

# Comparativo de la velocidad de transferencias digitales de archivos de métodos thread y multithread en red LAN

Angel Maldonado Sandez, M.C., Tecnología de Información, ing.sandez@gmail.com

Alejandro Guzmán Ocegueda, Dr., Consultor Empresarial, alex1guzman@yahoo.com

CETYS UNIVERSIDAD Tijuana, BC, México. 22210

#### **ABSTRACT**

Esta investigación aplica el enfoque DMAIC para el planteamiento del problema y estadística inferencial para el rechazo o aceptación de la hipótesis nula, abordamos una problemática en el ámbito de las tecnologías de la información sobre la eficiencia del método "thread" de trasferencia digital de archivos en un sistema cliente-servidor de desarrollo in-house.

#### I. INTRODUCCIÓN

Una empresa de manufactura utiliza un sistema in-house llamado archive record management (ARM), como repositorio digital de registros de producción. específicamente DHR's. contienen todos los registros relacionados con la producción de un dispositivo terminado dentro del alcance de la U.S. Food and Drug Administration (FDA). Esto conlleva una amplia documentación que incluye fechas de fabricación, cantidad fabricada y enviadas, registros de aceptación para mostrar que el dispositivo fue producido de acuerdo con los registros maestros del dispositivo (DMR),etiquetado de identificación identificación del dispositivo y números de lote [1]. Este proceso utiliza software e infraestructura para el registro, almacenamiento, propia búsqueda y visualización de los documentos digitalizados, ver Figura I, y actualmente la transferencia de archivos digitalizados al sistema decreció con el aumento del 100% de la demanda del 2017.

**RVP-AI/2018 PONENCIA** ES POR RECOMENDADA EL COMITE DE **APLICACIONES INDUSTRIALES** DEL CAPITULO DE **POTENCIA** IEEE DFI SECCION MEXICO Y, PRESENTADA EN LA REUNION INTERNACIONAL DE VERANO, RVP-AI/2018.

DMR: Cada fabricante debe mantener registros maestros del dispositivo, Device Master Record (DMR). Cada fabricante se asegurará de que cada DMR se prepare y apruebe de acuerdo con 820.40. El DMR para cada tipo de dispositivo incluirá o hará referencia a la ubicación de la siguiente información:

- (a) Especificaciones del dispositivo, incluidos dibujos apropiados, composición, formulación, especificaciones de componentes y especificaciones de software;
- especificaciones de (b) las del proceso producción. incluidas las especificaciones del equipo, apropiadas los métodos producción, los procedimientos de producción y las especificaciones del entorno de producción;
- (c) Procedimientos y especificaciones de garantía de calidad, incluidos los criterios de aceptación y el equipo de garantía de calidad que se utilizará;
  - (d) especificaciones de embalaje y etiquetado, incluidos los métodos y procesos utilizados; y
  - (e) Procedimientos y métodos de instalación, mantenimiento y servicio. [2]



**Figura I.** Descripción General del Proceso [3]

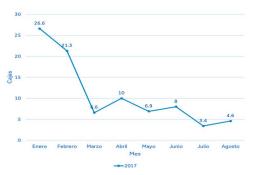
#### II. ANTECEDENTES

Una investigación previa con la metodología *DMAIC* abordaba perspectivas de infraestructura de red, servidores y métodos de copiado de la aplicación diagnosticó la causa raíz de los cuellos de botella en el proceso de trasferencia digital de archivos como principal factor.

DMAIC: (Definición del problema: El problema es identificado y los requerimientos del cliente son determinados, Medición: El proceso y sus defectos son cuantificados, Análisis: Los datos son estadísticamente revisados y la causa raíz es encontrada, Implementación: El proceso es modificado para remover la causa de defecto y Control: el proceso ahora mejorado es ejecutado de tal manera que el defecto no se repita.) [6]

El requerimiento de transferencia de archivos con línea base en el primer semestre del 2017 es un promedio de 45 cajas por semana conteniendo un promedio de 3,500 páginas cada caja. La línea base del nivel de desempeño en la velocidad de copiado de archivos digitalizados en formato PDF es desde 65 segundos hasta 2,722 segundos con un promedio de 996.3 segundos por caja de archivos con capacidad que van desde 70 MB hasta 1,537 MB con un promedio de 254.48 MB, este volumen digital representa físicamente en promedio 157,500 paginas por semana.

La capacidad promedio semanal de 22 cajas transferidas en 2016 decreció a 10 cajas promedio por semana en 2017 ver Figura II con la tendencia sin satisfacer el aumento en la demanda para transferir 45 cajas por semana causando un rezago de 1700 cajas no transferidas durante 2017; impactando en la disponibilidad digital de los DHR. Y requiriendo espacio para archivo muerto por 760 pies cuadrados, dos recursos asignados y cuatro horas para búsqueda física de cada DHR's cuando es requerido.



**Figura II.** Tendencia de la capacidad promedio semanal de cajas transferidas por mes entre enero y agosto 2017.

Considerando el rezago de la digitalización de documentos, esta investigación definió el siguiente problema de investigación.

## III. DEFINICIÓN DEL PROBLEMA

¿Qué método de copiado reduce el cuello de botella para transferir digitalmente Portable Document Format (PDF) al servidor desde una estación de trabajo por medio de sistema ARM en una red LAN?

Las hipótesis son:

H<sub>0</sub>: El método de copiado de archivos PDF thread en la aplicación ARM no influye sobre el cuello de botella en el proceso "Documento transferido a ARM".

H<sub>1</sub>: El método de copiado de archivos PDF thread en la aplicación ARM influye sobre el cuello de botella en el proceso "Documento transferido a ARM".

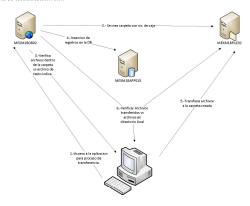
## IV. METODOLOGÍA

La metodología consistió en una investigación documental para identificar alternativas de métodos de copiado digital a través de una red computacional y una investigación de campo para la recopilación de tiempos de procesamiento en el ambiente de producción compañía para de una inferir estadísticamente la aceptación o rechazo de la hipótesis nula, con un intervalo de confianza del 95%, sobre el método de copiado más rápido. Y el modelo de solución de problemas aplicado fue DMAIC (Definición problema. Medición. Análisis. Implementación y Control) para este caso en Tecnología de la Información (TI). Sin embargo, la etapa de control, debido al tiempo de desarrollo para presentar este proyecto académico, será considerado posteriormente.

## V. DESARROLLO

Consideramos el flujo de comunicaciones entre dispositivos y la infraestructura de servidores existente en el ambiente de la empresa estudiada, representada en la Figura III, el experimento esta focalizado en el flujo de transmisión de datos de la estación de trabajo cliente hacia el servidor de archivos mexm1bfile02.

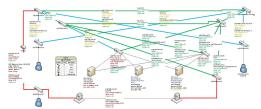
Ruta de comunicacion ARM



**Figura III.** Mapeo de flujo de comunicación utilizados por la aplicación cliente-servidor.

La representación de la infraestructura de red en la Figura IV, cuenta con las siguientes características de ancho de banda:

- Cliente-Switch de aplicación: 1Gbps
- Switch de aplicación-Switch de Distribución: 10Gbps
- Switch de Distribución-Switch de aplicación: 10Gbps
- Switch de aplicación-Servidor: 1Gbps



**Figura IV.** Mapeo lógico de infraestructura de red.

Durante el experimento observamos que el tráfico de red no excedió el 1.5% del ancho de banda de 1Gbps.

Con el dato histórico de 2,321 cajas trasferidas por medio del sistema ARM en el año 2017, la muestra utilizando la fórmula n =  $[z_{a>2}]^2p^2q^2$  /  $E_2$  [4] con un promedio semanal de 45 cajas, nivel de confianza de 95% y un margen de error de 5% resulta en 40.

La experimentación fue dos tecnologías de copiado para la transferencia de archivos de una estación de trabajo cliente a un servidor con las características mostradas en la Tabla I.

**Tabla I.** Características de hardware de cliente y servidor de archivos.

-	Cliente	Servidor	
OS	W7 x64	WS2008 R2 x64	
CPU	4 Cores @ 3.40GHz	8 Cores @ 3.50GHz	
RAM	8 GB	16 GB	
Red	1Gbps	1Gbps	

La masificación e incremento de potencia en arquitecturas hardware multinúcleo programas aumentó demanda la de multithread para aprovechar la capacidad ociosa y eficientizar el procesamiento de grandes cargas de trabajo [7, 9]. Por este motivo la elección fue la tecnología multithread como método alternativo y el segundo método utilizado no-multithread fue la línea base. La muestra de las 40 cajas fue utilizada para registrar los tiempos de transferencia de cada caja de muestras para analizarlo estadísticamente. Como lo han hecho otros departamentos de TI. [5]

## VI. RESULTADOS

la Tabla II muestra la velocidad de transferencia de 40 cajas con las siguientes características:

*ID Muestra*: Número identificador de la muestra y sus registros.

ID Caja: Nomenclatura interna asignada al folder digital para identificación de la caja digitalizada. Último renglón de la columna expresa la cantidad total de elementos en la tabla.

Cantidad de Archivos: Cantidad de archivos dentro del folder que representa la caja digitalizada. Último renglón de la columna expresa la suma total de los archivos en la columna.

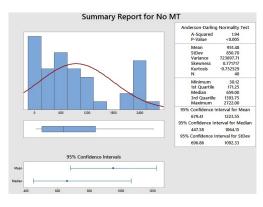
Transferencia MT Segundos: Registro de duración de la transferencia en segundos utilizando un método multithread como método alternativo. Último renglón de la columna expresa el promedio de la duración en esta columna.

Transferencia No MT Segundos: Registro de duración de la transferencia en segundos utilizando un método no multithread como método utilizando por el sistema en la línea base. Ultimo renglón de la columna expresa el promedio de la duración en esta columna. MB por caja: Tamaño del folder en disco duro expresado megabytes. Último renglón de la columna expresa la suma total de los megabytes en la columna.

**Tabla II.** Registros de datos sobre las muestras

ID ,	ID	MB por	Cantidad	Transferencia	Transferencia MT
Muestra 🕶	Caja 🐣	Caja	de Archivos	No MT Segundos *	Segundos
1	2277	117	146	490	15
2	345E	113	405	211	25
3	10232	142	226	2373	18
4	4164	355	165	919	49
5	388E	104	392	1258	21
6	732	152	142	158	31
7	10229	207	174	2184	45
8	9475	174	194	123	51
9	9065	155	91	222	17
10	9238	116	244	65	23
11	3660	297	178	626	55
12	0653	135	111	2370	25
13	407E	118	404	80	25
14	390E	180	405	692	38
15	725	136	122	105	25
16	722	143	134	94	26
17	931	196	120	2265	32
18	9045	182	102	614	29
19	2828	593	107	1164	83
20	3471	276	198	558	31
21	7364	127	790	131	32
22	8712	363	1,898	280	87
23	5006	797	2,204	966	200
24	3345	101	123	979	20
25	5579	195	131	1007	35
26	9254	332	124	2722	40
27	4819	1537	177	1107	153
28	7426	798	117	570	80
29	9247	267	132	2357	29
30	120P	320	115	1104	48
31	1253	84	154	1439	14
32	3915	163	78	832	29
33	4824	181	183	2038	39
34	872	179	122	2278	32
35	3655	126	145	418	27
36	980	163	165	600	34
37	3002	206	116	2481	43
38	2407	107	109	1812	21
39	10152	70	528	70	31
40	254E	172	166	79	53
	40	254.48	11,637	996.03	4

La Figura V muestra el resumen estadístico de los resultados en las 40 muestras para el grupo No MT el cual utilizó el método en el sistema ARM no multithread con una media  $\mu$  = 951.48 segundos. Y la Figura VI el resumen estadístico para el grupo MT el cual utilizó el método alternativo multithread con una media de  $\mu$  = 40.731 segundos.



**Figura V.** Resumen estadístico – Grupo No MT

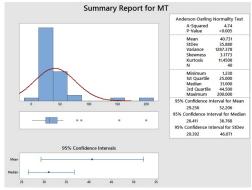


Figura VI. Resumen estadístico – Grupo MT

El análisis de las Figuras V, VI y VII, y la prueba t Student, Figura VIII comprobó estadísticamente las diferencias entre las medias de nuestras dos muestras, con un valor p de 0.000, menor al nivel de significancia de 0.05, permite rechazar la hipótesis nula asumiendo  $H_1$  la hipótesis alternativa.

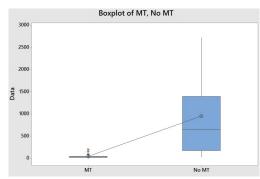


Figura VII. Análisis grafico - Boxplot

```
Two-Sample T-Test and CI: MT, No MT
```

```
Two-sample I for MI vs No MI

N Mean StDev SE Mean
MI 40 40.7 35.9 5.7
No MI 40 951 851 135

Difference = µ (MI) - µ (No MI)
Estimate for difference: (-1183, -638)
T-Test of difference: (0 vs #): T-Value = -6.76 P-Value = 0.000 DF = 39
```

Figura VIII. Comparación de 2 muestras

#### VII. CONCLUSIONES

El método de copiado de archivos PDF utilizado en la aplicación ARM influye significativamente en el cuello de botella del proceso "Documento transferido a ARM" por lo que la recomendación para el propietario de la aplicación es la modificación del método de copiado a multithread en la aplicación para reducir hasta un 95.7% el promedio del tiempo de ciclo en el proceso "Documento transferido a ARM".

#### VIII. REFERENCIAS

- Office of the Federal Register. (April 2017). Title 21, Volume 8, CITE: 21CFR820.184, February 2018, De U.S. Food & Drug Administration Base de datos.
- Office of the Federal Register. (April 2017). Title 21, Volume 8, CITE: 21CFR820.181, February 2018, De U.S. Food & Drug Administration Base de datos.
- Ramiri4. (2017). 10089322DOC Manual de Uso Para Sistema de Archive Records Management [Figura]. Recuperado de sistema electrónico de control de documentos Medtronic.
- 4. Mario F. Triola. (2009). Estadística. México: Pearson Educación.
- Chang-Tseh Hsieh, Bill Manduca, Binshan Lin. (05 Jan 2016). Information Technology and Six Sigma Implementation. Journal of Computer Information Systems, 27, p.1-10.
- Creveling, C., Slutsky, J., Antis, Jr, D. (2003). Design for Six Sigma in Technology and Product Development.

- Prentis Hall, Upper Saddle River, New Jersey.
- Tongping Liu, Charlie Curtsinger, Emery D. Berger. (2011). Dthreads: efficient deterministic multithreading. New York, NY, USA: O'Reilly.
- Magro, William; Petersen, Paul; Shah, Sanjiv. (2/14/2002). Hyper-Threading Technology: Impact on Compute-Intensive Workloads. Intel Technology Journal, ol. 6 Issue 1, p1. 9p.

## IX. AUTORES



M.C. Angel Maldonado es Ingeniero en Diseño Gráfico Digital y Maestría en Ciencias en Administración Industrial

por CETYS Universidad cuenta con más de 8 años de experiencia en las tecnologías de la información dentro de la industria de manufactura en giros metalmecánico, electrónico, automotriz y dispositivos médicos. Actualmente responsable de operaciones y proyectos de infraestructura de IT, liderando grupos multidisciplinarios geográficamente dispersos para entregar soluciones de negocios.



Dr. Alejandro Guzmán Ocegueda, Doctorado en Ingeniería, es consultor empresarial del Grupo ASIDE con más de 20 años de

experiencia, www.asidegroup.com y catedrático en CETYS Universidad en los programas de Maestría en Administración de Materiales y Logística, Maestría en Ciencias en Administración Industrial y Maestría en Ciencias de la Ingeniería.