Stories from the DBA trenches
Who am I?

- Silvia Botros/dbsmasher
- Principal Engineer @ SendGrid
- ~ 7 years
- Accidentally a DBA
- Initially a software engineer
Who is this talk for

You work at a rapidly growing company
Who is this talk for

The ops team is comprised of a single digit number of people
Who is this talk for

There is no dedicated DBA on staff….yet...ever?
Who is this talk for

Revenue multiplying year over year?
Who is this talk for

Marketing and sales teams ‘crushing it’?
Who is this talk for

How does engineering feel about stability though?
I KNOW THAT FEEL BRO
What we will talk about

- Database basics
- Architecture best practices
- When/Why do you need to hire DBRE
Schema design tips

- Primary keys are your friend
- Auto increment keys
  - Know your tradeoffs
- Always make your auto increment keys unsigned
Schema design tips

- Set sane default values. Null is *the worst*
- Timestamps are your friend
- Do not sweat the indexes
  - Iterate later using logs/metrics
- Know the MySQL types. AKA: The 6-way boolean trap
  - Column names matter
- One column/a dozen use cases
Metrics

- Metrics you should always track
  - Threads running/Threads connected
Metrics

- Disk IOPS (yes, even SSDs)
A big SSD array in tears
Metrics

- Disk space growth rate/disk space alerting threshold
- Instate data removal/age out where you can
- Your business critical database is not a data warehouse
- But…Aurora
  - Aurora has a maximum allowed size
  - Data scanning is a cost (Spoiler: Biggest cost)
Data rotation management

- Date based partitions
- pdb-parted (Credit: Palomino db team)
- Using sensu for DBA tasks
Architecture

- Beware of ORMs
- ProxySQL
  - Smarter read load management
  - Caching
  - Configuration changes with 0 downtime
  - Load shedding
  - Connection mirroring
Architecture

- Writes can fail. Design with that in mind
- Degraded mode, read only mode, serving stale data…are all things
- Shard your abstraction layer just as you do your clusters
- Lag will happen. Design with that in mind
- …no that doesn’t mean send all your read to the primary
Shard and shard again

If your product succeeds, you will have more than one DB cluster
Shortcuts you can take

- Agonizing over every column type *
- Tuning every config possible
- Indexing all the things
- Audit logging
Shortcuts you should *not* take

● The database timezone
● Your charsets
● Not tracking your auto increment key space
● Backup testing

“Do you test your backups? If you don’t, you have Schrödinger’s backups. Or Schrödinger’s cat ate your backups.. 😐” ~ accidental dba facts

3:09 PM - 22 Oct 2017
Shortcuts you should *not* take

- Not tuning certain configs
- Turn off the query cache….really
- Set your buffer pool size. Beware of connection overhead
- Set your mysql process file limits
- Set your max_connection
Lies of ‘The Cloud’

- Aurora’s storage tricks
- Cross region replication ≠ Cross AZ replication
- Parameter groups for the cluster vs instances
- Cloudwatch metrics
“We need a DBA”

- There are good reasons and bad reasons.
First, the bad reasons

- To run schema changes
- Get familiar with `pt-online-schema-change`, `gh-ost`, and `shift`
- Be ‘on call’ for databases
- ‘No one else has time to take care of this’
Good reasons

● Help set standards/best practices on how to choose/use datastore for your engineers
● Be a subject matter expert escalation point for your engineers when needed
● Your engineers are new to scaling datastores at the growth rate you are seeing
● You are solving really interesting, new, problems
Who to look for

- An enabler
- A teacher
- NOT a gardener for your walled garden of artisanal databases
Reading material

- The DBRE book - Charity Majors and Laine Campbell
- High Performance MySQL - Baron Schwartz
- “The accidental DBA” blog post by Charity Majors
- The SRE workbook
- State of DevOps Report, 2018
You got this