CLOUD COMPUTING: PROYECTO FINAL

PLAN DE MIGRACIÓN A LA NUBE DE AWS

APELLIDO: RIVAS NIETO

NOMBRE: CHRISTIAN JAVIER

COMISIÓN: 40520

TUTOR: NAYI JURDI CARVALHO

amazon

CODERHOUSE AÑO 2023

Cloud Computing: Proyecto Final: Plan de migración a la nube AWS

ÍNDICE

1.	Problemática del caso expuesto	2
2.	Descripción de la empresa	3
3.	Objetivos de la migración	3
4.	Descripción de las Tareas a Realizar con el Proyecto	5
5.	Diagrama de Servicios de AWS (Pipeline de AWS)	7
6.	Arquitectura de Servicios de AWS	8
7.	Cronograma de Implementación	10
8.	Estructura de Billing	11
9.	Ventajas e impacto en el negocio de la migración a Cloud	13
10.	. Buenas Prácticas	14
11.	. Conclusiones	15
12	. Referencias bibliográficas	16

1. Problemática del caso expuesto

La empresa "Legend" nos contrata como consultores para construir nuestra primera arquitectura de datos y servicios en AWS, y busca optimizar su proceso de Reporting para plasmar de manera eficiente su información en herramientas de Visualización.

Dado el nivel de crecimiento de la empresa, la misma busca seguir modernizando sus procesos para brindar mejores servicios para sus clientes y tomar las mejores decisiones para hacer seguir creciendo al negocio. El escenario es el que se detalla a continuación:

- La empresa extrae su información transaccional desde SAP CRM
 (Herramienta de Gestión Empresarial para los procesos) donde extrae información como por ejemplo de ordenes shippeadas, ordenes que se encuentran en backlog (que permanecen como ordenes abiertas), información de productos, información de clientes, etc.
- Sus bases de datos están en MySql y Microsoft SQL Server Studio.
- Tenemos mucha información que no está en el sistema de gestión SAP. Se usan muchos archivos de excel (.csv, .xml) para cruzar información adicional a las tablas que enriquecen los datos raw.
- La empresa utiliza distintas aplicaciones de BI como Power BI, Tableau,
 QLIKSense para mostrar la evolución del negocio en sus dashboards, y en base a los resultados mostrados en ellos, poder tomar las decisiones adecuadas en base al contexto.

2. Descripción de la empresa

La empresa "Legend" se dedica al rubro Tecnológico, específicamente se encarga de fabricar Computadoras (Desktops, Notebooks, Workstations, Tablets, Smart Phones) y Servidores. La empresa lleva funcionando varios años y utiliza como sistema transaccional SAP para la gestión de los procesos digitales de la compañía y de sus clientes. También posee muchos archivos de excel que se utilizan como medio de almacenamiento de datos. Si bien la empresa fabrica Servidores y los usa para gestionar su negocio, la misma busca migrar a Cloud Computing para aprovechar todos los beneficios que ella implica.

Dentro de este contexto, dado el gran volumen de datos que la organización se encuentra manejando y de la mano de la necesidad de aplicar un proceso de transformación digital que permita escalar las soluciones de negocios y ofrecer una respuesta rápida y ágil a los requerimientos de los clientes, la compañía busca poder aprovechar a su favor las bondades y beneficios que ofrece el cómputo en la nube en materia de gestión, almacenamiento, procesamiento y administración de los activos informáticos empresariales.

3. Objetivos de la migración

La empresa desea construir su arquitectura en la nube de AWS para hacer frente a los siguientes problemas:

1) Seguridad de los datos y la información: La empresa desea proteger sus datos frente a contingencias o accidentes naturales que en ocasiones han sucedido en el lugar que se alojan sus servidores físicos (como por ejemplo incendios que derivaron la perdida de información crucial), como también protegerlos de ataques maliciosos o ciberataques.

- 2) Tráfico de Datos Fluctuante en épocas de alta demanda, y estacional en épocas de baja demanda: El volumen de transacciones se multiplica en fechas especiales (Black Friday, época de Fiestas, cybermonday, etc) provocando un pico de demanda, mientras que en determinados meses (como en verano) el nivel de ventas disminuye considerablemente. La empresa busca incrementar el nivel de disponibilidad de los servicios críticos de la empresa para estar preparados para los picos de demanda, y usar los beneficios de la nube para adaptarse a las necesidades en épocas de baja demanda (comprar solo cuando es necesario). A su vez, la empresa busca pagar solo por lo que necesite, cuando lo necesite.
- 3) Confidencialidad de los datos y la información: La empresa busca otorgar distintos tipos de acceso a los empleados dependiendo el equipo en el que trabajen, para minimizar la probabilidad de errores causados por falta de conocimiento en determinadas tareas ajenas a cada equipo de trabajo.
- 4) Altos costos por tareas de mantenimiento de los Servidores, puesta a punto, configuración de los mismos, y costo de espacio inmovilizado: La empresa busca reducir costos por temas de mantenimiento de servidores físicos, costo de tiempo de configuración y puesta a punto, costo de espacio físico inmovilizado, etc.
- 5) **Operatividad en todo momento:** La empresa busca poder seguir operando aun cuando ocurran eventualidades como por ejemplo cortes de luz en las zonas donde se encuentran ubicados sus Servidores On Premise.
- 6) Aumento de Productividad: La empresa busca contar con una infraestructura que permita una organización ordenada y eficiente de los datos de la empresa, lo cual conlleva una mejor toma de decisiones a partir de dicho flujo de datos optimizado, y por ende un aumento de la productividad de la empresa.

7) **Modernizar los procesos de IT de la empresa:** Mediante una arquitectura que permita el escalado y crecimiento de la compañía. Organización de los datos de manera eficiente, modernización y optimización de los procesos.

4. Descripción de las Tareas a Realizar con el Proyecto

Los pasos a seguir para el re-host, es decir, pasar de On Premise (Servidores Físicos) a On Demand (Servicios de la Nube) son los siguientes:

- 1) Relevamiento de Datos y ANS (Acuerdo de Nivel de Servicio): Se establece el alcance del Servicio.
- Usaremos el servicio de IAM para (Identity and Access Management) para crear y administrar usuarios, políticas, permisos y roles con el objetivo de proteger los datos.
- 3) Haremos uso del servicio de Amazon VPC (Virtual Private Cloud) para crear subredes, para dirigir el acceso dentro de la base de datos a los perfiles, grupos o roles específicos dentro de la empresa que a su vez estarán respaldadas por grupos de seguridad para proteger las instancias.
- 4) Se usará el Servicio Amazon EC2 (Elastic Computing Cloud) para crear los servidores virtuales en base a la demanda y características de la empresa. Se usará AMI (Amazon Machine Image) como molde para replicar y tener de backup las configuraciones realizadas en Amazon EC2.
- 5) Crearemos un Bucket llamado "Raw Data" en Amazon S3 para alojar en carpetas la Raw Data que proviene de SAP. También podemos alojar en ese Bucket a los flat files que usualmente se usan en la Empresa. También crearemos otro Bucket en Amazon S3 con el nombre de "Processed Data" para alojar luego a toda la Data Raw que transformaremos en los pasos siguientes.
- 6) Usaremos el Servicio de AWS Lambda para que cuando automatizar el proceso de cuando se detecte que la data proveniente de SAP llega al Bucket de Amazon S3, automáticamente el serivicio de AWS Glue tome esa Raw Data para procesar en el paso siguiente.

- 7) Usaremos AWS Glue para realizar el proceso ETL (Extract, Transform and Load) de la Raw Data, es decir, para realizar la limpieza de los Datos y una vez finalizado el proceso, llevar a la Data transformada al Bucket llamado "Processed Data" de Amazon S3.
- 8) Usaremos el Servicio de AWS Lambda para automatizar con un Job que una vez que se detecte que la Data Transformada llega al Bucket "Processed Data" de S3, que sea llevada al servicio de Amazon Redshift donde el Data Warehouse almacenará toda esta información relacional estructurada en Bases de Datos.

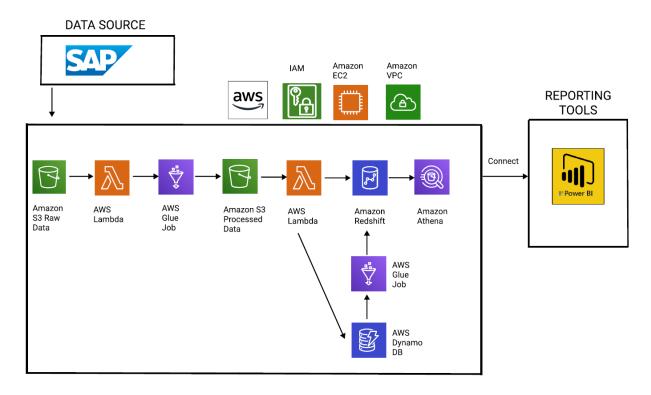
NOTA: Usaremos el servicio de AWS Redshift en lugar del servicio AWS RDS (Relational Database Service) ya que Redshift está orientada a analítica y Reporting como un Data Warehouse empresarial para bases de datos OLAP, y generalmente la empresa maneja Bases de Datos grandes con información histórica de varios años atrás para hacer las proyecciones de sus ventas. Mientras que Amazon RDS se usa generalmente como base de datos tradicional OLTP. Es decir:

Bases de Datos:

- OLTP: Bases de Datos medianamente chicas, pensadas para transacciones en las que a los usuarios les importa ver como mucho información histórica de 6 meses atrás. Para estas Bases de Datos se suele usar el servicio de RDS.
- OLAP: Son Bases de Datos más grandes que tienen información histórica de más tiempo (son de más largo plazo). Para estas Bases de Datos se suele usar el servicio de Redshift.
- 9) Para agilizar la velocidad de procesamiento en épocas de alta demanda de nuestros productos, podemos utilizar AWS Lambda para automatizar también que la data voluminosa sea llevada al servicio de Amazon Dynamo DB (Las Bases de datos SQL de Redshift están diseñadas para soportar una cantidad de cientos de miles de registros, pero no para millones de registros, como si lo soportan las Bases de Datos NoSQL de Dynamo DB. Es por este motivo que NoSQL es mucho más eficiente para grandes volúmenes de datos que SQL). Luego, para llevar la data no estructurada a las Bases de datos SQL de Redshift para hacer una sumarización de los datos (para sumarizar las ventas,

- costos, profits) por ejemplo, utilizaremos el servicio de AWS Glue con el cuál realizaremos el proceso ETL para transformar la data No Relacional y llevarla a las Bases de Datos Relacionales de Redshift.
- 10) Usaremos el Servicio de Amazon Athena para ejecutar queries de SQL para explorar la Data en las Bases de Datos de Amazon Redshift. Y usaremos también Amazon Athena para conectar AWS con Power Bl, y de esta manera poder llevar la Data Estructurada desde las Bases de Datos de Redshift a las aplicaciones de Visualización, como por ejemplo Power Bl, para luego ser usada la información para crear Dashboards visuales que permitan tomar las mejores decisiones para el negocio.

Diagrama de Servicios de AWS (Pipeline de AWS)



6. Arquitectura de Servicios de AWS

Arquitectura de servicios												
Servicio	Descripción	Justificación										
IAM	AWS Identity and Access Management (IAM) es un servicio web que nos ayuda a controlar de forma segura el acceso a los recursos de AWS.	IAM nos permite crear usuarios, administrar permisos y roles dentro de la nube, y por ende controlar de forma segura el acceso a los recursos dentro de la nube.										
EC2	Amazon Elastic Compute Cloud (Amazon EC2) proporciona capacidad de computación escalable en la nube de Amazon Web Services (AWS). Puede usarse para lanzar tantos servidores virtuales, configurar la seguridad y las redes, y administrar el almacenamiento.	El uso de Amazon EC2 elimina la necesidad de invertir inicialmente en hardware, de manera que puede desarrollar e implementar aplicaciones en menos tiempo y a un costo significativamente inferior. A su vez, puesto que la organización requiere el procesamiento de muchos datos, es factible integrar este servicio en clúster de máquinas virtuales.										
VPC	Amazon VPC (Virtual Private Cloud) permite crear una conexión de red privada virtual (VPN) de hardware entre el centro de datos de la empresa y la VPC, y utilizar la nube de AWS como una prolongación del centro de datos corporativo.	VPC nos permite contar con una red de acceso restringida para únicamente los empleados y específicamente para los usuarios de cada equipo de trabajo, y de esta manera prevenir también a la empresa de ciberataques para proteger su información.										
S 3	Amazon Simple Storage Service (Amazon S3) es un servicio de almacenamiento de objetos que ofrece escalabilidad, disponibilidad de datos, seguridad y rendimiento líderes en el sector.	S3 nos permite optimizar los costos, organizar archivos de todo tipo y configurar controles de acceso detallados para cumplir con los requisitos empresariales.										

Redshift	Servicio de base de datos relacional distribuida de Amazon Web Services. Es un servicio web que se ejecuta "en la nube" diseñado para simplificar la configuración, el funcionamiento y el escalado de una base de datos relacional para su uso en aplicaciones. Utiliza SQL para analizar datos estructurados y semiestructurados en almacenamientos de datos, bases de datos.	Dado que la organización cuenta con bases de datos relacionales, el servicio que nos permite gestionar motores relacionales para Bases de Datos grandes (OLAP, con data histórica de varios años atrás) en la nube de AWS es Amazon Redshift. En consecuencia, debe incluirse como un servicio clave dentro de la arquitectura propuesta.
Dynamo DB	Amazon DynamoDB es un servicio de base de datos no relacional rápido y flexible para todas las aplicaciones que requieren latencia uniforme en milisegundos de un solo dígito a cualquier escala.	Resulta clave contar con un servicio de almacenamiento de bases de datos no relacional para poder gestionar información no estructurada o para ser más eficiente el procesamiento de grandes volúmenes de data, y gracias a su escalabilidad horizontal, obtener grandes velocidades de procesamiento.
Lambda	AWS Lambda es un servicio que nos permite combinar con otros servicios de AWS para crear experiencias en línea seguras, estables y escalables.	Lambda nos permite crear funciones basadas en eventos para facilitar la comunicación entre servicios desacoplados ("Cuando pase A, que suceda B"), y de esta manera automatizar procesos.
Athena	Amazon Athena es un servicio de análisis interactivo y sin servidor creado en marcos de código abierto, lo que lo hace compatible con formatos abiertos de archivos y tablas. Athena proporciona un método simplificado y flexible de analizar petabytes de datos donde residan.	Athena nos permite ejecutar queries de SQL para explorar la info en nuestras Bases de Datos de RDS, y a su vez, conectar a la Base de Datos con aplicaciones de Visualización (como Power BI).

Glue	AWS Glue es un servicio de integración de datos sin servidor que facilita la detección, preparación, migración e integración de datos provenientes de varios orígenes para el análisis, machine learning (ML) y desarrollo de aplicaciones.	Gracias a Glue podremos hacer la extracción, transformación y la carga de los datos transformados al Data Lake de Amazon S3, y luego llevarlos al Data Warehouse de RDS.
------	---	---

7. Cronograma de Implementación

	. Actividad	Mes	Mes 1					Mes 3				Mes 3						
# Act.		Semana	S1	62	63	9/1	Q1	S2	co	64	C1	63	co	64	C1	52	63	64
		Duración	5	32	33	34	31	32	33	34	31	32	33	34	31	32	33	34
1	Relevamiento de Datos y ANS	3 Semanas																
2	Creación de Usuarios en IAM	2 Semanas																1
3	Creación de servidores EC2	2 Semanas																
4	Creación de VPC	2 Semanas																
5	Creación de Buckets en S3	1 Semana																
6	Automatización S3-Glue con Lamda	1 Semana																
7	Creación del ETL con Glue Raw - Processed data	2 Semanas																l l
8	Creación de Bases de datos en Redshift	2 Semanas																
9	Creación de Bases de datos en Dynamo DB	2 Semanas																
10	Automatización S3-Redshift con Lamda	1 Semana																
11	Creación del ETL con Glue Dynamo - Redshift	2 Semanas																
12	Creación Conexión Athena - Power Bl	2 Semanas			_	_		_										

8. Estructura de Billing

Se utilizará el servicio de Calculadora de AWS para estimar un costo aproximado por cada servicio. Recordemos que el Servicio de IAM es gratuito.

Para nuestro proyecto tendremos 3 ambientes distintos con los siguientes grupos:

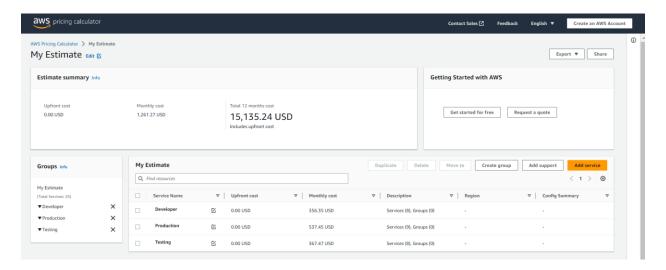
IAM: Creación de los siguientes grupos de usuarios:

- a. Developers
- b. Testing
- c. Production

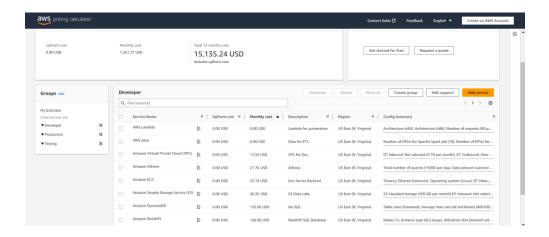
Link:

https://calculator.aws/#/estimate?id=e1de523eab3da0fa0c393c897aff0472 db19f1dd

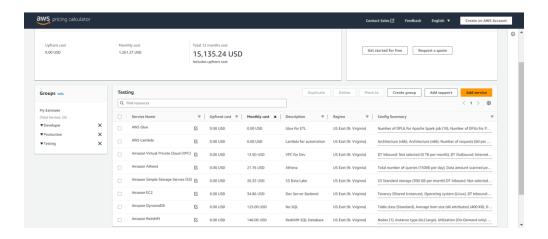
Costos Totales por Ambiente:



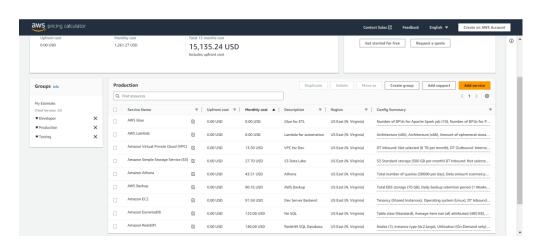
• Developers:



Testing:



• Production:



9. Ventajas e impacto en el negocio de la migración a Cloud

- Ambientes seguros: AWS nos ofrece un nivel muy alto de seguridad y sofisticación, para la administración y gestión de ambientes de trabajo.
- Soporte/mantenimiento: Estos costos suelen ser mensuales, permitiendo una presupuestación simple de los distintos ambientes de la empresa.
- Pasaje de Gastos CAPEX (Gastos de Capital como por ejemplo invertir en Servidores Físicos) a OPEX (Gastos Operativos que van variando según el consumo, como electricidad, agua, etc).
- Costo de almacenamiento ON DEMAND: Podemos agregar más capacidad sin necesidad de comprar hardware (Solo Pago por lo que Uso). La empresa puede pagar solo por lo que necesite, cuando lo necesite. Pasaje de On Premise (Hardware Físico) a On Demand (Hardware Virtual que Pago por el uso).
- Dejar de adivinar la Capacidad: No me tengo que comprometer en la compra de algo para estar atado cierto tiempo, sino que puedo comprar en base a la demanda.
- Aumenta la velocidad y agilidad: no tenemos tiempos de espera en ir a comprar el hardware, instalarlo, configurar todo, etc. En AWS, con un par de clicks levantamos un servidor. El tiempo de implementación (Time to Market) se reduce mucho.
- Ser Global en minutos: Si necesito tener un servidor en Europa porque mis principales clientes están en Europa, con AWS con dos clicks monto un servidor en Europa. Con On Premise sería un proceso costosísimo y que demoraría mucho.
- En caso de necesitar más recursos, facilidad de realizarlo con un solo par de clicks.
- Enfocarse en lo que importa: Voy a poder contratar servicios de la nube que me van a dar mantenimiento en todo momento, y nos permite concentrarnos

- y focalizar nuestra energía en tomar las decisiones para hacer crecer a nuestro negocio.
- Beneficios de enormes economías de escala: El mismo servidor corriendo para todos, la misma placa de red corriendo para todos, etc.

10. Buenas Prácticas

- 1. Utilizar AWS Trusted Advisor (AWS cuenta con un servicio de asistencia para buenas prácticas, reducción de costos y alertas. Trusted Advisor nos acompañará en el uso de nuestros servicios dándonos recomendaciones que podemos no encontrar a simple vista nosotros mismos) con el fin de utilizar comprobaciones de optimización de costos para lograr un nivel básico de ahorro de costos. Un ejemplo de comprobación de Optimización de Costos es la comprobación de Recursos Inactivos (instancias de Amazon EC2, Redshift, etc).
- 2. Etiquetado: Es una buena práctica para administrar los servicios de AWS ya que con los tags podemos identificar a qué área/equipo o Centro de Costos o Ambiente pertenece. Además, podemos luego usarlo para analizar e identificar más fácilmente el costo según el criterio de filtro por etiqueta.
- 3. Iniciar pequeños y crecer en función de la Demanda: Para cualquier Servicio de AWS, es recomendable que empecemos con la cantidad justa de recursos y conforme crezca nuestra Demanda (y además conozcamos nuestra aplicación), que vayamos aumentando la capacidad o tamaño para ir consumiendo los recursos justos o necesarios, y hacer uso del beneficio de "On Demand" (que la empresa pague solo por lo que necesite, cuando lo necesite).
- 4. Utilizar los Reportes: Utilizar Cost Explorer, Dashboard Billing, Cloudwatch
 y Budgets para armar presupuestos y crear alertas necesarias para estar

pendientes del consumo y evitar sobrecostos. Estas alertas nos permitirán tomar decisiones rápidamente, y no esperar a ver la factura para tomar decisiones. Mediante estas alertas podremos prevenir errores como detectar a tiempo un servicio mal apagado, así como también crear alertas para detectar actividades maliciosas a tiempo y prevenir algún intento de violación a la seguridad del servicio.

- 5. Utilizar la calculadora: https://calculator.aws/#/
 Usar la calculadora antes de crear un servicio nos ayudará para comparar el costo de distintas alternativas dependiendo los tamaños y tipo de servicios.
 Con la calculadora predeciremos qué costo aproximado vamos a tener por cada servicio de AWS.
- 6. Realizar copias de Seguridad (Backups) semanales en las Bases de Datos para prevenir la pérdida de información por alguna incidente interno o externo a la empresa.
- Cambiar las contraseñas de los usuarios cada cierto tiempo por cuestiones de seguridad.

11. Conclusiones

Con el presente plan de migración, podemos identificar en función de la problemática una arquitectura propuesta que ayudará a la organización en el logro de sus objetivos (aumentará su productividad a la hora de tomar decisiones, sus datos estarán seguros, optimizará sus costos pagando solo lo que se necesite cuando se necesite, estará preparada para los meses de alta demanda, se garantizará su operatividad en todo momento y modernizará sus sistemas informáticos) dentro de un marco del presupuesto y cronograma establecidos en los incisos anteriores.

12. Referencias bibliográficas

- AWS: https://aws.amazon.com/es/
- IAM: https://aws.amazon.com/es/iam/
- VPC: https://aws.amazon.com/es/vpc/
- EC2: https://aws.amazon.com/es/ec2/
- S3: https://aws.amazon.com/es/s3/
- Lambda: https://aws.amazon.com/es/lambda/
- Glue: https://aws.amazon.com/es/glue/
- Redshift: https://aws.amazon.com/es/redshift/
- Athena: https://aws.amazon.com/es/athena/
- DynamoDB: https://aws.amazon.com/es/dynamodb/