Proceedings of the 7th International Conference on Applied Informatics
Eger, Hungary, January 28–31, 2007. Vol. 1. pp. 15–23.

Requirement analyses and a database model for the project EGERFOOD Food Safety Knowledge Center

Tibor Radványi, Gábor Kusper

Institute of Mathematics and Informatics
Eszterházy Károly College
e-mail: {dream,gkusper}@aries.ektf.hu

Abstract

In this article we present the background of the food safety tracking system of Regional Knowledge Center (EGERFOOD), which can be found in Eger, Hungary at the Eszterházy Károly College. We analyze its requirements for the underlying information system. To build a user friendly system, which serves quickly and cost effectively the costumers, the providers, and the effected authorities by information, is a complex task. Furthermore, the system has to fulfill the strict requirements which one put up for data-safety and -encryption in case of a tracking system. We considered also these ones by setting up the EGERFOOD database model.

Keywords: food safety, tracking, database systems

1. Introduction

In the focus of the research-service activities nowadays stays the environment protection and the food analytical researches, from which the most attractive (from the viewpoint of R+D, economy, and society) works deals with the food analytical and food safety.

We have understood this at the Eszterházy Károly College and we have decided to setup a food safety and analytical monitoring center, which is the next logical and continuous step after the work we have done, but in a bit different way, because it also takes account economic expectations.

Before that in Hungary there were only segregated attempts to boost the safety some well known products. But these detached examinations could not achieve a new quality. There was no new quality- or safety-parameter introduction. But there were some results but these ones were not yet integrated in a coherent and comprehensive food tracking system. In Hungary there is so far only one food tracking system. This is for red paper. There was not nearly any attempt to adapt this for other products.

Our goal regarding informatics: Building up a system, which sits the costumer in the center, and which is able to process and send food safety information (firstly) to the costumers and (secondly) to the food producers and effected authorities in a fast, cost effective, and reliable way.

The Regional Knowledge Center, which established in our institution, at the Eszterházy Károly College, has a serious goal, to increase the competitiveness of those foods, which had been produced in our country. This plan is in line with the North Hungarian Innovation Strategy and with that connection; it tries to develop the R+D innovative abilities. Furthermore, with the coordination of the national food safety developing activities it will give valuable results to the economical sphere, too.

The following companies are taking part in the development with their products.

These are their analyzed products: Egri Bikavér wine, Detki household biscuit, Chilled fresh strudel-sheet, Tóth pungent sausage, Csiperke canned mushroom, Canned fish.

The information technology is a really important tool in every aspects of the project. From the communication that exists between the collaborating partners, through the food tracking system, to the food safety communication with the customers. There will be tasks for both the device developing (for example for solving the signal transferring problems) the software developing (for example the internet based framework of the food tracking system).

Tasks of information technology are: Create and support continuously the web system, which operates the inner communication of the project. Its goal is to ensure the information-flow between those who work in the project.

Determinate the structure of the food tracking database and create the hardware and software side of the data transmission system. The backbone of the informatics system is the database of the food tracking system.

We analyze the collected data and requirements, and by that, we create the data model of the informatics system.

Considering the long term developing strategy of the food tracking system, we have created an algorithm and a code system that is capable to identify a product. By using that we make a code which will be carried by a guarantee ticket that certifies that the actual product is in the food tracking system. It is important that the data should be stored and moved encrypted by the proper cryptographic method in the food tracking system.

Our task is the following: Develop and create the user interfaces for the data collection and query activities.

We create the aspects of the communication through WAP and Internet, by that we keep the connection with the customers. Afterwards, the general information about the food safety will be displayed on user's platform. We plan the safety requirements and the safety methods of the whole informatics system.

The installation of the data-collector hardware devices at the involved food provider companies. We connect the new gauging devices – which are in experimental stage – into the communication network of the project. The general goal: Make as many online and automated measurements as we can, regarding the critical parameters.

2. Requirement analyzes

The central data warehouse will finally store the data and serve the client modules. The first question is the passing of the data what we collect during the production procedure. For that we can choose many ways. The first is that with the help of the pre-built network system the online stated data sources constantly sending the data to the central. We should take a look at the data sources:

The data comes from two big areas: the results of the college's research lab and the data comes from the remote users, and those data we can collect from the industrial parks as a result of the production. They could come 24 hour per day while the results of the lab might come fractionally. The outer parks belong to six different food companies. They have absolutely different geographical locations, equipments, and opportunities. The products they produce are different too, so as their manufacturing technologies. Because of that the composition and the formation of the inspected data will be various. To that deflection the informatics system, which achieve the storing of the network and the data, must be prepared. As one can see the expected amount of data demands the creation of a big central data warehouse. The incoming data require a network with big enough bandwidth.

Next to the basic parameters the forwarding and the keeping of the data with high safety level is a very important requirement too. It means that data loosing cannot be happened due to hardware or network error. And it also means that the moved data must be properly encrypted, because there are industrial-like data too, which cannot get into a 3rd party company. So we must pay attention for the proper use of the fast and effective encrypting methods.

The central data warehouse's tasks will be the serving of the outgoing data and the managing the queries. We expect this task to fade in as the project goes on, because the server will get many queries, which have to be gratified. The outgoing data will be available with thick clients through Web, Wap, and inner communication channels so the rate of charging will grow dramatically as number of the users grows.

Because of the requisites and the expected development we must secure greatly the paths of the incoming data and the superstructure wisely. So it is reasonable that the data have to be buffered and not being sent directly to the central server. For these two tasks we set up temporary or buffer servers into the system. So it is important to find the proper place of the buffer servers then circumscribe and realize it. We marked the places of the buffer servers as the co-operating companies' inner parks where they produce the inspected products. That kind of server is keeping together the work of the College's researcher lab.

The servers has to preprocess the data from the computers and measurement-

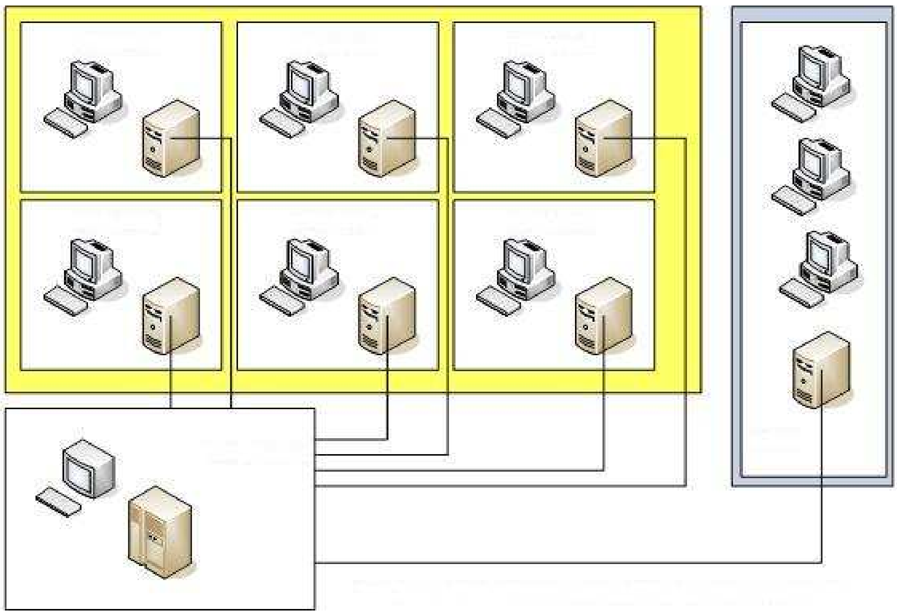


Figure 1: The way of data from the co-operating companies and from the inner labor

tolls in there domains. Then send the preprocessed data in packages to the central server. The encryption is also done in the servers at the partners. The encryption software will be developed in the framework of this project. There is an important question, how can we avoid data-loss with the highest probability. This question has two sides. On one side, we have to solve the long term protection of data already in the central server. This can be done with a good archive making policy. On the other side, we have to solve the long and also the short term protection of data which was "born" but not yet sent to, or received by the central server. The short term protection should mean that the first data saving term should be as close as possible, i.e. should be saved on the computer where it is arose. Then the long term protection means that on the server the preprocessed, encrypted and ready to send data should be again saved before sending to the central sever. This means that the data, which has not been sent yet, is saved on two independent places. This kind of data redundancy seems to be very resource consuming, but to fulfill the requirements this steps seems to be crucial. To be able to safe resources we have to make an archive making policy at each storing level. This means that we have to find a suitable time period, which depends on the speed and storing capacity of the storing machine and the time and size intensity of data to store, and after this time we make an archive and delete archived data.

First level of storing: The first level of the data storing system is the computers attached to the measurement tools and the data storing computers. The

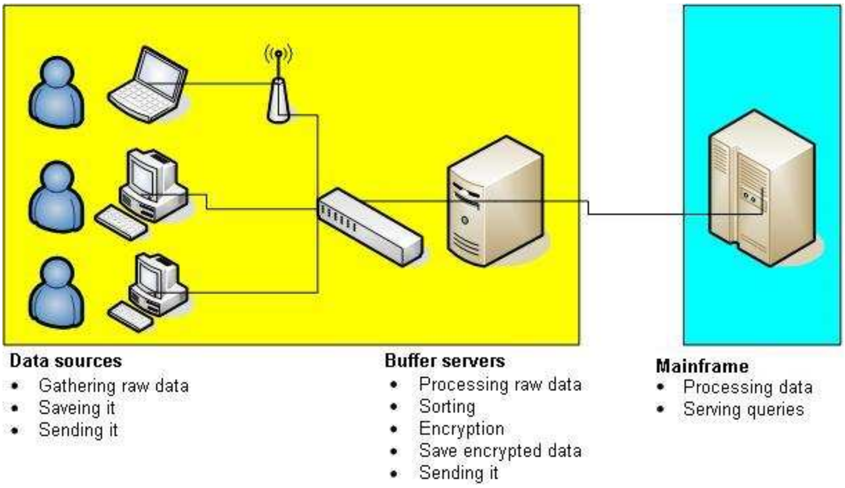


Figure 2: The way of the data and the role of nodes from the data source to the central data warehouse

format of storing is XML, because it is a text file, i.e. it is very mobile and easy to use, and it can be stored easily in a database.

Second level of storing: The level of buffer servers. Here takes place the preprocess tasks and evaluations (e.g. computing average value) and here are stored the data waiting for a possibility to be sent to the server. These data can be easily exported if it is necessary to XML or to any other portable format.

Third level of storing: The level of the central data warehouse. This is a Linux server running an Oracle Standard Edition database server program.

There was a question about the setup of the buffer servers, which operation system and database server program should be installed on them. The answer depends on the local specialties. It is important to play upon existing hardware and software elements (e.g. at the food providers) and integrate those. This means that each buffer servers may have a different operation system and database management program.

3. Conclusions

Since the whole system is very complex, very different platforms, operation systems, and database management systems will be used, we have drawn the following conclusions:

1. The requested data safety can be guaranteed with the help of data redundancy.

2. It is necessary to develop a component which makes possible the appropriate communication between the different operation systems.

3. There is a software development task: Make transparent the differences between the possible database systems and to solve the data conversions.

4. There is a mathematical and software development task: Find the appropriate coding algorithms and implement it, furthermore, the encryption of the encoded data by the appropriate cryptography methods.

5. There is a research task: Programming and fine tuning the central data warehouse in such a way that while we insert the (possibly great) input data and select the answers for the (possibly great numerous) incoming requests both the insertions and selections should remain quick, although, the optimization of these process works against each other (e.g. the number of indices and the inner structure of them).

4. Plans for the setup of the database model

First we considered the following analyzes steps:

1. We have prepared a questionnaire and sent it to each data provider partner. Afterwards, we have interviewed them at their sites. We found that each data provider (most of them are a food provider company) stores production data in a notebook like minutes of production. The format of this document is fixed in the quality management documentation nearly everywhere.

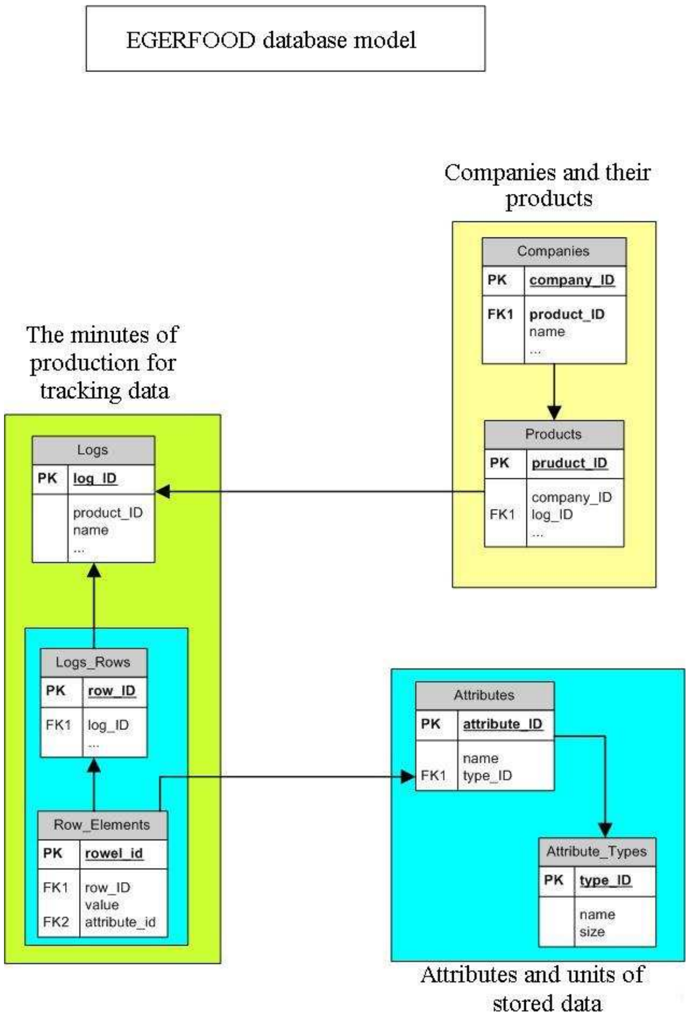
2. We considered the question:

    - Is it possible to create a database model which can store any important data (regarding food safety) of any considered products (or possibly later introduced) or
    - it is better to make a specialized database model for each considered product and take the union of them.

The second solution means that the database will have fractions without any link to other ones. In this case we gain the advantages of static databases, but the clients and data processing software models will be more complicated, thus, will be more costly and time consuming regarding the development time.

After considering the data from the first step we made the decision that we setup a general database model, which can store any minutes of production, thus, it can store data for any production process. The development of this data model is more time consuming but afterwards the development of the software components will be dramatically easier.

The structure of the data model is shown on the following drawing:



The model consists of the following parts:

Two tables contain the data of the companies and the data of the products produced by them (Companies, Products). This solution allows that later other companies could take part in the project or an other product can be introduced in the system.

Two tables contain the two main attributes of a stored data. These attributes are type and size (Attributes, Atributes_Type). These information makes possible the data conversion of the stored data.

Three tables contains the minutes of the production, the rows of them, the elementary data stored in the rows (Logs, Logs_Rows, Row_Elements).

In the Logs table data of a work-period of a company are stored. This means that it contains the company id, the product id and the id of the work-period. There is no strict definition for work-period. It can mean sorter or longer time period depending on the production process. For one Logs entry could be attached more Logs_Rows entries. These ones together mean one data storing process. During a data storing process we may store for example that Mary Smith on December 10. 2006 in the 2^{nd} work-period in the 10^{th} tent harvested 3 kg of mushroom. We may attach also a comment to this Logs entry.

The Row_Elements Table stores the elementary data. An entry of this table contains the also the attribute id.

By this scenario we can ensure that either for the input forms or for the analytical questions we can select the necessary information.

5. Future tasks

1. Create the database from this database model.
2. Performance analyzes of the database using deferent database servers, like PostgreSQL, Oracle, and MSSQL. This analyzes should cover both data insertion and query. Develop the software which is needed for this test. Analyzes of the test results and publication of the conclusions.
3. From the information gathered in the step above we shell refine the database model.
4. Setup the encryption system, selecting the most suitable algorithm(s) and implementing it (them).
5. Develop the web clients (thin clients) and the stand alone clients (fat clients).
6. Test the software and create the programmer's and the user's guide.

References

[1] Ailamaki, A., Shao, M., DBMbench: Microbenchmarking Database Systems in a Small, *Yet Real World in Confidential* – Submitted to ICDE (2004).

[2] Microsoft, Improving .NET Application Performance and Scalability, (2004), 639–682.

[3] Ruthruff, M., (Microsoft Co.), Microsoft SQL server 2000 Index Defragmentation Best Practices, (2003).

[4] Gray, J., The Benchmark Handbook for Database and Transaction Processing Systems, *Morgan Kaufman Publishers*, Inc. 2nd edition, (1993).

[5] Gray, J., `http://research.microsoft.com/~gray`

[6] Beeler, D., Rodriguez, I., Optimizing your Database Performance, *BMC Software* (2002).

[7] Tsai, M., Kulkarni, C., Sauer, C., Shah, N., Keutzer, K., A Benchmarking Methodology for Network Processors 1st Network Processor Workshop, 8th Int. Symp. on High Performance Computer Architectures (HPCA), Feb. 3rd 2002 Boston, MA.

[8] Lindström, J., Niklander, T., Benchmark for the Real-Time Database Systems for Telecommunications University of Helsinki, Department of Computer Science.

[9] Boncz, P. A., Riihl, T., Kwakkel, F., The Drill Down Benchmark Data Distilleries EN.

[10] Gulutzan, P., Pelzer, T., SQL Performance Tuning.

Tibor Radványi
Eszterházy Károly College
Computer Science Department
H-3300, Eger, Leányka u. 6–8.
Hungary

Gábor Kusper
Eszterházy Károly College
Information Technology Department
H-3300, Eger, Leányka u. 6-8.
Hungary