Sine Nomine Associates

October 2000

LINUX for S/390: Technical Solutions for Competitive Advantage

Contents

Executive Summary

This white paper presents the technical advantages of LINUX® solutions deployed on S/390® processors as opposed to Linux solutions on discrete physical servers of other architectures. This paper specifically addresses technical implementation and support issues as viewed from a management perspective.

A companion white paper, Linux for S/390: Scalability and Competitive Advantage, addresses the total cost of ownership issues of large scale Linux for S/390 solutions versus discrete physical servers.

A brief for engineers, programmers and technical professionals titled Technical Notes on a Large Scale Implementation is also available.

Technical Solution Applicability

This section discusses technical benefits of the Linux for S/390 solution. Each of these elements should be included in an evaluation of the applicability of Linux for S/390, and form the beginnings of a decision matrix of what applications are appropriate for consideration as candidates for this environment.

Solution Benefits

Pure Linux – Not a Alternative Personality on Top of Another OS

The Linux implementation on S/390 is a complete self-contained operating system, similar to other S/390 operating systems like VSE/ESA™ and OS/390® in that it provides editors, program development tools and communications facilities such as IP connectivity and Internet services. Compared to other earlier IBM® efforts to provide a UNIX® environment on the S/390 (such as AIX/ESA®), this adherence to the Linux API and operational standards implies several things:

Pure ASCII – No EBCDIC

Linux for S/390 is ASCII-based, which allows applications written for other UNIX systems to use the same character sorting and processing code without the modifications required to cope with a noncontiguous character set such as EBCDIC provides. Applications do not incur the overhead of translating char-

acter sets for all I/O operations flowing over real devices or network connections, thus improving application performance

Direct Source Portability from Other Linux and UNIX Implementations

With the Linux API and toolset identical to Linux on other platforms, most applications that do not require specific kernel information or platform-specific extensions are directly source compatible. Applications can be moved to the Linux for S/390 environment by copying the source code to the Linux for S/390 system and recompiling.

Large Number of Mainstream Internet-Oriented Applications Immediately Available

Much of the current development on Internet services and applications is being performed on Linux systems due to the wide acceptance of Linux and the easy availability of system and application source code. New research and new services are introduced daily that support Linux on many platforms, providing Linux a status as a de-facto cross-platform coding environment for diverse architectures.

Increasing Acceptance by Commercial Software Vendors

An increasing number of commercial software vendors (led by database vendor Oracle and others) are providing support for Linux on many platforms, and most are considering or have announced support for Linux on S/390 in addition to their support for Linux on other platforms. Industry leaders in network management and database applications such as BMC, Software AG and IBM have announced commercial support for applications on Linux for S/390 available in late 2000.

Growing Availability of Linux-Knowledgeable Staff

Universities and technical training programs are currently providing training and system administration courses for Linux on the Intel platform; this skill transfers directly to Linux on S/390. This abundance of training programs

implies that the cost of management staff for Linux systems will continue to decrease as the number of trained staff available in the industry continues to increase.

This fact also correlates with the current predominance of Sun hardware in enterprise deployments; Sun pursued an aggressive policy of generous hardware and software grants to universities in the 1980s and 1990s, producing a generation of Sun-aware students which is reaching management positions in the early decades of the new century. This approach has been successful for Sun, and viewing the production of Linux-savvy students, the next generation is likely to support a Linux strategy and solution space on a more multivendor hardware base, such as the S/390.

Fast Availability of Patches and Solutions for Defects

The direct integration of the S/390 support into the mainline Linux source tree and the availability of source code allows the vendor (IBM), system integrators such as SuSE and TurboLinux, and the general user community to collaborate on defect resolution in a proactive manner. Defects discovered in the core code are often repaired within hours of detection.

Availability of Full IBM Defect and Deployment Support

By offering full IBM defect and deployment support on selected machines and distributions, IBM has provided a credible environment to consider enterprise use of Linux. The same proven support structure available for traditional IBM operating systems is being made available to Linux for S/390 customers running specific hardware and software configurations. The configurations IBM offers support for are the Generally Available versions of SuSE or TurboLinux offerings of Linux for S/390 running on G5, G6, or Multiprise® 3000 processors.

IBM has publicly committed to returning fixes generated for the core Linux components back to the open source community for inclusion in the master source tree.

Virtual Systems and VM/ESA® Provide Powerful End-to-End Systems Testing Capability

The virtual machine environment makes it possible to create not only virtual servers, but entire application environments for testing and system evaluation purposes. Production data need not be placed at risk of damage by testing applications and test systems can be created and destroyed at need without consuming resources or expense in software or hardware commitments.

Powerful System Management and Measurement Tools and Capabilities When Running under VM/ESA

When running under VM, all I/O and application operations can be instrumented and measured by the VM hypervisor and can be analyzed by traditional mainframe performance tools. This capability is unique in the UNIX environment, and provides a number of interesting opportunities, such as reliable billing for system resource utilization and resource usage analysis and capacity planning on a large scale.

VM includes a number of sophisticated automation and system management tools such as a common console management facility that integrates all messages written to the Linux system console into a single console log with timestamps and system identification, and a programmable operator that can interpret messages and react using user written action routines. These tools are part of the basic VM operating system, and do not require additional application purchases or installation.

Resource utilization and management constraints are available as single commands for virtual systems running under VM/ESA. Constraints on I/O, memory, CPU timeslice and share of the total system resources can all be controlled directly from an appropriately privileged VM user ID and dynamically changed on the fly.

Research Areas In Current Code Release (2.2.16)

Naturally, there are some areas requiring additional research in the current Linux for S/390 code. The issues discussed in this section outline a few of the more serious issues.

Idle Task Timer Resolution Consumes Significant Resources in Current Code

The current Linux kernel design assumes that Linux is the only operating system active on the physical machine. The Linux kernel uses a simple method to handle waits for work when idle: the kernel sets a timer interrupt to awaken it periodically. Upon receipt of the interrupt, the kernel scans a queue for events to process. If no events are found, the kernel resets the timer interrupt and quiesces again. This process consumes several tens of thousands of machine instructions per instance – essentially wasting the cycles if no work is to be performed.

In the default Linux distribution, the timer interval for idle scans is set at 100 Hz. Normally, this would not be a problem on a relatively capable machine, however in a virtual server environment, these constant wakeup interrupts consume valuable CPU resources that cannot be delivered to other virtual servers. To successfully deploy large numbers of virtual servers, it is necessary to adjust this interval to reduce the number of wasted cycles used on idle interrupt processing.

Note that adjusting this timer also affects the interactive performance of the Linux system. The idle timer is also used to compute the timeslices given to other processes in the Linux scheduler and reducing the number of timer interrupts lengthens the time slice per process at the expense of system responsiveness. This is generally not a problem for server-only applications, but will significantly impact interactive use.

S/390 Architecture Not Optimized for Computational Tasks

The S/390 hardware architecture is primarily optimized for I/O-bound tasks due to its history as a business processing system environment. While computationally significant, it does not compare favorably to other platforms

for floating point intensive processing. As many UNIX applications rely on inexpensive floating point math (such as graphical X Window System applications) to complete tasks, this design decision will affect the performance of graphics-intensive and math-intensive applications on S/390 Linux implementations without IEEE floating point support (i.e., G5 class systems or above). Note that integer performance on the S/390 is relatively comparable to other Linux platforms, and the the I/O capabilities more than make up for the lack in floating point performance for I/O and network intensive applications.

Some S/390 Architecture Functions Not Yet Supported

Currently the Linux for S/390 system code is not yet aware of some of the specialized hardware assistance available in the S/390 hardware architecture, such as the integrated cryptography CPU or the dynamic I/O definition facility which allows new I/O devices to be added without a power-on reset. IBM is cooperating with the open source community to provide information on exploiting these features, however some of the interface capabilities may be limited due to a desire to protect IBM intellectual property with regard to proprietary hardware interfaces.

Installation Skill Sets Not Yet Completely Customized and Documented to Comparable IBM OS Levels

Currently a substantial amount of both S/390 and Linux skills are required to install and configure Linux for S/390. The installation process requires knowledge of the S/390 environment to bring up the basic Linux kernel and access to a working UNIX or Linux system to complete the installation from either the Internet or a locally mounted CDROM over a network device. The traditional IBM installation documentation is not yet complete, but is under development in concert with the open source community.

Commercial distributions of Linux for S/390 are expected to include installation tools provided by the distributor to ease installation effort.

Limited Support for Tape and Disk Device Types

Linux for S/390 currently has limited support for DASD and tape devices. In the current release, only CKD and ECKD DASD are supported when running in basic mode or in LPARs. If running under VM/ESA, all DASD supported by VM/ESA are available.

Linux for S/390 currently has no tape driver support, however development is underway to add drivers for 3420, 3480 and 3490/3590 tape devices as a collaborative project with the open source community. Currently this limits the backup capability of the system to full volume dumps using DDR or other system specific utilities

Prime Solution Opportunities

Internet Services

Given that Linux for S/390 provides direct source compatibility with a wide range of open source applications, the virtual server provides a cost-effective means of deploying large scale Internet support services based on open source tools. System administrators can take advantage of widely used tools and techniques to manage UNIX based services and system images and add the tools in the CMS environment to provide powerful programmable operator and console management services.

"Appliance" Servers

As a subclass of the above Internet services, a specialized but growing need in the server environment is the area of computing support for dedicated function Internet appliances such as set-top Internet boxes like those used by WebTV. These "appliance" servers require large scale deployments and are generally single function servers supporting geographically distributed clients. The resource controls and massive scalability of the Linux for S/390 (as well as the close coupling with other more traditional IBM operating systems), lend additional integration possibilities for application data and billing solutions.

Development and Testing Environments

When running Linux for S/390 under VM/ESA, the inherent ability to bring up complete test environments without maintaining separate systems is a significant cost savings. New or upgraded applications can be tested within a complete copy of the entire production environment without risking production data. Virtual servers can be created to model the effects of hardware upgrades or to simulate the use of hardware that is not physically present in the environment to demonstrate concretely the benefits of a proposed upgrade.

Enterprise Server Farms

Just as with the Internet services mentioned above, in many enterprise environments, the proliferation of special-purpose servers for single functions presents a significant management burden on central IT staff. The common console management and system management capabilities of Linux for S/390 under VM make this solution attractive for similar reasons.

This area also presents some interesting financial benefits, as open source replacements for LAN file and print service are already available and can be deployed at minimal cost to serve the needs of departmental file servers or other utility functions within an enterprise environment. This can be a significant operational savings, as organizations can retain central management of operational functions such as backup and recovery planning, but still allow distributed control by providing dedicated Linux instances on demand.

Internet Data Centers and High-Density WWW Hosting Facilities

As more service providers enter the application server provider (ASP) marketplace and following an increasing wave of interest in Internet-connected data center facilities offering contract WWW hosting services, we see increasing demand for resource accounting and management on an individual user basis. The combination of VM/ESA and Linux for S/390 provides a completely instrumented solution capable of recording network and application resource usage in a large scale environment. The accounting and resource utilization data is recorded in a standard format understood by most major billing systems, and can easily be processed using standard tools such as SAS or Perl.

Areas to Avoid In Current Code Releases

Computationally Intensive Applications on Systems without IEEE Floating Point Support

While Linux for S/390 will run on G2 class systems and above, S/390 hardware without IEEE floating point hardware (G4 and below) suffer a significant performance hit (approximately 15-20%) on applications using extensive floating point computation. The design of Linux assumes IEEE floating point support rather than the traditional S/390 floating point format, and the implementors of the Linux for S/390 port complied with the Linux design specification to ensure that applications written for Linux on other platforms run on S/390 returned the same results as on the other platforms.

On G4-class systems and below, IEEE floating point support is handled by a program exception routine that intercepts the calls to the hardware IEEE FP processor and simulates the calls using software routines. This contributes to some issues with performance for X Window System applications, as X uses extensive floating point math as part of the display generation and imaging routines.

Large Numbers of Systems with Real-Time Interactive Response

As mentioned above, the idle timer interrupt problem consumes significant resources, and thus limits the number of systems that can be operated due to the overhead of processing high-frequency timer interrupts for a large number of systems. For server applications (where real-time interactive response time is not critical), adjusting the timer frequency is an acceptable solution. If real-time response to a command-line user is required, the maximum number of systems supported must be limited to ensure sufficient resources to support sub second response time. Careful analysis of system performance data will be required to maintain good interactive response time for large numbers of systems. This scenario is particularly difficult to manage with basic mode or LPAR deployments, as these environments do not offer performance instrumentation beyond the limited data available in the Linux environment.

Sizing A Linux for S/390 Solution

The recommendations below describe some empirical observations of Linux sizing characteristics in a large scale deployment.

Operating Environment

Linux for S/390 can be deployed in three possible configurations: native, in an LPAR, or using VM/ESA or IBM Virtual Image Facility for Linux™ (a limited implementation of some of the VM hypervisor technology). Each has some benefits, but we find the VM or IBM S/390 Virtual Image Facility for Linux deployment to be the most attractive arena given the flexibility and the number of systems supported. Some specific comments follow.

Native

In most cases, running Linux for S/390 as the only operating system on an S/390 is probably not practical or cost effective for most users. Linux does not yet support many of the required system configuration tools (such as IOCP or EREP), and without tape support, it is difficult to perform production services due to an inability to back up the system. This mode is most attractive for sites with concerns about licensing costs for S/390 operating systems or hobbyists experimenting with S/390 simulators and looking for a low cost operating environment to support tinkering with the S/390 architecture.

LPAR

For sites with only OS/390 or VSE/ESA, this is the most likely initial configuration, however the current architectural limitation of 15 LPARs per physical system makes the comparison with non-System/390® hardware unfavorable due to the high initial cost of S/390 hardware. Adding additional virtual systems normally requires a VM or Virtual Image Facility–based solution, however this configuration may be useful for introducing Linux-based front-ends to existing OS/390 or VSE/ESA-based applications.

VM/ESA or Virtual Image Facility

In most cases, this is the most flexible and desirable solution. VM provides substantial resource management and system management capabilities that are necessary in a large-scale deployment, as well as a simple backup and recovery

scheme that would need to be invented or licensed for the basic mode or LPAR solution.

Virtual Image Facility offers a low-cost introduction to the virtual environment intended to introduce users unfamiliar with the virtual system environment to system management, however it lacks the individual resource management and automation capabilities of VM/ESA. In most significant deployments, Virtual Image Facility presents a method for sites to deliver quick deployment of Linux for S/390 systems as part of a planned small to medium-size deployment that requires more images than an LPAR-based or native S/390 solution can deliver, but that cannot cost-justify a full VM/ESA license.

Number of Server Instances

Native/LPAR

The native or LPAR environments support a hard limit of one and up to fifteen instances respectively. In most cases, this is not suitable for enterprise deployment in production due to the limitations in management and resource management, however sites experimenting with enterprise Linux services and making a case for expanding services or consolidating existing UNIX environments with more capable management services.

VM/ESA or Virtual Image Facility Guests

The number of Linux instances under VM or Virtual Image Facility is limited only by the physical memory and CPU resources of the underlying processor complex. In most cases, the most significant limitation will be network bandwidth if internal OSA adapters are used (a maximum of 16 adapters are currently supported).

Physical CPU Resources

Linux for S/390 supports single and multiple CPU environments up to the 64 CPU S/390 architectural limit. Each Linux instance consumes an average of .01% of a single engine when idle; consumption when active is dependent on

application and number of additional systems active (observed average in large deployments is 20% of defined instances are active at any given interval; the remainder idle performing only timer interrupt processing).

Storage

Native/LPAR

Storage requirements for LPAR Linux instances are a minimum of 64 MB central storage per partition.

VM/ESA

Allocating a minimum of 128 MB central storage for VM and as much expanded storage as is available allows VM to manage resources more efficiently than Linux for S/390 can do so, as the VM/ESA virtual storage simulation system is capable of using central and expanded storage for paging and other functions. Observation indicates that it is considerably more effective to define 2 GB virtual machines for Linux for S/390 and allow VM to do virtual memory management, as this allows transfer of control to another dispatchable virtual system during page fault processing in a specific virtual machine.

DASD Resources

ECKD DASD

For LPAR installation, reserve an entire 3380 or 3390 volume for Linux installation. For installation under VM, initial installation will require 1000 cylinders of 3380 or 3390 DASD, however future installations can share approximately 700 cylinders of the space between instances (/usr, /bin, /lib, and /src can be shared using R/O minidisks).

FBA DASD

FBA DASD is not supported for LPAR or basic mode installation. Under VM, FBA DASD may be used for data or minidisk access with the DASD device driver, but IPL must be performed from a CKD or ECKD disk.

LAN Access

3172 LCS

The 3172 LCS and LCS-compatible devices are fully supported for connectivity to Ethernet, Token-Ring and FDDI connectivity.

OSA/OSA-Express

All OSA adapters are supported (Ethernet, Token-Ring, ATM, Fast Ethernet). The OSA-Express driver is still under development and does not yet support gigabit Ethernet or other QDIO-based access methods.

Real or Virtual CTC Connection to IBM TCP/IP Implementation

Linux for S/390 can use a real or virtual CTCA connection to an existing IBM TCP/IP implementation. In this configuration, any network adapter supported by IBM TCP/IP may be used. Up to approximately 200 CTC interfaces can be supported on a single Linux for S/390 interface.

IUCV to IBM TCP/IP Implementation (VM Only)

When running under VM/ESA, an IUCV connection to VM TCP/IP may be used. IUCV offers a singificant performance boost (measured throughput of more than 350 MB/sec), but is limited to a single interface using the current device driver.

CLAW (Channel Attached Workstation)

A CLAW driver is under development in the open source community at this time. This driver would allow direct Linux interoperability with devices such as the Cisco CIP or CPA or use with a channel-attached RS/6000®. Expected availability of the driver is late 2000.

Performance Characteristics

Linux for S/390 is currently a work in progress, so performance numbers can be somewhat misleading at this time. Some observations from a large deployment are shown below.

Normal WSS

In most cases, the normal working set is approximately 1500 4 KB pages per instance, varying from a low of 1220 pages to as high as 1750 pages during heavy I/O periods where significant caching is performed (under VM).

CPU Consumption

CPU consumption for an idle Linux instance is approximately .001% of a G5 CPU per instance. Active instances can consume up to all available CPUs for a heavily threaded application (tested on a virtual 4-way system). Use of VM or Virtual Image Facility allows simple distribution and load balancing across CPU engines, providing significantly better throughput than LPAR or basic mode.

I/O Rates and Scheduling

Idle Linux instances perform 1-2 I/O operations/minute to maintain timer processing. Active instances can drive the I/O system to full capacity (in the native LPAR, VM or Virtual Image Facility scenarios).

Conclusions

Customers entering the Internet-driven marketplace need to react quickly and provide very reliable services to distinguish themselves from the worldwide melee of competitors in the marketplace. The cost and time to market advantages provided by the S/390, VM/ESA and Linux for S/390 solutions provide a powerful and flexible alternative to traditional solutions for businesses wanting the legendary reliability of the S/390 coupled with cutting-edge Internet technology. In turn, the same solution allows business value by allowing existing systems to be fully brought to bear on the same problems – there is no requirement to convert or replace existing systems with new, untested hardware or software solutions. Finally, the total operational cost of the environment is the most convincing argument for the S/390 solution – the solution optimizes resources, staff, and capabilities to provide the most powerful, reliable and flexible Internet solution available.

IBM Corporation
Integrated Marketing Communications,
Server Group
Route 100
Somers, NY 10589

Produced in the United States of America
10-00

G326-4010-00