



Fundamentos de MongoDB Enterprise 4.4.1

Instructor: Carlos Carreño
Email: ccarrenovi@gmail.com

Modulo 5 Escalabilidad Horizontal

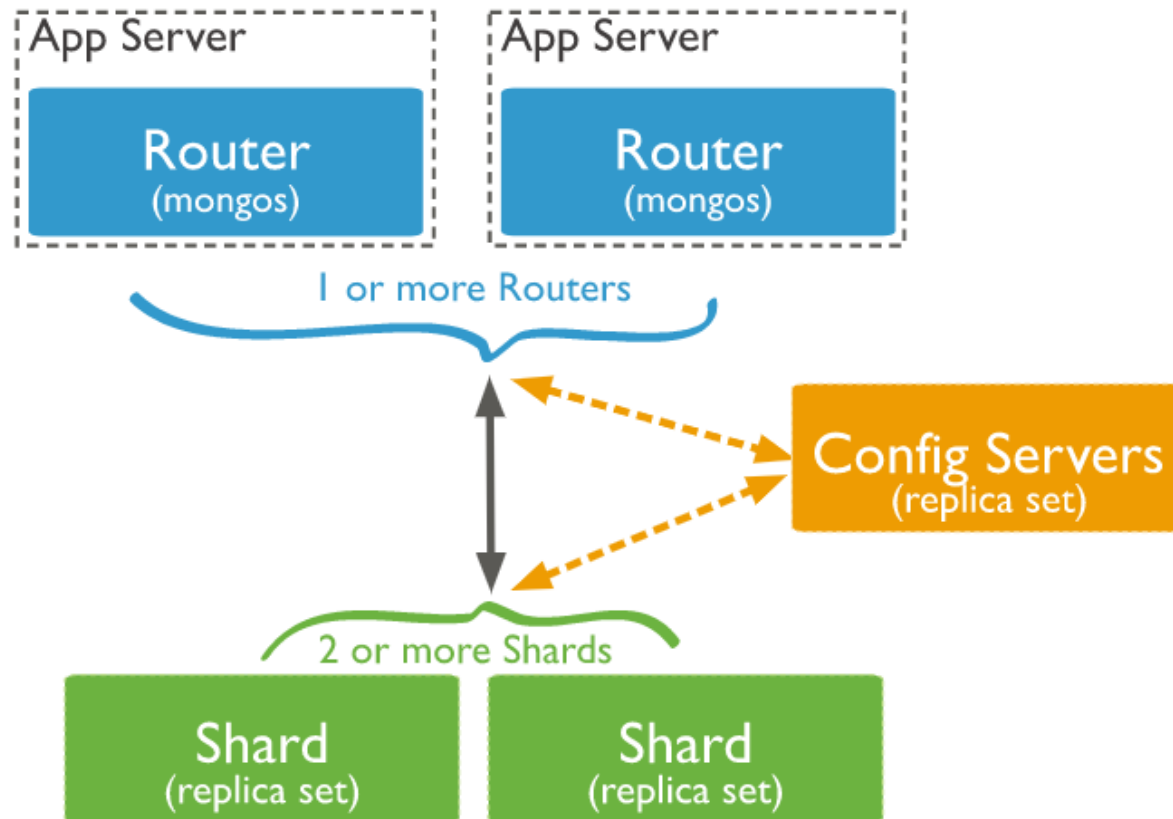
- Sharding
- Cluster
- Balanceo

Sharding

- Sharding es un método para distribuir datos a través de múltiples máquinas.
- El sharding se utiliza con grandes cantidades de datos y alto throughput

Cluster

- En un Cluster de MongoDB son tres los componentes principales:



Componentes del Cluster

- **Shard**: cada shard contiene un subconjunto de los datos fragmentados. Cada shard se puede implementar como un conjunto de réplicas.
- **mongos**: el mongos actúa como un enrutador de consultas, proporcionando una interfaz entre las aplicaciones del cliente y el clúster fragmentado.
- **servidores de configuración**: los servidores de configuración almacenan metadatos y ajustes de configuración para el clúster

Shard keys

- MongoDB usa la **shard key** para distribuir los documentos de la colección entre shards. La shard key **consiste en un campo o campos** que existen en cada documento en la colección de destino.
- Usted elige la shard key cuando fragmentas una colección. **La elección de la shard key no se puede cambiar después de fragmentar** la colección. Una colección fragmentada solo puede tener una shard key.
- Para fragmentar una colección no vacía, **la colección debe tener un índice** que comience con la shard key.

Chunks

- Las particiones MongoDB fragmentan los datos en chunks.
- Cada chunk tiene un rango superior e inferior inclusivo basado en la shard key.
- Cada chunk por defecto es de 64 MB

Balanceo

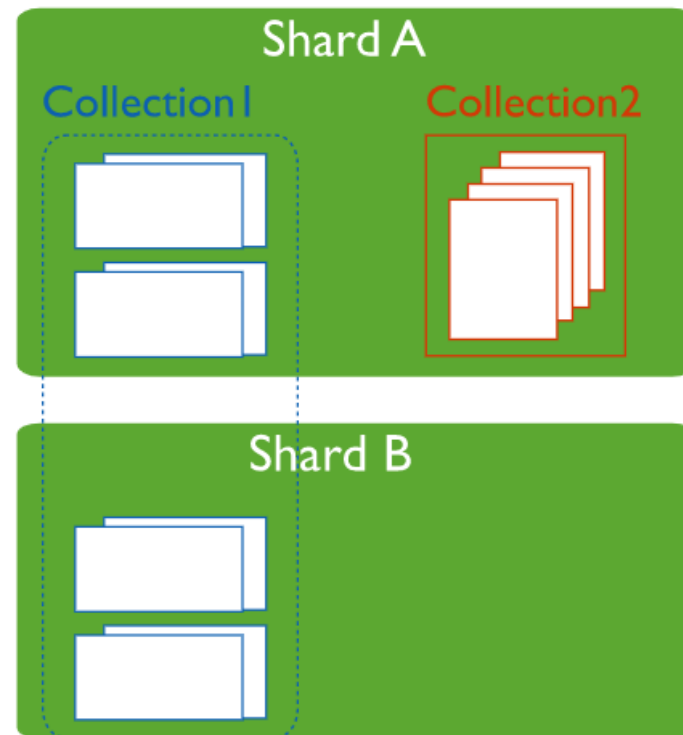
- En un intento por lograr un balanceo y una distribución uniforme de los **chunks** en todos los **shard** del clúster, se ejecuta un balanceador en segundo plano para migrar chunks en los shards del Cluster.

Ventajas del Sharding

- Mejora el desempeño en operaciones de lectura y escritura
- Implementa una mejor gestión de la capacidad de almacenamiento
- Proporciona una solución de alta disponibilidad

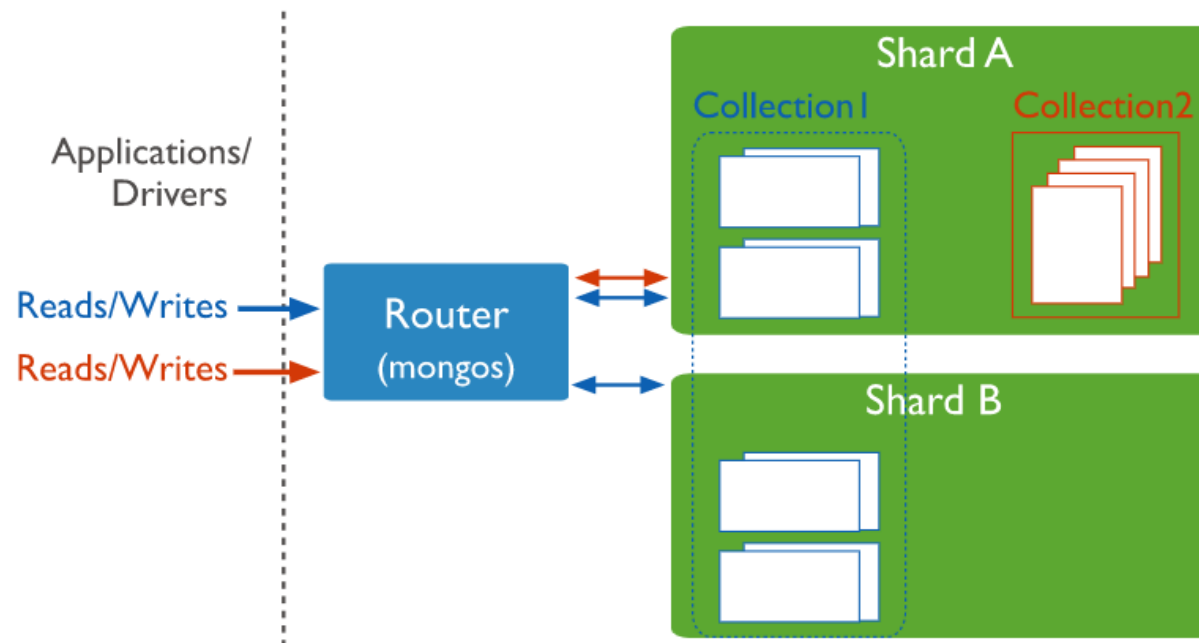
Sharded and Non- Sharded collection

- Una base de datos en mongo puede trabajar en modo mixto puede tener algunas colecciones en sharded y otras en modo no sharded



Conexión al sharding cluster

- El cliente debería conectarse a un mongos router para poder interactuar con el sharding cluster, tanto con colecciones sharded como con no sharded.

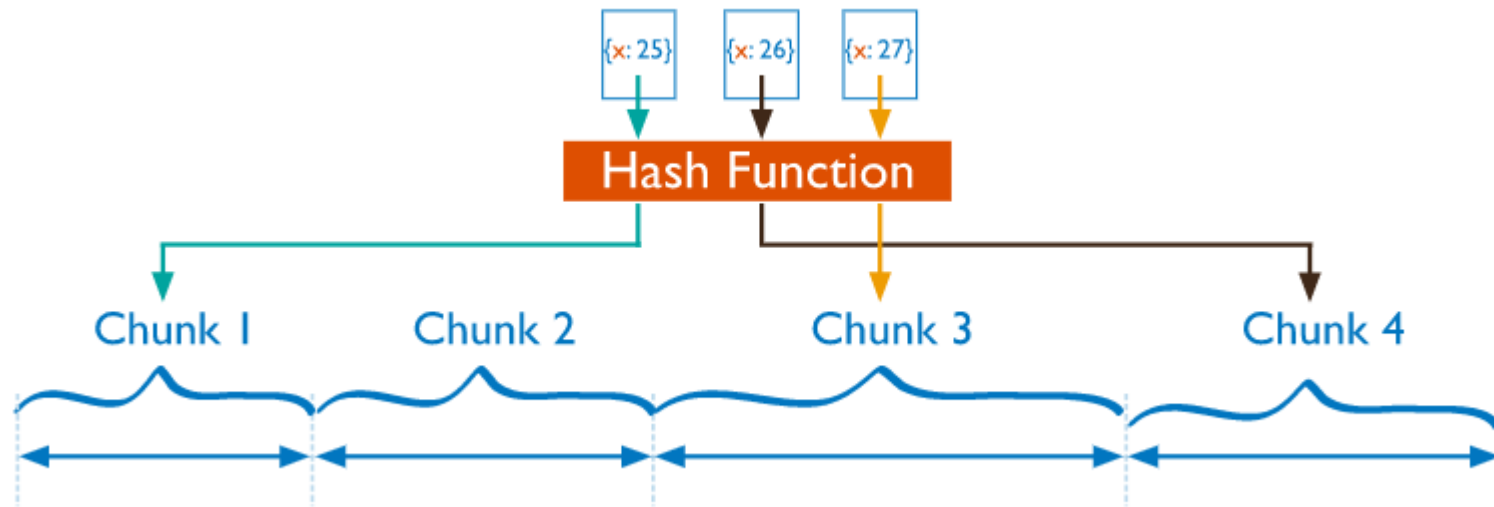


Sharding Strategy

- Mongo soporta dos estrategias para distribuir los datos en el cluster:
 - Hashed Sharding
 - Ranged sharding

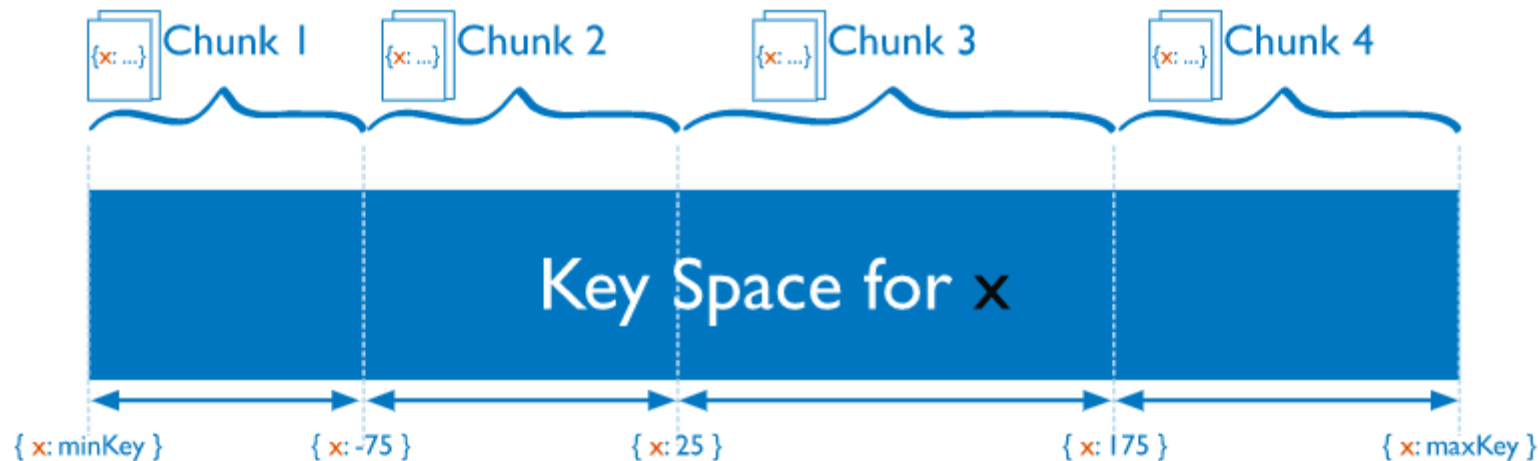
Hashed Sharding

- Hashed Sharding implica calcular un hash del valor del campo shard key. A cada chunk se le asigna un rango basado en los valores del shard key hash.



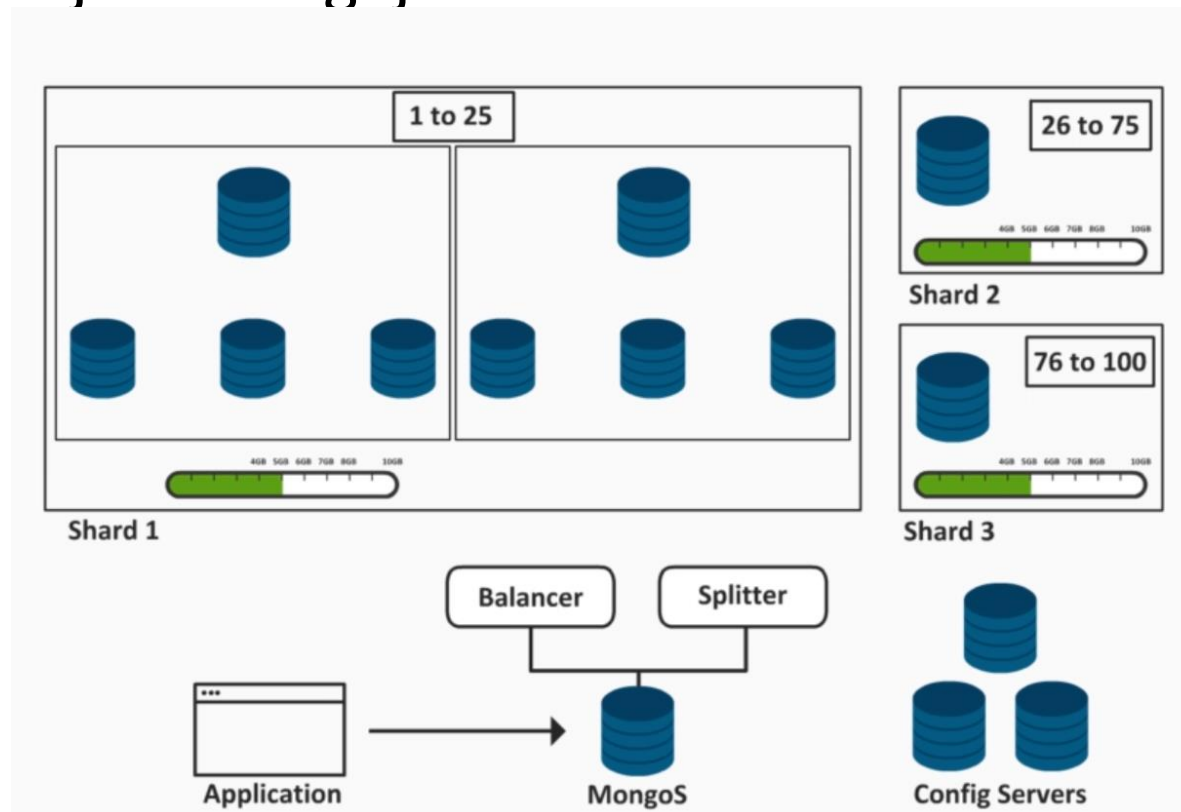
Ranged sharding

- Ranged sharding divide los datos en rangos basados en el valor del shard key. Cada chunk es asignado a un rango basado en el valor del shard key.



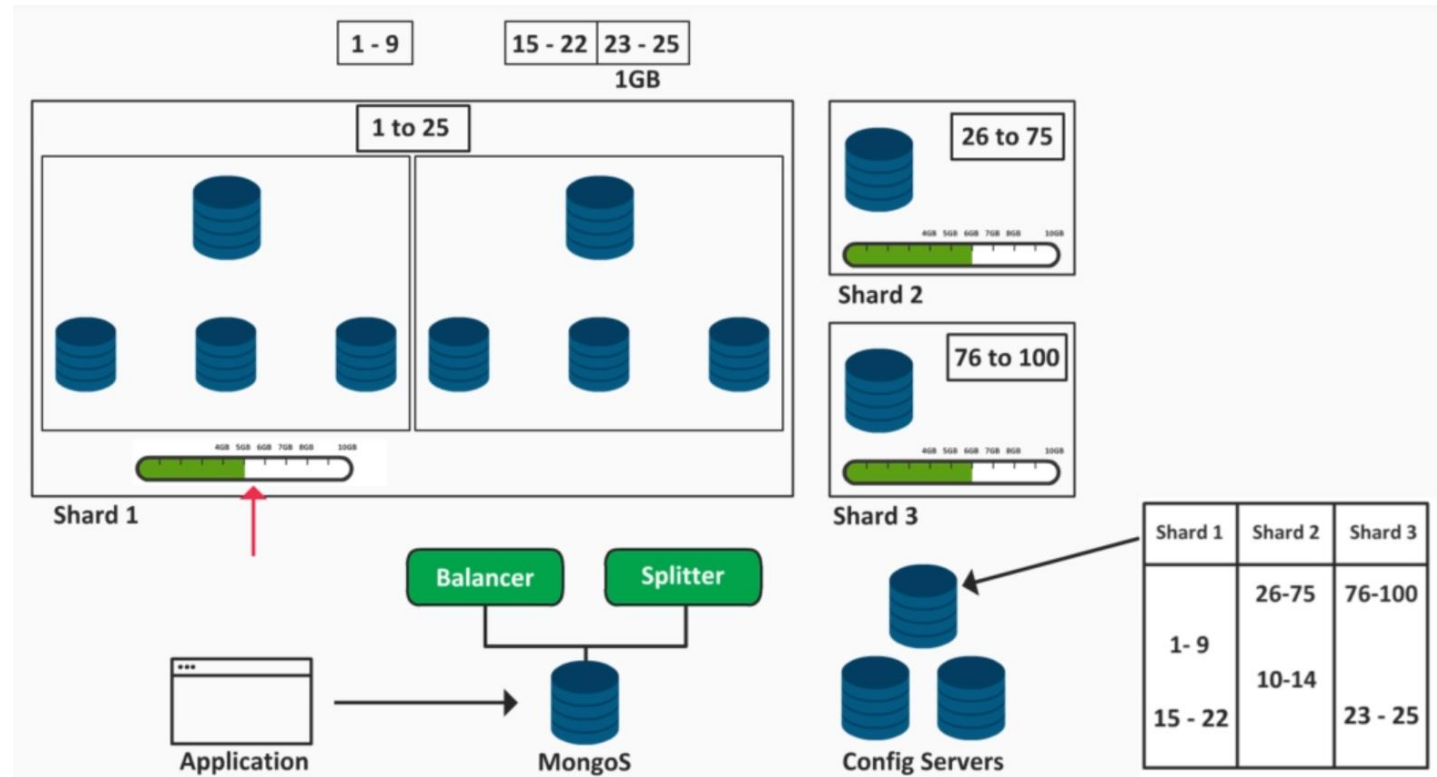
Sharding Workflow

- Flujo de trabajo de una solución de sharding



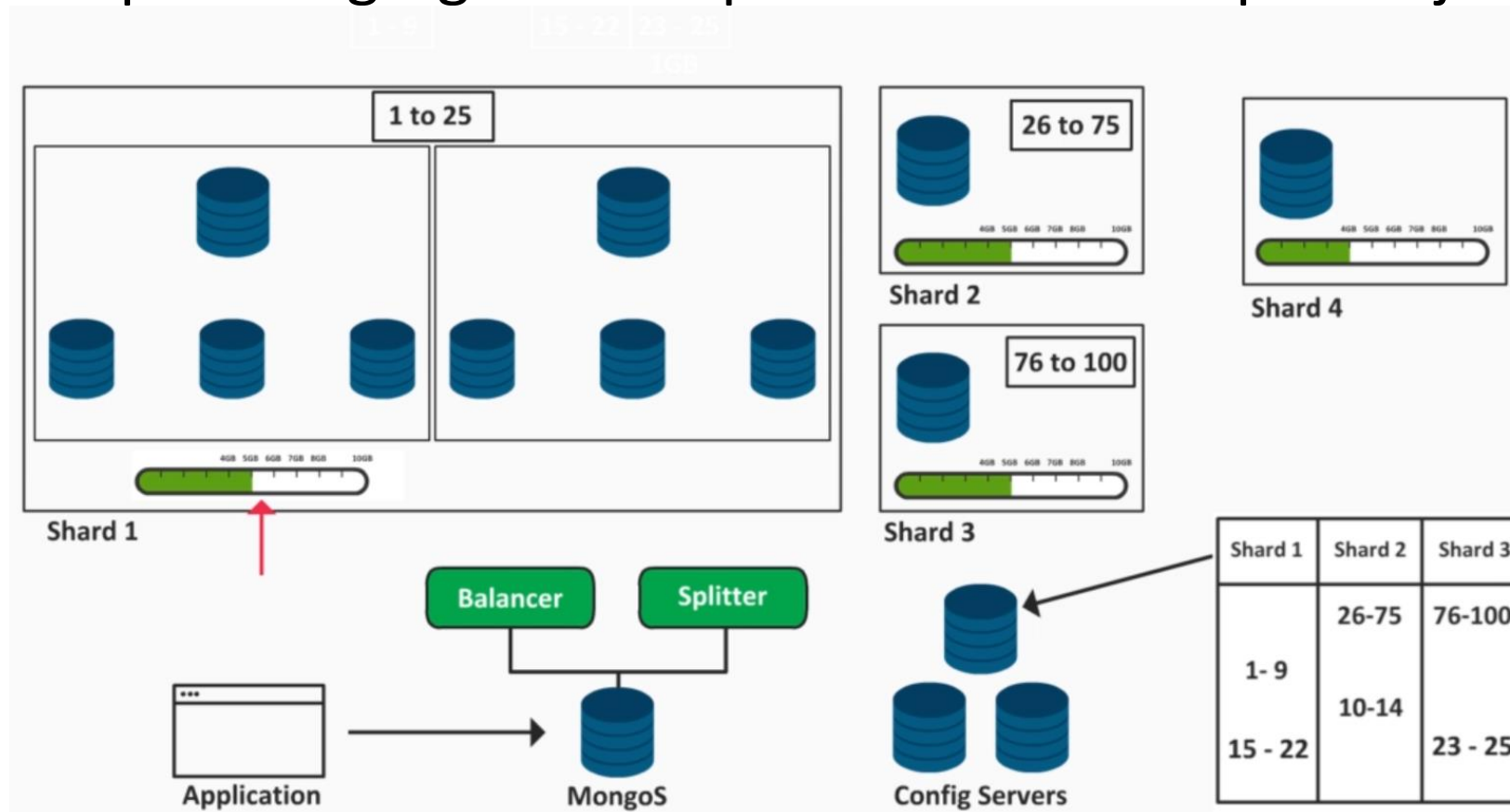
Sharding Workflow & Balancer

- Si algún shard está con sobre carga esta se puede balancear entre los Sharp disponibles



Sharding Workflow & Add replica set

- Se puede agregar una replica set adicional par mejorar el desempeño



Production configuration

- En un clúster de producción, asegúrese de que los datos sean redundantes y que sus sistemas estén altamente disponibles.
Considere lo siguiente para una implementación de un shard clúster:
 - Implementar servidores de configuración como un conjunto de réplicas de 3 miembros
 - Implemente cada shard como un conjunto de réplicas de 3 miembros
 - Implemente uno o más enrutadores mongos

Referencias

- <https://docs.mongodb.com/manual/sharding/>
- https://www.youtube.com/watch?v=ifZHWvS_Gig&t=3433s