

Business Intelligence Solution Providers Specialized in creating talent resource pool

## IBM Data Stage Lab Guide#13

### Loading Unique Records from Flat Files to SQL Server

### **Description:**

BISP is committed to provide BEST learning material to the beginners and advance learners. In the same series, we have prepared a list of beginner's guide and FAQs for IBM Data Stage. We have built complete financial Data Model and various data transformation techniques. Download many such learning documents, student guide, Lab Guide and Hands-on practice materials. This document guide you through how to Load sorted data into SQL Server from Flat File. Join our professional training to learn from Experts.

History:

Version 0.1 0.1 Description Change Initial Draft Review#1 Author Varun Khare Amit Sharma Publish Date 12th Aug 2012 18<sup>th</sup> Aug 2012

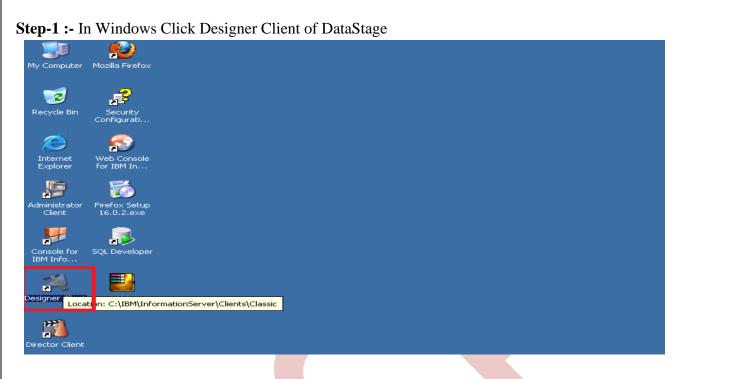
www.bispsolutions.com

www.bisptrainings.com 1

# **Mapping Sheet**

Unique Value Load. Source- Flat File Target- Sql Server

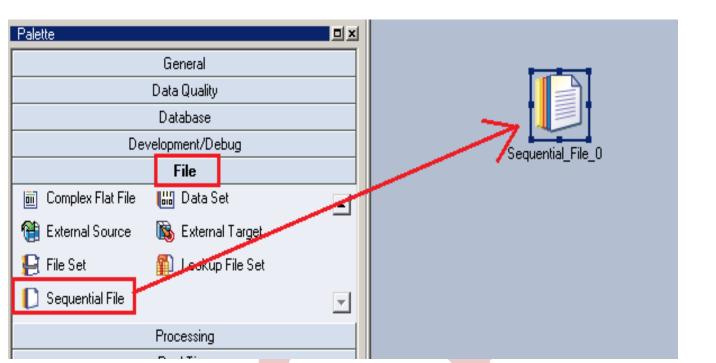
Transformation	Unique Value Load					
Source Name	Flat File					
Target Name	Sql Server					
Flat File	Source Details		u		SQL SERVER Details	
Entity Name	Data Types	ls Null	Expression	Destination Entity Na	DestinationField Name	ls Null
BRANCH_CD	NUMBER(5)	N	l	sa	BRANCH_CD	Ν
CREATED_BY	VARCHAR2(30)	у		sa	CREATED_BY	у
CREATION_DATE	TIMESTAMP(6)	у		sa	CREATION_DATE	у
DEFINITION_LANGUAGE	VARCHAR2(4)	у	Load Unique Value into Sql	sa	DEFINITION_LANGUAGE	у
ENABLED_FLAG	VARCHAR2(1)	у	Server from Flat File	sa	ENABLED_FLAG	у
LAST_MODIFIED_BY	VARCHAR2(30)	у		sa	LAST_MODIFIED_BY	у
LAST_MODIFIED_DATE	TIMESTAMP(6)	у		sa	LAST_MODIFIED_DATE	у
LEAF_ONLY_FLAG	VARCHAR2(1)	у		sa	LEAF_ONLY_FLAG	у
BRANCH_DISPLAY_CD	VARCHAR2(10)	у		sa	BRANCH_DISPLAY_CD	у



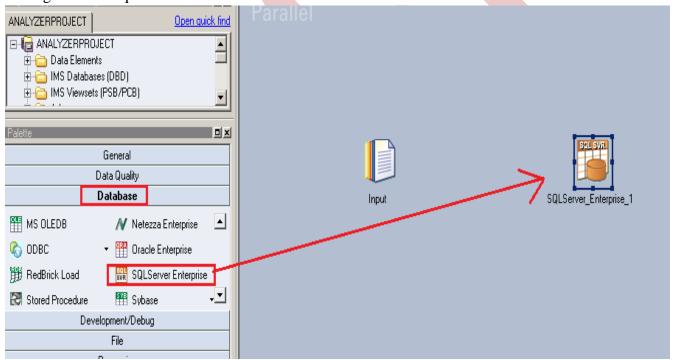
**Step-2 :-** New Window opens, then click Jobs ,then click Parallel Job because we are using parallel jobs.

WebSphere DataStage and O	QualityStage Designer
Eile ⊻iew <u>R</u> epository Import E≥	port Tools Help
🏮 - 🧭 🖬 🕼 🖪 🏁 🗊	
ANALYZERF Open quick find ANALYZERF Open quick find ANALYZERPROJECT Data Elements Data Elements IMS Databases (D IMS Viewsets (PS	New
Palette	Control Stage Types Control Other Control Assistants
Favorites This group is initially empty and is intended to contain shortcuts to commonly used repository items. Shortcuts can be created by	Jobs
dragging items from the Repository Tree or by using the Customize option from the menu.	OK Cancel Help

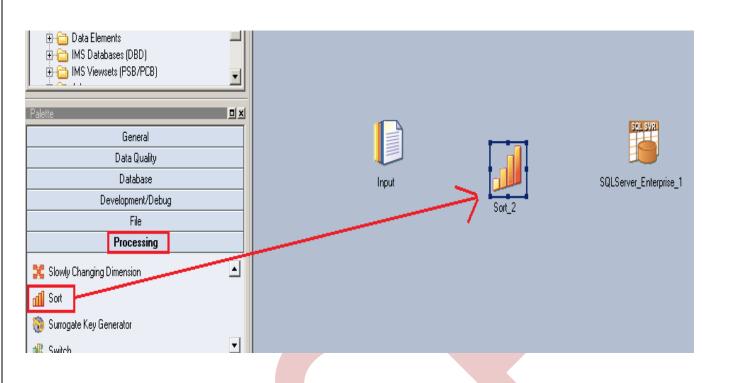
**Step-3 :-** First go on File Palette then scroll the cursor and choose Sequential File and drag it to Parallel pane and rename it. This file is used for input data.



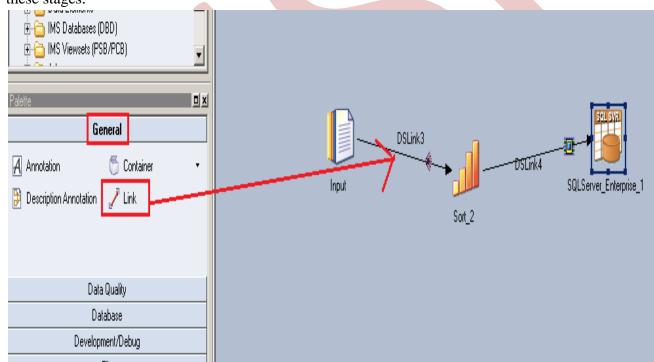
**Step-4 :-** Choose Sql Server Enterprise from Database Palette and drag it to parallel pane. This is used for loading data into sql server from flat files.



**Step 5 :-** Choose an sort stage from processing palette. This sort process is used when we want to sort any number of values according to ascending or descending order than we apply this on a particular column and output is based on key column in which we apply sorting we get result.



**Step-6:-** Now make an connection with these stages that's why we have to use this link stage for connecting these stages.



**Step-7**: Now first before loading data into output, we have to create a table in sql server. So, we write a create table command and make sure to give same column name as you have in flat file. Goto SQL tab than write query and remember the username and password

```
CREATE TABLE DBO.SRC_BRANCH_CD
(
BRANCH_CD NUMERIC(5,0) PRIMARY KEY,
CREATED_BY VARCHAR(30),
CREATION_DATE TIMESTAMP6),
DEFINITION_LANGUAGE VARCHAR(4),
ENABLED_FLAG VARCHAR(1),
LAST_MODIFIED_BY VARCHAR(30),
LAST_MODIFIED_DATE TIMESTAMP (6),
LEAF_ONLY_FLAG VARCHAR(1),
BRANCH_DISPLAY_CD VARCHAR(10) DEFAULT NULL
);
```

Step-8:- Double click on Input file and browse an input file that can be .txt,.csv any test file.

rallel	INPUT - Sequential File	
	Stage       Qutput         Output name:       DSLink2         General       Properties         File       Columns         Adyanced         Source         Source <tr< th=""><th>Columns       View Data         File:       Switch to multiline editor         \Creditcard\SRC_BRANCH_CD_UNIQ.txt       Switch to multiline editor         Information:       Insert job parameter         Type: Pathname       Prowse for file         Name of a file that data will be read from.       se_3         Available properties to add:       V         OK       Cancel       Help</th></tr<>	Columns       View Data         File:       Switch to multiline editor         \Creditcard\SRC_BRANCH_CD_UNIQ.txt       Switch to multiline editor         Information:       Insert job parameter         Type: Pathname       Prowse for file         Name of a file that data will be read from.       se_3         Available properties to add:       V         OK       Cancel       Help

Step-9 Go to Format tab and Remove double Quote because we don't need them then Click on OK. Note:- If you use date as a input than specify here which type of format you used

Stage Output	
Output name: DSLink11	Colu <u>m</u> ns ⊻iew Data
Properties:	[No property selected]         Properties that apply to date-type fields unless overridden at the field level.         Available properties to add:         Image: Second sec
	Load De <u>f</u> aults ►

Step-10 :- Here I'm using this date format so we specify here than we able to load the date from flat file.

🚺 Sequential_File_7 - Sequential File	
Stage Output	
Output name: DSLink11	Colu <u>m</u> ns <u>V</u> iew Data
<u>G</u> eneral <u>Properties</u> Fo <u>r</u> mat <u>C</u> olumns Advanced	
Properties:	Format string         %dd-%mm-%yyyy         The string format of a date. By default, the format of the date string is %yyyy-%mm-%dd.         Available properties to add:
	Load Defaults >
	OK Cancel <u>H</u> elp

**Step-11:-** Go to Columns tab and enter same column names as declared in Input file and put datatype, length than click on View Data. This will show your Input Data.

ut name: DSLink2 neral   <u>P</u> roperties   Fo <u>r</u> mat [ <u>C</u> ol	umns 🗛	d <u>v</u> anced	Colu	<u>m</u> ns		⊻iew Dal	ia
Column name	Key	SQL type	Extended	Length	Scale	Nullable	script
BRANCH_CD		Integer		10		No	
CREATED_BY		VarChar	Unicode	255		No	
CREATION_DATE		VarChar	Unicode	255		No	
DEFINITION_LANGUAGE		VarChar	Unicode	255		No	
ENABLED_FLAG		VarChar	Unicode	255		No	
LAST_MODIFIED_BY		VarChar	Unicode	255		No	
LAST_MODIFIED_DATE		VarChar	Unicode	255		No	
LEAF_ONLY_FLAG		VarChar	Unicode	255		No	
BRANCH_DISPLAY_CD		Integer		10		No	

**Step-12:-** After Clicking on View Data if your data shows like this than our Data is valid and if some error comes than we have to rectify them and check our column names.

BRANCH_CD CREA	ATED_BY	CREATION_DATE		DEFINITION_LANG -						
18		11-JUN-10 05.22.42.000000 P	M	US						
17		11-JUN-10 05.22.42.000000 P	M	US		1		<u> </u>		_
16		11-JUN-10 05.22.42.000000 P	Ħ	US		Colu	mns		View Dat	ð
15		11-JUN-10 05.22.42.000000 P	M	US	Level 1					
14		11-JUN-10 05.22.42.000000 P	m	US	ced					
13		11-JUN-10 05.22.42.000000 P	M	US	SQL type	Extended	Length	Scale	Nullable	script
12		11-JUN-10 05.22.42.000000 P	M	US	eger		10		No	
11		11-JUN-10 05.22.42.000000 P	M	US	Char	Unicode	255		No	
10		11-JUN-10 05.22.42.000000 P	M	US	Char	Unicode	255		No	
9		11-JUN-10 05.22.42.000000 P	M	US	Char	Unicode	255		No	
8		11-JUN-10 05.22.42.000000 P	m	US	Char	Unicode	255		No	
7		11-JUN-10 05.22.42.000000 P	M	US	Char	Unicode	255		No	
6		11-JUN-10 05.22.42.000000 P	M	US	Char	Unicode	255		No	
5		11-JUN-10 05.22.42.000000 P	M	US	Char	Unicode	255		No	
4		11-JUN-10 05.22.42.000000 P	M	US	eger		10		No	
3		11-JUN-10 05.22.42.000000 P	M	US	_		_	-	_	
2		11-JUN-10 05.22.42.000000 P	M	US						•
6		11-JUN-10 05.22.42.000000 P	m	US						
0		11-JUN-10 05.22.42.000000 P	M	US			Sav	/e	Load	1
19		11-JUN-10 05.22.42.000000 P	H	US				_		
20		11-JUN-10 05.22.42.000000 P	M	US 🔹			1 .		1	
d in the second se						OK	C	ancel	100	Help

**Step-13:-** After choosing file, select 'True' from First line in column Names dropdown for removing first line from Input file.

Output         name:       DSLink2         I       Properties         Format       Columns         Advanced         Source         File = C:\Creditcard\SRC_BRANCH         First Line is Column Names:         True         False		
name: DSLink2 ▼ Columns View Data al Properties Format Columns Advanced Source Source File = C:\Creditcard\SRC_BRANCH Read Method = Specific File(s) Options First Line is Column Names = True False Type: List If set true, the first line of a file contains column names on writing and is ignored on reading. Report Progress = Yes	r - Sequential File	×
al       Properties       Format       Columns       Advanced         Source       Source       File = C:\Creditcard\SRC_BRANCH       First Line is Column Names:         Source       First Line is Column Names:       True         Options       First Line is Column Names = True       False         Source       First Line is Column Names:       True         Source       First Line is Column Names:       True         Source       First Line is Column Names:       True         Source       Source       Source       Source         Source       Source       Source       Source         Source       Source       Source       Source       Source         Source       Folder       Source       Source       Source         Source       First Line is Column Names:       True       Source         Source       Source       Source       Source       Source         Source       Source       Source       Source <td< td=""><td><u>O</u>utput</td><td></td></td<>	<u>O</u> utput	
<ul> <li>Source</li> <li>File = C:\Creditcard\SRC_BRANCH</li> <li>Read Method = Specific File(s)</li> <li>Options</li> <li>First Line is Column Names = True</li> <li>False</li> <li>Full</li> <li>Folder</li> <li>Folder</li> <li>Folder</li> <li>Folder</li> <li>First Line is Column Names:</li> <li>Folder</li> <li>F</li></ul>	name: DSLink2	Colu <u>m</u> ns   <u>V</u> iew Data
<ul> <li>Source</li> <li>File = C:\Creditcard\SRC_BRANCH</li> <li>Read Method = Specific File(s)</li> <li>Options</li> <li>First Line is Column Names = True</li> <li>False</li> <li>Full</li> <li>Folder</li> <li>Folder</li> <li>Folder</li> <li>Folder</li> <li>First Line is Column Names:</li> <li>Folder</li> <li>F</li></ul>	ral Properties Format Columns Advanced	
	<ul> <li>File = C:\Creditcard\SRC_BRANCH</li> <li>Read Method = Specific File(s)</li> <li>Options</li> <li>First Line is Column Names = True</li> <li>Keep File Partitions = False</li> <li>Missing File Mode = Depends</li> <li>Reject Mode = Continue</li> </ul>	True False True Type: List If set true, the first line of a file contains column names on writing and is ignored on reading.
		OK Cancel <u>H</u> elp
OK Cancel <u>H</u> elp		
OK Cancel <u>H</u> elp		
- Make input file and remember column names.	CREATED_BY,CREATION_DATE,DEFINITION_LANGUAGE,ENABLEI 11-JUN-10 05.22.42.00000 PM,US,Y,RTUSER,11-JUN-10 1 11-JUN-10 05.22.42.00000 PM,US,Y,RTUSER,11-JUN-10 1 1-JUN-10 05.22.42.00000 PM,US,Y,RTUSER,11-JUN-10 1 1-JUN-10 05.22.42.00000 PM,US,Y,RTUSER,11-JUN-10 0 1-JUN-10 05.22.42.00000 PM,US,Y,RTUSER,11-JUN-10 0 1-JUN-10 05.22.42.000000 PM,US,Y,RTUSER,11-JUN-10 0 11-JUN-10 05.22.42	05.22.42.00000 PM, Y, 17, 05.22.42.00000 PM, Y, 15, 05.22.42.00000 PM, Y, 14, 05.22.42.00000 PM, Y, 14, 05.22.42.00000 PM, Y, 12, 05.22.42.00000 PM, Y, 11, 05.22.42.00000 PM, Y, 11, 05.22.42.00000 PM, Y, 9, 05.22.42.00000 PM, Y, 8, 05.22.42.00000 PM, Y, 14, 05.22.42.00000 PM, Y, 19, 05.22.42.00000 PM, Y, 19, 05.22.42.00000 PM, Y, 21, 05.22.42.00000 PM, Y, 22, 05.22.42.00000 PM, Y, 23, 05.22.42.00000 PM, Y, 24, 05.22.42.00000 PM, Y, 24, 05.22.42.00000 PM, Y, 26, 05.22.42.00000 PM, Y, 26, 05.22.42.00000 PM, Y, 27,

**Step-15:-** Now double click on sort stage and see here there are some field showing like key field here we have to specify in which column we want to apply sorting so we select here and we also specify in which order we want our target data.

Parallo	Key: BRANCH CO BRANCH CO BRANCH DISPLAY_CD CREATED_BY CREATION_DATE DEFINITION_LANGUAGE ENABLED_FLAG LAST_MODIFIED_DATE LEAF_ONLY_FLAG Case Sensitive Nulls Position Sort as EBCDIC
	OK Cancel <u>H</u> elp

**Step-16:-** Now click to input and choose hash key partition because these are key based stage so we have to choose hash and check on unique key with the help of this only unique values will store in sql server.

1 Sort_4 - Sort	
Stage Input Dutput	
Input name: DSLink2	Colu <u>m</u> ns
General Partitioning Columns Advanced	
Partitioning / Collecting	Sorting
Partition type: Hash	✓ Perform sort     Stable       ✓ Unique
Available:	Selected:
BRANCH_DISPLAY_CD	Kev Usage Option 🛧
CREATED_BY CREATION_DATE	T BRANCH_CD Sorting, Partitioning Ascer
ENABLED_FLAG LAST_MODIFIED_E LAST_MODIFIED_DATE	<b>₽</b>
LEAF_ONLY_FLAG	±
	OK Cancel <u>H</u> elp

Step-17:- Click on Input and see all columns are available or not

	iame: DSLink2 ral P <u>a</u> rtitioning <u>Columns</u>	vhA i	 anced		Colu <u>m</u> ns.		]	
	Column name	Key	SQL type	Extended	Length	Scale	Nullable	Description
1	BRANCH CD		Integer		10		No	
2	CREATED_BY		VarChar	Unicode	255		No	
3	CREATION_DATE		VarChar	Unicode	255		No	
4	DEFINITION_LANGUAG		VarChar	Unicode	255		No	
5	ENABLED_FLAG		VarChar	Unicode	255		No	
6	LAST_MODIFIED_BY		VarChar	Unicode	255		No	
7	LAST_MODIFIED_DATE		VarChar	Unicode	255		No	
3	LEAF_ONLY_FLAG		VarChar	Unicode	255		No	
Э	BRANCH_DISPLAY_CD		Integer		10		No	
T								Þ
-								

**Step-18** Now go to output tab than select all links and drag them into OutLink for sending data which satisfies filter condition into output links..

	name: DSLink6	Colu <u>m</u> ns
<u>à</u> ene	aral Mapping Columns Advanced	Drag all columns into output tab
	Columns	DSLink6
	Expression Column N	Derivation Col
	DSLink2.BRANCH_CC BRANCH_	DSLink2.BRANCH_CD BRA
	DSLink2.CREATED_B CREATED	DSLink2.CREATED_BY CRE
	DSLink2.CREATION_[ CREATION	DSLink2.CREATION_DATE CRE
	DSLink2.DEFINITION_DEFINITIC	DSLink2.DEFINITION_LANGU/ DEF
	DSLink2.ENABLED_FI ENABLED	DSLink2.ENABLED_FLAG EN/
<b>T</b>	DSLink2.LAST_MODIF LAST_MO	DSLink2.LAST_MODIFIED_BY_LAS
		·
-		Eind Auto-Mate

**Step-20:-** Now double click on Sql Server enterprise stage then it looks like below screenshot. Here we have to enter our table name where we want to insert our data. Multiple options are there for write mode dropdown. If we directly want to load data than we simple used write method as load otherwise we manually

put queries on it. Also we have to specify username and password and this password should match with oracle username and password. Below we specify Replace as write Mode that means if already data is present in table than replace the data with new data.

rt Tools Window Hel	🔐 SQLServer_Enterprise_1 - SQLServer Enterprise		
Parallel	Stage Input Input name: DSLink4	Colu <u>m</u> ns Write Mode: Append	
Sequential	<ul> <li>♦ Write Method = Write</li> <li>♦ Write Mode = Append</li> <li>♥ Data source = ?</li> <li>♥ Password = ?</li> <li>♥ User = ?</li> <li>♥ Options</li> <li>♥ Truncate Column Names = False</li> </ul>	Append Freplace Truncate Replace by dropping table and creating new; Truncate table records only. Available properties to add:	<b>_</b>
SOLServer Enterprise 1		OK Cancel	<u>H</u> elp

Step-21:- Click on column tab and see all the columns are showing or not.

General Properties Partitioning Columns Advanced								
	Column name	Key	SQL type	Extended	Length	Scale	Nullable	escriptio
1	BRANCH_CD		Integer		10		No	
2	CREATED_BY		VarChar	Unicode	255		No	
3	CREATION_DATE		VarChar	Unicode	255		No	
4	DEFINITION LANGUAGE		VarChar	Unicode	255		No	
5	ENABLED_FLAG		VarChar	Unicode	255		No	
6	LAST_MODIFIED_BY		VarChar	Unicode	255		No	
7	LAST_MODIFIED_DATE		VarChar	Unicode	255		No	
8	LEAF_ONLY_FLAG		VarChar	Unicode	255		No	
9	BRANCH_DISPLAY_CD		Integer		10		No	
•						Save		•

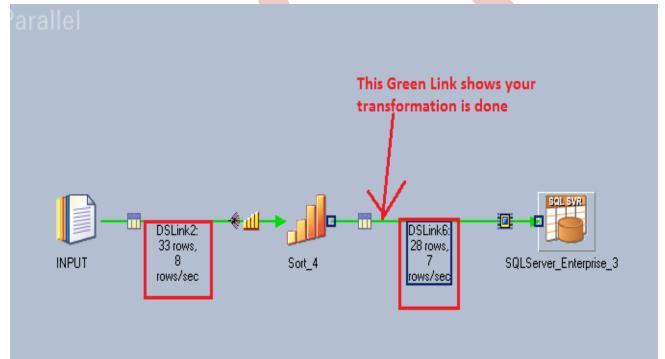
**Step-22:-** Now here we have to specify the table name and schema name and make an ODBC connection. System DSN and all the other entries should be same as where your table is situated and whatever you specify in System DSN.

QLServer_Enterprise_3 - SQLServer	Enterprise		
nput name: DSLink6		Colu <u>m</u> ns	
General       Properties       Partitioning       Column         Target       Table = SRC_BRANCH_CD_I         Table = SRC_BRANCH_CD_I         Write Method = Write         Write Mode = Replace         Connection         Partitioning         Data source = mas         Partitioning         Partitioning         Connection         Partitioning         Partitioning         Connection         Partitioning         Parti		Table: SRC_BRANCH_CD_UNI Information: Type: String Name of the table, if reading from it you must have SELECT privileges on it. Available properties to add:	
	_		
•23:- Now compiling the job, click of ⓐ  ▷ []] []] []] []] []] []] []] []] []] [			
		utton.	

**Step-24:-** If this is shown that means your compilation is done otherwise it shows error. Now click on run button or (CTRL+F5).

🌂 Compile Job - load1	×
Compilation Status	<u>R</u> e-Compile
Job successfully compiled with no errors.	<u>S</u> how Error
	More
	<u>C</u> lose
	<u>H</u> elp

Step-26:- After click on run wait for a while than it shows GREEN line that means your transformation is successfully done otherwise if it shows RED Line that means not Done and BLUE Line means Under Process.



**Step-27:-** Now Goto SQL Console and connect with the same login credentials as I mentioned in sql server enterprise stage sa/password123 so you can choose your own username and password. Before loading, remember to check structure of this file is available and then you should be able to load the data, then simple query:

<SELECT \* FROM SRC\_BRANCH\_CD\_UNI;>

#### MITESHA.mast...QLQuery1.sql\* Summary

SELECT \* FROM SRC BRANCH CD UNI;

¢.	)										
🔜 Results 👔 Messages											
	BRANCH_CD	CREATED_BY	CREATION_DATE	DEFINITION_LANGUAGE	ENABLED_FLAG	LAST_MODIFIED_BY	LAST_MODIFIED_DATE	LEAF_ONLY_FLAG	BRANCI		
1	0		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	0		
2	2		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	2		
3	3		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	3		
4	4		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	4		
5	5		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	5		
6	6		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	6		
7	7		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	7		
8	8		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	8		
9	9		11JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11-JUN-10 05.22.42.000000 PM	Y	9		
10	10		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	10		
11	11		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	11		
12	12		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	12		
13	13		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	13		
14	14		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	14		
15	15		11-JUN-10 05.22.42.000000 PM	US	Y	RTUSER	11JUN-10 05.22.42.000000 PM	Y	15		

www.bispsolutions.com

•