DEPARTMENT OF COMMERCE

WORSE CASE SCENARIO IN THE DATABASE

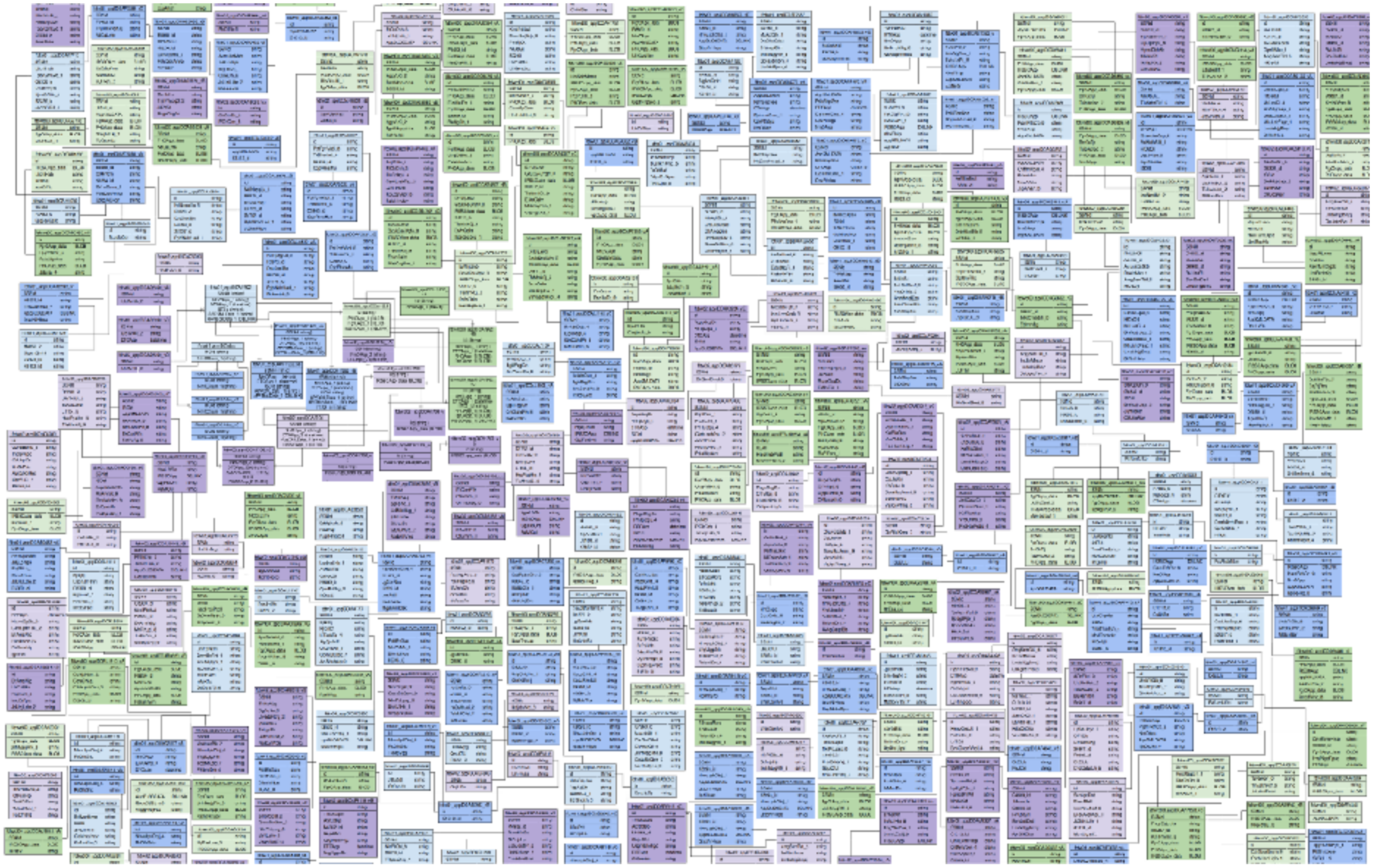
GIVE ME YOUR WORST

WORSE CASE SCENARIO IN THE DATABASE

WE'RE HAVING SERIOUS PERFORMANCE ISSUES. CAN YOU LOOK AT OUR DB?

Important Client

WORSE CASE SCENARIO IN THE DATABASE



@bellmar

WORSE CASE SCENARIO IN THE DATABASE

@bellmar

IT'S NOT THE NUMBER OF TABLES

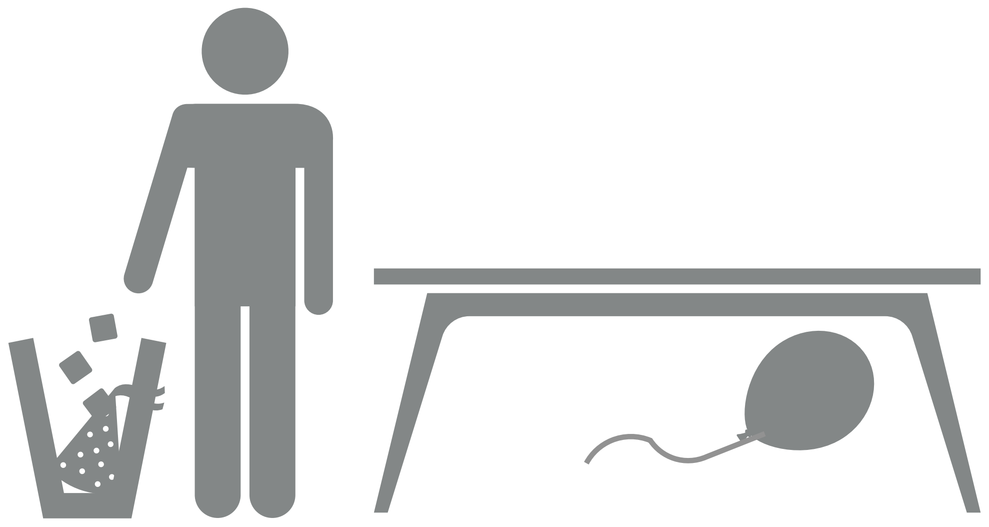
IT'S WHY YOU NEED SO MANY

TECHNICAL DEBT

- "Technical debt" is not limited to application code
- How often do we change things?
    - Application code: Often
    - Infrastructure: Rarely
    - Data model and schemas: Variable

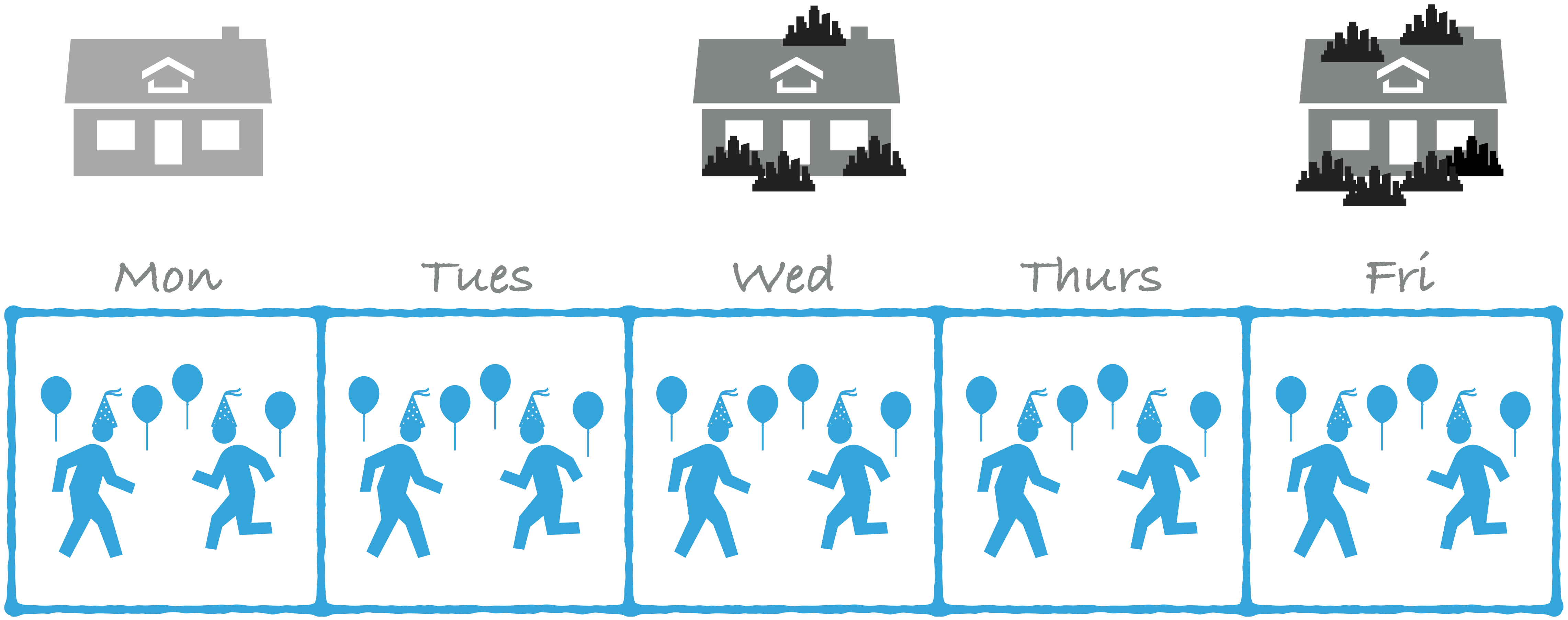
TECHNICAL DEBT

New Feature


Refactoring

WORSE CASE SCENARIO IN THE DATABASE

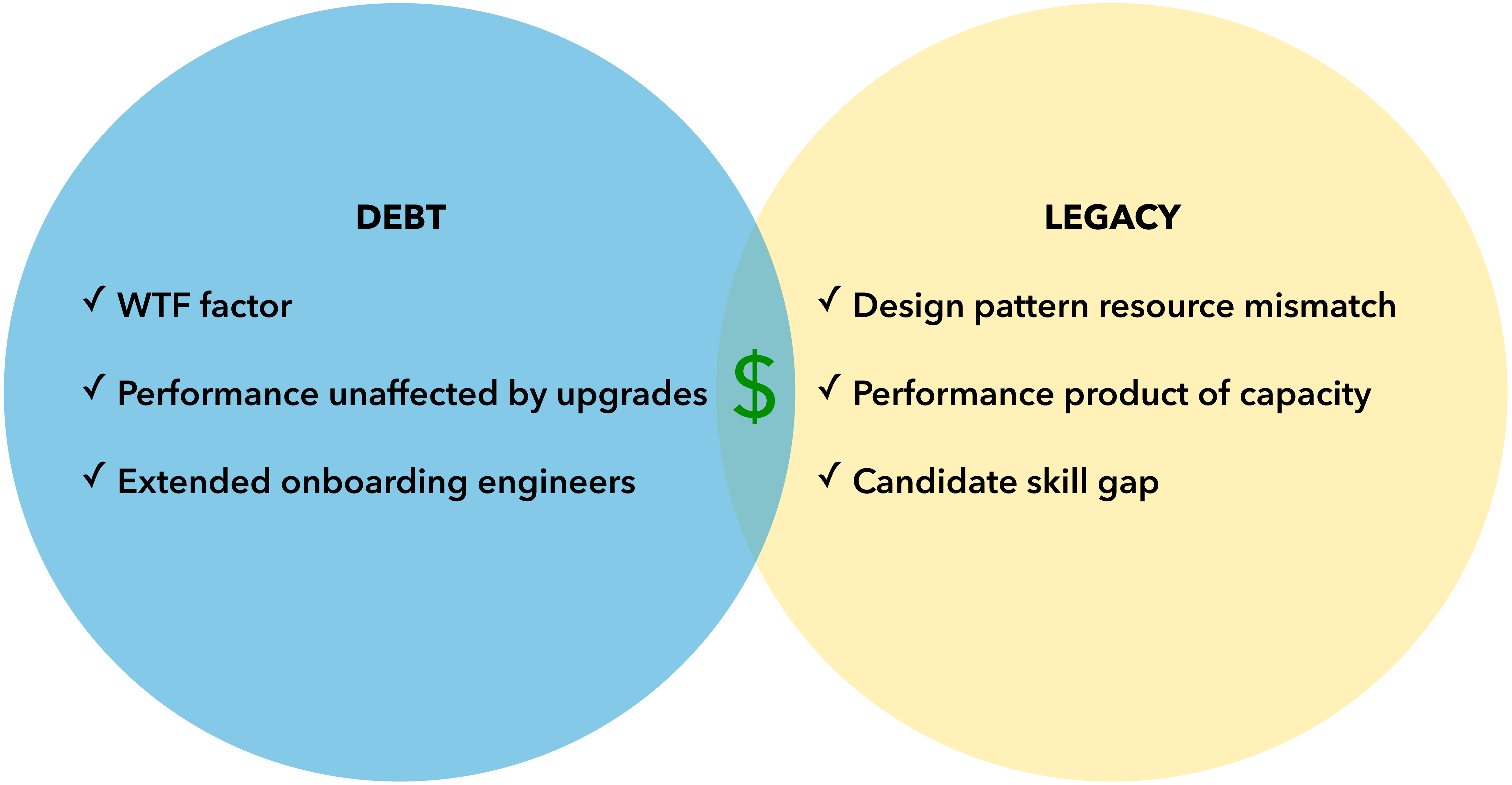
TECHNICAL DEBT



@bellmar

WORSE CASE SCENARIO IN THE DATABASE

DEBT -VS- LEGACY

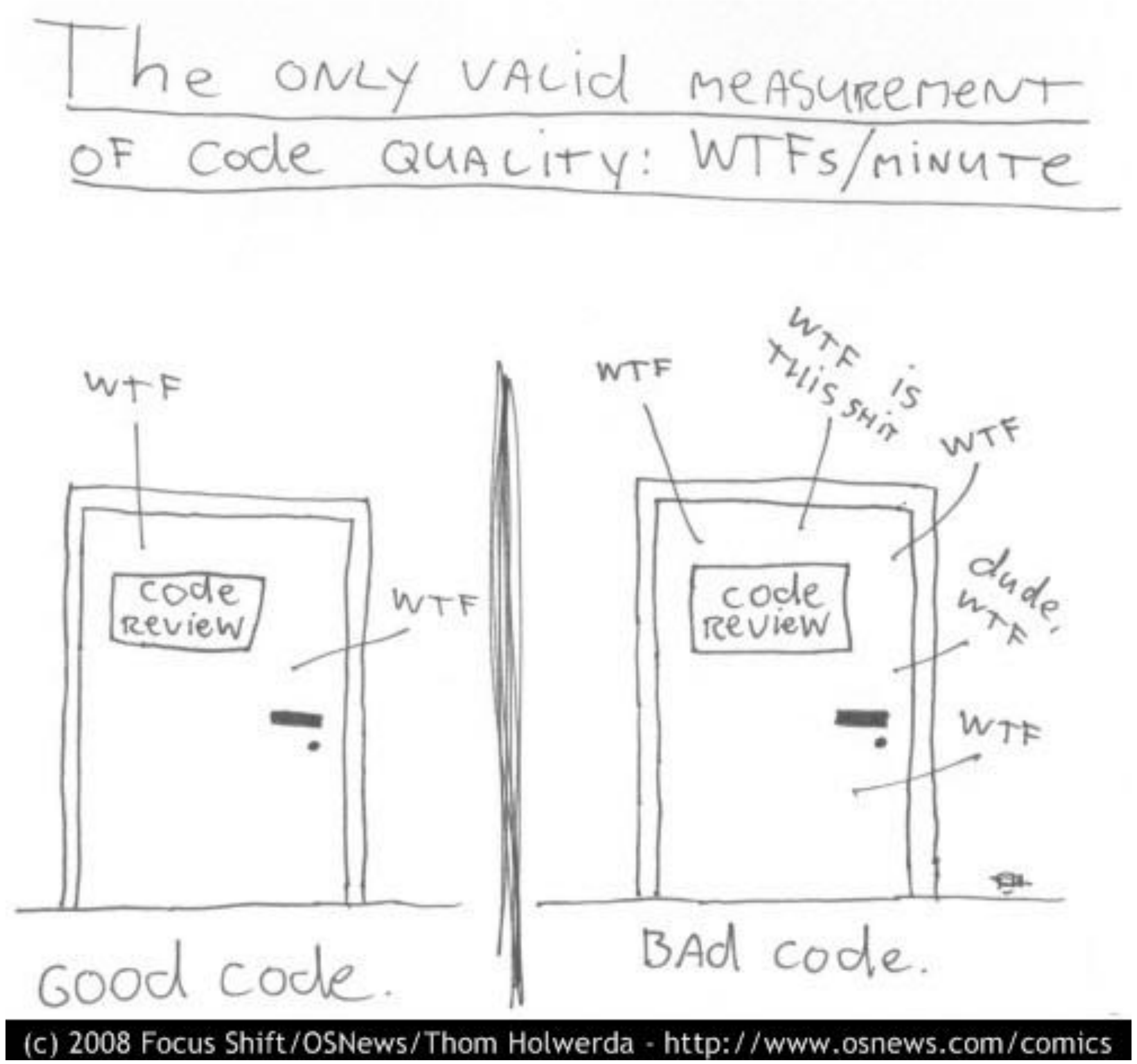

DEBT
✓ WTF factor
✓ Performance unaffected by upgrades
✓ Extended onboarding engineers
$
LEGACY
✓ Design pattern resource mismatch
✓ Performance product of capacity
✓ Candidate skill gap

@bellmar

WITHOUT A WAY TO MEASURE DEBT

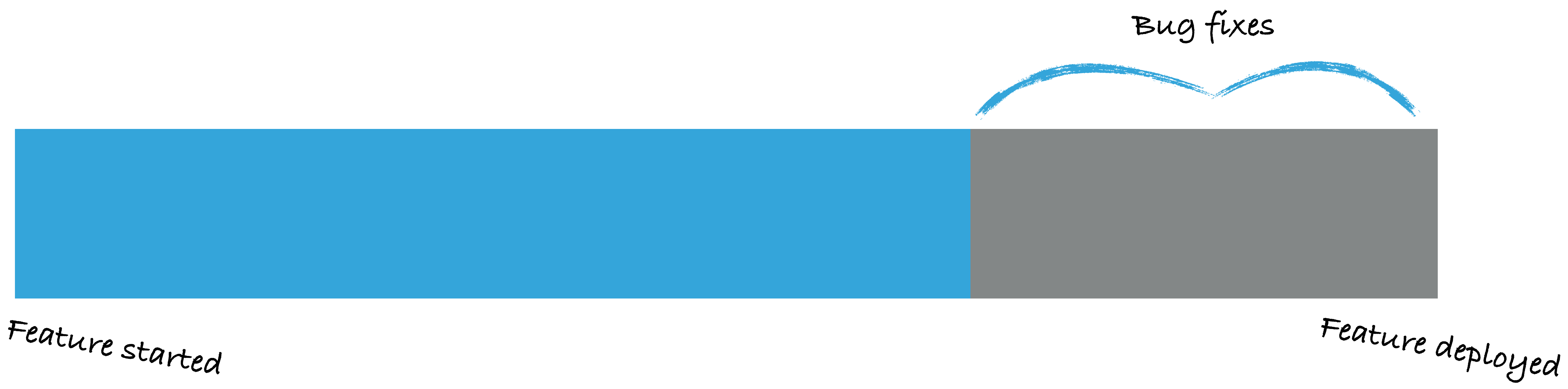
THE BEST TIME TO PAY IT DOWN IS ALWAYS TOMORROW

MEASURING TECHNICAL DEBT


The only valid measurement of code quality: WTFs/minute
WTF
code review
WTF
Good code.
WTF
WTF is this shit
WTF
code review
dude, WTF
WTF
BAd code.
(c) 2008 Focus Shift/OSNews/Thom Holwerda - http://www.osnews.com/comics

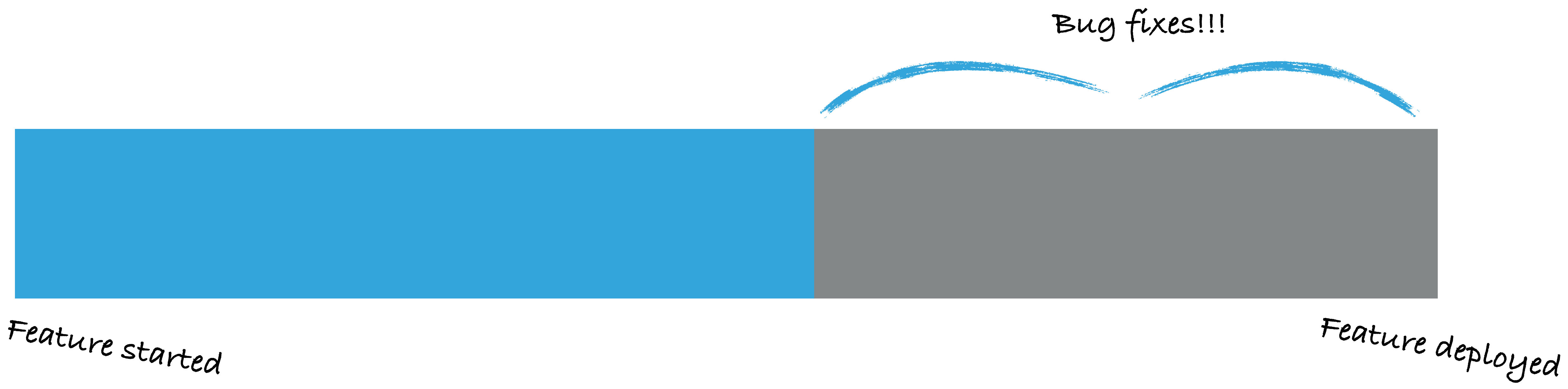
@bellmar

MEASURING TECHNICAL DEBT



@bellmar

MEASURING TECHNICAL DEBT



@bellmar

MEASURING TECHNICAL DEBT

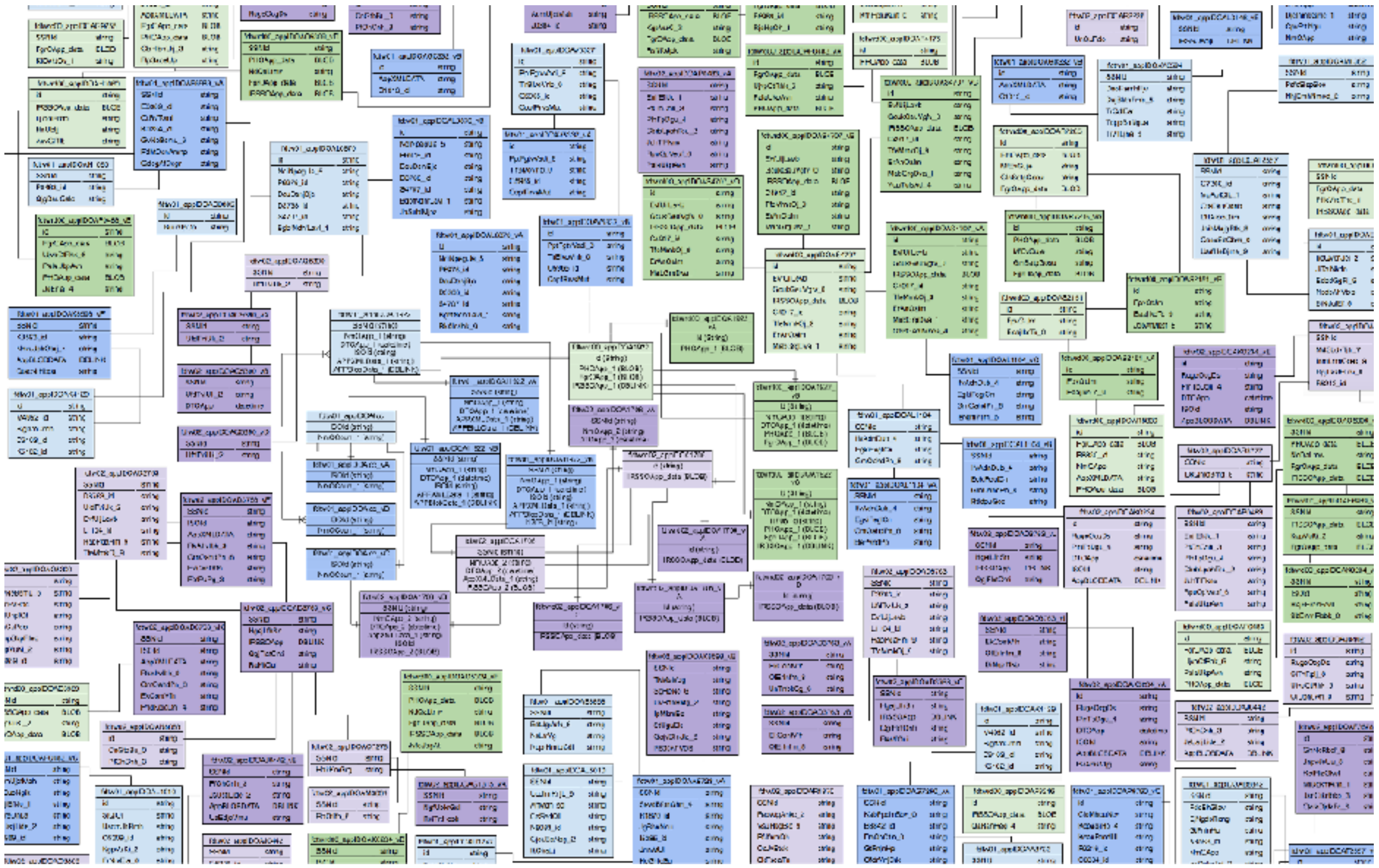
- Increase in operation costs
- Static code analysis
- Test coverage

WHY DON'T WE TALK ABOUT TECHNICAL DEBT IN THE DB?

- Most businesses only live 10 years (Time Magazine, 2015)
- Silos between DBA and Data Engineering
    - Backups? Software upgrades?
    - Normalization? Queries?
    - "From my experience, a DBA maintains existing infrastructure and a Data Engineer designs new/expanding databases"

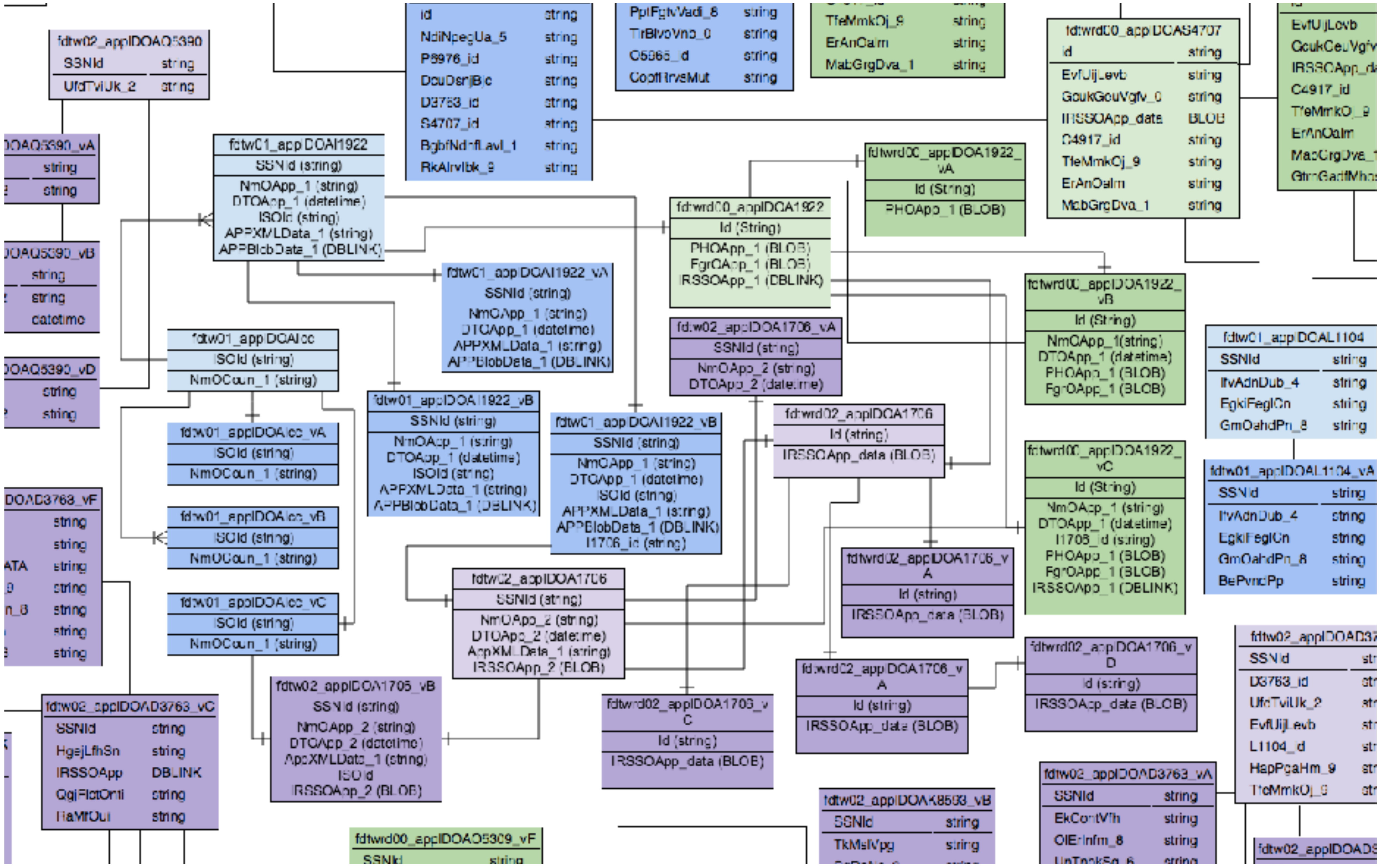
umm...what? O.o

WORSE CASE SCENARIO IN THE DATABASE

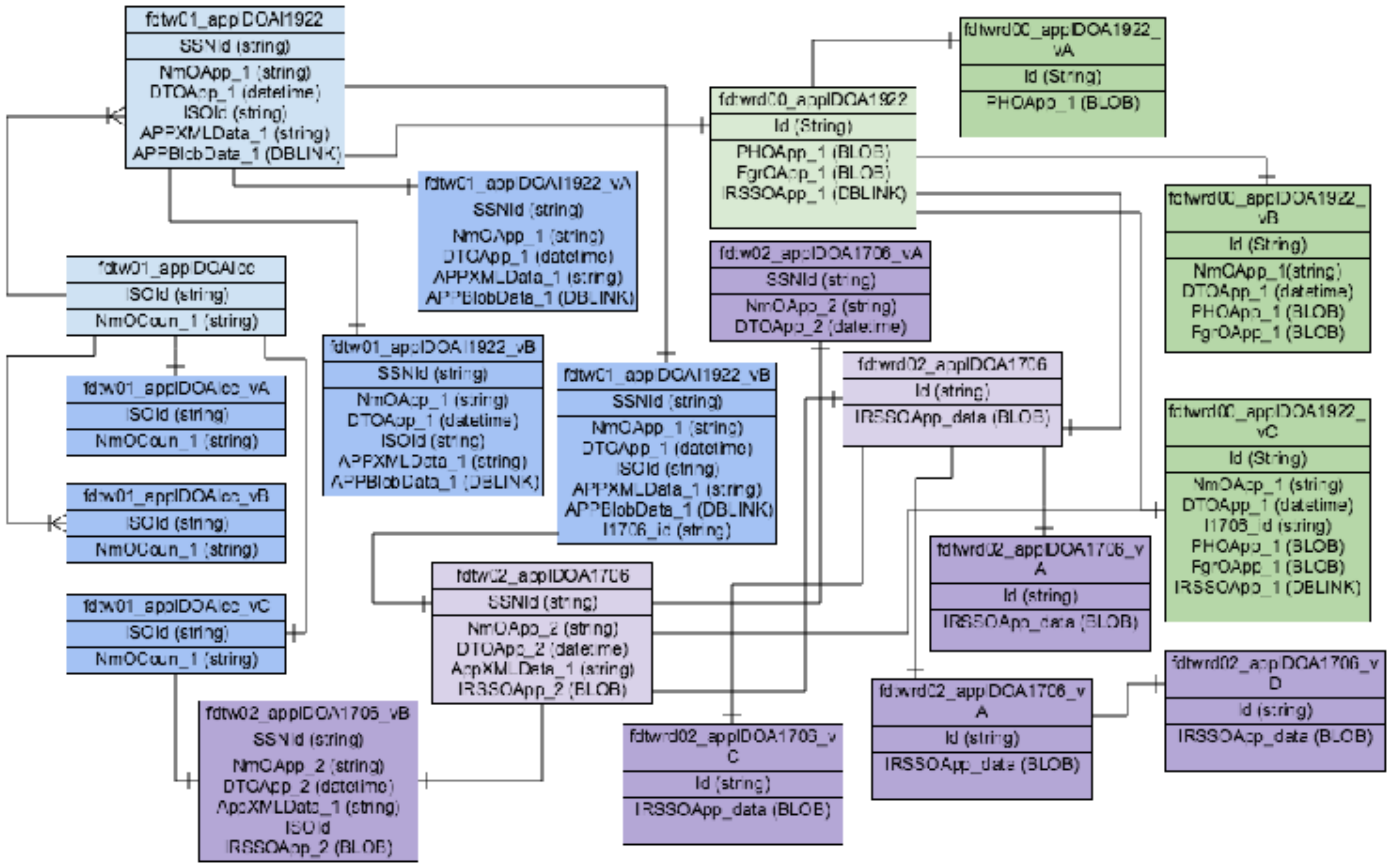


@bellmar

WORSE CASE SCENARIO IN THE DATABASE



WORSE CASE SCENARIO IN THE DATABASE



@bellmar

WORSE CASE SCENARIO IN THE DATABASE

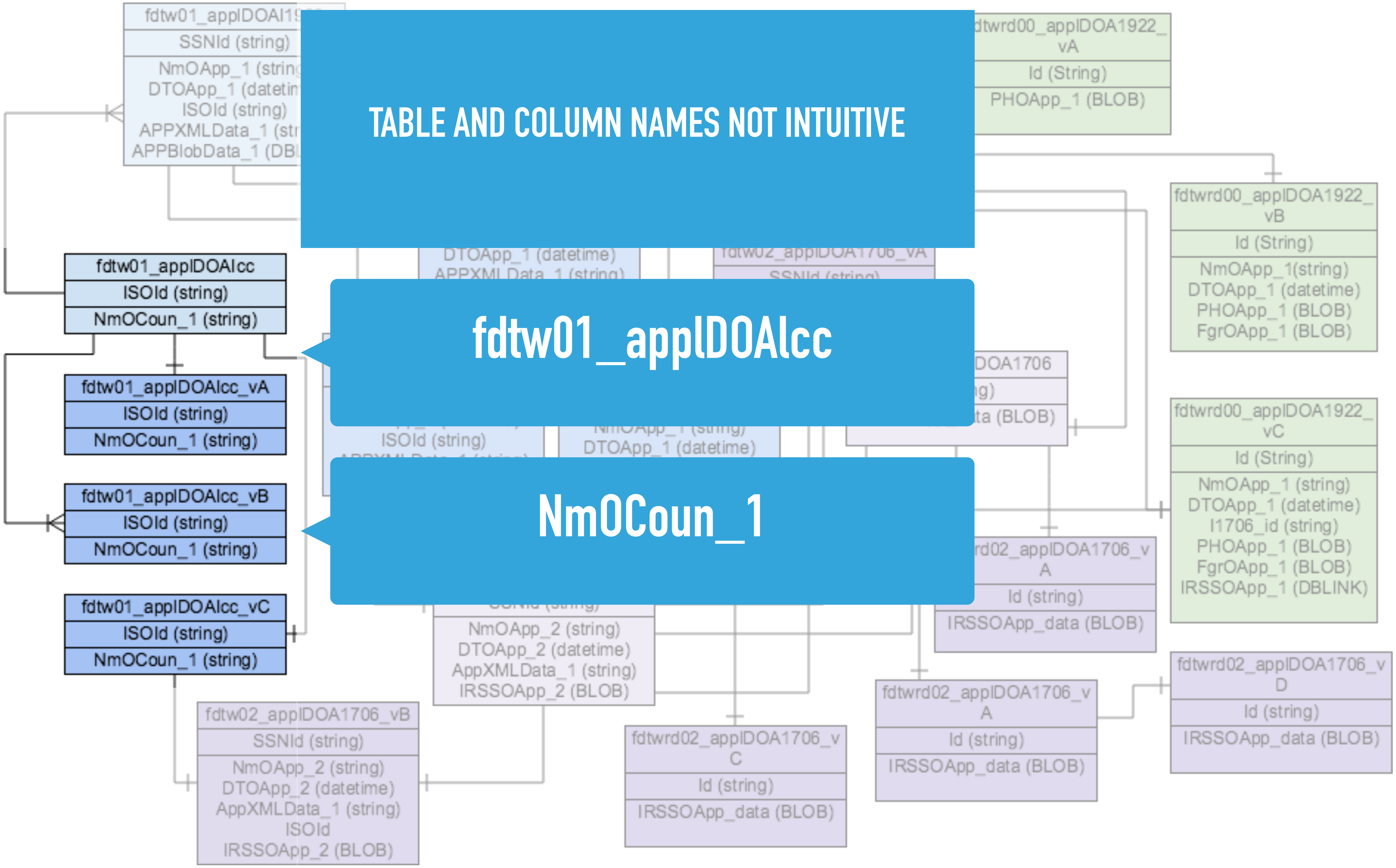


TABLE AND COLUMN NAMES NOT INTUITIVE

fdtw01_applDOAlcc

NmOCoun_1

@bellmar

WORSE CASE SCENARIO IN THE DATABASE

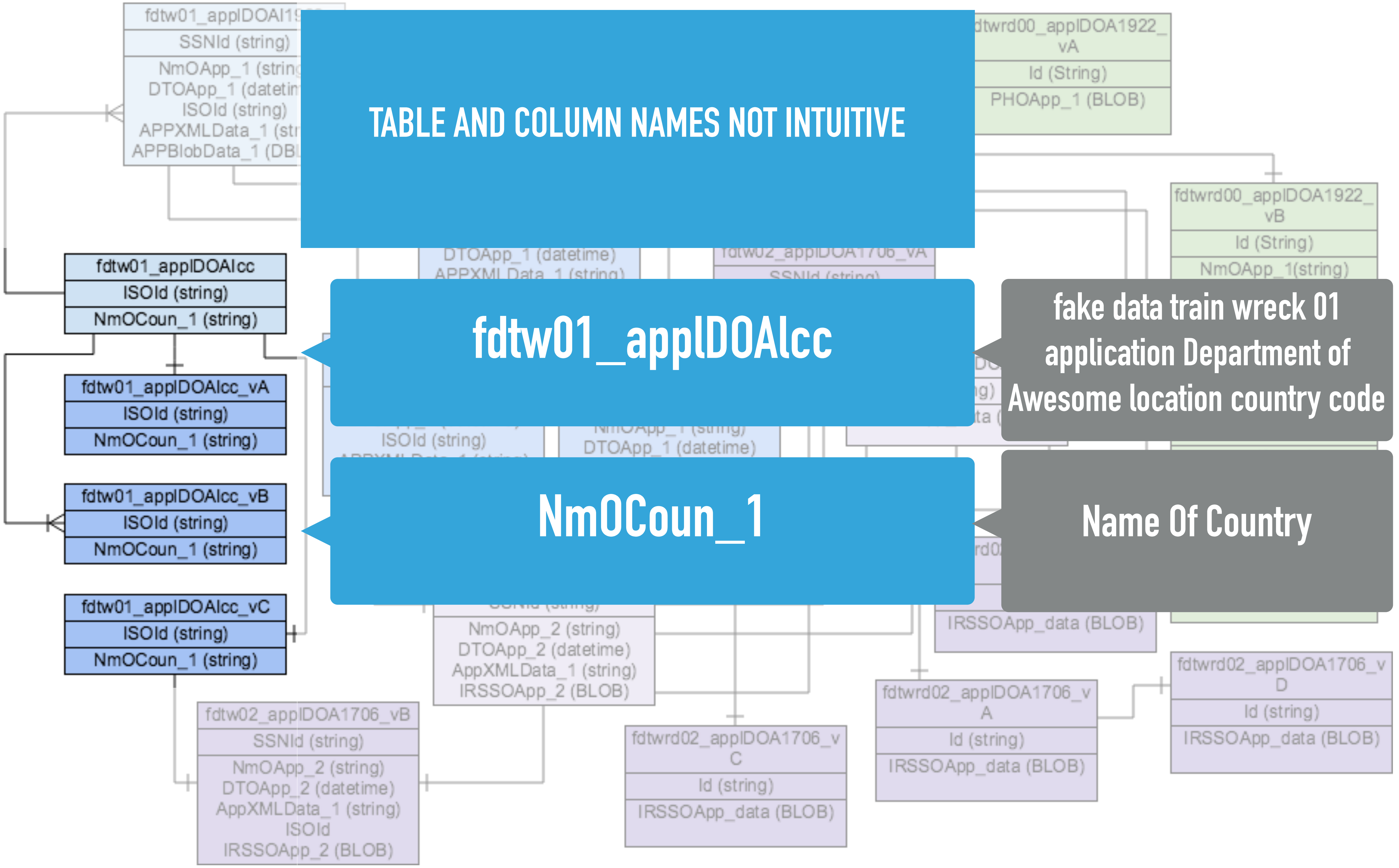


TABLE AND COLUMN NAMES NOT INTUITIVE

fdtw01_applDOAlcc

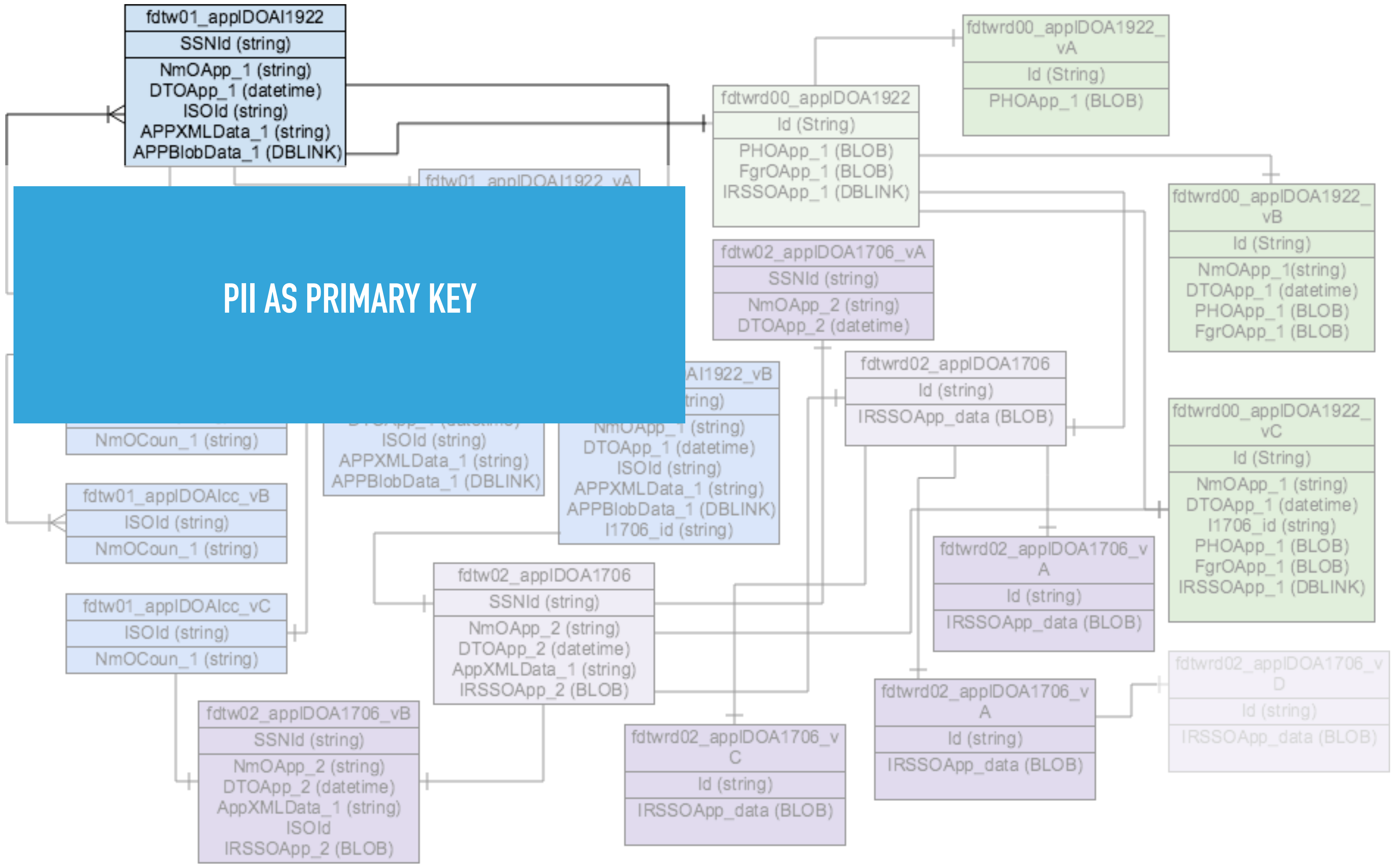
fake data train wreck 01 application Department of Awesome location country code

NmOCoun_1

Name Of Country

@bellmar

WORSE CASE SCENARIO IN THE DATABASE



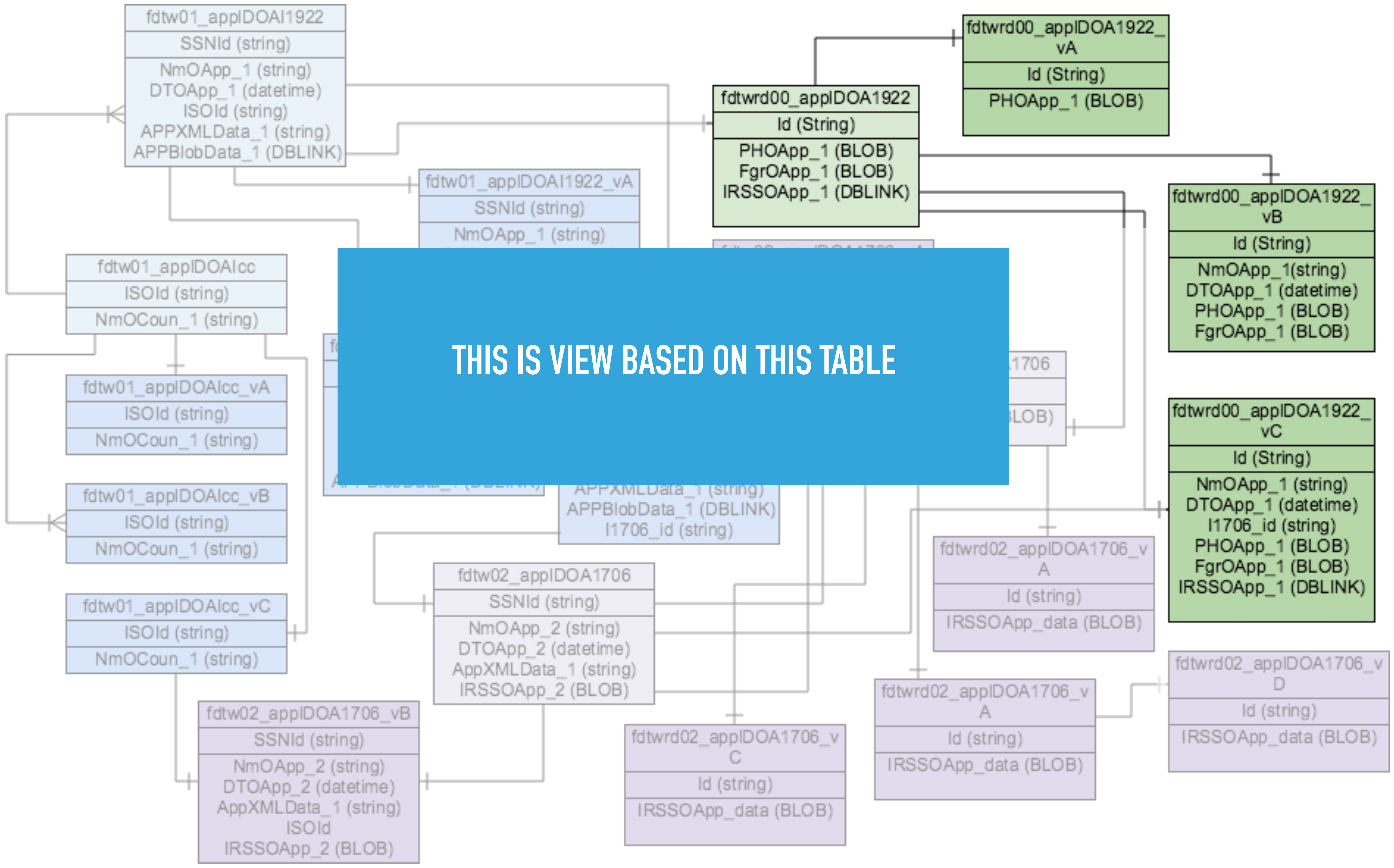
PII AS PRIMARY KEY

@bellmar

IMPROPER PRIVACY/SECURITY

- Restricting tool options (test data when PK is unencrypted PII)
- “Temporary” roles that have too much access
- Not upgrading hashing algorithms (MD5 ➔ SHA-1 ➔ SHA-256)
- Not maturing architecture as organization matures (message queues!)

WORSE CASE SCENARIO IN THE DATABASE



THIS IS VIEW BASED ON THIS TABLE

@bellmar

DATABASE VIEWS

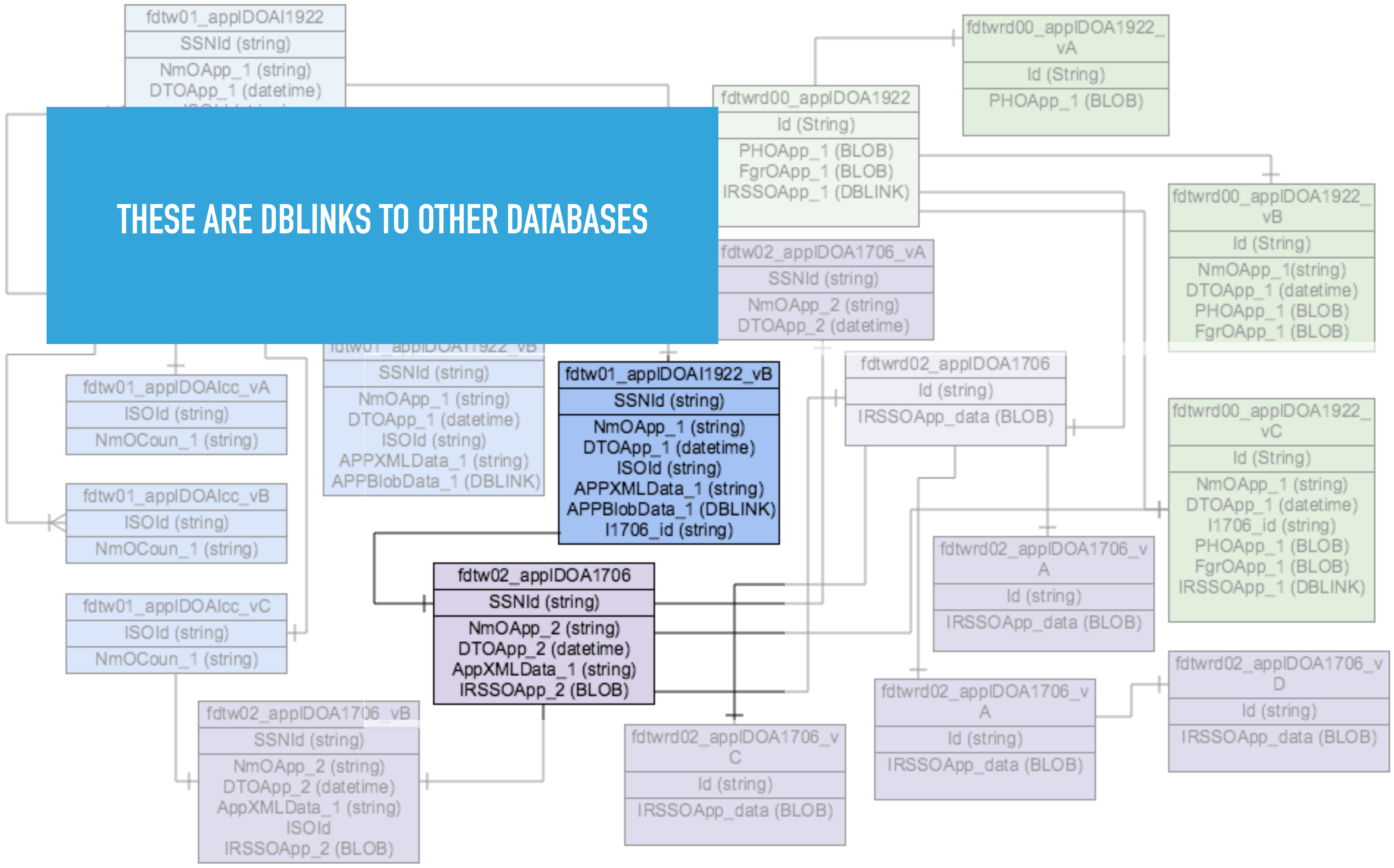
USEFUL

- Regularly joining multiple tables
- Subsets of data, better access control
- Routine db calculations (sums, geo)
- Feature flagging

DYSFUNCTIONAL

- Developer silos (my views, your views)
- Hides complexity
- Application logic in the db
- Small but not nonexistent

WORSE CASE SCENARIO IN THE DATABASE



THESE ARE DBLINKS TO OTHER DATABASES

@bellmar

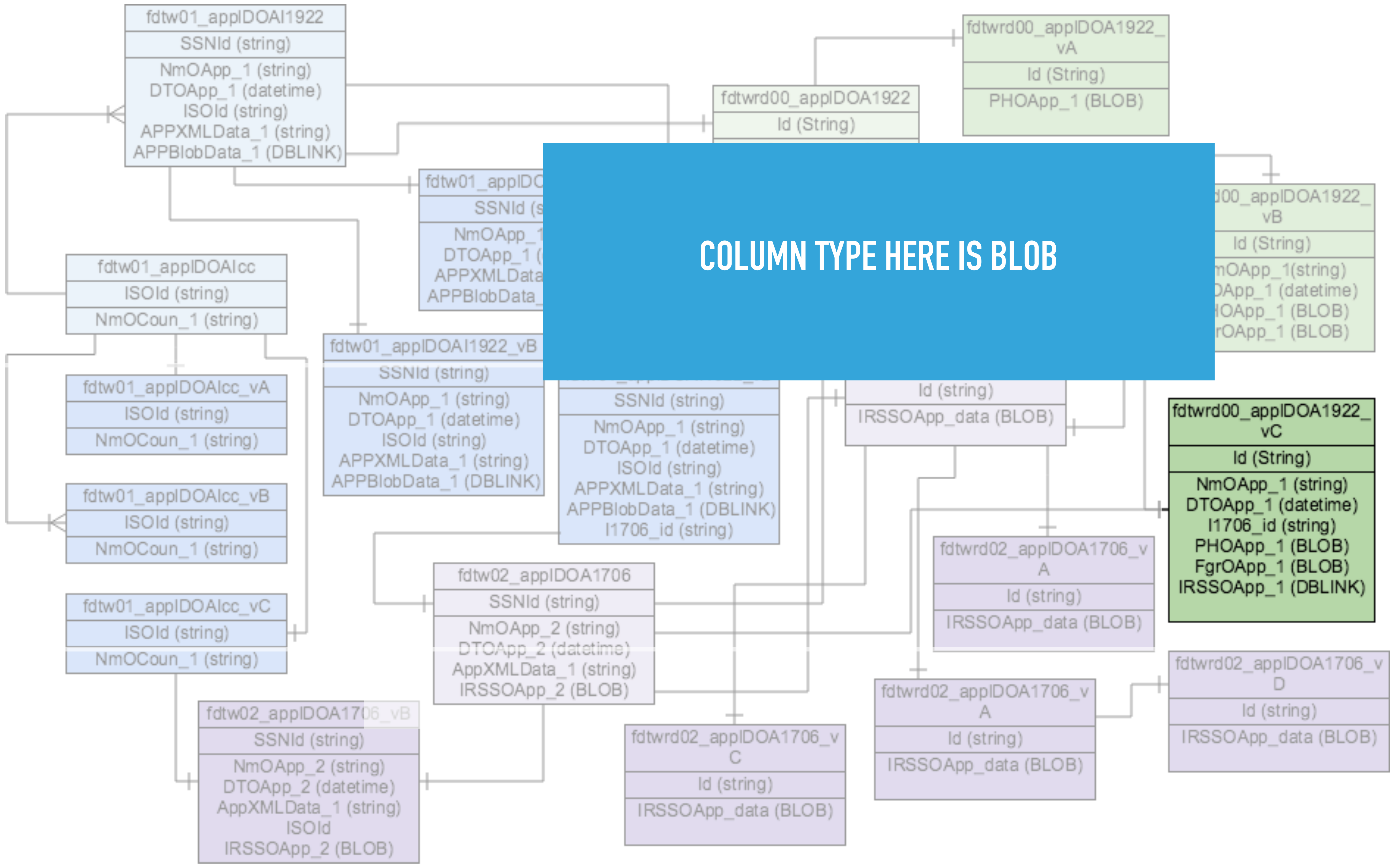
DBLINKS AND OTHER MAGIC TRICKS

- DBlinks: joins across databases
- Query time + network speed
- Directionality: Query from which table? Can affect performance
- Complicates security
- Why was this data separate in the first place?

STORED PROCEDURES

- Scripts inserted and run on the db itself
- Code will usually run faster on the db than the application
- Application logic kept away from where application teams can see it
- “We don’t need to version control it because it’s in our backups”
- Harder to trace or predict impact of changes

WORSE CASE SCENARIO IN THE DATABASE



COLUMN TYPE HERE IS BLOB

@bellmar

BLOBS: WHAT DATA IS DATA?

- BLOB = Binary Large Object
- Images, audio, executables … these things are not queryable
- Popular as storage became cheap, but inflates the size of the database
- As connection speeds increase, cloud file storage (AWS S3) preferred

WHAT SHOULD YOU DO?

- Audit the queries?
- Migrate to NoSQL?
- Rewrite and simplify the applications using this db?
- Light the thing on fire and go home?

INCREMENTAL FAILURE IS SOLVED WITH

INCREMENTAL IMPROVEMENT

RESPONSIBILITY GAP

Backups

Schema design

Upgrading

Query optimization

Analysis

Security

Normalization

DBAS
DEVS
OPS

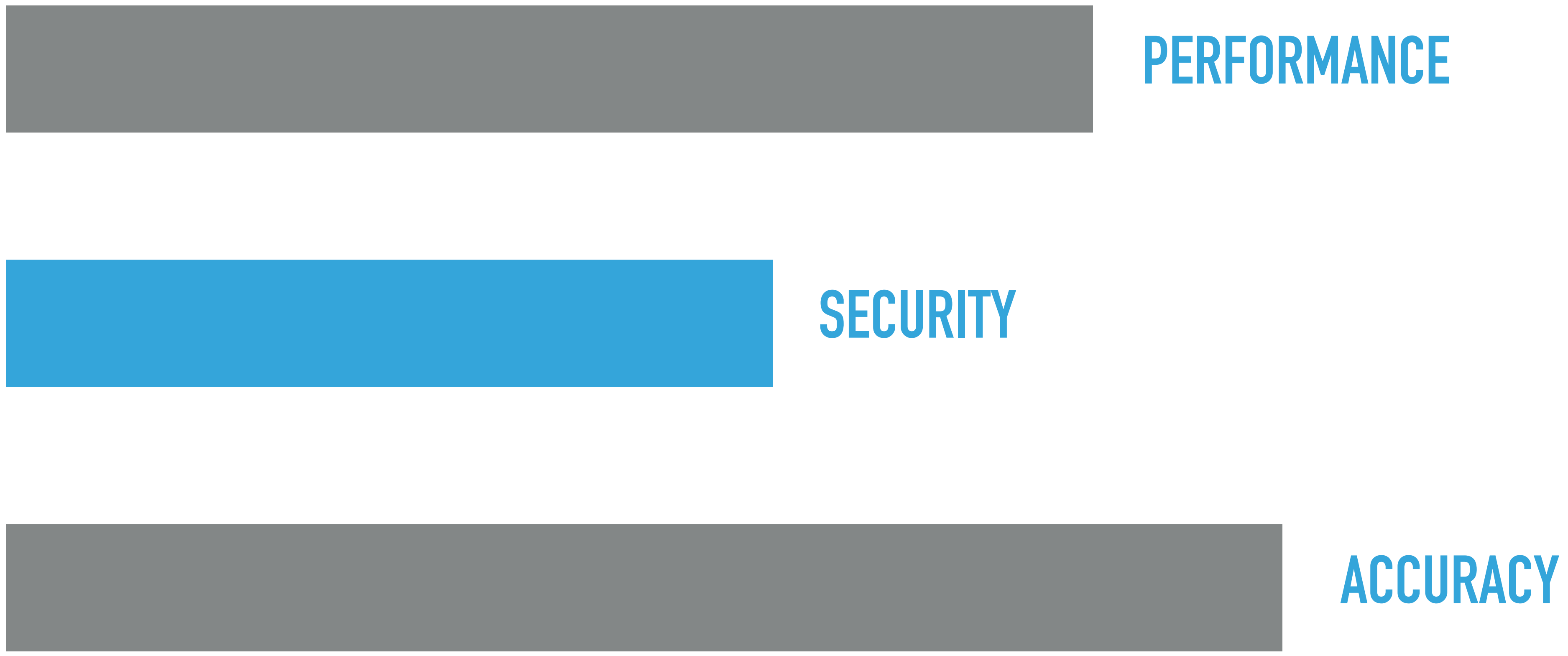
ARE ENGINEERS FIRST CLASS CITIZENS?

Oh stop being so dramatic. It's just one little change

COMMUNICATION TOOLS

- Organized chat: Can people figure out who each other are and reach out quickly?
- Automation: Readable configuration scripts and immutable architecture means devs can see hidden logic. Dev environments easier to setup
- Documentation: How does your data dictionary relate to your code documentation? Are you using ORM?

DEFINE YOUR GOAL



PRIORITIZE BASED ON WORST QUERIES

TIME PER EXEC	TOTAL TIME	EXECUTIONS	% CPU
88.89	1,155.52	13	78.09
6.33	43,742.23	6,914	86.07
0.55	792.04	1,442	83.38

PRIORITIZE BASED ON WORST QUERIES

TIME PER EXEC	TOTAL TIME	EXECUTIONS	% CPU
88.89	1,155.52	13	78.09
6.33	43,742.23	6,914	86.07
0.55	792.04	1,442	83.38

PRIORITIZE BASED ON WORST QUERIES

TIME PER EXEC	TOTAL TIME	EXECUTIONS	% CPU
88.89	1,155.52	13	78.09
6.33	43,742.23	6,914	86.07
0.55	792.04	1,442	83.38

PRIORITIZE BASED ON WORST QUERIES

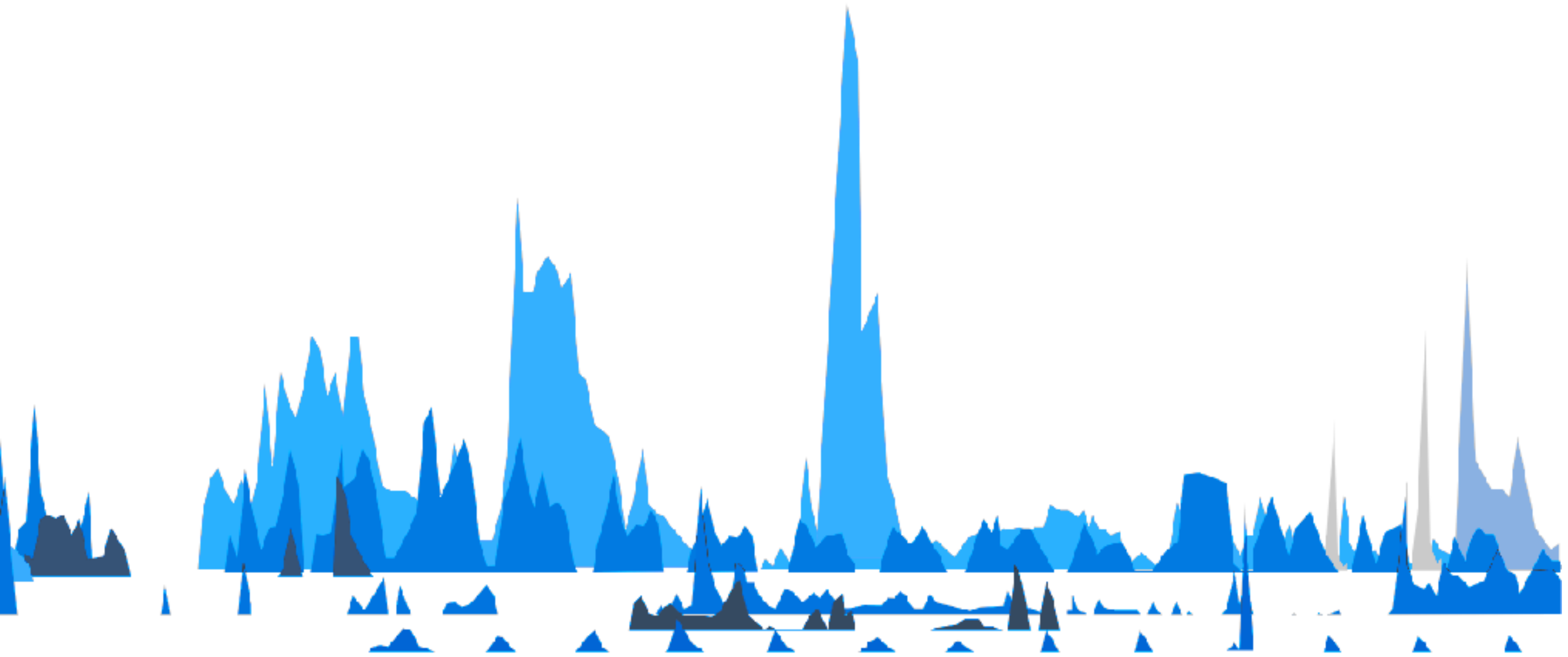
TIME PER EXEC	TOTAL TIME	EXECUTIONS	% CPU
88.89	1,155.52	13	78.09
6.33	43,742.23	6,914	86.07
0.55	792.04	1,442	83.38

@bellmar

PRIORITIZE BASED ON WORST QUERIES

TIME PER EXEC	TOTAL TIME	EXECUTIONS	% CPU
88.89	1,155.52	13	78.09
6.33	43,742.23	6,914	86.07
0.55	792.04	1,442	83.38

MINIMIZE VECTORS

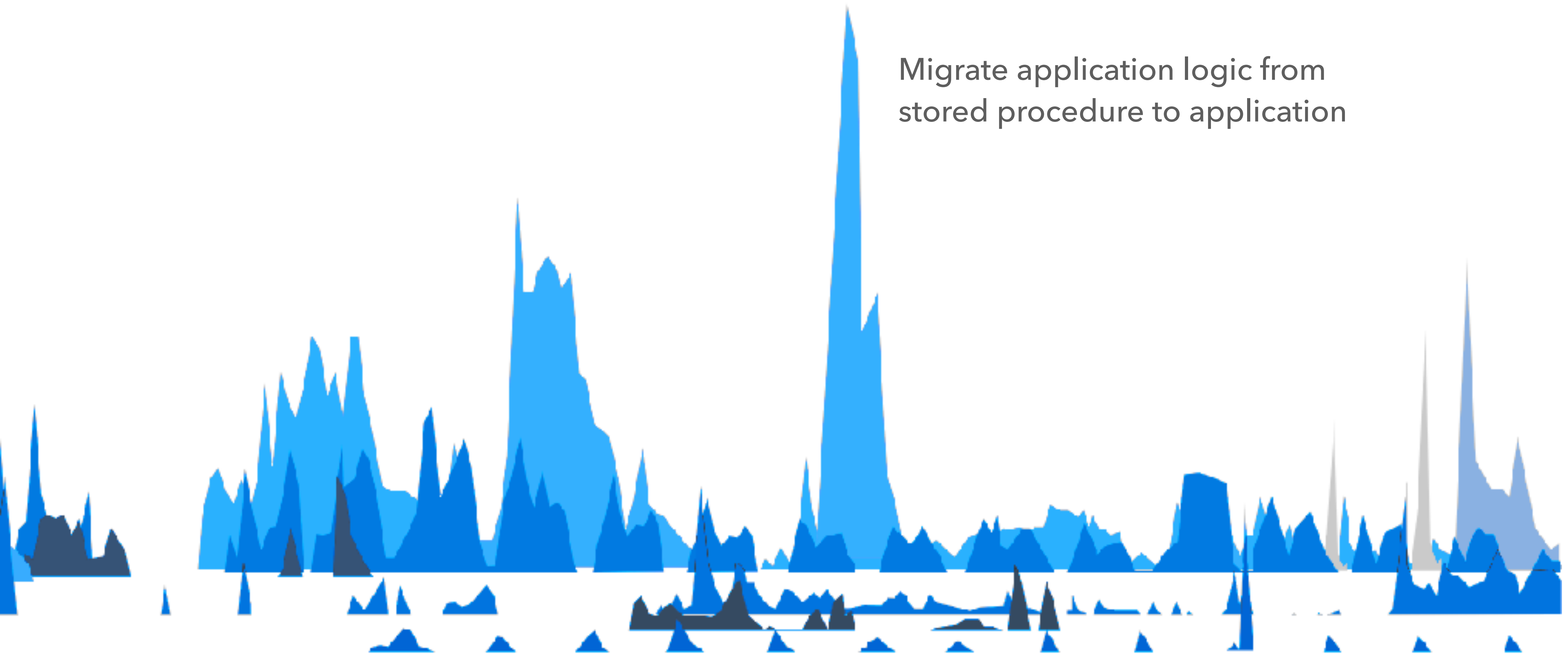


@bellmar

WORSE CASE SCENARIO IN THE DATABASE

MINIMIZE VECTORS

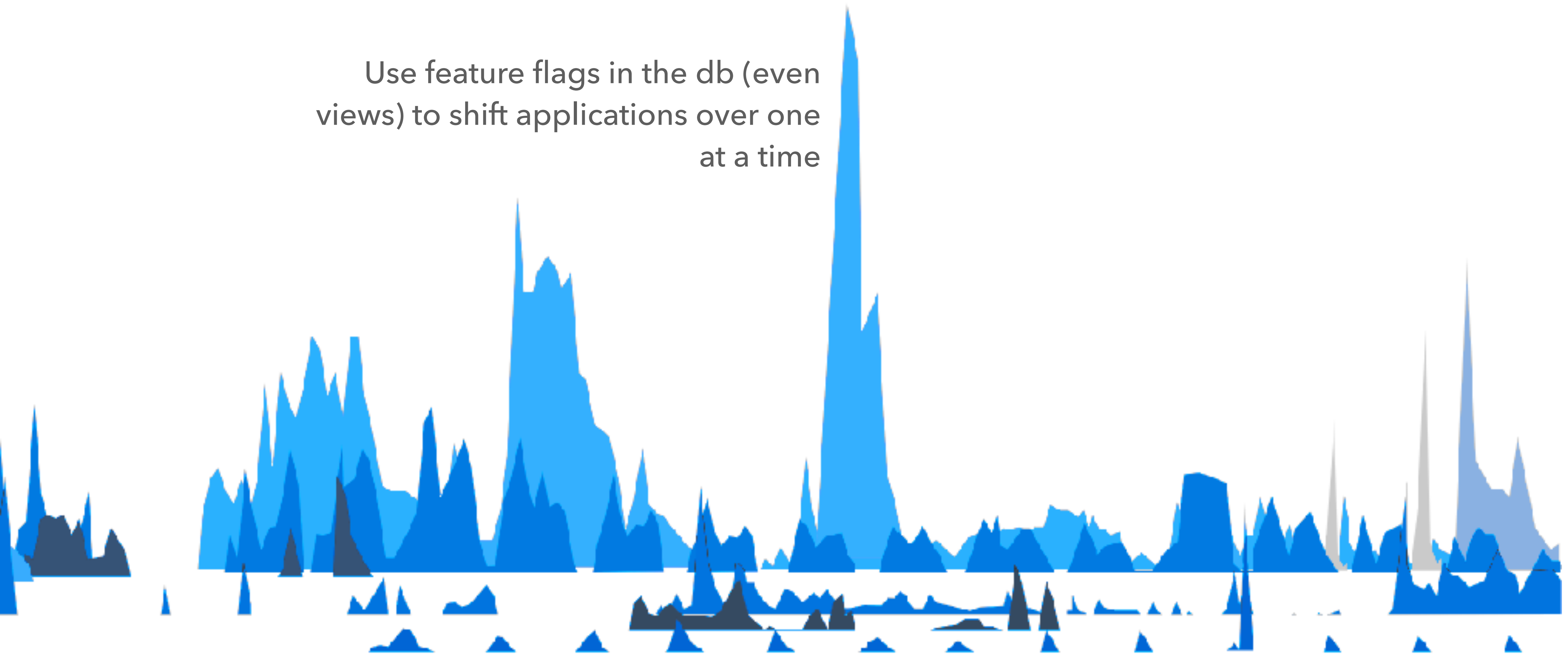
Migrate application logic from stored procedure to application



@bellmar

MINIMIZE VECTORS

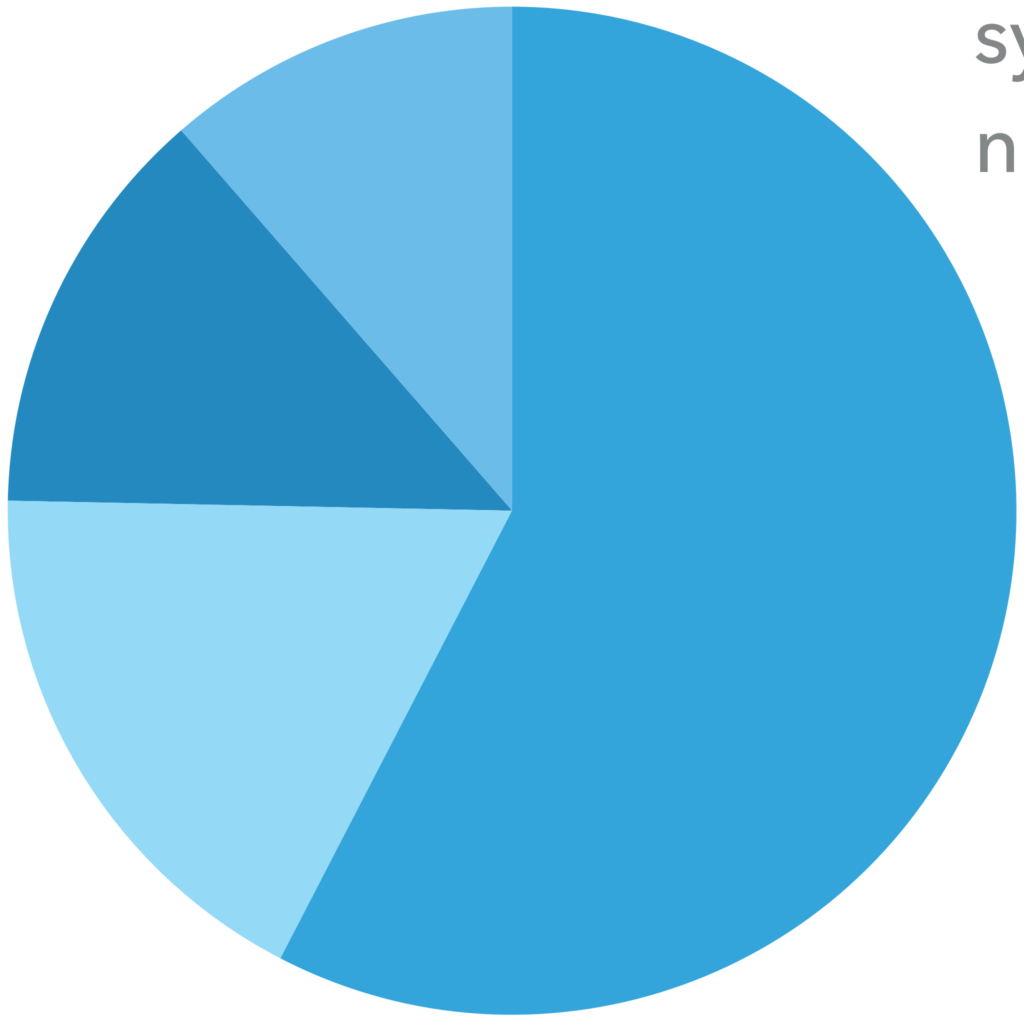
Use feature flags in the db (even views) to shift applications over one at a time



@bellmar

WORSE CASE SCENARIO IN THE DATABASE

REDUCE OVERALL DATABASE SIZE



"After 3 months the data the user has entered into our system for their shipment is not accessed again"

← 6+ months old

@bellmar

WORSE CASE SCENARIO IN THE DATABASE

OK, BUT I WOULD NEVER WORK FOR A COMPANY LIKE THIS. WHY SHOULD I CARE?

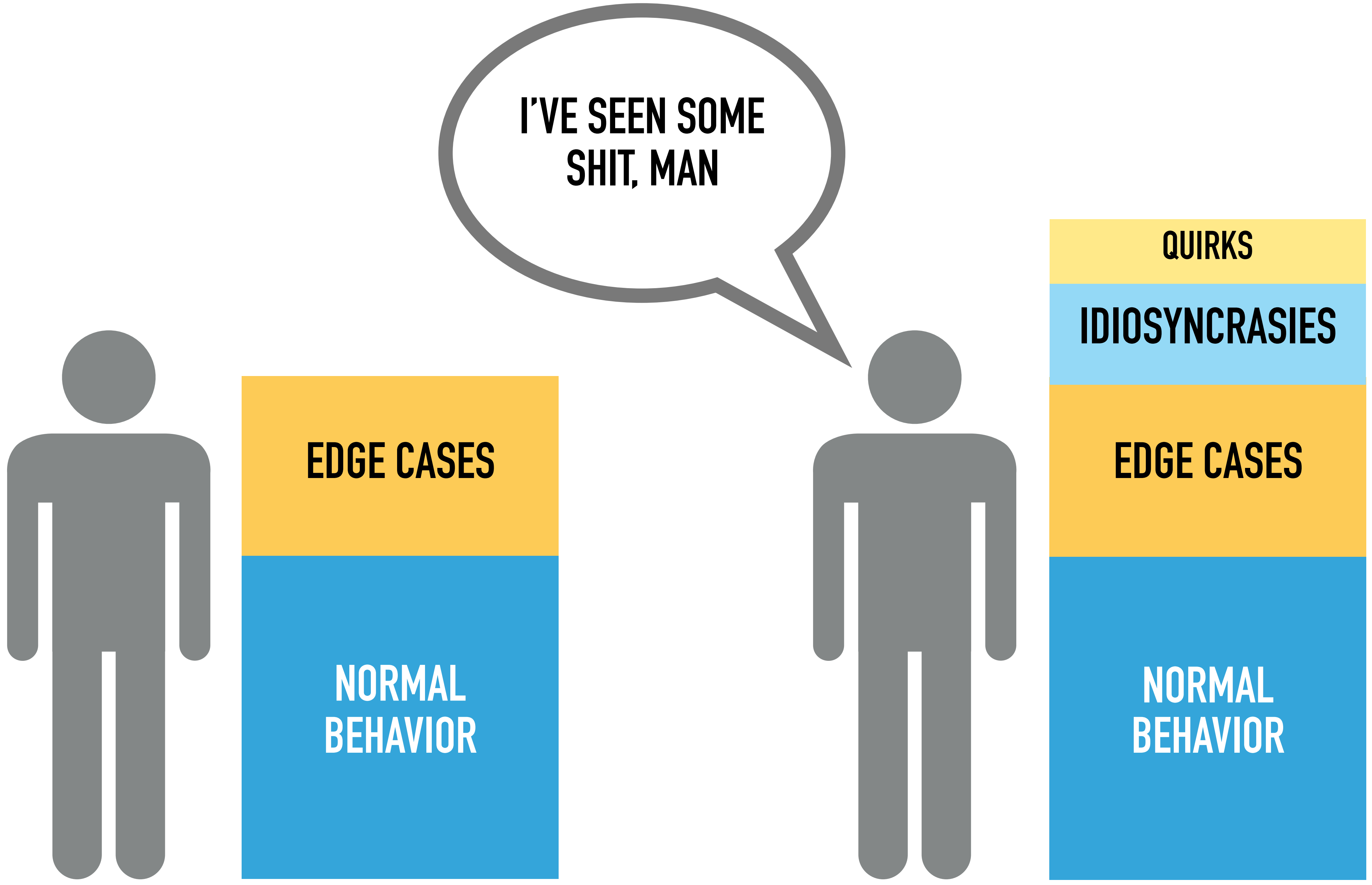
You, and thousands of other engineers

THIS DATA IS IMPORTANT

- Background checks for visa applications
- Shipping logistics for military families
- Financial data that informs decisions made by pensions and investment managers
- Election records

“I HAVE [X] YEARS OF EXPERIENCE AND I’M STILL TERRIFIED OF BEING ON CALL”

HOW WE JUDGE EXPERTISE



@bellmar

TECHNICAL RESPONSE UNIT

THANK YOU!

COME TALK TO US ABOUT OPPORTUNITIES

THE U.S. DIGITAL SERVICE