

The Design of Digital Campus Shared Data Center Based on ETL

Jian Zhao^{1, a}, Xiaofeng Wang^{2, b}, Puhong Tang^{1, c}

¹Academic Administration, Jiaxing Vocational Technical College, Jiaxing, 314036, China

²Library and Information Center, Jiaxing Vocational Technical College, Jiaxing, 314036, China

^aemail: 260462422@qq.com, ^bemail:121478825@qq.com, ^cemail:tangpuhong@126.com

Keywords: ETL; Shared Data Center; DTS; Synchronization

Abstract. In order to build digital campus, integrate data resources, and realize data sharing, a scheme of building shared data center to realize data integration based on ETL is put forward. The structure of shared data center is introduced, and the design of ETL between business systems and shared data center is discussed.

Introduction

Extraction, transformation and loading (ETL) is the core process of data integration [1]. It is a technology that integrates the sparse and disorderly data of different standards and codes into the target database. There are three steps: extracting data from variety of original business systems, then transforming data according to certain rules, finally loading the transformed data into the target database as planned [2]. With the expansion of school scale, requirements of the information construction in colleges and universities are increasingly high. The management systems have been used one after another at various functional departments in the process of information construction. With the rapid development of application in the system, management efficiency has been greatly improved, but the problem has been arising at the same time. Owing to the limitation of use or development time and funds and purpose, different business systems are in a state of independent, which makes unable to achieve better data exchange and data sharing between the various business units, resulting in the formation of “islands of information” between the departments. Therefore, integrating different business systems efficiently to realize the integration of management and sharing of resources by constructing digital campus has become the focus of the construction of University information.

The architecture of shared data center

The shared data center is a platform that can collect, store and process all kinds of shared data, at the same time, it provides information sharing service for all. Shared data center is a comprehensive technology scheme based on the data management and use. It can not only store large amounts of data and manage data efficiently to realize the integration and data sharing between various systems, but also provide data access methods to ensure the timeliness, integrity and consistency of data.

The benefits of using Shared data center for data integration is that, the original systems of various departments still keep running independently, and business database and Shared database communicate with each other by XML interface. The shared data just as information of students and teachers is loaded into shared database, and data sharing between various business departments is based on data subscription to shared data center, without getting data from certain department by hand or mapping [3]. The data belonged to shared data center comes from authoritative department. Every data has only one maintainer, so the accuracy and authority of data can be ensured. Therefore, building shared data center is a better scheme of data integration.

The structure of shared data center is shown in Fig. “1”.

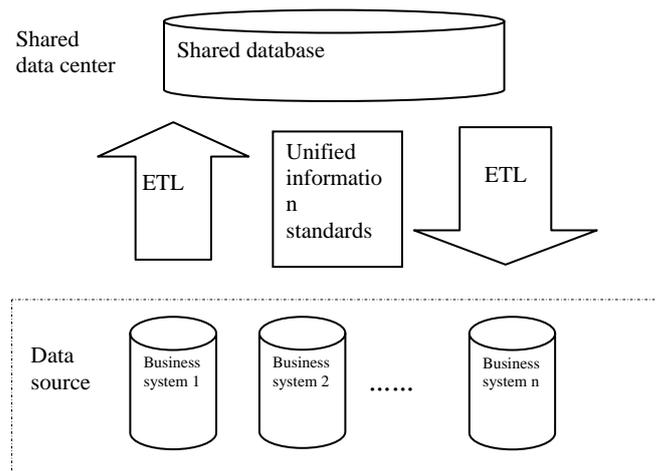


Figure 1 The structure of shared data center

The design of ETL

1. The ETL of historical data

At the beginning of the establishment of shared data center, the public data of business systems should be extracted to the shared data center as historical data. These ETL processes require appropriate ETL tools. Self-developed tools can meet the special demands of the project but with long period, poor universality and low efficiency. The General ETL tools focus only on business logic without extra development time, but usually costly [4]. On comprehensive comparison on ETL tools, we finally choose Microsoft SQL Server 2000's ETL tool DTS (Data Transformation Services) to realize the ETL process of historical data from business systems to shared data center [5].

As a tool of SQL Server, DTS can support ODBC or JDBC data interface. DTS is mainly used for data transfer between different data sources. It has a graphical interface and is also a programmable object. The source database and target database supported by DTS are all based on OLE DB, including SQL Server, ORACLE, DB2, SYBASE, Access, Excel and so on. The transformation forms include import, export and exchange of the data. During the transformation, the data with different names, sizes, types and precision can be copied among various databases by specifying different transformation rules.

In realizing ETL through DTS, first create a DTS package, then choose the needed source database and target database. When the business system connects the shared data center, create a "data conversion task" and set its configurations, then extract the data from the business system and import them into the data preparation area.

"Conversion data task" is mainly used to copy data between source database and target database. It can realize not only data copying and transformation among various databases that support OLE DB but also most fundamental data extraction in DTS by optionally applying column level transformation to data.

The configuration steps of "Conversion data task" are as follows:

- 1) Establish connection.
- 2) Set the attributes of source data and objective data
- 3) Set the attributes of transformation

The default conversion mapping is a direct data copy. If there is heterogeneous between source data and target data, the attributes of transformation should be set, such as rematching the source column and target column, or choosing another type of conversion.

If the field cannot be matched directly, scripting language can be written to solve the problem. For example, the value of field "xb" from "xsjbx" (the table of student's basic information) of educational administration system is "male" or "female", but the value of field "XB" from "stu_information" in the shared data center is "1" or "0".

Following codes can be entered into the code box:

```
Function Main ()
if DTSSource (“xb” )=“male” then
DTSDestination (“XB”) =“1”
Else DTSDestination (“XB”) =“0”
end if
Main=DTSTransformStat_OK
End Function
```

The students' gender can be transformed into gender code “1” or “0” in this way before it is loaded into the shared database.

2.ETL of real-time synchronous data

It is easy to realize the ETL process by DTS for historical data, but in the case of high real-time requirements, this is inappropriate. For example, student's basic information in educational administration management system has changed, but the relevant data of shared data center can be updated only after the execution of DTS package, which usually takes place only in some particular time. If the student's basic information is extracted from the shared data center by any business system before the update, the data will be inaccurate. Therefore, real-time synchronization of data should be considered. If the shared data of authorities change, these changes should be submitted to the shared database timely to ensure the accuracy and completeness of shared data. Meanwhile, these changes should be updated to the business departments who have subscribed the shared data [6].

The ETL process of real-time synchronous data can be divided into three steps:

4) *Create insert, modify, and delete triggers to obtain update data in real-time with conversion of data types and data cleaning.*

5) *Store the update data to “update data record” with operation sequence and operation type(Insert, Modify, Delete).*

6) *The data synchronization service connects the “update data record” through Web Service to realize the synchronization between business database and shared database.*

The task of “update data record” is to record all the update operations of the business system, including data increase, delete and modify. No matter what update operations, there will be a record to represent this update of data, including operation types, and the specific content of the update data.

“Update data record” is created at business system end, and it is composed of the fields of business data table and control fields including the update time (CREATE_TIME), update type (CHG_TYPE) and update key (SDC_ID) [7].

The following example introduces the ETL process of real-time synchronous data from the educational administration systems to the shared data center.

The trigger (insert, delete or modify) is created in business system, and it will be triggered to add records to “update data record” when the data in business system are updated. Part of the codes to create a trigger for the “xsjbxxb” of educational administration management system is as follows:

```
Create trigger trg_xsjbxxb
on xsjbxxb
for insert as
declare @xh char (10) declare@xm char (10)declare@xb char (2)declare@mz char
(10)declare@csrq char (10)declare@xymc char (20)declare@zymc char (30)declare@bjmc char
(20)declare@nj char (4) declare@xz char (1) declare@xjzt char (10)declare@rxrq char (10)
declare@byxx char (2)declare@syszd char (10)declare@jg char (10)declare @jtdz char
(50)declare@yzbm char (6)declare@lxdh char (25)declare @zzmm char (8)declare@sfzh char
(18)declare@xslb char (8)declare@zydm char (4)declare@ksh char (20)declare@bz char (20)
select
@xh=xh,@xm=xm,@xb=xb,@mz=mz,@csrq=csrq,@xb=xb,@csrq=csrq,@xymc=xymc,@zym
c=zymc,@bjmc=bjmc,@nj=nj,@xz=xz,@xjzt=xjzt,@rxrq=rxrq,@byxx=byxx,@syszd=syszd,@jg=
jg,@jtdz=jtdz,@yzbm=yzbm,@lxdh=lxdh,@zzmm=zzmm,@sfzh=sfzh,@xslb=xslb,@zydm=zydm
```

```
,@ksh=ksh,@bz=bz
    from inserted Inst
    insert
    into tb_xsjbxxb values
(@xh,@xm,@xb,@mz,@csr,@xymc,@zymc,@bjmc,@nj,@xz,@xjzt,@rxrq,@byxx,@syszd,@jg
,@jtdz,@yzbm,@lxdh,@zzmm,@sfzh,@xslb,@zydm,@ksh,@bz,'insert')
```

Part of the codes to create web service of releasing update data for the educational administration management system is as follows:

```
string sql="select * from tb_xsjbxxb";
SqlDataAdapter da = new SqlDataAdapter (sql, con);
DataSet ds = new DataSet ();
da.Fill (ds," tb_xsjbxxb");
con.Close ();
return ds;
```

Part of the codes to create a web service to clean “update data record” for the educational administration management system is as follows:

```
string sql="delete from tb_xsjbxxb";
SqlCommand com=new SqlCommand (sql, con);
com.ExecuteNonQuery ();
string a;
a="The data of shared data center data has been updated successfully! The data of update data
record has been cleared!"
return a;
```

Conclusion Remarks

Building digital campus is a long-term system engineering of which the shared data center forms its important infrastructure. The scheme of building shared data center based on ETL can realize the safe sharing of public data among various business systems and eliminate the previous “information islands”. It will greatly enhance the coordination and flexibility in all university departments and improve their efficiency.

References

- [1] V. Gour, S.S.Sarangdevot, G.S. Tanwar, A. Sharma, “Improve Performance of Extract, Transform and Load (ETL) in Data Warehouse,” International Journal on Computer Science and Engineering, March 2010, p.786.
- [2] W.H. Inmon, Building the Data Warehouse, Wiley Publishing ,Inc, 2005,pp. 68-70.
- [3] Z.H. Dang, J. Zhao, W.H. Xiao, “Study on the data integration of campus network,” Science and Technology Information, December 2010, pp. 249-252.
- [4] R. Kimball., “The Data Warehouse ETL Toolkit,” Wiley Publishers, September 2004, pp: 8-12.
- [5] F. Liu, L. Luo, G.C. Wu, “The Research and Application on ETL Tools in Construction Industry Decision System Support,” Energy Procedia, February 2012, pp. 21-23.
- [6] L.F. Wang, “The Research of ETL Model Based on Event-Driven,” Proceedings of 2011 3rd IEEE International Conference on Information Management and Engineering(ICIME 2011) VOL.02, May 2011, pp. 344-347.
- [7] C.J. Kong, A Research on Construction the Digital Campus Shared Data Center--Take Xinjiang Normal University for Example, Xinjiang Normal University, 2009,pp. 43-52.