New Approaches of Storing ERP Data

Attila Selmeci and Tamás Gábor Orosz

To install an earlier ERP system we had to design the disk layout accurately keeping in mind the behavior, functionality, and mechanism of many different hardware and software components. Some of these components should be mentioned, like the ERP solution itself, the underlying database management system solution and the operating system. The business data is stored generally always in database tables with primary keys. In some cases the relation between the tables are defined in the database level, but there are ERP solutions, where the table correlations are defined on application level having own dictionary for that. In both cases the data is read from (and stored in) tables, which are not always accessed by unique key attribute values, but several times by not key attributes. To have an acceptable response time from the database layer indexes should be created for the frequently searched attribute sets. The first database system based ERP systems required special attention to isolate table data and index parts to separate disk areas. The main reason was the speed, because the database can read the index and data parts simultaneously in that case. The data store of ERP systems as technique and method can be divided into three main areas:

- Business data representation in the ERP system (logical view)
- Database storage of the internal representation (technical view)
- Data or disk storage technology (physical view)

These different approaches cannot be handled as separated, unequally replaceable methods, but they are rather layers, which not in any case are really distinguishable. In the hardware world the physical layer was improved in the last two decades. The two main technical goals were generally the reliability (e.g. redundancy, high availability) and the response time, speed (e.g. load distribution, bandwidth). The early data stores used different RAID (redundant array of independent disks) technologies to approximate both goals together. The capacity and the speed of the disks increased and it was available to build and use larger data stores. The first RAID controllers were hardware components only, but afterwards some software configuration tools were offered to manage, design the disk distribution in a server. This direction founded the early storage system and solutions having specially connected disks with special-purpose hardware and an operating system offering many different features and services.

According to our study the storage vendors did not really follow the new hardware developments, options, but only the palette of services grew in a wide range. Nowadays the storage systems offer several redundancy, high availability, data handling (like de-duplication), backup, fast copy, read cache and other options. These features are very useful and required in some cases, but the costs are much more higher and the value that they provide. On the other hand the speed as a goal is not in the main focus of the vendors, because they advice only faster disks or even flash drives (SSD-s), and of course more disks to distribute the load among the spindles. These directions are correct but with having many features, special connectivity layers, like Fiber channels and switches, the disks, the real data storing surfaces are far away from the database server or even application server CPU-s. The newer and faster hardware elements made easier to use bigger, faster and reliable 'local' disks in a server. The local disks or even PCIe storage cards provide data in a much higher speed limit, which are sometimes necessary. Our measurements proved that less CPU power is enough if the data storage layer can faster provide data for the ERP system (in case of commercial RDBMS it could decries license amount as well).

The top layer of the above mentioned three data handling approaches determine the mechanism of data management and data store in lower layers as well. Each ERP system contains

several business modules and functions, which handles business objects, like accounting documents, employees, sales or purchase orders, etc. The business objects can be defined as object oriented entities with status (defined via properties), object modification functions (declared through class methods), and of course published status modifications (managed by events) to inform other objects about status change (like employees were hired or even fired). The business objects on technical level in an ERP system are not always defined as real object oriented entities, but only logical correspondence and relationships are defined between data or even tables. Many ERP environments can manage them as objects in higher level, but each time these objects are stored in database tables separating the current status to several tables connected in the database (or as described above in ERP system) level via foreign key relationships. The object as a unit or even entity cannot be recognized on database level.

According to our study for ERP and other kind of OLTP (Online Transaction Processing) systems the object based data handling on logical and technical level would be more effective that the currently widely used table based representation. This direction goes forward to the database layer, where the object oriented database solution would handle better the ERP or OLTP requirements than the relational database management systems.

Our paper collects the availability of the offered technologies, which could be used for ERP systems and architectures as future environment and technical solution. We try to refer to data-warehouse systems as well pointing out the differences in architecture and requirements positioning the possible directions. We figure out the advantages and disadvantages of the different technologies, approaches and try to sketch a way, which leads to now opportunities of data store for ERP systems.

Acknowledgements

This work was supported by TAMOP-4.2.2/B-10/1-2010-0012 Programme.