E-COMMERCE - Promise, Illusion and Reality

by Bruce Taylor

The consequences of e-commerce for the IT architecture and operations of an organisation are, at best, poorly understood. Even when understood in theory, they are rarely appreciated or accepted or applied in practice. Our first hypothesis:

A prerequisite for any e-commerce venture is a robust and scalable on-line business transaction processing capability with high availability!

The terms "robust", "scalable", "on-line transaction processing" and "high availability" are standard, stock phrases in every salesman's and every consultant's vocabulary and are applied liberally and indiscriminately to every product or architecture currently in fashion. In reality, the vast majority of products as delivered do not deserve a single one of these descriptions and architectures are just conceptual simplifications not inherently possessing any such characteristics. In theory, any consistent and comprehensive set of products arrayed in any reasonable architectural blueprint can be made to conform to all of the required characteristics, given sufficient application of human intelligence, perseverance, dedication, feedback mechanisms, control mechanisms and, last but by no means least, the passage of time. In practice, most attempts are expensive failures or inadequate and unstable approximations to the desired goal.

In its most extreme, but increasingly common form, e-commerce must service a user community of indeterminate size, indeterminate location (thus indeterminate time zone) and indeterminate behaviour. The customers are the users, not the (hopefully) trained and disciplined business employees of yesteryear. Thus the users use the system as and when and how they choose to do so, not as your architects and designers thought, hoped or assumed they would do so. Not only do they do whatever they want, whenever they want to, but they regularly abort their interaction part way through a process, because the connection has dropped, or they became fed up, or were called away for dinner, or hit the wrong button and closed the browser. To service this community your system must be available all the time (i.e. 24x365 operation is essential) and be capable of dealing with loads and work patterns that are unpredictable, unstable and constantly shifting and fluctuating and aborting. If your e-commerce venture is successful, then your peak loads could be increasing at exponential rates, at least in the early phases of its existence.

Another persistent characteristic of successful e-commerce ventures is that the "look-to-book" (i.e. queries versus sales) ratio goes up and up and up. With traditional distribution channels the look-to-book ratio is usually a few-to-1, but with successful e-commerce sites it easily becomes 50-to-1 and could even go to hundreds-to-1. If you are successful, then people want to have a look at what you have and how you present it, even though they have no intention whatsoever of actually buying anything you have on offer. This implies that the transaction volumes your business processor must handle are at least one order of magnitude higher to generate the same volume of sales. A system capable of behaving itself at a few transactions a second does not necessarily do so when forced to cope with many tens of transactions per second, even if you do upgrade the hardware. Hence, we have to realise and accept that doing business on the web involves consuming a much larger amount of processing capacity. I.E. if doing $1M of business via traditional channels requires the ability to process 10 transactions per second; doing the same amount of business via

the web will require the ability to process hundreds of transactions per second. This frequently leads to what one travel industry expert (expert not consultant, since he does real work and knows what he is talking about), Novak Niketic of Sutra in Boston, calls "Death by Booking Engine".

For at least the past decade, the IT industry has been dominated by fashion and illusion, not by practical reality. The dismal failure or phenomenal cost overruns of IT projects was far from unusual before e-commerce arrived on the scene. With e-commerce the operational requirements are significantly more difficult to achieve and the failure rate will increase dramatically.

In our opinion, the most demanding requirement of e-commerce is the reliable 24x365 operation of the IT business transactions. In the majority of the successful e-commerce ventures the IT business transaction processing capability is provided by a legacy system, which had been around for years, and has been doing 24x365 for a long time already. Legacy systems are, of course, not fashionable (i.e. they do not contribute to the six-figure incomes of IT consultants, gurus, salesmen and marketers). The legacy systems have been misrepresented by the purveyors of fashion to mean: obsolete, bad, expensive, inefficient and in dire need of replacement. However, the true meaning of the word "legacy" is "something of value that you inherit from your forbearers" and that is exactly what many of these systems in reality are. With very few exceptions, only legacy systems achieve true high availability 24x365 operations and they achieve it at an integral cost per transaction that is far less than their so-called "new technology" competitors.

The business sector with overwhelming success in the e-commerce field is travel; the legacy system making that possible is IBM's TPF (Transaction Processing Facility, program product number 5748-T14), which has been around for 35 years! TPF has been doing 24x365 for decades and getting better and better at it with the passage of time. Current well-run TPF installations achieve total system availability levels of 99.94% to 99.96% as a percentage of total time: not scheduled time, not planned time, not expected time, but TOTAL TIME, ACTUALLY MEASURED OVER ONE YEAR. This is true for the smaller installations running on average a few tens of transactions per second up to the larger systems running many thousands of transactions per second. With current I-stream technology TPF is scalable up to somewhere between 15,000 and 20,000 transactions per second, depending on the travel business applications in use. The largest system is already running in excess of 9000 transactions per second.

The proponents of new technology systems dismiss the claimed achievements of legacy systems as being the irrelevant ranting of obsolete bigots, who are incapable of change and of appreciating the overwhelming logic of maintaining a six-figure income, no matter what it costs the customers. The legacy system bigots, outnumbered, outgunned, mostly older and cynical, not as vocal and completely on the defensive, simply do not comprehend why the lessons learned over the past three decades are not applied to the new technology environment. By trial and error, they lived with the incremental improvement of hardware and software technology through the 1970's and 1980's, to the stage where true 24x365 operations were a fact and commonplace. They know the dedication and the effort it takes to achieve such a goal, but seem incapable of passing that message on to the new generation, whom they dismiss as "low volume, low availability, low quality experts, with big mouths, big incomes and no achievements". However, quite apart from the generation gap and the mutual antagonism, the world of new technology ICT is very large, very diverse and is driven by completely different priorities than quality and performance. It would probably be logistically impossible

for such a knowledge transfer to occur.

As was previously stated, we believe that it is possible to make any software environment designed for transaction processing meet the 24x365 goal. However, what is required to actually achieve that? What can we learn from the experience of the legacy bigots who actually did it and on a scale beyond the wildest dreams of any competent new technology performance analyst?

First of all, we believe we can safely state that hardware reliability and performance are no longer primary issues. From the 1960's through the 1980's hardware reliability was a major concern and consumed a large part of the software engineer's and application designer's effort to circumvent; hardware performance was the constraint limiting the growth of systems. Those constraints are no longer applicable, as TPF so amply demonstrates. In addition, it is interesting to review how long it took TPF to get where it currently is. In the beginning, the late 1960's when the first dozen or so TPF systems became operational, a total system availability of 95+% was achieved. In the 1970's this grew to 98+%; through the 1980's to 99+%; and only in the 1990's to 99.9+%. It took 30 years, but the first 2 decades were primarily a struggle against the hardware. The 1990's were a struggle against the software and that was a collaborative effort between the supplier, IBM, and the highly competent and knowledgeable TPF technicians spread across several dozen TPF installations scattered around the world. All the major TPF installations have a kernel of "TPF Gurus" who have worked on the system for decades and know it inside out.

Hence, the problem is software and, as such, it is a people and an organisation problem; thus much more difficult to address. This leads to our second hypothesis:

The primary prerequisite for achieving a true high availability 24x365 operation is an established culture dedicated to that goal!

What do we mean by culture? It means having a kernel of bigots of sufficient mass and in long-term employment, who are dedicated to the challenge of making their system number one functionally and number one in the 24x365 league. This group, depending on the type and size of system, could be anywhere from a small handful to several dozen people and they are all technical experts in the systems and applications software of "their system", plus extremely knowledgeable about the business environment in which it operates. They need not be in the same department nor under the same manager, since if they are good they will ignore the organisation, ignore the whims and fantasies and reorganisations of ephemeral management and do their own thing. They will work in an underground network, paying lip service to their managers, but in fact only really communicating with and obeying the wishes of their kernel peers. They pursue what they believe is the overriding logic of their software and the business they feel responsible for. They are the antithesis of the "dedicated followers of fashion": IT Consultants and professional IT career managers.

This kernel is the embodiment of the culture. It takes time to establish itself and represents a wealth of ever-increasing cumulative working knowledge dedicated to the incremental improvement of the system performance in all senses of that word "performance". Without it no system will achieve high availability 24x365 operations. Hence, if you have by design or accident acquired such a culture, then destroy it at your peril…

Having achieved reliable 24x365 operations, the issue of scalability remains. Every computing environment is scalable to a certain limit, which goes up with each hardware improvement and approximately follows Moore's Law. That any new technology system will come within one order of magnitude of the scalability of TPF in the next decade, we do not believe.

Bruce L.M. Taylor
Amsterdam, October 2001
