



# AI From Cloud to Edge

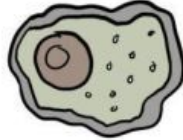
Role of DevOps in this changing world

Gautam R | Regional Solutions Engineer | ANZ

May 2024

# EVOLUTION OF OPERATIONS

OPS



- PRIMORDIAL, PROTOZOIC
- BORN IN THE SWAMPS OF PERL
- OPERATES IN A SINGLE-CELL SILO
- SURPRISINGLY RESILIENT

DEVOPS



- A CROSS-FUNCTIONAL MARVEL
- VASTLY INCREASED AGILITY
- SECRETLY JUST A BUNCH OF SINGLE CELLS THAT HAVE LEARNED NOT TO KILL EACH OTHER

DEVSECOPS



- MORE ADVANCED, MORE PARANOID
- SECURITY IS AUTOMATED RIGHT INTO ITS DNA
- KNOWS THAT SHARED RESPONSIBILITY IS THE ONLY ESCAPE FROM FOSSILIZATION

DEVSECMLOPS



- WHAT EVEN IS THIS?
- IS IT A FISH WITH FEET?
- WE SHOULD PROBABLY LEAVE IT ALONE FOR A FEW MILLION YEARS AND SEE WHAT HAPPENS

TRICERATOPS



- DOES NOT CARE ABOUT YOUR ORG STRUCTURE
- VULNERABLE ONLY TO DIRECT METEOR STRIKES
- WHAT WERE WE TALKING ABOUT, AGAIN?

@acloudguru



## **Regional SE - ANZ, Couchbase**

Over a decade of experience as Consultant,  
Architect, DevOps & MLOps Specialist



**Connect with me  
on LinkedIn**

# What happens on the Internet in 60 seconds

- 5.9 million Google searches take place.
- More than 231 million emails are sent.
- Around 500 hours of videos are uploaded on YouTube.
- Instagram users share around 66,000 photos.
- 1.7 million content pieces are shared on Facebook.
- More than 4.3 million Snaps are sent on Snapchat.
- 16 million text messages are sent.
- More than \$443,000 is spent by consumers on Amazon.
- Around 347,200 tweets are tweeted on Twitter.
- Over 120 professionals join LinkedIn.



# A day in Life

>



# Innovation Leaders



**FICO**

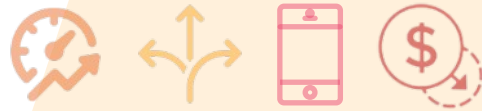
Global fraud detection  
and prevention

**3B+**

payment accounts protected

**65%**

of world's card accounts



**Carnival**

City-at-sea experience via IoT  
& edge-based mobile devices

**15k**

mobile sensors per voyage

**100's M**

operations per/day



**LinkedIn**

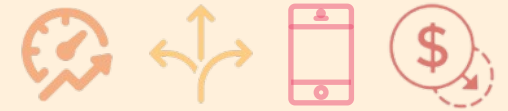
High performance  
caching for social  
network

**10M+**

requests / second

**800M+**

users



**STAPLES**

Product catalog, pricing,  
shopping cart, daily deals,  
inventory management

**4,000**

stores worldwide

**1.6B**

rules applied in real-time

# What customers want?

>



# Customers & The Market has changed

- 65% of a company's revenue comes from the repeat business of existing customers.
- Over 57% of the data worldwide is generated by internet users worldwide.
- Almost 90% of the data stored today is unstructured.
- 1 in 3 customers will leave a brand they love after just one bad experience.
- 57% of customers won't recommend a business with a poorly designed website on mobile.
- 50% of customers will stop visiting a site that isn't mobile-friendly, even if they like the business.
- By end of 2023, there were 1.6 networked mobile devices and connections per person.

# Customers & The Market: Application Needs Have Changed



## Deliver Great Experiences

- Personalized & responsive
- Anywhere & everywhere
- Real time info & inventory
- Ensuring dependable transactions

## Develop Efficiently

- Simplify & accelerate development
- Tap my skills (SDKs & SQL)
- Support best practices
- Avoid data sprawl

## Deploy Effectively

- Support hybrid clouds, Edge, 5G
- 100% uptime & global scale
- Flexible management options
- Cost effective

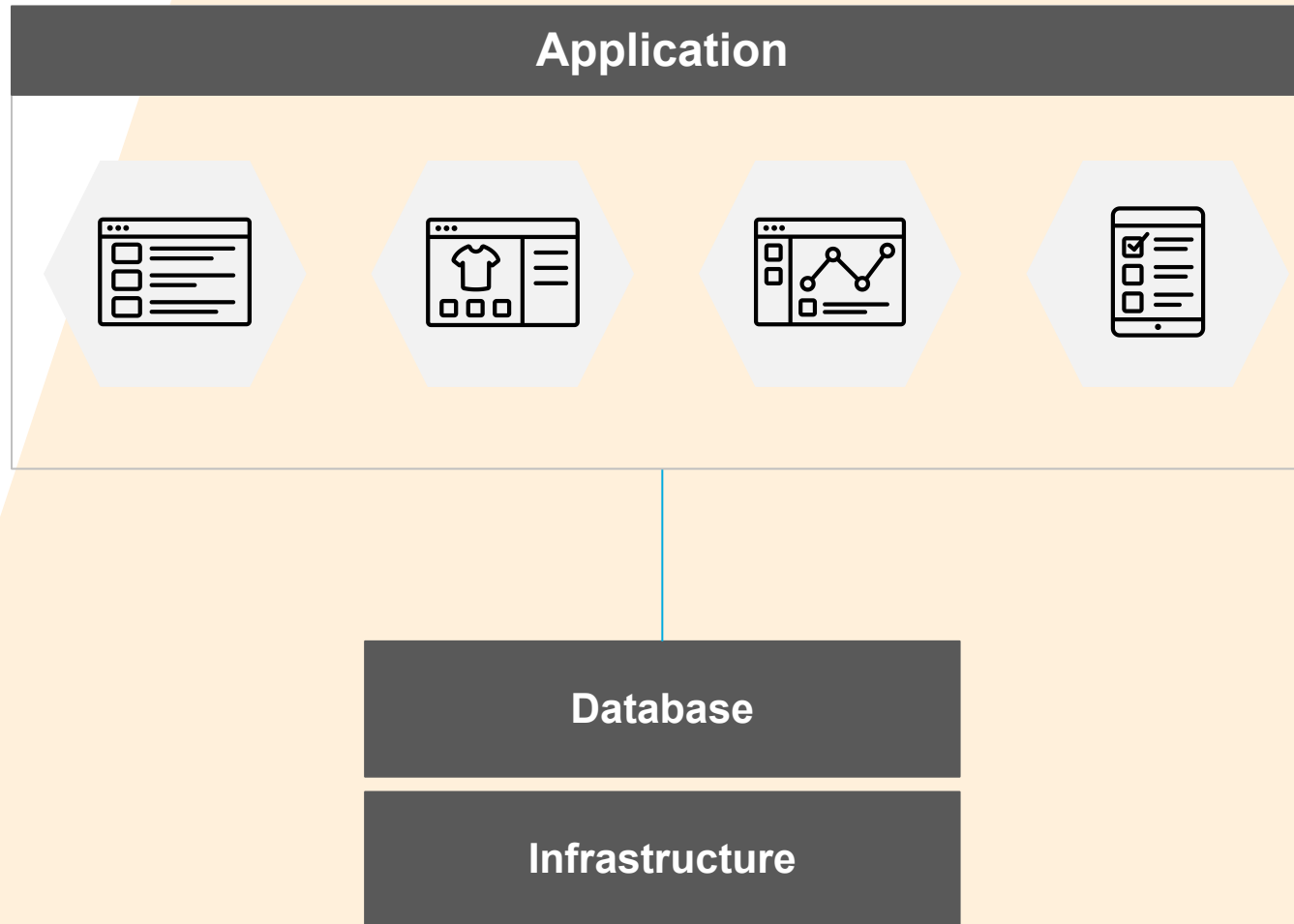
# The Reality

>

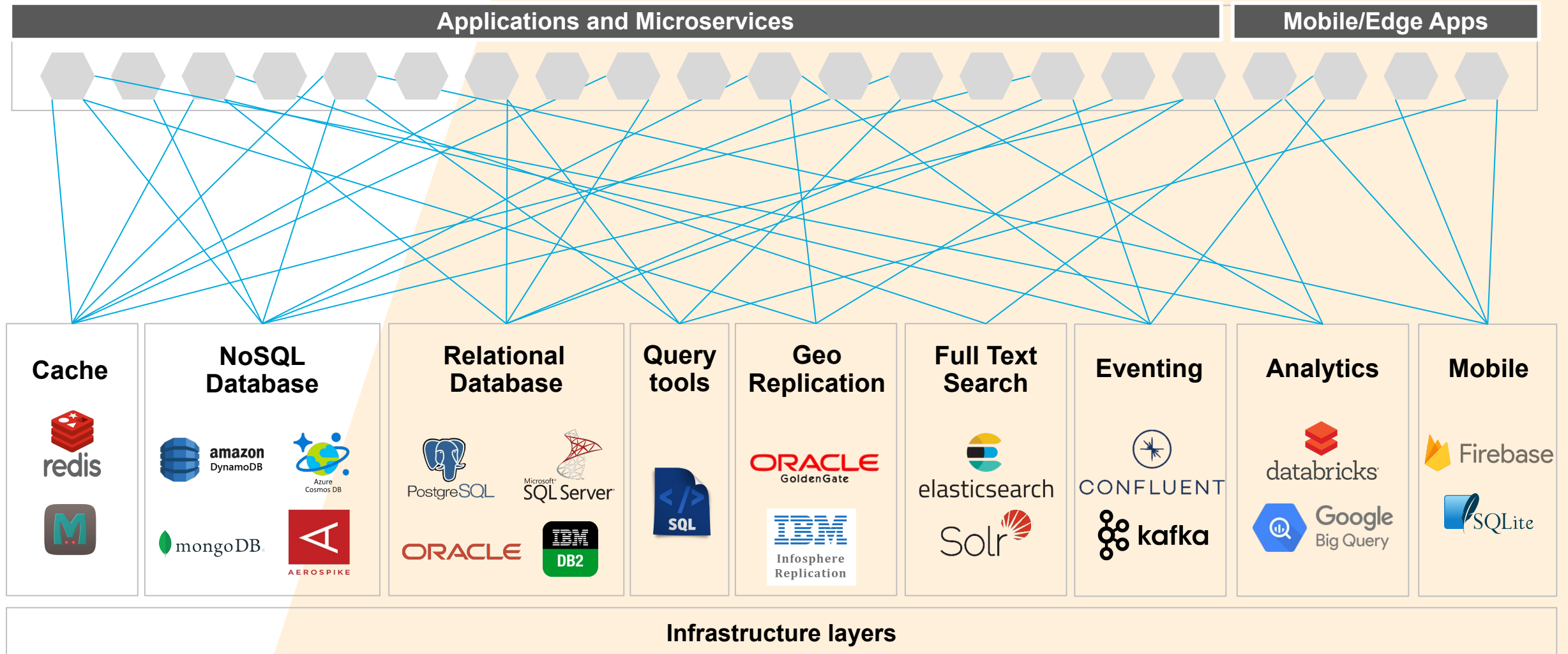


# Simplistic View: Data Management For Applications

**Functional  
Requirements**

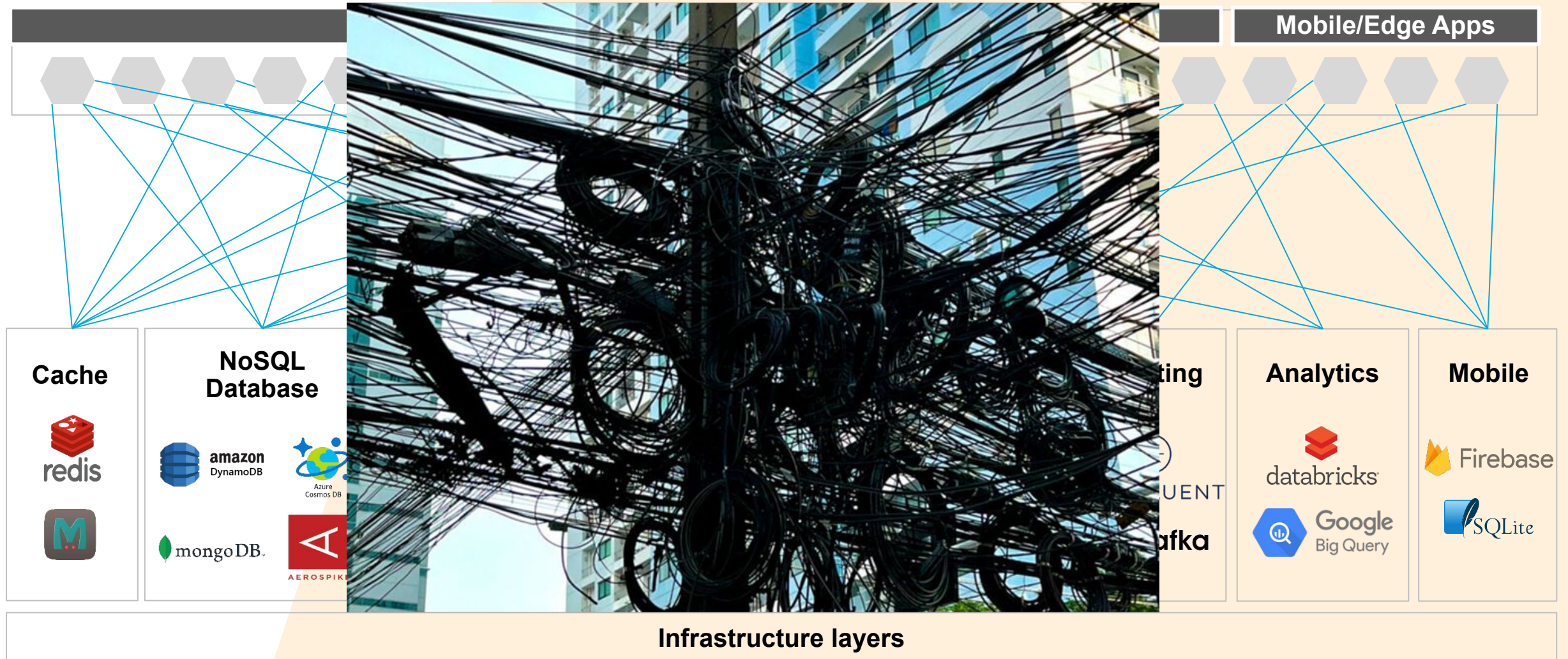


# Reality: Data Sprawl & Management Challenges





# Reality: Data Sprawl & Management Challenges



# DevOps to Rescue

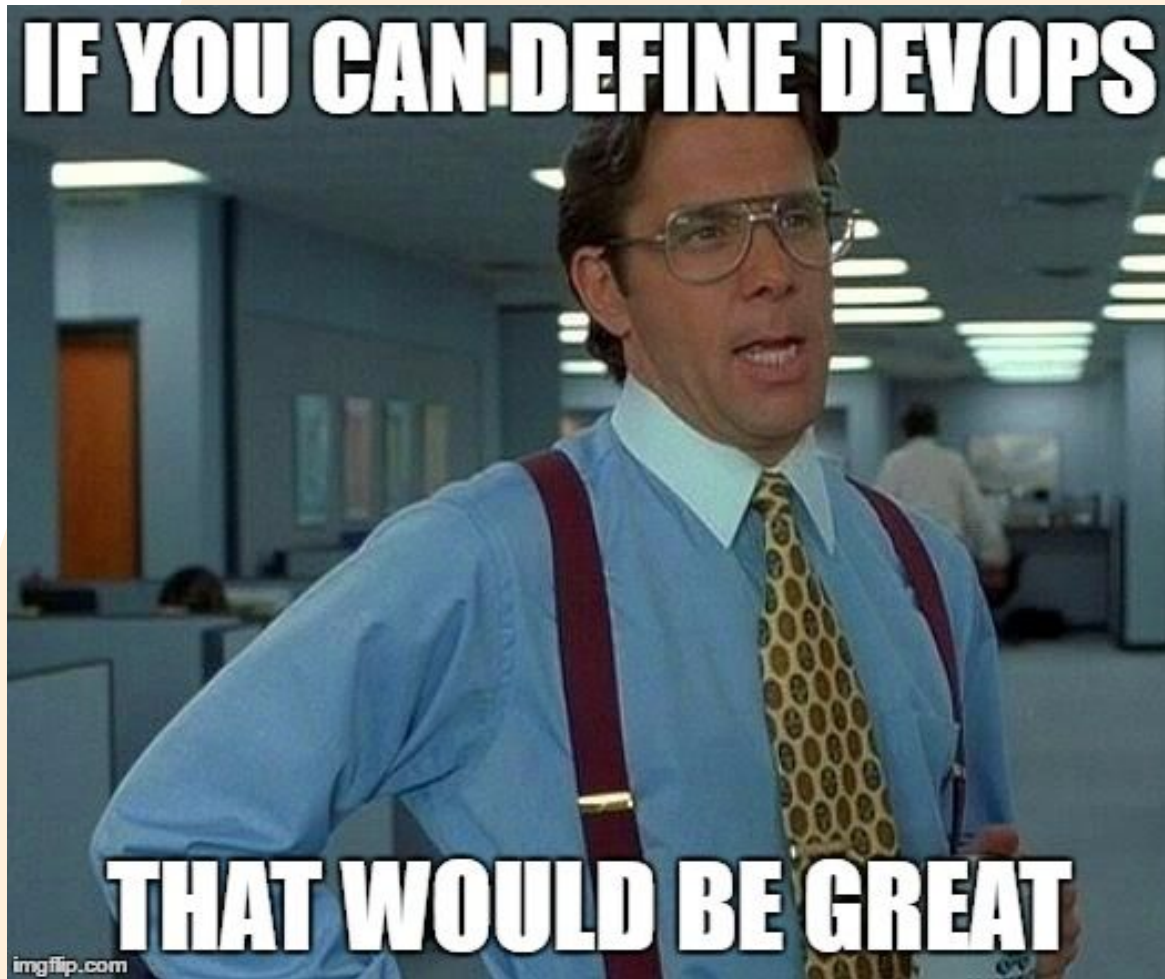
>



**TIME FOR**



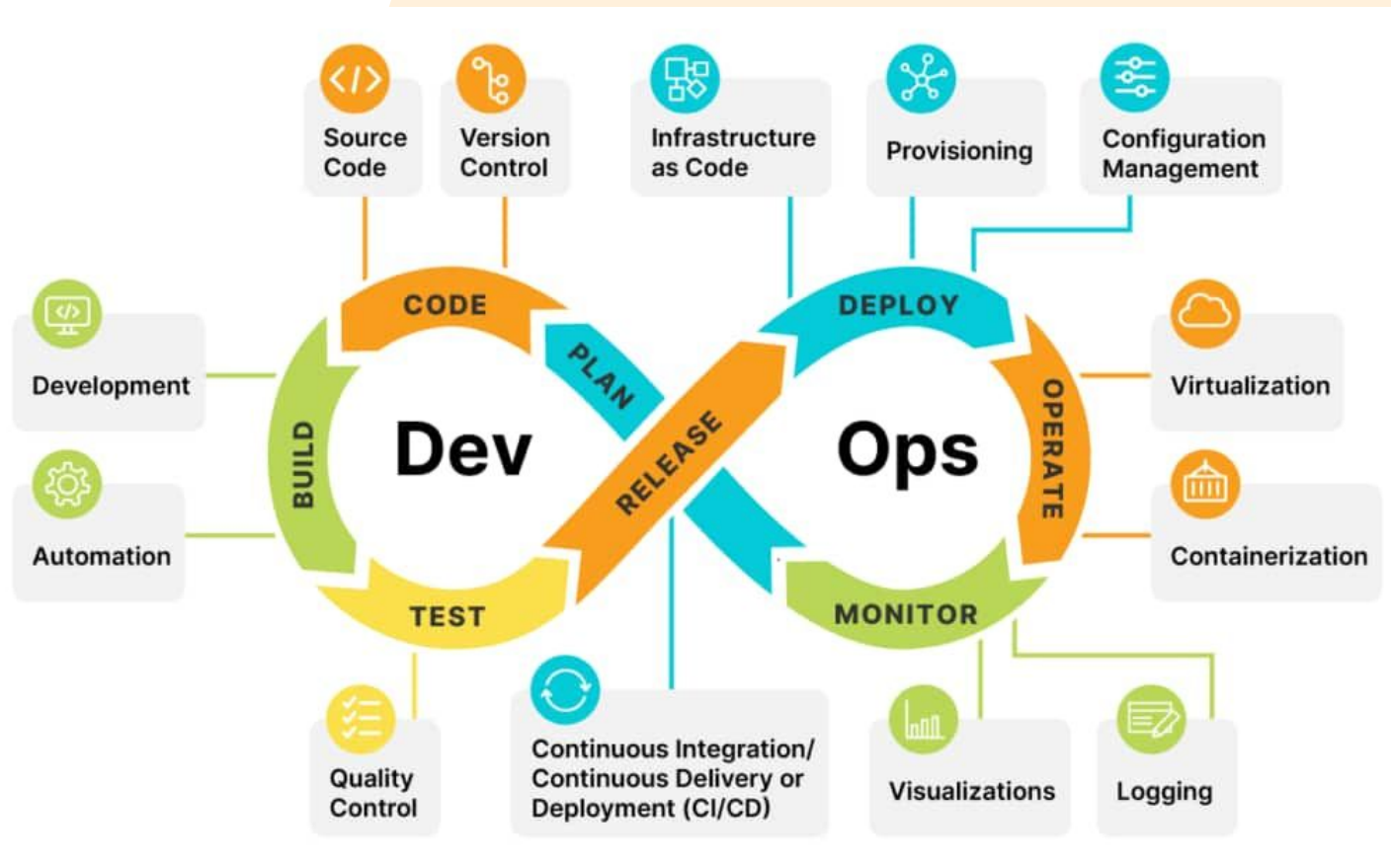
**DEVOPS**







# What is DevOps? In one word its **Simplification**

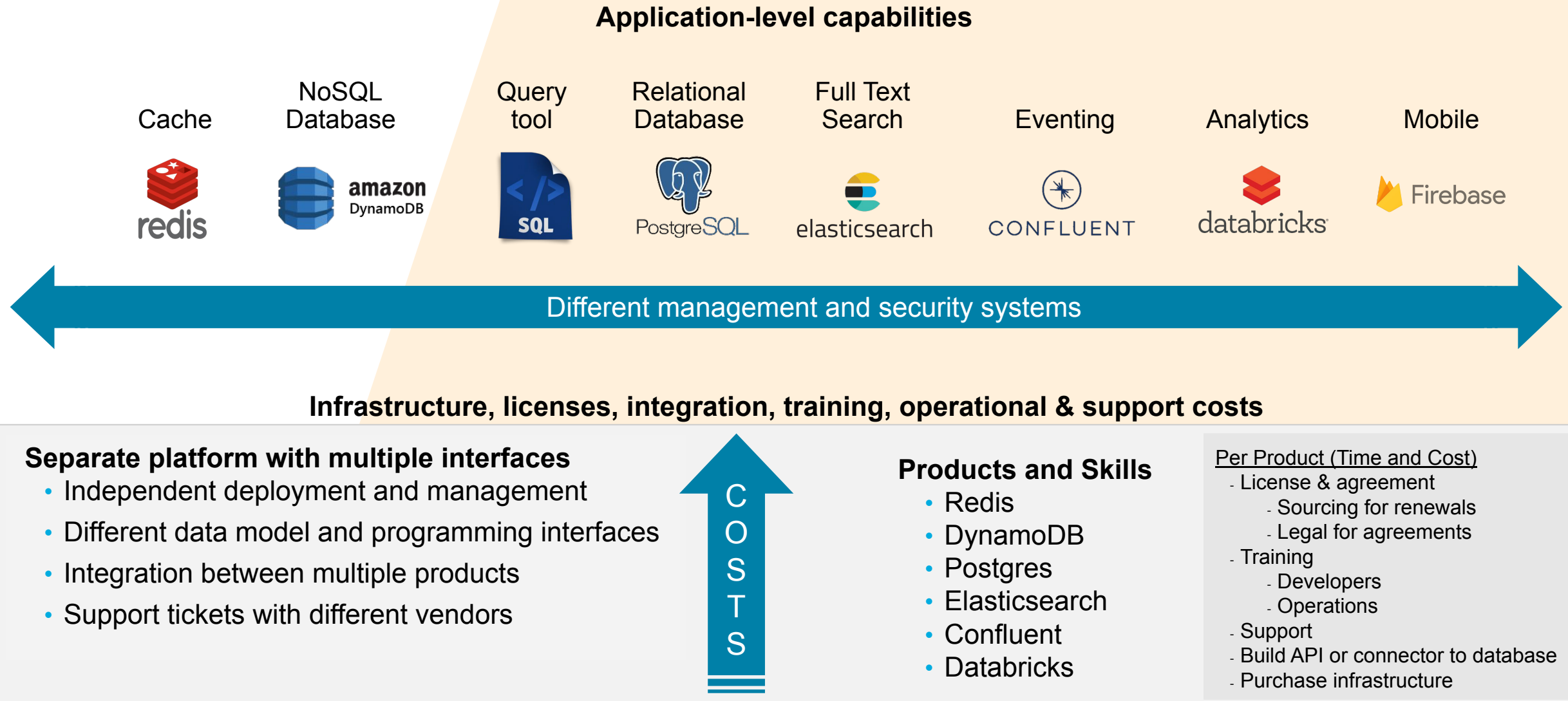


DevOps (a portmanteau of “development” and “operations”) is the combination of practices and tools designed to increase an organization's ability to deliver applications and services faster than traditional software development processes.

# How Couchbase addresses the complexity challenges ?



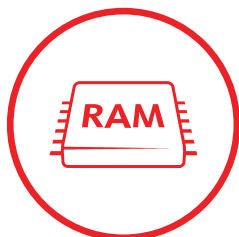
# Polyglot Persistence Design Architectures are Inefficient





# What is Couchbase

Performance &  
High Availability  
Architecture



SQL ,  
Full Text &  
Vector Search



Replication Across  
Data Centers



Workload  
Isolation



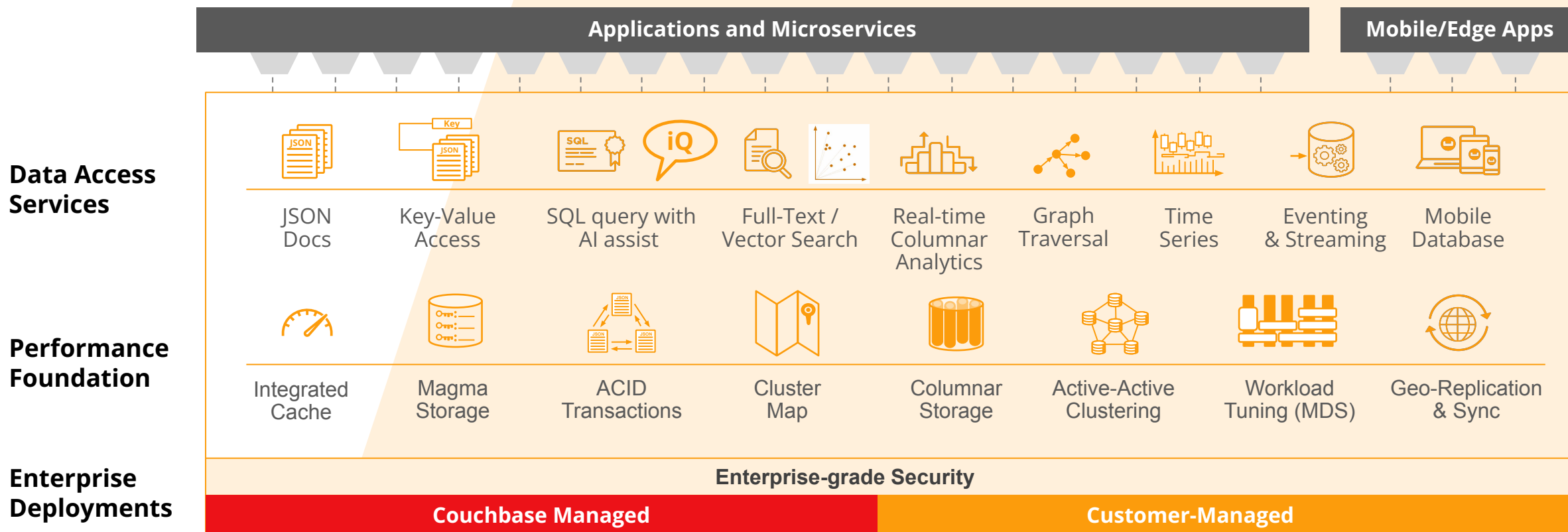
Cloud-agnostic  
& Containers



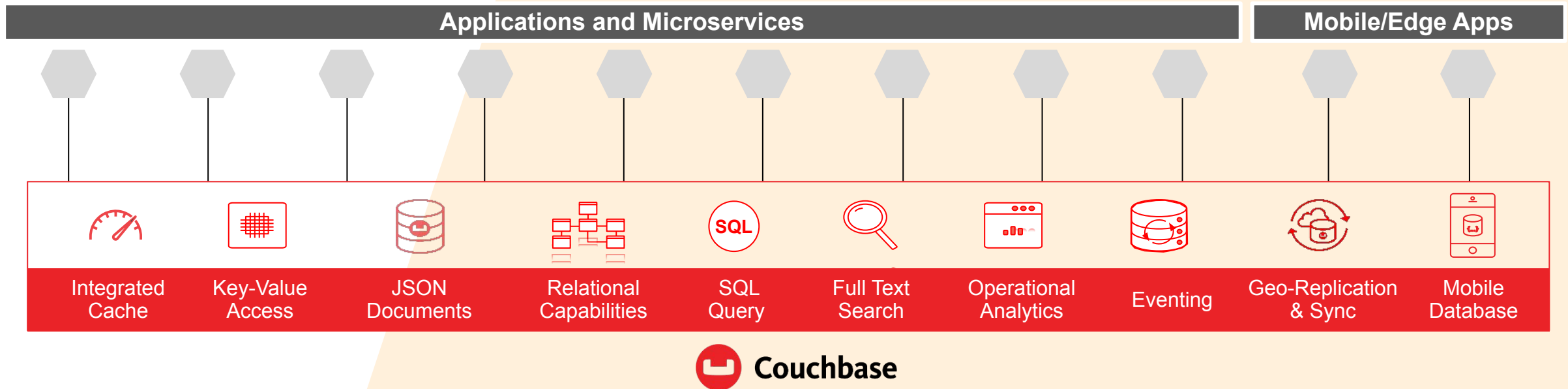
Offline-First  
Mobile



# Couchbase: Consolidate Data Sprawl With A Single Platform



# What is Needed to Achieve The Convergence Effect



## Speed

- Memory-first design
- Cloud-native scale
- Geo-replication via XDCR
- HA, DR & backup
- Low latency Cloud to Edge

## Flexibility

- JSON document
- Multimodel services
- Cloud deploy anywhere
- Mobile & Edge ready

## Familiar Skills

- SQL++ query language
- Dynamic Schema
- ACID SQL Transactions
- Cost-based optimizer
- SDKs for 12+ languages

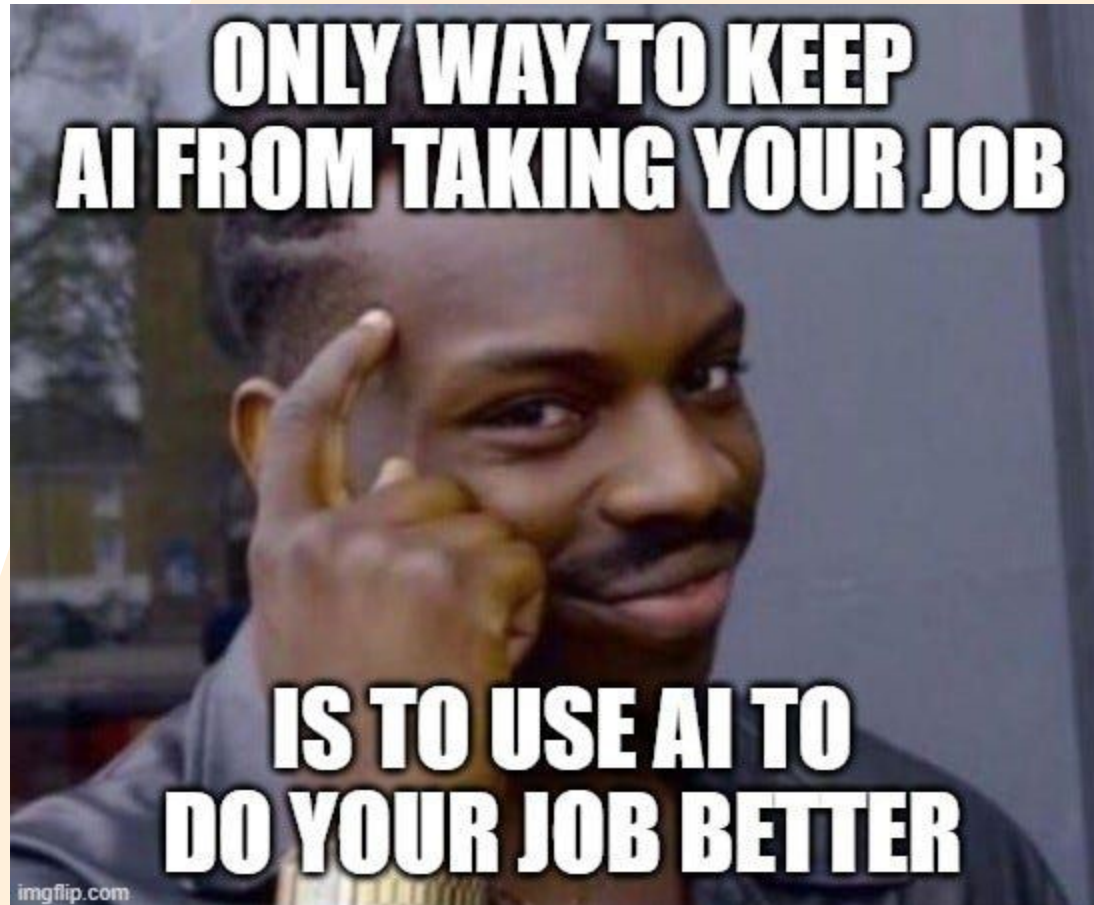
## Affordability

- Elastic scaling, sharding & rebalancing
- Multidimensional scaling
- High-density storage
- Incredible price/performance

# AI from Cloud to Edge

>





# AI-driven Adaptive Apps are the future



## Hyper Personalized

Accurate responses that fit the situation perfectly



## In context

Inputs: who, what, where, why, when, weather, real-time awareness



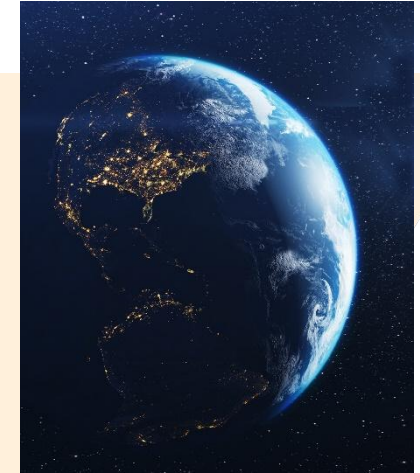
## Innovative

The best adaptations are fueled by trustworthy data



## Responsive

Latency is unacceptable



## Everywhere

Interactions are mobile





# Adaptive App use cases

## Content Generation

BLACK+DECKER 12-Cup Digital Coffee Maker, CM1160B, Programmable, Washable Basket Filter, Sneak-A-Cup, Auto Brew, Water Window, Keep Hot Plate, Black



### Customers say

Customers like the ease of use of the coffee maker. They say it's very simple to set and use. Customers are also satisfied with ease of cleaning, value, and speed. However, some customers have reported issues with drips. They mention that the inside will flood over with coffee grounds. Customers disagree on performance, quality, and temperature.

AI-generated from the text of customer reviews

## Data Analysis: Classification / Anomalies



## Advanced Semantic / Hybrid Search



## AI-powered Chatbots and Applications



# Security, Governance, Privacy, Control, Accuracy & Trust

How to Solve for These Issues

## Sharing proprietary and sensitive data



Vector Search and RAG techniques solve for sharing what the model needs without remembering it

*Why you need Vector Search*

## Sharing data that induces hallucinations

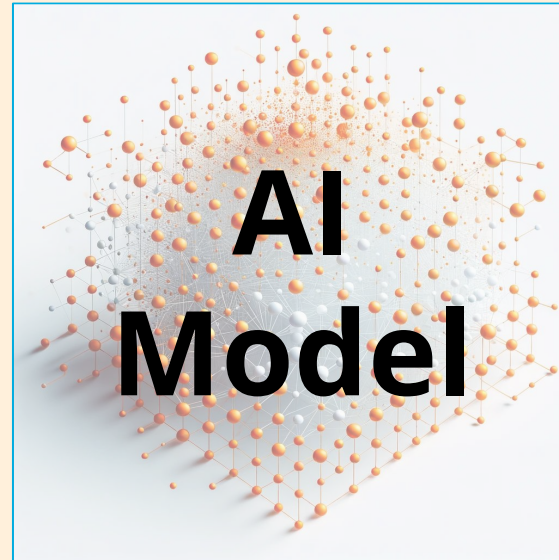


Simplifying your data architecture and building prompts from a single, trustworthy source, including Vector Search, dramatically improves response accuracy

*Why you need Couchbase*



# What are vectors?



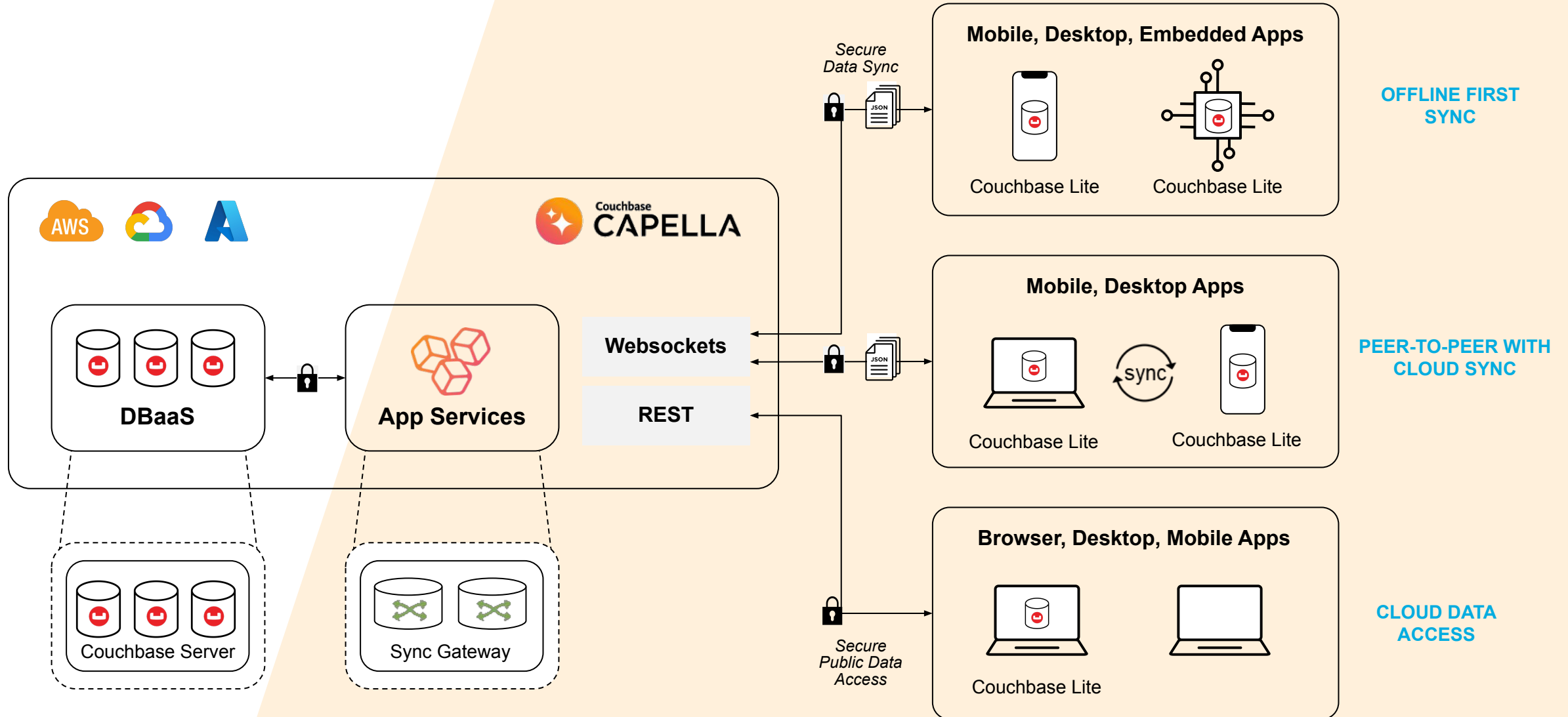
## Arrays of Floats

[0.9229, 0.7824, 0.194, 0.0402, 0.206,  
0.4446, 0.5875, 0.4214, 0.2371,  
0.4038, 0.9229, 0.7824, 0.194, 0.0402,  
0.206, 0.4446, 0.5875, 0.4214, 0.2371,  
0.4038]

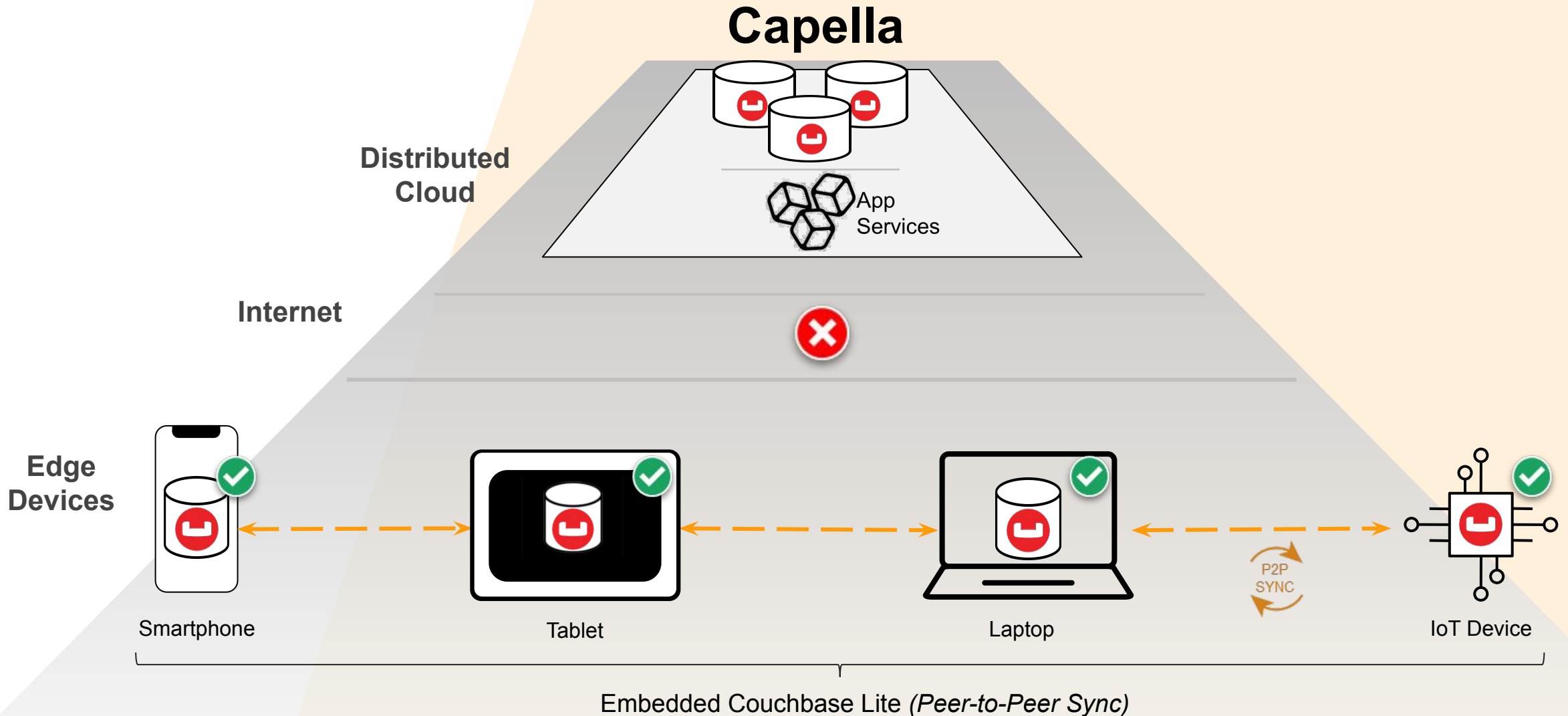
[0.0509, 0.8862, 0.179, 0.4116,  
0.5995, 0.769, 0.5282, 0.8079, 0.0468,  
0.272, 0.0509, 0.8862, 0.179, 0.4116,  
0.5995, 0.769, 0.5282, 0.8079, 0.0468,  
0.272]

[0.3819, 0.1375, 0.0828, 0.0856,  
0.9081, 0.1556, 0.6251, 0.7356,  
0.2964, 0.2198, 0.3819, 0.1375,  
0.0828, 0.0856, 0.9081, 0.1556,  
0.6251, 0.7356, 0.2964, 0.2198]

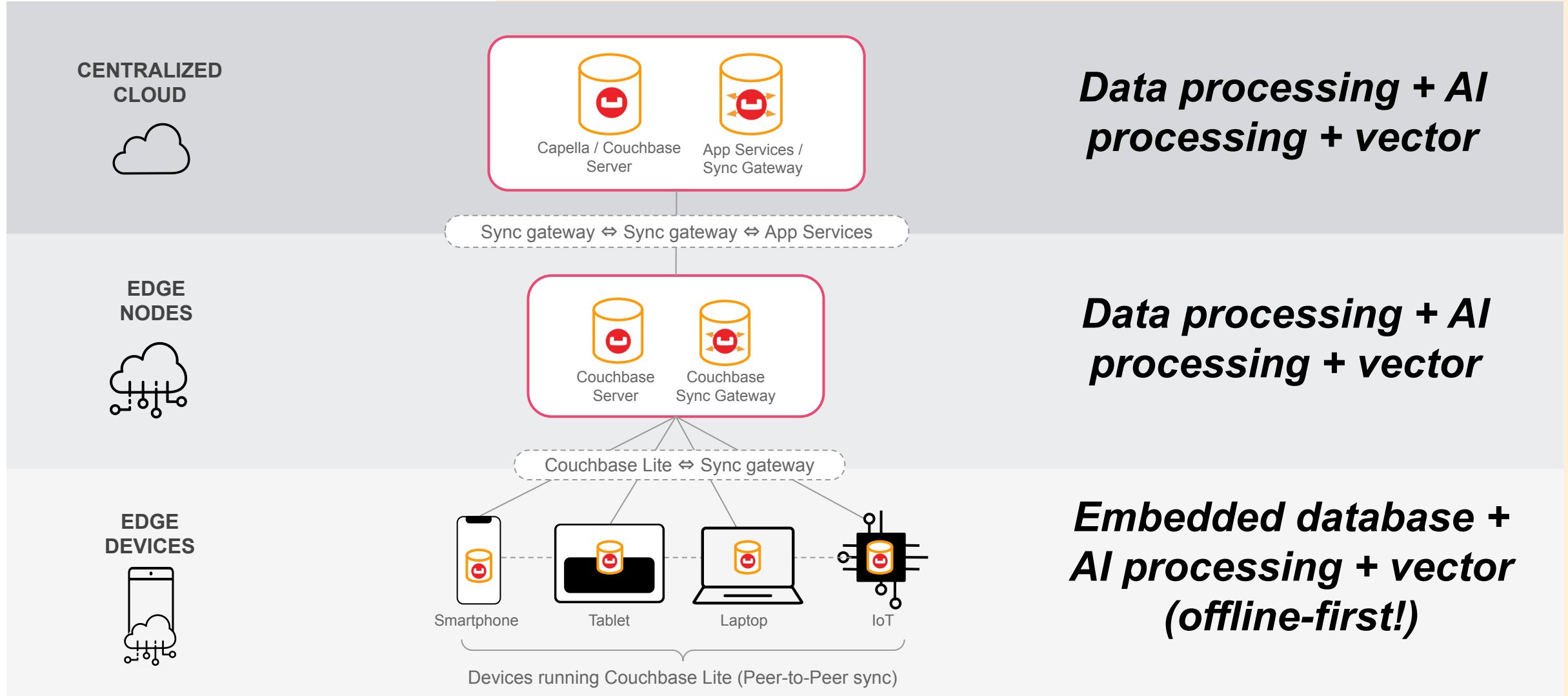
# Couchbase Mobile: common topologies



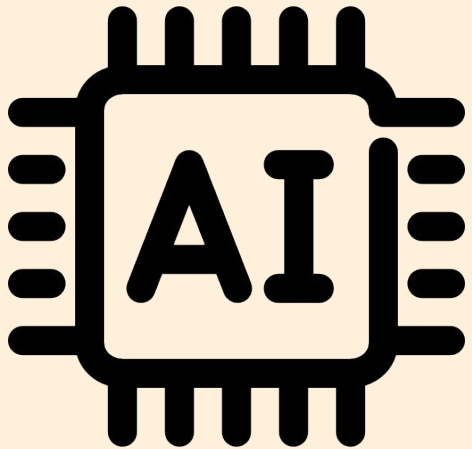
# Offline First: Eliminate Dependencies On the Internet



# Couchbase Mobile: cloud to edge AI



# Couchbase Capella Is AI-Ready



## COLUMNAR SERVICE

Aggregation of data across large data sets

Multiple sources: Capella, RDBMS, NoSQL, S3, Parquet

Fast processing for speedy ML training iterations

JSON accommodates multiple types and formats of data

Python UDF for calling trained ML models

## VECTOR SEARCH

Similarity

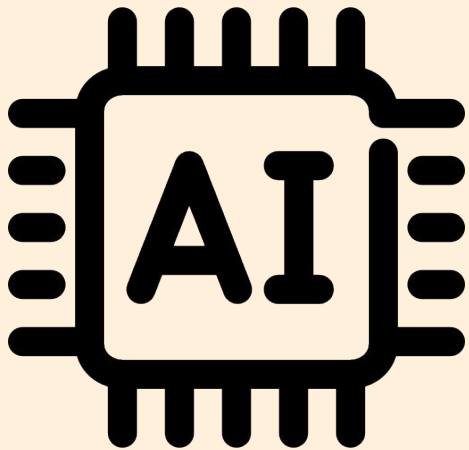
Semantic search

Retrieval Augmented Generation

Hybrid-search (vector, FTS, range, predicate, aggregations)

# Couchbase Lite Is AI-Ready

Embedded database  
for mobile and IoT  
devices



## PREDICTIVE QUERY API

Call optimized ML models to make predictions on Couchbase Lite data.

OFFLINE-FIRST

## VECTOR SEARCH

Similarity

Semantic search

Retrieval Augmented Generation

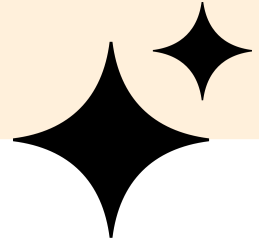
Hybrid-search (vector, FTS, range, predicate, aggregations)

OFFLINE-FIRST

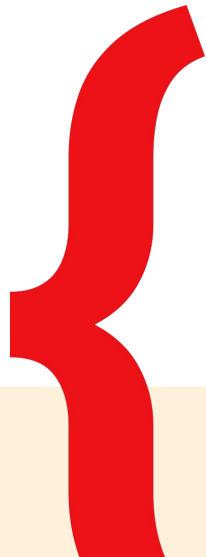
Demo

V





# Thank you!



**Couchbase**