# How to Measure an Information System's Efficiency?

**András Keszthelyi**

Budapest Tech, Hungary
keszthelyi.andras@kgk.bmf.hu

*Abstract: The use of computer-based information systems has become part of our everyday life in the administration in the higher education as well while end-users usually do not satisfied with these scholar information systems (SIS). Our everyday experience is that our SIS having been grown up in the past years has some annoying problems even today. E.g. it is usually too slow. Beyond our personal experiences our students also reckon that it is too slow according to a student questionnaire. The low speed can have a number of reasons. I investigated the role or, in particular, the necessity of the three-level data modelling. In this paper I describe the circumstances of a measurement which has been executed by a colleague of mine and me. This measurement shows that better performance can be reached based on a 'good' data model even in a poorer environment.*

*Keywords: database efficiency, scholar information system*

## 1 What is a SIS?

The administration of the scholar records of the students has become a great task for today which needs great resources. This is caused by not only the increasing number of students but the complexity of the credit system as well.

In the „old" times the curriculum was a well-defined one which was mandatory for all students in only one possible way. Nowadays the curriculum is well defined, too, but there are a large number of possible ways to fulfill the requirements them. Students can decide their own way and velocity of their studies. The only main rule is a logical one: a student is allowed to register for a course if the prerequisite(s) of that course is fulfilled. E.g. one is not allowed to register for Mathematics II while didn't succeeded Mathematics I.

There are some other, practical rules in the credit system. According to the above it is natural that the students select not only the objects they want to study in a given semester but they choose one of the courses of that object as well personally. While the number of the subscribed students to the given course is less than or equal to the maximum allowed number.

So to manage the administration by hand can hardly be imagined. Almost everywhere this task is solved by computer based information systems. Such a system can be called a 'Scholar Information System'.

## 1.1 Problems of a SIS

There are many problems can be mentioned in connection with a SIS. These problems can be grouped by many ways. From the point of view of the end-users the main problems are the following:

Is the system able to manage a large number of administrative tasks simultaneously? In this case the expression „large number" means that large enough in order to serve all or nearly all the students in a given time period who want to or are obliged to do something in the administrative filed. There are two typical situations when this problem can occur: the registration for examinations at the end of the semesters and the registration for courses at the beginning of the semesters.

Is the system realistic? Does it work in accordance with the real life? This means at least two different fields. First: does it know and serve the administrative rules of the given institute? Second (last but not least): Are these rules themselves realistic? Are they in accordance with the rules of the logic, are they practical?

Is the system ergonomic? How much time is needed to perform a given task? How many mouse-click is needed to perform the most frequent activities?

In this paper I examine the first mentioned problem, the loadability of a SIS-like database in a typically problematic situation. I have tried to make a quantitative measurement in order to determine that how many students and/or tasks can be performed almost simultaneously. I have chosen the testing of the registration for examinations process which is a bottleneck usually.

## 1.2 Existing Scholar Information Systems

In Hungary there are two well-known scholar information systems are used in the higher educations. We at Budapest Tech have used one of these two ones for nearly a decade.

In the beginning there were serious problems. The database could not cope with the load caused by the registration periods as it would have been expected. After nearly a decade we use the third main version of this SIS. Of course it has grown up since then but it has it's own weaknesses in the field of loadability even today. Our nearly ten thousand students are divided into three sets and each set of students can start their registration on different days even today in order to lower the load of the SIS.

According to a student questionnaire made by me in 2008 the most frequent problems observed by the students were: aborting the connections, short timeout and slowness.

# 2 Which Factors Affect Efficiency

Database efficiency, i.e. it's capability to cope with high loads are determined by some very different factors. Of course the influence of the hardware environment is very important. This is the first circumstance which comes into one's mind, but it must be declared that to develop the performance of the hardware in order to have a higher software performance is a 'brute force' method: the more money you have the higher performance you will have.

There are more sophisticated and, of course, more cheaper methods which results in a higher software performance.

Let's see the software environment. The operating system, the relational database management system and the application itself are the most important elements in this field. The first two can only be chosen from a given set based on some ratings of their most important technical and co-operational features. How some technical aspects influence the performance I had investigated before and was lucky to present it here. [F4369]

At the third one you have more possibilities to influence the performance. After choosing the programming language and tools there are two main fields which determine the performance of the developed program(s). These are the quality of the applied algorythms and the quality of program coding. In case of databases the 'algorythm' has a more special meaning than in general: the quality of the data model is included as the most important, necessary but not enough circumstance.

## 2.1 The Quality of Data Models

The main steps of the developing an information system are: to determine what we want as exactly as possible, data modelling (i.e. determining the data structure), determine the functions operating on the data structure. In case of data-intensive systems the data structure is more important and determines the functionality. [F4282 p. 541]

So data modelling is the basis, which is necessary but of course not enough for succeeding, on which a good IR can be constructed. In order to succeed we need three level data modelling and planning according to Halassy. [F4292 pp. 28-33] These levels are the conceptual, the logical and the physical levels. The names of these levels or even the „three level" label have widely been used but in most

cases without the appropriate meaning. In the early times of the databases Codd wrote about that even the SPARC committee of ANSI uses these words without defining them precisely. „The definitions of the three levels supplied by the committee in a report were extremely imprecise, and therefore could be interpreted in numerous ways." [F4296 p. 33])

As Halassy states the conceptual level data model is the one in which we describe the entities of the reality, their properties and relations or linkages in natural concepts and corresponding to the reality. The logical level is the one where the data structure of the database is planned according to the circumstances and constraints of the technical aspects, the accessibility and the efficiency. Defining the exact type and size of the data elements, the way they are stored in the storage equipments, the way they are accessed are described at the physical level plan.

There are general prerequisites of the goodness of data models. At the conceptual level a good data model needs to be understandable, unambiguous, realistic, full and minimal. [F4260 p. 192] Of these properties minimality is the one which can be examined exactly by mathematical tools.

Minimality is a very important property because redundancy is dangerous. If a data structure is redundant the database being built upon it needs (much) more storage. If it needs more storage it will need more time to be handled. These problems can be solved by 'brute force', by quicker storage equipments and processors. The biggest danger of redundancy is the possibility of data errors: redundancy causes certain undesirable characteristics, the so called insertion, update, and deletion anomalies that could lead to a loss of data integrity. We can suppose that at least some of the experienced problems of the SIS used by us is rooted in model level errors. I focused on the efficiency.

I was wandering whether the data model of the SIS used by our Budapest Tech meets the above requirements. Of course I was not given the model documentation itself because it is a commercial software so I had to try another way. I made a data model for such a scholastic system which is considered to be good enough, at least by me, to examine that instead of the original one. If I can reach better, or at least not worse, results in a poorer environment than I can state that the reason of the difference can be identified in the differences of the data models.

My concept was to identify a function which is critical from the aspects of the response time and of the number of concurrent users to be attended. There are two such functions in a scholar system: registration for examinations and for courses at the end and at the beginning of the semesters, when nearly ten thousand pupils would like to be attended by the system each of whom needs to register for about four examinations or for about twelve courses.

I chose the registration for examinations. I made the conceptual data model carefully and the logical and physical level plans based upon it. A colleague of mine implemented the plan and developed the part of the application which is needed to do some efficiency measurements. [F4367]

# 3 How to Decide the Object of the Measurement

## 3.1 What to Measure?

I chose above the registration for exams as a critical field to investigate. First I had to decide what I wanted to measure at all in this field. The exact response times of each registration of each students? The number of retries and/or the response times? The number of successful and unsuccessful tries in a certain time-period? Do I need to make an ABC-assay and to rank the responses into three sets, one of them is called 'very good', the other is called 'acceptable' and the third one is called 'poor'? At what values can I mark the boundaries of these sets?

## 3.2 Errors are not Mistakes

There are numerous random factors which can and, of course, do influence the measuring. Let's see at least some of them.

Since the measurement is done in a working computer environment all the other possible activities of the operating system would be taken into account. E.g. to save some data as a response of a given query to a local file needs some (a little but significantly greater than zero) time. This is a random error because the moving of the read-write heads of the hard disks and the puffer usage is unpredictable.

The network traffic which is not part of the measurement activities could be eliminated closing the subnet for the time of the measurement, but even in this case there are some factors at ethernet level which could influence the measurement. This is a quasi-random error because it increases if the network traffic increases.

Last but not least the measurement itself influences itself. To measure some computer activities by computers needs one or more, less or more complex programs to run. These programs also need some or more resources while the total amount of resources is a given constant.

Beyond errors like the above there are observational and computational errors, too, to cope with.

## 3.3 The Object of the Measurement is Decided

To know the response times to two or to four significant digits (in seconds) is not important. The important questions: Could the response times be tolerated by an average human student or not? How many of the registration attempts are fulfilled?

To try to answer these two questions the influence of the above mentioned measurement errors are negligible. The borderline between the 'tolerable' and 'intolerable' time requisites cannot be defined exactly in a mathematical manner because it is a subjective opinion of the end-users, in this case the students.

So I decided to measure the average and the maximum response times and the successfulness of the registration attempts. If the response times and the numer of the unsuccessful attempts are lower than in case of the real system, even in a poorer environment, my above statement of the difference of the data models could be proven.

# 4 The Measurement

The test environment consists of PC computers and free software. The database server has an Intel processor of four cores and 8 GB of RAM. It runs Linux as the operating system, httpd server Apache with PHP as an application interface between the users and the database, and MySQL as a relational database management system. Workstations are simple computers equipped with processors Intel P-IV of 2.6 GHz clock speed.

The test database contains 8192 students, about as many as our active students, four examinations for each of them. The number of places is one and a half times bigger than the amount of the students' examination. The registration itself is made by PHP scripts randomly for the test user currently called it via the offline browser of the test client workstation. Each test user registered a date for all her examinations. We logged the client system time at the beginning of the connection to the database server and when the response was saved to a local file by the offline browser. The server load was also logged. Test registrations were started almost simultaneously, with a two second pause after every one hundred starts. The settings of the offline browser wget were: max 4 retries, 30 seconds timeout, 10 to 30 seconds between to retries.

The test environment and application is described more precisely in [F4367], [F4368/a].

## 4.1 The Results

The results were better than I had expected before.

All the registration of all the pupils were successful.

The total time needed for the 32.768 registrations of the 8.192 students was 3 minutes and 7 seconds, so the average time needed for a test student is 0.0228 second with a maximum value of 1 (one) second because the offline browser wget

logs it's activities in hh:mm:ss format so fractions of seconds cannot be taken into account.

I think that these values good enough, so I my statement can be considered as proved.

The maximum value for the server load (1 min load) was 2.85, with an interesting, staircase-of-staircases like diagram as described in [F4367].

**Conclusions**

Summarizing the above I can state that (much) better results can be achieved based on a 'good' data model even in poorer hardware and/or software circumstances. The quality of the data model influences the quality of a database at user level so much that three-level data modelling ought to be considered much more important as stated by Halassy. [F4314 p. 32]

**References**

[F4282] Raffai Mária dr.: Információrendszerek fejlesztése és menedzselése. Novadat Bt., 2003

[F4292] Halassy Béla dr.: Adatmodellezés. Nemzeti Tankönyvkiadó Rt., 2002

[F4296] Codd Edgar Frank: The Relational Model for Database Management - version 2. Addison-Wesley Publishing Company, 1990

[F4260] Halassy Béla dr.: Az adatbázistervezés alapjai és titkai. IDG Magyarországi Lapkiadó Kft., 1995

[F4314] Halassy Béla dr.: Ember - információ - rendszer. IDG Magyarországi Lapkiadó Vállalat., 1996

[F4367] Szikora Péter: Measured Performance of an Information System. 7th International Conference on Management, Enterprise and Benchmarking, Budapest, 2009

[F4368/a] Szikora Péter: The Role of the Tools and Methods of Implementation in Information System Efficiency. 2nd International Conference for Theory and Practice in Education, Budapest, 2009

[F4368/b] Keszthelyi András: The Role of Data Modeling in Information System Efficiency. 2nd International Conference for Theory and Practice in Education, Budapest, 2009

[F4369] Keszthelyi András: Information Management in the Higher Education -- the Role and Importance the of the Different Technologies. 3rd International Conference on Management, Enterprise and Benchmarking, Budapest, 2005