7Sistema Gestor de Bases de Datos Riak

\*Note: Sub-titles are not captured in Xplore and should not be used

line 1: 1st Given Name Surname   
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 2nd Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

line 1: 3rd Given Name Surname  
line 2: *dept. name of organization   
(of Affiliation)*  
line 3: *name of organization   
(of Affiliation)*line 4: City, Country  
line 5: email address or ORCID

*Abstract*—El texto describe cómo Riak ofrece soluciones escalables y altamente disponibles para gestionar datos no estructurados en entornos empresariales. Riak simplifica la integración y el despliegue de cargas de trabajo activas para aplicaciones de Big Data, IoT y nubes híbridas. Se destaca la capacidad de Riak para garantizar alta disponibilidad, tolerancia a fallos y escalabilidad, siendo utilizado por empresas líderes en Fortune 50. Además, se presenta el viaje de Riak desde su fundación en 2008 hasta convertirse en una referencia en sistemas distribuidos. El texto también detalla las características y beneficios de las bases de datos Riak KV, Riak TS y Riak S2, destacando su capacidad para gestionar datos clave-valor, series temporales y almacenamiento de objetos grandes, respectivamente.

Keywords— Riak, Sistemas distribuidos, Escalabilidad, Alta disponibilidad, Datos no estructurados, Cargas de trabajo activas, Big Data, IoT, Nube híbrida, Riak KV, Riak S2, Riak TS, Apache Spark, Redis, Apache Solr, Integración, Simplificación, Tolerancia a fallos, Escalado rápido

# Introducción

# En este artículo, se explora cómo Riak está revolucionando la gestión de datos en el mundo empresarial, encontrando un equilibrio entre la complejidad del Big Data y la simplicidad operativa. Desde su fundación en 2008, Riak se ha destacado como una fuerza pionera en sistemas distribuidos, desarrollando soluciones escalables y altamente disponibles para almacenar y administrar datos no estructurados ().

# En este análisis, se examina en detalle la filosofía de Riak y sus diversas plataformas, como Riak KV, Riak TS y Riak S2, diseñadas para abordar las necesidades específicas de Big Data, IoT y almacenamiento de objetos. Además, exploraremos cómo Riak ha atraído a una amplia gama de clientes de diferentes industrias, consolidando su posición como líder en la gestión de bases de datos distribuidas e inteligentes ().

# Historia

Riak fue creado en 2007 por Basho Technologies, una empresa fundada por Rich Murphey y Alex Polvi. El objetivo era crear un sistema de gestión de bases de datos NoSQL que fuera escalable, confiable y fácil de usar.

Desde su lanzamiento inicial, Riak ha experimentado un desarrollo continuo. Se han añadido nuevas características y funcionalidades, y se ha mejorado el rendimiento y la escalabilidad. Riak se ha utilizado en una amplia gama de aplicaciones, desde sitios web de comercio electrónico hasta aplicaciones de análisis de big data.

En 2017, Cockroach Labs adquirió Basho Technologies y Riak. Cockroach Labs es una empresa de software que desarrolla una base de datos SQL distribuida y tolerante a fallos. La adquisición de Riak permitió a Cockroach Labs ampliar su oferta de productos y ofrecer una solución completa para la gestión de bases de datos NoSQL y SQL.

Riak sigue siendo un sistema de gestión de bases de datos NoSQL popular y ampliamente utilizado. Se utiliza por empresas de todos los tamaños para almacenar grandes cantidades de datos de forma escalable y confiable.

# Concepciones

Riak se destaca como una solución líder en el panorama de gestión de datos al ofrecer sistemas distribuidos escalables y altamente disponibles. Su enfoque se centra en simplificar la complejidad de la gestión de datos en entornos empresariales, permitiendo a las aplicaciones almacenar y recuperar datos no estructurados de manera eficiente y confiable.

La plataforma de datos de Riak sirve como base sólida para abordar los desafíos actuales en la integración y despliegue de cargas de trabajo activas en aplicaciones de Big Data, IoT y nube híbrida. Riak KV, conocido por su flexibilidad y disponibilidad, es respaldado por Riak S2, una base de datos optimizada para series temporales, y Riak TS, diseñado específicamente para datos de IoT y series temporales. Además, Riak ofrece complementos para tecnologías ampliamente utilizadas como Apache Spark, Redis y Apache Solr, mejorando aún más su versatilidad y utilidad en diversos casos de uso.

La capacidad de Riak para trascender el caos del Big Data no estructurado en entornos empresariales de TI radica en su capacidad para garantizar alta disponibilidad, tolerancia a fallos, escalabilidad y simplicidad operativa. Las organizaciones pueden confiar en Riak para almacenar y gestionar datos de sesión, usuario, dispositivo y una variedad de otros usos de manera eficiente y segura.

El éxito de Riak se refleja en su base de clientes diversa y en constante crecimiento. Más de un tercio de las empresas Fortune 50 han implementado Riak para impulsar sus aplicaciones de Big Data, lo que demuestra su confiabilidad y eficacia en entornos empresariales críticos. Además, un número creciente de usuarios recurren a Riak en momentos en que necesitan escalar rápidamente aplicaciones de misión crítica y generadoras de ingresos, confiando en su capacidad para ofrecer un rendimiento predecible y confiable.

Desde su fundación en 2008, Riak ha estado a la vanguardia del desarrollo de tecnologías disruptivas para abordar los desafíos más críticos en la gestión de datos en sistemas distribuidos. Su equipo de ingenieros altamente talentosos y dedicados trabaja incansablemente para resolver problemas complejos y mejorar continuamente la plataforma Riak. Además, Riak ha cultivado una comunidad activa de seguidores entre programadores, arquitectos y académicos influyentes, y es el anfitrión de la conferencia RICON, que reúne a destacados expertos en el campo de la gestión de bases de datos distribuidas e inteligentes. En resumen, Riak representa una opción sólida y confiable para empresas y organizaciones que buscan gestionar eficazmente sus datos en entornos distribuidos y exigentes.

En 2009, lanzada la primera versión de Riak basada en la investigación y el desarrollo de la empresa, Además, Riak fue creada desde el principio para ola disponibilidad, tolerancia a los fallos. Asi mismo, se basa en el principio del modelo clave-valor donde los datos se almacenan en pares de clave-calor distribuido entre diferentes nodos.

En 2010 a 2015, Riak comenzó a ganar bastante publicidad entre las empresas que necesitaban una Base de Dato altamente escalable y tolerante a los fallos. también, su capacidad para el control de altos volúmenes de Dato y ofrecer una alta disponibilidad los convirtió en una opción llamativa para aplicaciones web y móvil como para casos de uso en la nube y el análisis de datos.

# Arquitectura

\*RIAK KV

Riak KV es una base de datos clave-valor NoSQL distribuida que se destaca por su capacidad de replicación avanzada y multiclúster, garantizando la disponibilidad incluso en situaciones de fallos de hardware o particiones de red. Su diseño clave/valor, que ofrece modelos de datos potentes pero simples, lo convierte en una solución ideal para almacenar grandes volúmenes de datos no estructurados y abordar una amplia gama de desafíos en aplicaciones de Big Data.

Una de las características sobresalientes de Riak KV es su capacidad para automatizar la distribución de datos en todo el clúster, lo que asegura un rendimiento óptimo y una sólida continuidad del negocio. Esto se logra mediante una arquitectura sin maestro que escala casi linealmente, permitiendo la fácil adición de capacidad con hardware básico. Además, Riak KV es altamente resistente ante interrupciones imprevistas, garantizando que las operaciones de lectura y escritura puedan continuar incluso en situaciones adversas.

* Características del Riak KV:

Resiliencia: Riak KV elimina el impacto de lo inesperado, permitiendo que las operaciones continúen incluso si se producen particiones de red o fallos de hardware.

Escalabilidad masiva: Su arquitectura escalable permite agregar capacidad sin problemas, mejorando el rendimiento de manera casi lineal.

Simplicidad operativa: La adición de nodos a Riak KV es fácil y automática, lo que garantiza una distribución uniforme de los datos en todo el clúster.

Replicación inteligente: Riak KV está diseñado para replicar y recuperar datos de forma inteligente, asegurando que las aplicaciones de Big Data, IoT o nube híbrida estén siempre disponibles.

Compatibilidad con consultas complejas: Ofrece diversas opciones para consultar datos, incluyendo búsqueda de texto completo, índices secundarios y reducción de mapas.

Caducidad global de objetos: Permite especificar cuándo se eliminan los datos antiguos de la base de datos de manera automática y eficiente.

Vectores de versión punteada (DVV): Riak KV realiza un seguimiento del tiempo lógico para resolver conflictos de manera rápida y automática.

Tipos de datos distribuidos: Proporciona tipos de datos predefinidos para las estructuras de datos más comunes utilizadas en cargas de trabajo activas distribuidas.

API y bibliotecas de cliente robustas: Facilitan la integración con una amplia gama de lenguajes de programación, como Java, Ruby, Python, C#, Erlang, Node.js y .NET.

Además, Riak KV ofrece integraciones con otras tecnologías clave, como Apache Spark, Apache Mesos y Redis, lo que amplía aún más su funcionalidad y versatilidad en entornos de Big Data. Con Riak KV, las organizaciones pueden construir soluciones completas y robustas para satisfacer sus necesidades de gestión de datos en entornos distribuidos y exigentes.

\*RIAK TS

Riak TS es una base de datos NoSQL especialmente diseñada para manejar datos de series temporales en entornos empresariales, con un enfoque particular en las necesidades de las aplicaciones de IoT (Internet de las cosas) y análisis de series temporales. Sus características y funcionalidades se han diseñado cuidadosamente para garantizar un rendimiento óptimo, una alta disponibilidad y una fácil escalabilidad:

Garantía de alto rendimiento para aplicaciones de IoT:

Empresas que utilizan Riak TS para gestionar datos de series temporales disfrutan de una serie de beneficios cruciales que incluyen la capacidad de tomar decisiones en tiempo real, reducir costos operativos, aumentar la escalabilidad y el rendimiento, minimizar el tiempo de inactividad, acelerar el tiempo de comercialización y gestionar eficientemente grandes volúmenes de datos.

* Características de Riak TS:

Resiliencia: Riak TS está diseñado con una arquitectura sin maestro que replica automáticamente los datos para garantizar que los datos de series temporales estén siempre disponibles, incluso en situaciones de fallos de hardware o de red. Esto elimina el tiempo de inactividad y garantiza la continuidad del negocio.

Escalabilidad: La arquitectura distribuida de Riak TS permite escalar fácilmente para satisfacer picos de demanda repentinos. Puede agregar o eliminar capacidad sin problemas utilizando hardware básico, lo que garantiza un rendimiento casi lineal incluso en entornos de alta carga.

Simplicidad operativa: Riak TS simplifica las operaciones a escala empresarial. Agregar nuevos nodos al clúster, distribuir datos y realizar actualizaciones son tareas sencillas, lo que reduce la carga operativa y facilita la gestión del sistema.

Co-Ubicación de datos: Agrupar los datos relacionados temporalmente en la misma parte física del clúster mejora significativamente el rendimiento de las consultas y análisis de series temporales, lo que garantiza un acceso más rápido a la información relevante.

Comandos SQL: Riak TS admite consultas SQL, lo que facilita la manipulación y análisis de datos semiestructurados de series temporales. Esto permite a los usuarios realizar consultas complejas y extraer información valiosa de sus conjuntos de datos.

Consultas de intervalo SQL: Las consultas basadas en el tiempo permiten a los usuarios recuperar datos dentro de un rango específico, lo que facilita el análisis temporal y la identificación de tendencias o patrones.

Agregaciones: Riak TS ofrece funciones de agregación integradas que permiten a los usuarios consolidar y analizar grandes conjuntos de datos de manera eficiente, reduciendo el tiempo necesario para realizar operaciones complejas.

Caducidad de los datos: La función de caducidad de datos de Riak TS permite a los usuarios especificar cuándo se eliminan automáticamente los datos obsoletos, optimizando el almacenamiento y mejorando el rendimiento del sistema.

Conector Apache Spark: La integración con Apache Spark permite realizar análisis operativos eficientes de datos de series temporales, lo que facilita la integración con otras herramientas analíticas y la generación de información valiosa a partir de los datos.

API y bibliotecas de cliente robustas: Riak TS proporciona soporte para una variedad de lenguajes de programación, lo que facilita el desarrollo de aplicaciones de IoT y series temporales en diferentes entornos de desarrollo.

Replicación de varios clústeres: Riak TS permite replicar clústeres local o globalmente, lo que mejora la geolocalización de datos, el análisis secundario y garantiza la continuidad del negocio incluso en entornos distribuidos.

Marco Apache Mesos: Riak Meso Framework proporciona una gestión eficiente de recursos y escalado vertical/reducido para los nodos de Riak, lo que mejora su rendimiento y eficiencia operativa en entornos empresariales.

\*RIAK S2

Riak S2 es una solución de almacenamiento de objetos que se destaca por su alta disponibilidad, escalabilidad y facilidad de operación. Está diseñado específicamente para manejar archivos grandes como videos, imágenes y otros tipos de datos. Veamos con más detalle las características clave de Riak S2:

1. Alta Disponibilidad:

Riak S2 garantiza que los objetos almacenados estén siempre disponibles para las aplicaciones, incluso en caso de fallos de hardware o de red. Utiliza una replicación inteligente de objetos en el clúster para asegurar esta disponibilidad continua.

2. Compatibilidad con Amazon S3:

Una de las ventajas de Riak S2 es su compatibilidad con las API de Amazon S3, lo que facilita a los desarrolladores el uso de herramientas y bibliotecas existentes. Además, es compatible con OpenStack Swift, lo que amplía aún más sus capacidades de integración.

3. Corrección Automática de Datos:

Riak S2 supervisa continuamente los datos almacenados y repara automáticamente cualquier inconsistencia que pueda surgir entre las réplicas. Esto garantiza la integridad de los datos almacenados en todo momento.

4. Escalabilidad Masiva:

La arquitectura de Riak S2 permite escalar fácilmente para manejar grandes volúmenes de datos. Puede aumentar la capacidad de almacenamiento agregando nodos al clúster de manera casi lineal, lo que proporciona un rendimiento predecible y escalable.

5. Arrendamiento Múltiple:

Riak S2 ofrece informes por inquilino sobre el uso de datos y estadísticas. Esto es útil en entornos de múltiples inquilinos, ya que facilita la facturación y la medición de recursos utilizados por cada usuario.

6. Almacenamiento de Varios Niveles:

Para optimizar el rendimiento, Riak S2 permite priorizar el acceso a los datos colocando los más frecuentemente utilizados en medios más rápidos. Esto ayuda a garantizar una baja latencia y una respuesta rápida a las solicitudes de los usuarios.

7. Almacenamiento de Objetos Grandes:

Con la capacidad de carga de varias partes, Riak S2 puede manejar eficientemente archivos de gran tamaño, desde gigabytes hasta terabytes. Esto lo hace adecuado para aplicaciones que requieren almacenamiento de objetos grandes, como videos de alta definición o copias de seguridad de bases de datos.

8. Simplicidad Operativa:

La instalación y configuración de Riak S2 son sencillas y requieren solo tres componentes. Además, agregar capacidad adicional al clúster se realiza fácilmente mediante la incorporación de nuevos nodos, lo que simplifica la gestión operativa.

9. Replicación de Varios Clústeres:

Para garantizar la continuidad del negocio, Riak S2 ofrece almacenamiento de baja latencia que mantiene la disponibilidad para los usuarios en caso de fallo del sitio. Esto se logra mediante la replicación de datos en varios clústeres, lo que garantiza la redundancia y la resistencia a los fallos.

Riak S2 es una solución completa y robusta para el almacenamiento de objetos, diseñada para satisfacer las demandas de aplicaciones que manejan grandes volúmenes de datos en entornos de nube privada, pública e híbrida. Su combinación de alta disponibilidad, escalabilidad y facilidad de operación lo convierte en una opción atractiva para una amplia gama de casos de uso.

Ventajas y Desventajas

**Ventajas.**

* Escalamiento Sencillo.
* Buen Rendimiento en entornos con aplicaciones de alta disponibilidad.
* Muchos Api`s.
* MapReduce en Erlang y JavaScript.
* Enterprise Edition.
* Escabilidad horizontal.
* Tolerancia a fallos.
* Flexibilidad en el esquema de datos.
* Modelo de consistencia configurable.

**Desventajas.**

* Complejidad de configuración.
* Consistencia eventual.
* Curvas de aprendizaje.
* Herramientas de administración limitadas.
* Requisitos de hardware y escalabilidad.
* Documentación limitada.
* Requisitos de mantenimiento.
* Integración con herramientas de terceros.
* Rendimiento en cargas de trabajo especifica.

# Referencias

Home page. (s. f.). Riak. Recuperado 12 de abril de 2024, de https://riak.com/index.html

Wikipedia contributors. (2024, febrero 2). Basho Technologies. Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Basho\_Technologies&oldid=1202398319

*Home page*. (s. f.). Riak. Recuperado 12 de abril de 2024, de https://riak.com/

Riak documentation. (s. f.). Riak.com. Recuperado 12 de abril de 2024, de https://docs.riak.com/

*Riak KV*. (s. f.). Riak.com. Recuperado 12 de abril de 2024, de https://docs.riak.com/riak/latest/

https://es.slideshare.net/andreicito1/bases-de-datos-nosql-riak