

Thirty Years of Server Technology — From Transaction Processing to Web Services

Klaus Meyer-Wegener

Friedrich-Alexander-University of Erlangen-Nuremberg, Germany
kmw@informatik.uni-erlangen.de

Abstract. Server technology started with transaction-processing systems in the sixties. Database Management Systems (DBMS) soon adopted mechanism like multi-process and multi-threading. In distributed systems, the remote procedure call also needed process structures at the server side. The same is true for file servers, object servers (CORBA), Web servers, application servers, EJB containers, and Web Services. All these systems support a request-response behavior, sometimes enhanced with a session concept. They are facing thousands of requests per second and must manage thousands of session contexts at the same time. While programming the applications that run on the servers and actually process the requests should be as simple as possible, efficiency must still be very high. So a general programming environment should be defined that is easy to use and, on the other hand, allows for the efficient execution of thousands of program instances in parallel. This contribution will identify mechanisms that have been developed in the context of transaction processing and database management. It will then generalize them to server processing of any kind. This includes program structures, context management, multi-tasking and multi-threading, process structures, program management, naming, and transactions. The driving force behind the discussion is to avoid the re-invention of the wheel that far too often occurs in computer science, mostly in ignorance of older and presumably outdated systems.

1 Introduction

Servers are ubiquitous today, and they can be quite small. The services they offer span from remote procedure calls, HTTP requests, and application invocations to Web services. New kinds of server software are being developed, often without knowing that the history goes back into the sixties, when transaction-processing systems were created for the first time. Who among the server implementers of today has ever heard of CICS, CINCOM, DATACOM/DC, UTM, and all the other systems from that time? Most of

them have not, and the consequence is a repeated re-invention of the wheel. CICS is still in operation today, but it has long been restricted to the domain of IBM mainframe computers. These computers are not very well known among the developers of today. In fact, some of them even refuse to learn about these systems. For them, the world of computing is limited to Windows and Linux.

The purpose of this contribution is to identify the common techniques of all these systems and to describe them in a neutral terminology. This should enable developers of server software to pick from a well-established set of techniques. It also means to identify the “invariants” needed in any kind of server—irrespective of the plethora of names. It has been very hard to find names that are not occupied yet by any of the various servers, and it has not always been successful. In these cases, however, the different meaning is stated explicitly.

2 Kinds of Servers

Transaction-processing systems, also called *online transaction-processing* (“OLTP”) systems, are the oldest class of server systems, going back into the sixties. Typically, they were airline reservation systems like the TWA reservation system [12] or banking systems [5]. The first system to mention here is Sabre. The Brainy Encyclopedia describes it as follows: “Sabre (Semi-Automated Booking and Reservation Environment) was the world’s first online airline reservations system. Developed through the joint efforts of IBM and American Airlines, it first went online in the fall of 1962, running on an IBM 7090 computer. After going through a series of system upgrades, including a relocation from Westchester County, NY to Tulsa, OK, the system remains operational today (2003) and was the prototype for virtually every mainframe-based online system that followed. Sabre was based on real-time computing advances made by the US Air Force in the development of their SAGE radar-coordination and target tracking system.”¹

The notion of “transaction” here is not used in the sense of databases yet, but simply names a processing step. While in the first approach systems were limited to single-step transactions, they very soon progressed to multi-step transactions. Single-step transactions are much easier—not surprisingly, Web servers also started with them before cookies and other mechanisms were introduced. No context needs to be managed. If information from the last transaction was needed in the next, it had to be re-entered by the user—the trick of including it invisibly in the next input message has been used very early in these systems and is by no means an invention of Web-site developers.

The middleware that has been introduced for this is called a *transaction-processing monitor*, or TP monitor for short [3, 8, 9]. These systems have rarely been investigated in the scientific community, but they have—and still are—used extensively in commer-

1. http://www.brainyencyclopedia.com/encyclopedia/s/sa/sabre_airline_reservations_system.html