

# Solutions for integrating data in healthcare Using Microsoft technologies

Pham Van Tap, Ponomarev A.A. (Faculty of Automation and Computer Engineering)

Adviser: Maria Yurova, Senior English Teacher

Tomsk Polytechnic University

Tomsk, Russia

Email: vantap2002@yahoo.com, boss@aics.ru, mary173@mail.ru

**Abstract** — *Introduced is an actual issue in the integration of health data, which is still a matter of great difficulty for the developers. Different approaches to integrate medical data using various Microsoft technologies are proposed, depending on their specifications. Each approach has been outlined on the tools used to develop the project and on the process of designing software for data integration in medicine. In addition, were shown the advantages and disadvantages in each of the proposed method of data integration. In the research were used three different databases in order to demonstrate the results of this work.*

## I. INTRODUCTION

It is well known, that different kind of software programmes are used in any medical organization for running a business. In fact, these software products are very specific as they have been developed using different software environments, technologies and moreover such programmes are used in various fields in the implementation of various infrastructure using diverse data sources. For healthcare organizations it is important to take issues of data integration with the related agencies. In such a situation there are many tasks, which are necessary to be solved during the process of integrating information systems. Even for the corporate users this kind of problem is highly topical, since different applications by default are not able to communicate among themselves. Such a circumstance could lead to duplication of data, system conflicts, etc. From the perspective of those, conducting a business, these mentioned situations can result in unnecessary working hours, increase of medical services cost and lack of patient information.

The main part of this work is facing a situation, in which there are three clinics and a universal information portal between them. The challenge is that the registries of services for patients and medical staff should be synchronized. To resolve the assigned task of the given article the following Microsoft technologies have been examined: Microsoft BizTalk Server, SQL Server Integration Services tools and Replication of Distributed Databases. To illustrate a

solution of the mentioned above issues, three sources of data have been used:

- Medical information portal – *Med Portal* (Microsoft SQL Server);
- Medical information system – *Aurora* (Microsoft SQL Server);
- Other medical information system – *Stat Main* (Firebird SQL Server).

## II. BIZTALK SERVER

Microsoft BizTalk Server – a server designed to integrate the software applications and create an information infrastructure interaction between them [1, 2]. When using BizTalk Server, organizations can build distributed business processes that integrate different applications, as well as implement a reliable and secure communication with partners through the Internet. In a set of supporting transport services for exchanging data among scattered sources, BizTalk Server includes a huge number of drivers, such as: HTTP, FTP, FILE, SQL, Firebird, Oracle and so on [1, 2]. The interesting moment of using this method is that BizTalk project allows organizing data flow without any coding. During this process users can define how to translate data from the source format to the target by using BizTalk Mapper. In the given work a task of translating medical staff data from *Med Portal* database to *Aurora* database, which has completely other structure of data, has been considered. Fig. 2 shows the scheme of converting data from table, which contains the registries of physicians in *Aurora* database, into a format corresponding to the structure of the table in *Med Portal* database.

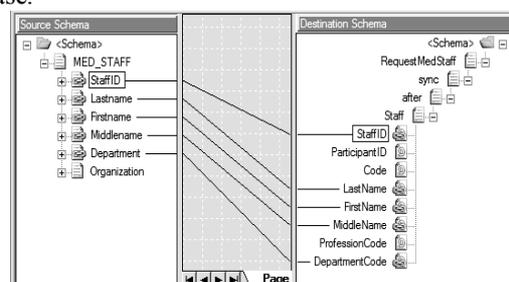


Figure 1: BizTalk Mapper for converting data formats of physicians

### III. SQL SERVER INTEGRATION SERVICES

Most of the data warehouses are still based on the integration of data from different sources using traditional systems of extraction, transformation and receiving data. However, the need to obtain data from multiple sources, dynamically changing requirements and availability of global and online transactions rapidly changes the demands for data integration process. SQL Server Integration Services offer flexible, productive and scalable engine architecture for the effective data integration in modern conditions of running a business [3].

The Business Intelligence Development Studio is used to design SQL Server Integration Services projects. The architecture of the projects includes the control flows and data flows. Each data flow is divided into three main types of control elements, such as data provider, elements for the transformation of data formats and data sink.

SQL Server Integration Services can retrieve data from various sources, including OLE DB, ADO.NET, ODBC data sources, databases providing by Firebird or Oracle, text files, excel files and etc. It is necessary to use the corresponding adapters in order to work with such kinds of data sources. In addition, SQL Server Integration Services supports a set of methods for data transformation, from which it is possible to make all the data manipulations necessary to create a data warehouse. Elements, providing functionality of data managements are listed below:

- **Lookup:** provides flexible cached-search in the related data sets;
- **Merge, Merge Join and Union All:** perform operations of connection and integration;
- **Data Conversion:** converts data into different types (numeric, string, etc.).

Fig. 2 presents the data stream for conversion and transfer of medical data from the data source *Stat Main* to the data sink *Med Portal*.

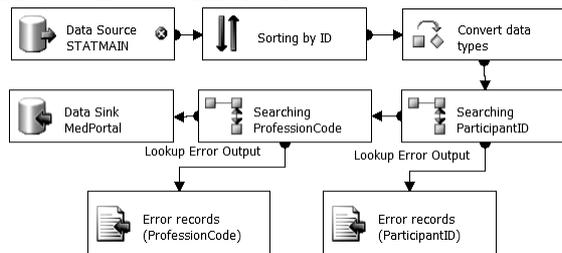


Figure 2: Data stream describing the process of converting data formats

In the above process of converting data formats, it is necessary to change the data types of some fields as required in the destination data source. There are two lookup blocks corresponding with two foreign keys in the source table used due to searching validate data in the related tables. As the result of these steps, there are possible error records return saved in separated files.

### IV. REPLICATION OF DISTRIBUTED DATABASES

Data replication is a process of bringing these two spreadsheets or databases in an identical status. The replication process is based on the concepts of “publisher” and “subscriber”. The publisher is a publish-

ing server, i.e. server sends the information. Respectively subscriber is a receiving server, called server subscription [4]. There are several ways to classify the types of data replication based on the diverse categories, but it could be classified into the following types:

- **Snapshot:** a type of replication, when information on the subscriber is just overwritten by information from the publisher at regular intervals of time. It’s the most simple type and usually used when the changing requirements are rarely met;
- **Transactional replication:** when using this type of replication a snapshot of the data source from the publisher is initially applied to the subscriber, and then information on the changes occurring at the publisher-transactions is transferred and applied to the subscriber at regular intervals of time. Using current type of replication allows users only to change data from publisher;
- **Merge replication:** using this type of replication, user can make changes at the both sites – publisher and subscriber. All changes are aggregated at the publisher, which resolves conflicts when they arise.

There are some following conditions, which must be satisfied when using the concept of data replication:

- Database objects are present on the both of publisher and subscriber;
- SQL Server has an account with the necessary rights to run the replication agent.

When developing a new information system technological nuances of the distributed databases should be taken into account. And also, depending on the purposes of the business the right type of data replication should be chosen.

### V. CONCLUSIONS

The problems of data organization face the IT-companies almost daily. Currently, with the advent of new technologies, there are various ways to integrate data. Each approach to realize such tasks has its advantages and disadvantages. In order to select the most appropriate solutions, organizations need to choose the tools for integrating data that will satisfy the current and also the future requirements.

### REFERENCES

- [1] Brian E Travis. XML and SOAP programming for BizTalk Servers. / B. Travis. - Moscow: Russian edition, 2001. – 469 pp.
- [2] G. Dunphy, S. Moukhmitski, C. Constant and others. Pro BizTalk 2009. / Publisher: Apress L. P., August 2009. - 768 pp.
- [3] B. Knight, A. Mitchell. Professional SQL Server 2005 Integration Services. May 2007. – 692 pp.
- [4] Product Documentation for Microsoft SQL Server 2005.