

Estudio de plataforma de alto rendimiento como soporte de negocio en las empresas peruanas

Study of a high platform of performance as a business support in Peruvian companies

Recibido: abril 11 de 2017 | Revisado: mayo 10 de 2017 | Aceptado: junio 15 de 2017

VERÓNICA CASTRO AYARZA¹
NORMA LEÓN LESCANO²

ABSTRACT

This research raises the need for processing and storage of information in Peruvian companies, which make use of specialized centers with the necessary infrastructure that allows their services and core business processes to operate uninterruptedly. The method used for this research has been the analysis of the need for storage of eleven 11 Peruvian companies that are in the ranking of the top 500 in Latin America, including retail, financial services, education, media, Steel industry / metallurgy, agribusiness, health service, construction, pharmaceutical chemistry and manufacturing industry. The results obtained show that cluster technologies are needed, which allow a large amount of data processing and high availability of its services to reduce the risk of unavailability.

Key words: company, storage, information, technology, cluster

RESUMEN

Esta investigación plantea la necesidad del procesamiento y almacenamiento de información en las empresas peruanas, las cuales hacen uso de centros especializados con infraestructura necesaria que permita a sus servicios y procesos core de negocio, funcionar de forma ininterrumpida. El método seguido para esta investigación ha sido el análisis de la necesidad de almacenamiento de once empresas peruanas que se encuentran en el ranking de las quinientas mejores de América Latina, entre las que se encuentran empresas de retail, servicios financieros, educación, medios, siderurgia/metallurgia, agroindustria, servicio de salud, construcción, química farmacéutica e industria manufacturera. Los resultados obtenidos evidencian que, efectivamente, necesitan de tecnologías clúster, que les permitan realizar una gran cantidad de procesamiento de datos y tener alta disponibilidad de sus servicios, disminuyendo así el riesgo de indisponibilidad de los mismos.

Palabras clave: empresa, almacenamiento, información, tecnología, clúster

¹ Universidad de San Martín de Porres, Perú
Centro de Investigación FIA, USMP
Correo: veronica_castro2@usmp.pe
² Correo: nleonl@usmp.pe

La innovación empresarial es factor clave para garantizar la competitividad o supervivencia de las empresas. Las nuevas tecnologías de información y comunicación son fundamentales como soporte en el ámbito empresarial (Ochoa, Parada, & Verdugo, 2008). Competir en el mundo empresarial es un reto que las empresas peruanas asumen diariamente (González & Pacheco, 2015); por tanto, necesitan que sus procesos “Core” de negocio estén soportados por tecnología de información que aseguren la alta disponibilidad de sus servicios y aplicaciones con el objetivo de prevenir transacciones fallidas y capacidad de escalabilidad para desplegar rápidamente nuevos servicios.

Una de las soluciones tecnológicas existentes es el uso de un “clúster” que consiste en la unión de varios servidores, conocidos como nodos (IBM, 2012), que autorizan a las empresas a incrementar la disponibilidad de aplicaciones y servicios, permitiendo la contingencia con un segundo servidor en caso que el primer servidor haya fallado (IBM, 2017). Los clúster son la base de las costosas supercomputadoras, las cuales brindan soluciones para problemas de alta complejidad en las áreas de la ciencia e ingeniería. Sin embargo, al ser las supercomputadoras demasiado costosas para asuntos de negocio, las empresas están utilizando clúster. En el Perú, las principales empresas hacen uso intensivo de clúster, tanto para el procesamiento como el almacenamiento de la información de su negocio.

Un clúster consiste en la unión de varios servidores, conocidos como nodos, esto para proveer una mayor confiabilidad y capacidad. Existen dos tipos de clúster: el primero de ellos es llamado clúster failover o de alta disponibilidad, en que se tienen servidores de respaldo en caso que el servidor principal falle

y no se encuentre disponible. El segundo tipo de clúster es el de balanceo de carga, que como su nombre lo indica se encarga de distribuir el procesamiento de la información en varios servidores, evitando así la saturación de uno de ellos.

Los clúster están constituidos por servidores. Un servidor es un sistema que opera en una base hardware, la cual sirve como plataforma para implementar un servicio o una serie de servicios, según sea la necesidad del cliente. En los servidores las empresas suelen tener sus bases de datos, aplicaciones empresariales, educativas, páginas web, entre otros. Como ejemplo de una gama de servidores se tiene a los IBM Power System, los cuales son servidores que utilizan la tecnología de procesamiento Power, desarrollada por IBM. Con el fin de otorgar un alto rendimiento, escalabilidad, fiabilidad y capacidad de manejo para las cargas de trabajo comerciales (Bicas Caldeira, Cho, Cruickshank, & Grabowski, 2017). Esta gama de servidores han sido diseñados para las exigencias con respecto a almacenamiento y seguridad de los datos que necesitan hoy en día las empresas. Además se puede implementar servicios en la nube o realizar análisis para big data.

En la Figura 1 se muestra las características de un servidor Power System 795. Entre las más importantes se encuentra la cantidad de sockets, los cuales determinarán el total de procesadores que puede tener el servidor. También se encuentra la capacidad del procesador, la cual se mide en Gigahertz. Este servidor puede tener desde 24 hasta 192 núcleos (cores) de 3.7 GHZ cada uno o desde 32 hasta 256 núcleos de 4 Ghz cada uno. Así mismo, las capacidades de Ram se miden en bytes, esta capacidad puede llegar hasta 16TB. La capacidad de discos que también utiliza la misma medida puede llegar hasta 3628 TB.

IBM Power Systems

Power 795

Product Line	IBM Power 795
Machine type	9119-FHB
System packaging	24" system frame (+ expansion frames)
Microprocessor type	64-bit POWER7
# of processor sockets per server	4 – 32 (4 sockets per processor book)
Processor options GHz (cores/socket) # of cores Min. 24 active	3.7 GHz (6) 24 - 192 4.0 GHz (8) 32 - 256 4.25 GHz (4) ¹ 24 – 128 (TurboCore mode)
EnergyScale	Y
Level 2 (L2) cache per core	256 KB
Level 3 (L3) cache per core	4 MB (MaxCore mode) 8 MB (TurboCore mode) ¹
System memory (min physical/min active/maximum)	64 GB/32 GB/16 TB (1066 MHz DDR3)
Active Memory Expansion	Optional
Reliability, availability, serviceability	
Chipkill memory	Y
Service processor	Redundant SP and clock with failover
Hot-swappable disks in I/O drawer	Y
Dynamic Processor Deallocation	Y
Processor Instruction Retry	Y
Alternate Processor Recovery	Y
Dynamic deallocation: PCI bus slots	Y
Hot-plug PCI slots in I/O drawer	Y
Blind-swap PCI slots in I/O drawer	Y
Active Memory Mirroring	Y
Redundant hot-plug power	Y
Redundant hot-plug cooling	Y
Node Add, Node Repair, Memory Upgrade	Y
Dual VIOS	Optional
Capacity and expandability	
Capacity on Demand (CoD) functions	Y ^{M, M+}
PowerVM Express Edition	-
PowerVM Standard Edition	Optional
PowerVM Enterprise Edition	Optional
Maximum logical partitions/micro-partitions ^r	1000
Maximum system unit PCI slots	Use I/O drawers
Max PCI slots: system unit + PCI-X 12X I/O drawers	600 PCI-X
Max PCI slots: system unit + PCIe 12X I/O drawers	640 PCIe
System Unit: Disk bays media bays	Use I/O drawers
Maximum disk storage in system unit	Use I/O drawers
Maximum I/O loops (12X)	32
Maximum PCI-X 12X I/O drawers	30
Maximum PCIe 12X I/O drawers	32
Max disk drives (system unit+I/O drawers) Storage	4032 ⁿ 3628 TB with 900 GB drives
Maximum Ultra SSD I/O Drawer TB	N/A

Systems, IBM PureFlex and Power Blades

Fuente: (IBM, 2013)

Con el conocimiento de la capacidad que puede tener un servidor físico se puede tener una idea de la capacidad que puede llegar a tener un clúster considerando que, normalmente, cada servidor físico es virtualizado permitiendo así tener varios servidores virtuales dentro. Por lo tanto, de toda la capacidad del servidor físico se asigna cierta cantidad de recursos a cada servidor virtual. Un

nodo del clúster puede estar conformado por un servidor virtual dentro de un servidor físico.

Para asegurar la alta disponibilidad de los servicios en un clúster existen diversos software. Uno de ellos es PowerHA SystemMirror, que provee un ambiente de computación comercial, el cual asegura la rápida recupera-

Figura 1. IBM Power F

ción de los servicios ante fallas de hardware o software y permite acceso completo a los nodos, interfaces de red y al almacenamiento del servidor (IBM, 2014). PowerHA está basado en clientes VIO (Figura 2) el cual permite usar componentes virtuales como adaptado-

res Ethernet virtuales, discos SCSI virtuales y NPIV. Además, el almacenamiento compartido es solo presentado a los nodos del clúster y no a los servidores virtuales que se encuentran en el servidor físico (Quintero & Bodily, 2014).

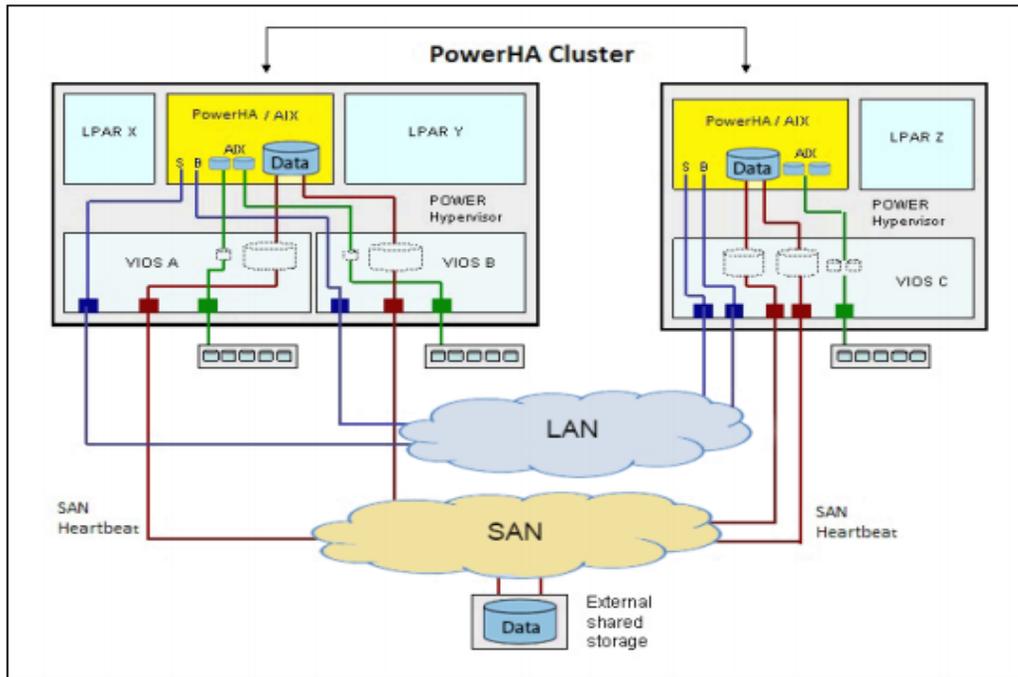


Figura 2. Cliente VIO Power HA

Fuente: (IBM, 2014)

Existen dos tipos de clúster: Failover clúster o de alta disponibilidad y clúster de balanceo de carga. El Cluster Failover cluster o de alta disponibilidad, cuya característica permite mover los recursos del nodo caído del clú-

ter al nodo que sí se encuentra activo. Como recursos del nodo del clúster nos referimos a IP de servicio, filesystem, entre otros (Fernández Troncoso, 2016).



Figura 3. Clúster Failover

Clúster de balanceo de carga, cuya característica permite distribuir todas las peticiones entrantes entre sus diferentes nodos, los cuales se utilizan como plataforma para un mismo servicio de aplicación o base de datos.

Si uno de los nodos cae, los otros nodos se encargarán de mantener el servicio disponible. Esta distribución se utiliza típicamente para ambientes de web-hosting (Werstein, Situ, & Huang, 2017).

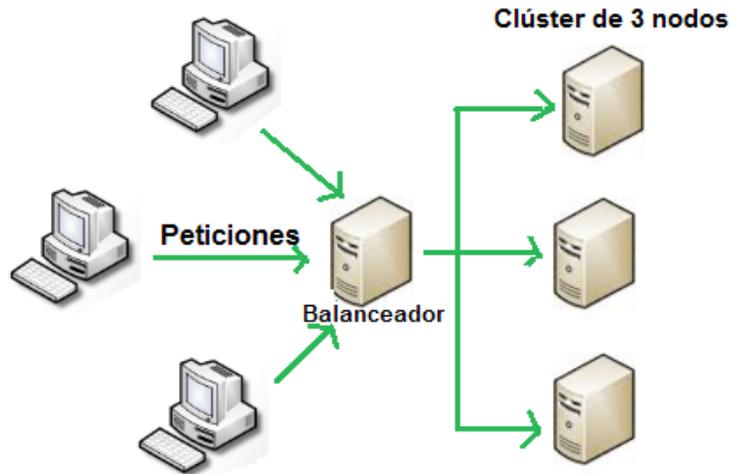


Figura 4. Clúster balanceo de carga

Los clúster se pueden clasificar según las soluciones de disponibilidad en base al tiempo de respuesta ante los fallos de un nodo de la

siguiente manera: Stand alone, enhanced Stand-alone y clúster de alta disponibilidad y tolerancia a fallos. (Quintero, Baeta, & Bodily, 2016)

Solution	Downtime	Data availability	Observations
Stand-alone	Days	From last backup	Basic hardware and software
Enhanced stand-alone	Hours	Until last transaction	Double most hardware components
High availability clustering	Seconds	Until last transaction	Double hardware and additional software costs
Fault-tolerant	Zero	No loss of data	Specialized hardware and software, very expensive

Figura 5. Comparativo de clasificación de clúster
Fuente: (Quintero & Bodily, 2014)

- a. Stand Alone. Tipo de clúster en el que se puede tardar días para que el servicio se encuentre disponible nuevamente. Además, la data que se tendrá será la del último backup realizado.
- b. Enhanced Stand-alone. Tarda horas para que el servicio se encuentre disponible en el clúster de contingencia.

La data disponible será la de la última transacción; es decir, hasta antes de que el nodo primario falle.

- c. Clúster de alta disponibilidad. El tiempo que se tarda en pasar los recursos de un nodo a otro es de segundos. Es decir, el servicio se encuentra disponible en segundos, en caso el nodo primario falle.

d. Tolerancia a fallos. El servicio se encuentra disponible todo el tiempo, ya que el tiempo de indisponibilidad de servicio es de cero. En este tipo de clúster no se pierde data.

de negocios América Economía, se estudió a las empresas peruanas cuyos servicios de sistemas de información utilizan como plataforma al clúster. Se tomó una empresa por cada sector de salud, retail, siderurgia/metallurgia, medios, servicios financieros, petróleo/gas, construcción, industria manufacturera, educación, agroindustria y química farmacéutica. Sumaron un total de 11 empresas las cuales se muestran en la Tabla 1.

Las empresas peruanas y el uso de los clúster

Del ranking de las 500 empresas más grandes del Perú, publicado en el 2014 por la revista

Tabla 1
Posición empresa peruana por sector/rubro

Empresa	Puesto en el ranking de las 500 empresas más grandes del Perú	Sector/Rubro
ESSALUD	6	Salud
INTERCORP	24	Retail
FERREYROS	20	Siderurgia/metallurgia
EL COMERCIO	134	Medios
SCOTIABANK	53	Servicios financieros
PERÚ LNG	30	Petróleo/Gas
JJC	139	Construcción
CELIMA	178	Industria Manufacturera
PUCP	173	Educación
CAMPOSOL	153	Agroindustria
UNIQUE	111	Química Farmacéutica

Fuente: (América Economía, 2014)

Se estudió los tipos de servicios que utilizan clúster en cada empresa y se priorizó a los servicios de almacenamiento de data, sistemas

transaccionales, sistemas de apoyo a la toma de decisiones. La clasificación se muestra en la Figura 4.

Empresa	Sector/Rubro	Servicios en los clúster
ESSALUD	Salud	Base de datos, aplicaciones del negocio
INTERCORP	Retail	Base de datos, ERP
FERREYROS	Siderurgia/metallurgia	Base de datos, ERP
EL COMERCIO	Medios	Base de datos, ERP
SCOTIABANK	Servicios financieros	Base de datos
PERÚ LNG	Petróleo/Gas	Base de datos, ERP
JJC	Construcción	Base de datos, ERP
CELIMA	Ind. Manufacturera	Base de datos, ERP
PUCP	Educación	Base de datos, Software BI empresariales
CAMPOSOL	Agroindustria	ERP
UNIQUE	Química Farmacéutica	Data empresarial, Base de datos

Figura 4. Servicios por empresa

Se realizó un estudio acerca de la cantidad de clúster y nodos por clúster que tiene cada empresa seleccionada. Las empresas del sec-

tor salud, farmacéutica y retail son las que las usan más nodos de clúster.

Tabla 3

Cantidad de clúster por empresa

Empresa	Sector/Rubro	Nº de clústers	Nº de nodos por clúster	Total de nodos
ESSALUD	Salud	5	2	10
INTERCORP	Retail	10	2	20
FERREYROS	Siderurgia/metalurgia	2	2	4
EL COMERCIO	Medios	4	2	8
SCOTIABANK	Servicios financieros	2	2	4
PERU LNG	Petróleo/Gas	3	2	6
JJC	Construcción	1	2	2
CELIMA	Ind. Manufacturera	1	2	2
PUCP	Educación	2	2	4
CAMPOSOL	Agroindustria	1	2	2
UNIQUE	Química Farmacéutica	6	2	12

Se realizó una investigación de la capacidad de almacenamiento en promedio por nodo de clúster utilizados por las empresas seleccionadas y se calculó la cantidad total de almacenamiento obtenida de la multiplicación

entre la cantidad de nodos y capacidad de cada nodo. Se encontró que el sector salud lidera esta clasificación, seguida por el sector de retail y, en tercer lugar, se encuentra el rubro de siderurgia/metalurgia.

Tabla 4

Capacidad de almacenamiento por nodo de clúster

Empresa	Sector/Rubro	Nº de clústers	Total de nodos	Capacidad de Total de almacenamiento (en GB)	Capacidad de almacenamiento por nodo (en GB)
ESSALUD	Salud	5	10	106860	10686
INTERCORP	Retail	10	20	115460	5773
FERREYROS	siderurgia/metalurgia	2	4	11280	2820
EL COMERCIO	Medios	4	8	27392	3424
SCOTIABANK	Servicios financieros	2	4	4384	1096
PERU LNG	Petróleo/Gas	3	6	8106	1351
JJC	Construcción	1	2	2954	1477
CELIMA	Ind. Manufacturera	1	2	4148	2074
PUCP	Educación	2	4	1972	493
CAMPOSOL	Agroindustria	1	2	1202	601
UNIQUE	Química Farmacéutica	6	12	15048	1254

Conocida la capacidad de almacenamiento que necesitan las empresas, se hace el estudio de las características adicionales de la capaci-

dad del nodo, como el número de cores de CPU por nodos, la velocidad de procesamiento de los datos y el tamaño de la memoria RAM.

Tabla 5
Características de nodos del clúster

Empresa	Sector/Rubro	Nº de Clústers	Total de nodos	Nº de cores de CPU (por nodo)	Velocidad de Procesador en MHz (por nodo)	Tamaño de memoria RAM en GB (por nodo)
ESSALUD	Salud	5	10	2	3500	110
INTERCORP	Retail	10	20	12	3612	64
FERREYROS	siderurgia/meta-lurgia	2	4	3	3550	40
EL COMER-CIO	Medios	4	8	5	4060	43
SCOTIA-BANK	Servicios financieros	2	4	3	3550	32
PERU LNG	Petróleo/Gas	3	6	4	3525	27
JJC	Construcción	1	2	5	4060	40
CELIMA	Ind. Manufacturera	1	2	3	3425	40
PUCP	Educación	2	4	2	3525	30
CAMOSOL	Agroindustria	1	2	3	4060	48
UNIQUE	Química Farmacéutica	6	12	5	3612	10

De la cantidad de core de CPU, la velocidad de procesamiento y la capacidad de memoria necesaria para soportar solo el servicio de almacenamiento de las empresas en estudio, se necesitaría aproximadamente 74 nodos de clúster conectados. Además, para el procesamiento de datos que utilizan estos servicios se estaría utilizando un total de 47 cores de CPU y 484 Gb de memoria RAM.

Conclusiones

Los resultados obtenidos evidencian que, efectivamente, se requieren de tecnologías clúster para soportar el almacenamiento de data de las empresas del medio. Además, se visualiza en la muestra que las empresas de salud son las que necesitan almacenar más información, ya que esta es útil para futuros

estudio sobre la real capacidad necesaria para el sector salud cuando toda ella se encuentre digitalizada e integrada a nivel nacional.

Sí, es necesario un esfuerzo de un súper computador para este sector. Las empresas del sector de retail y medios también ameritan ponerles atención; sin embargo, por ser empresas de sectores comerciales, se considera que la necesidad futura será absuelta con el aumento de clúster en la nube.

Referencias

América Economía. (2014). *Las mayores empresas del Perú y America Latina*. Obtenido de <https://rankings.americaeconomia.com/las-500-mayores-empresas-de-peru-2014/ranking-500/>

- Bicas Caldeira, A., Cho, Y., Cruickshank, J., & Grabowski, B. (2017). *IBM Power Systems E870 and E880*. IBM RedBooks.
- Cao, J. (2016). *IBM Power 7*. Houston: Rice University.
- Fernández Troncoso, C. (2016). *Clúster de Alta disponibilidad en Promox VE 4.2 con balanceador de carga e instalación de Pydio sobre LXC*. PROMOX.
- Gonzales, C., & Pacheco, F. (2015). *Políticas públicas de tecnologías de la información y comunicación para la internacionalización de las empresas peruanas de software : estudio de caso del período 2003-2013*. Lima: PUC.
- IBM. (3 de Abril de 2012). *Clustering: A basic 101 tutorial*. Obtenido de <https://www.ibm.com/developerworks/aix/tutorials/clustering/clustering-pdf.pdf>
- IBM. (10 de Julio de 2013). *IBM Power Facts and Features*. Obtenido de https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?infotype=PM&subtype=BR&appname=STGE_PO_PO_USEN&htmlfid=POB03022USEN&attachment=POB03022USEN.PDF
- IBM. (2014). *IBM PowerHA SystemMirror for AIX Cookbook*. Obtenido de <http://www.redbooks.ibm.com/redbooks/pdfs/sg247739.pdf>
- IBM. (2014). *PowerHA SystemMirror concepts*. IBM.
- IBM. (2013 de Mayo de 2017). *IBM*. Obtenido de <https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=POD03053USEN>
- IBM. (14 de Julio de 2017). *Using clusters for workload management*. Obtenido de https://www.ibm.com/support/knowledgecenter/en/SSFKSJ_7.5.0/com.ibm.mq.pro.doc/q002750_.htm
- Ochoa, J., Parada, E., & Verdugo, M. (2008). El uso estratégico de las tecnologías de información y comunicación para incrementar la innovación en las empresas. *Red Internacional de Investigadores en Competitividad*, 05-07.
- Quintero, D., & Bodily, S. (2014). *IBM PowerHA SystemMirror for AIX Cookbook*.
- Quintero, D., Baeta, S., & Bodily, S. (6 de Julio de 2016). *IBM PowerHA SystemMirror V7.2 for IBM AIX Updates*. Obtenido de <http://www.redbooks.ibm.com/abstracts/sg248278.html?Open>
- Werstein, P., Situ, H., & Huang, Z. (5 de Mayo de 2017). *University of Otago*. Obtenido de <http://www.cs.otago.ac.nz/staffpriv/hzy/papers/pdcat07.pdf>