Universidad Mayor de San Andrés Facultad de Ciencias Puras y Naturales

Informe sobre CouchDB

DAT251 - Base de Datos III



Docente:

Lic. Celia Elena Tarquino Peralta

Estudiante:

Gabriel Muñoz Marcelo Callisaya

Dupla:

23

Fecha de entrega:

7 de mayo del 2025

La Paz - Bolivia

Análisis de CouchDB: Fundamentos y Aplicaciones

1. Marco Teórico

1.1. Concepto

CouchDB es una base de datos NoSQL orientada a documentos, desarrollada bajo el auspicio de la Apache Software Foundation. Creada en 2005 por Damien Katz, su diseño responde a la necesidad de sistemas distribuidos que prioricen la disponibilidad y la tolerancia a particiones. En 2008, se convirtió en un proyecto de Apache, consolidándose como una solución robusta para entornos distribuidos. CouchDB almacena datos en documentos JSON, identificados por un campo _id único, lo que permite estructuras flexibles sin un esquema rígido. Su arquitectura se basa en una API RESTful, que utiliza solicitudes HTTP para operaciones CRUD (Crear, Leer, Actualizar, Eliminar). El lema «relax» refleja su enfoque en la simplicidad, escalabilidad y facilidad de uso.

La evolución de CouchDB incluye mejoras en replicación, consultas avanzadas (Mango Queries, introducidas en 2015) y soporte para entornos móviles. Su diseño distribuido lo distingue de bases de datos relacionales y lo alinea con otras soluciones NoSQL como MongoDB y Cassandra, aunque con un enfoque único en la sincronización bidireccional.

1.2. Utilidad

CouchDB es ideal para aplicaciones que requieren alta disponibilidad y operación en entornos con conectividad intermitente, como aplicaciones móviles, sistemas de gestión de contenido (CMS) y plataformas web escalables. Según el teorema CAP (Consistencia, Disponibilidad, Tolerancia a Particiones), CouchDB prioriza la disponibilidad y la tolerancia a particiones sobre la consistencia inmediata, adoptando un modelo de consistencia eventual. Esto permite que los nodos funcionen de forma independiente y sincronicen datos cuando sea posible, lo que es crucial para sistemas distribuidos globalmente.

Por ejemplo, CouchDB es utilizado en aplicaciones como IBM Cloudant y en proyectos de sincronización offline, donde los usuarios pueden trabajar sin conexión y sincronizar datos al reconectarse. Su capacidad de replicación bidireccional lo hace adecuado para entornos donde los datos deben estar disponibles en múltiples ubicaciones sin un punto central de fallo.

1.3. Herramientas

CouchDB ofrece dos interfaces principales:

- Interfaz de línea de comandos (CLI): Utiliza comandos curl para enviar solicitudes HTTP a la API RESTful, permitiendo operaciones como creación de documentos y consultas.
- **Fauxton**: Una interfaz gráfica web, accesible en http://localhost:5984/_utils/, que simplifica la gestión de bases de datos, documentos y consultas.

Los datos se organizan en documentos JSON, almacenados en bases de datos. Cada documento es un objeto autónomo con un _id único y campos arbitrarios.

CouchDB soporta **Mango Queries**, un sistema de consultas declarativas similar a MongoDB, que utiliza selectores JSON para buscar documentos. Las consultas se optimizan mediante índices, que deben crearse manualmente para mejorar el rendimiento.

Funcionamiento de las consultas: Las Mango Queries operan sobre la API _find, evaluando condiciones definidas en un selector (por ejemplo, {"precio": {"\$gte": 10}}). La respuesta incluye solo los campos especificados, reduciendo la carga de datos. La eficiencia depende de índices predefinidos; sin ellos, CouchDB realiza búsquedas completas, lo que puede ser lento en bases de datos grandes. Este enfoque combina flexibilidad con control, aunque requiere planificación para consultas complejas.

1.4. Problemática a abordar

CouchDB aborda la necesidad de bases de datos distribuidas en entornos donde la conectividad es inestable o los datos deben replicarse globalmente. Su arquitectura permite nodos independientes que sincronizan datos mediante protocolos de replicación, resolviendo problemas de latencia y disponibilidad. Sin embargo, presenta desafíos:

Ventajas:

- Alta disponibilidad y tolerancia a particiones.
- Replicación bidireccional para sincronización offline.
- API RESTful accesible desde cualquier lenguaje de programación.
- Tolerancia a fallos mediante diseño distribuido.

Desventajas:

- Consistencia eventual, que puede generar conflictos en datos.
- Consultas complejas menos intuitivas que en bases de datos relacionales.
- Rendimiento limitado en operaciones de scritura intensiva debido a la arquitectura de almacenamiento.

CouchDB es óptimo para aplicaciones como CMS, aplicaciones móviles con sincronización offline y sistemas IoT, pero no para escenarios que requieren transacciones ACID estrictas o joins complejos, donde bases de datos relacionales o MongoDB pueden ser más adecuadas.

2. Marco Aplicativo

El marco aplicativo detalla la implementación práctica de CouchDB, incluyendo instalación, configuración, gestión de datos y consultas. Cada proceso se presenta como una solución práctica, con ejemplos ilustrativos y resultados verificables.

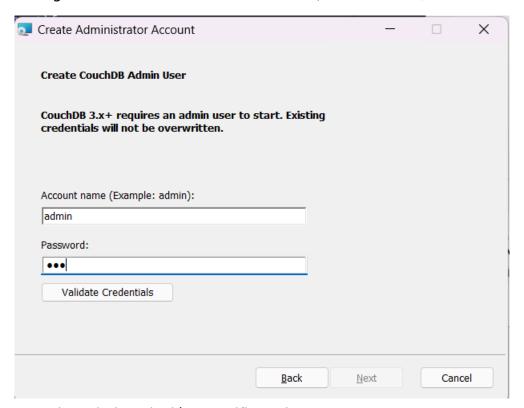
2.1. Instalación

CouchDB 3.4.3 se instaló en Windows usando el archivo apache-couchdb-3.4.3.msi. Los pasos clave fueron:

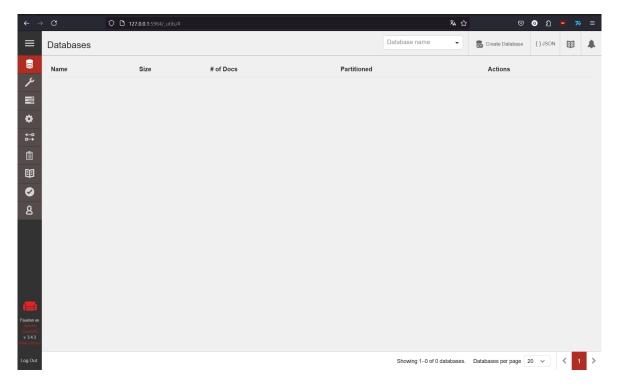
1. Ejecutar el instalador y aceptar los términos.



2. Configurar una cuenta de administrador (usuario: admin, contraseña: 123).



3. Completar la instalación y verificar el acceso a Fauxton en http://localhost: 5984/_utils/.



Resultado: CouchDB se ejecutó correctamente, con Fauxton mostrando una interfaz vacía lista para crear bases de datos.

2.2. Configuración de puerto

Si el puerto predeterminado (5984) está ocupado, se puede cambiar:

1. Detener el servicio:

```
net stop CouchDB
```

2. Modificar C:\Program Files\Apache CouchDB\etc\couchdb\local.ini, ajustando port
= 9548 en [chttpd].

```
[chttpd]
port = 9548 ;o el puerto que quieras
:bind address = 127.0.0.1
```

3. Reiniciar el servicio:

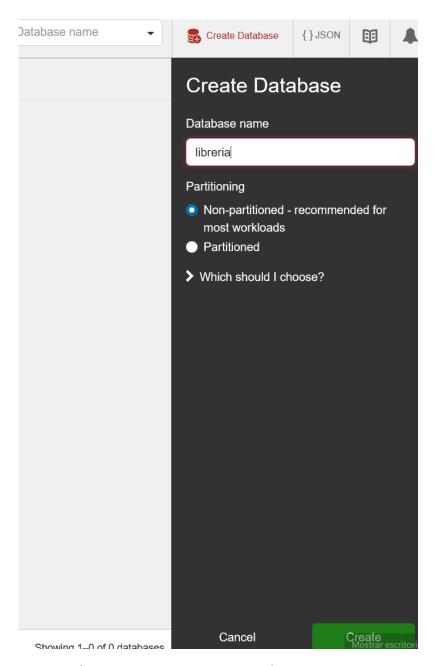
```
net start CouchDB
```

Resultado: Fauxton fue accesible en http://localhost:9548/_utils/, demostrando flexibilidad en la configuración.

2.3. Creación de base de datos y documentos

La creación de una base de datos y documentos en CouchDB es se realizó utilizando Fauxton y la CLI, demostrando la interoperabilidad de las interfaces.

Creación de base de datos: En Fauxton, se creó una base de datos llamada libreria accediendo a http://localhost:5984/_utils/, iniciando sesión con las credenciales de administrador, y seleccionando «Create Database» en la interfaz gráfica. Fauxton valida el nombre de la base de datos (sin caracteres especiales) y la crea mediante una solicitud HTTP PUT interna al endpoint /{db}. Este proceso es intuitivo, ya que la interfaz guía al usuario a través de un formulario simple.



Inserción en Fauxton: Se insertó un documento manualmente en la base de datos libreria utilizando el botón «Create Document». Fauxton proporciona un editor JSON donde se ingresa el documento, que debe cumplir con el formato JSON válido y contener un campo _id único. El documento insertado fue:

```
{
  "_id": "libro_001",
  "titulo": "Cien años de soledad",
  "autor_id": "autor_001",
  "editorial_id": "editorial_001",
  "genero": "Realismo mágico",
  "anio_publicacion": 1967,
  "precio": 19.99
}
```

Sintaxis y propósito:

- _id: Identificador único del documento, obligatorio para todas las operaciones CRUD.
- titulo, autor_id, editorial_id, genero, anio_publicacion, precio: Campos personalizados que definen los atributos del libro, flexibles según las necesidades de la aplicación.
- **Funcionamiento**: Al guardar, Fauxton envía una solicitud HTTP PUT al endpoint /libreria/libro_001, almacenando el documento en la base de datos. La interfaz valida el JSON y muestra el documento creado en la vista «Table».



Inserción vía CLI: Un segundo documento se creó utilizando un comando curl que interactúa directamente con la API RESTful:

```
curl -X PUT http://admin:123@127.0.0.1:5984/libreria/libro_002 -H "Content-Type:
application/json" -d "{\"_id\": \"libro_002\", \"titulo\": \"El amor en los
tiempos del cólera\", \"autor_id\": \"autor_001\", \"editorial_id\":
"editorial_002\", \"genero\": \"Romance\", \"anio_publicacion\": 1985,
\"precio\": 15.50}"
```

Desglose de la sintaxis:

- curl: Herramienta de línea de comandos para enviar solicitudes HTTP.
- -X PUT: Especifica el método HTTP PUT, usado para crear o actualizar un documento en el endpoint especificado.
- ► http://admin:123@127.0.0.1:5984/libreria/libro_002: URL del endpoint, donde:
 - admin: 123: Credenciales de autenticación (usuario:contraseña).
 - 127.0.0.1:5984: Dirección y puerto del servidor CouchDB.
 - libreria: Nombre de la base de datos.
 - libro 002: ID del documento a crear.
- ► -H "Content-Type: application/json": Header que indica que el cuerpo de la solicitud es JSON.
- -d "{...}": Cuerpo de la solicitud, un objeto JSON con los mismos campos que el documento de Fauxton.
- **Funcionamiento**: El servidor valida el JSON, asigna un campo _rev (versión del documento) y responde con un mensaje {"ok":true, "id":"libro_002", "rev":"1-..."}, indicando éxito. La CLI permite automatización y scripting, a diferencia de la interfaz gráfica.



Resultado: Ambos documentos fueron visibles en Fauxton, confirmando que las operaciones en CLI y Fauxton son equivalentes, ya que ambas interactúan con la misma API RESTful subyacente.

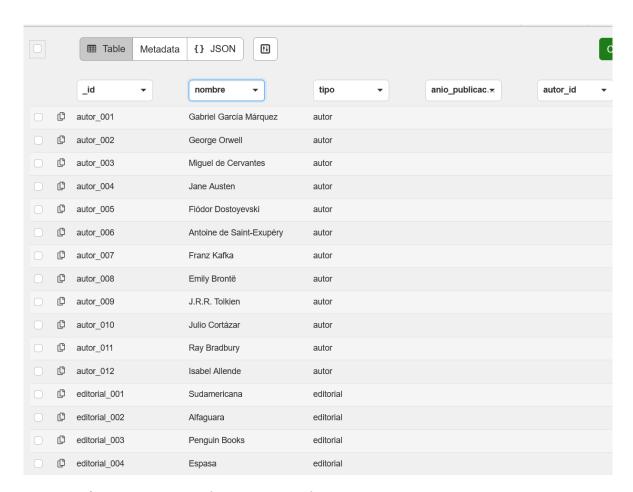
2.4. Importación masiva

La importación masiva permite insertar múltiples documentos en una sola solicitud, optimizando el rendimiento. Se importaron documentos (libros, autores, editoriales) desde un archivo libreria.json usando el endpoint _bulk_docs:

```
curl -X POST -H "Content-Type: application/json" --data @libreria.json http://
admin:123@127.0.0.1:5984/libreria/_bulk_docs
```

• Desglose de la sintaxis:

- ► -X POST: Método HTTP POST, usado para enviar datos al endpoint bulk docs.
- ► -H "Content-Type: application/json": Indica que el cuerpo es JSON.
- --data @libreria.json: Lee el contenido del archivo libreria.json, que debe contener un objeto JSON con una clave "docs" y un arreglo de documentos.
- http://admin:123@127.0.0.1:5984/libreria/_bulk_docs: Endpoint que procesa múltiples documentos en una transacción atómica.
- **Funcionamiento**: CouchDB valida cada documento en el arreglo, asigna un _rev a los documentos creados y responde con un arreglo de resultados (por ejemplo, [{"ok":true, "id":"autor_001", "rev":"1-..."}, ...]). Esto reduce la sobrecarga de múltiples solicitudes individuales.



Verificación: La importación se confirmó con:

curl -X GET http://admin:123@127.0.0.1:5984/libreria/_all_docs

• **Desglose**: Este comando envía una solicitud GET al endpoint _all_docs, que retorna un objeto JSON con una lista de todos los documentos en la base de datos, incluyendo sus _id y _rev. La respuesta confirma que los documentos de libreria.json se añadieron correctamente.

Resultado: La operación _bulk_docs permitió la inserción eficiente de múltiples documentos, escalable para grandes volúmenes de datos.

2.5. Consultas

CouchDB utiliza **Mango Queries** para realizar búsquedas declarativas, similares a MongoDB, ejecutadas a través de la CLI o Fauxton. Se implementaron dos consultas para demostrar su funcionalidad.

Consulta CLI: Obtener el nombre y nacionalidad del autor con _id «autor 001»:

```
curl -X POST http://admin:123@127.0.0.1:5984/libreria/_find -H "Content-Type:
application/json" -d "{\"selector\": {\"_id\": \"autor_001\"}, \"fields\":
[\"nombre\", \"nacionalidad\"]}"
```

- Desglose de la sintaxis:
 - ► -X POST y -H "Content-Type: application/json": Indican una solicitud POST con un cuerpo JSON.
 - ► http://admin:123@127.0.0.1:5984/libreria/_find: Endpoint para Mango Queries.
 - ► -d "{...}": Cuerpo JSON con dos claves:

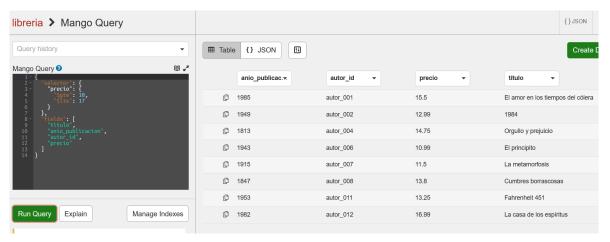
- "selector": Define las condiciones de búsqueda. Aquí, {"_id": "autor_001"} busca documentos con id exacto.
- "fields": Lista de campos a retornar (["nombre", "nacionalidad"]), optimizando la respuesta.
- **Funcionamiento**: CouchDB evalúa el selector contra los documentos de la base de datos. Si un índice existe para _id, la consulta es eficiente; de lo contrario, realiza una búsqueda completa. La respuesta es un objeto JSON con una clave "docs" que contiene los documentos coincidentes, por ejemplo:

```
{"docs": [{"nombre": "Gabriel García Márquez", "nacionalidad": "Colombiana"}]}
```

Consulta Fauxton (Mango Query): Obtener título, año de publicación, autor y precio de libros con precio entre 10 y 17:

```
{
    "selector": {
        "precio": {
            "$gte": 10,
            "$lte": 17
        }
    },
    "fields": ["titulo", "anio_publicacion", "autor_id", "precio"]
}
```

- Desglose de la sintaxis:
 - ► "selector": Condiciones de búsqueda. El objeto "precio" usa operadores:
 - \$gte: Mayor o igual a 10.
 - \$lte: Menor o igual a 17.
 - "fields": Campos a incluir en la respuesta, reduciendo el tamaño del resultado.
- Funcionamiento: En Fauxton, el usuario ingresa esta consulta en el «Mango Query Editor». Fauxton traduce el JSON a una solicitud HTTP POST al endpoint _find, enviándola al servidor. CouchDB evalúa la condición, filtra los documentos y retorna solo los campos especificados. La interfaz muestra los resultados en una tabla, facilitando la visualización.



• **Por qué funciona**: Las Mango Queries son eficientes cuando se usan índices predefinidos para los campos consultados (por ejemplo, precio). Sin índices, CouchDB escanea todos los documentos, lo que puede ser lento. La sintaxis declarativa permite consultas complejas (por ejemplo, \$or, \$and) sin necesidad de vistas MapReduce, simplificando el desarrollo.

Resultado: Ambas consultas retornaron resultados precisos, demostrando la flexibilidad de Mango Queries. La CLI es ideal para automatización, mientras que Fauxton ofrece una experiencia visual para usuarios no técnicos.

3. Conclusiones

CouchDB se consolida como una base de datos NoSQL robusta para entornos distribuidos, gracias a su arquitectura orientada a documentos y su enfoque en disponibilidad y tolerancia a particiones. Su API RESTful y herramientas como Fauxton y CLI facilitan la gestión de datos, mientras que las Mango Queries ofrecen un equilibrio entre simplicidad y potencia para búsquedas. Sin embargo, su modelo de consistencia eventual plantea desafíos en aplicaciones que requieren datos siempre coherentes, como sistemas financieros. Comparado con MongoDB, CouchDB destaca en sincronización offline, pero puede ser menos eficiente en consultas complejas. Frente a Cassandra, ofrece una interfaz más accesible, aunque con menor escalabilidad en escritura.

La implementación práctica reveló la facilidad de instalación y uso de CouchDB, aunque la configuración de índices para consultas optimizadas requiere planificación. Su diseño es ideal para aplicaciones modernas como IoT, CMS y aplicaciones móviles, pero su adopción depende de la comprensión de sus limitaciones. Este análisis subraya la importancia de seleccionar bases de datos según las necesidades específicas del proyecto, equilibrando factores como disponibilidad, consistencia y rendimiento.

4. Bibliografía

- Apache CouchDB. (s.f.). Documentación oficial. Recuperado de http://docs. couchdb.org/
- Apache Software Foundation. (s.f.). CouchDB. Recuperado de https://couchdb. apache.org/
- Documentación de Fauxton. Recuperado de https://docs.couchdb.org/en/stable/ fauxton/
- Instalador CouchDB 3.4.3. Recuperado de https://couchdb.apache.org/#download
- IBM. (2019). What Is CouchDB? Recuperado de https://www.ibm.com/topics/ couchdb
- Cloudant. (s.f.). CouchDB and Cloudant Documentation. Recuperado de https://cloud.ibm.com/docs/Cloudant?topic=Cloudant-couchdb
- CouchDB Blog. (s.f.). Time to Relax! Recuperado de https://blog.couchdb.org/
- GitHub Apache CouchDB. (s.f.). Repositorio oficial de CouchDB. Recuperado de https://github.com/apache/couchdb