**Estudio de simulación de un sistema de generación de reportes**

***Universidad Tecnológica Nacional, Facultad Regional Buenos Aires***

**Abstract**

*El presente estudio parte de la necesidad de obtener información sobre las modificaciones necesarias a implementar en un sistema de generación de reportes de una importante entidad financiera, con el fin de optimizar su funcionamiento y disminuir los tiempos en los que este sistema debe ser apagado porque los mensajes encolados son mayores al límite aceptado por la entidad en cuanto a desfase de información. Basándonos en la utilización de una simulación computacional, con la metodología Δt constante y el análisis de datos mediante la herramienta EasyFit buscaremos determinar cual es la cantidad óptima de instancias a utilizar en este sistema, como información para la toma de decisión también buscaremos obtener el porcentaje de tiempo en que el generador estuvo apagado, el tiempo máximo en el que estuvo apagado y el mayor y menor lag que se tuvo en la cola de procesamiento.*

**Palabras Clave**

Instancia. Metodología ΔT Constante, Simulación, Computacional, Generador de reportes, EasyFit.

**Introducción**

Nuestro modelo de estudio tiene su origen en un sistema de generación de reportes, detallaremos el funcionamiento actual de dicho sistema y la problemática que deseamos solucionar.

Este sistema de generación de reportes genera reportes solicitados por los usuarios de forma online y se alimenta de los movimientos de una cola, uno de los reportes más importantes, llamado “Todas las Transacciones” muestra todos los movimientos realizados y para mantener la consistencia del mismo, cuando en la cola de procesamientos, el número de transacciones a procesar supera una cierta cantidad de mensajes, el cual llamamos “lag alto”, se deben apagar los generadores de reportes.

Actualmente el sistema cuenta con cuatro instancias de procesamiento activas, las cuales procesan todas las transacciones y se puede observar que durante la hora pico, cuando las colas de procesamiento superan el lag máximo admisible se deben apagar los generadores de reportes a fin de asegurar la consistencia de los datos mostrados en los mismos.

Cuando las colas bajan la cantidad de mensajes en espera de procesamiento a un lag mínimo, los generadores de reportes se vuelven a encender.

Este proceso de prendido y apagado de los generadores se da de forma manual, por lo que a través de la simulación buscamos automatizar este proceso, para estos fines necesitamos poder determinar la cantidad de instancias posibles a tener activas, el threshold máximo y mínimo a setear.

Otro punto a tener en cuenta, es el tiempo máximo en que los generadores estén apagados, dado que no se desea que superen un cierto tiempo, porque también afecta la consistencia de los informes, por lo que este dato se debe obtener en la simulación.

**Elementos del Trabajo y metodología**

Para el análisis del caso de estudio, se aplicó la técnica de simulación con la metodología de avance en el tiempo Δt constante, de esta manera se pudo predecir el comportamiento a partir de diferentes factores externos.

Se logró la mayor representación del modelo a partir de la combinación de simplicidad y realismo.

1. Metodología

Para el avance del tiempo, en esta simulación, se utilizó la metodología Δt constante, la cual se caracteriza porque los datos se miden en densidades.

A partir del análisis previo y la clasificación de variables datos, y cómo las mismas estaban dados en volúmenes por minuto, decidimos optar por esta metodología.

1. Formulación del problema

El presente estudio busca resolver dos grandes problemas, por un lado necesitamos poder automatizar el prendido y apagado de los generadores de reportes. Y por otro lado, se necesita reducir los tiempos en los que los generadores de reportes se encuentran apagados con el fin de mantener la consistencia de los informes que estos generan.

Una de las condiciones por parte de la entidad es que el número de instancias no debe crecer tanto, dado que se suman costos de implementación y mantenimiento, los cuales no están dispuestos a solventar, para esto debemos lograr presentar una opción que sea viable tanto a nivel consistencia de datos de los informes, como los costos que la solución planteada pueda generar.

1. Recolección y procesamiento de información

Partimos de un trabajo de recolección de datos, con fin de lograr resultados que sean válidos para la toma de decisiones, se tomó un número de registros lo suficientemente grande.

Como datos de interés, obtuvimos los registros de los mensajes que llegan y se procesan en las instancias, relevamos la información referente al funcionamiento de las mismas, el RPM (Request Por Minuto), que hace referencia a la cantidad de registros que cada instancia puede procesar por minuto, cantidad de instancias disponibles en el sistema actual, umbrales mínimos (TLL) y máximos (THL) que hoy tienen configurados.

Para obtener las funciones de densidad de probabilidad que representa a los mensajes que llegan en la cola, se utilizó el software EasyFit.

Durante este procesamiento de información, nos encontramos con tres funcionamientos distintos en la llegada de movimientos a la cola.

Por un lado, pudimos observar que tenemos una diferenciación de hora pico (9 hs a 23 hs de lunes viernes), hora no pico (el resto del día y fines de semana) y dentro de la hora pico, observamos que los jueves ingresa un flujo muy diferente de movimientos, por lo que decidimos dividir el procesamiento de datos en estas tres situaciones.

1. Formulación del modelo

Basándonos en los datos obtenidos, se realizó el análisis previo del modelo, donde podemos diferenciar las distintas variables.

Las variables que intervienen se clasifican en dos grupos:

* Exógenas: son las independientes o de entrada del modelo. Estas variables actúan sobre el sistema, pero no reciben acción alguna por parte de él. Se dividen en:
  + Datos: Las variables no controlables son los datos. Los datos se toman de la realidad y son estudiados y expresados como fdp.
  + Control: Las variables controlables son susceptibles de manipulación o control por quienes toman decisiones o crean políticas para el sistema. Estas variables permiten ensayar diferentes escenarios.
* Endógenas: Las variables endógenas son variables que se generan dentro del modelo, dependen de la evolución del sistema, varían con el transcurso del tiempo. Se dividen en:
  + Estado: Estas variables describen el estado del sistema en la línea de tiempo, son aquellas que se quiere monitorear, que evolucionan y esa evolución es importante estudiar a lo largo del tiempo.
  + Resultado: Las variables de resultado son las de salida del sistema. Son generadas por la interacción de las variables exógenas con las de estado, durante la evolución del sistema simulado.

En nuestro sistema encontramos la siguiente clasificación de variables:

|  |  |
| --- | --- |
| **Clasificación de variables** | |
| Exógenas | |
| Datos | VMP: Volumen de Movimientos por minuto.  VPMJ: Volumen de Movimientos por minuto del día jueves.  VMPNP: Volumen de Movimientos por minuto en horas no pico |
|
|
| Control | CI: Cantidad de instancias de procesador de movimientos (Unidades)  TLL: Threshold low lag (Unidades)  THL: Threshold high lag (Unidades) |
| Endógenas | |
| Estado | PTP: Porcentaje de tiempo que los generadores estuvieron pausados  TMP: Tiempo máximo que el generador estuvo pausado.  MYL: Mayor lag en la cola de movimientos  MNL: Menor lag en la cola de movimientos |
|
|
| Resultado |
|

*Tabla 1: Clasificación de variables*

1. Diseño del modelo

Se va a modelar el sistema de generación de reportes considerando las solicitudes que ingresan a la cola y deben procesarse en las instancias de generación.

Los mensajes llegan a la cola de procesamiento con un determinado volumen por minuto, las instancias pueden procesar los datos de acuerdo a los request por minuto que tienen configurado.

De esta forma, podemos identificar un ingreso de movimientos y un consumo de movimientos dentro de la cola de procesamiento.

A partir del análisis del problema, podemos generar la tabla de clasificación de eventos y la tabla de eventos futuros, ambas correspondientes a la metodología delta t constante.

|  |  |
| --- | --- |
| **Eventos propios** | **Consumo de Movimientos**  **Llegada de Movimientos** |
| **Eventos comprometidos en ΔT anteriores** | - |
| **Eventos comprometidos para ΔT futuros** | - |

*Tabla 2: Clasificación de Eventos*

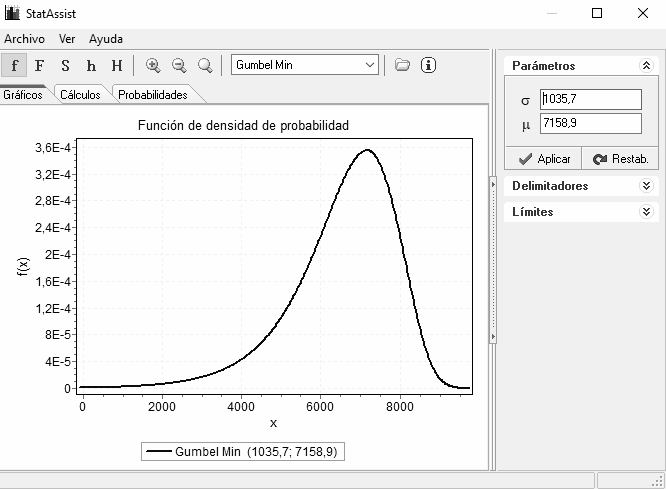
A partir de la determinación de la tabla de clasificación de eventos, podemos observar que al no tener eventos comprometidos en Δt anteriores y por ende, no se tienen eventos comprometidos para Δt futuros, la tabla de eventos futuros no es necesario realizarlo dado que no va a tener valores.

El objetivo de esta simulación es obtener el número adecuado de instancias de generación de reportes necesarias, el threshold máximo y mínimo, que hacen referencia a la cantidad máxima y mínima de movimientos esperando a ser procesados en la cola, a fines de apagar o prender los generadores, para que estos estén el menor tiempo posible apagados, sin aumentar en gran medida los costos.

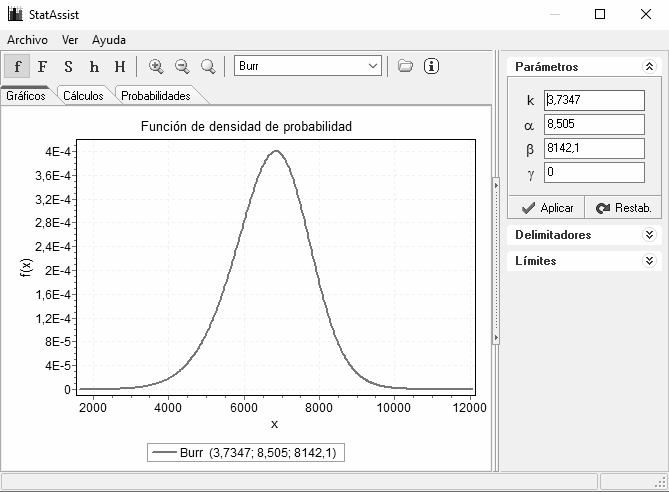
Luego de haber realizado el análisis correspondiente, se procede a realizar el diagrama de flujo de la simulación.

1. Funciones de densidad de probabilidad:

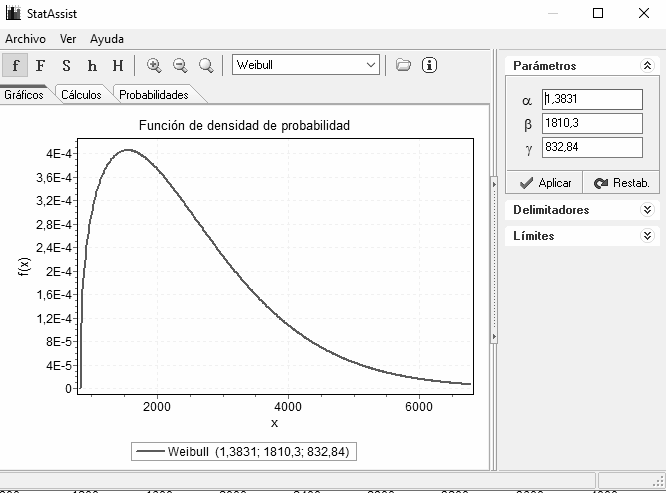
Las variables VMP, VPMJ y VMPNP responden a funciones de densidad de probabilidad (fdp). Tomando como punto de partida datos obtenidos del sistema, pudimos realizar el cálculo de la fdp utilizando la herramienta EasyFit. Estas fdp nos permitirán realizar el avance de tiempo en nuestro modelo.



*Figura 1: Ingreso de movimientos a la cola de procesamiento los lunes, martes, miércoles y viernes en hora pico. (VMP)*

**

*Figura 2: Ingreso de movimientos a la cola de procesamiento los jueves en hora pico. (VMPJ)*



*Figura 3: Ingreso de movimientos a la cola de procesamiento en hora pico. (VMPNP)*

1. Escenarios estudiados:

Se definieron los diferentes escenarios posibles para poder realizar las pruebas utilizando el modelo desarrollado y poder conocer el número adecuado de instancias de generación de reportes necesarias para satisfacer las necesidades de los usuarios.

En los diferentes escenarios iremos modificando la cantidad de instancias de generación activas para ir observando los diferentes comportamientos del sistema, también modificaremos el threshold máximo y mínimo.

El primer escenario a analizar será el de la realidad, con el fin de poder comparar los resultados obtenidos.

En la siguiente tabla se muestran los escenarios planteados:

|  |  |  |  |
| --- | --- | --- | --- |
| **Escenario** | **CI** | **TLL** | **THL** |
| 1 | 4 | 2500 | 1000 |
| 2 | 6 | 2500 | 1000 |
| 3 | 6 | 1000 | 500 |
| 4 | 8 | 1000 | 500 |
| 5 | 4 | 5000 | 1000 |
| 6 | 4 | 5000 | 5000 |
| 7 | 5 | 2500 | 2000 |

*Tabla 4: Escenarios estudiados*

**Resultados**

A partir del análisis de los diferentes escenarios simulados, obtuvimos diversos resultados.

Como indicamos en secciones anteriores, buscamos mantener la consistencia de los reportes, por lo que para la toma de decisión necesitamos medir el tiempo promedio en que las instancias estuvieron apagados, la mayor cantidad de registros esperando en la cola cuando se apagaron las mismas y la cantidad óptima de instancias teniendo en cuenta que no incrementen el costo de implementación y mantenimiento de las mismas en gran escala porque no está dentro de lo admisible por la entidad.

Considerando el primer resultado que nos interesa estudiar, TMP, podemos observar que al incrementar la cantidad de instancias y variando los THL y TLL de generación el tiempo promedio en que los generadores estuvieron apagados se modifica drásticamente. Podemos observarlos en los siguientes resultados:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Escenario** | **CI** | **THL** | **TLL** | **TMP** | **PTP** |
| 1 | 4 | 2500 | 1000 | 73 min | 7.81% |
| 2 | 6 | 2500 | 1000 | 7min | 0.31% |
| 3 | 6 | 1000 | 500 | 8min | 1.23% |
| 4 | 8 | 1000 | 500 | 3min | 0.18% |
| 5 | 4 | 5000 | 1000 | 54min | 3.02% |
| 6 | 4 | 5000 | 5000 | 26 min | 1.35% |
| 7 | 5 | 2500 | 2000 | 16min | 0.98% |

*Tabla 5: Resultados obtenidos para TMP(Mayor tiempo pausado) y PTP(porcentaje de tiempo pausado)*

En la tabla anterior podemos observar que cuando el sistema muestra distintos resultados de acuerdo a las combinaciones que se den entre CI, TLL y THL.

En la siguiente tabla, se anexan los resultados obtenidos para los mismos escenarios para las variables MYL y MNL.

|  |  |  |  |
| --- | --- | --- | --- |
| **Escenario** | **CI** | **MYL** | **MNL** |
| 1 | 4 | 21858 | 832 |
| 2 | 6 | 18770 | 833 |
| 3 | 6 | 22907 | 833 |
| 4 | 8 | 17533 | 833 |
| 5 | 4 | 21685 | 834 |
| 6 | 4 | 22841 | 833 |
| 7 | 5 | 21046 | 832 |

*Tabla 5: Resultados obtenidos para MYL(Mayor lag en al cola de movimientos) y MNL(Menor lag en al cola de movimiento)*

**Discusión**

A partir de los resultados indicados anteriormente, podemos observar que en la situación actual, correspondiente al escenario 1, nos encontramos un tiempo máximo de pausado del generador (TMP) de 73 min que representa un 7.81% (PTP) del tiempo total simulado y el número máximo de movimientos esperando en la cola cuando se pausó fue de 21858 (MYL), en base a estos valores que consideramos altos, principalmente el TMP, iniciamos las distintas simulaciones.

Podemos observar que si mantenemos la misma cantidad de instancias y variamos los threshold(escenarios 5 y 6) disminuye el TMP, siendo para el escenario 5 54min y para el escenario 6 26 min, en ambos casos podemos ver que el MYL es parecido al escenario actual.

Si aumentamos las instancias el TMP disminuye considerablemente, esto lo podemos observar en el escenario 2, donde al aumentar las instancias a 6, vemos que baja a 7min con un tiempo total de pausado durante la simulación de 0.31%, lo cual lo puede convertir en un escenario ideal, pero es probable que sea rechazado por la entidad dado que se incrementa en 2 la cantidad de instancias a implementar.

Si bien, sabemos que no es opción para la fintech aumentar las instancias, decidimos realizar 3 simulaciones donde aumentamos en 2 y 4 instancias, para demostrar que el problema se puede resolver de más de una forma.

Por último, en el escenario 7, podemos observar que solo tuvimos que aumentar 1 instancia y obtenemos una disminución del TMP de mas de un 80% en comparación a la situación actual.

Dado que el MYL es una variable que hoy utilizan como señal de alerta para realizar el apagado manual de los generadores de reportes, que este resultado no varíe en drásticamente en los distintos escenarios simulados, no es para preocuparse dado que es el lag admisible por la entidad, por lo que no haremos mención del mismo a la hora de las conclusiones y toma de decisión.

**Conclusión**

Dado que el objetivo principal del presente estudio es poder tener la información necesaria para la automatización del apagado y encendido del generador de reportes, concluímos que con pocos cambios, el escenario 7, que contiene solo una instancia más a la actual, y modificando los umbrales para que tengan números más cercanos (oscile rápidamente el apagado y el encendido del generador) nos va a dar un sistema con poco tiempo de apagado y consistente a la hora de generar reportes y está dentro de los márgenes aceptables por la entidad, tanto a nivel costos como así también a nivel consistencia de los informes y respetando el mayor lag admisible por los usuarios.

**Agradecimientos**

A la profesora Ing. Silvia Quiroga, al ayudante Rubén Flecha y a la cátedra de Simulación de UTN FRBA por ofrecernos las herramientas necesarias para realizar el presente estudio, el cual nos otorgó experiencia sobre el trabajo de un simulador.

A Laura Gerber, por la charla sobre la expresión de nuestras ideas de forma clara.

**Datos de Contacto:**

*Pompey Pablo Sebastián - UTN FRBA -*

[*pompeypabl@frba.utn.edu.ar*](mailto:pompeypabl@frba.utn.edu.ar)

*Sanchez Mauricio Ramon - UTN FRBA -*

[*maurisanchez@frba.utn.edu.ar*](mailto:maurisanchez@frba.utn.edu.ar)

*Villegas Maryann - UTN FRBA -*

[*mvillegasromero@frba.utn.edu.ar*](mailto:mvillegasromero@frba.utn.edu.ar)