields as True. Below, change the data type for each field from **Unicode string [DT_WSTR]** to **string [DT_STR]** to convert these fields to a string from nvarchar.
8. From the toolbox, drag a Derived Column task onto the workspace below the Data Conversion task. Drag and drop the green arrow from the Data Conversion task onto the Derived Column task to connect them in the flow.
9. Double click the  icon to open the Derived Column Transformation Editor. In the field under Derived Column Name, enter **EntityType**. Under Expression for *EntityType*, enter **"CO"** and set the data type to **string [DT_STR]**. Click **OK**.
10. Next, create the Entity Destination. From the toolbox, drag an OLE DB Destination to the workspace beneath the Derived Column task and rename it **Entity Destination**. Drag the green arrow from the Derived Column task onto the Entity Destination to connect them in the data flow.
11. Double-click the  icon to open the OLE DB Destination Editor. In the OLE DB Connection Manager, select your destination source **MSRBDMS**. In the Name of the table, select **[dbo].[Entity]**.
12. In Mappings in the OLE DB Destination Editor, match the input and destination columns using the dropdown boxes as follows:
 - a. Copy of *CoName* to *EntityName*
 - b. Copy of *Stat* to *Status*
 - c. *EntityType* to *EntityType*
 - d. *CoNo* to *CoNo*
13. Save your work.

Develop the Address Data Flow

Next, we'll create another data flow task for Address. The SSIS data flows also show the importance of staging the data import to begin with data from the core RBDMS tables. When the exercise is complete, the data flow will look similar to the image shown in Exhibit 3. Note how the translation and importation of address data is tied to the results of the previous SSIS package runs:

1. Return to the Control Flow (see Exhibit 1 to regain your bearings).
2. From the toolbox, drag a Data Flow Task onto the workspace below Entity and rename it **Address Data**. Drag the green arrow from the Entity data flow task onto the Address data flow task so that they are connected.

3. Double-click the  icon on the Address Data flow task to enter the workspace to begin defining the steps in the flow.

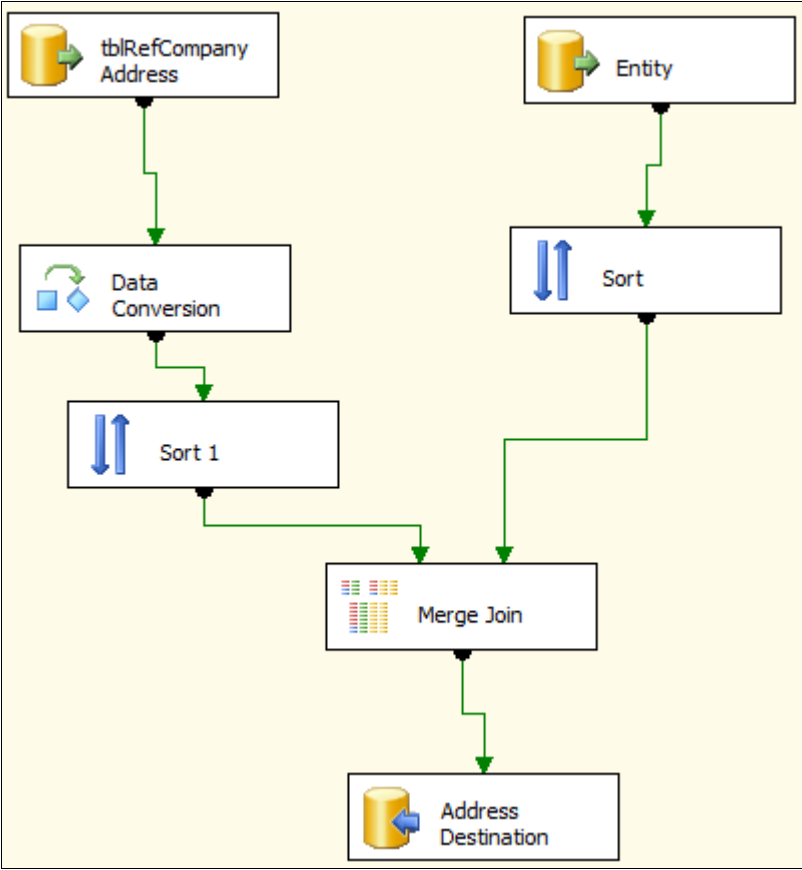








EXHIBIT 3.
Data Flow for Address.

4. Drag an OLE DB Source from the toolbox onto the workspace and rename it **tblRefCompanyAddress**.


5. Double-click the  icon to open the OLE DB Source Editor. The OLE DB connection manager should be set to your source database **MS_RBDMS**. For the Data access mode, select **SQL command** and either build or copy and paste the following SELECT statement and click **OK**:

```

SELECT      CoNo, Addr1, Addr2, City, State, Zip1 + '-'
+ ISNULL(Zip2, '') AS Expr1, Country
FROM        tblRefCompany
  
```

6. Check the Columns to make sure the external column and the output column are aligned and click OK.
7. Add a Data Conversion task to the workspace below the tblRefCompanyAddress OLE DB Source. Drag the green arrow from tblRefCompanyAddress to the Data Conversion task so that they are connected.
8. Double-click the  icon to open the Data Conversion Transformation Editor. In the top pane, make sure that all of the fields are checked except *CoNo*. In the lower pane, change the data types for each of the input columns from **Unicode string [DT_WSTR]** to **string [DT_STR]**. Then click **OK**.
9. Next add a Sort data flow transformation task below the Data Conversion task and link them by dragging the green arrow from Data Conversion to Sort.
10. Double-click the  icon to open the Sort Transformation Editor. Select the *CoNo* column for sorting and click **OK**.
11. Drag an OLE DB Source from the toolbox onto the workspace next to the one for tblRefCompany Address and rename it **Entity**.
12. Double-click the  icon to open the OLE DB Source Editor. Set the OLE DB connection manager to your destination database **MSRBDMS**. For the Data access mode, leave the default selection of **Table or view**. For the name of the table, select **[dbo].[Entity]**.
13. In the OLE DB Source Editor, click Columns and uncheck everything in the Available External Columns list except for *PKey* and *CoNo*. Click **OK**.
14. Next add a Sort data flow transformation task below the Entity Data Conversion task and link them by dragging the green arrow from Data Conversion to Sort.
15. Double-click the  icon to open the Sort Transformation Editor. Select the *CoNo* column for sorting and click **OK**.
16. From the toolbar, drag a Merge Join data transformation task onto the workspace. Drag the green arrow from the Sort task below tblRefCompanyAddress onto the Merge Join. When prompted to edit the Input Output Selection, for Input, select **Merge Join Left Input**. Repeat for the Sort task beneath Entity as **Merge Join Right Input**.
17. Double-click the  icon in the Merge Join task to open the Merge Join Transformation Editor. In the top pane for the left sort add the following inputs

by checking all of the fields except for *CoNo*. For the right Sort task, check only the *PKey*.

18. In the lower pane, locate the Input Column for the PKey and change its Output Alias to **EntityPkey**. Click **OK**.
19. End by dragging an OLE DB Destination task from the toolbar to the bottom center of the workspace. Rename the task **Address Destination**. Connect the Merge Join task to the Address Destination by dragging the green arrow from Merge Join onto Address Destination.
20. Double-click the  icon to open the OLE DB Destination Editor. In the OLE DB Connection Manager, select your destination source **MSRBDMS**. Leave the data access mode set as **Table or view**, and for the Name of the table, select **[dbo].[Address]**.
21. In the OLE DB Destination Editor, click Mappings and then match the input and destination columns using the dropdown boxes as follows and then click **OK**.
 - a. <ignore> to PKey
 - b. Entity Pkey to Entity Key
 - c. Copy of Addr1 to Address1
 - d. Copy of Addr2 to Address2
 - e. Copy of City to City
 - f. Copy of State to State
 - g. Copy of Expr1 to PostalCode
 - h. Copy of Country to Country
 - i. <ignore> to TempID
22. Save your work.

Develop the EntityType Data Flow

The EntityType data flow is the most involved component of the ImportEntity.dtsx package merely because of the repetition required to develop it. The left side sort in the diagram splits out the various entity types, while the right side sort assigns the keys in the resulting import operation. Since the individual steps are as straightforward as the others demonstrated in this exercise, the class session will track only one of the fourteen paths, which will be typical for the thirteen the others. The participants are welcome to carry on the pattern independently. The completed data flow will look like Exhibit 4 in the BIDS workspace.

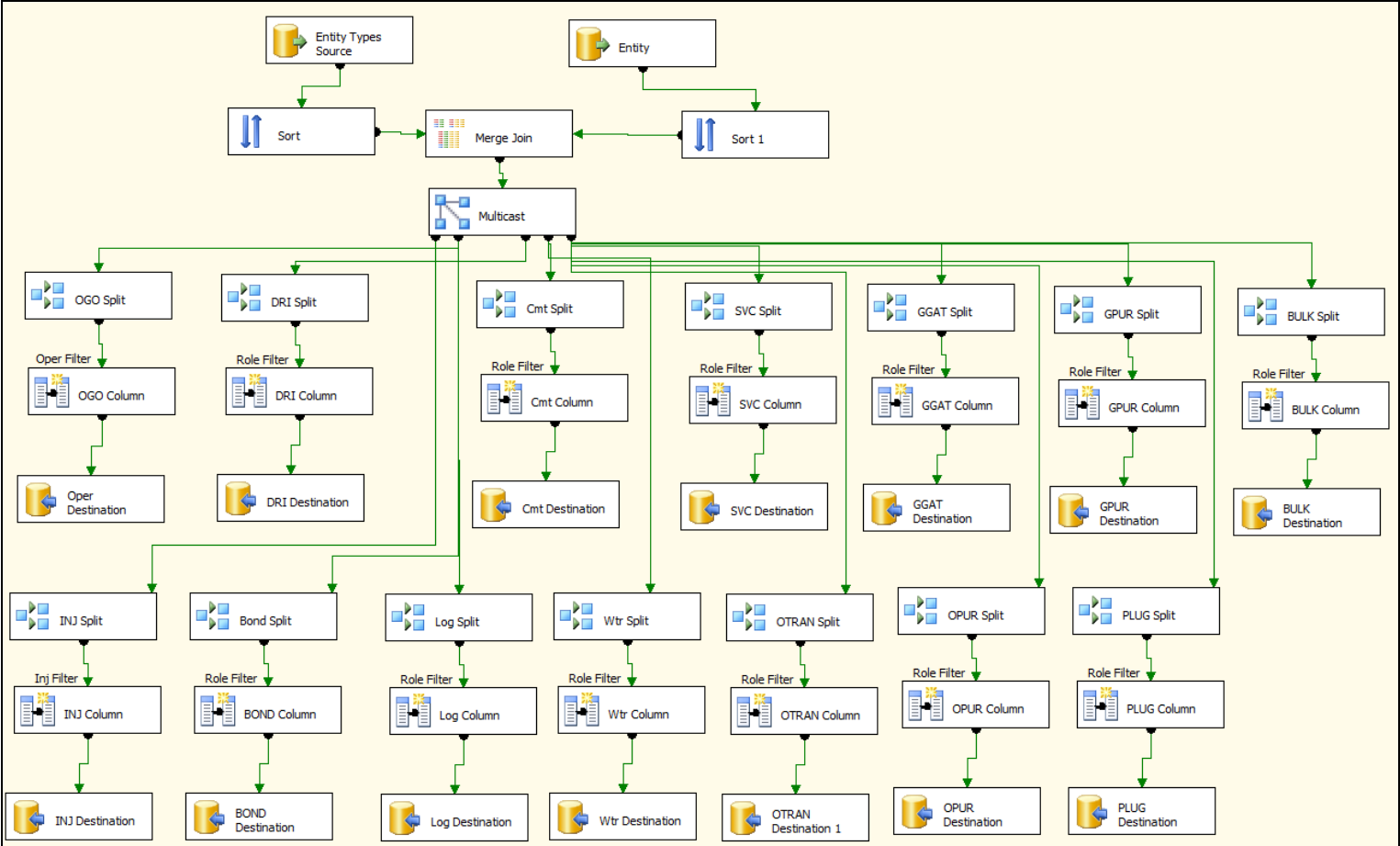




EXHIBIT 4.
The EntityType Data Flow.

1. Return to the Summary Control Flow. Add a data flow task from the toolbox to the workspace beneath Address Data. Rename the new data flow task **EntityType**. Connect Address Data to EntityType by dragging the green arrow from Address Data onto EntityType.

2. Double-click the  icon on the EntityType data flow task to enter the workspace to begin defining the steps in the flow.

3. Drag an OLE DB Source from the toolbox onto the workspace and rename it **Entity Types Source**.


4. Double-click the  icon to open the OLE DB Source Editor. The OLE DB connection manager should be set to your source database **MS_RBDMS**. For the Data access mode, select **SQL command** and either build or copy and paste the following SELECT statement:

```
SELECT      Oper, Inj_Oper, Driller, Bond_Co,
Logg_Contr, Cement_Contr, Watr_Haul, Service, Docket,
Misc2, NaturalGasOperator, Oil_Trans, Gas_Gathr,
                Oil_Purchr, Gas_Purchr,
Plugging, Bulk_, OilCompany, CoNo
FROM        dbo.tblRefCompany
```

5. Check the Columns to make sure the external column and the output column are aligned and click **OK**.

6. Double-click the  icon to open the Sort Transformation Editor. Select the *CoNo* column for sorting and click **OK**.

7. Drag another OLE DB Source from the toolbox onto the workspace to the right of the Entity Types Source and rename the new OLE DB Source **Entity**.

8. Double-click the  icon to open the OLE DB Source Editor. The OLE DB connection manager should be set to your destination database **MSRBDMS**. Leave the Data access mode set to **Table or view**, and for the Name of the table or view, select **[dbo].[Entity]**.

9. In the OLE DB connection manager, click Columns. For the Available External Columns list, uncheck all of the named columns except for *PKey* and *CoNo*. Then click **OK**.


10. Next add another Sort transformation task below Entity and link them by dragging the green arrow from Entity to Sort.



11. Double-click the  icon to open the Sort Transformation Editor. Select the *CoNo* column for sorting and click **OK**.

12. From the toolbar, drag a Merge Join data transformation task onto the workspace between the two Sort tasks. Drag the green arrow from the Sort task below Entity Types Source onto the Merge Join. When prompted to edit the Input Output Selection, for Input, select **Merge Join Left**. Repeat for the Sort task beneath Entity as **Merge Join Right**.



13. Double-click the  icon in the Merge Join task to open the Merge Join Transformation Editor. In the top pane for the left sort check all of the inputs except for *CoNo*. For the right Sort task, check only the *PKey*.


14. In the lower pane, locate the Input Column for *PKey* and change its Output Alias to **EntityPKey**. The other rows will stay the same. Click **OK**.

15. From the toolbox, drag a Multicast data flow transformation task onto the workspace below the Merge Join task. Connect the Merge Join with the Multicast by dragging the green arrow from Merge Join onto Multicast.



To this point, your work in the Entity Type data flow should look like the top portion of Exhibit 4. The next few steps will create the first of fourteen copies of the dataset we need to import the roles from the de-normalized source database to the RBDMS.NET Entity module. The pattern described below for the Oil Gas operator role is typical for the other thirteen roles represented in the classic MS RBDMS database.

1. From the toolbar, drag a Conditional Split data flow transformation task onto the lower left portion of the workspace and rename it **OGO Split**. Drag a green arrow from the Multicast task onto the OGO split task to connect them.



2. Double-click the  icon to open the Conditional Split Transformation Editor. In the lower portion of the Editor pane, for the Output Name enter **Oper Filter** and for the Condition enter **Oper == TRUE**. Change the Default output name to **Oper Default Output**. Click **OK**.

3. From the toolbar, drag a Derived Column data flow transformation onto the workspace below the OGO Split task. Drag the green arrow from the OGO Spilt onto the Derived Column to connect them. When prompted for the Input Output Selection, select **Oper Filter for the Output** and click **OK**.

4. Double-click the  icon to open the Derived Column Transformation Editor. In the lower pane of the Editor, for the Derived Column Name, enter **Role**. The Derived Column will be **<added as new column>**. For Expression, enter **“OGO”** and change the data type from **Unicode string [DT_WSTR]** to **string [DT_STR]**. Click **OK**.
5. From the toolbox, drag an OLE DB Destination task onto the workspace beneath the Derived Column task and rename it **Oper Destination**. Drag the green arrow from the Derived Column task onto the Oper Destination task to connect them.
6. Double-click the  icon to open the OLE DB Destination Editor. In the OLE DB Connection Manager, select your destination source **MSRBDMS**. Leave the data access mode set as **Table or view**, and for the Name of the table, select **[dbo].[EntityAddress]**.
7. In the OLE DB Destination Editor, click Mappings and then match the input and destination columns using the dropdown boxes as follows and then click **OK**.
 - a. **<ignore>** to PKey
 - b. **Entity PKey** to Entity Key
 - c. **<ignore>** to AddressKey
 - d. **Role** to Role
 - e. **<ignore>** to FirstName
 - f. **<ignore>** to LastName
 - g. **<ignore>** to Active
 - h. **<ignore>** to StartDate
 - i. **<ignore>** to EndDate
8. Save your work. Repeat steps 1 through 7 above for the other 13 roles (DRI, Cmt, SVC, GGAT, GPUR, BULK, INJ, BOND, Log, Wtr, OTRAN, OPUR, and PLUG).

Develop the Entity Communications Data Flow

Finally, we will create the data flow definition for importing data into the Communications table of the Entity module in RBDMS.NET. Again, the process will be somewhat complex because of the need to import data from a de-normalized structure into the more relational RBDMS.NET Entity module. The final Entity Communications data flow is shown in Exhibit 5.

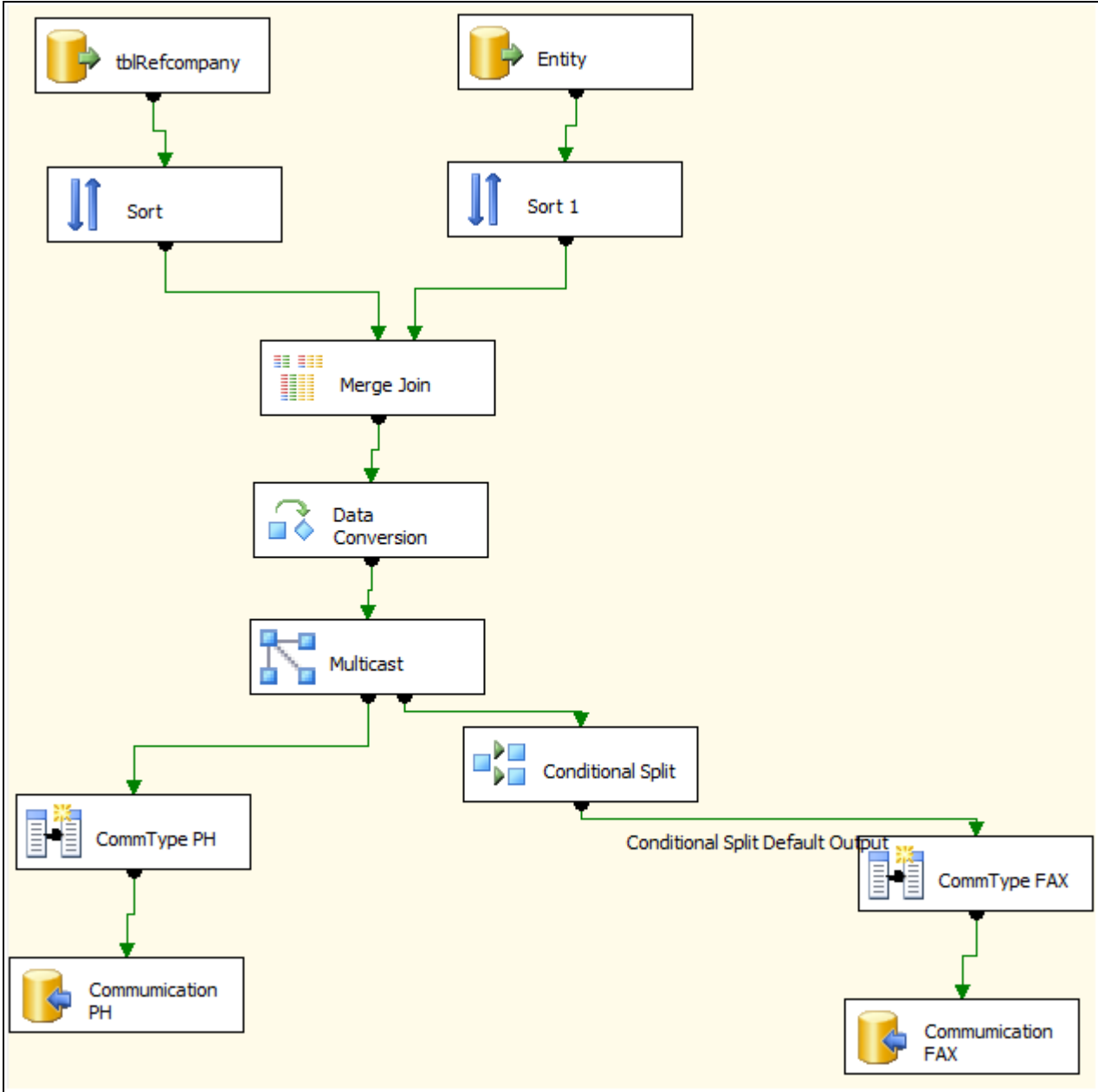




EXHIBIT 5. Entity Communications Data Flow.

1. Return to the summary Control Flow and add a data flow task from the toolbox to the workspace beneath Entity Type. Rename the new data flow task **Entity Communications**. Connect Entity Type to Entity Communications by dragging the green arrow from Entity Type onto Entity Communications.



2. Double-click the  icon on the Entity Communications data flow task to enter the workspace to begin defining the steps in the flow.
3. Drag an OLE DB Source from the toolbox onto the workspace and rename it **tblRefCompany**.



4. Double-click the  icon to open the OLE DB Source Editor. The OLE DB connection manager should be set to your source database **MS_RBDMS**. For the Data access mode, **select SQL command** and either build or copy and paste the following SELECT statement:

```
SELECT          CoNo, Phone + ISNULL(N'x' + Ph_Ext, N'')
AS PhoneExt, Fax
FROM            dbo.tblRefCompany
```


5. Check the Columns to make sure the external column and the output column are aligned and click **OK**.
6. Next, add a Sort task beneath tblRefCompany and drag the green arrow from tblRefCompany onto Sort to connect them.








7. Double-click the  icon to open the Sort Transformation Editor. Check *CoNo* and click **OK**.

8. Drag an OLE DB Source from the toolbox onto the workspace to the right of tblRefCompany and rename it **Entity**.





9. Double-click the  icon to open the OLE DB Source Editor. The OLE DB connection manager should be set to your destination database **MSRBDMS**. Leave the Data access mode set to **Table or view** and for the Name of the table, select **[dbo].[Entity]**.
10. In the OLE DB Source Editor, click Columns and in the list of Available External Columns, uncheck everything except *PKey* and *CoNo*. Click **OK**.
11. Next, add a Sort task beneath Entity and drag the green arrow from Entity onto Sort to connect them.


12. Double-click the  icon to open the Sort Transformation Editor. Check *CoNo* and click **OK**.
13. Add a Merge Join task from the toolbox below and between the two Sort tasks. Drag the green arrow from the Sort task below Entity Types Source onto the Merge Join. When prompted to edit the Input Output Selection, for Input, select **Merge Join Left**. Repeat for the Sort task beneath Entity as **Merge Join Right**.
14. Double-click the  icon in the Merge Join task to open the Merge Join Transformation Editor. In the top pane for the left sort check all of the inputs except for *CoNo*. For the right Sort task, check only the *PKey*.
15. In the lower pane, locate the Input Column for *PKey* and change its Output Alias to **EntityPKey**. Change the Output Alias for *PhoneExt* to **Phone**. The Fax row will stay the same. Click **OK**.
16. Add a Data Conversion task below the Merge Join task and use the green arrow from Merge Join to connect it to Data Conversion.
17. Double-click the  icon to open the Data Conversion Transformation Editor. In the top pane, make sure that all of the fields are checked except *EntityPKey*. In the lower pane, edit the Output Aliases by changing *Copy of PhoneExt* to **strPhone** and *Copy of Fax* to **strFax**. Change the data types for each of the input columns from **Unicode string [DT_WSTR]** to **string [DT_STR]**. Then click **OK**.
18. Add a Multicast task below the Data Conversion task and use the green arrow from Data Conversion to connect to the Multicast.
19. From the toolbar, drag a Derived Column data flow transformation onto the workspace below and to the left of the Multicast. Rename the Derived Column task **CommType PH**. Drag the green arrow from the Multicast onto the CommType PH to connect them.
20. Double-click the  icon to open the Derived Column Transformation Editor. In the lower pane of the Editor, for the Derived Column Name, enter **CommType**. The Derived Column will be **<added as new column>**. For Expression, enter **"PH"** and, for the data type, enter **string [DT_STR]**. Click **OK**.
21. Add an OLE DB Destination task to the workspace below the CommType PH and connect the two tasks with the green arrow. Rename the OLE DB task **Communication PH**. In the Properties pane, locate FastLoadOptions in the Custom Properties list and turn **off** the TABLOCK;CHECK_CONSTRAINTS option.

22. Double-click the  icon to open the OLE DB Destination Editor. In the OLE DB Connection Manager, select your destination source **MSRBDMS**. Leave the data access mode set as **Table or view**, and for the Name of the table, select **[dbo].[Communication]**.
23. In the OLE DB Destination Editor, click Mappings and then match the input and destination columns using the dropdown boxes as follows and then click **OK**.
- <ignore> to PKey
 - Entity PKey to Entity Key
 - CommType to CommType
 - strPhone to CommAddress
 - <ignore> to TempID
 - <ignore> to Comment


Save your work. Then begin again at the Multicast task you added above Comm Type PH.

1. From the toolbar, drag a Conditional Split data flow transformation task onto the lower portion of the workspace to the right of CommType PH. Drag a green arrow from the Multicast task onto the Conditional Split task to connect them.
2. Double-click the  icon to open the Conditional Split Transformation Editor. In the lower portion of the Editor pane, for the Output Name enter **Case 1** and, for the Condition, enter **(ISNULL(strFax))**. Click **OK**.
3. From the toolbar, drag a Derived Column data flow transformation onto the workspace below and to the right of the Conditional Split. Rename the Derived Column task **CommType FAX**. Drag the green arrow from the Conditional Split onto the CommType FAX to connect them. When prompted for the Input Output Selection, for Output, select **Conditional Split Default Output** and click **OK**.
4. Double-click the  icon to open the Derived Column Transformation Editor. In the lower pane of the Editor, for the Derived Column Name, enter **CommType**. The Derived Column will be **<added as new column>**. For Expression, enter **"FAX"** and, for the data type, select **string [DT_STR]**. Click **OK**.
5. Add an OLE DB Destination data flow task below CommType FAX and rename the OLE DB Destination to **Communication FAX**. Drag the green arrow from CommType FAX to Communication FAX to connect them. In the Properties

pane, locate FastLoadOptions in the Custom Properties list and turn **off** the TABLOCK;CHECK_CONSTRAINTS option.

6. Double-click the  icon to open the OLE DB Destination Editor. In the OLE DB Connection Manager, select your destination source **MSRBDMS**. Leave the data access mode set as **Table or view**, and for the Name of the table, select **[dbo].[Communication]**.
7. In the OLE DB Destination Editor, click Mappings and then match the input and destination columns using the dropdown boxes as follows and then click **OK**.
 - a. <ignore> to PKey
 - b. Entity PKey to Entity Key
 - c. CommType to CommType
 - d. strFax to CommAddress
 - e. <ignore> to TempID
 - f. <ignore> to Comment
8. Save your work.

Test and Debug the ImportEntity.dtsx

When you are ready to test and debug your SSIS package, click the  **Start Debugging** button from the application toolbar. As your various tasks are validated and completed, their flow chart images will change background color.

To view the imported data, open your destination database in SQL Server Management Studio and open the various tables (Entity, EntityAddress, Address, and Communication) with which you have been working.

PSST, Look Here!

If you need to check the answers “in the back of the book,” the training disk provided to participants includes both this document and a sample BIDS solution with the results of this exercise.