

# The Flipkart journey through Scaling and building platform Resilience



regunathb

# I'll talk about

- Architecture evolution
- Scaling journey, patterns and systems built
- Dimensions to consider in Scaling
  - Protecting customer experience
  - Linear scalability and reducing cost of linear scaling
- Resilience
- Key Platform Services
- Data Stores evolution

# Architecture Evolution

June 2007

**FlipKart.com**

is coming soon ...

# Nov 2007

**flipkart**™ India beta

[Login/Signup](#) | [Help](#)  
[Home](#) [My Account](#) [Shopping Cart \(0 items\)](#)

**Find Books**    
(e.g dan brown, kite runner, isbn 0143029347)

**Search**  
Simple and Fast  
Search from over 1 Lakh Books

**Buy**  
Delivered at your doorstep

**Save**  
[Shop Securely](#) more...  
[anywhere in India](#) more...  
[Great Discounts](#) on in-store prices

**Why shop at flipkart.com?**



Contact our customer service  
[cs@flipkart.com](mailto:cs@flipkart.com)

---

**Our prices are all inclusive**  
No Extra Taxes  
No Shipping Charges

---

Politics Entertainment Children Books Competition Exams Business & Investing Professional & Technical Biographies & Memoirs Cooking Non-Fiction Textbooks Literature & Fiction Philosophy Photography Computers & Internet Education Architecture History Science Economics Health & Fitness Crafts & Hobbies Best-Sellers Religion & Spirituality Sports Outdoors & Nature Medical Body & Mind Arts & Music Travel Science Fiction & Fantasy Romance Management Engineering Reference Mystery & Thrillers

---

**New Releases**



# Dec 2012

flipkart.com

24x7 Customer Support - Contact Us | Home | Account | Wishlist | Login | Signup  
e-Gift Voucher | Wallet | Digital Cart (0)

All Categories

- Clothing
- Footwear
- Mobiles & Accessories
- Computers
- Watches, Bags & Wallets
- Cameras
- Books, Pens & Stationery
- Home & Kitchen
- Beauty & Personal Care
- Gaming
- TV, Video & Audio
- Music, Movies & Posters
- Baby Care & Toys
- Sports & Fitness

**flyte** The Digital Store

eBooks (Beta) NEW

MP3 Downloads

Sending request...

Search: All Categories

Search for items

Cart (0)

New at Flipkart

Zero Fee EMI on Laptops

HP »

Lenovo »

Samsung »

Men's Clothing

Sweaters »

Sweatshirts »

Jackets »

Best of 2012

Books »

Movies »

Games »

New Year Diaries

Premium Pens »

Special Offers »

Stationery Best Sellers »

Mobiles

View More Mobiles »

Combo of Samsung Galaxy S Duos S7562 (Pure White)  
Rs. 17700

HTC Explorer (Black)  
Rs. 7899

Android v2.3 OS

Nokia Asha 205 (Cyan)  
Rs. 3349

2.4-inch LCD Screen

Samsung Galaxy Tab 2 P3100 (Pure white, Wi-Fi, 3G, 16 GB)

# Flipkart today

Flipkart Plus  Regunath More 

Electronics TVs & Appliances Men Women Baby & Kids Home & Furniture Sports, Books & More Grocery Offer Zone

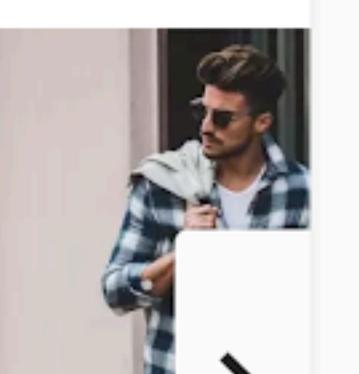
**FLIPSTART DAYS** 1st-3rd April

Daily Essentials For Home From ₹49 Towels, Bottles & more

AXIS BANK 10% Instant Discount\* on Axis Bank Credit and Debit Cards

\*T&C Apply

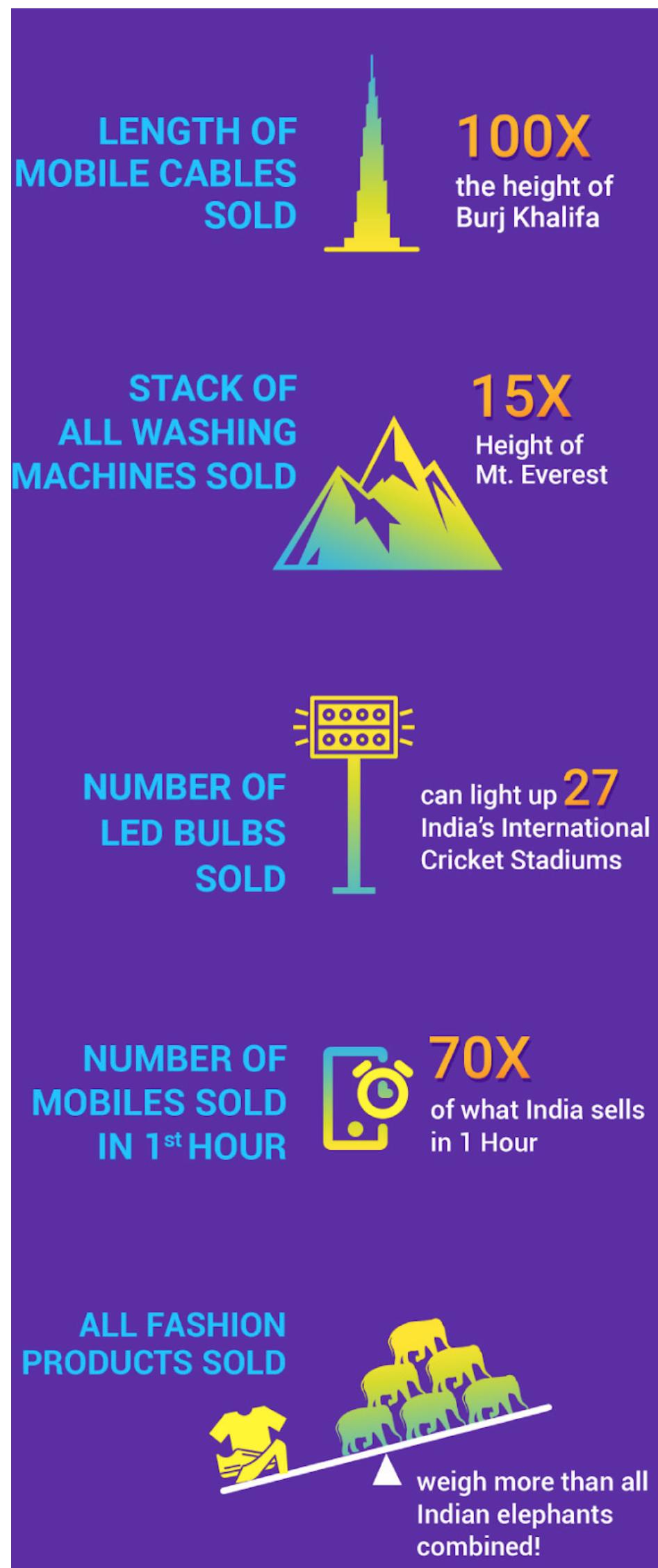
**Deals of the Day** [VIEW ALL](#)

					
<b>Ducati &amp; Ed Hardy</b> Min. 50% Off	<b>Fila</b> Min. 50% Off	<b>LED Bulbs</b> From ₹89	<b>Skybags, Puma, America...</b> 55-80% Off	<b>Adidas, Wrangler &amp; more</b> Up to 70% + Extra 10%	<b>Arrow, Lee &amp; T...</b> Min. 50% O

**HOT DEALS** New Deals Every 8 Hours Extra 10% Off Grab Now 

[https://www.flipkart.com/home-april-payday-store?otracker=hp\\_bannerads\\_6\\_payda-homep-b83d7\\_a\\_TJE14HFXJ914](https://www.flipkart.com/home-april-payday-store?otracker=hp_bannerads_6_payda-homep-b83d7_a_TJE14HFXJ914)

# eCommerce has arrived!

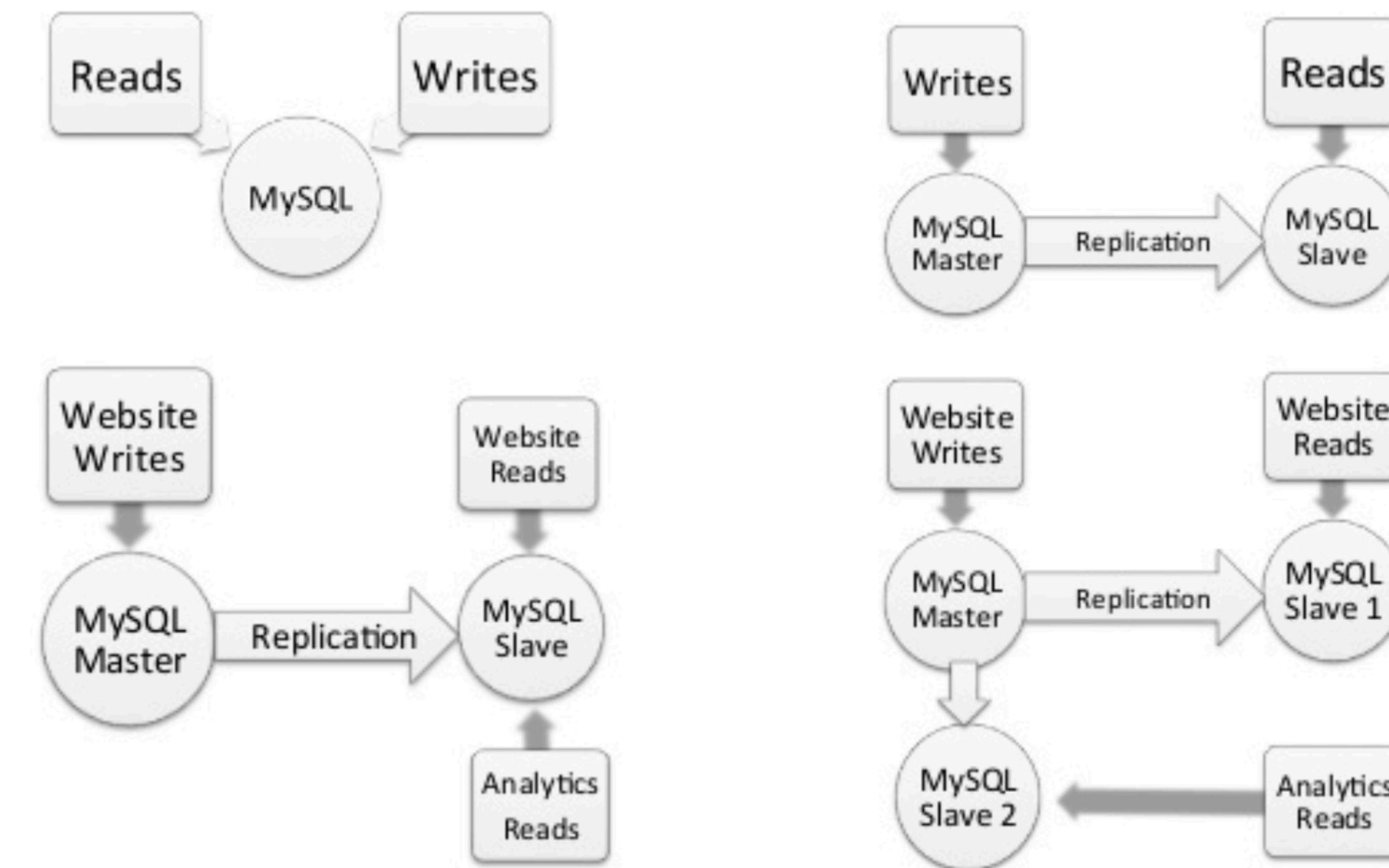


# Architecture and Tech stack - 1st Generation

- Two tier architecture - PHP(Interpreted, Process per Request) and MySQL

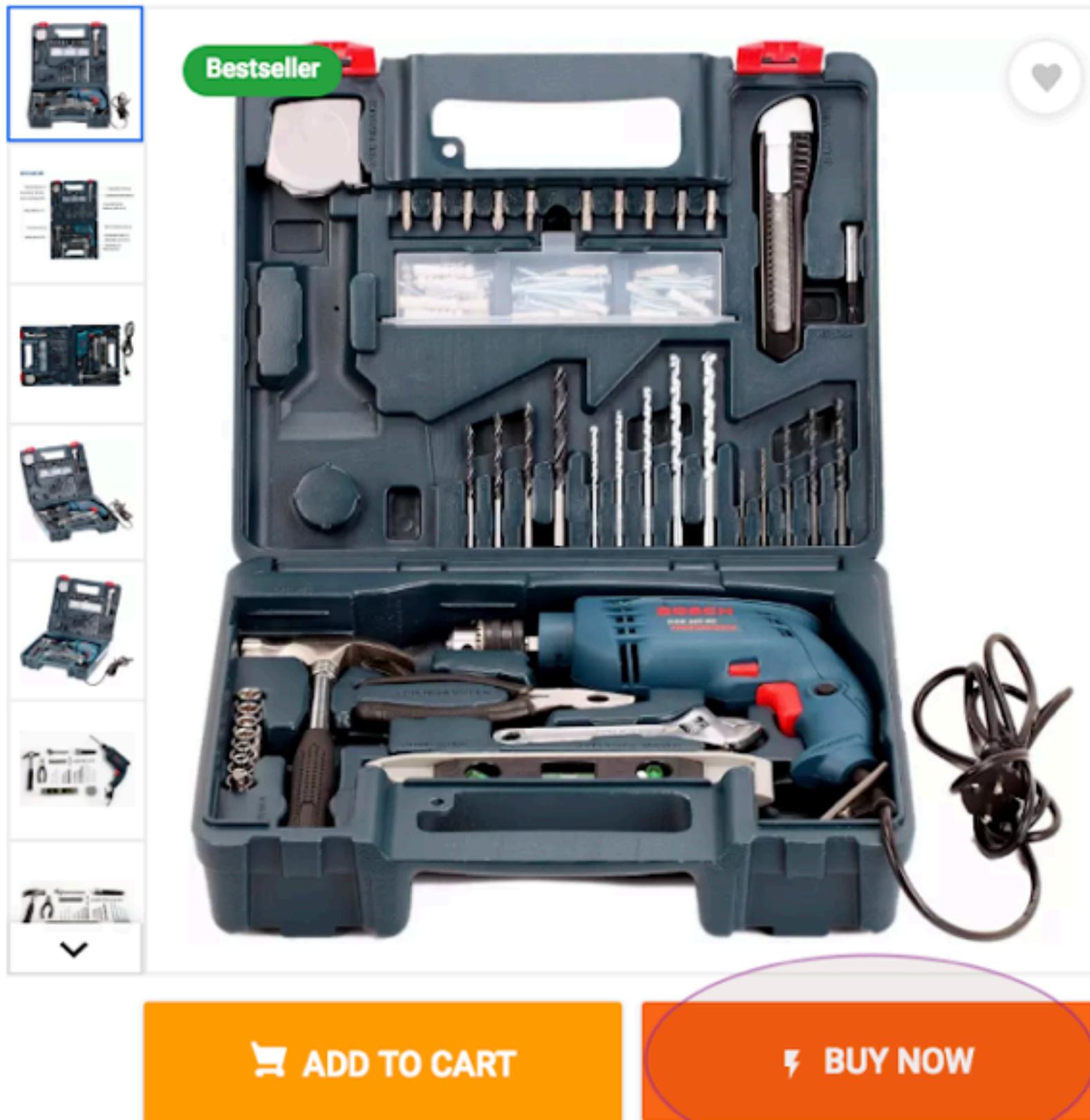


- MySQL scaling through adding slaves (Moved from Strong Data Consistency to Eventual Consistency)



Scaling journey, patterns and systems built

# Scaling - Polyglot persistence, SOA



Home > Home Improv... > Tools > Power Tools > Power & Han... > Bosch Power ... >

**Bosch GSB 500 RE Power & Hand Tool Kit (92 Tools)**

4.4★ 48,582 Ratings & 7,983 Reviews Assured

**₹3,498** ₹5,320 34% off

Eligible for Flipkart Pay Later ?  
Bank Offer 10% off\* with Axis Bank Buzz Credit Card [T&C](#)  
Debit Card EMI ₹1,209/month. Standard EMI also available [View Plans](#) >

Delivery: 641603 Change check pincode X

Delivery by 4 Mar, Monday | Free ₹40 ?  
If ordered before 4:25 PM

[View Details](#)

Highlights: 500 W  
Power Requirement: 500 W

Services:

Important Note: Flipkart customers get a warranty of 2 years on this product, availing the warranty at any Bosch-authorized service center.

Seller: RetailNet 4.5★

Data	Store
User	Vitess (sharded MySQL)
Product	HBase, Elastic Search, Redis
Cart/Checkout/ Payments	YAK(HBase)
Search	Solr, Redis
Pricing, Promotions	HBase, Hazelcast, Aerospike
Promise (Inventory)	Redis, Aerospike, MySQL
User Session	Memcached, HBase

Patterns & Choices : CQRS, different Consistency levels, Data durability guarantees

# Latency & Failures in SOA Fanout

The screenshot shows a product page for a Bosch GSB 500 RE Power & Hand Tool Kit (92 Tools). The page includes a large image of the tool kit, a price of ₹3,498 (34% off), and various purchasing options like EMI and bank offers. It also displays delivery information (Delivery by 4 Mar, Monday | Free ₹40) and highlights the 500 W power requirement.

**Product Details:**

- Bosch GSB 500 RE Power & Hand Tool Kit (92 Tools)
- 4.4 ★ 48,582 Ratings & 7,983 Reviews
- ₹3,498 ₹5,320 34% off
- Eligible for Flipkart Pay Later
- Bank Offer 10% off\* with Axis Bank Buzz Credit Card T&C
- Debit Card EMI ₹1,209/month. Standard EMI also available [View Plans](#)
- Delivery: 641603 (Change, Check pincode)
- Delivery by 4 Mar, Monday | Free ₹40 (If ordered before 4:25 PM)
- View Details
- Highlights: 500 W, Power Requirement: 500 W
- Services

**Buttons:**

- ADD TO CART
- BUY NOW

**Important Note:** Flipkart customers get a warranty of 2 years on this product. Availing the warranty at any Bosch-authorized service center.

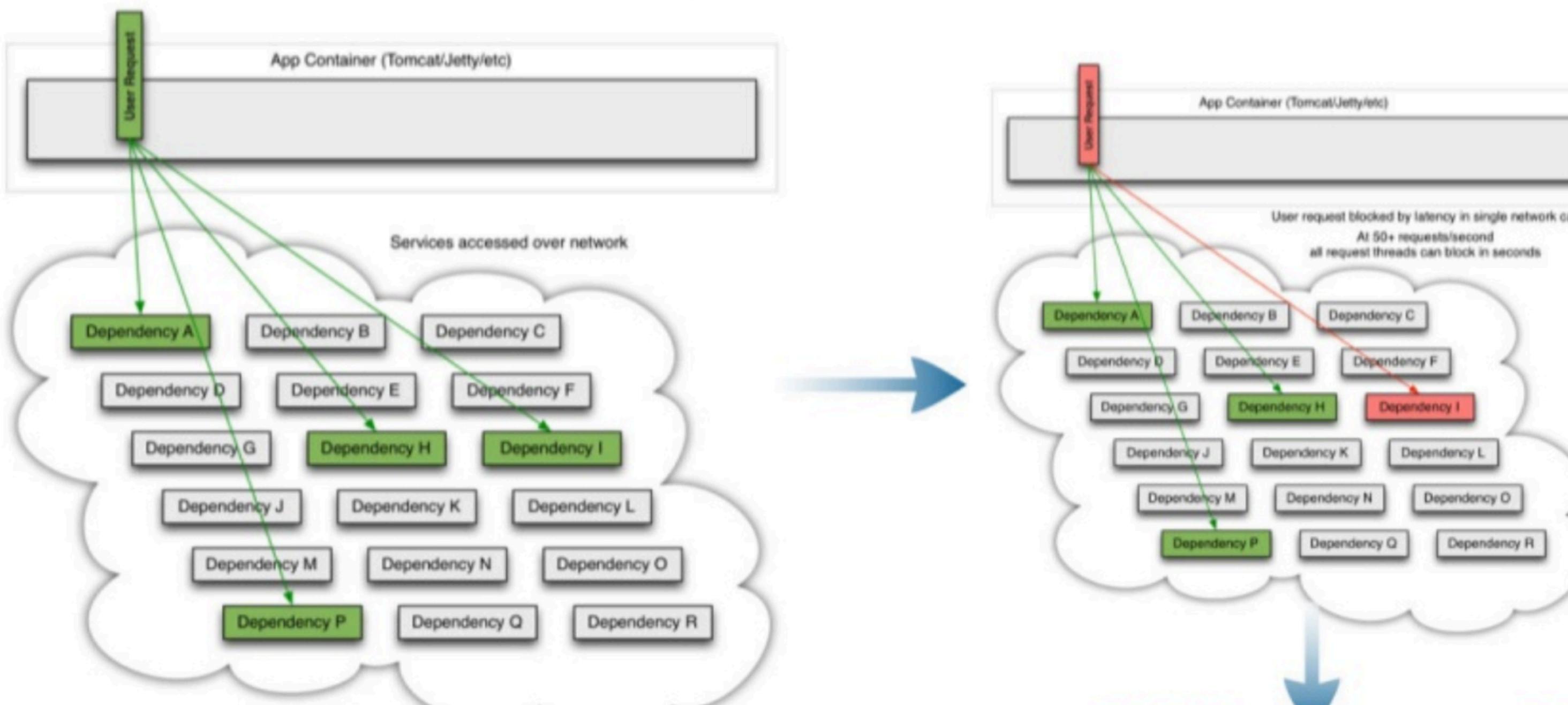
**Seller:** RetailNet 4.5★

Running a website that depends on 15 services, each with 99.9% availability, we get

**99.9% ^ 15 = 98.5% uptime**

*(probability of all services providing 99.9% availability at the same instance of time)*

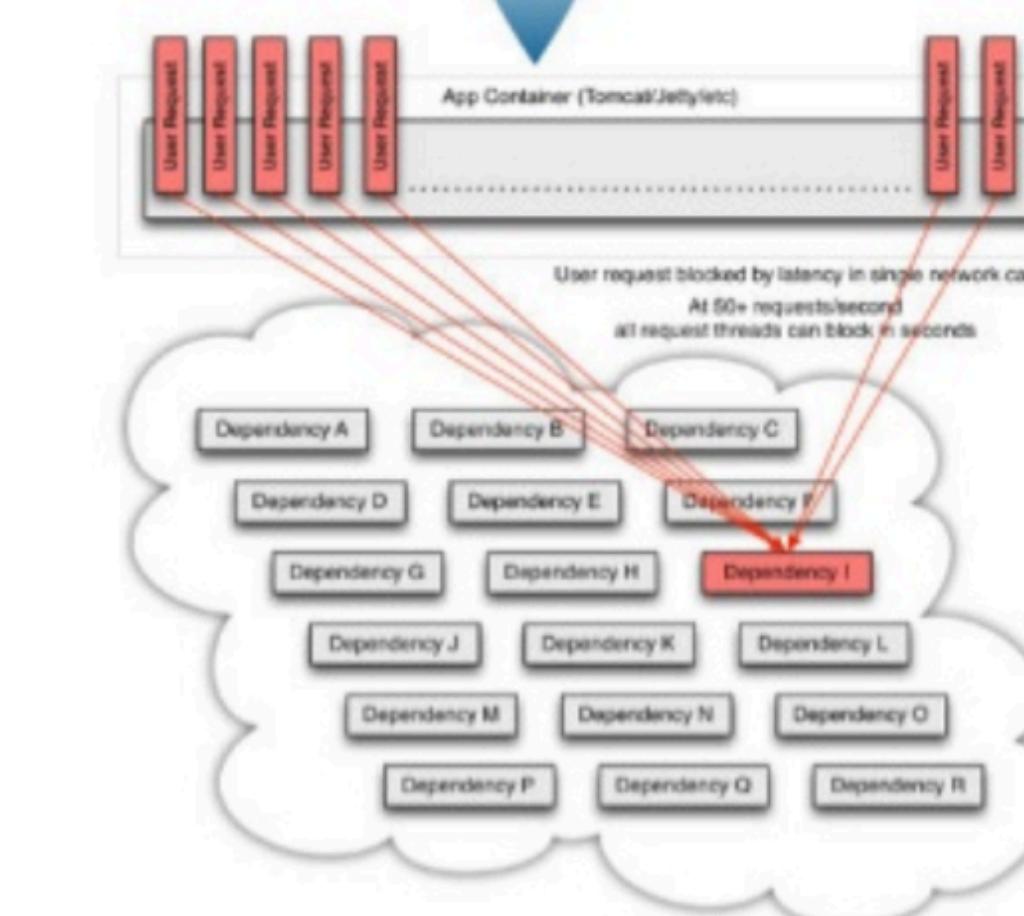
# Latency & Failures in SOA Fanout (contd.)



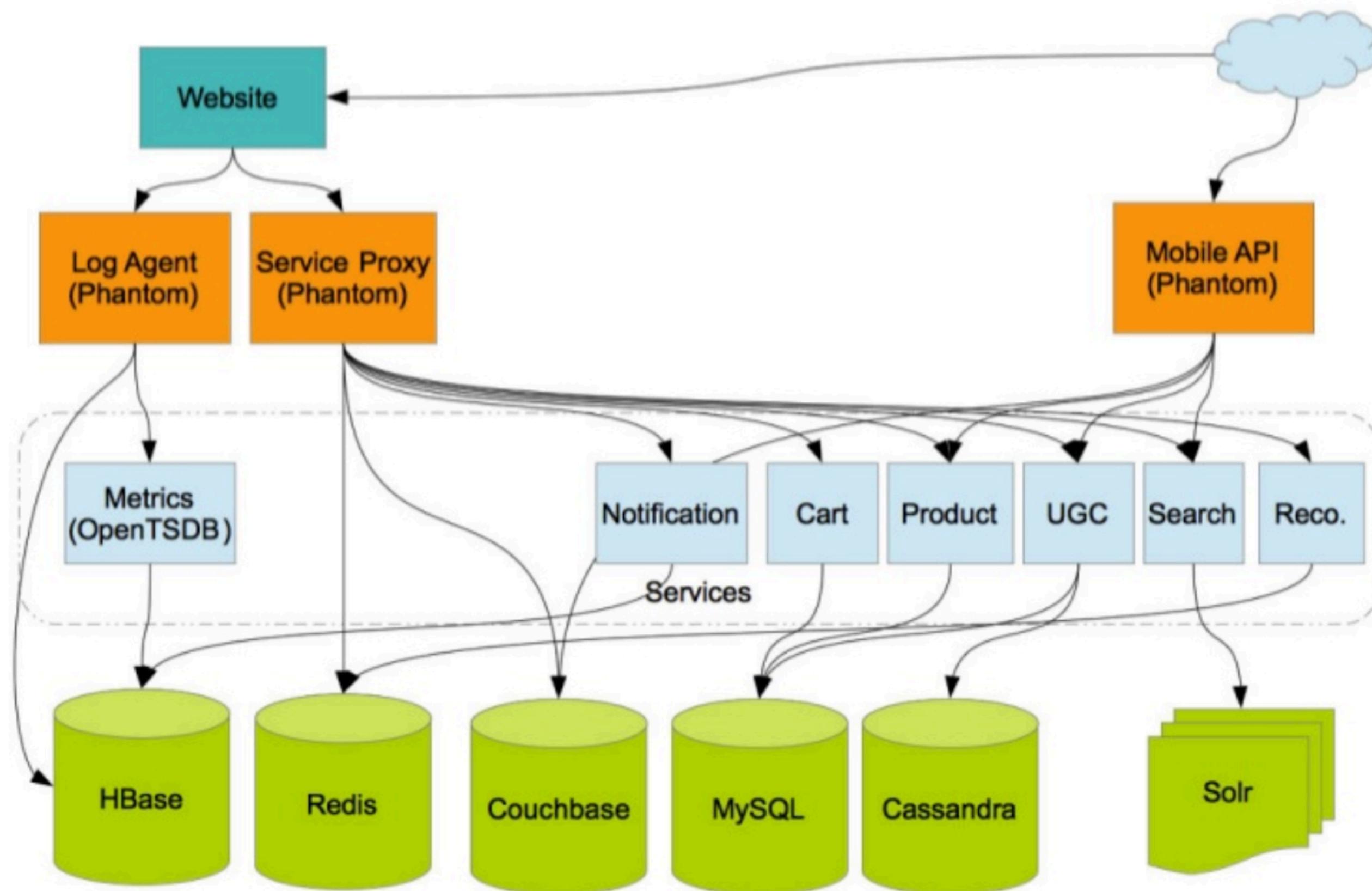
source: [github.com/Netflix/Hystrix/wiki](https://github.com/Netflix/Hystrix/wiki)

Delay	User reaction
0-100ms	Instant
0-300ms	Feels sluggish
100-1000ms	Machine is working...
1s+	Mental context switch
10s+	I'll come back later...

source: [www.igvita.com/slides/2012/railsconf-making-the-web-faster/#4](http://www.igvita.com/slides/2012/railsconf-making-the-web-faster/#4)



# Fault tolerance in SOA (Fanout), Sidecars, Circuit-breakers



Flipkart open source:

- <https://github.com/flipkart/phantom>
- <https://github.com/flipkart-incubator/Poseidon>

- 1 billion semaphore isolated calls
- 0.5 billion thread isolated calls
- Smart auto-fallback for latent, failed calls

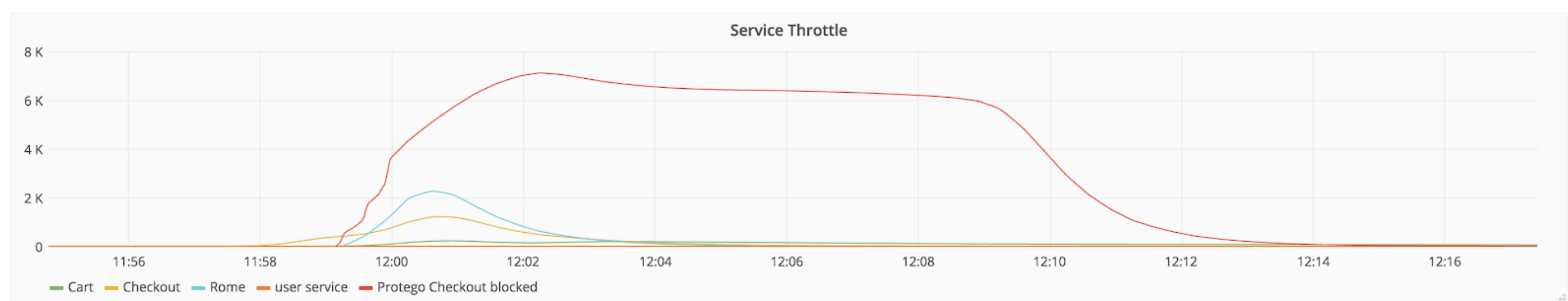
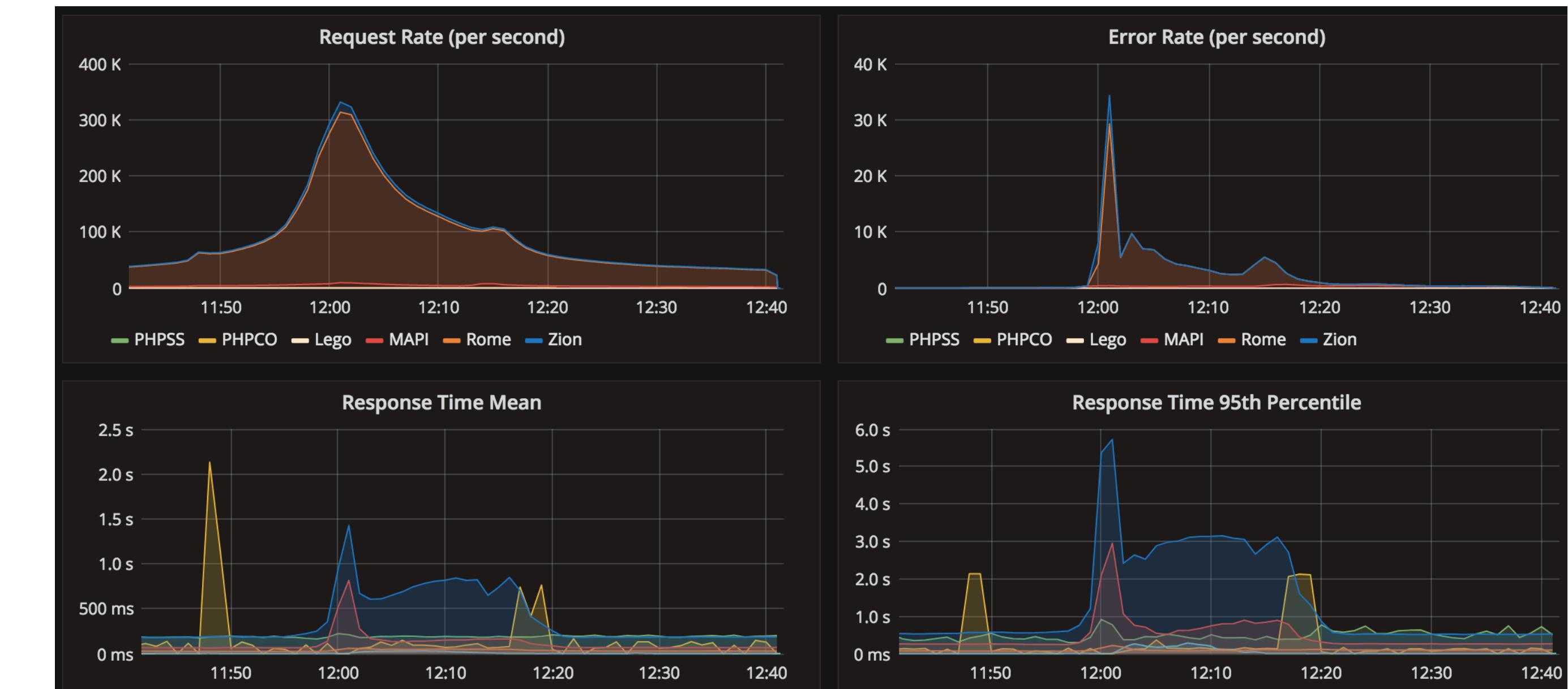
# Runtime containers (Closed source)

- Excalibur
  - Dropwizard powered lightweight web services framework
  - Integration with Argus to provide schema driven API development
  - Annotation driven Hystrix integration with dynamic configs
  - Google guice integration to enable DI pattern
  - Pushing to FDP via Specter
  - Dropwizard metrics at all level and support for custom metrics
  - ConfigService integration with support for hierarchical configurations and virtual hosts
  - Debugability via annotation driven API
  - Integration with Audit systems like FoxTrot
  - Swagger integration for API documentation
  - Bill-of-Materials pattern for dependency management
  - Standardized debian packaging
  - Customized instance create templates and build scripts
- nettyio - Dropwizard like container on Netty

# Dimensions to consider in Scaling

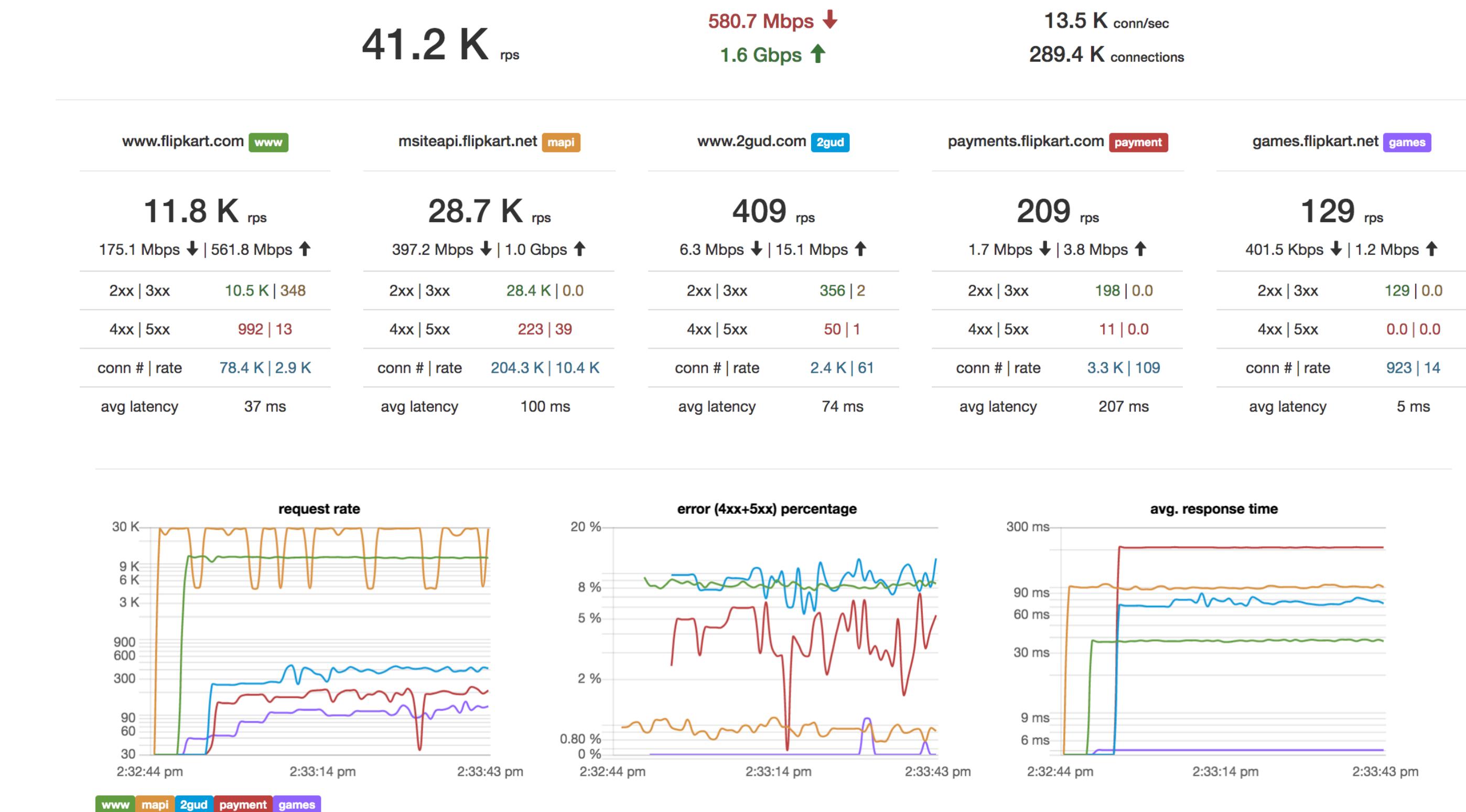
# Protecting customer experience

- Spike in traffic - e.g. Flash sales, BBD
  - Service-wise throttling
  - Isolated service clusters
  - Differentiated Data consistency guarantees
- Handling Bots/Plug-in traffic



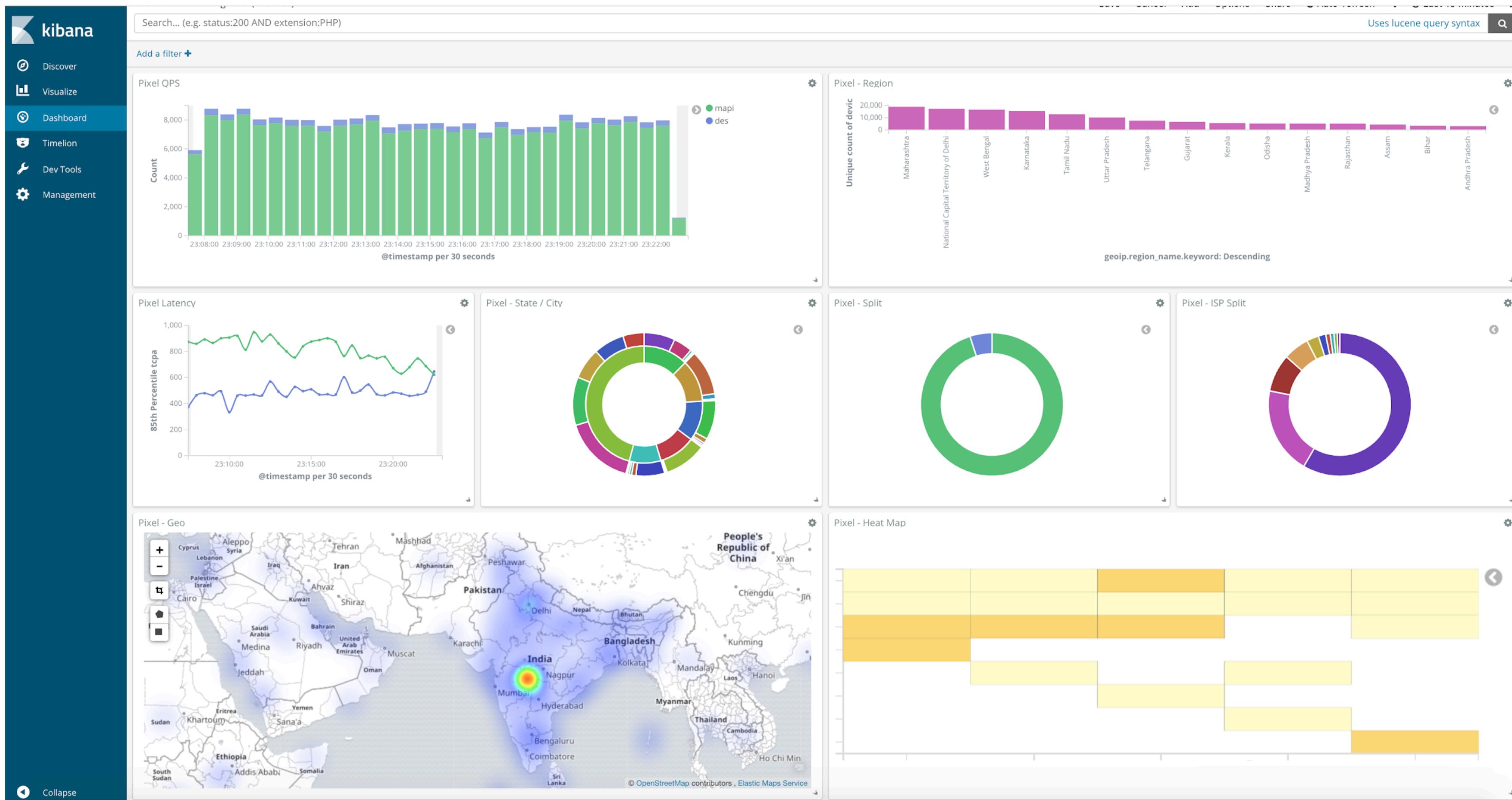
# Linear scalability and reducing cost of linear scaling

- Flipkart Cloud Platform
  - Private cloud built ground-up, offers Virtualization on VMs/Containers
  - 10000's of bare metals (Custom hardware directly from ODMs)
  - Self-serve IaaS, PaaS services
  - Elastic Load Balancer, SDN
    - Scalable, Secure, Multi-protocol proxying capabilities, Cluster management & Traffic routing capabilities



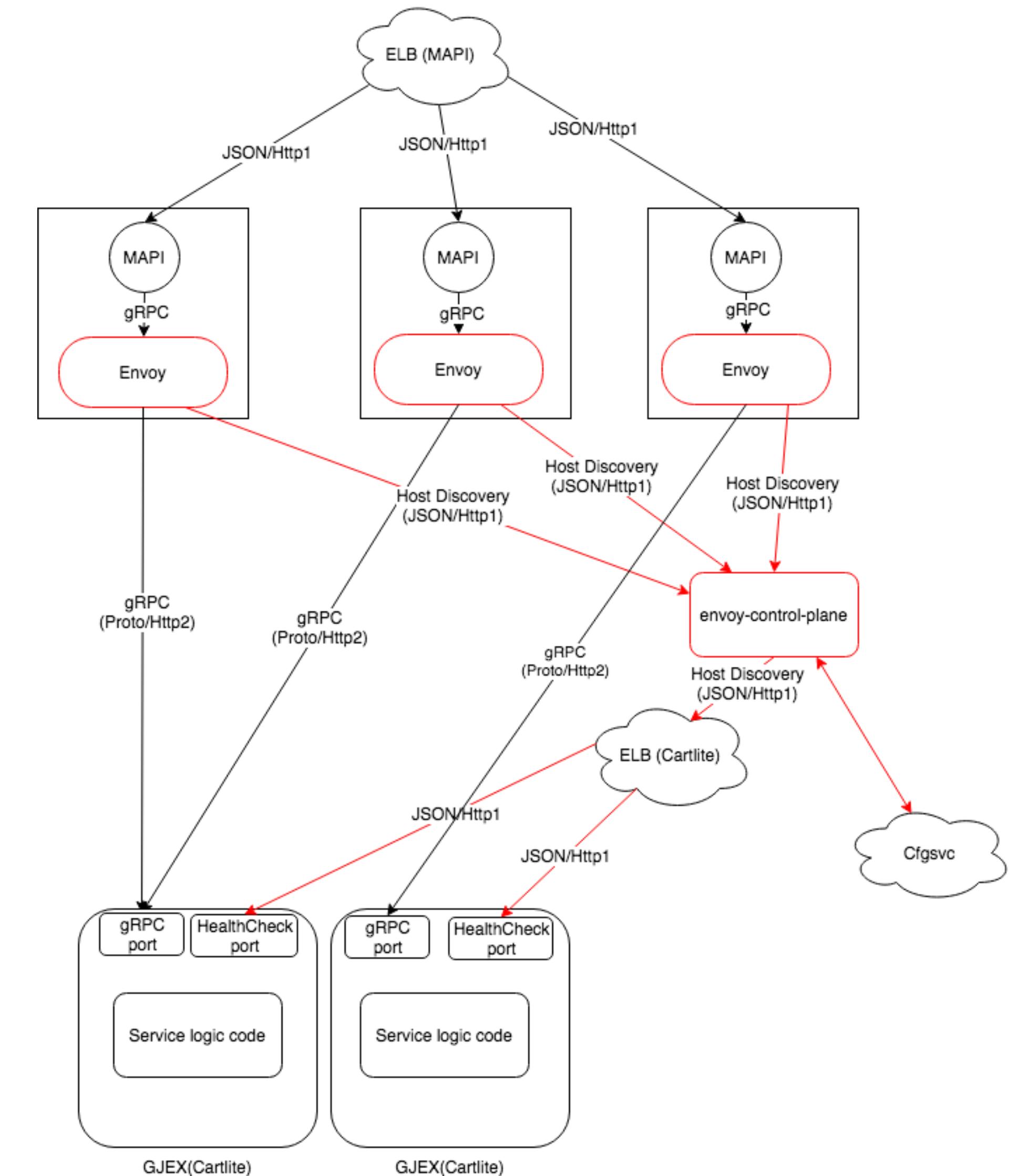
# Linear scalability and reducing cost of linear scaling

- FCP Real User Monitoring



# Linear scalability and reducing cost of linear scaling

- Some of our largest services
  - 3 ~ 5 million requests per second
  - Deployed on ~1000 nodes
  - Need optimisation (CPU, Network, Transport, Memory etc.)
- gRPC Service Mesh
  - Sidecar - Envoy
  - Custom Lookaside Load-balancer
  - GJEX Service Container  
(Dropwizard for gRPC)
  - Flipkart open source : <https://github.com/flipkart-incubator/grpc-jexpress>
  - 30% to 40% hardware reduction



# Resilience

# Resilience Journey (WIP)

- Handling spike in traffic, differentiating User intent, Throttling **[Done]**
- Better customer experience
  - Higher overall availability, Data Consistency tradeoffs, Throttling, Bots/Reseller handling **[Done, In Progress]**
- Lower MTTR, Isolation of failures
  - Circuit Breaker **[Done]**
  - Degraded modes of execution, Ability to work with Partial Results **[Done]**
  - Smart retries **[In Progress]**
- Multi-homed Services **[In Progress]**
  - Active-Active deployments across multiple DCs
  - Custom Strongly Consistent Data store(s)
- Resilience in non-User facing systems
  - Backoffs & Retries **[Done]**
  - Enhanced Observability, Early Warning, Throughput SLA guarantees **[In Progress]**

# Key Platform Services

# IaaS services (Generic)

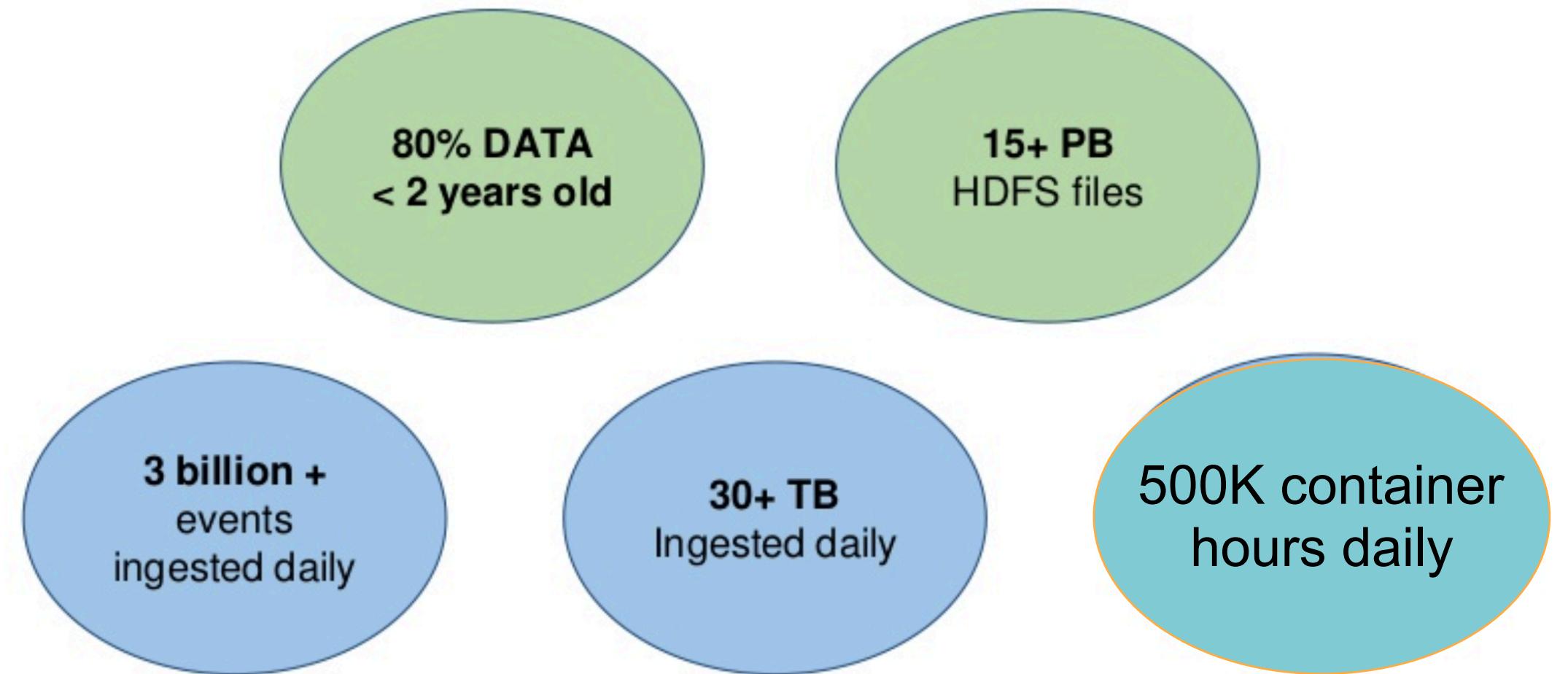
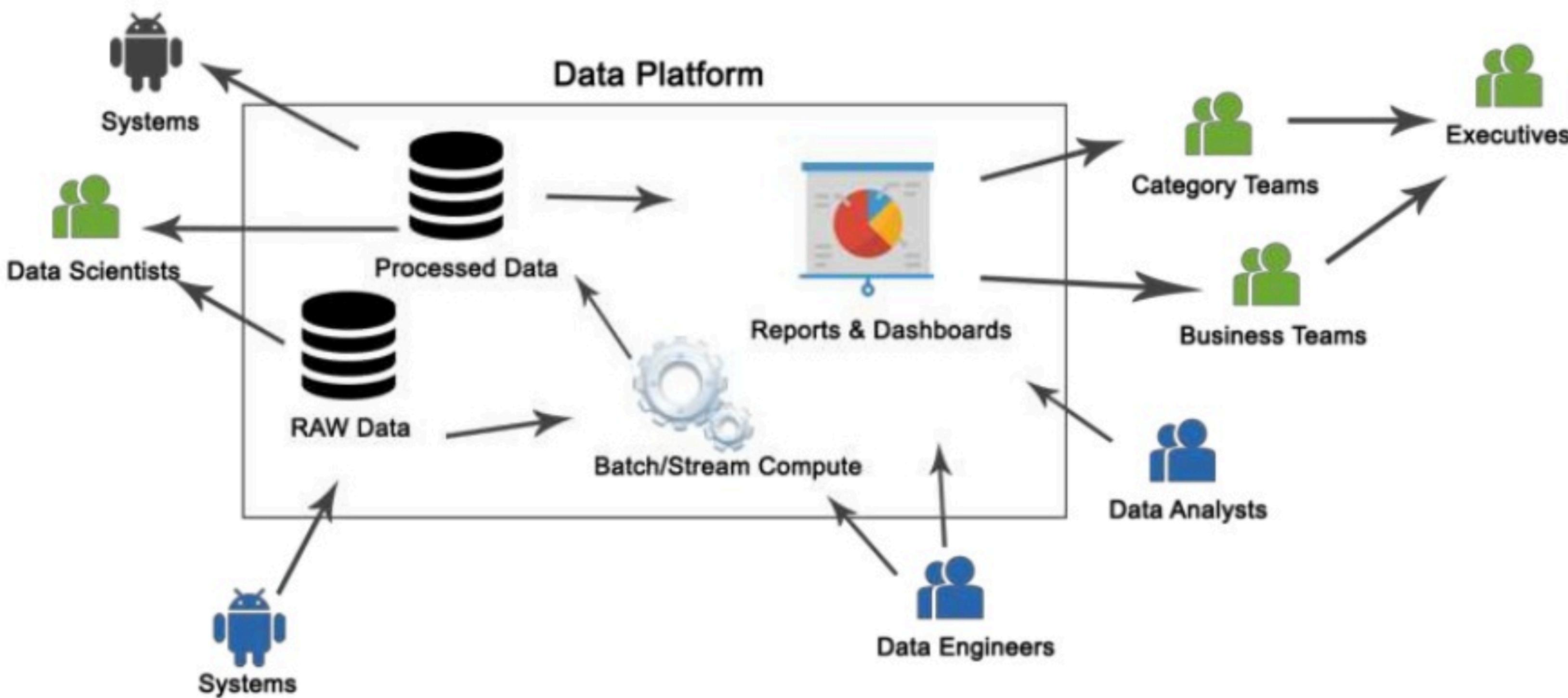
- Instance Provisioning - VM packing on Bare metal across Racks & Regions, Resource, Instance isolation
- Maintenance - VM movement, Baremetal upgrades
- ELB and SDN control plane
- Instance metrics, alerting
- Support for multiple instance types - API services, Caches, Transactional Datastores, Messaging, Analytics etc.
- OS upgrades, Network, Storage drivers for ODM hardware
- Deployment service
- Containerisation
- VPC, Forward proxy

# PaaS services (Generic)

- Metrics - OpenTSDB based custom solution with quotas, cross-DC replication
- Configuration - ZK based store with custom solution for pushing config changes across 1000's of instances
- Blob Store - Ceph based, S3-like tiered blob store
- Hosted MySQL - Managed Operations (Upgrades, Backup/Restore, Alerting)
- AuthN, AuthZ
- Logging service - Hot (ES backed) and Cold store(HDFS, Hive queried)
- Distributed Tracing - Dependency Fanout(from Network), Log based

# PaaS services (Generic)

## Flipkart Data Platform



# PaaS services (Generic)

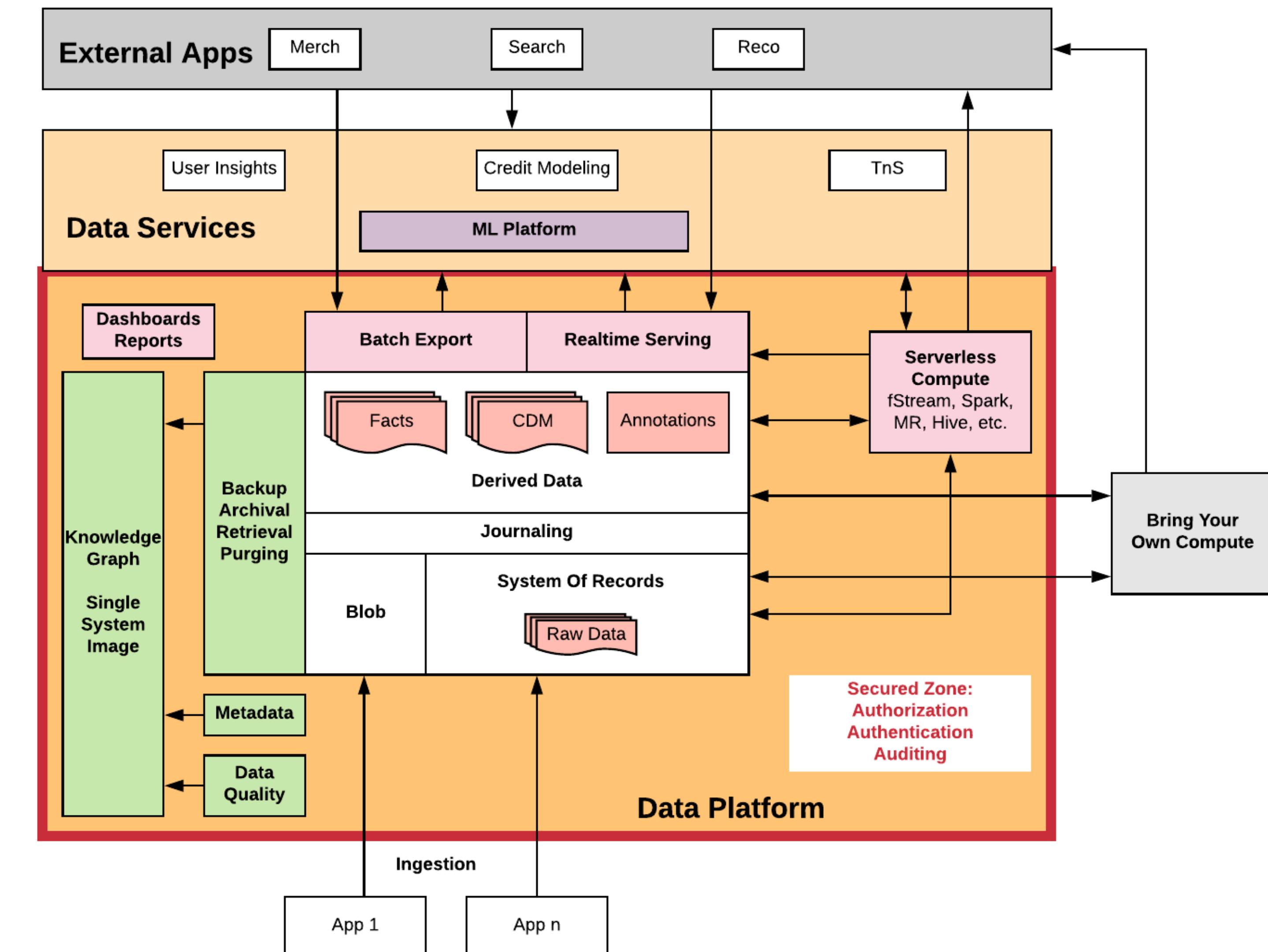
Flipkart Data Platform

## Features & Optimizations

- FStream
- Dr Elephant - Job Analysis
- Tez - Compute Engine
- ORC - Storage Format

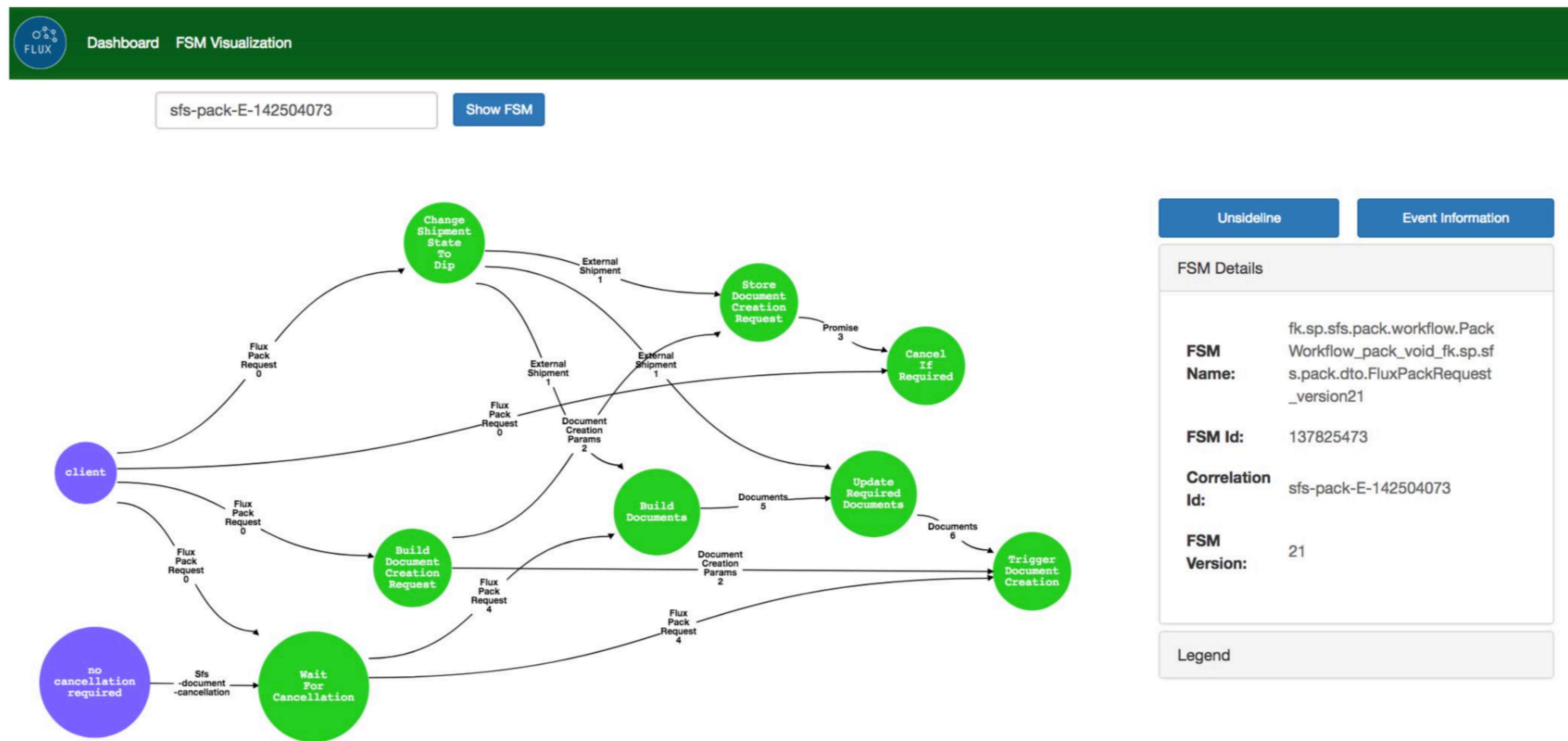
## Data Governance

- Dependency Lineage
- Schema Tightening
- DQ Asserts
- Org Queues



# PaaS services (Generic)

- Flux (AWS Lambda, Step-Functions like - Hosted, Stateful, Event driven)
- <https://github.com/flipkart-incubator/flux>



# PaaS services (Generic)

- Flux programming primitives (Code as DSL)

## WorkFlow

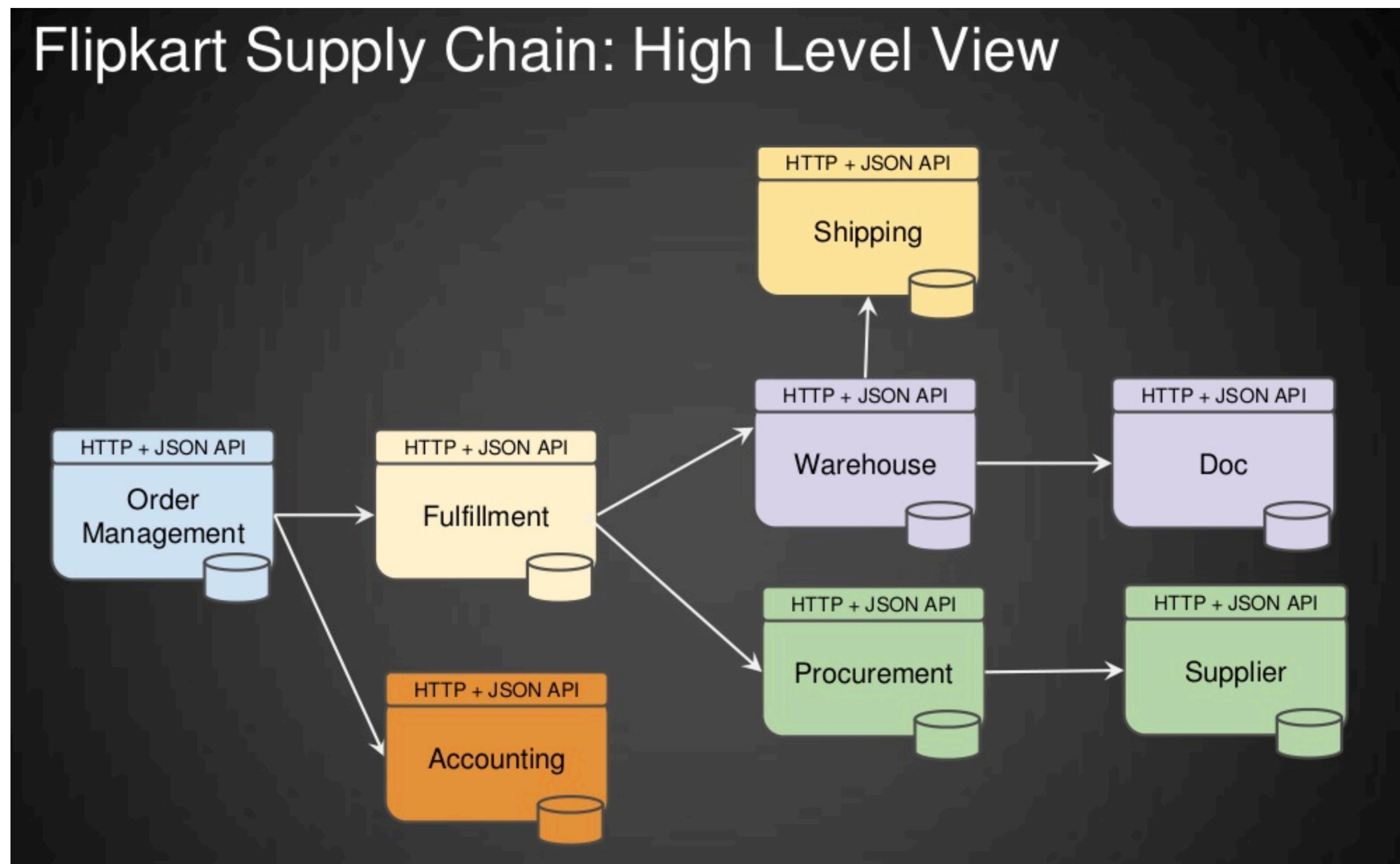
```
@Workflow(version = 1)
public void verifyUser(UserId userId) {
    UserData userData = userDataService.retrieveUserData(userId);
    final UserVerificationStatus userVerificationStatus = userVerificationService.verifyUser(userData);
    checkVerificationStatus(userVerificationStatus);
}
```

## Task

```
@Task(version = 1, timeout = 1000l, retries = 2)
public void checkVerificationStatus(UserVerificationStatus userVerificationStatus) {
    if (userVerificationStatus.isVerifiedUser()) {
        notificationService.sendWelcomeEmail(userVerificationStatus.getUserId());
    } else {
        notificationService.notifyCustomerSupport(userVerificationStatus.getUserId());
    }
}
```

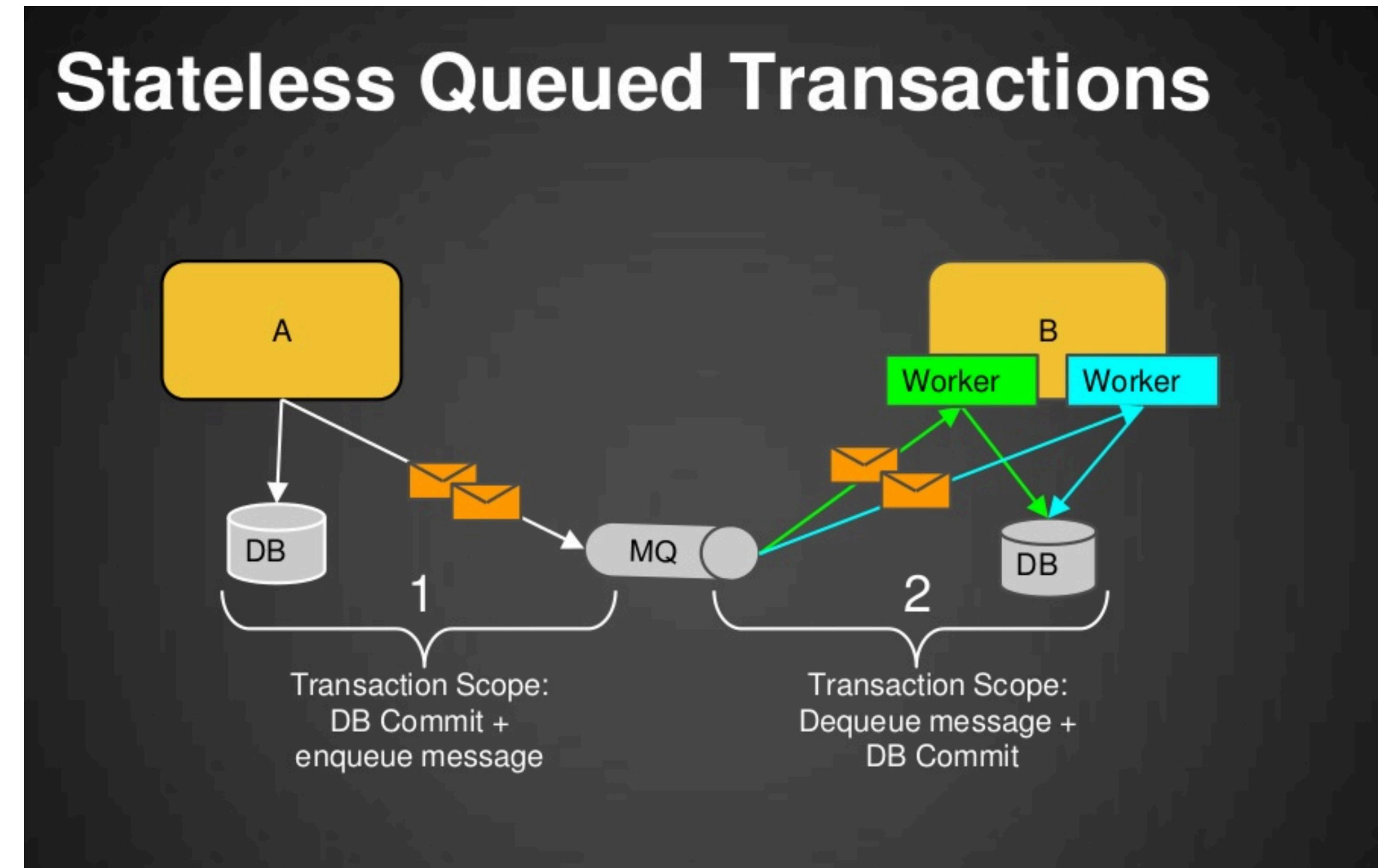
# PaaS services (Domain specific)

- Need for Stateful,  
Reliable  
orchestration



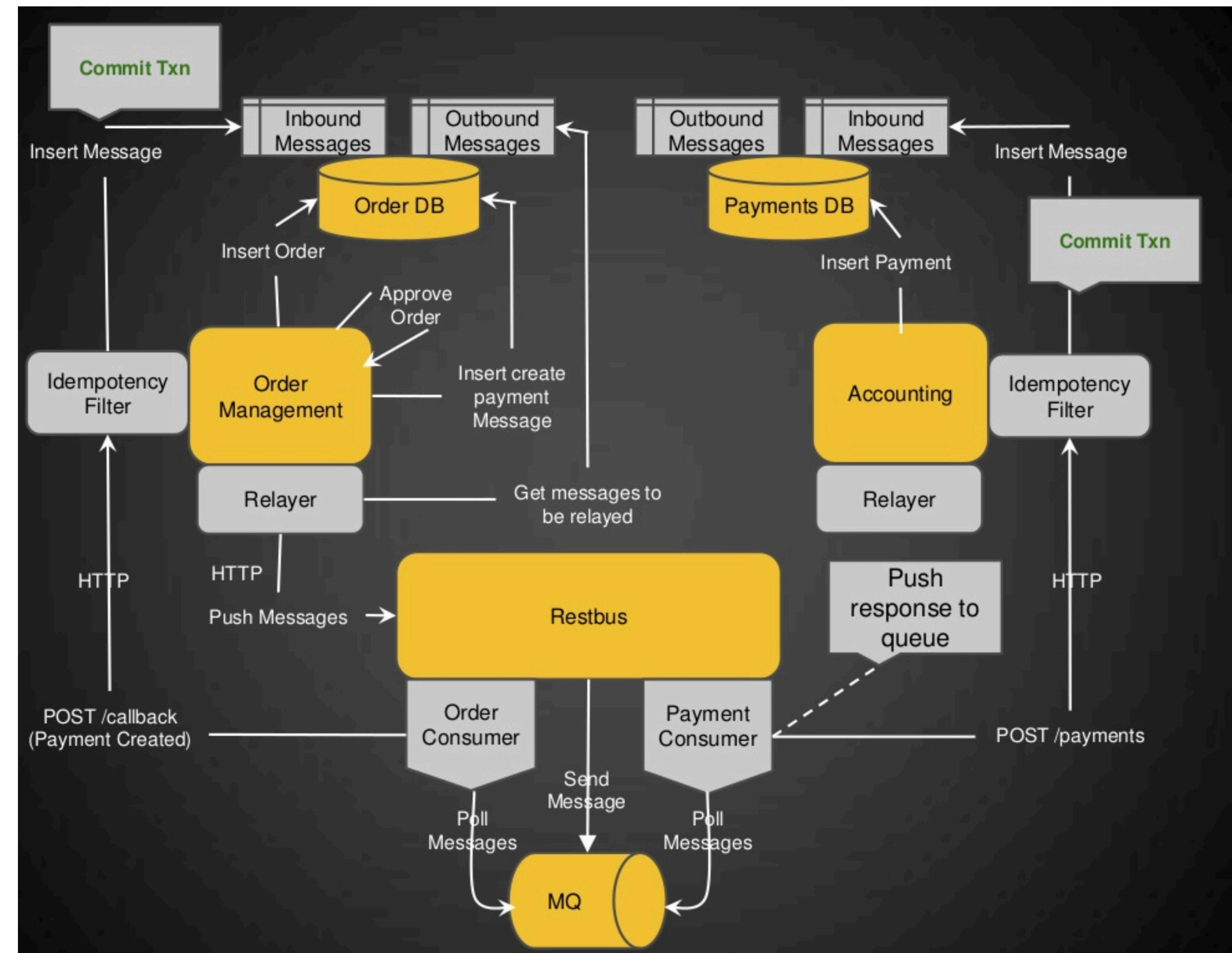
# PaaS services (Domain specific)

- Distributed Transactions



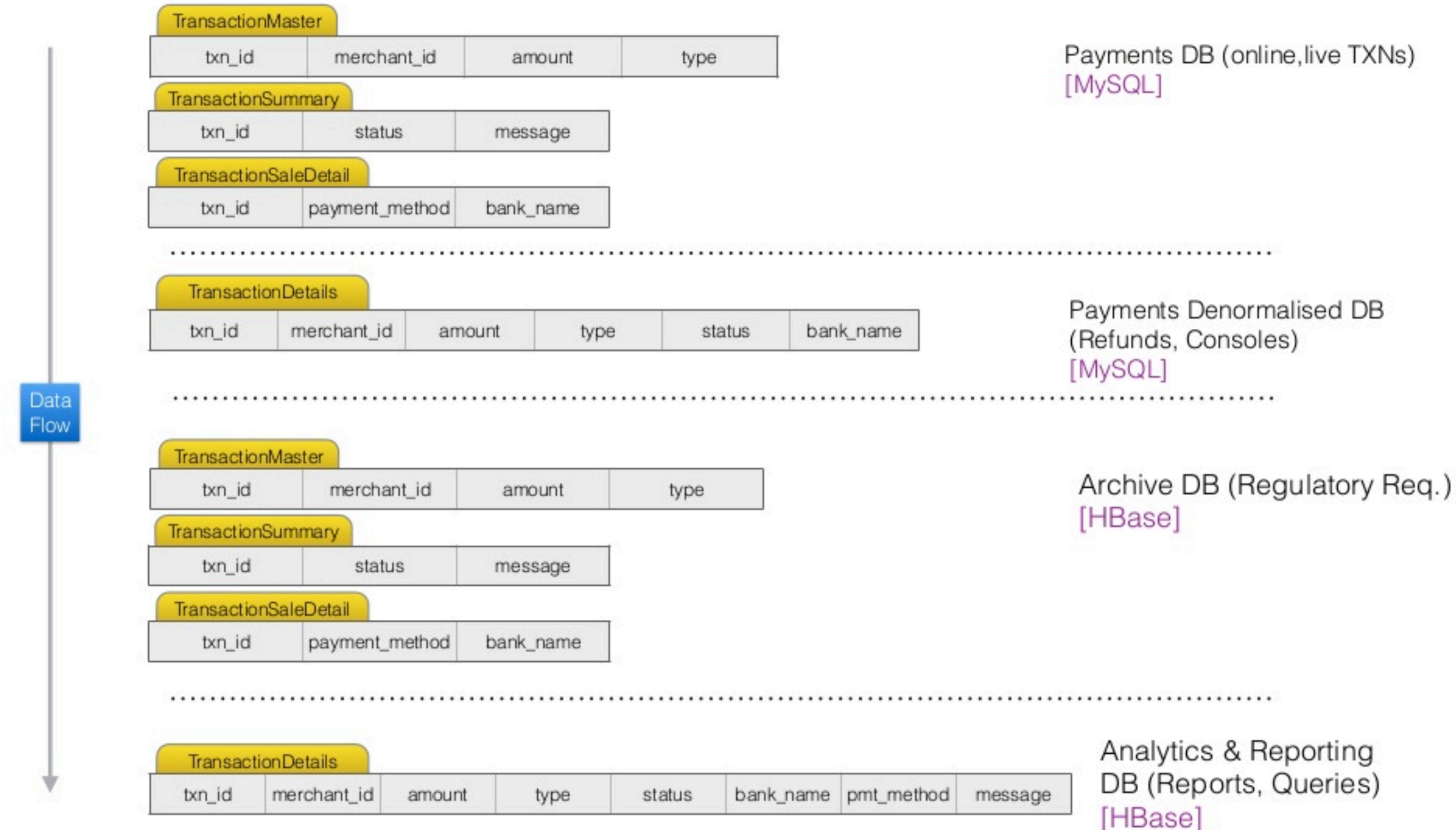
# PaaS services (Domain specific)

- RESTBus, Varadhi
  - Transactional
  - Relying in SOA



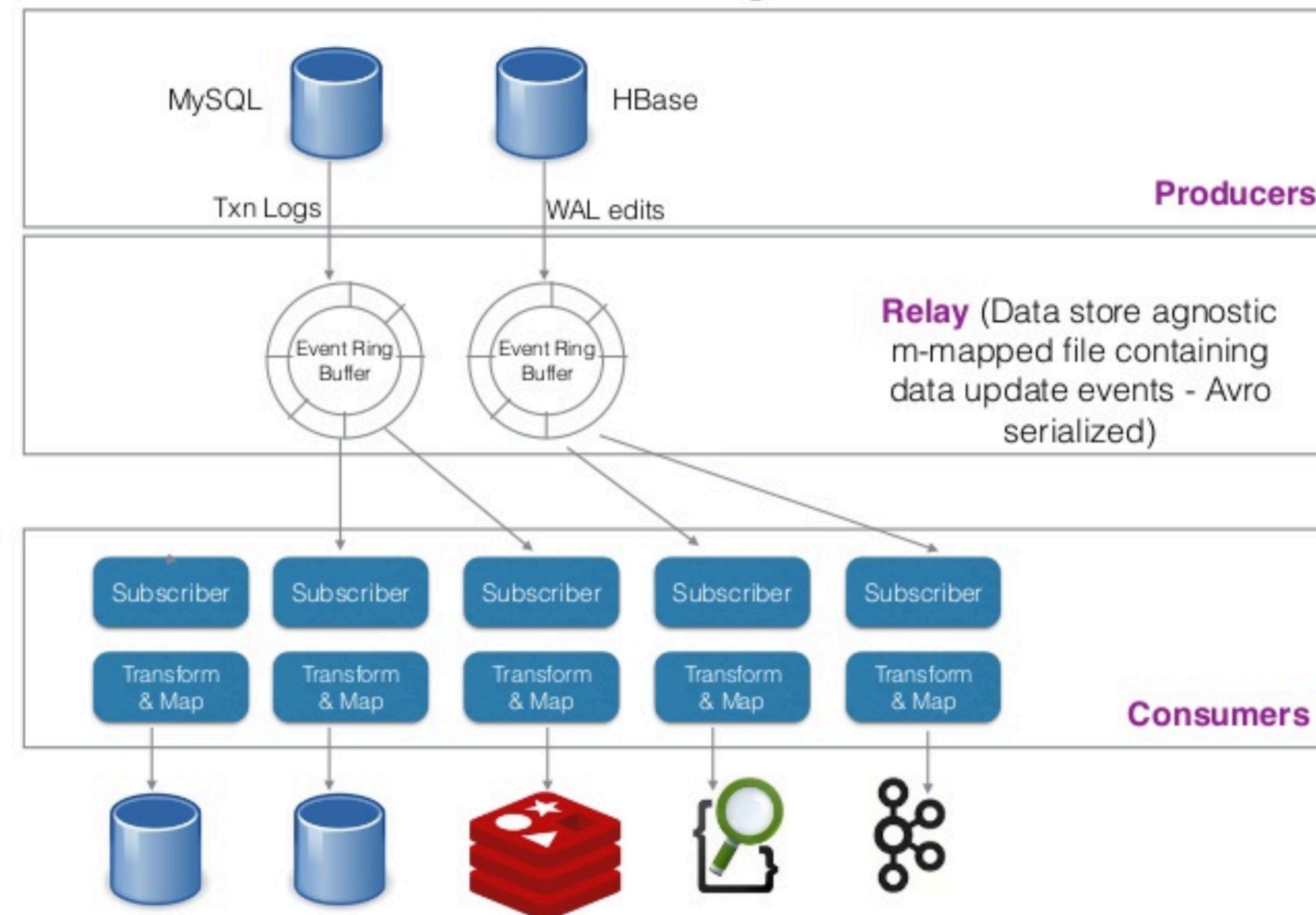
# Data Stores evolution

# Tiered-Data Stores



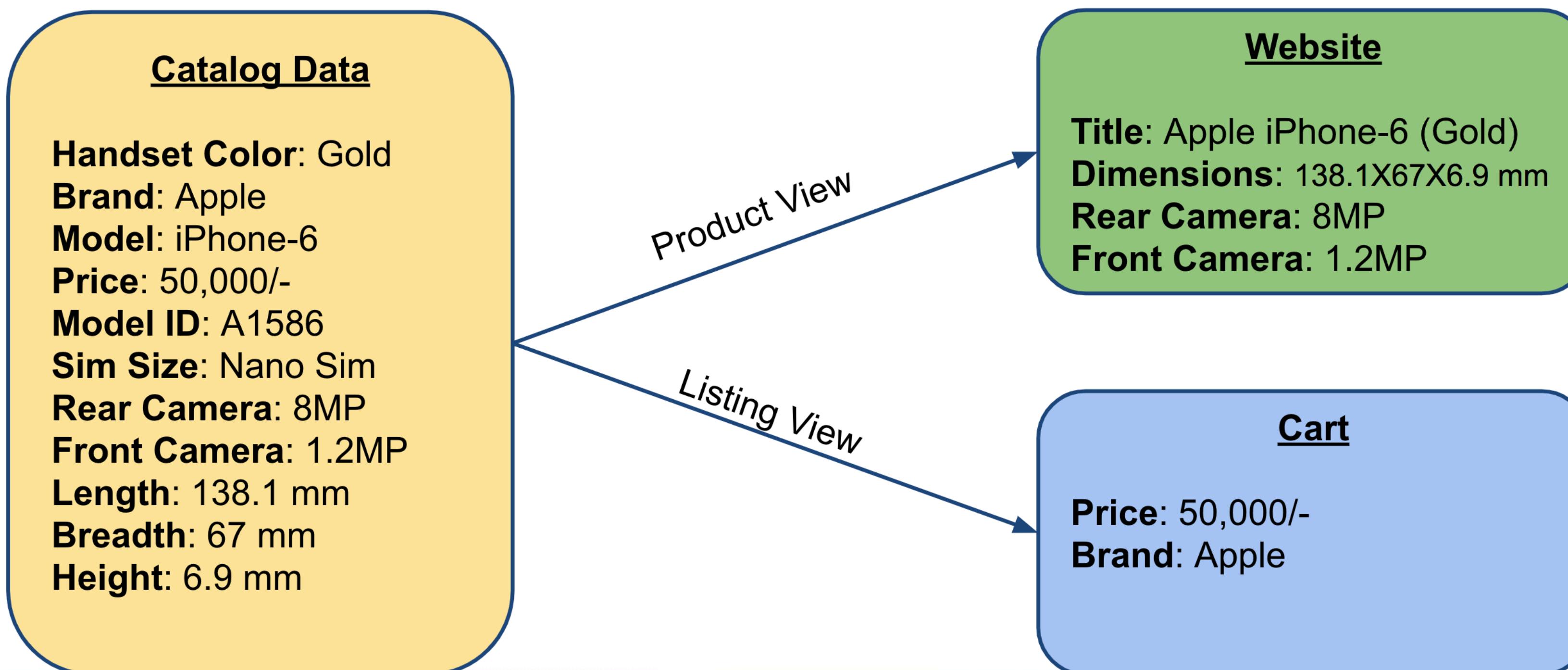
# Change Data Capture - tapping Data Replication

- Aesop (<https://github.com/Flipkart/aesop>)



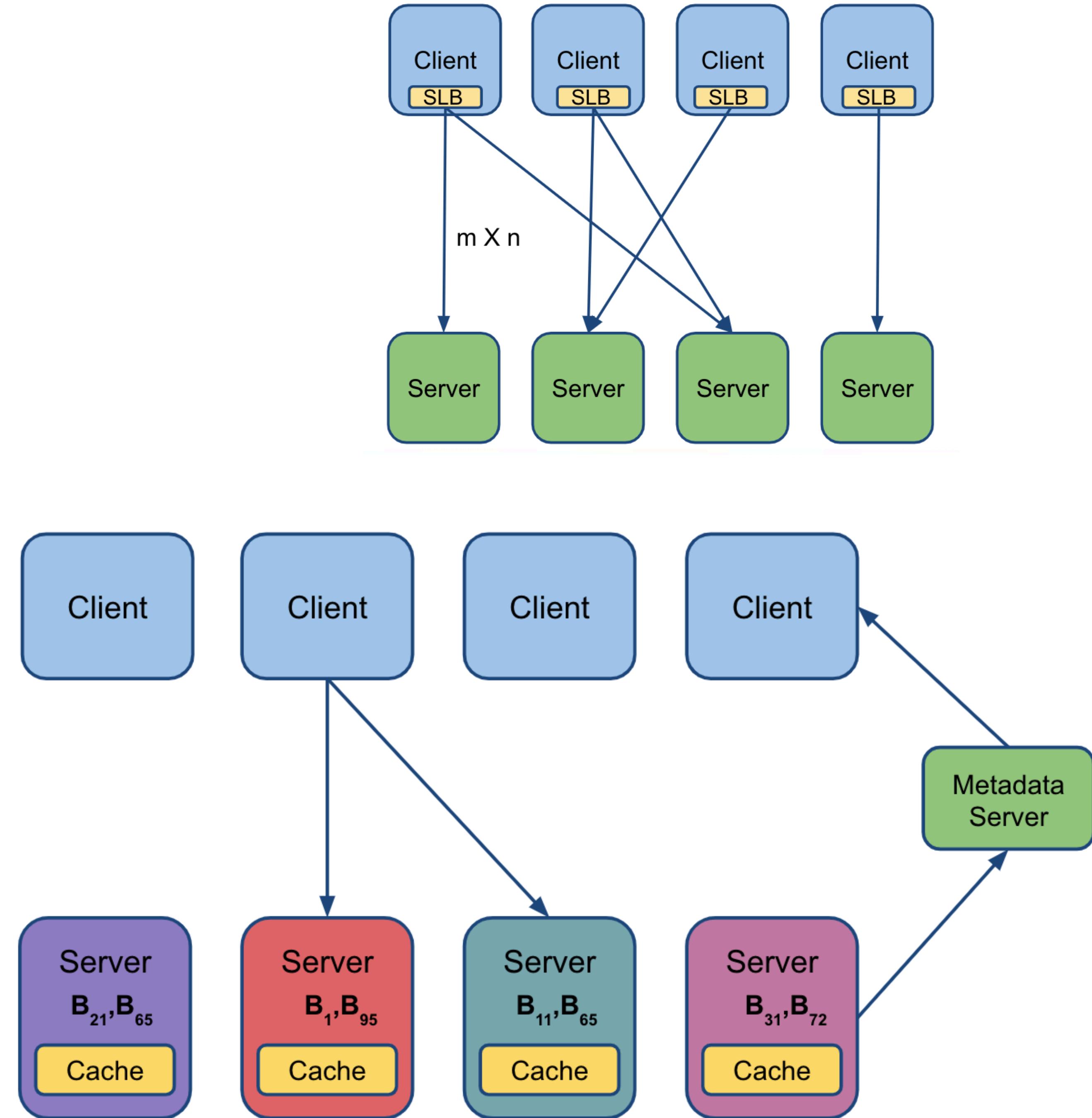
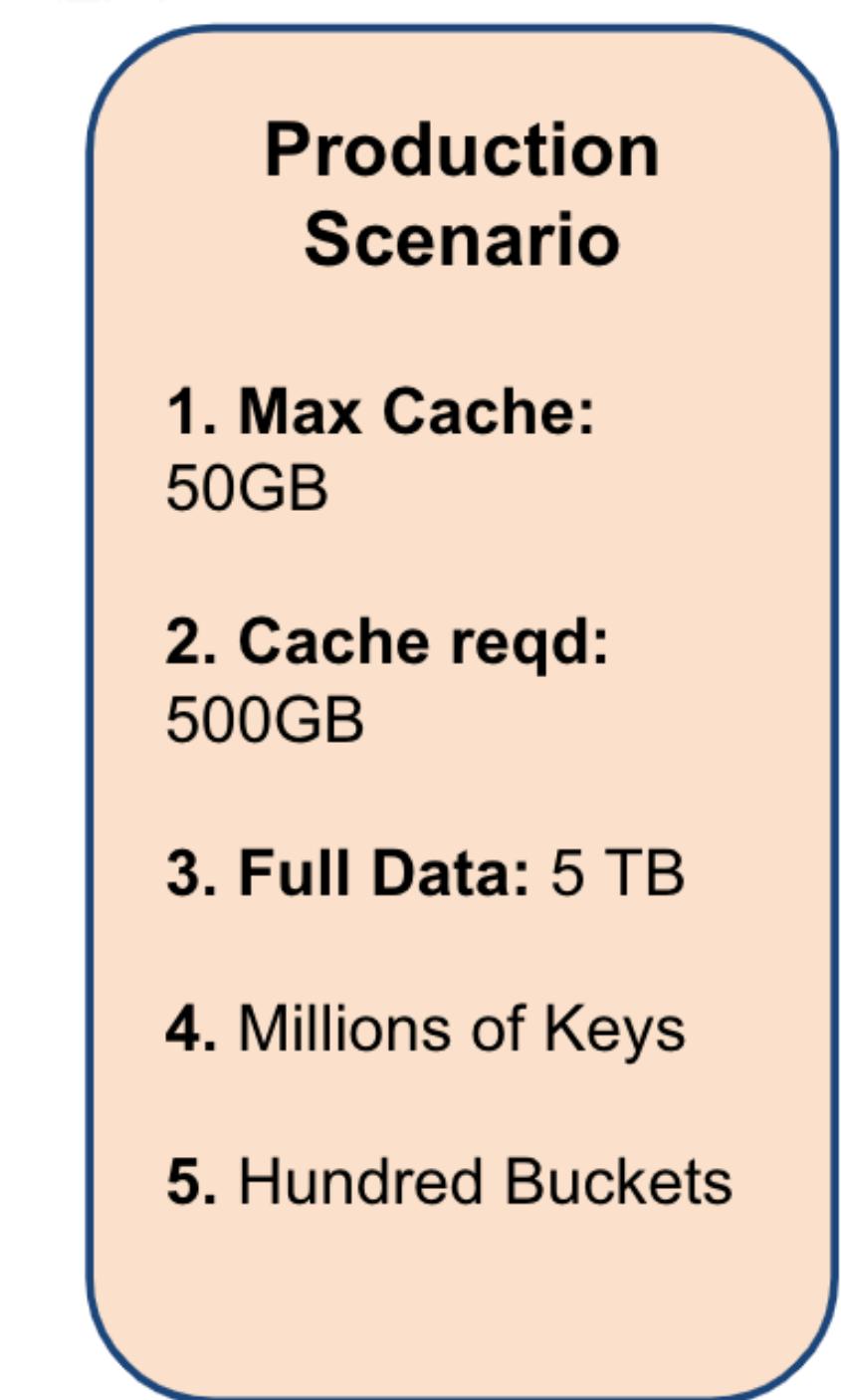
# Catalog Service

- Diverse set of data attributes (iPhone vs Baby Diaper)
- Large scale (3M “View” Reads/Sec)
- Hot spotting of data - specific Products (e.g. during Flash sale)



# Catalog Service

- Conventional proxy load balanced service does not work (requires a 190+ Gbps network)
- Data location aware Smart client, Tuneable replication factor (via vBucket)
- CQRS - Separate write, read Data stores



# Data Store for write-heavy OLTP workloads

The screenshot shows a shopping cart interface from the Flipkart website. At the top, there's a search bar and a user profile dropdown for "Gokulvanan". The main area displays "MY CART (1)" with a single item: "Flipkart SmartBuy Cotton Braided Charge & Sync USB ...". The item details include a small image of a black USB cable, the price "₹179", and a "55% Off" discount from "₹399". It also mentions "2 Offers Available". Below the item are quantity controls (-, 1, +) and buttons for "MOVE TO WISHLIST" and "REMOVE". To the right of the item, delivery information shows "Delivery by Tomorrow, Sun: ₹40" and a "10 Days Replacement Policy". The "PRICE DETAILS" section lists the "Price (1 item)" as ₹179, "Delivery Charges" as ₹40, and "Amount Payable" as ₹219. A note indicates "Your Total Savings on this order ₹180". At the bottom, there are buttons for "[CONTINUE SHOPPING](#)" and a large orange button for "[PLACE ORDER](#)". A promotional offer for "Add Worth ₹321 more for FREE delivery" is shown, noting it's "Eligible only for Assured products". A "Browse Super Value store" link is at the bottom right.

## Shopping Cart

Type of Store:  
CP

Access patterns:  
K-V, Write and Read  
heavy  
Hot Reads (Time,  
frequency)

Read Write Ratio:  
2:1

Throughput:  
100K QPS

# Data Store for write-heavy OLTP workloads

The screenshot shows a shopping cart on the Flipkart website. The cart contains one item: a Flipkart SmartBuy Cotton Braided Charge & Sync USB cable. The item is priced at ₹179, down from ₹399, with a 55% discount. It has a 10-day replacement policy. The delivery address is set to 560103, with delivery by tomorrow at ₹40. The total amount payable is ₹219, with a savings of ₹180. There is an offer to add Worth ₹321 more for free delivery, eligible only for Assured products. Buttons for 'CONTINUE SHOPPING' and 'PLACE ORDER' are visible.

MY CART (1)

560103 Change

Flipkart SmartBuy Cotton Braided Charge & Sync USB ...  
Flipkart SmartBuy EU31M USB Cable, Black, 1 Mtr  
Seller: RetailNet Assured

₹179 ₹399 55% Off 2 Offers Available

MOVE TO WISHLIST REMOVE

CONTINUE SHOPPING PLACE ORDER

PRICE DETAILS

Price (1 item)	₹179
Delivery Charges	₹40
Amount Payable	₹219

Your Total Savings on this order ₹180

Add Worth ₹321 more for FREE delivery  
Eligible only for Assured products

Browse Super Value store >

## Checkout

Type of Store:  
CP

Access patterns:  
K-V, Write and Read heavy  
Hot Reads (Time, frequency)

Read Write Ratio:  
2:1

Throughput:  
100K QPS

# Data Store for write-heavy OLTP workloads

My Account > My Orders

OD211684454285350000

ADK LR01&02 New Stylish Attractive Combo... ₹318 Cancelled  
Seller: IndianWatches OFFERS: 2 Your item has been cancelled

Ordered On Fri, Mar 2nd '18 Order Total ₹318

OD211607291974542000

Laxmi Ganesh Billiard SNOOKER CUE STICK ... ₹290 Cancelled  
Color: Black Size: 2 Seller: LaxmiGaneshBilliard OFFERS: 2 Your item has been cancelled

Laxmi Ganesh Billiard SNOOKER CUE STICK ... ₹290 Cancelled  
Color: Black Size: 2 Seller: LaxmiGaneshBilliard OFFERS: 2 Your item has been cancelled

Laxmi Ganesh Billiard SNOOKER CUE STICK ... ₹290 Cancelled  
Color: Black Size: 2 Seller: LaxmiGaneshBilliard OFFERS: 2 Your item has been cancelled

Ordered On Wed, Feb 21st '18 Order Total ₹870

OD211526930326422000

Titan NE2370YM01 Raga Watch - For Women ₹3,105 Delivered on Fri, Feb 16th '18  
Seller: EMAZING Your item has been delivered

Ordered On Mon, Feb 12th '18 Order Total ₹3,105

## Orders

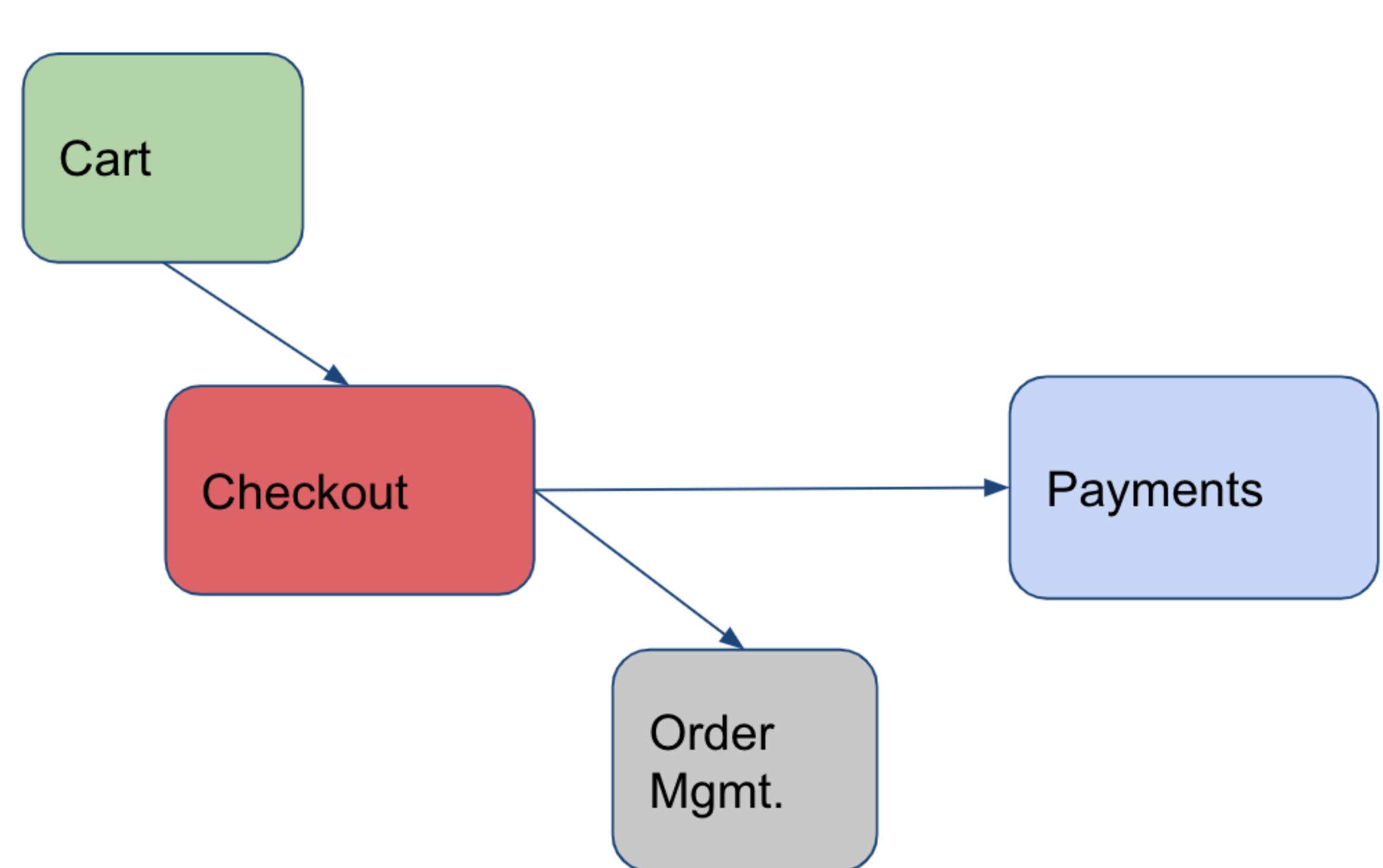
Type of Store:  
CP

Access patterns:  
K-V, Index reads  
Read Heavy, Random reads

Read to Write Ratio:  
10:1

Throughput:  
120K QPS

# Order management data stores (multiple)

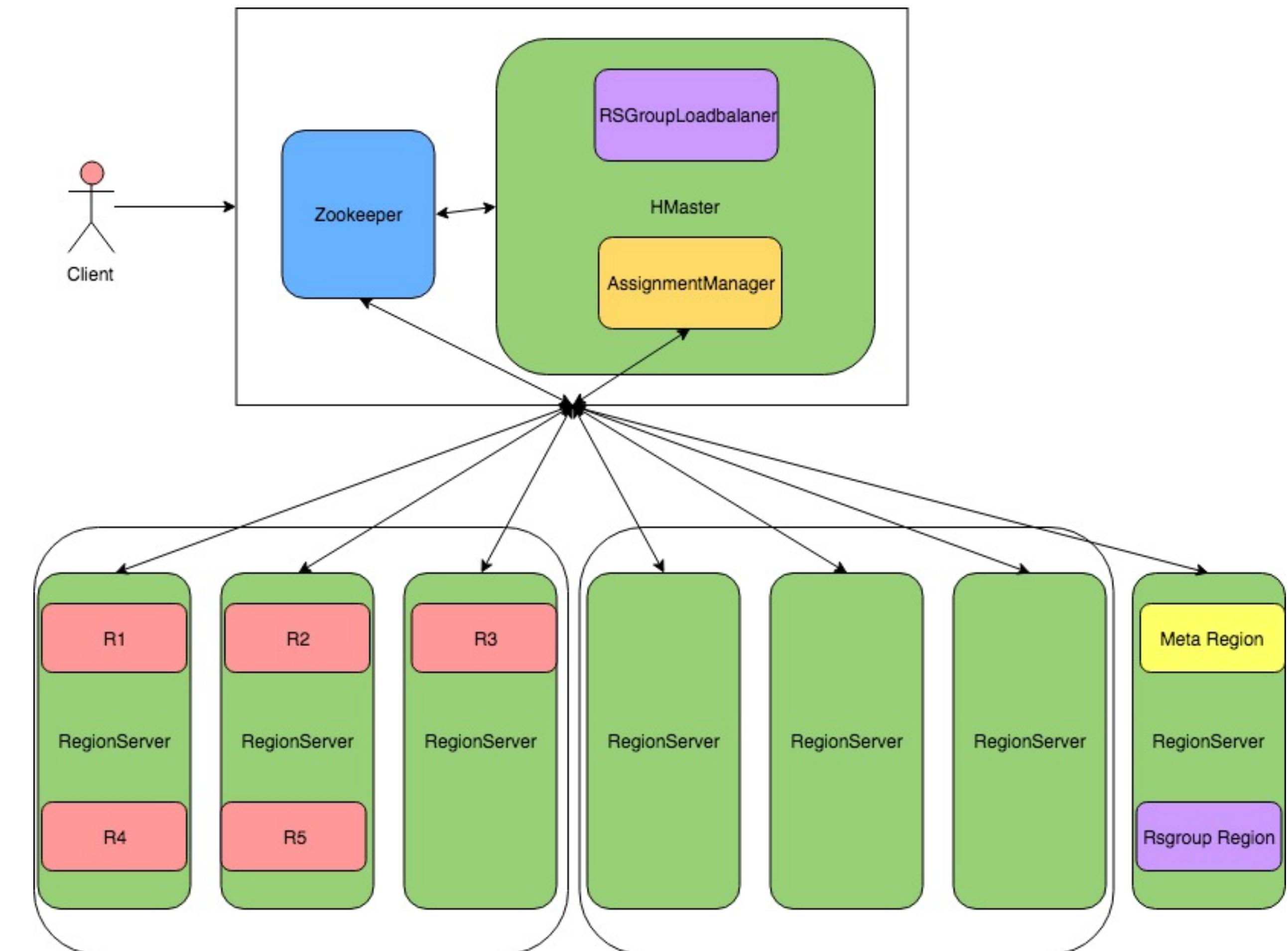


# YAK - Custom Multi-tenant HBase cluster with tenant isolation

- Multi-tenancy : Backported RSGroup to HBase 1.x
- Custom data placement LoadBalancer

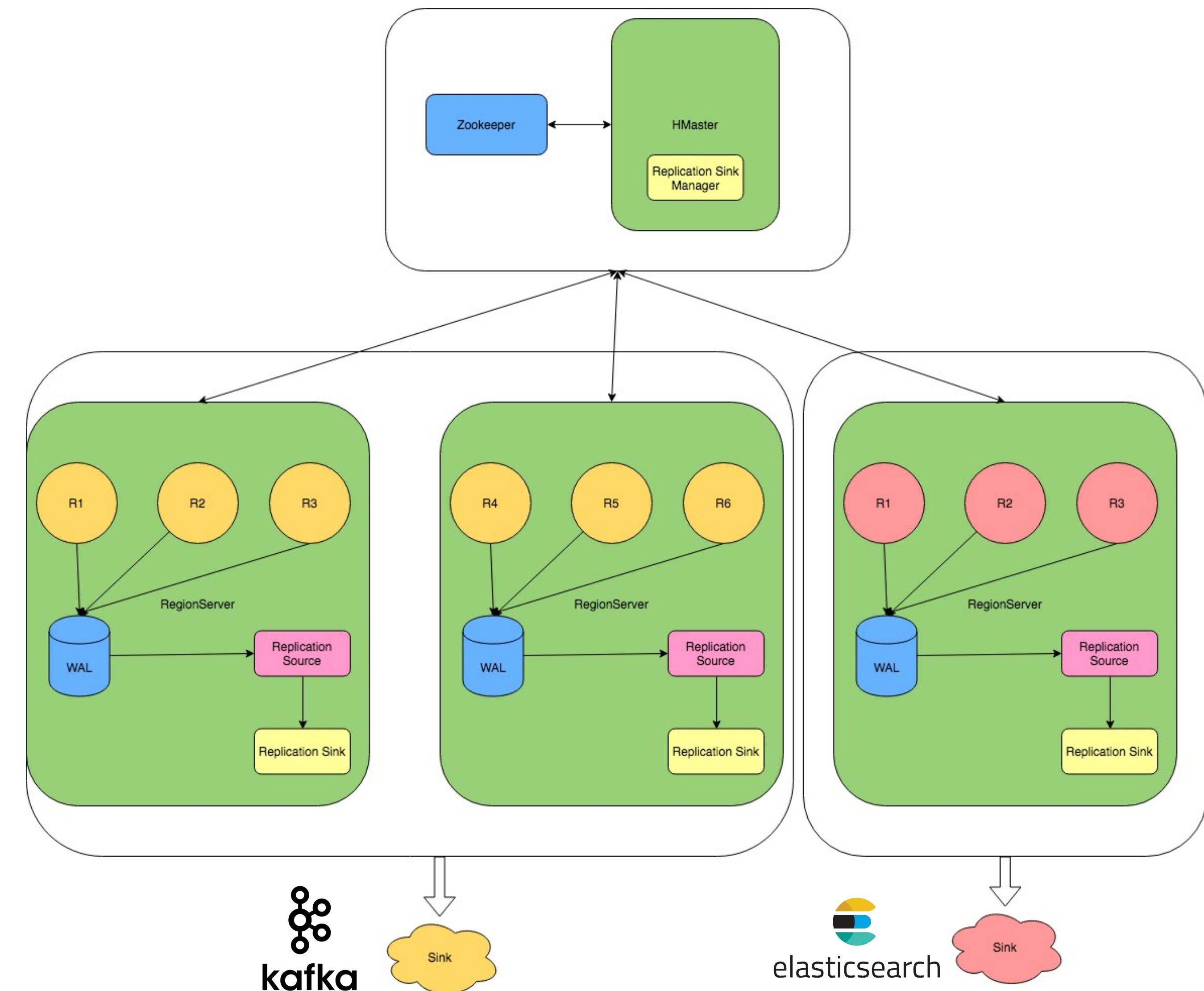
Objective of Multi-Tenancy:

- Avoid regions on one table/tenant interspersed with other table/tenant
- Define logical subclusters on regionserver per table/tenant



# YAK - Custom Multi-tenant HBase cluster with tenant isolation

- Eventual Consistency,  
Change Data Capture :  
ReplicationSink(s) for Kafka,  
Elastic Search



# YAK - Custom Multi-tenant HBase cluster with tenant isolation

Simple indices : Dual write to Index table and Data table

Create/update a key - dual writes:

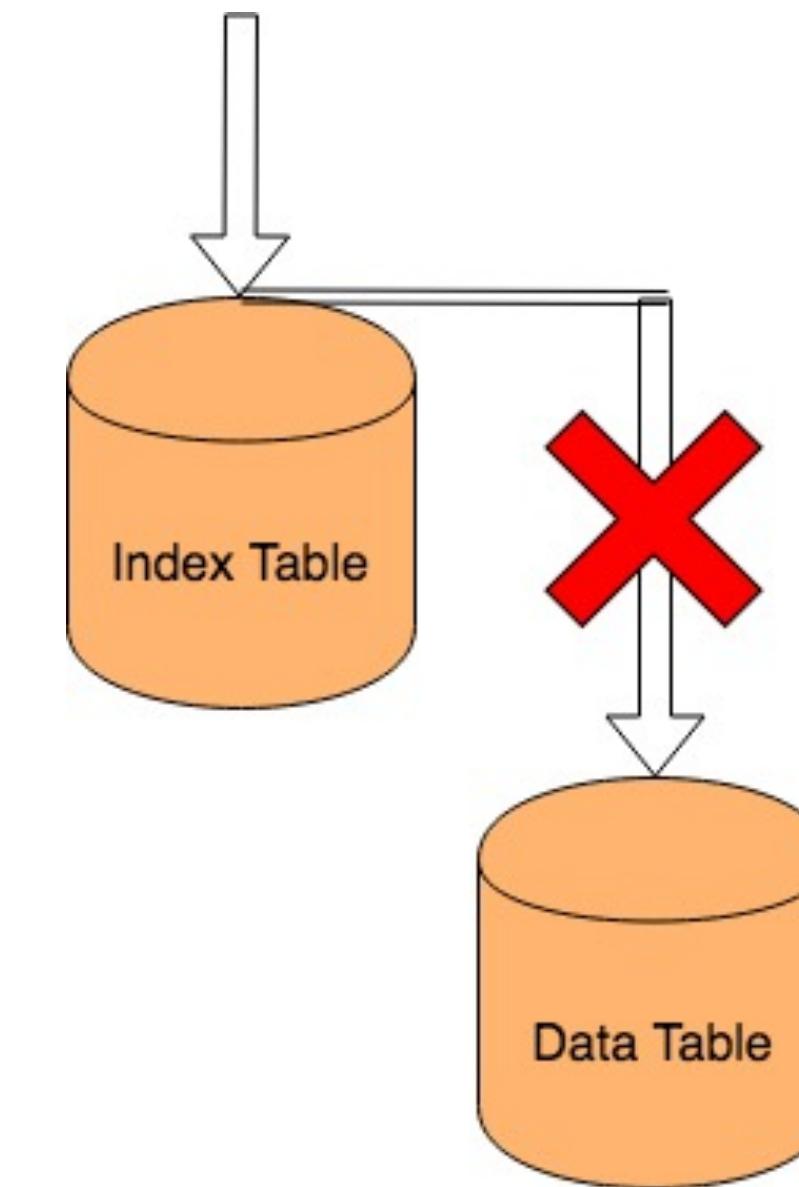
- write to index (transactionId => orderId)
- write to data (orderId => data)

Handle distributed transaction using MVCC trick with read repair.

Client-end dual read of index & data offers full consistency.

Trade-offs:

- Additional call for Index write/read
- False positive index keys is possible



## Summary:

Single HBase cluster with multi-tenancy

- Predictable horizontal scalability for tenant
- Lesser operational overhead

Software oblivious to hardware choices: works on SSD, SATA

Handles throughput of 500k/s at the cluster level

# Data Stores for Multi-homed Services

- Services running across multiple DCs in Active-Active mode. For BCP, what is ideal experience in these scenarios
  - User logs into DC1, changes password. Failed over to DC2 immediately after
  - User orders products in DC1. Failed over to DC3 immediately after
  - Flash sale for a Product where users can buy across DC1, DC2

Varied Data stores : Vitess MySQL(User), YAK (Orders), Redis/Aerospike (Inventory)

Tune-able consistency : Strong, Eventually Consistent

Low Latency, High Read throughput, High write throughput (Inventory Counters)

No Data Loss, ideally want to preserve existing DC-local data stores i.e. MySQL, YAK, Redis/Aerospike , patterns like CQRS, Data operations (backup/restore)

# Work In Progress...

- Understanding Distributed Consensus - e.g. using Raft
- Nexus - Data replication library built on Raft
- Pluggable storage layer - MySQL, YAK, dkv
- RocksDB as KV storage engine - durable, backup/restore, LSM based (SSD friendly)
- dkv (<https://github.com/flipkart-incubator/dkv>) - Distributed KV data store with tunable consistency, replication
- Data consistency checks using techniques like Merkle Trees, MVCC based updates to address concurrency issues etc.
- Hardware/Cloud instance types to support newer Data Store implementations - e.g. Raft log persistence, Off-heap HBase Block cache.

Thank You  
Questions?