Data Mesh or Data Mess?
Separating Reality from Hype

Donna Burbank
Global Data Strategy, Ltd.

March 23, 2023

DATA ARCHITECTURE STRATEGIES

Global Data Strategy
The Business of Data

Couchbase

The Modern Cloud Database Platform for Distributed Applications

March 2023

The Market: Application and Database Needs Have Changed

Applications

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
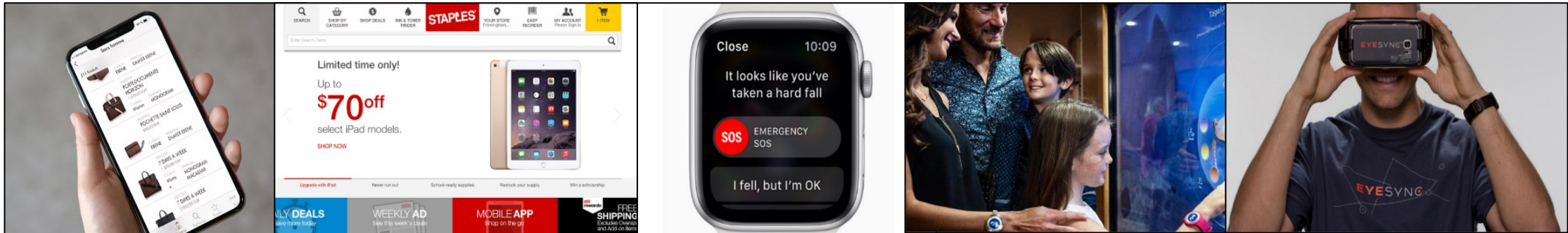
Databases

Legacy Expectations
- Schema & SQL familiarity
- ACID transactions
- Consistency and integrity
- Storage efficiency

+

Modern Demands
- Massive data and dynamic users
- Scale & performance
- Flexibility for agile development
- Microservices architecture
- Web, mobile & IoT experiences

Helping Customers Innovate Faster

Performance

Databases are failing modern demands

Flexibility

Significantly improve agility

Mobile/IoT

Deliver incredible experiences

Cloud cost

Drive down cloud TCO

Speed Up Legacy

Profiles & Personalization

Operational Analytics

Product Catalog

Customer 360

Mobile & IoT Applications
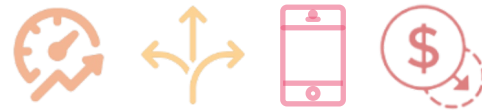
Industry & Innovation Leaders

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
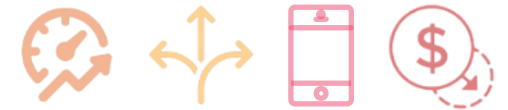
100's M
operations per/day

Linked in

High performance caching for social network
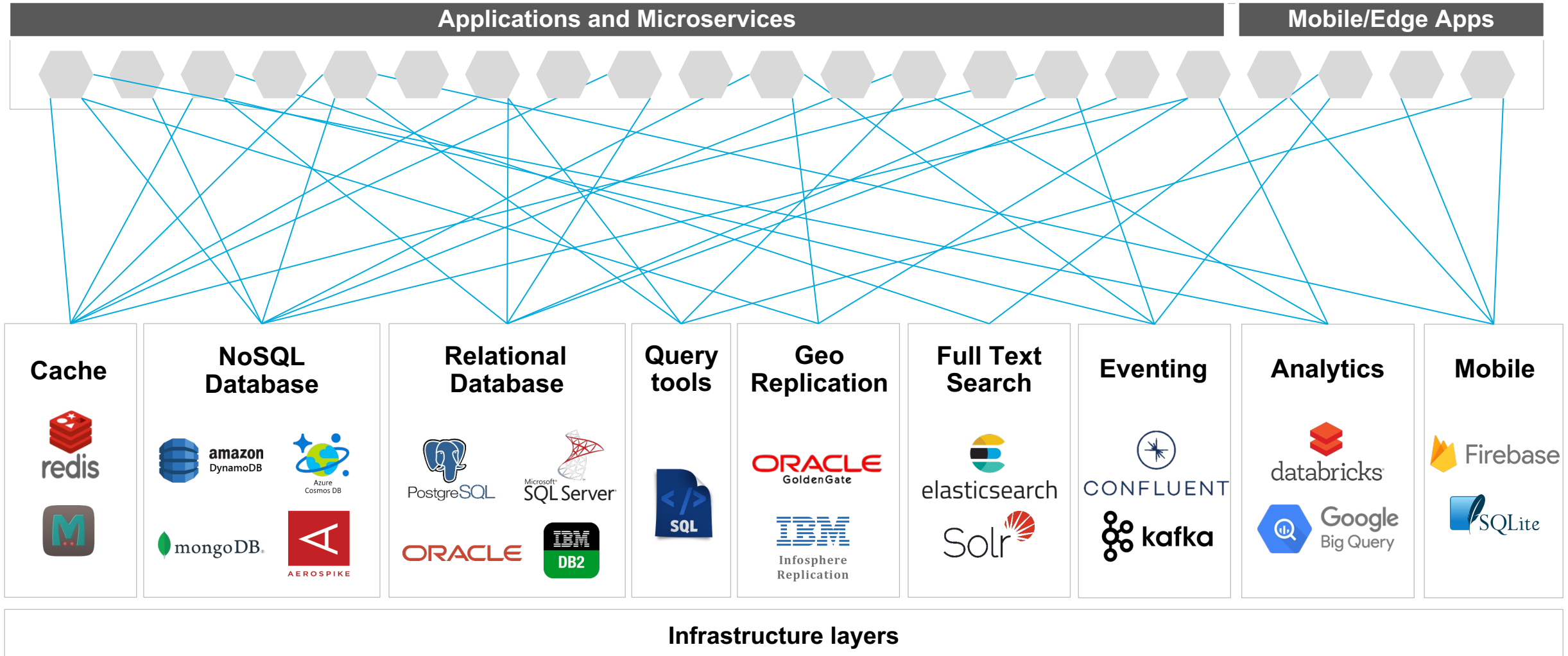
10M+
requests / second

800M+
users

STAPLES

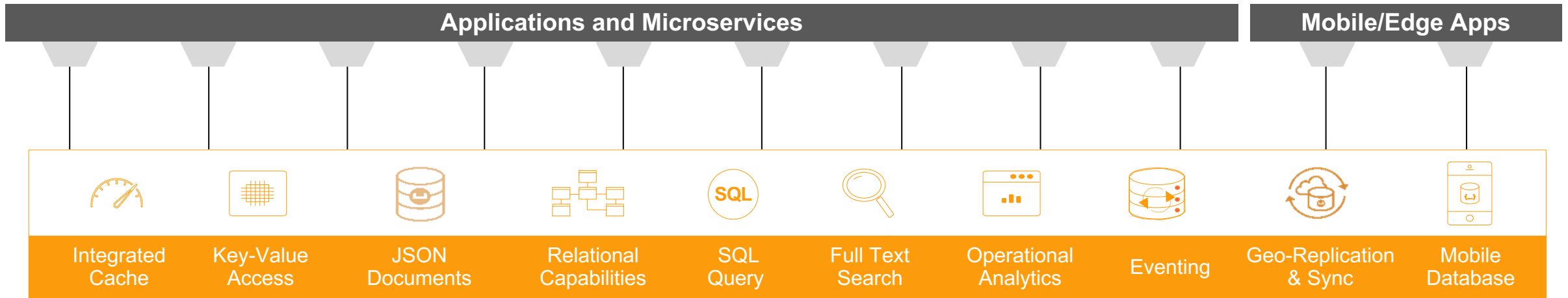Product catalog, pricing, shopping cart, daily deals, inventory management

4,000
stores worldwide

1.6B
rules applied in real-time

Reality: Data Sprawl, Messy Meshes & Management Challenges

Couchbase Capella is a Consolidated Database Platform

Applications and Microservices	Mobile/Edge Apps

Integrated Cache	Key-Value Access	JSON Documents	Relational Capabilities	SQL Query	Full Text Search	Operational Analytics	Eventing	Geo-Replication & Sync	Mobile Database

Couchbase CAPELLA

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

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

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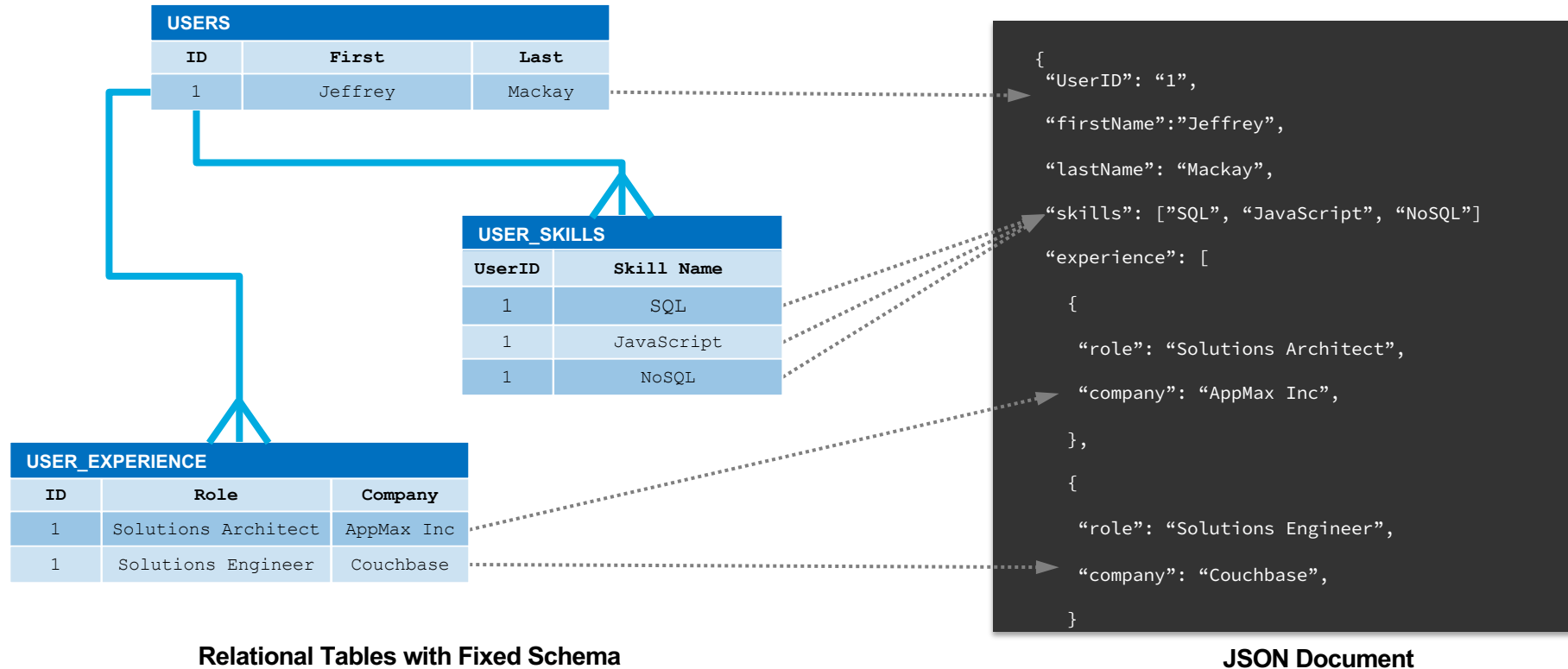- Cost-based optimizer
- SDKs for 12+ languages

Affordable
- Elastic scaling, sharding & rebalancing
- Multidimensional scaling
- High-density storage
- Incredible price/performance that lowers TCO

JSON is Flexible, its Structure can be Changed by Everyone

JSON makes changing dynamic user profiles and complex catalog entries very easy

USERS

ID	First	Last
1	Jeffrey	Mackay

USER_SKILLS

UserID	Skill Name
1	SQL
1	JavaScript
1	NoSQL

USER_EXPERIENCE

ID	Role	Company
1	Solutions Architect	AppMax Inc
1	Solutions Engineer	Couchbase

Relational Tables with Fixed Schema

```
{
 "UserID": "1",

 "firstName":"Jeffrey",

 "lastName": "Mackay",

 "skills": ["SQL", "JavaScript", "NoSQL"]

 "experience": [

  {

   "role": "Solutions Architect",

   "company": "AppMax Inc",

  },

  {

   "role": "Solutions Engineer",

   "company": "Couchbase",

  }
```

JSON Document

Store data in logical ways:
- Denormalized single document
- Normalized with references
- Add new values when needed
- Support for binary values
- As relational schema using Scopes & Collections
- Include metadata about the data

Access Data in multiple ways:
- Direct Key-Value
- SQL++ querying
- Full-Text Search
- MPP analytics
- As Events

Unlike document databases, RDBMSs store JSON as a column datatype, therefore it is still attached and bound to the schema, and updating JSON values is…"a PITA," and limited to the JSON extensions added to the vendor's query language.

Unlike relational databases, the application developer controls the structure of the data (and its metadata) rather than the DBA who owns the relational schema, thereby dictating the application's structure.

SQL++ is EASY   SQL for JSON Query Language

Easy as SQL, leveraging skillsets: Developers, Architects, DBAs, Analysts, IT

- Known query syntax
- Supporting
  - ANSI standards
  - JOINs
  - Sub-queries
  - Nested objects
  - Arrays
- Distributed ACID Transactions
- User Defined Functions
- MPP Analytics

Relational SQL

```
SELECT ac.industry,
  SUM(CASE WHEN a.activitytype = 'Task'
    THEN 1 ELSE 0 END ) task,
  SUM(CASE WHEN a.activitytype
='Appointment'
    THEN 1 ELSE 0 END ) appts
FROM crm.activity a
  INNER JOIN crm.account ac
    ON (a.accid = ac.id)
WHERE a.startdate BETWEEN '2018-10-01'
    AND   '2018-12-31'
GROUP BY ac.industry
```

Couchbase - SQL++

```
SELECT ac.industry,
  SUM(CASE WHEN a.activityType = 'Task'
    THEN 1 ELSE 0 END) task,
  SUM(CASE WHEN a.activityType =
'Appointment'
    THEN 1 ELSE 0 END ) appts
FROM crm a
  INNER JOIN crm ac ON a.accid = ac.id
    AND ac.type='account'
WHERE a.type='activity'
  AND a.startDate BETWEEN '2018-10-01'
    AND '2018-12-31'
GROUP BY ac.industry
```

MongoDB Query Language

```
db.activity.aggregate(
{ $match: { startDate: { $gt: '2018-01-01',
                $lt: '2018-12-31' } } },
{$lookup: {
      from: "account",
      localField: "accid",
      foreignField: "id",
      as: "account_docs"
      }
},
{ $match: { "account_docs": { $ne: [] } } },
{ $unwind: "$account_docs" },
{
  $project: {
    item: 1,
    task: { $cond:{if:
      { $eq: ["$activityType", "Task"] }, then: 1,
        else: 0 } },
    appt: { $cond: { if:
      { $eq: ["$activityType", "Appointment"] },
then: 1,
        else: 0 } } }
  },
  {
    $group: {
      _id: "$account_docs.industry",
      tasks: { $sum: "$task" },
      appointments: { $sum: "$appt" }
  } }
);
```

Customers are Seeking Offline-first Applications

Field Services
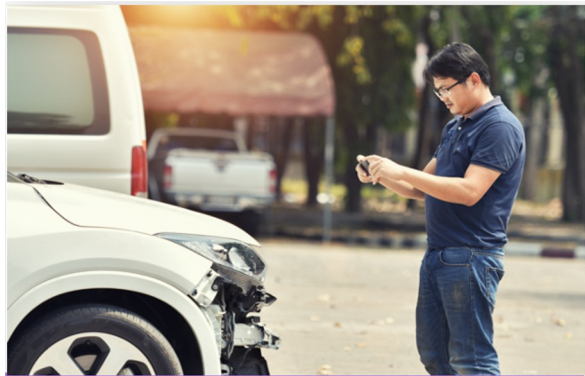
File claim reports, update customer policy information on mobile devices even with no network connectivity

Utility Workers

Access and update task lists, infrastructure maps, inspection reports, etc on handheld devices even with no network connectivity

Restaurants

Expedite ordering process and reduce wait times with order entry kiosks in fast food restaurants

Mobile Clinics

Access & update patient health records in rural areas even with no network connectivity

Q: How do you **guarantee speed and uptime** for apps especially when operating in internet dead zones?

Data Mesh at the Edge

- Bring data processing to the near side of the network — close to the users who need it
- Embed a database to edge clients, reducing dependencies on distant cloud data centers
- Sync data from the edge to the cloud for consistency and integrity
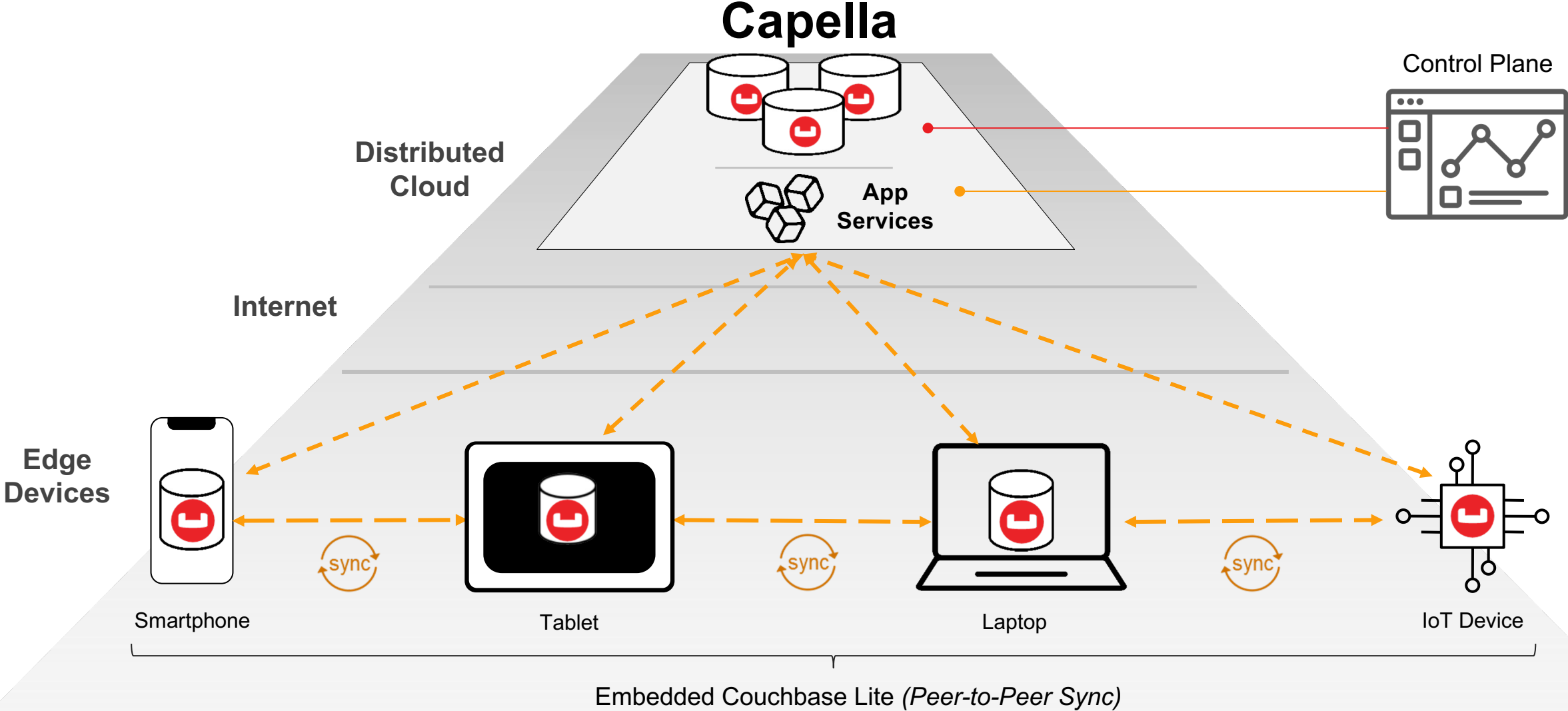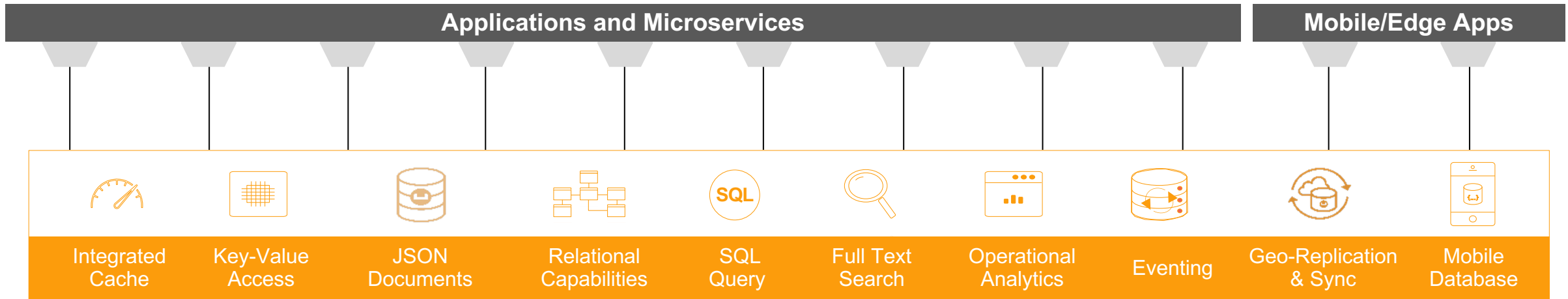
The Promise Of
Edge Computing

Speed

Reliability

Privacy

Architecture Overview

Capella

Control Plane

Distributed Cloud

App Services

Internet

Edge Devices

Smartphone

sync

Tablet

sync

Laptop

sync

IoT Device

Embedded Couchbase Lite *(Peer-to-Peer Sync)*

Integrated Capabilities: Benefits for Customers

Applications and Microservices								Mobile/Edge Apps	
Integrated Cache	Key-Value Access	JSON Documents	Relational Capabilities	SQL Query	Full Text Search	Operational Analytics	Eventing	Geo-Replication & Sync	Mobile Database

Couchbase CAPELLA

Innovate Faster & Spend Less

- Faster release cycles
- Less duplication & data sprawl
- Scale more easily

- Less time on maintenance
- Lower operational costs
- Lower infrastructure costs

- Easier data catalog management
- Easier analytic and AI integration
- Happier users

THANK YOU

Couchbase

Couchbase Capella: Key Capabilities

Fully-Managed Database

Enterprise-Grade Performance

Development Flexibility

High Availability
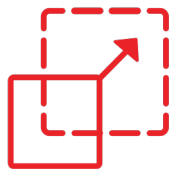
Automated Scaling In/Out & Up/Down

Single Pane for Multi-cluster, Multi-region

Mobile and IoT app synchronization
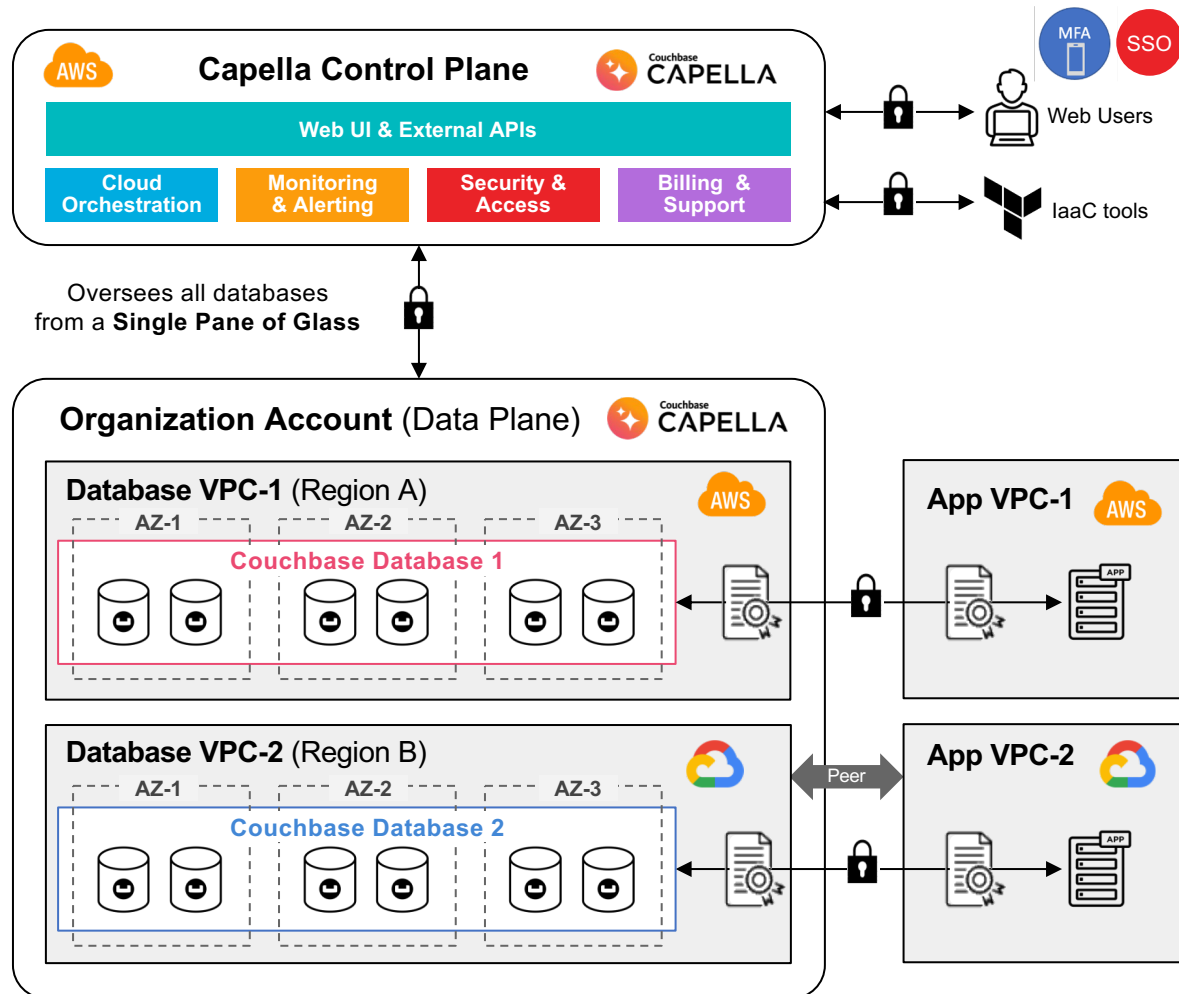
Security Peace of Mind

Couchbase Capella

A Better, Easier, more Flexible and Affordable Database as a Service

November 2022

Architecture | Overview

Fully Managed, Secure and Highly Available Infrastructure

Capella Control Plane

- **Manages** the Cloud Orchestration, Monitoring & Alerting, Security & Access, Billing & Support
- Is the Access Point for Organization **Web UI** Users
- Allows Infrastructure as Code Tools such as **Terraform**

Organization Account (Data Plane)

- **Account Isolation**: 1 Account per Organization
- **Cluster Isolation**: 1 VPC per Couchbase Database
- **Multiple Clouds**: AWS, Google and Azure

Applications

- **Connect** directly to Databases (Data Plane)
- **Multiple** connectivity options - Over Public Connection, through VPC Peering or PrivateLink
- All communications are **encrypted**

Couchbase Drives Down TCO

Customers' Reduced TCO

More than 100%	
75-100%	
50-74%	
25-49%	
Up to 25%	

57%

Customers who had
50%+ TCO reduction

[Tech Validate Survey]
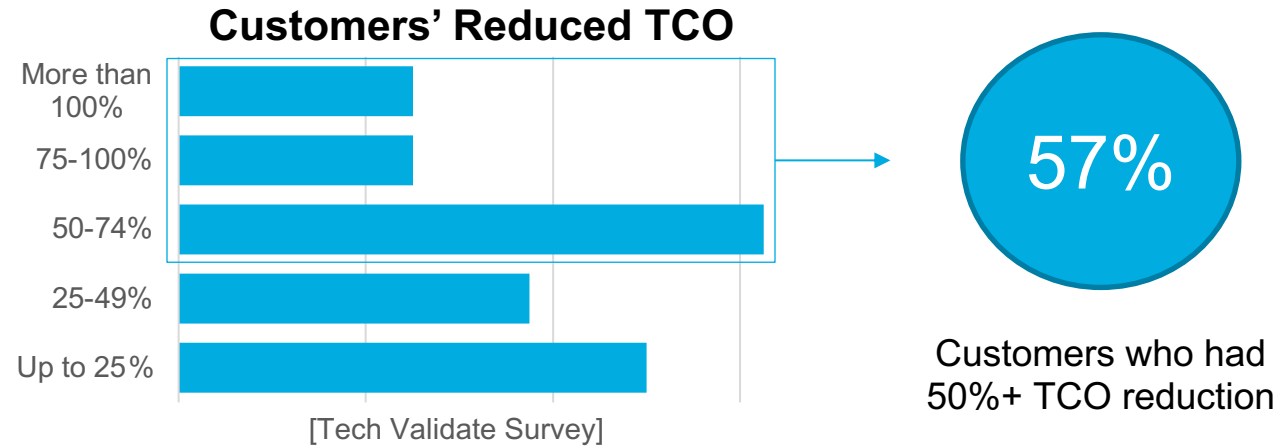
Cloud Advantages

- OPEX vs CAPEX
- Add servers as needed
- Centralized management

Single Platform

- Reduce license & hardware
- Less to learn, code, integrate
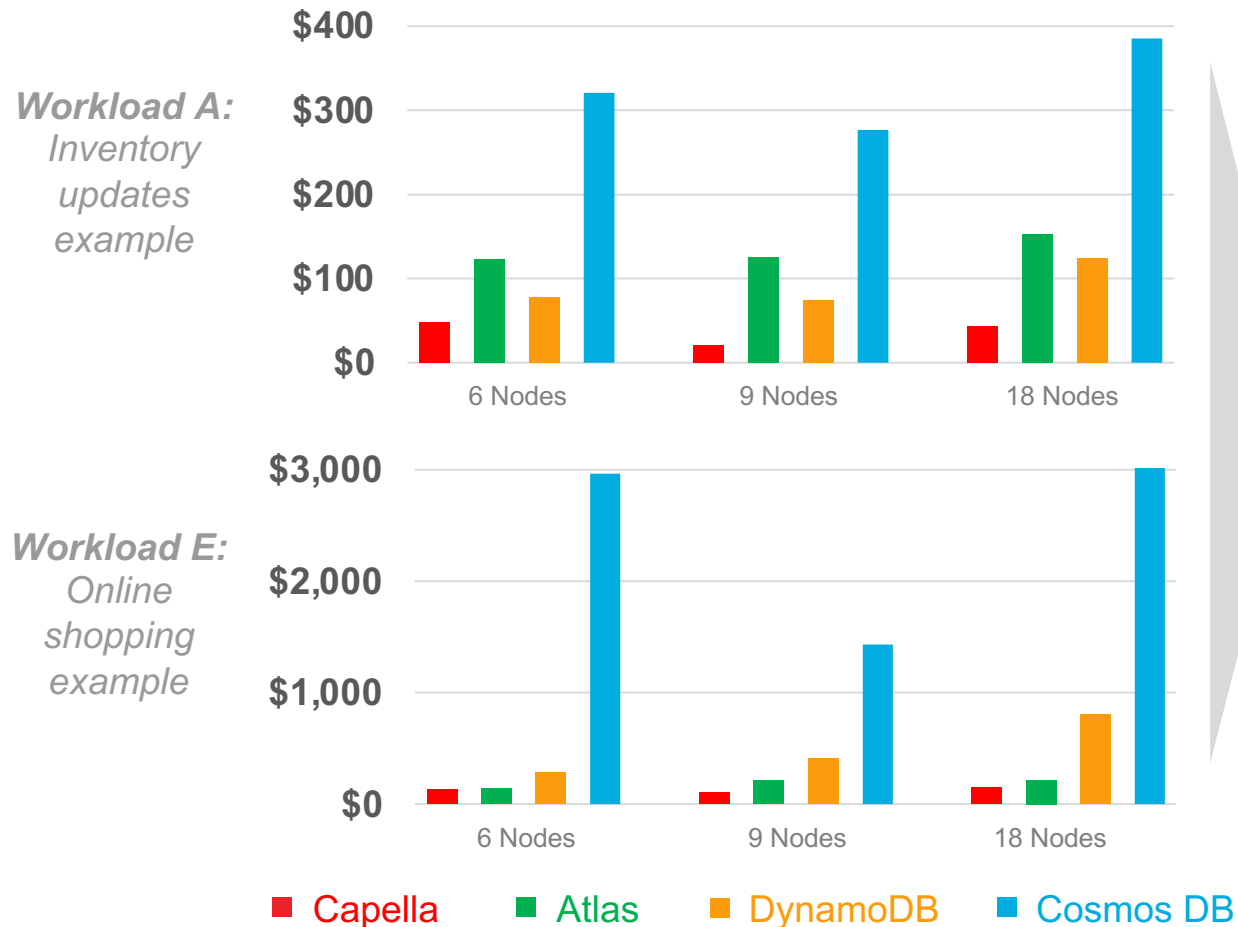- Leverage SQL skillsets

Performance at Scale

- Optimize services
- Easy scale up/down/in/out
- Dollar for dollar winner

Industry-Leading Price Performance: Cost of 1B Operations

Third party YCSB benchmarks

Workload A: Inventory updates example

Workload E: Online shopping example

■ Capella ■ Atlas ■ DynamoDB ■ Cosmos DB

Capella Delivers on Average

- **2-12X** faster throughput
- **10-50X** lower latency
- **50-500%** cost savings

(Node for Node)

"At BroadJump, we help healthcare providers control expenses, lower service costs, and operate efficiently.

With Couchbase we were able to dramatically improve our own TCO by reducing our storage needs by over 50%."

Philip Lupercio, VP of technology at BroadJump

Advantages of Couchbase for You

Flexibility of NoSQL fused
with strengths of RDBMS

A simple platform to design
sophisticated, scalable applications

Rapidly develop, test and deploy
via cloud-native automation

Developers

- **Familiar and flexible**: SQL for JSON
- **Structured**: Multi-level data controls
- **Dependable:** Transaction support
- **Programmable:** 10 SDKs and UDFs
- **Fast**: In-memory design
- **Multi services**: Build more, faster
- **Distributed**: Mobile, geographic & edge

Architects

- **Multi services**: Improve time to market
- **Dependable:** ACID transactions
- **Structured**: Multi-level data organization
- **High performance & high availability**
- **Cloud-native:** Kubernetes
- **Deployability**: However, Wherever

DevOps

- **Self-managed in the cloud**
  - Very high degree of control
  - Flexibility
- **Fully-managed in the cloud**
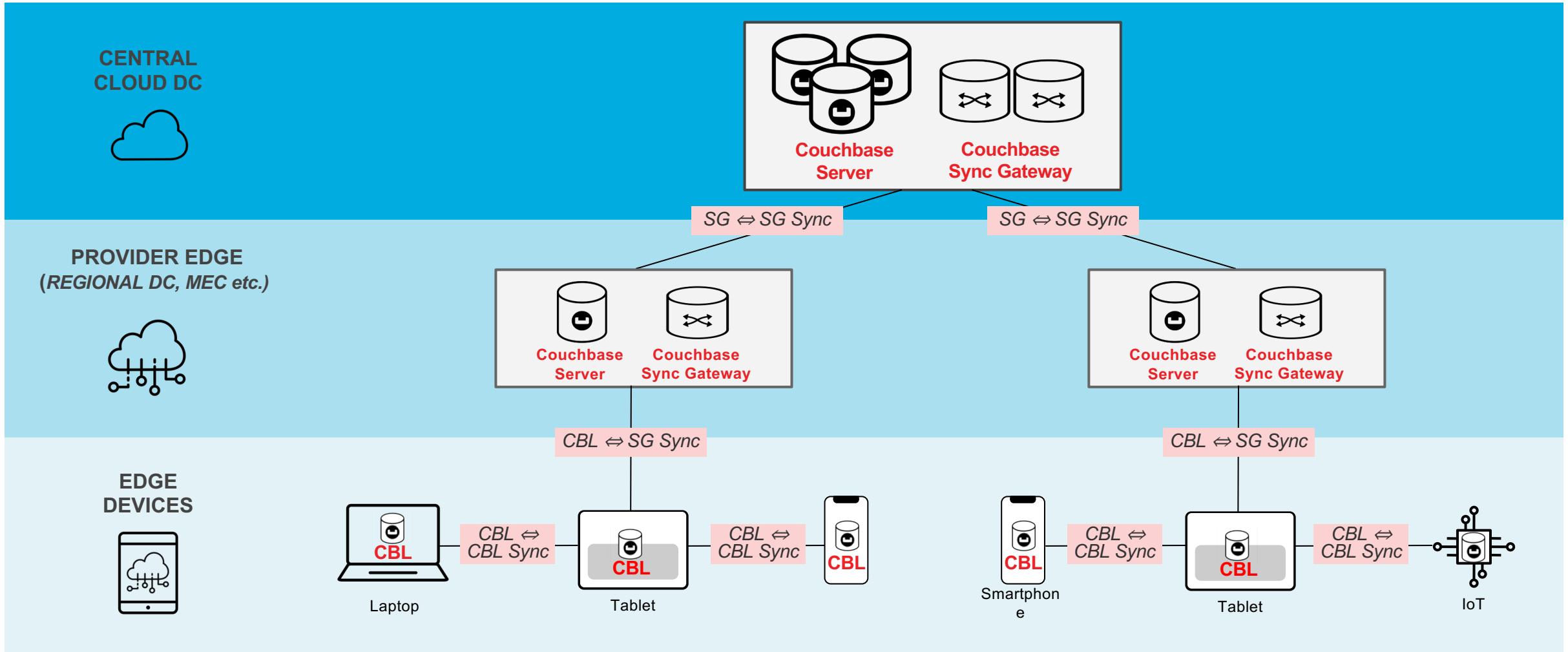 - For convenience
  - Speed to start
  - Smaller teams

Architecture Overview with Multidimensional Scaling
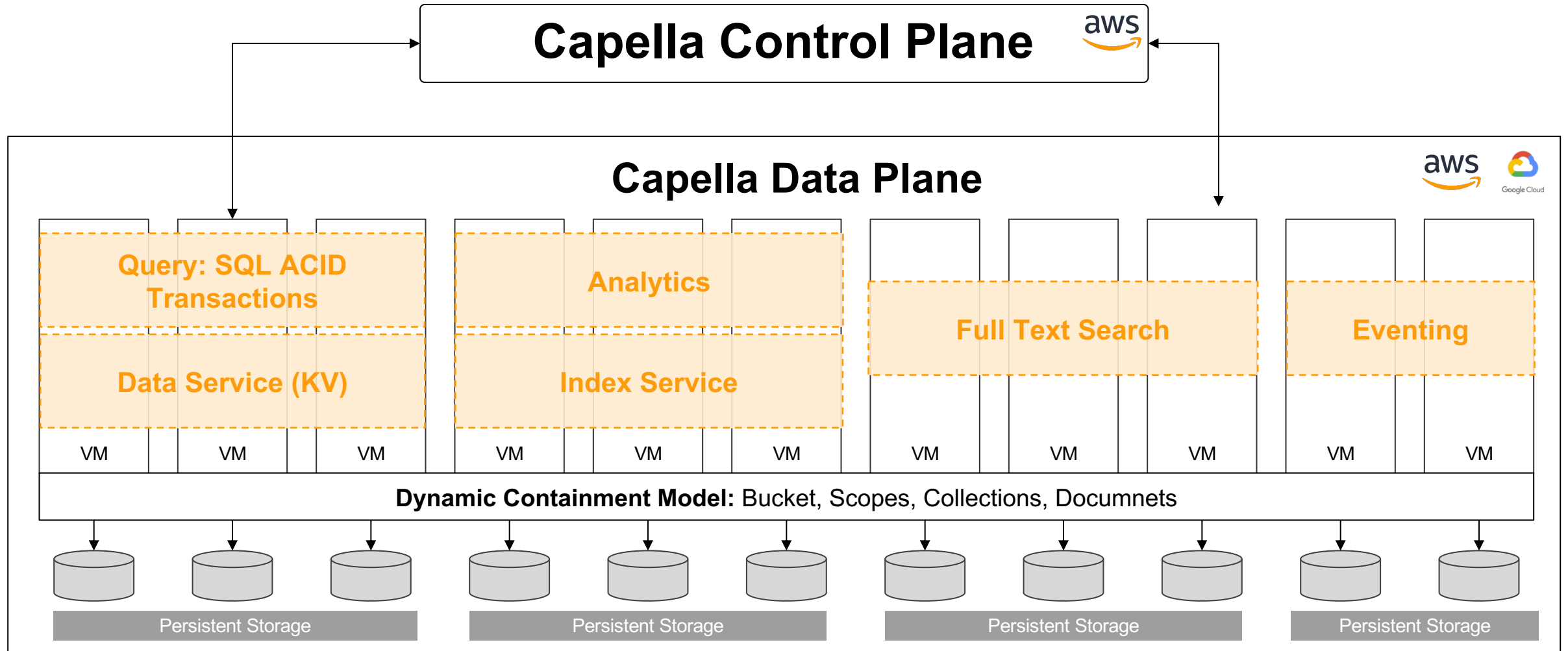
Capella App Services
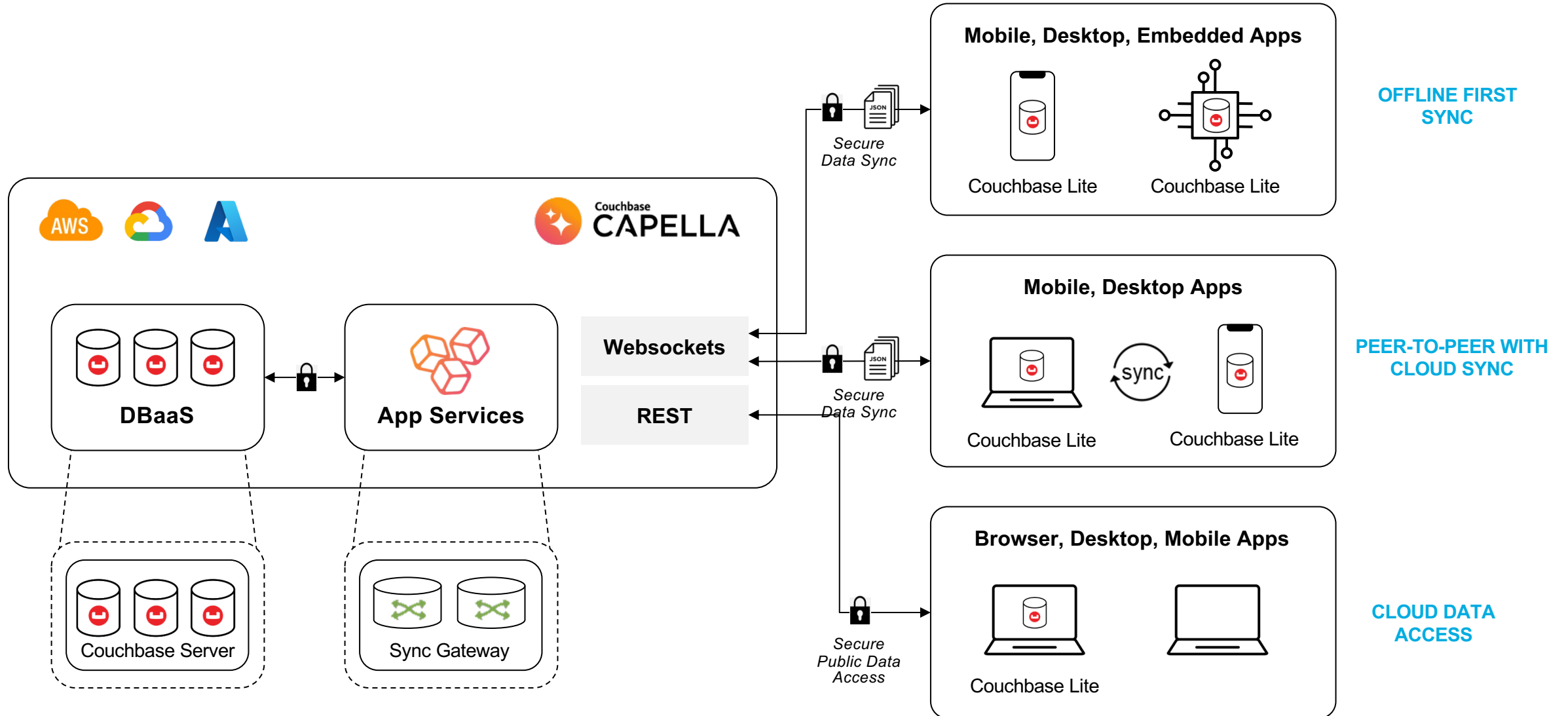
Peer-to-Peer Sync at the Edge

Data Observability

Unlocking the Power of Your Data

Shane Murray

MC

Shane Murray

Field CTO @ Monte Carlo (2022–)
SVP Data & Insights @ The New York Times (2013–2021)

MC MONTE CARLO

$236m in funding by backers of the world's best enterprise companies,

including **Accel, GGV, Redpoint, ICONIQ, and IVP**

Creator of the **Data Observability category**
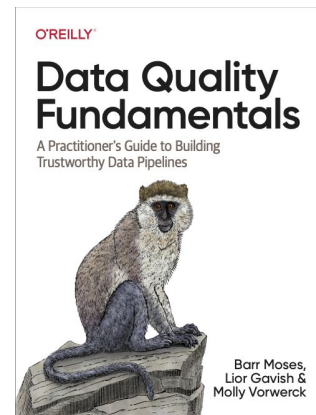
Why is data downtime a problem?

Can't see downstream

Can't predict all the ways data will break

Can't know when data is bad or who to ask for help

Data producers
(source systems)

Data platform
(warehouses, lakes)

Data consumers
(BI tools, data products)
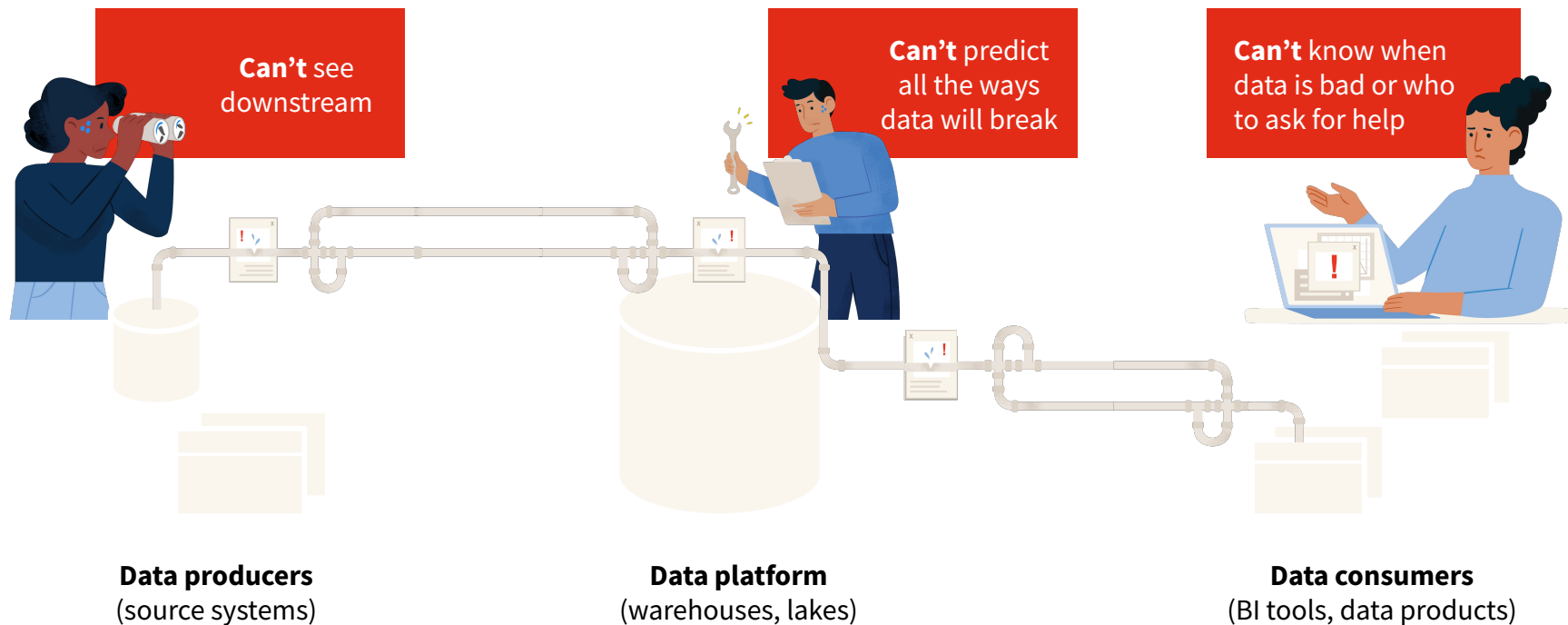
How we started:
Manual tests were the only means to ensuring data quality

Testing

Detected by downstream data consumers

Typical distribution of data incidents

10%

90%
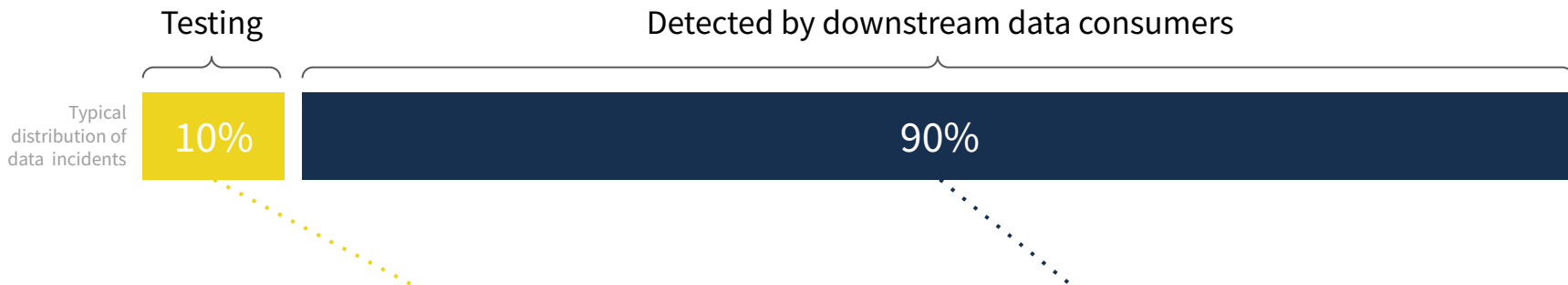
Common tests

- Column must have 0% nulls

- Transaction price must be >$0

- …and other predictable anomalies

Why testing falls short

- Can't test what you can't predict

- Can't see downstream

- Direct queries become costly

- Data and stack are constantly changing

- Doesn't scale

MC

Consequences of data downtime range from trivial to existential
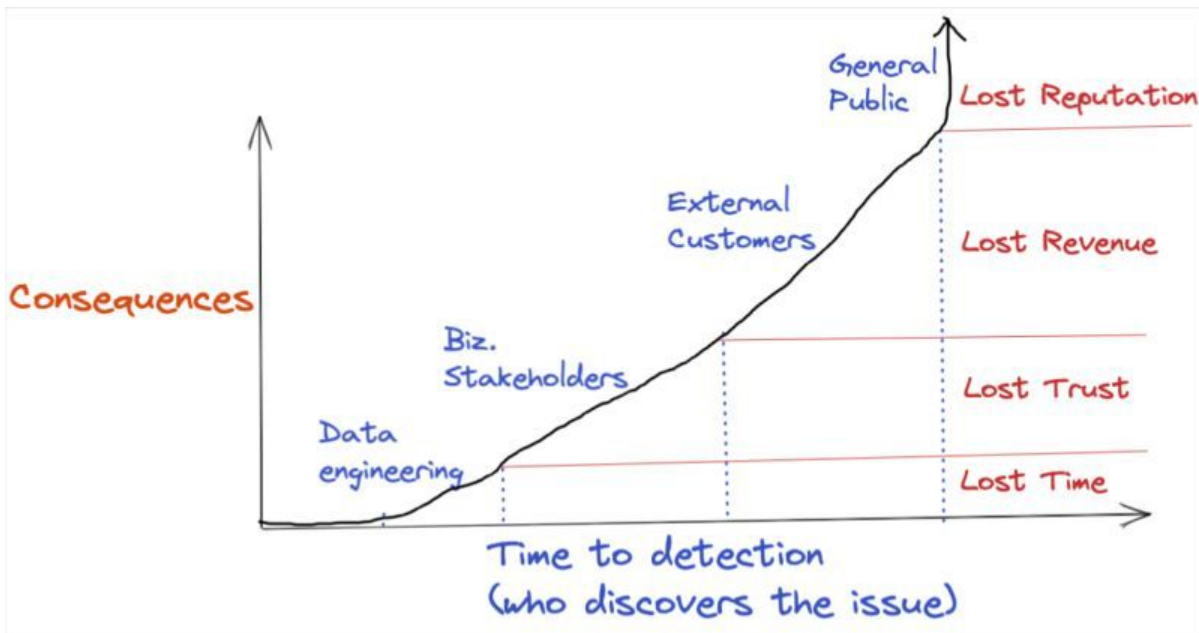
~70

high severity events each year per every 1k tables [1]

30-50%

data engineering time spent on fire drills [2]

12-27%

avg. annual revenue lost for companies resulting from poor data quality [3]

1. Benchmark data based on Monte Carlo customer production deployments
2. Monte Carlo market research and customer-reported benchmarks
3. Experian Data Quality; MC research via Wakefield Research Survey

Data downtime looks similar at all companies

- Is this data up-to-date?

- Why does this data size look off?

- Isn't this value suspiciously high?

- Why are there so many nulls?

- Why do we have duplicate IDs?

- What reports will I break with this schema update?

- Why are there 0s on tiles that usually show 100s?

  …

DATA OBSERVABILITY PILLARS

Freshness

Volume

Quality

Schema

Lineage

What is Data Observability?

Detect

- ML-powered anomaly detection
- Rule-based detection
- Targeted alerts to impacted owners & downstream users

Resolve

- Automated field-level lineage
- Impact radius assessment
- Code, data, and operational diagnostics

Prevent

- Auto-generated and on-demand insights
- Schema change notifications
- Automated circuit breakers

DATA OBSERVABILITY PILLARS

Freshness | Volume | Quality | Schema | Lineage

Observability equips your team with automated monitoring across 100% of production tables and critical fields

Powered by ML

1 Freshness, volume & schema monitoring **on all production tables**

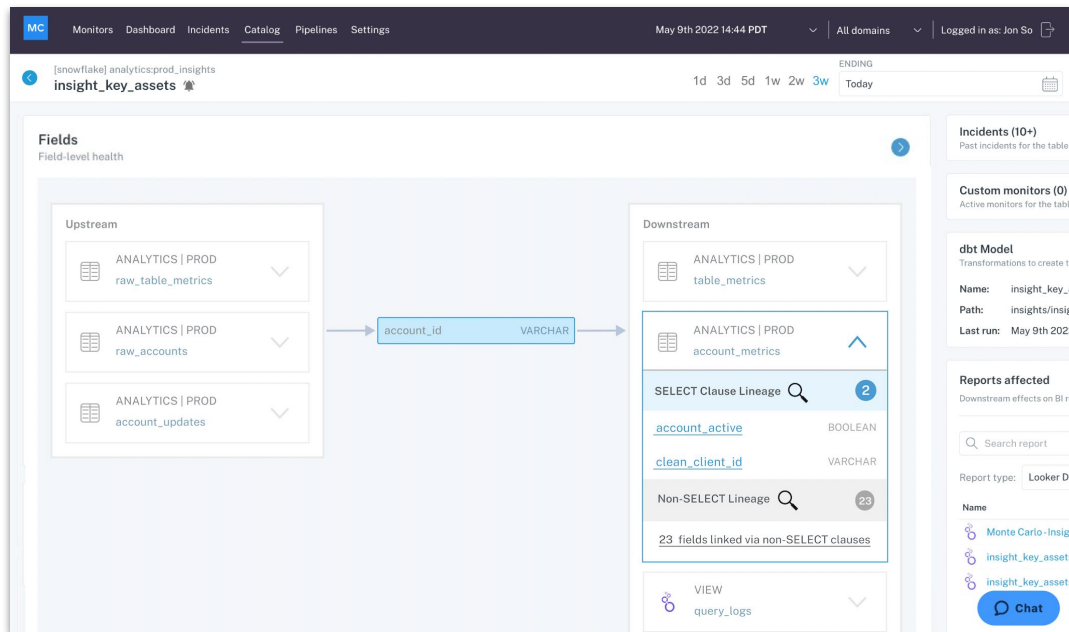**2** Data quality monitoring **for critical fields**
% null, median, p20-p80, and 20+ metrics

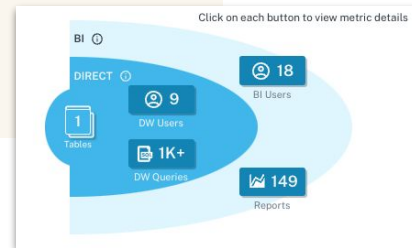**3** Custom rules & tests **for critical tables & fields**

Data warehouse, lake, or lakehouse

MC
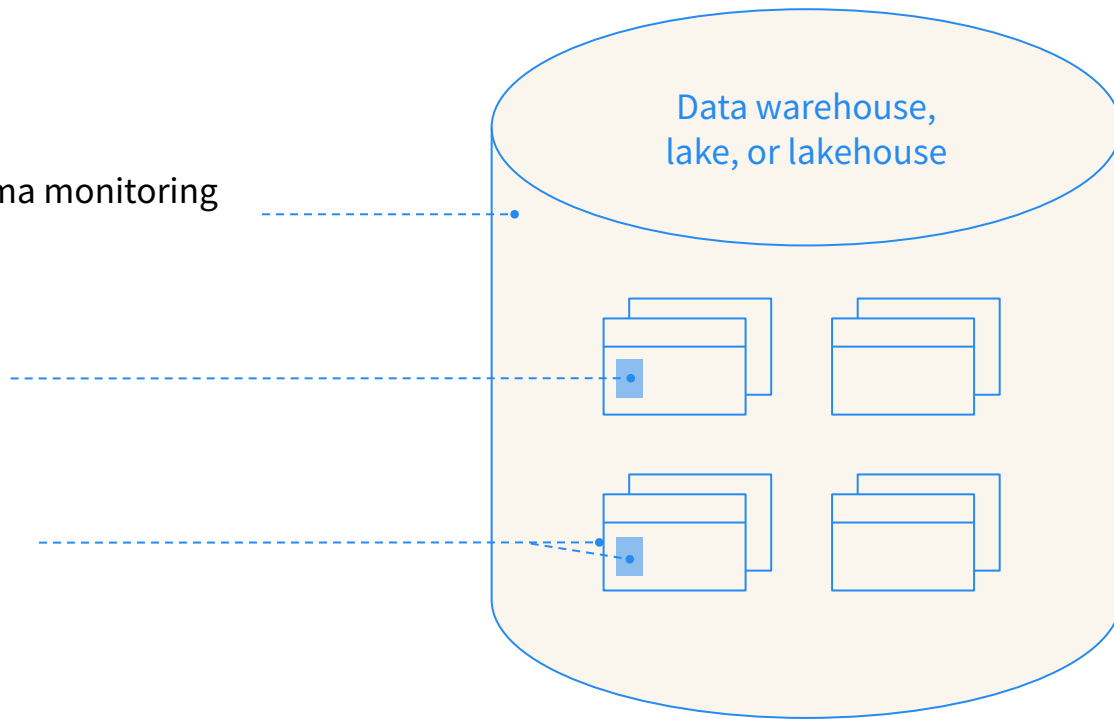
Triage, find root cause, and manage incident workflows
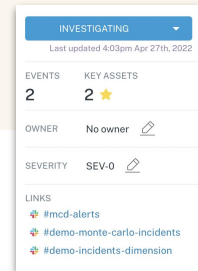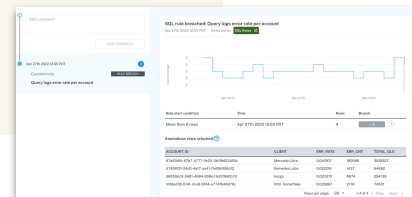
O'Reilly's
Data Quality Fundamentals

Learn what it takes to build more reliable data pipelines at scale.

https://www.montecarlodata.com/oreilly-data-quality-fundamentals-early-release/
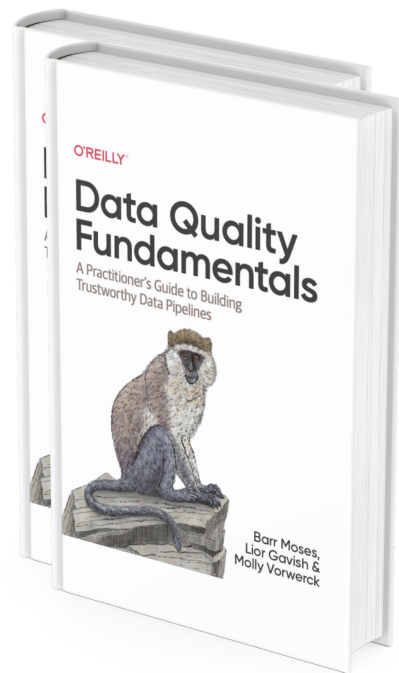
Thank you

smurray@montecarlodata.com

MC

Donna Burbank

Donna is a recognised industry expert in data management with over 25 years of experience in data strategy, data governance, data modeling, metadata management, and enterprise architecture. Her background is multi-faceted across consulting, product development, product management, brand strategy, marketing, and business leadership.

She is currently the Managing Director at Global Data Strategy, Ltd., an international data management consulting company that specializes in the alignment of business drivers with data-centric technology.

In past roles, she has served in key brand strategy and product management roles at CA Technologies and Embarcadero Technologies for several of the leading data management products in the market.

As an active contributor to the data management community, she is a long time DAMA International member, contributor to the DMBOK 2.0, Past President and Advisor to the DAMA Rocky Mountain chapter, and was awarded the Excellence in Data Management Award from DAMA International.

She has worked with dozens of Fortune 500 companies worldwide in the Americas, Europe, Asia, and Africa and speaks regularly at industry conferences. She has co-authored several books and is a regular contributor to industry publications. She can be reached at donna.burbank@globaldatastrategy.com Donna is based in Boulder, Colorado, US.
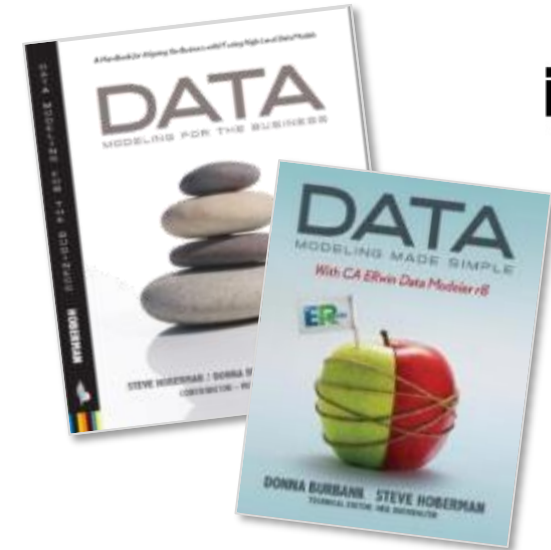
DATAVERSITY Data Architecture Strategies

This Year's Lineup

- **January**   Emerging Trends in Data Architecture – What's the Next Big Thing?

- **February**   Building a Data Strategy - Practical Steps for Aligning with Business Goals

- **March**   Data Mesh or Data Mess? Separating the Reality from the Hype

- **April**   Master Data Management - Aligning Data, Process, and Governance

- **May**   How do Data Governance & Data Architecture Support Each Other?

- **June**   Why You Need Data Management – Getting Executive Buy-In

- **July**   Artificial Intelligence and Machine Learning – Building the Right Architectural Foundation

- **August**   Data Quality Best Practices (with Nigel Turner)

- **September**   Best Practices in Metadata Management

- **October**   Designing Data for Business Intelligence & Analytics – Where the Star Schema Fits in a Modern Data Architecture

- **December**   Enterprise Architecture vs. Data Architecture

What We'll Cover Today

- Discussions around **data mesh and data fabric** are ubiquitous in the data industry, and data mesh is often touted as the next "Big Thing"

- Are these promises well founded?

- How does a **data mesh approach fit into today's modern Data Architecture**, and how does it align with more fundamental approaches such as Master Data, Data Quality, and Data Governance.

- This webinar will provide a practical, real-world approach to data mesh and its alignment with other disciplines.

Let's Start with some Core Definitions

Data Mesh

Data mesh is a **cultural and organizational shift** for data management focusing on federation technology that emphasizes the authority of **localized data management**. Data mesh is intended to enable easily accessible data by the business.

Data assets are analyzed for usage patterns by subject matter experts, who determine data affinity, and then the **data assets are organized as data domains**.

Domains are contextualized with business context descriptors. **Subject matter experts use the patterns and domains to define and create data products.** Data products are registered and made available for reuse relative to business needs.

*- Gartner IT Dictionary*

Data Fabric

A data fabric is an **emerging data management design** for attaining flexible, reusable and augmented data integration pipelines, services and semantics.

A data fabric **supports both operational and analytics use cases** delivered across multiple deployment and orchestration platforms and processes.

Data fabrics support a combination of different data integration styles and leverage active metadata, knowledge graphs, semantics and ML to augment data integration design and delivery.

- Gartner IT Dictionary

Word Salad or Buzzword Bingo?
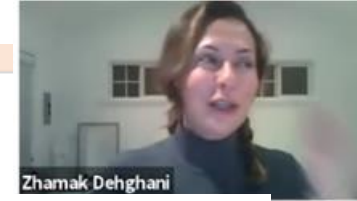
A different language is a different vision of life - Federico Fellini

"We need to create a new language" – Zhamak Dehghani

"Many of these concepts exists already but we need new names"

– Zhamak Dehghani

Data as a by-product

Data as a product

FROM		TO
IGESTING	▶	SERVING
EXTRACT | LOAD | ONBOARD	▶	DISCOVER | CONSUME | LINK
FLOW DATA THROUGH PIPELINES	▶	PUBLISH DATA VIA PORTS
CENTRALIZED LAKE | WAREHOUSE | PLATFORM		ECOSYSTEM OF DATA AS PRODUCTS

Never-changing

Rigid   Fragile

Monolithic   Broken

Orthogonal to the Axis Change

Disconnected Execution

Tension   Everyone is Unhappy

Full of Friction

Architectural Quantum

Polyglot app data ports

Higher-order value

Intelligently Empowered

Ecosystem

Bootstrapping

Trustworthy   Serving Data on the Same Plane

Delightful Experience   Convergence   Continuous Flow

Paradigm Shift

Lower cognitive load   Decentralized Socio-technical approach

Keynote by Zhamak Dehghani, Thoughtworks Australia 2021

Keep it Simple

Eschew Obfuscation

Data Mesh: Four Core Principles

Data mesh's objective is to create a foundation for getting value from analytical data at scale
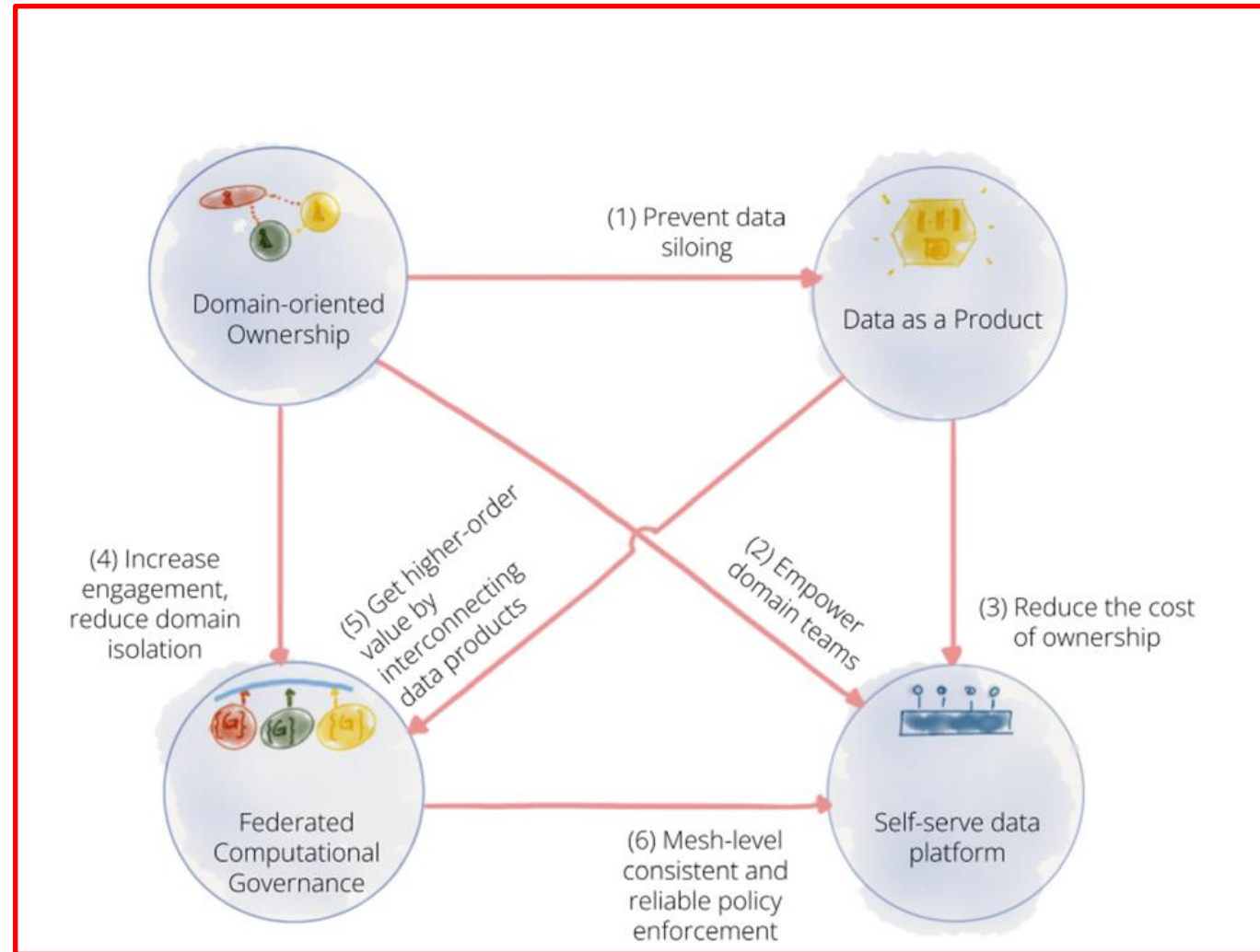
Domain Ownership

Data as a Product

Self-serve Data Platform

Federated Computational Governance

Source: 'Data Mesh: the Architecture Essentials', Zhamak Deghgani

Data Mesh: Domain Ownership

A Primary principle that data (particularly analytical data) is better organized and managed in Domains

- A domain can be:
  - An organizational unit (e.g. Human Resources, Sales)
  - Business function (e.g. Production, Supply Chain Management)
  - A logical data domain, as identified in a Business / Conceptual data model

- Each domain should have a formal Domain Data Owner
  - Enshrines the concept of devolved data ownership
  - Places data ownership on data producers

- A Domain should manage:
  - Operational Data (e.g. supplier details, customer name & address etc.)
  - Analytical Data (e.g. historical data, aggregated data etc.)
  - Metadata (e.g. business metadata, technical metadata etc.)
  - Computational capability needed to store, process and provide access to the data

Data Mesh definition of Domain:
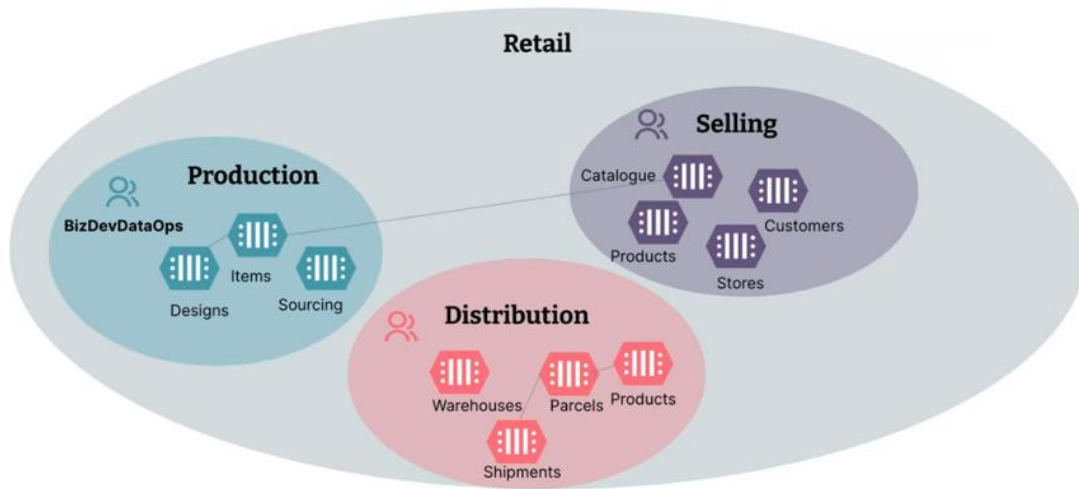" An aggregation of people organized around a common functional business purpose "

Potential Data Domains

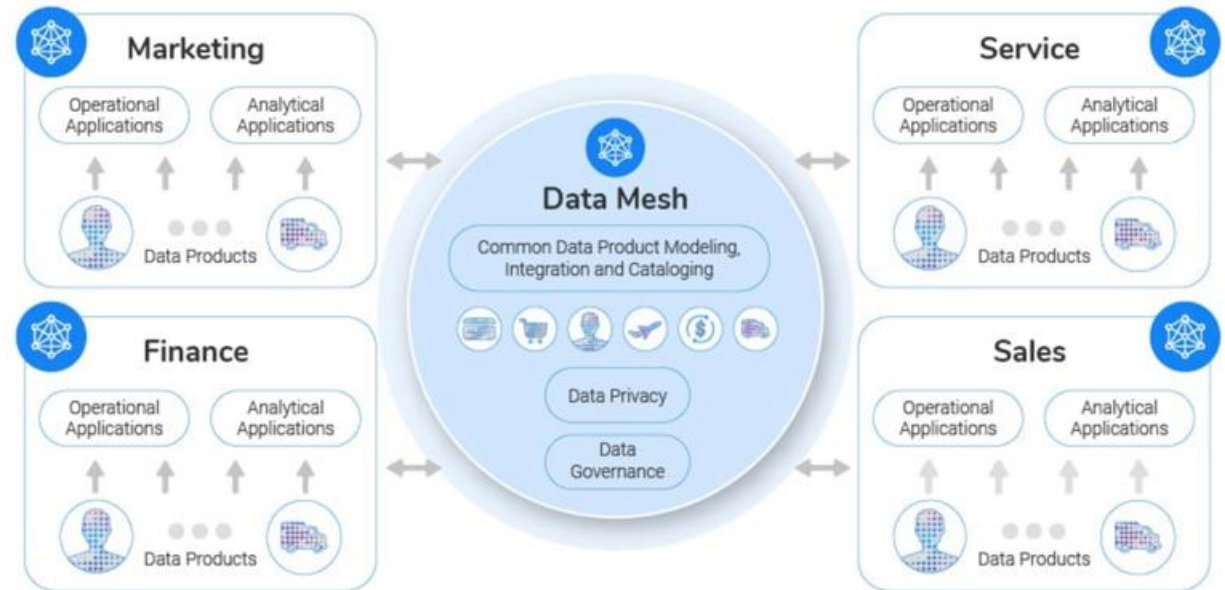|		
Customer	Sales	Supply Management
Marketing	HR	Product
Finance	Podcasts	Social Media

What is a Domain?

Domains deconstruct areas of the business...which is a loosely-defined concept.

Introduction to Data Mesh with Zhamak Dehghani, Stanford Deep Data Research Center

What is Data Mesh? A Market Primer' K2View

"Instead of flowing the data from domains into a centrally owned data lake or platform, domains need to host and serve their domain datasets in an easily consumable way."
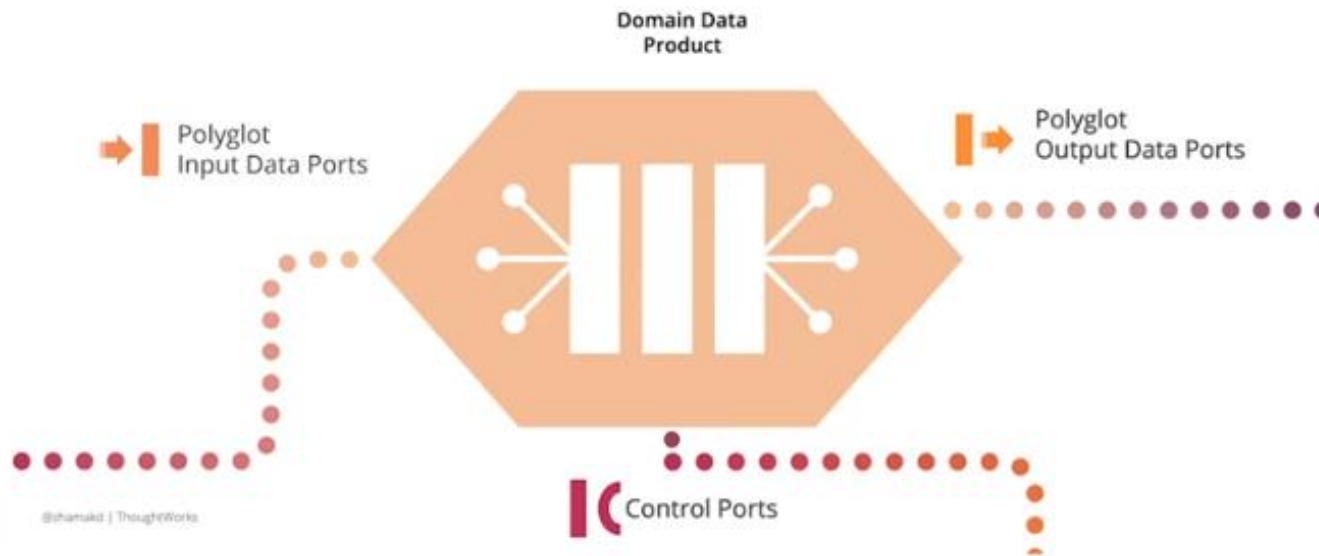
– Zhamak Dehghani

DATA PRODUCT IS the *ARCHITECTURE QUANTUM*

Always historical and read only access to data

Domain Data Product

Polyglot Input Data Ports

Polyglot Output Data Ports

Control Ports

@zhamakd | ThoughtWorks

Architectural quantum, as defined by Evolutionary Architecture, is **the smallest unit of architecture that can be independently deployed with high functional cohesion**, and includes all the structural elements required for its function.

This quantum includes:
1. Code
2. Data and metadata
3. Infrastructure

Code

Data product

Polyglot Data

Infrastructure
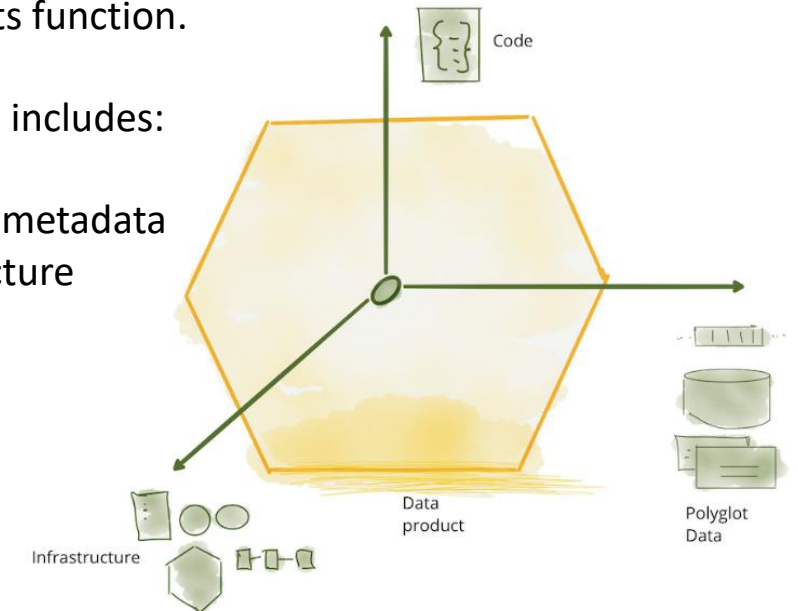
Data Mesh: Data as a Product

- The concept of Data as a Product aligns with a standard product lifecycle
  - Data is not seen as a by-product of operational processes but **data is the product**
  - Data should therefore be managed via a **Product Management Lifecycle**
  - Data consumers and users should be treated as **customers** so their needs are paramount
  - Domain customers can be internal or external (e.g. Open Data)

- Taking a Product Management approach will ensure each Domain produces data that is:
  - Well understood in terms of its scope, scale and lineage
  - High quality
  - Trusted by its customers
  - Accessible and discoverable through standard interfaces and publish & subscribe models
  - Meaningful, with appropriate business and technical context and metadata
  - Secure
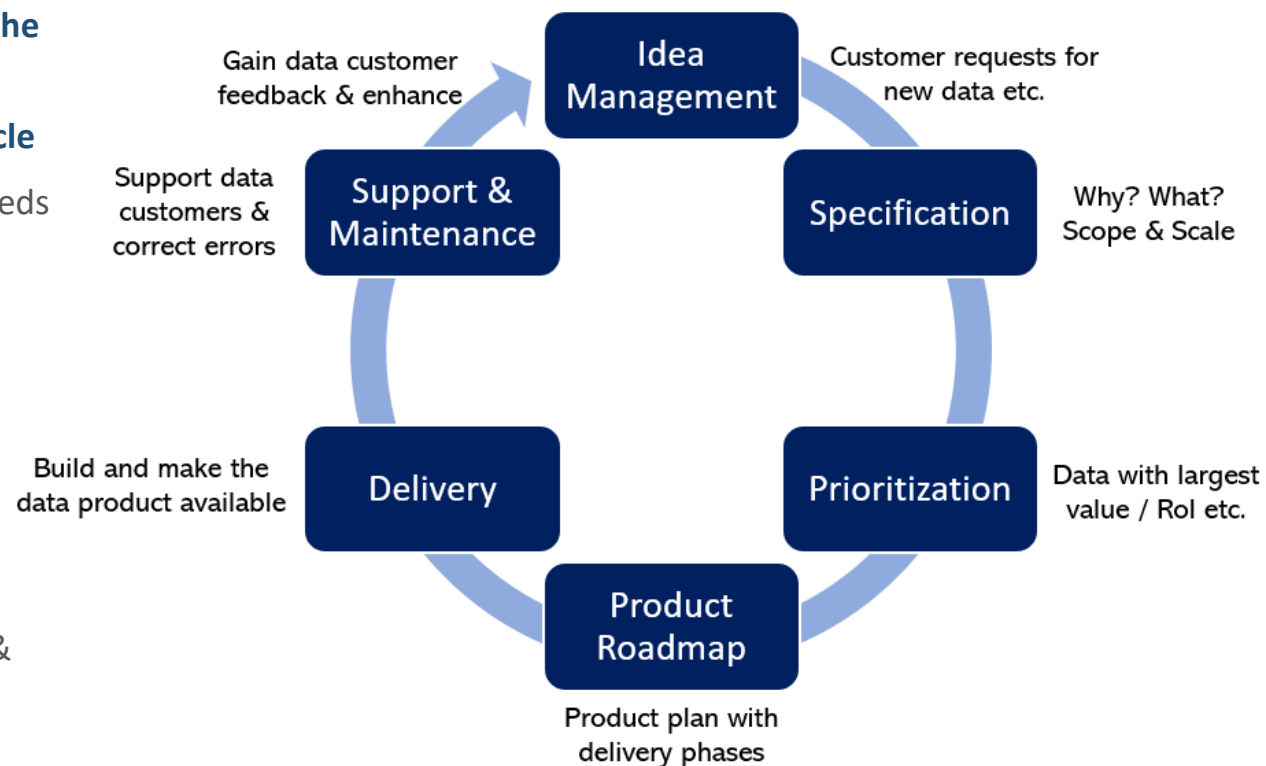
Data Mesh: Self-Serve Data Platform

- **Each domain should develop a data platform to support its data products** and ensure its autonomy

- The platform connects to existing operational and analytical platforms to source its data, but **hides complexities from its data customers**

- It should be specifically designed to allow its customers to search, access and process its data via a "digital storefront"

- Data customers should be able to self-serve data from the platform

- The platform consists of three primary layers, known as planes:

  - **Data Infrastructure** – file storage, access control, query engines etc.

  - **Data Product Developer** – workflow, code generation etc.

 - **Data Supervision** – knowledge graph of data products and relationship to connected products

Plane 3: Supervision

Plane 2: Developer

Plane 1: Infrastructure

Data Mesh: Federated Computational Governance

- Data Mesh requires a balance between:
  - The need for the domains to operate autonomously but cooperate when required
  - The need for the domains to interoperate to enable customers to:
    - Create data joins and unions for analytics and data science
    - Discover data intersections that derive wider insights and knowledge

- Its key principles for delivering this are:
  - Decentralization and self sovereignty
  - Information quality managed at the domain level
  - Automated execution to minimize manual efforts
  - Some cross-domain standardization to allow interoperability, e.g. APIs, data standards (format & content) etc.

- Data Mesh claims a new Federated data governance paradigm is needed as:
  - Traditional data governance is too centralized and stifles innovation and change
  - Does not meet the needs of its data product customers as focused mainly on operational data

Data Governance with Data Mesh
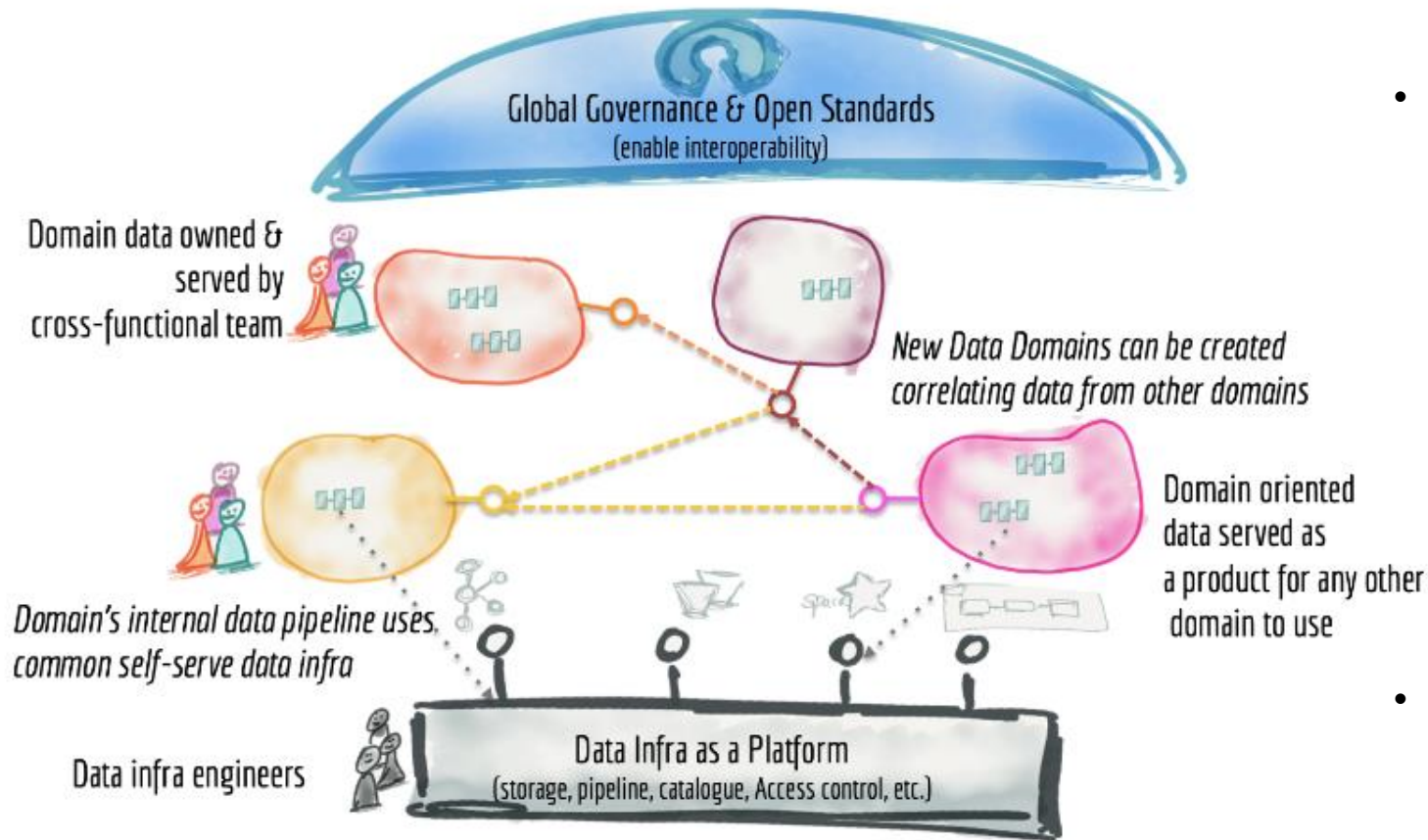
Source: Data Monolith to Mesh, Zhamak Dehghani

- **Global Governance & Standards**
  - "The key for an effective correlation of data across domains is following certain standards and harmonization rules."
  - "Such standardizations should belong to a global governance, to enable interoperability between polyglot domain datasets."
  - "Common concerns of such standardization efforts are field type formatting, identifying polysemes[1] across different domains, datasets address conventions, common metadata fields, etc."
- **Enterprise Data Stores**
  - "Where does the data lake or data warehouse fit in this architecture? They are simply nodes on the mesh."

[1] Polyseme: word with different meaning in different domains

Some Issues with the Assumptions of Data Mesh

Data Governance

Assumption: Data Governance is a rigid construct where everything goes through a central committee.

Counterpoint: There are many models of data governance and most successful data governance programs have a successful mix of local vs. global decision-making.

Data Domains

Assumption: Data Domains are loosely-defined components of the business.

Counterpoint: Enterprise Architecture and Data Governance models pay close attention to the definition of domains by:

- Process
- Organization
- Data Domain
- Business Capability

Architectural Quantum

Assumption: Data Products by Domain are the lowest-level architectural quantum.

Counterpoint: Data itself is the architectural quantum, e.g. Master Data objects (Customer, Product), or even critical data elements (customer email, product number).

Analytical Focus

Assumption: Federated data ownership focuses on published analytical data sets.

Counterpoint: True ownership and accountability for data needs to include operational data, where data is created and managed.

Socio-Technical Approach

Assumption: Enterprise Data Management is full of friction. Everyone is unhappy.

Assumption: It takes friction to light a fire. If you're not having the hard, cross-functional conversations to resolve key "polysemes" or "architectural quanta", you're simply avoiding the hard issues that need to be resolved.

"Data is a representation of the real world. If you want your data to be simple, make the world simple and get back to me." – Karen Lopez
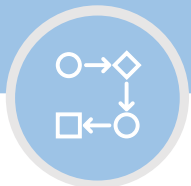
Models of Data Governance & Stewardship

There are diverse ways to implement governance, unique to each organization.

Many implementations have a *Blended approach*.

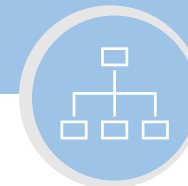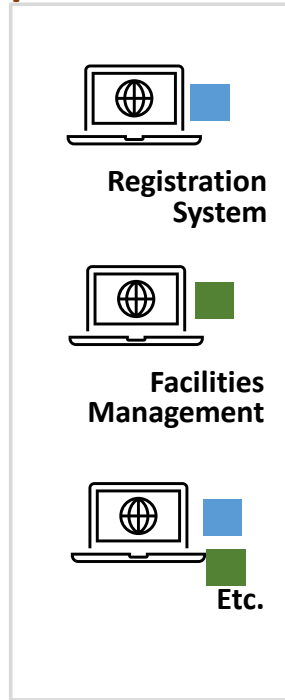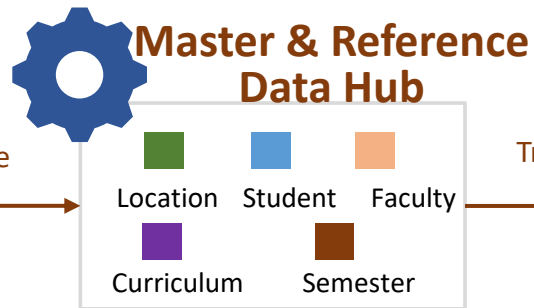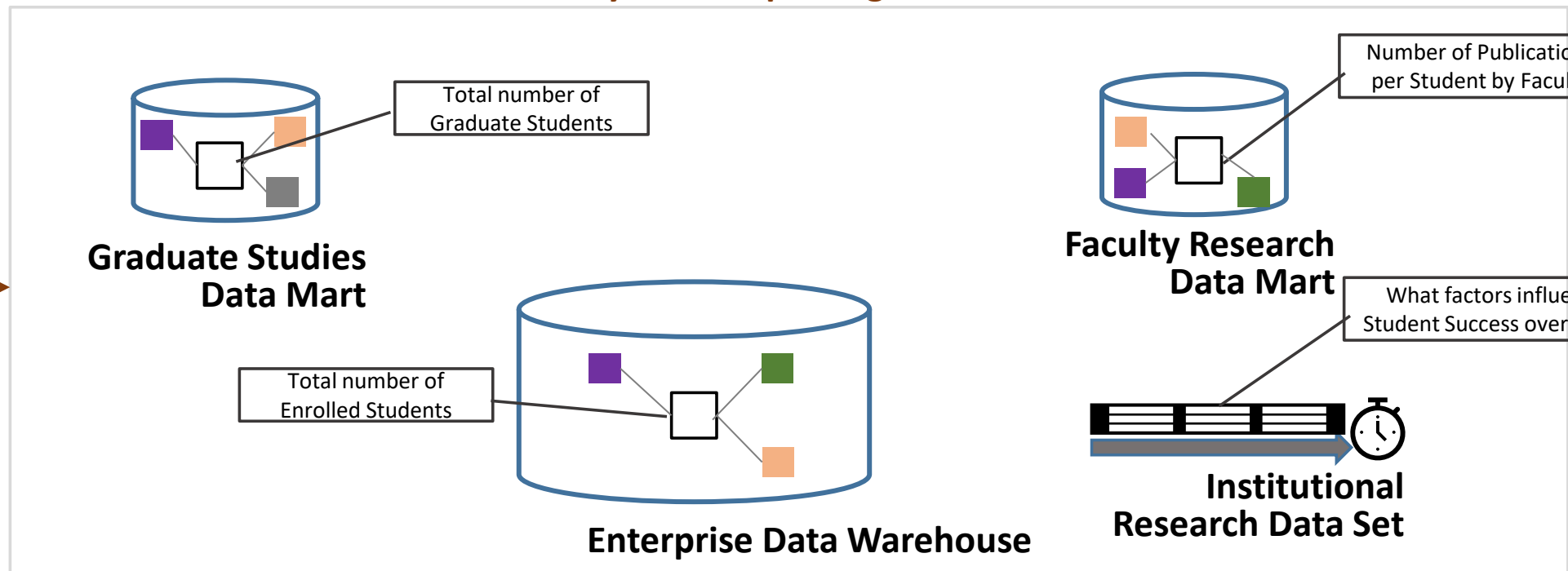| PROCESS CENTRIC	SYSTEMS CENTRIC	DATA DOMAIN CENTRIC	ORGANIZATION CENTRIC	CAPABILITY CENTRIC
Process owners become the data owner for all data created, amended & deleted by the business process for which they are responsible	**System owners** become the data owner for all data created, amended & deleted by the IT system for which they are responsible	Business appointed roles accountable for improvement of **key data domains** used across an organization	Business appointed roles accountable for improvement of key data domains on the basis of **departmental boundaries or geographic locations**	Business accountability based on **core business capabilities**, which are separated from organizational hierarchies or departments.
(e.g. Order to Cash, Customer Onboarding, etc.)	*(e.g. SAP, Peoplesoft, etc.)*	*(e.g. Customer, Supplier)*	*(e.g., Finance, Supply Chain, EMEA Sales)*	*(e.g., Finance, Supply Chain, Sales)*

Global Data Strategy, Ltd. 2023

17

Another Approach to Data Democratization

Operational Data

Analytics & Reporting Data Sets

Registration System

Facilities Management

Etc.

Total number of Graduate Students

Graduate Studies Data Mart

Total number of Enrolled Students

Enterprise Data Warehouse

Number of Publications per Student by Faculty

Faculty Research Data Mart

What factors influence Student Success over time?

Institutional Research Data Set

Master & Reference Data Hub

Location Student Faculty
Curriculum Semester

Bi-directional Publish & Subscribe of core data "quanta"

Trusted data sets as building blocks for reporting

Data Governance and Accountability exist across all phases of the data lifecycle – from operational to reporting

Data Governance & Accountability

The Concept of Data as a Product Exists in Other Areas

Open Data Sets Published for Consumption & Data Democratization

Be Careful of Assumptions

- Data Warehouses and Enterprise Data Management have **had numerous failures** over the past 20 years.

- Data Mesh has none of these failures.

- Michael Phelps has **lost more swimming races** than Donna Burbank.

- Donna has never lost a swimming race.

Summary

- Data Mesh stresses the principles of:
  - Domain Ownership
  - Data as a Product
  - Self-serve Data Platform
  - Federated Computational Governance
- Mesh is a combination of people and technology – a Socio-technical approach
- While many of the objectives of Data Mesh are laudable, there are some mis-alignments with core data management principles
  - Data Governance best practices
  - Master Data Management
  - Enterprise focus to include operational, analytical, and cross-functional "friction"

DATAVERSITY Data Architecture Strategies

Join Us Next Month

- **January**   Emerging Trends in Data Architecture – What's the Next Big Thing?

- **February**   Building a Data Strategy - Practical Steps for Aligning with Business Goals

- **March**   Data Mesh or Data Mess? Separating the Reality from the Hype

- **April**   Master Data Management - Aligning Data, Process, and Governance

- **May**   How do Data Governance & Data Architecture Support Each Other?

- **June**   Why You Need Data Management – Getting Executive Buy-In

- **July**   Artificial Intelligence and Machine Learning – Building the Right Architectural Foundation

- **August**   Data Quality Best Practices (with Nigel Turner)

- **September**   Best Practices in Metadata Management

- **October**   Designing Data for Business Intelligence & Analytics – Where the Star Schema Fits in a Modern Data Architecture

- **December**   Enterprise Architecture vs. Data Architecture

Global Data Strategy
The Business of Data

DATA
ARCHITECTURE
STRATEGIES

In today's business environment, showing **rapid time to value** for any technical investment is critical.

But technology and data can be complex. At Global Data Strategy, **we help demystify technical complexity** to help you:

- Demonstrate the ROI and **business value of data** to your management

- Build a data strategy **at your pace to match your unique culture** and organizational style.

- Create an **actionable roadmap for "quick wins"**, which building towards a long-term scalable architecture.

Global Data Strategy's shares experience from some of the largest international organizations scaled to the pace of your unique team.

Global Data Strategy has worked with organizations globally in the following industries:

Finance · Retail · Social Services · Health Care · Education · Manufacturing · Government · Public Utilities · Construction · Media & Entertainment · Insurance …. and more

Global Data Strategy, Ltd. 2023

www.globaldatastrategy.com

Thoughts? Ideas?
Questions?