**Caso 4:** Infraestructura Computacional

# Entendimiento del Problema

Identificación de carga de trabajo

El primer paso para el entendimiento del problema es identificar el tipo de transacciones que va a realizar el sistema, para esto realizamos un análisis de los stakeholders y los factores que consiste en determinar quién lo hace, qué hace y cómo lo hace. A continuación, se presenta una tabla con la información de los tipos de transacciones realizadas por la organización.

|  |  |  |
| --- | --- | --- |
| **Quién lo hace** | **Qué hace** | **Cómo lo hace** |
| **Entidad Externa** | Reporta periódica y electrónicamente la información que corresponde a las afiliaciones que se alojan en la aplicación de afiliación. | Una entidad externa subcontratada maneja la aplicación y la información de las afiliaciones, esta procede a enviar dicha información para ser procesada en forma batch y posteriormente se ingresa en la base de datos de la empresa. |
| **Operadores de recaudo** | Reportar diaria y periódicamente la información relacionada con los recaudos de dinero de la organización. | Son entidades encargadas del manejo y procesamiento de la recaudación, requiere hacer conciliaciones tripartías entre diferentes procesos en la historia laboral. Implica realizar verificaciones en las cantidades recaudadas y realizan el cobro de aportes no realizados por las empresas. |
| **Afiliados** | Puede realizar diferentes labores, puede usar los servicios de la aplicación de historia laboral para que puedan consultar toda su historia laboral, también puede consultar el estado de un proceso de tutela o de un derecho de petición que realiza el afiliado. | Los afiliados entran al portal web de historia laboral para realizar la consulta de su información laboral y a través de la aplicación de manejo de tutelas, pueden consultar o crear diferentes tutelas o derechos de petición. Este procedimiento lo pueden realizar en cualquier lugar del país. |
| **Empleados** | Los empleados cumplen diferentes labores dependiendo de la aplicación que estén utilizando. En el caso de la aplicación de Historia laboral, consultan la historia laboral de los clientes para realizar las actualizaciones y los cambios en la misma. Por otro lado, en la aplicación de tutelas tienen que revisar las diferentes tutelas y derechos de petición. | Los empleados reciben la petición de actualización o para mostrar el historial laboral de la persona mediante la aplicación o los centros de atención, los empleados proceden a solicitar dicha información al sistema, esta solicitud tiene que realizarse en un tiempo de respuesta mínimo para la satisfacción del cliente. Por otro lado, la aplicación de tutelas es sumamente importante, ya que los empleados tienen que recibir los diferentes derechos de petición y tutelas, toca revisarlos de acuerdo con las normativas legales y otorgar una respuesta dentro de los plazos establecidos. |

Requerimientos de cada aplicación

1. **Afiliación**
   1. **Disponibilidad:** Según los requerimientos del negocio, este servicio debe estar disponible los siete días de la semana entre 7 a.m y 10 p.m. En este caso el total de horas en el que el servicio estaría disponible serían 3285 horas.

* 1. **Capacidad:** a partir de los requerimientos del negocio se tiene que la carga que debe soportar en promedio, en términos de productividad, es de 100 transacciones por día.
  2. **Escalabilidad:** Teniendo en cuenta la capacidad que debe garantizar la aplicación y a partir del crecimiento calculado históricamente de la cantidad de afiliados que maneja la empresa, se tiene que con un crecimiento de 5,73% la carga que debe soportar en términos de productividad es de:

De este modo, la carga que debe soportar hasta una carga de 175 transacciones al día, proyectando el negocio hasta a 10 años.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Año** | **2019** | **2020** | **2021** | **2022** | **2023** | **2024** | **2025** | **2026** | **2027** | **2028** | **2029** |
| **Trs/día** | 100 | 105,73 | 111,78 | 118 | 124,96 | 132,12 | 139,69 | 147,7 | 156 | 165 | 175 |

Igualmente, se debe tener en cuenta un crecimiento en horas pico donde la proporción de carga puede ser hasta 2 veces la usual manejada por la aplicación, de este modo:

Es decir, que en el caso de que se encuentre en un horario pico, en el cual se realicen el doble de transacciones a lo estandar, se requerirá una capacidad hasta de 200 transacciones por día.

* 1. **Desempeño:** Debido a que el proceso se encuentra subcontratado con una entidad externa, la cual realiza el procesamiento de las peticiones de forma batch y realiza reportes periódicamente a la empresa, no existe un requerimiento en cuanto al tiempo de respuesta que se debe garantizar para la aplicación. No obstante, se espera que el usuario que se intenta afiliar a la compañía obtenga una respuesta de comprobación en menos de 1 segundo dada la prioridad del servicio que se ofrece (el cual es prioridad para los objetivos del negocio).

1. **Recaudo**
   1. **Disponibilidad:** Según los requerimientos del negocio, este servicio debe estar disponible los siete días de la semana entre 7 a.m y 5 p.m. En este caso el total de horas en el que el servicio estaría disponible serían 5110 horas.

Por otra parte, este sería el mínimo requerido que debe garantizar la aplicación. No obstante, entre los objetivos de la aplicación se encuentra el recaudo de aportes no realizados por aportantes, lo cual en términos del negocio son de importancia significativa dado a que contribuye al funcionamiento operacional. Igualmente, se debe tener en cuenta que la aplicación realiza los procesos en modo batch y dependen de entidades externas para llevar a cabo las tareas asignadas. De este modo, la disponibilidad de esta aplicación debe cumplir con un mínimo de 24 horas, 7 días a la semana, es decir, que se debe garantizar una disponibilidad de 99,9%.

* 1. **Capacidad:** A partir de la carga que soporta la aplicación en un horario pico, es decir 30.000 transacciones por día, asumiendo que en este caso se procesa hasta 2 veces la carga usual, se tendría una capacidad en términos de productividad de:

Es decir, que la aplicación debe soportar una carga de 15000 transacciones al día.

* 1. **Escalabilidad:** A partir de la información suministrada, la aplicación debe funcionar normalmente hasta con una carga de 30.000 transacciones al día en términos de productividad.
  2. **Desempeño:** Dado lo establecido por requerimientos del negocio, se tiene que la respuesta del servicio ofrecido por la aplicación debe ser menor o igual a 4 segundos. Igualmente, se debe tener en cuenta el alto nivel de procesamiento que debe manejar esta aplicación; si bien esta se encarga de realizar cobro de aportes no realizados, también debe realizar diariamente conciliaciones tripartitas en las que se debe verificar el monto recaudado y realizar reportes a partir de la información obtenida de las inconsistencias entre las transacciones recibidas y el log bancario, el log bancario y el saldo del día de la cuenta de recaudo de cada entidad financiera y entre el log bancario y los recaudos imputados en la historia laboral. Debido a esto, dicha aplicación requerirá un alto rendimiento para realizar la respuesta requerida para cumplir con el requerimiento de tiempo de respuesta.

1. **AFE**
2. **Disponibilidad:** Según los requerimientos del negocio, este servicio debe estar disponible los siete días de la semana entre 7 a.m y 10 p.m. En este caso el total de horas en el que el servicio estaría disponible serían 3285 horas.

No obstante, esta aplicación se encarga de la comunicación entre todas las aplicaciones que conforman el sistema y realizar la validación de datos entre estas por lo que es de suma importancia garantizar una disponibilidad de 99,9%. Si bien, esto teniendo en cuenta la prioridad de la disponibilidad del servicio para el funcionamiento de la totalidad del sistema, por lo que como mínimo debe garantizar la recepción de tareas 24 horas, 7 días a la semana con una indisponibilidad de hasta de 8 horas y 5 minutos en el año.

1. **Desempeño:** El tiempo de respuesta de la aplicación debe ser menor o igual a 4 segundos, esto de acuerdo con los requerimientos del negocio. No obstante, dada la arquitectura “stored procedures” de la aplicación, los servidores de las bases de datos se encuentran en sobrecarga por lo que se da un nivel de saturación en el que el rendimiento no cumple con el requerimiento del negocio; esto debido a que no se da una distribución de la carga computacional requerida y, como consecuencia, el procesamiento de las peticiones y la respuesta son ineficientes para el negocio.

1. **Capacidad:** La aplicación maneja un promedio de 5000 transacciones por día, de este modo como requerimiento esta debe garantizar el funcionamiento normal del sistema para finalizar hasta 5000 tareas por día.

1. **Escalabilidad:** Teniendo en cuenta la capacidad que debe garantizar la aplicación y a partir del crecimiento calculado históricamente de la cantidad de afiliados que maneja la empresa, se tiene que, con un crecimiento de 5,73% (proporción de crecimiento anual) la carga que debe soportar en términos de productividad es de:

De este modo, la carga que debe soportar hasta una carga de 8729 transacciones al día, proyectando el negocio hasta a 10 años.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Año** | **2019** | **2020** | **2021** | **2022** | **2023** | **2024** | **2025** | **2026** | **2027** | **2028** | **2029** |
| **Trs/día** | 5000 | 5286 | 5589 | 5910 | 6248 | 6606 | 6985 | 7385 | 7808 | 8256 | 8729 |

Igualmente, se debe tener en cuenta un crecimiento en horas pico donde la proporción de carga puede ser hasta 2 veces la usual manejada por la aplicación, de este modo:

Es decir, que la carga que debe soportar la aplicación en términos de productividad debe ser hasta de 10000 transacciones por día para soportar los requerimientos del negocio en horarios pico.

1. **Historia laboral y Portal web**
   1. **Disponibilidad:** La disponibilidad es uno de los requerimientos más importantes para la aplicación de historia laboral, debido a que los empleados tienen que consultar la historia laboral desde las sucursales para realizar sus actividades entre las 7:00am y las 5:00pm y adicionalmente los afiliados deben tener una disponibilidad entre 7:00 Am a 10:00pm para consultar en la página web. A raíz de la información proporcionada por Colpensiones nos dimos en la labor de calcular el porcentaje de disponibilidad y la cantidad de días que puede estar indisponible el sistema. La información de la historia laboral se encuentra en una misma base de datos, tanto para la consulta por parte de los empleados y por el portal web, por lo que calculamos la disponibilidad de la plataforma basado en la página web, donde se necesita una mayor disponibilidad del sistema. Al tener 15 horas de disponibilidad diaria, se calcula la disponibilidad anual en segundos, la cual es 5475 horas al año. Sabemos que en un año hay 8760 horas por lo que procedemos a calcular la disponibilidad:

Tiempo de indisponibilidad: 3285 horas.

* 1. **Escalabilidad**: La escalabilidad es otro tema sumamente importante para la organización, ya que como nos menciona la organización de Colpensiones, el promedio en procesamiento en horas pico es de 72% llegando a picos de hasta el 92% dependiendo de los días y la ubicación geográfica de los centros de atención, donde en algunos puntos centrales, hay una afluencia mayor de afiliados. Adicionalmente realizando diferentes supuestos socioculturales, la cantidad de afiliados crece con respecto al crecimiento poblacional y laboral, por lo que se espera que, en los siguientes años, haya mayor afluencia de afiliados y por ende de transacciones.
  2. **Capacidad**: Otro factor de gran importancia es la capacidad del sistema, ya que tiene que soportar 6.500 transacciones al día en las cuales se incluye el portal web y la historia laboral, las cuales tienen una intensidad alta de entrada y salida, tanto en lectura como en escritura, ya que los empleados registran diferentes cambios a los afiliados y los afiliados consultan mediante la plataforma web o mediante las sucursales su historial laboral. Por otro lado, se ha determinado que la infraestructura actual no soporta una carga grande de transacciones, debido a la infraestructura de stored procedures la cual genera una sobrecarga en los servidores y eventualmente un cuello de botella en los mismos lo cual incumple el requerimiento de desempeño eventualmente.

al día.

* 1. **Desempeño**: Un aspecto importante no solo para la aplicación de historial, si no para la organización, es el desempeño de sus aplicaciones. El portal web y los servicios solicitados de historial por parte de los empleados tienen que asegurar un tiempo de respuesta menor o igual a 4 segundos y es de suma importancia mejorar la infraestructura actual, ya que como se menciono anteriormente, los cuellos de botella por parte de los servidores generan que no se cumpla este requerimiento. Es de suma importancia que esta aplicación tenga un tiempo de respuesta mínimo, como nos lo comunica Colpensiones.

1. **Nómina de pensionados**
   1. **Disponibilidad**: Uno de los procesos mas importantes para Colpensiones y que contiene mayor riesgo, es el procesamiento de las nominas de los pensionados. Un fallo en este proceso puede ser fatal para todos los afiliados y puede llegar a tener unas consecuencias tanto económicas como legales importantes para la organización. Este proceso dura 10 días y se procesa automáticamente sin necesidad de un agente externo, es decir es un proceso automático mensual. Por lo que necesitamos que tenga una disponibilidad de 120 días al año para que mensualmente pueda procesar la nomina de los pensionados. El porcentaje de disponibilidad que requiere el sistema es igual a:

.

Sin embargo, tenemos que ser muy cautelosos con este proceso, ya que su información es lo mas importante en la organización y una perdida en la misma seria extremadamente costosa, por lo que se decide aumentar la disponibilidad y tener una tolerancia a fallos del alta, por lo que determinamos que su disponibilidad sea del 99.9%, es decir una indisponibilidad de 8 horas al año.

* 1. **Escalabilidad:** La aplicación de la nomina de los empleados solo se ejecuta durante 10 días en modo batch, es decir es automatizado, por lo que la carga de información y de transacciones varia mucho durante ese tiempo, donde llega hasta unos picos de 900.000 transacciones en días pico por lo que es de suma importancia que la infraestructura de esta aplicación soporte esos picos.
  2. **Capacidad:** Este aspecto está ligado con la escalabilidad de la aplicación, ya que esta infraestructura tiene que soportar una gran carga de datos en 10 días, mayormente, la cantidad de afiliados que son 1.900.000 en 10 días que en promedio serian 190.000 afiliados. Esa es la carga que tiene que soportar el sistema, y si la comparamos con las demás aplicaciones, es significativamente mayor. Y mas teniendo en cuenta los estancamientos o cuellos de botella que actualmente los servidores están teniendo.
  3. **Desempeño**: El tiempo de respuesta de la aplicación debe ser menor o igual a 4 segundos, esto de acuerdo con los requerimientos del negocio. No obstante, dada la arquitectura “stored procedures” de la aplicación, los servidores de las bases de datos se encuentran en sobrecarga por lo que se da un nivel de saturación en el que el rendimiento no cumple con el requerimiento del negocio; esto debido a que no se da una distribución de la carga computacional requerida y, como consecuencia, el procesamiento de las peticiones y la respuesta son ineficientes para el negocio.

1. **Tutelas** 
   1. **Disponibilidad**: La aplicación de tutelas es sumamente importante ya que un mal manejo de la información traería problemas legales a la organización, es por eso que no se puede admitir ningún tipo de fallos del sistema, es por eso que hemos determinado una disponibilidad del 99.9% equivalente a 8 horas y 45 minutos de indisponibilidad.
   2. **Capacidad**: Este requerimiento es sumamente importante para la organización, ya que tiene que soportar toda la carga que le llega que son mas o menos 80.000 procesos de tutelas y derechos de petición que realizan los afiliados por la plataforma y soportar en promedio unas 3500 transacciones al día. Puede que no necesite mucha capacidad en cuanto a entrada y salida, pero si para soportar la carga y la cantidad de procesos que recibe. Teniendo en cuenta el servidor de apoyo a la base de datos actual, el cual se encuentra a su máxima capacidad con un 72% de uso de CPU con picos mayores al 90%, se necesita aumentar la capacidad de el servidor para soportar las transacciones. Por otro lado, hay que tener en cuenta que esta la entrada de datos a esta aplicación se debe hacer desde cualquier parte del país, por lo que la infraestructura debe soportar la carga desde diferentes partes de Colombia.
   3. **Desempeño**: El requerimiento de desempeño es sumamente importante para la organización y en especial para la aplicación de tutelas, ya que los clientes tienen que recibir su respectiva respuesta en un tiempo estrictamente establecido, de lo contrario incurriría a algunas faltas legales lo que podría causar problemáticas en la organización. Es muy difícil cuantificar el tiempo de desempeño, debido a que los diferentes derechos de petición y tutelas tienen diferentes plazos de respuesta que varían entre un día y diez días por lo que es variable su tiempo de respuesta.

# Selección de la infraestructura

Identificación de las problemáticas más relevantes de cada aplicación

1. **Afiliación**

Basándonos en el análisis realizado previamente, se tiene que los requerimientos más relevantes son el desempeño, así como la escalabilidad, así como la disponibilidad. No obstante, los procesos que maneja la aplicación son realizados por terceros de modo que se deben realizar las exigencias debidas para que se dispongan de los debidos planes de contingencia y continuidad documentados debidamente. Si bien, tanto el desempeño como la disponibilidad son de suma importancia debido a que estos garantizan un buen servicio a los usuarios potenciales e igualmente se debe tener en cuenta que esta es una de las aplicaciones misionales al contribuir con los objetivos del negocio. Por otra parte, se debe tener en cuenta que se debe poder expandir la capacidad del servicio de acuerdo al crecimiento de la empresa.

1. **Recaudo**

A partir del análisis anterior se identificaron tres requerimientos que se deben garantizar: alta disponibilidad, un buen tiempo de respuesta e igualmente un alto procesamiento. Es indispensable además el buen funcionamiento de esta aplicación teniendo en cuenta que de esta depende brindar un buen servicio a los afiliados. En cuanto a la selección de los requerimientos que son prioritarios, la disponibilidad se toma como uno de estos debido a que esta aplicación maneja procesos con entidades financieras e igualmente, se manejan logs bancarios y cuentas financieras de modo que es de vital importancia que no halla fallos. Por otra parte, para llevar a cabo las tareas de los recaudos se necesita un alto poder de procesamiento en orden para cumplir con el requerimiento de negocio de tiempo de respuesta. Es por esto que se debe tomar en cuenta las medidas adecuadas para la infraestructura de manera que el rendimiento sea óptimo.

1. **AFE**

La aplicación de AFE es una de las más importantes dado a que realiza la verificación de datos entre todas las aplicaciones e igualmente lleva a cabo la comunicación entre estas. Es por ello que es de suma importancia garantizar la disponibilidad de esta aplicación sobre todos los requerimientos, puesto a que una falla podría generar fallas en los procesos llevados acabo por las demás aplicaciones.

1. **Historia Laboral y Portal Web**

Se han identificado tres problemáticas sumamente relevantes que podrían afectar a la organización, estas son la capacidad y el desempeño de la aplicación. La problemática más relevante es el desempeño de la aplicación, es de suma importancia que los empleados tengan un tiempo de respuesta mínimo para lograr atender correctamente a los afiliados en las oficinas de servicio público y, por otro lado, los afiliados deben tener respuesta por parte del servidor en el portal web de la forma más rápida posible. Se ha identificado una problemática en la base de datos, la cual genera colas de espera una saturación en el servidor, lo que puede afectar el desempeño de las transacciones incumpliendo el requerimiento de desempeño que consiste en un tiempo de respuesta menor a 4 milisegundos.

Por otro lado, pueden presentarse problemas de indisponibilidad causados por la infraestructura que actualmente está prestando el servicio. Esta infraestructura cuenta con más de 10 años de uso, por lo que son más propensas a fallos y cuentan con una disponibilidad sumamente baja. Una problemática de disponibilidad es sumamente grave, ya que afectaría directamente las transacciones relacionadas con el historial laboral de los afiliados bloqueando así las oficinas de atención y el portal web.

Por último, hemos encontrado una problemática en cuanto a la capacidad. El servidor manejador de las bases de datos está en su límite de capacidad y de uso de CPU, que en promedio en las mediciones de la carga de procesamiento del servidor el uso de CPU está en 72% con algunos núcleos llegando al 92% de su uso, causando retrasos en los procesos de recaudo y en la historia laboral un bajo desempeño.

1. **Nómina de pensionados**

Se han identificado tres problemáticas relevantes que podrían afectar la organización, estas son la disponibilidad, la escalabilidad y la capacidad. La nomina de pensionados es una aplicación sumamente importante para la organización, ya que maneja la información y los balances pertenecientes a las operaciones principales de Colpensiones. Teniendo en cuenta la información otorgada por Colpensiones, la indisponibilidad del sistema es la problemática principal, al ser una aplicación en “*batch*”, automatizada, toca asegurar que el sistema esté disponible durante el lapso de 10 días que se realizan los procesos por lo que es sumamente importante que este sistema tenga un plan de contingencia y una tolerancia a fallos alto, que minimiza el impacto de la indisponibilidad. Otro aspecto sumamente importante es la escalabilidad de la aplicación, esto es por que la aplicación es susceptible a incrementos de carga significativos y es sumamente importante que el sistema logre soportar dichos aumentos de carga sin afectar su operación. Ya que un fallo en alguna transacción o un cuello de botella en un proceso automatizado suele ser fatal generando una perdida de información y saturaciones en el sistema. Por ultimo se pueden presentar problemas de capacidad, donde el sistema no esta soportando en algunas ocasiones la carga generada por el BPM y la ejecución de los procesos por el “batch”, las cuales generan una saturación del sistema donde la CPU alcanza un uso mayor al 90%, lo que provoca tiempos de respuesta mucho mayores.

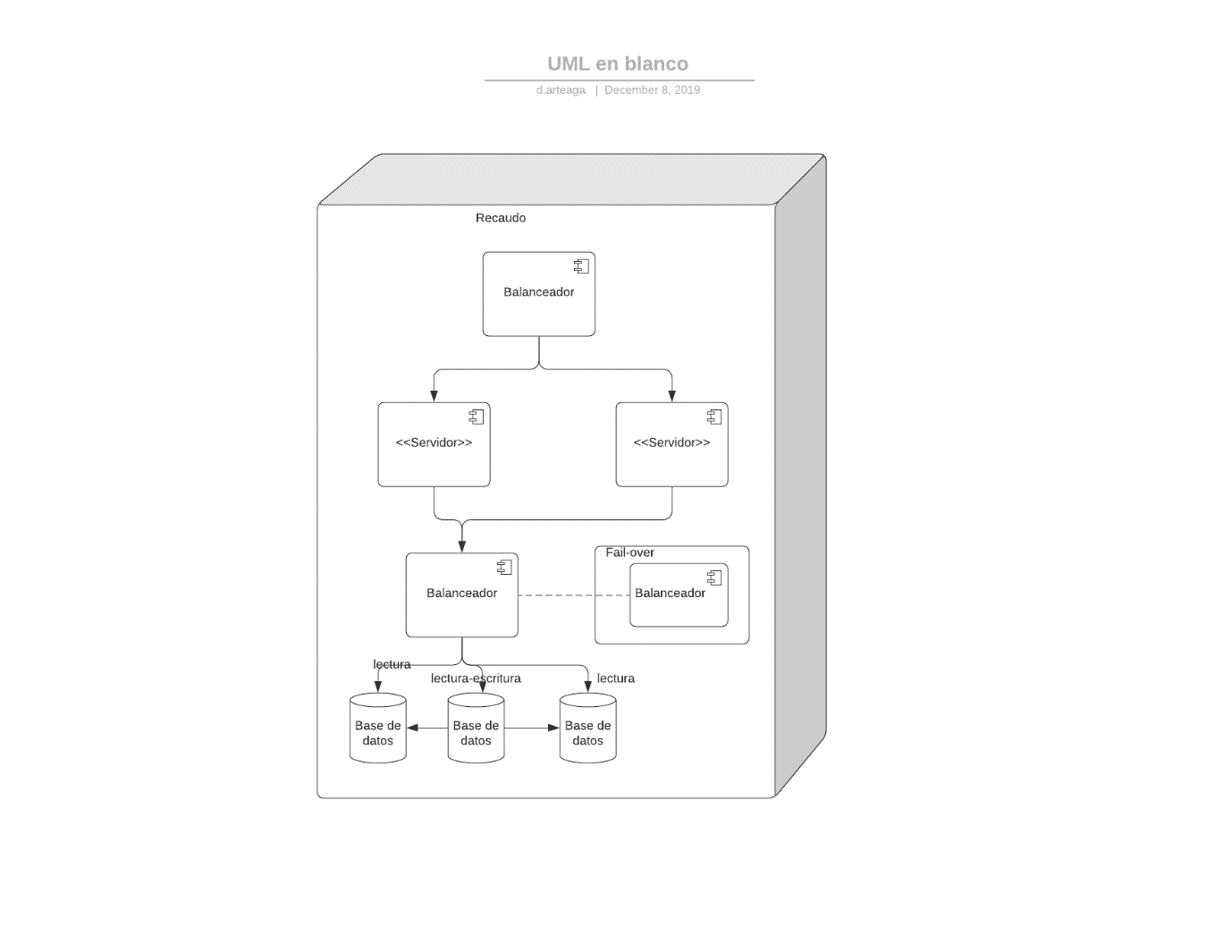
1. **Tutelas**

La aplicación de tutelas es una de las más críticas para Colpensiones, ya que está sujeta a ámbitos legales y una mala manipulación de la información conllevaría a problemas legales para los empleados, es por eso que determinamos que la disponibilidad, capacidad y eficiencia de la aplicación son los requerimientos mas importantes y más susceptibles a fallos. La disponibilidad del sistema es sumamente importante, y por la importancia de la información manipulada en la aplicación se necesita que el sistema este todo el tiempo disponible para manipular correctamente la información, y evitar perdidas o daños a los datos. Por otro lado, es sumamente importante que el sistema tenga la capacidad adecuada, evitar cuellos de botella y demoras por cargas mayores, ya que esto causaría una demora en el sistema y posibles fallos en la manipulación de la información, se tiene que asegurar los 80.000 procesos. Por último, el desempeño de la aplicación se puede ver afectado por una saturación del sistema, toca asegurar que los procesos sean resueltos en los plazos establecidos, si no, traería problemas legales a Colpensiones.

Infraestructura utilizada para cada aplicación

1. **Recaudo**

Se decidió utilizar un cluster de balanceo de carga, de este modo se garantiza una alta disponibilidad y la capacidad necesaria para satisfacer los requerimientos de alto procesamiento; este se implementará por medio de proxis inversos con una selección de servidor de tipo balanceo adaptivo. Igualmente se decidió utilizar un failover de tipo N+1 para garantizar un buen desempeño y establecer una buena respuesta ante fallas. Igualmente se utilizó un cluster de bases de datos con un patrón maestro-esclavo para optimizar la carga y garantizar disponibilidad.



1. **Nómina de pensionados y Afiliación**

Para el diseño de la infraestructura de las aplicaciones de nómina de pensionados y afiliación, tenemos que garantizar los requerimientos establecidos en la sección 1.2 que son, disponibilidad, capacidad, escalabilidad y desempeño. Encontrar un clúster adecuado que cumpla perfectamente los requerimientos es una labor ardua por lo que dividimos en importancia los requerimientos y determinamos las problemáticas. Lo mas importante es la Disponibilidad y la capacidad, ya que la información que se maneja es de suma importancia y sensible por lo que fallos durante el proceso serian inaceptables, adicionalmente la escalabilidad es un tema de suma importancia, ya que al ser una aplicación “*Batch”* es automatizada por lo que no hay patrones de carga, es decir, la carga es variable entonces el sistema necesita escalar hasta ciertos picos de carga. El clúster que decidimos fue un clúster de aplicaciones multi maestro, ya que es de suma importancia la lectura de las nóminas de los pensionados para realizarles la consignación. En este caso, vamos a tener un servidor de aplicaciones el cual le solicita a un balanceador de cargas y de tareas que le solicite al servidor de base de datos la información y la procese y almacene. Adicionalmente se va a tener un failover para los servidores de aplicaciones y para el balanceador. Ya que es de suma importancia garantizar la disponibilidad de los sistemas en los horarios establecidos. Para garantizar el desempeño usamos los balanceadores los cuales distribuyen la carga a los servidores dependiendo de la carga de los estos.

Por otro lado, tenemos que asegurar la disponibilidad y la capacidad de almacenamiento, es por eso que para el clúster se decidió utilizar el sistema de RAID 0-1, el cual consiste en un espejo de stripes, esto me garantiza la mayor disponibilidad del sistema con un desempeño inigualable, ya que escribe el doble de rápido en dos discos con stripe.

1. **AFE, Tutelas, Historia Laboral y portal web**

Para el diseño de la infraestructura de las aplicaciones, tendremos en cuenta los requerimientos descritos en la sección 1.2. Uno de los aspectos fundamentales al determinar la infraestructura que se va a implementar fue separar las aplicaciones que funcionan tipo “Batch” y tipo “on line” ya que el procesamiento de la información y los requerimientos de dichas aplicaciones son diferentes. En este caso agrupamos las aplicaciones AFE, Tutelas, Historia Laboral y el portal web en un solo clúster, esto debido a que las transacciones están encadenadas y pueden ser dirigidas por un BPM el cual es el AFE. El AFE encadenará las aplicaciones de Tutelas, Historia Laboral y el portal web en un solo clúster que vamos a explicar a continuación.

En Las cuatro aplicaciones cuentan en conjunto con los 4 requerimientos, de disponibilidad, escalabilidad, desempeño y carga. Nos dimos en la labor de optimizar los recursos monetarios, ya que no era viable financieramente realizar un clúster por aplicación si realmente se puede realizar un clúster que unifique todas las demás aplicaciones. El clúster consiste en un balanceador de carga mezclado con un clúster shared nothing. Que significa este balanceados de carga en shared nothing, significa que las transacciones realizadas tanto por los clientes como por los usuarios las capta la aplicación AFE, la cual es encargada de encadenar las aplicaciones anteriormente mencionadas, para procesar la información y otorgar la respuesta indicada al solicitante. El servidor de AFE se encarga de realizar las peticiones a los diferentes servidores de aplicaciones, los cuales individualmente tienen su propio sistema de discos. Unos aspectos sumamente importantes que toca mencionar es que el sistema shared nothing nos garantiza que se cumplan los diferentes requerimientos, pero con unas condiciones dadas.

En este tipo de sistemas el desempeño se ve sumamente mejorado dependiendo del shipping y la distribución de los datos, al nosotros distribuir la información uniformemente en la aplicación correspondiente, nos garantiza un desempeño superior a diferentes alternativas. Hemos encontrado en este sistema dos inconvenientes, los cuales consisten en la disponibilidad del sistema y la carga. La disponibilidad, ya que, si contamos no mas con un servidor por aplicación, un eventual fallo afectaría de manera importante una sección del clúster, el cual sigue funcionando, pero un servicio no estaría disponible.

En cuanto la carga, en los clústeres shared nothing, dependen de los patrones de carga acceso por lo que en algunos casos se vería afectada. La solución para ambos requerimientos esta en la implementación de un servidor adicional como si fuera multi maestro, es decir, ahora hay dos servidores conectados a la base de datos, esto me garantiza tanto disponibilidad como la carga. Por último, al ser un clúster shared nothing, con distribución uniforme en los servidores, la escalabilidad es la ideal, ya que cada aplicación contiene sus volúmenes de discos, es sencillo escalar por aplicación.

Por otro lado, es de suma importancia la configuración de almacenamiento por aplicación, ya que se necesitan cumplir diferentes requerimientos. A continuación, enumeramos las diferentes aplicaciones con su respectiva configuración de almacenamiento:

1. Tutelas: La configuración ideal para la aplicación de tutelas es RAID 6, la cual me garantiza la mayor disponibilidad y tolerancia a fallos de las configuraciones. Es de suma importancia la información en la aplicación de tutelas, no se puede permitir una perdida de información por algún fallo en los discos ya que causaría unos problemas legales para la organización. Al ser un Stripping con doble paridad distribuida un fallo en un disco puede ser recuperado con la paridad calculada en los demás, adicionalmente ofrece un buen rendimiento en lectura, pero se degrada un poco en escritura.
2. Historial laboral y portal web: Unificamos estas dos secciones en un mismo disco, ya que maneja la misma información. El servidor web maneja la información de la historia laboral únicamente, como menciona Colpensiones, por lo que es viable utilizar un disco unificado para ambas aplicaciones. En este caso es de suma importancia la disponibilidad y el desempeño de las plataformas, por lo que la configuración ideal seria RAID 0-1, esta consiste en un espejo de stripes, lo que mejora significativamente la lectura, el cual es el requerimiento mas importante de estas aplicaciones, las cuales tienen que cumplir un tiempo menor a 4 segundos y adicionalmente mejora la disponibilidad de la plataforma por si un disco se llega a dañar, hay una copia exacta que puede reemplazar su función mientras se recupera el disco dañado.

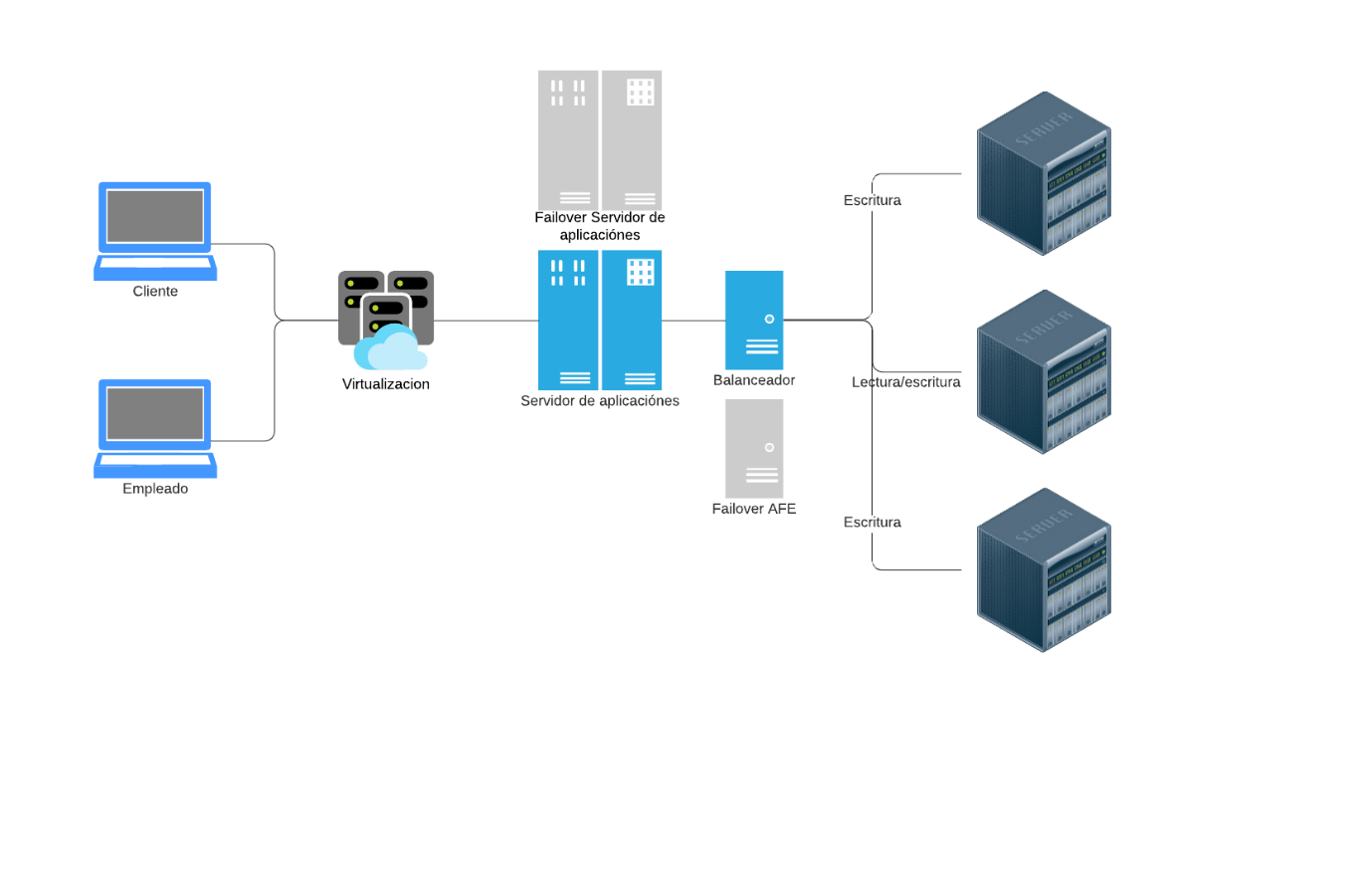
Diagramas de la infraestructura

Imagen que contiene captura de pantalla

Descripción generada automáticamente

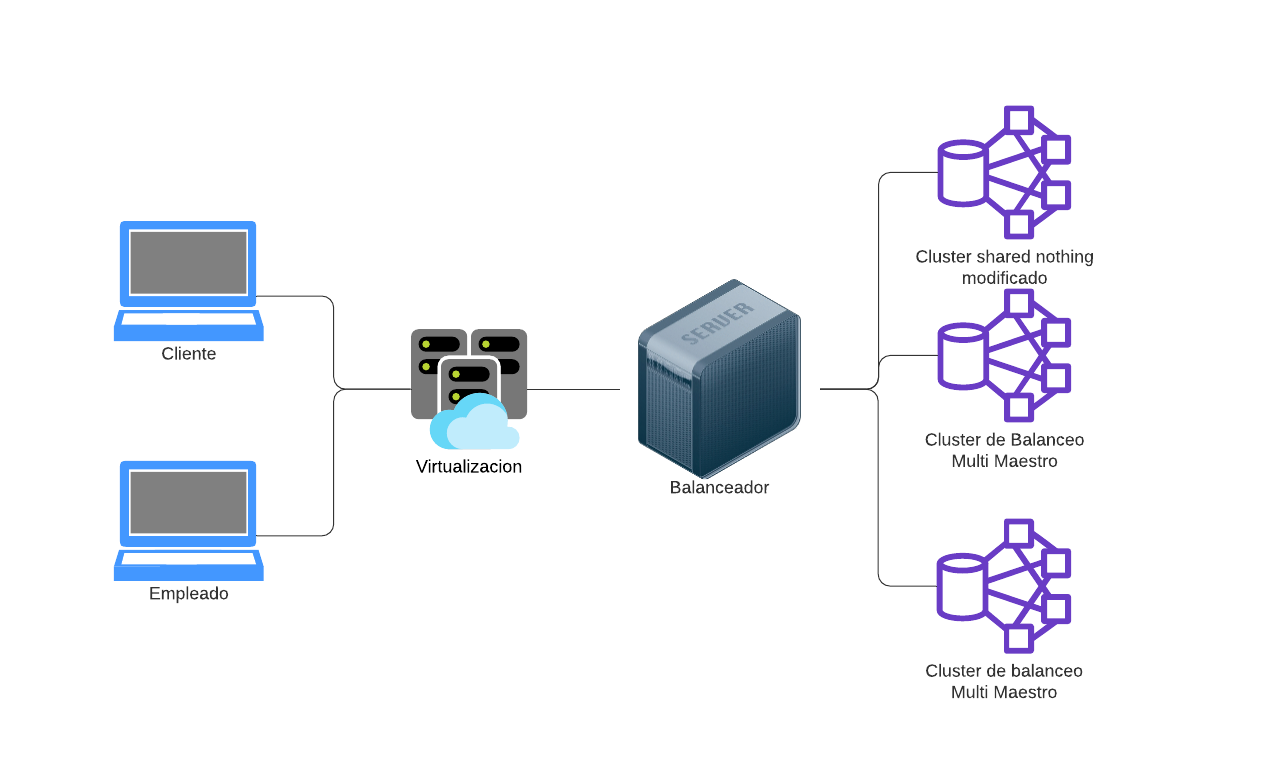
*Diagrama 1. Clúster Shared Nothing modificado*

Como se puede observar, este es el diagrama que representa el clúster de pagina web, historial laboral y tutelas. Para mayor seguridad, se van a virtualizar las conexiones al BPM el cual es la aplicación AFE. Como se menciona en el enunciado de Colpensiones, el AFE encadena diferentes aplicaciones relacionadas con el BPM, las cuales son Pagina web, historia laboral y tutelas. Como se puede evidenciar, se tiene un clúster de balanceo donde el AFE actúa de balanceador activo/pasivo y se comunica como si fuera un clúster shared nothing con los diferentes servidores de aplicaciones. Así mismo para la redundancia se tienen dos servidores por aplicación, para garantizar una mayor disponibilidad del sistema y que puedan soportar una carga mayor los servidores. Adicionalmente para las aplicaciones página web e historia laboral se determino usar un solo rack de servidor ya que manejan la misma información, que es la del historial laboral de los empleados. En el nombre del rack, aclaramos que es un RAID 6. Por otro lado, el servidor de aplicaciones de tutelas contiene un rack de tipo RAID 0-1 para garantizar sus requerimientos.



*Diagrama 2. Clúster de servidores con balanceo para Nomina y Recaudo*

Como se puede evidenciar, para las aplicaciones de nómina y afiliación, las unificamos en un servidor de aplicaciones. Para garantizar, se asegura la virtualización del sistema, donde el servidor de aplicaciones se conecta con un balanceador de carga, el cual distribuye las transacciones de lectura en el servidor que tenga mayor capacidad. Nosotros sabemos que estas aplicaciones son tipo “Batch” por lo que están automatizadas las transacciones, por lo que tenemos que garantizar tanto la disponibilidad del sistema como la capacidad de este. Es por eso que implementamos un sistema de almacenamiento de RAID 0-1 dentro de cada servidor de archivos o datos, esto me garantiza la mayor disponibilidad, una capacidad y desempeño extremadamente altos y lo mas importante, el sistema me provee escalabilidad, ya que, al ser automatizado, la carga varia dependiendo de los procesos.



*Diagrama 4. Infraestructura completa de Colpensiones*

Este es el diagrama completo de la infraestructura que requiere Colpensiones, es necesaria la virtualización para garantizar la seguridad de la infraestructura. El balanceador distribuye las tareas solicitadas dependiendo del servidor de aplicaciones que desee acceder. El balanceador está conectado a los tres clústeres diseñados anteriormente para las diferentes aplicaciones.