Matthew Eastwood
Group Vice President and General Manager, Enterprise Platforms

Mission-Critical Business Applications: The Need for Always-On Servers

November 2013

Social media, mobility, and cloud services are driving up user expectations for always-on access to business services and important data — when, where, and how they want it. Always-on requirements significantly affect business organizations and the IT departments that support those businesses. When downtime is not an option, organizations are increasingly turning to fault-tolerant systems to keep their business up and running.

The following questions were posed by HP to Matthew Eastwood, group vice president and general manager of IDC's Enterprise Platforms service, on behalf of HP's customers.

Q. **How are the always-on requirements for social media, mobility, big data, cloud services, and security impacting IT departments' decisions today?**

A. IDC defines the rapid growth of cloud, mobile, social, and big data analytics as the emergence of a new IT paradigm called the 3rd Platform. The 3rd Platform has accelerated the pace of change in organizations of all types as business leaders pressure CIOs to deliver new IT services quicker and more efficiently than ever before. As new users come online and application portfolios are modernized for the 3rd Platform, IT departments must consider the platform and architectural needs of their workloads more carefully than ever before. IDC believes the high business value workloads — including line-of-business applications, OLTP, data warehouses, and data analytics — will continue to be hosted on-premise for many years to come. These high-value workloads continue to have unique service requirements, including reliability, availability, and serviceability necessary to meet stringent service-level agreement (SLA) requirements. The emergence of the 3rd Platform is driving robust transaction growth, including mobile transactions. The back-end needs of these mobile applications should be carefully considered, and fault-tolerant systems will continue to serve an important purpose in many enterprises.

Q. **How has the need for high availability changed over the past decade? Is it expanding beyond traditional transaction-based applications?**

A. One of the biggest IT impacts associated with the 3rd Platform is availability. As users spend more time online, IT services must be available around the clock. Windows for planned downtime become increasingly difficult to manage, and users are unwilling to accept unplanned downtime. Downtime of any kind results in a loss of confidence and competitive advantages in the marketplace. Revenue impacts from IT outages can be significant, but damage to an organization's reputation can be longer lasting and potentially more impactful. When evaluating the implications of downtime, one must consider revenue loss, productivity impact,

compliance issues, and damaged reputation with customers, partners, and employees. Impacts are not limited to traditional mission-critical workloads such as OLTP, business applications, and database; instead, they increasingly extend up and down the IT application portfolio.

Q.   **With today's user demands for continuous access to information and services, is the risk of downtime increasing for the business?**

A.   CIOs are responsible for ensuring that an organization's people, budget, and implementation schedules are carefully managed, particularly in the highly dynamic world of the 3rd Platform. Today's enterprises face unique challenges, which are exacerbated by businesses seeking to modernize while embracing mobile, social, big data analytics, and cloud. This new style of IT introduces a number of potential areas of expanded risk to businesses of all types. History tells us that organizations that prepare well for risk will continue to thrive, while those that don't face increased irrelevance in the marketplace, including almost certain failure at some point in the future. The IT organization is almost always seen as having an important role in the mitigation of business risk of all types. Potential areas of business risk are wide ranging and generally include cash flow, legal, regulatory, and operational hazards. Savvy management and sound IT platforms are commonly seen as the best defense against potential business failures associated with business risks, which can be appropriately managed. For many years, IT has been pressured to reduce complexity and lower delivery costs through consolidation and standardization. However, it is important to note that heterogeneity in the datacenter continues for very important business reasons, and ensuring appropriate availability levels for a given workload is an important reason IT infrastructures are not completely standardized. In fact, many organizations maintain a complex application portfolio that is closely aligned with specific business processes that impact infrastructure decisions and datacenter design.

Q.   **What do you see as the primary challenges for IT departments trying to meet SLAs for higher availability while balancing budgets through efficiency and standardization?**

A.   SLAs spell out a critical element of the effective provisioning of IT services to the business. These SLAs identify the appropriate key performance indicators (KPIs) necessary to provide transparency between the business and IT that in turn ensure what is actually being delivered. Unfortunately, SLAs often lose their initial orientation aimed at mutual agreement between IT and the business and instead become weapons for the business to challenge internal IT delivery or external service providers. Ultimately, the SLA needs to be written in a way that is most relatable to the business, including enhanced performance to the business (i.e., high availability), constraints removed from the business (i.e., flexibility), and IT performance (i.e., throughput). Again, this needs to be done in a way that is respectful of workload, making particular considerations for high-value applications that typically have significant interdependencies with critical business processes. For high-value applications where enhanced performance (i.e., high availability) is a requirement, the availability of sufficient IT budget to sustain the SLA is equally important. Standardizing on common modular components, such as those utilized within an x86 architecture, can contribute to the economics and efficiencies needed to ensure critical business processes and to balance a tight budget.

Q.   **How should enterprises measure the value of fault-tolerant systems (IDC Availability Level 4) against other alternatives such as virtualization or Linux failover clusters?**

A.   The pace of change in organizations of all types continues to accelerate, and this is prompting business leaders to pressure CIOs to deliver new IT services quicker, more reliably, and at lower costs than ever before. This is happening alongside the rapid acceptance of new mobile technologies and cloud delivery models. These models in turn command the development of

new types of enterprise applications that further enable robust data creation. The 3rd Platform is also impacting the staid world of datacenter design as IT seeks new levels of application availability, reliability, flexibility, scalability, and utilization from systems. This can often be confused with the move toward standardization, when, in fact, growing interest in integrated and converged infrastructure offerings is beginning to reverse prevalent thinking over the past 20 years — namely that open system design and standardization and the resulting independent development paths for hardware and software are in the best interest of virtually all datacenter workloads. IDC has long believed that a strategic fit exists between server platform and workload. Understanding this relationship can help IT organizations target hardware solutions to their businesses more effectively and efficiently.

Fault tolerance is an architectural property that enables a system to continue operating (generally with imperceptible degradation of service) in the event of a failure of one or more of its components. Fault-tolerant computer systems have been around for nearly 30 years. In such systems, two common architectures are employed to ensure that the system is distributed to an array of subsystems. Fault tolerance is about much more than redundant hardware design, with full fault tolerance achieved only when data integrity is maintained at all times. Fault tolerance must also take into consideration planned and unplanned downtime where scheduled downtime is the function of a management-led event and unscheduled downtime is the result of a physical event such as a power outage, hardware failure, software failure, human error, or other datacenter anomaly. Because fault-tolerant systems deliver resources that are optimized end to end for reliability, availability, and serviceability, the system cannot easily be confused with a more general-purpose system serving a workload with potentially lower business value. This could be the case with a more mainstream failover Linux cluster. Additionally, it's important to remember that in general, high-availability fault-tolerant systems require less human intervention to restore operations following a failure, with the most common cause for an IT service outage still being human error.

A B O U T   T H I S   A N A L Y S T

In his role as group vice president and general manager of IDC's Enterprise Platform service, Matthew Eastwood leads a team of analysts responsible for identifying and analyzing vendor strategies as well as technology, market, and customer trends affecting enterprise platforms and datacenters worldwide. In this role, he is responsible for managing market analysis, forecasting, and consulting for servers, workstations, server and client operating environments, virtualization/cloud software, systems management software, datacenter trends, infrastructure channels, and IT leasing and financing worldwide.