



Linux – scaling to data center performance levels

by
Mark Lillycrop
Chief Analyst
Arcati Research
mark@arcati.com
www.arcati.com/ml



Linux – scaling to data center performance levels

From humble origins in 1991 as Linus Torvalds' personal Unix implementation, Linux has taken the IT world by storm. The Open Source operating system accounts for a very large proportion of the Web server and hosting market, and is experiencing huge growth in print/file server applications, e-mail and firewall systems, and numerous distributed office roles. Figures from analyst IDC for the third quarter of 2003 revealed a Linux server revenue increase of 49.8%, to \$743 million. Moreover, this represents six consecutive quarters of year-on-year growth. In view of the static economic conditions witnessed elsewhere in the IT market, it's clear that Linux is expanding at the expense of other platforms.

Since Linux entered the IT environment initially as a platform for Internet applications, it has experienced widespread adoption in small and large organizations alike, but it is still struggling to make a real impact in the data center. To understand the way that the system is evolving and developing, it is important to separate its role as a commodity system in the smaller organization from its potential impact in the larger enterprise and as a cornerstone for business-critical applications.

The Linux server

For small and medium-sized organizations, and for distributed enterprise applications, Linux has some very powerful benefits in a server role:

- it offers a truly open approach to software development, with a clearly defined feature set and a range of well-supported distributions (which have to adhere to a formal standards base) from companies such as Red Hat, SuSE, Turbo Linux, and Debian;
- support is plentiful and, like the software itself, relatively cheap (though rarely free!), one of the system's overwhelming advantages;
- performance and stability are strong and predictable, thanks primarily to the system's Unix heritage, and the effort expended by the army of enthusiastic programmers who respond speedily to every reported bug;
- it has an excellent security record, so far; there are conflicting views as to whether it is really much more secure than Windows, or whether the lack of exposure is due mainly to its lower profile. However, the Linux developers' rapid and open response to security bugs has helped to build confidence in the system's resilience to hackers and viruses;
- the system offers a burgeoning range of packaged applications, making it suitable for the vast majority of standard business requirements;
- it offers a consistent software platform for a growing range of hardware;



Arcati Research Bulletin

- it has the backing of nearly all the major IT vendors, to the extent that it is the focus for many key technological innovations, including blade server development.

Research conducted recently by Vanson Bourne on behalf of IBM found that 26% of small businesses are already using Linux (42% having moved from a Windows server or NT environment), and 55% believed that the operating system was robust enough to run core business applications. Respondents cited a range of reasons for making the move, including cost, performance, security and reliability.

The signs are that, as the Linux application portfolio continues to grow, and the drift from Windows and Unix gathers further momentum, the operating system will garner further market share in commercial and government organizations.

The Linux client

Linux on the desktop has taken considerably longer to gain acceptance; the sheer logistics involved in replacing Windows applications on the desktop, and the cultural implications of converting office workers to an alternative product suite (such as Star Office), have restricted the impact of the Linux client.

Nevertheless, there have been some notable successes in recent months, particularly for Sun Microsystems, which has won some significant business with its Java Desktop System (a combination of the Mozilla browser and Star Office applications). Sun has a large-scale trial lined up with the UK government, and has recently secured a contract with the Chinese government which potentially involves millions of Linux seats. China, of course, offers a particularly interesting opportunity, as many of these licences are new, with no Microsoft legacy to displace; but the UK trial is just one indication of the enthusiasm among European government departments to adopt an Open Source strategy across the board – right down to the desktop.

Scaling up to the data center – ‘business critical’ drivers

To date Linux has been most successful in the commodity server market and, despite the impressive list of benefits, it has faced some significant obstacles in climbing up to the business-critical data center environment. With the system’s limited ability to scale beyond four processors at the top end, and some general concerns about the Open Source movement’s ability to drive the software up to the levels of performance and availability of the more mature Unix implementations, Linux has to prove itself outside its core market.

We believe, however, that it is only a matter of time before Linux gains the credibility and technical scalability to support 24x7 CRM, ERP, and transactional applications. Ultimately Linux offers the large enterprise two irresistible advantages: reduction in cost and reduction in complexity. The ability to shift between an Intel box and the mainframe without changing a single line of application code, and the opportunity to take advantage of a wealth of free (or nearly free) Open Source tools and utilities are benefits that can’t easily be dismissed in the data center environment.

Much relies on general market confidence, and this is constantly increasing – though SCO’s high-profile copyright crusade is not very helpful. The speed of Linux’s upward progress will depend not so much on the operating system’s technical characteristics *per*



Arcati Research Bulletin

se, but on the determination of larger organizations to use Linux as a way of achieving tighter integration of applications and reducing cost across the enterprise. The attractions of supporting the Open Source movement, with its perceived liberation from lock-in to proprietary standards, are also of mounting significance.

In our view, there are three principal factors that will determine how rapidly Linux is taken up in the data center:

- experience of pioneering users
- scalability and resilience
- maturity of management tools.

Early 'mission-critical' users

Large IT shops are notoriously cautious in their acceptance of new technologies, since every sizeable change in the data center causes a ripple effect throughout the whole IT environment. Before adopting Linux in business-critical roles, many CIOs are waiting to see how well their competitors and business partners manage the transition before making the jump themselves.

Some Linux consolidation stories have enjoyed a very high profile. When Amazon.com migrated to Linux from a complex configuration of Unix systems, porting its key applications from 64-bit Unix to 32-bit Linux in the process, it claimed to be making savings of around \$17 million. The project, which was well supported by Red Hat, and by a number of enterprise systems management specialists, showed just how cost-effective an Open Source solution could be.

While Amazon is now an extremely well-known name, its e-business infrastructure is relatively simple compared with those of its brick-and-mortar counterparts. More established companies in the financial, retail, and utilities sectors have a far more complex mix of legacy systems to consider, and their integration and management requirements are more difficult to resolve. For such companies, acceptance of Linux within the data center is necessarily a slow process. The New York Stock Exchange, which announced a far-reaching Linux strategy in 2001 (involving the replacement of numerous Unix servers), marked a turning point for 'big iron' Linux. Since then, there has been a slow trickle of genuinely heavyweight Linux implementations, including a number of large financial institutions and airlines, but many of these stories have failed to reach the headlines.

Without doubt, the most enthusiastic supporters of business-critical Linux are currently government and education departments, which see Open Source as a way of achieving independence from vendor lock-in. National and regional authorities in the USA, UK, France, Germany, Finland and others are going for Linux in a big way, as are a number of Mid-European and Far Eastern governments (which in many cases are utilizing the operating system to support green-field applications). From the perspective of market momentum, governments are in a strong position to dictate Open Source support to the organizations that work with them. This in turn is contributing to Linux penetration on the periphery of the government sector.

IBM, one of the greatest advocates for Linux in the data center, has been working hard to generate greater market activity with its 64-bit z/OS and 31-bit OS/390 versions of the operating system. For IBM, Linux on the zSeries mainframe offers a new lease of life for its top-end machines, partly because it opens up a large portfolio of packaged applications



Arcati Research Bulletin

on the zSeries, and partly because it gives the mainframe a unique role as a hugely scalable hosting environment. By running multiple Linux images under z/VM, customers can consolidate tens (potentially hundreds) of distributed servers onto the mainframe.

IBM has also been at pains to make Linux more cost-effective for customers who want to experiment with the system alongside existing mainframe applications. The Integrated Facility for Linux (IFL) has proved extremely popular on smaller mainframe systems, and many users are now opting to run z/OS and Linux applications in neighbouring LPARs, the two partitions interoperating via Hipersockets.

The technology for supporting Linux apps in the data center is maturing fast, and our own experience suggests that interest is rising fast. However, we still need to see more 'bet-your-business' case-studies in the headlines, and more evidence of the ways that Linux is being integrated into the broader environment of the heterogeneous data center.

Scalability, resilience, and the 2.6 kernel

If there is one single factor that will encourage the take-up of Linux in the data center in the next few months, it's the impending release of the 2.6 kernel, Torvalds' core code on which all the commercial distributions are based.

Linux on the mainframe is great for customers who want to scale out – ie to run a very large number of relatively small applications in one place, with centralized management. However, it still needs to scale upwards, beyond the four processor limit which has so far restricted its appeal at the high-performance end of the market.

2.6 covers a lot of ground in terms of new features. At the low end, there are substantial improvements to the way that embedded processors are supported (2.6 incorporates the UCLinux technology for managing multiple low-end microcontrollers). The new version of the Linux kernel is also designed with 64-bit desktop applications in mind, and offers more functionality for graphics workstations.

However, it's in the server world that the biggest impact will be felt. 2.6 is designed to support more memory, more disks, and more processing power, offering support for 16-way processing, with scope for scaling up to 32-way and even 128-way in due course. The new release is also optimized for NUMA (Non-Uniform Memory Access) systems, and has been substantially re-written to improve the way that processors, memory pools, and I/O devices interact. Moreover, Linux will now provide more efficient management of the hyperthreading performance feature on Pentium 4 machines.

All of these features add up to a Linux kernel that is headed straight for the data center. Of course, the kernel alone won't be sufficient – this extra function will also need to be supported in the separate distributions and applications – but it is a sign that Linus Torvalds, Andrew Morton, and their team of kernel custodians are pushing the operating system in the right direction.

The only danger is that potential users will be concerned about the informality of the timescale for the next release. At the time of writing, its launch is imminent; but Torvalds is quoted in the press as saying simply that it will be launched when it is ready. Unlike more commercially-minded developers, the Open Source movement seems to be allowed considerable latitude within the marketplace in rolling out new products!



Arcati Research Bulletin

Managing Linux within the heterogeneous enterprise

Technically, then, there is no practical reason why Linux will not be able to scale up to the most demanding applications over time.

The third driver that we mentioned earlier for business-critical Linux is mature management tools. The biggest challenge for large data centers today is capturing consistent information on the performance and availability of all their platforms, and presenting that data in a form that is meaningful to both a business and technical audience, as a way of aligning IT operations and business processes more closely.

All of the major systems management vendors – IBM/Tivoli, HP, BMC, CA etc – are focused on this area, and each has a slightly different emphasis. Incorporating Linux fully into the enterprise-wide systems management environment is an essential prerequisite to establishing the system's credibility within the data center.

For customers, the choice of management products depends largely on their existing mix of technologies, and on the degree of comprehensiveness that they require from their tools. The prime considerations for all large organizations are how to measure the impact of system performance on all affected applications and processes, and how to rectify any problems transparently. The relative merits of all-singing, all-dancing management frameworks and the more fragmented but flexible best-of-breed solutions have been debated for years, but as Linux integration becomes a key priority, such questions need to be re-examined carefully.

Linux and OMEGAMON

One company that has a very valuable contribution to make to today's enterprise management environment is Candle. Candle was one of the earliest entrants into the market for third-party mainframe performance management tools with the OMEGAMON family, and it says a lot for the reputation of those tools that Candle is still using the OMEGAMON name to market its data center products.

In recent years, Candle has diversified into new areas of system management, focusing on application integration in general and WebSphere in particular, an area where it has developed some unparalleled technical expertise. But its core products for the data center have retained a simple focus and strategy – to manage system, workload, performance and availability; and to use these functions to automate alerts and event management across the enterprise.

The OMEGAMON XE (eXtended Edition) family of performance and availability monitors are now available for a wide range of platforms, including the mainframe, Windows, Unix, and a number of databases. OMEGAMON XE for Linux, currently at version 120, is one of the most mature enterprise Linux performance management tools available, and runs on OS/390, zSeries, and Intel platforms, and on the SuSE, RedHat and TurboLinux distributions.

The function of OMEGAMON XE for Linux basically comes down to alerts and reports. The alerts are intelligent warnings on impending problems or trouble-spots that need to be addressed by an operator or an automated procedure. The vast range of reports include information on disk I/O and utilization, file system, network connections, and system-wide performance. The Linux version of XE is a powerful performance



Arcati Research Bulletin

management tool in its own right, but where Candle is really planning to make its technology count is in the level of integration it can achieve between the monitors for each platform. This is where OMEGAMON DE (Dashboard Edition) comes into the picture.

OMEGAMON DE has grown out of Candle's existing OMEGAVIEW platform for integrating system and performance data from multiple sources. DE, however, is specifically intended to combine traditional systems management functions with the relatively immature area of application performance management, and to bring all that information together at a single console for performance and availability.

In the last few years, Candle has become increasingly focused on service management as a high-level discipline. DE is a key part of this strategy, allowing customers to build a flexible view of enterprise-wide performance data, and of the impact that any problems or bottlenecks are having on applications. Using its Universal Agent to draw data from multiple sources, such as SNMP, File/Log, IP Socket, HTTP and API, DE permits on-the-fly analysis, and helps analysts to align system management information with application impact and change management in a unique way. This in turn can play a major part in helping businesses to align technical and business performance objectives, and even to automate a number of areas of software change management.

Candle is the first to admit that its solution is not a universal one; its application management capability is currently focused strongly on the IBM WebSphere environment, and the company's real strengths lie firmly in the mainframe world. But one of the great advantages of Candle's approach is that it has extended its highly mature mainframe management capabilities out to other platforms, and can gather detailed, consistent performance data from every platform that it supports.

As Candle builds full support for Linux into its DE integration console, it will be able to offer a new level of integration, not just between Linux system and application performance data, but also between Linux and other key data center platforms. This will give the company a very strong hand to play among large and medium-sized corporates.

The future for Linux in the data center

The bottom line is that the barriers to adoption of Linux for data center applications are gradually lowering. The benefits – in terms of openness, cost, and breadth of applications – are being articulated by a growing number of 'pioneer' users, and concerns over the ownership and long-term prospects of the technology (SCO notwithstanding) are now far less significant than they were five years ago.

In terms of technical scalability, the Linux community is rapidly removing the constraints to SMP and NUMA growth beyond the four-processor level; while IBM is working hard to increase the resilience and power of mainframe Linux.

The final issue is integration and management of Linux applications within the broader data center environment. With the emergence of solutions such as Candle's OMEGAMON and robust tools from all the heavyweight enterprise management vendors, it is clear that – in this respect too – Linux is now coming of age.

Mark Lillycrop is Chief Analyst of Arcati Research Ltd and an Associate of Valley View Ventures, Inc. For further information on this paper or Arcati services, visit www.arcati.com/ml.html or e-mail mark@arcati.com.