

USO DE LA TEORÍA DE RESTRICCIONES Y MANEJO DE COLAS AL INICIO DE LA GESTIÓN DE LA FUNCIÓN MANTENIMIENTO

Rafael Suárez

rafael_suarez@cargill.com

Cuando se nos asigna la dirección de un departamento de mantenimiento cuyo estado se caracteriza por ser los héroes del sobretiempo y el arreglo rápido de los problemas, por fallas que, en general, se atienden después de que ocurren, por empleados que están exhaustos y por costos excesivamente altos, podemos afirmar que el departamento se encuentra en una condición de “mantenimiento reactivo”. Una de las mejores maneras de afrontar los inicios de la gestión para pasar a la condición de “mantenimiento planificado” es utilizando métodos sencillos de optimización de cadenas de valor como la teoría de restricciones y los modelos de colas.

La teoría de restricciones se centra, básicamente, en la búsqueda del flujo perfecto de bienes o servicios a través de una cadena de valor balanceada, coordinada y sincronizada de estaciones de trabajo, logrando así, bajar los costos de operación, reducir los inventarios y aumentar las ventas.

En el caso de un departamento de mantenimiento aplica todo lo anterior, ya que por un lado, la función mantenimiento no es otra cosa que un conjunto ordenado de actividades que agregan valor a un “servicio prestado”, desde una condición inicial conocida hasta una condición final que debe cumplir con los parámetros de calidad y seguridad establecidos o convenidos entre el ente encargado de agregar valor (Departamento de Mantenimiento) y el que lo recibe (clientes internos) en un tiempo determinado. Siendo el servicio prestado un bien *intangibile* podemos tomar la “solicitud de trabajo” como elemento tangible del servicio en su condición inicial, la “orden de trabajo en proceso” como elemento tangible del servicio en tránsito y la “orden de trabajo culminada” como elemento tangible del servicio en su condición final.

Por otro lado, en cuanto a los resultados esperados, podemos establecer una relación directa, ya que, los costos de operación son los gastos de mantenimiento, los inventarios estarían representados por las ordenes de trabajo (sin atender y en tránsito) y las ventas se pueden asociar directamente con el nivel de servicio (cantidad de solicitudes atendidas en una cantidad de tiempo determinada).

Según la teoría de restricciones, el primer paso para la implantación es identificar los cuellos de botella (estaciones de servicio que determinan la velocidad del flujo en el sistema), luego se asignan las prioridades a los trabajos a ser ejecutados y finalmente se balancea el sistema. Este proceso es reiterativo, ya que al cambiar las condiciones externas o internas del sistema, van apareciendo nuevos cuellos de botella, lo que trae consigo nuevas acciones correctivas.

En general, las solicitudes de trabajo provienen de tres grandes fuentes: resultados del mantenimiento preventivo y predictivo, requisiciones o necesidades de las áreas productivas o de servicios generales de planta y modificaciones dirigidas a atender las mejoras en los procesos que estén relacionadas con el incremento en la producción, la seguridad y la calidad o, la disminución de costos o riesgos de impacto ambiental.

Lo primero que se debe hacer es construir el mapa del proceso de la función mantenimiento y simular el flujo de elementos tangibles a través de él, luego, se deben identificar las diferentes estaciones de trabajo y seleccionar la más lenta de ellas; esta estación será identificada como “cuello de botella” y determinará la velocidad para procesar ordenes de trabajo del sistema completo, por lo que usted, debe asegurarse de mantenerla siempre ocupada y con una cantidad de órdenes de trabajo esperando a su entrada para ser procesadas. Seguidamente, se debe medir el tiempo promedio que tarda el cuello de botella para procesar las órdenes de trabajo que llegan a la estación. Si el tiempo de procesamiento del cuello de botella coincide con la demanda de trabajos a ser realizados, no hay de que preocuparse, pero si la demanda de servicio es superior a la capacidad del cuello de botella para procesarlo, se debe realizar un balance en el sistema asignando parte del trabajo a otras estaciones o buscar la manera de ampliar la capacidad del cuello de botella otorgándole más recursos (humanos, técnicos o económicos).

Como el cuello de botella es la estación de trabajo más lenta, es lógico pensar que a la entrada de la misma se formará una fila o “cola” de elementos tangibles para ser procesados, la cantidad de elementos presentes en esta fila es directamente proporcional a la velocidad de las estaciones anteriores al cuello de botella. Según la teoría de restricciones la cola de elementos tangibles de nuestro sistema no es más que inventario, el cual, debemos reducir pero en ningún momento eliminar, ya que dejaríamos nuestro cuello de botella sin trabajo que hacer y esta situación repercutiría de manera directa en nuestro nivel de servicio. Si nosotros conocemos la rata de procesamiento de elementos tangibles de una estación de trabajo, no tiene sentido colocar a la entrada de ésta más trabajo del que ella puede realizar porque se acumulará un número excesivo de órdenes de servicio en tránsito, las cuales, tienen costos asociados como material en almacén, tiempo de planificación, mano de obra, anticipos de contratos de servicio, etc. Es preferible tener la estación de trabajo aguas arriba paralizada y no procesando más inventario. Sin embargo, el “tiempo libre” de cada estación, puede utilizarse asignando sus recursos a otras estaciones para balancear el sistema o aumentar el flujo de adición de valor de nuestra función de mantenimiento.

¿Cual es el número de elementos en cola que debemos esperar a la entrada de una estación de trabajo?

Si conocemos el tiempo de servicio promedio real o ideal de la estación y la rata de llegada de elementos tangibles, podemos calcular el número de elementos en espera real o ideal para ser atendidos utilizando el modelo de cola simple (fórmula de Little) para estaciones de trabajo únicas y el modelo M/M/S para estaciones de trabajo con varios servidores. Es importante acotar que los elementos tangibles que llegan a la cola no necesariamente deben esperar que todos los elementos que se encuentran delante de él

sean procesados; recordemos que según la teoría de restricciones, después de identificar el cuello de botella, se deben establecer las prioridades para la ejecución del servicio, lo cual, en nuestro caso, se puede hacer aplicando los estudios de análisis de criticidad en los activos productivos, estos estudios le darán un índice de criticidad o prioridad a los elementos, ya que éstos están asociados de manera directa a intervenciones o servicios dirigidos a asegurar la continuidad operativa de los equipos.

El número calculado es un indicador que nos permitirá conocer si la estación de trabajo está funcionando correctamente, o si por alguna variación en la distribución estadística que caracteriza la llegada de elementos tangibles, debemos balancear el sistema completo. Un número de elementos en cola controlado alrededor de un valor establecido significa una buena gestión de la función mantenimiento en el estado planificado, ya que todas las estaciones están cumpliendo con su trabajo en el tiempo esperado y la demanda de servicios ha sido controlada.

Seguramente al principio de su gestión, usted tendrá que trabajar mucho para bajar el número de elementos en cola al valor calculado, pero al poco tiempo, el valor de la cola estará muy por debajo de lo que se esperaba, entonces; ¿Que sucede si el número de elementos en cola disminuye a un valor muy inferior al deseado?. Esta condición tiene dos interpretaciones, primero, indica que su estación está trabajando más eficientemente, de hecho, ahora es capaz de atender más demanda y por ende de manejar más flujo. En segundo término significa que algo anda mal con las estaciones aguas arriba, puede ser que aparezcan nuevos cuellos de botella, incluso, las fuentes generadoras de solicitudes pueden convertirse en uno de ellos. Por esta razón, usted debe estar pendiente de establecer estrategias que le permitan atender demandas no satisfechas de los clientes internos de forma anticipada. Es posible que para lograr más solicitudes de trabajo usted deba comenzar a aplicar técnicas de mantenimiento predictivo o incrementar las existentes, realizar inspecciones formales y rutinarias para detectar fallas o aumentar sus planes de mantenimiento preventivo. Parece contradictorio, pero usted no tendrá que aplicar estrategias como las mencionadas anteriormente porque tiene muchas ordenes que atender sino porque no tiene la cantidad suficiente de órdenes para que la capacidad de su sistema sea totalmente utilizada.

Recuerde siempre los objetivos de la teoría de restricciones, bajar inventarios, reducir costos y aumentar el nivel de servicio. Toda acción que se tome para asignar o redistribuir recursos (humanos, técnicos o económicos) debe estar orientada y alineada para lograr estos objetivos.

USO DE LA TEORÍA DE RESTRICCIONES Y MANEJO DE COLAS AL INICIO DE LA GESTIÓN DE LA FUNCIÓN MANTENIMIENTO (Segunda Parte: Balanceo y Sincronización del Sistema)

Una vez que entendemos y aceptamos que la función de mantenimiento no es más que un conjunto ordenado de actividades que agregan valor a un “servicio prestado”, desde una condición inicial conocida hasta una condición final que debe cumplir con los parámetros de calidad y seguridad establecidos o convenidos entre el ente encargado de agregar valor (Departamento de Mantenimiento) y el que lo recibe (clientes internos) en un tiempo determinado, y que, este conjunto de actividades puede ser modelado de acuerdo a la teoría de restricciones, nos topamos con la siguiente pregunta: ¿Que metodología debo aplicar para balancear y sincronizar el proceso de adición de valor de tal forma que funcione de manera óptima?

La respuesta a ésta pregunta se encuentra en el sistema *amortiguador – tambor - cuerda*. Este sistema fue desarrollado por E. Goldrat, el mismo promotor de la teoría de restricciones, y se centra básicamente, en desarrollar metodologías que permitan establecer vínculos entre las estaciones de trabajo que conforman el mapa de procesos de la función mantenimiento. Estos vínculos deben estar destinados a hacer de la cadena de adición de valor un sistema tan rígido como para evitar la acumulación excesiva de elementos tangibles (órdenes de trabajo) a la entrada de las estaciones de servicio, pero al mismo tiempo, lo suficientemente flexible como para que las estaciones no se queden sin trabajo para realizar. Por otro lado, debe existir una forma de garantizar que el orden de entrada de los elementos tangibles a las estaciones sea el adecuado para garantizar un buen nivel de servicio.

En la primera parte del desarrollo del tema se menciona el uso de modelos de cola para calcular las cantidades óptimas de elementos tangibles a la entrada de las estaciones de servicio. Estas cantidades, de hecho, son los amortiguadores de nuestro sistema. Para el cálculo del amortiguador en las estaciones simples (o de un solo servidor) utilizaremos la fórmula de Little, la cual se muestra a continuación:

$$Lq = Wq * \lambda$$

donde,

Lq es la cantidad de elementos en cola, esperando entrar a la estación para ser procesadas

Wq es el tiempo promedio de servicio para atender un elemento tangible

λ es la rata promedio con que llegan los elementos a la entrada de la estación de servicio.

Si la estación de servicio está conformada por varios servidores, podemos utilizar el modelo M/M/S, el cual se desarrolla a continuación:

En general, la función mantenimiento está regulada por dos distribuciones estadísticas, una que gobierna la llegada de ordenes de trabajo o elementos tangibles (Distribución de Poisson) y otra que es la cantidad con que cada servidor de la estación de trabajo es capaz de procesar un elemento tangible (Distribución exponencial).

El modelo M/M/s supone un patrón de llegada Poisson y una rata de servicio exponencial con s servidores. Para calcular el número de elementos tangibles en cola tenemos:

$$Lq = (1/s!) (\lambda / \mu)^s [\rho / (1 - \rho)^2] p_0$$

Donde

λ es la rata promedio con que llegan los elementos a la entrada de la estación de servicio.

μ es el tiempo promedio de procesamiento de un elemento tangible de una estación de trabajo.

$$\rho = \lambda / s\mu$$

$$p_0 = 1 / \{ \sum_{n=0}^{s-1} (1/n!) (\lambda/\mu)^n + (1/s!) (\lambda/\mu)^s [1/(1-\rho)] \}$$

$$Wq = Lq / \lambda$$

s es el número de servidores

Si usted no posee la información suficiente para hacer los cálculos mostrados anteriormente, puede hacer una aproximación que, en la práctica, funciona bastante bien. Mida durante un tiempo aproximado de dos meses la cantidad de elementos tangibles que procesan las estaciones de servicio que conforman su sistema de mantenimiento, calcule el promedio de elementos por estación y sume una desviación estándar, ese será el número ideal para comenzar, después, usted tendrá los datos suficientes para realizar un cálculo más formal y exacto.

Para rigidizar o colocar una “cuerda” a la cadena de adición de valor podemos utilizar el modelo “Kamban”, lo cual obliga a cada estación a no procesar más ordenes de las que fueron calculadas para la estación que se encuentra aguas abajo en el período de tiempo correspondiente al ciclo de procesamiento de los elementos tangibles, el cual se recomienda que comprenda 7 días (una semana). Esto permite a las estaciones de procesamiento más rápido invertir el tiempo libre en otras actividades que agreguen valor a la función mantenimiento o simplemente distribuir sus recursos entre las estaciones cuello de botella.

El tambor (ritmo de funcionamiento del sistema), además de ser caracterizado por la demanda, también está determinado por la criticidad de los equipos, el orden de entrada y el tipo de procesamiento que se le dará a cada elemento tangible estará regulado por la metodología de “Mantenimiento Centrado en Confiabilidad”, lo que permite garantizar el funcionamiento adecuado de los activos productivos en su contexto operativo en un tiempo determinado.

Para establecer el ritmo de servicio de su sistema usted debe seguir básicamente los siguientes pasos:

1. Escoger un área piloto para la implementación del mantenimiento centrado en confiabilidad tomando en cuenta criterios que le permitan hacer una matriz de decisión.
2. Realizar un inventario de equipos mecánicos y/o elementos de transmisión y control de parámetros operacionales del área piloto.
3. Identificar la función de cada equipo y/o elemento.
4. Identificar el modo y el tipo de falla para cada uno de los componentes que conforman los equipos y/o elementos.
5. Estudiar los efectos, las consecuencias y las alteraciones en el contexto operacional del equipo y/o elemento debido a la ocurrencia del evento falla.
6. Determinar del índice de criticidad de los equipos y/o elementos mencionados en el punto 1

7. Determinar el tipo de mantenimiento para cada caso dependiendo del tipo de falla y la criticidad.
8. Elaborar las rutas y planes de mantenimiento
9. Calcular el inventario de repuestos asociados al los planes y rutas.
10. Crear códigos de falla
11. Hacer diagramas de Pareto de fallas crónicas y utilizar la metodología de análisis causa-raíz.

Estos 11 pasos le permitirán establecer un ritmo de prestación de servicio para los equipos que conforman el área piloto, los órdenes de trabajo correspondientes a las demás áreas deben ser atendidas por orden de llegada y según la gravedad del caso hasta que cada una de las áreas vaya siendo integrada a la metodología del mantenimiento confiable.

Para medir los avances y el desempeño de su sistema, se presentan a continuación los indicadores de gestión más adecuados para una función basada en teoría de restricciones y modelos de cola.

Parámetros

- La unidad de tiempo en la gestión de mantenimiento es la semana.
- **Órdenes de trabajo en cola (ODTC):** Ordenes de trabajo que esperan por entrar al proceso que define la función mantenimiento.
- **Órdenes de trabajo planificadas (ODTP):** Ordenes de trabajo que se realizarán la semana siguiente a la semana en la cual fue asignada la prioridad, los materiales y los estándares de ejecución.
- **Órdenes de trabajo no planificada (ODTNP):** Ordenes de trabajo que no cumplen con al menos un requisito para poder ser considerada como ODT planificada.
- **Orden de trabajo ejecutada (ODTE):** Orden de trabajo que se realiza cumpliendo con los estándares de ejecución y es aceptada como satisfactoria por el cliente.
- **Tiempo de servicio (TS):** Es en tiempo promedio (en semanas) que tarda una solicitud de trabajo desde que llega a la cola hasta que es ejecutada.

Indicadores de Gestión.

- ODTC
- ODTP ejecutadas vs. ODTE : Mide el porcentaje de ordenes de trabajo planificadas del total ejecutadas - *prioridad que se le da a la planificación* -
- ODTP ejecutadas vs. Total de ODTP: Mide el porcentaje de ordenes de trabajo planificadas ejecutadas del total de ordenes de trabajo planificadas – *eficiencia de la planificación* -
- ODTE vs. ODTN: Mide la relación entre las ordenes de trabajo ejecutadas y las que están esperando por ser atendidas – *capacidad del sistema* -
- TS

Estos cinco indicadores son más que suficientes para evaluar el desempeño de su sistema, ya que lo describen de manera completa, más aún, si usted debe calcular y reportar otros indicadores por políticas de la empresa, se dará cuenta que la relación entre éstos y los recomendados en el artículo es directamente proporcional.

La aplicación de la teoría de restricciones en la gestión del mantenimiento industrial es algo nuevo y poco desarrollado en la práctica, es por ésto, que su aplicación rompe con muchos paradigmas hasta ahora tomados como ciertos, entre ellos, la necesidad de una programación estricta, la recopilación de grandes historiales de fallas y extensas inversiones de recurso humano en mantenimiento predictivo (horas hombre). Si usted toma la decisión de adoptar el modelo propuesto para el desarrollo de la función mantenimiento, es muy probable que obtenga excelentes resultados en muy poco tiempo, sin embargo, tenga en cuenta que lo más importante es creer en la teoría y estimular a su equipo a seguir las pautas. Como se refleja en el libro “La Meta” de E. Goldrat, el diseñador del sistema debe ser usted mismo, atendiendo las necesidades particulares de su planta. Existen muchos detalles que no se muestran en ninguna de las dos partes del artículo “Uso de la Teoría de Restricciones y Manejo de colas al Inicio de la Gestión de la Función Mantenimiento”, pero seguramente aparecerán conforme se vaya implementando el sistema, recuerde siempre que lo más importante, es tomar decisiones que bajen los inventarios de elementos tangibles por procesar y en tránsito, reduzcan los costos de mantenimiento e incrementen el nivel de servicio.

Elaborado por
Rafael Suárez