

# **Sharding in 20 minutes**

## **Why;Who;When;Where;**

David Murphy , Mongo Master  
Lead DBA, ObjectRocket  
@dmurphy\_data @objectrocket

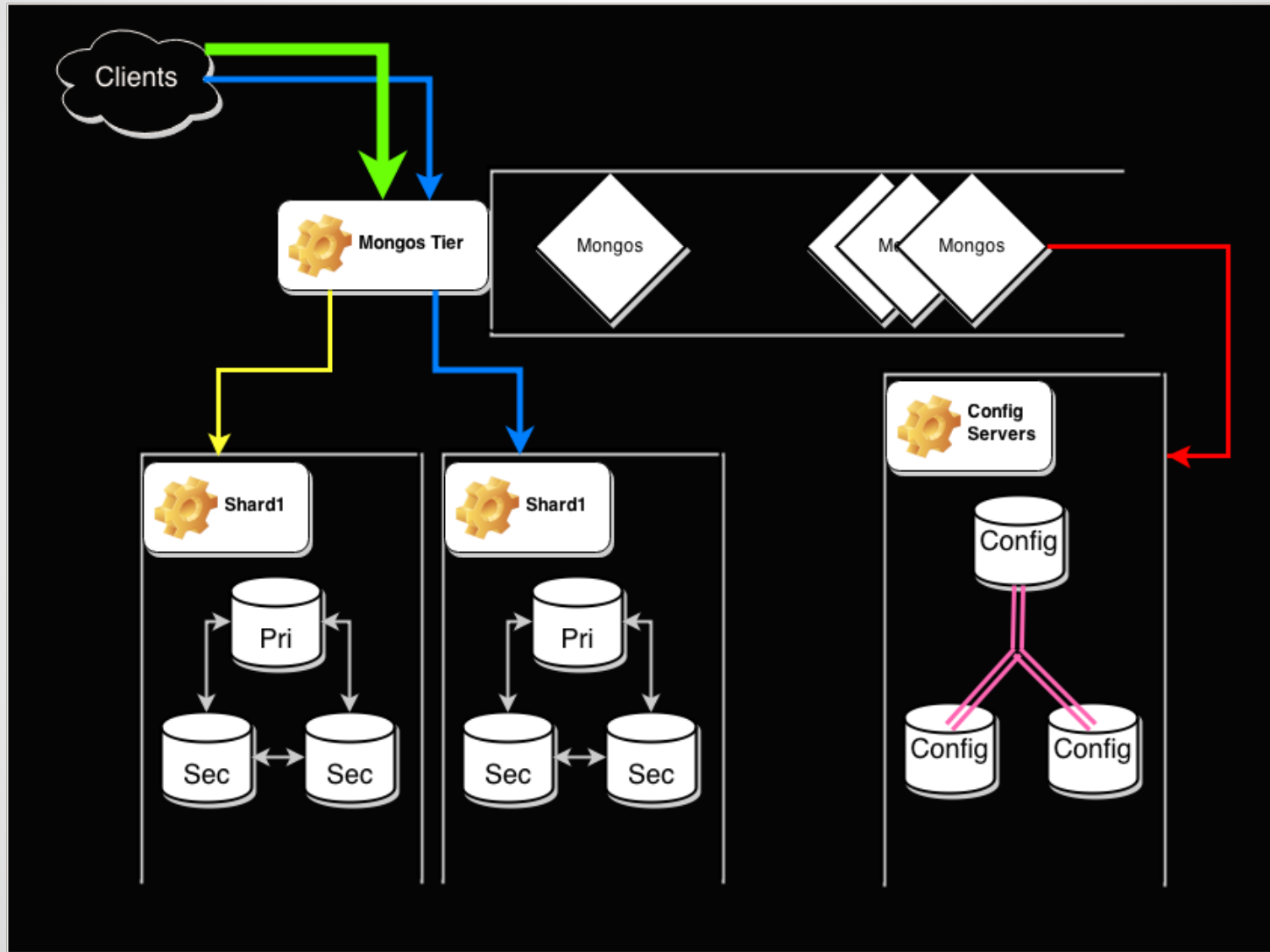


# Background

- **16 yrs in databases, development, & system engineering**
- **Lead DBA @ ObjectRocket**
- **Mongo Master with a focus on sharding, chunks, and scaling mongo beyond normal means.**



# What does a sharded cluster look like?



# Why;Who;When;Where;

*i) Why do we shard?*

*ii) Who should I shard?*

*iii) When do we shard?*

*iv) Where is my shard key*



# Why do we shard?

- **Scaling out write locks**



# Why do we shard?

- **Scaling out write locks**
- **Small dataset to search per node**



# **Why do we shard?**

- **Scaling out write locks**
- **Small dataset to search per node**
- **Getting more connections to the data**



# Why do we shard?

- **Scaling out write locks**
- **Small dataset to search per node**
- **Getting more connections to the data**
- **More smaller node vs Scaling up to expensive nodes**





# Who should I shard?

- **Biggest Collections by Size**



# Who should I shard?

- **Biggest Collections by Size**
- **Busiest Collection by changes**



# Who should I shard?

- **Biggest Collections by Size**
- **Busiest Collection by changes**
- **Groupings of data (example):**
  - **State/Country**
  - **UserID**
  - **Company**
  - **Category**



# When do I shard?

**ALWAYS as early as possible!**



# **When do I shard?**

**ALWAYS as early as possible!**

**Reasons:**

- **Not all commands work**



# **When do I shard?**

**ALWAYS as early as possible!**

**Reasons:**

- **Not all commands work**
- **Future Proof - No recoding**



# **When do I shard?**

**ALWAYS as early as possible!**

**Reasons:**

- **Not all commands work**
- **Future Proof - No recoding**
- **Adding index once your live can take time you don't have!**



# **Where *(and what)* is my shard key?**

**You have to pick your own :/**

**But there are some quick hints...**





# Where (*and what*) is my shard key?

## Sharding Quick Hints:

- **Hashed Shard keys**

Great for even disk usage

Uses Scatter-Gathers == More Conns

Dates, Increasing IDs , and text are great here



# Where *(and what)* is my shard key?

## Sharding Quick Hints:

- **Hashed Shard keys**

Great for even disk usage

Uses Scatter-Gathers == More Conns

Dates, Increasing IDs , and text are great here

- **Non-Hashed Keys**

Use profiler\_level:2 & review ALL operations

Things you wont change only

No Dates

No Increasing ID numbers

No Text



# **Why mongo sharding/balancing**

## **Modulus Sharding with MySQL:**

**Hard to rebalance online**

**Requires application coding to support**

## **Ring Topologies like Cassandra -**

**Cant change schema online**

**Hard to rebalance online**



# Further Reading

## **Presentations:**

Kenny Gorman Sharding - [bit.ly/1oXYDfm](http://bit.ly/1oXYDfm)

David Murphy - Adv Sharding for Operations - [bit.ly/1oXYDfm](http://bit.ly/1oXYDfm)

Other Sharding MongoDB Links - [bit.ly/ZTtDI1](http://bit.ly/ZTtDI1)

Picking a shard key (manual) - <http://bit.ly/1ozuzMH>

Choosing a shard key - <http://slidesha.re/1nBnGtq>



# Contact

@dmurphy\_data

@objectrocket

david@objectrocket.com

<https://www.objectrocket.com>

**WE ARE HIRING! (DBA,DEVOPS, and more)**

**<https://www.objectrocket.com/careers>**

