# Leveraging workload consolidation for lower IT costs

## Why IT organizations use workload consolidation to resolve data center constraints and mitigate cost

### Workload consolidation reduces cost

Whether large or small, companies are seeking solutions to simplify IT operations and reduce cost. For many, consolidating workloads onto denser, centralized computing platforms is an effective way to decrease IT expense.

A major savings driver is the decrease in software costs. Typically, Linux workloads running on centralized servers such as LinuxONE and Integrated Facility for Linux (IFLs) on IBM Z® require fewer per core licenses due to per core pricing.[1]

*LinuxONE III and IBM z15 servers use **78%** fewer licenses for a competitive database versus the compared x86 environment by reducing the number of cores required to run the same transactional workload.[1]*

Another savings driver is energy efficiency. Workloads on IFLs on IBM Z and LinuxONE consume less energy compared to distributed server environments[2], reducing data center carbon footprint and improving Power usage effectiveness (PUE).

Consolidating x86 workloads onto a fewer physical servers also lowers floor space costs.[3] As distributed server environments grow to meet new business demands, floor space can become a significant expense, particularly when an IT organization has reached the physical limits of its data center and is considering a move to a larger facility.

*A single frame LinuxONE III or z15 saves an estimated **50%** in power consumption per year than compared x86 systems running workloads with the same throughput.[2]*

*A LinuxONE III LT2 or z15 single frame system requires **75%** less floor space than compared x86 2U servers in racks, running the same workloads and throughput.[3]*

Not only can workload consolidation lower software and data center costs, it can lower administrative overhead. Fewer physical servers can mean less hardware maintenance, less network management, and simpler software patching.

For most organizations, workload growth is inevitable. Centralized servers simplify the task of workload provisioning and deprovisioning by leveraging available capacity within the same physical server.

Most IBM Z and LinuxONE systems provide dormant capacity that can be activated on demand for rapid provisioning of new LPARs versus setting up a distributed server that requires procurement, installation, configuration, security administration, and workload deployment. Reliable disaster recovery for a distributed server environment can also become difficult as more servers with potentially different components, hundreds or thousands of parts, and new configurations are added over time. In contrast, a condensed server environment comprised of one or a few servers can facilitate replication for disaster recovery.

### Which workloads consolidate well

Organizations opting for workload consolidation to relieve cost and IT complexity tend to look for the following types of workloads.

1. **Workloads with per core pricing**
   Linux® workloads that have a software license price per unit of compute power (processor or socket) are strong candidates for consolidation on LinuxONE or IFLs on IBM Z from a financial perspective. This is due to differences in centralized versus distributed server architecture such as processor speeds, caching, HiperSockets™ for in-memory communication across LPARs, high levels of sustained CPU utilization and workload management capabilities. In general, distributed servers require considerably more processor cores to run the same Linux workloads than LinuxONE or IFLs on IBM Z. IBM internal tests and data from client environments show core consolidation ratios ranging from 10 to 32.5 distributed cores to one IFL[4], yielding dramatically lower software costs.
2. **Workloads with variable resource requirements**
   Linux workloads with activity fluctuations are very well suited for LinuxONE and IFLs. Centralized servers provide compute elasticity, or resource sharing, so that memory, CPU and I/O can be allocated to workloads with diverse timeline requirements over a 24-hour period.
3. **Workloads with I/O demands**
   Most business workloads consistently use I/O to perform their tasks (for example databases, messaging, and stream processing workloads). These workloads tend to be I/O driven and can accelerate response times by leveraging LinuxONE and IFL FICON® [5] or FCP protocols designed to enhance data transfer and to increase sustained CPU utilization through advanced workload management capabilities. FICON I/O capabilities such as multipathing[6] that automatically switches to an alternate path in event of an interruption, can alleviate administrative overhead for maintenance and network bottlenecks.

1. This is an IBM internal study designed to replicate a typical IBM customer workload usage in the marketplace. It consists of IBM LinuxONE III with 50 cores, 2,048 GB memory, z/VM, RHEL, and competitive database, compared to a comparably tuned x86 configuration with a total of sixteen x86 systems, each with 28 Intel Broadwell cores, using 768 GB memory, RHEL Linux, and competitive database executing a materially identical order fulfillment database workload in a controlled laboratory environment. The test for the database workloads, each running as a guest on z/VM in a logical partition, executed an identical SQL query transaction mix for a total throughput of 48,974 transactions per second. For the x86 configuration, the test measured the same number of database workloads, each running on bare metal and executing an identical SQL query transaction mix at a total throughput of 48,974 transactions per second. The results were obtained under laboratory conditions, not in an actual customer environment. IBM's internal workload studies are not benchmark applications.
2. Compared LinuxONE III LT1 model consists of 3 CPC drawers containing 108 IFLs, and one I/O drawer to support both network and external storage. Power consumption for the LinuxONE III LT1 is estimated using the Power Estimation Tool for 8561 https://www01.ibm.com/servers/resourcelink/hom03010.nsf/pages/pet8561v21507opendocument assuming a "Normal" workload. x86 systems ran at various CPU utilizations according to 15 customer surveys, representing Development, Test, Quality Assurance, and Production levels of CPU utilization and throughput. Three workloads were tested, consisting of a mix of leading databases and application servers. Each workload ran at the same throughput and SLA response time on LinuxONE and x86. Power consumption on x86 was measured while each system was under load. LinuxONE III LT1 performance data and number of IFLs were projected from actual LinuxONE Emperor II performance data including a performance improvement of 10% on LinuxONE III LT1. Compared x86 models were 78 2-socket servers containing a mix of 8-core, 12-core and 14-core Xeon x86 processors. External storage is common to both platforms and is not included in power consumption. Assumes LinuxONE and x86 are running 24/7/365. Power consumption may vary depending on factors including configuration, workloads, etc. Energy cost savings are based on a U.S. national average commercial power rate of $0.10 per kWh based on U.S Energy Information Administration (EIA) data, https://www.eia.gov/electricity/monthly/epm_table_grapher.php?t=epmt_5_6_a. Individual rates may vary. Savings assumes a power usage effectiveness (PUE) ratio of 1.66 to calculate additional power for data center cooling. PUE is based on IBM and the Environment - Climate protection - Data center energy efficiency data, https://www.ibm.com/ibm/environment/climate/datacenter_energy.shtml
3. Actual floor space covered by the systems includes doors and covers.The LinuxONE III LT2 consists of two CPC drawers containing 64 IFLs, and one I/O drawer containing 7 FCP and 3 OSA adapters versus 4 x86 racks, each occupying 16 2U slots to run the comparable workloads, consisting of a mix of databases and application servers. Each workload ran at the same throughput and SLA response time on LinuxONE and x86. x86 systems ran at various CPU utilizations according to 15 customer surveys, representing Development, Test, Quality Assurance, and Production levels of CPU utilization and throughput. 16 x86 2U form factor servers populated a standard 32U rack. Other 10 2U slots contained PDU, network switches, SAN switches and allowed space for air circulation. External storage floor space is not included. LinuxONE III LT2 performance data and number of IFLs was projected from actual LinuxONE Rockhopper II performance data including a 3% lower throughput using MIPS ratio on LinuxONE III LT2 versus LinuxONE Rockhopper II with high availability. Assumes LinuxONE and x86 are running 24x7x365 with 42 Development, Test, Quality Assurance, and Production servers and 9 High Availability servers. Compared x86 models were all 2-socket servers containing a mix of 8-core, 12-core and 14-core Xeon x86 processors.
4. 20 IT Economics assessments involving analysis of x86 workloads for consolidation onto IFLs on IBM Z or LinuxONE were selected from diverse industries (35% financial, 25% government, 5% healthcare, 10% retail, 10% technology, 10% transport, 5% utilities), and different geographies (5% North America, 15% Latin America, 30% Europe, 20% Asia Pacific, and 30% Greater China Group). The assessments included were performed for clients with business critical workloads running in production and non-production environments. The workloads targeted for consolidation from x86 and distributed servers were IBM and third party proprietary and open source databases, application server middleware and industry specific solutions running on different types of x86 and distributed servers. Each client engaged the IT Economics team to evaluate the distributed workloads and the proposed IFL or LinuxONE environment for the consolidation. For each assessment, IT Economics consultants met with the client to discuss consolidation planning and execution, analyzed the client's current total cost of ownership, and provided a projected total cost of ownership with workload consolidation based on estimated core consolidation ratios for the client's workloads. Consolidation analysis identified potential financial savings and IT efficiencies that enabled the client to move their workloads to IFLs or LinuxONE. For additional information on x86 workload analysis contact the IBM IT Economics team, IT.Economics@us.ibm.com.
5. IBM z15 FICON Performance, https://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/WP102812
6. I/O multipathing with LinuxONE and IFLs https://www.ibm.com/support/knowledgecenter/linuxonibm/com.ibm.linux.z.ldsg/ldsg_c_multipathing.html

4. **Workloads with high availability requirements**
Business critical workloads that require 24x7 availability are often placed in IFLs on IBM Z or LinuxONE to leverage built-in redundancy and resiliency.[7] Capacity Backup (CBU) allows hardware engines to be used for disaster recovery without incurring additional software charges if a server is temporarily unavailable.[8] Unlike a distributed architecture DR environment in which DR servers must remain online (and incur license costs), a CBU environment can remain offline and be brought up in minutes in the event of an outage.

   Additionally, IFLs and LinuxONE hardware used for disaster recovery environments cost less than hardware for production environments, resulting in increased disaster recovery savings.

5. **Workloads with low latency and high transaction requirements**
Many IT organizations keep their critical system of record data on IBM Z and leverage other platforms for their applications. If the data on z/OS® is used from applications on distributed servers, latency increases as the data is accessed by an off-platform environment. Overall application performance is reduced since the data must constantly access the system of record over TCP/IP.

   These applications are best collocated with the data on the same physical server as the system of record. The applications can run on IFLs on the same server and leverage HiperSockets[9] or Shared Memory Communication through TCP/IP, enabling greater bandwidth and lower latency compared to accessing the data over TCP/IP from distributed servers.

6. **Workloads with high security requirements**
Workloads that access sensitive data are typically placed on IBM Z or LinuxONE to minimize the possibility of a security event. Both IFLs on IBM Z and LinuxONE provide unique security benefits to lower the risk of a data or privacy breach with:

   - Hardware Security Module Crypto Express card certification at highest level 4 of FIPS 140-2[10]
   - Pervasive encryption features with HSM-based key management [11] and Secure Service Containers [12] to reduce security risks
   - Cryptographic coprocessors to deliver high throughput for cryptographic functions in crypto workloads [13]
   - z/VM® security features for virtualized workloads such as LDAP, RACF® and cryptography for Linux guests on z/VM [14]
   - IBM Data Privacy Passports to encrypt eligible data, grant, control, and revoke access to it, even as it moves off the system of record within your enterprise [15]

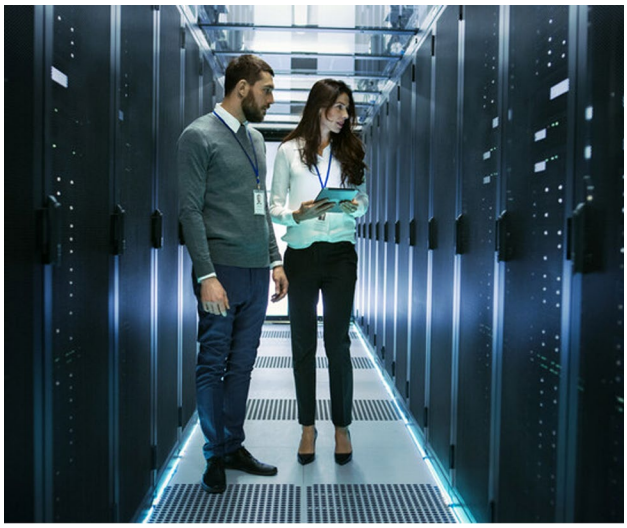
7. **Workloads headed toward the cloud**
Both new cloud native and existing workloads targeted for modernization for the cloud are good fits for IFLs on IBM Z or LinuxONE using IBM Cloud Paks™.

   IBM Cloud Paks allow new and existing workloads to be containerized and prepackaged using IBM Cloud Pak unique capabilities on the Red Hat® OpenShift® Container Platform. Each IBM Cloud Paks includes containerized IBM middleware and common software services for development and management, on top of a common integration layer designed to reduce development time and operational expenses.

   With Red Hat OpenShift Container/Kubernetes technology, containerized workloads can be densely packed to lower infrastructure costs and be easily managed, reducing operations expense. DevOps automation across the application delivery lifecycle brings higher productivity and efficiencies resulting in higher business values. Learn more about the capabilities of each IBM Cloud Pak for:

   - **Applications**
   IBM Cloud Paks for Applications helps to accelerate the build of cloud-native apps by leveraging built-in developer tools and processes, including support for microservices functions and serverless computing.
   - **Data**
   IBM Cloud Paks for Data helps to unify and simplify the collection, organization and analysis of data. Enterprises can turn data into insights through an integrated cloud-native architecture.
   - **Integration**
   IBM Cloud Paks for Integration helps support the speed, flexibility, security and scale required for all of your integration and digital transformation initiatives. It comes pre-integrated with a set of capabilities including API lifecycle, application and data integration, messaging and events, high-speed transfer and integration security.
   - **Automation**
   IBM Cloud Paks for Automation helps you deploy on your choice of clouds anywhere Kubernetes is supported, with low-code tools for business users and real-time performance visibility for business managers.
   - **Multicloud management**
   IBM Cloud Paks for Multicloud Management helps to provide consistent visibility, automation and governance across a range of hybrid, multicloud management capabilities such as event management, infrastructure management, application management, multicluster management, edge management and integration with existing tools and processes.
   - **Security**
   IBM Cloud Paks for Security helps to uncover hidden threats, make informed decisions about the risks they pose, and then respond faster to those threats — while leaving data where it is.

7. ITIC 2019 Global Server Hardware, Server OS Reliability Survey Mid-Year Update for LinuxONE and IBM Z found 0% annual unplanned server downtime of >Four Hours in 2019, https://itic-corp.com
8. https://www.ibm.com/it-infrastructure/z/software/pricing-resources, Backup, Disaster Recovery, and Capacity Backup Upgrade video https://www.youtube.com/watch?v=XPvd5OJTJ-Y
9. HiperSockets, http://www.redbooks.ibm.com/redbooks/pdfs/sg246816.pdf
10. https://www.ibm.com/blogs/systems/security-considerations-for-critical-environments/
11. https://www.ibm.com/downloads/cas/3V7EY7N9
12. http://www.redbooks.ibm.com/redbooks/pdfs/sg248447.pdf
13. https://www.ibm.com/security/cryptocards/hsms
14. http://www.redbooks.ibm.com/redbooks/pdfs/sg247471.pdf
15. https://www.ibm.com/us-en/marketplace/data-privacy-passports



**Workload consolidation in your IT organization**
Most IT organizations find that they can easily implement workload consolidation by using a phased approach so that over time, the majority of distributed workloads are converged onto one or a few centralized servers. If your organization is looking at how to get started with workload consolidation, or wants help analyzing their workloads for IT efficiencies, contact the IBM IT Economics team at IT.Economics@us.ibm.com for a no-charge assessment.

**Authors**
David Anderson
Chief IT Optimizer Architect and Software Designer
dfa@us.ibm.com

Christian Meimeic
Executive Consulting IT Economic Analyst
cmiemiec@us.ibm.com

Greg Sechuga
Executive IT Architect, Manager of Americas IT Economics Team
gsechuga@us.ibm.com

John Gustavson
CTO, Worldwide IBM IT Economics Consulting & Research
jgustav@us.ibm.com

Susan Proietti Conti
IBM Executive Project Manager, PMP®
sconti@us.ibm.com


IBM Corporation
New Orchard Road
Armonk, NY 10504
U.S.A.
0420



IBM IT Economics Disclaimer
IBM IT Economics Studies are presented as illustrations of how clients can estimate their associated costs for IBM products and non-IBM products. Study results are projections using a total cost of ownership analysis, including costs such as hardware, software, network, storage, facilities and other costs over multiple years. Most costs and performance characteristics are provided by you, the client.

If costs and performance characteristics are obtained from a supplier, it is based on published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the vendor of those products.

Actual costs and performance characteristics will vary by client. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the data used in the study.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

IBM IT Economics Study content is provided "AS IS" without warranty of any kind. Any reliance by you on the study and any results from the study is at your sole risk and will not create any liability or obligation for IBM.

19032019-USEN-03