***Grill Fresh Online***

Contents

[1. Introduction 4](#_Toc58000845)

[2. Context 4](#_Toc58000846)

[2.1 Grill Fresh Online (GFO) 4](#_Toc58000847)

[3. Functional Overview 5](#_Toc58000848)

[4. Quality Attributes 6](#_Toc58000849)

[5. Constraints 7](#_Toc58000850)

[6. Architecture Principles 7](#_Toc58000851)

[6.1 Microservices Architecture 7](#_Toc58000852)

[6.2 API Gateway / BFFs 7](#_Toc58000853)

[6.3 Communication 8](#_Toc58000854)

[6.4 Diagnostics and Observability 9](#_Toc58000855)

[7. Architecture 9](#_Toc58000856)

[7.1 Applications 9](#_Toc58000857)

[7.2 Shopping Gateway / BFFs 10](#_Toc58000858)

[7.2.1. Identity Service 10](#_Toc58000859)

[7.2.2. Catalog Service 10](#_Toc58000860)

[7.2.3. Shopping Cart Service 11](#_Toc58000861)

[7.2.4. Order Service 11](#_Toc58000862)

[7.2.5. Payment Service 11](#_Toc58000863)

[7.2.6. Store Service 11](#_Toc58000864)

[7.3 Executive Gateway / BFFs 12](#_Toc58000865)

[7.3.1. Inventory Service 12](#_Toc58000866)

[7.3.2. Insights Service 12](#_Toc58000867)

[7.3.3. Sentiments Service 12](#_Toc58000868)

[8. Sequence Diagrams 12](#_Toc58000869)

[8.1 Order workflow 12](#_Toc58000870)

[8.2 Insights Analytics workflow 13](#_Toc58000871)

[8.3 Sentiments workflow 13](#_Toc58000872)

[9. Design Diagrams 14](#_Toc58000873)

[9.1 User Authentication and Authorization 14](#_Toc58000874)

[9.2 Order Capture (Shopping) 14](#_Toc58000875)

[9.3 Order Fulfillment 14](#_Toc58000876)

[9.4 Inventory Management 14](#_Toc58000877)

[9.5 Insights and Analytics 14](#_Toc58000878)

[10. Deployment 14](#_Toc58000879)

[10.1 API Gateways 14](#_Toc58000880)

[10.2 Auto-scaling on demand 15](#_Toc58000881)

[10.2.1 Scaling-out 15](#_Toc58000882)

[10.2.2 Scaling-in 15](#_Toc58000883)

[10.3 Blue / Green Deployment 16](#_Toc58000884)

[10.4 Deployment Landscape 16](#_Toc58000885)

[Compute instances for event bus 16](#_Toc58000886)

[Compute and storage instances for application services 17](#_Toc58000887)

# Introduction

This document offers a high-level overview and explains the proposed architecture for “Grill Fresh Online”, a pizza online ordering system for Grill Fresh Inc.

# Context

Grill Fresh Inc wants to grow its business across the country in a gradual manner, in terms of customer base, volume of business and number of stores. It wants to achieve this goal through the following accelerators of business:

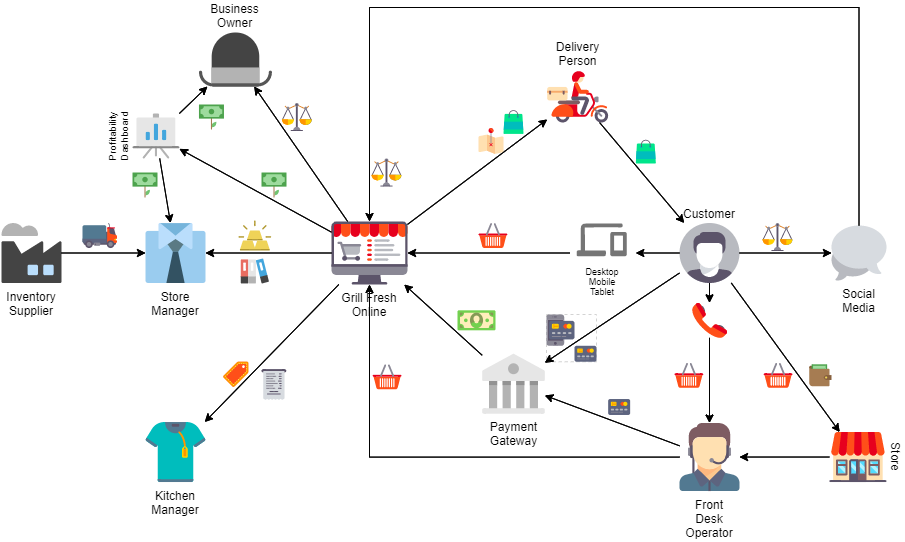
* A customer places orders for pizzas through online with any nearby store – instant and easy
* A customer receives delivery through additional modes like Take away and Home delivery, apart from dine-in option
* The business owner gathers a holistic view of profitability at business level or individual store level – to locate customer interests and expansion opportunities
* A store owner gathers a holistic view of profitability at store level – to understand business health and to act
* The business owner and store owners prefer to understand the sentiments of customers about business – to collate and correlate feedbacks (in near real-time, if possible)

Grill Fresh Inc is convinced that the potential enabler for this goal is to embrace the online platform opportunities.

## Grill Fresh Online (GFO)

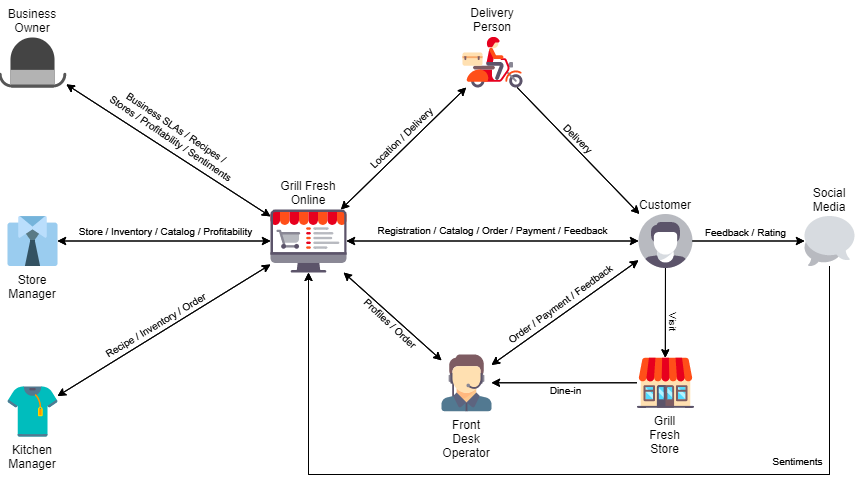
GFO system is a unified platform for all stakeholders of business – customers, business owner, store manager, front desk operator, kitchen manager and delivery person – catering to their varied interests. It is accessible through various devices – desktops, laptops, tablets, smart phones and Point-of-Sale.

At a high level, this system along with its neighborhood systems:



# Functional Overview

Various major workflows fulfilled by the GFO system are as described below:



* Customer:
  + Register and manage profile – contact details and preferences
  + Browse through catalog of available items – pizzas, toppings, sides, desserts, beverages
  + Place orders for take away and home delivery
  + Pay for orders using varied secure payment options
  + Track status updates for orders placed
  + Provide sentiments about a store, item or overall business by direct feedback or over social media
* Store Front Desk Operator:
  + Browse through catalog and place orders on behalf of a customer – for dine-in, take away and home delivery
  + Approve and move orders into queue for Kitchen Manager
  + Track status updates for all orders placed in store
* Kitchen Manager:
  + Browse through recipes and ingredients for preparing items
  + Check inventory of ingredients and notify purchase needs
  + View and receive notification about incoming orders in queue
  + Process and notify status updates for orders
* Delivery Person:
  + Pick-up out-for-delivery orders from delivery queue
  + View order and delivery information for customer
  + Confirm after delivering orders
* Store Manager:
  + Manage store – address, serving radius, front desk operators, kitchen managers and delivery persons
  + Manage catalog of items serving in store
  + Manage inventory and purchases for ingredients
  + Find profitability of store
  + Find sentiments from customers about store
* Business Owner:
  + Set business SLAs – order fulfillment time, max inventory levels, delivery charges, etc.
  + Manage stores – stores and store managers
  + Manage catalog of items serving in business and prices
  + Manage recipes and ingredients for preparing items
  + Find profitability of stores
  + Find sentiments from customers about business and stores

# Quality Attributes

GFO needs to fulfill following quality expectations of stakeholders to offer a flawless, convenient and delightful experience:

* Availability
  + System should be available 24x7 to fulfill important workflows:
    - Receive orders from customers
    - Fulfill business activities inside stores – recipe, inventory, order and payment management
  + A downtime of “< 1.5 minutes per day” can be tolerated {[99.9% (3N) availability](https://uptime.is/three-nines) 🡺1m26s per day downtime}
  + System shall continue being available even during deployment, if possible.
* Security
  + System should be available only to authorized users reaching through internet cloud.
  + System should support role-based authorization.
  + System should support different authentication mechanisms for easy and intuitive onboarding experience:
    - Internal identity provider based on stored credentials
    - External authentication providers: Google, Facebook and Twitter
  + System should be deployed within the country for statutory data compliance requirements.
* Scalability
  + System should initially support workflow traffic happening across 2 existing outlets.
  + The workflow traffic is expected to increase gradually with an estimated growth plan for next 5 years, as described below (doubling every year), with number of customers growing to 80 lakhs:

|  |  |  |  |
| --- | --- | --- | --- |
| **Year** | **# of cities** | **# of stores** | **# of users (lakhs)** |
| 1st | 5 | 15 | 5 |
| 2nd | 10 | 30 | 10 |
| 3rd | 20 | 60 | 20 |
| 4th | 40 | 120 | 40 |
| 5th | 80 | 240 | 80 |

* + There should not be any impact in system response time due to such planned increase in workflow traffic.
* Reliability
  + System should 100% reliably capture shopping cart details even on client-session expirations.
  + System should 100% reliably capture orders once the cart is checked-out and payment is made.
  + System should 100% reliably track and notify order status.
  + System should 100% reliably alert on lower inventory levels to raise purchase orders on time.
* Accuracy
  + System should 100% accurately calculate prices, taxes and other charges for every order.
  + System should 100% accurately pass approved orders to kitchen queue in same sequence as received.
  + System should 100% accurately calculate inventory levels based on orders and wastages.
  + System should 100% accurately calculate profitability at store-level and at business-level.
* Latency
  + System should notify about order status updates within 2 seconds.
  + System should pass approved orders to kitchen queue with zero-time delay.
  + System should fetch customer profile, wish-list items and respective availability-in-store details within 1 second for front desk operator.
  + System should fetch catalog details and respective availability-in-store details within 2 seconds for customer.
* Observability
  + System should provide continuous updates about and ability to monitor system health.
  + System should automatically restart on any failures.
  + System should provide system-wide activity logs.

Any other attributes can be evaluated periodically and decided based on more understanding of requirements.

# Constraints

* IT infrastructure – stakeholder wants to invest in IT infrastructure incrementally in concurrence to business expansion.
* Point of entry – customers shall be able to place orders from variety of devices – laptops, tablets and smart phones
* Point of service – front desk operators shall be able to place orders from point-of-sale devices and smart phones, kitchen managers shall be able to update orders from point-of-service devices and business owner and store managers shall be able to access system from laptops.
* Technology stack – modern trending development stack to be used for system to portray as technology innovator or pioneer in domain and to attract new tech savvy generation.

# Architecture Principles

## Microservices Architecture

The architecture proposes a microservices architecture implementation with multiple autonomous microservices in order to ensure:

* Highly scalable – scale and continue to offer same latency and response-time, with increasing or spike load
* Independently scalable – scale-out business capabilities on high demand (add / remove resources on need)
* Services aligned to business capabilities – ownership of business capabilities with small focused teams
* Loosely coupled – clear microservices boundaries and interaction across boundaries through well-defined APIs
* Highly cohesive and isolated – each microservice owns its data and database
* Highly resilient and non-disruptive deployment

Each microservice is developed using .Net Core framework following MVC pattern. [Polly](https://github.com/App-vNext/Polly) is used to support resiliency measures using patterns like retries and circuit breakers.

Each microservice runs inside a separate container. Each database runs inside a separate container for isolation and resilience needs.

## API Gateway / BFFs

**API gateway**

There are few concerns in exposing APIs from individual microservices directly to client:

* Typically, business capabilities to support a single functionality are spread across multiple focused microservices. For e.g. a functionality for placing an order involves the services namely “Catalog”, “Cart”, “Payment” and “Order”. APIs are very fine-grained in nature, while the client needs are much larger. Consequently, the client code that offers this functionality needs to fetch information by calling APIs from multiple services. This results in multiple roundtrips to server, thus increasing latency.
* When API endpoints change, client dependencies get disrupted and consequently need to be synchronized.

In order to address these concerns, the architecture places the APIs behind an API Gateway that takes care of fanning out a client request to multiple required microservices appropriately.

Since the API gateway acts on behalf of server by consolidating all APIs, it encapsulates all boilerplate functionalities like load balancing, authentication, authorization and more in a single place.

**Backends for Frontends (BFFs)**

There are few significant differences in supporting disparate devices like laptops and mobile devices:

* Data types – due to smaller size of screens, mobile devices tend to show a subset of data than laptops.
* Data formats – since only a subset of data needed, mobile devices can use efficient techniques like GraphQL than REST APIs.
* Network speeds – due to slower mobile network speeds, mobile devices prefer lesser requests and roundtrips.

In order to serve for these differences efficiently, the architecture defines a separate API gateway for each kind of client, in other words, a separate backend for each frontend – laptops and mobile devices. Each BFF provides a unique endpoint for its clients and then forwards the call to specific microservice or custom aggregator.

There are 4 BFFs – one for customer-facing shopping functionalities and another for internal business operation functionalities (like business and sentiments insights), one variant each for different clients.

The architecture proposes an implementation using [Envoy](https://www.envoyproxy.io/) to employ its advanced capabilities towards load balancing, health-check observability, dynamic routing, JWT and external authentications, role-based access controls.

Each BFF runs inside a separate container.

## Communication

**Angular Push Notifications**

The architecture proposes to use Angular Push Notifications to notify about status updates and other business events.

**Event-driven workflows**

This system exchanges messages (or events) between microservices to establish an orderly, reliable and efficient communication mechanism. It uses publish/subscribe and topic-based communication through messaging channels of RabbitMQ to exchange these messages.

RabbitMQ runs inside a separate container.

**Data interchange format**

The architecture proposes JSON as the preferred data interchange format for major communications between microservices and between client/server.

## Diagnostics and Observability

**Logging**

The architecture proposes a system-wide, centralized and structured logging using [Serilog](https://serilog.net/) such that it becomes easy and robust to diagnose failures faster.

Serilog sinks based centralized logging service runs inside a separate container.

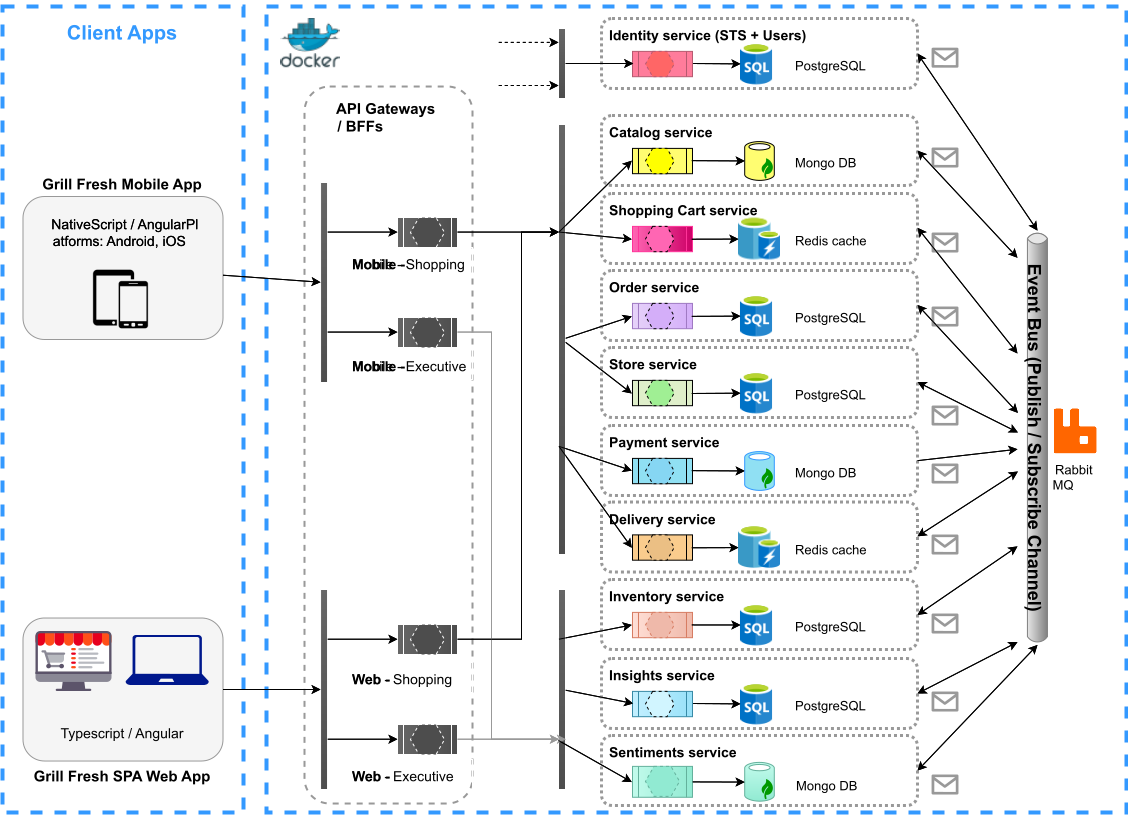
**Observability**

All applications and microservices expose endpoints that check the application and all its dependencies.

Health check APIs run available from every microservice and health check monitoring service runs on a separate container.

# Architecture

**Anatomy of the proposed architecture**



## Applications

The architecture proposes two variants of the application that can be respectively accessed from mobile devices and laptops.

**SPA (Single Page Application) Web App**

The architecture proposes to offer the laptop client app as a SPA style Web App. It is developed using Typescript in Angular framework.

**Native Mobile App**

The architecture proposes to offer the mobile client app for both Android and iOS platforms. It is developed using NativeScript in Angular framework.

Each client application runs from a different container.

**Swagger UI (for Open API)**

Also, all microservices allow anyone (with right credentials) to access the APIs. Swagger UI for a microservice is accessible from the Swagger UI page for the respective microservice. It is turned off in production environment.

## Shopping Gateway / BFFs

The shopping gateway organizes all the APIs required to fulfill customer-facing shopping functionalities. This aggregator serves into 2 BFFs, one variant each respective for 2 different client apps (mobile and laptop devices). The microservices under this gateway are described in this section below.

### Identity Service

The architecture proposes to have a separate Identity Management Service. Identity STS (Security Token Service) generates a JWT per every successful login session. Auto logoff is activated on session expiry (set to 20 minutes presently).

The architecture proposes to implement JWT and external authentications and role-based access controls (claims-based) inside STS service. Also, it proposes to use [Envoy](https://www.envoyproxy.io/) at Shopping Gateway level such that it is straightforward to decide whether to forward the requests further based on the identity and role. At present, external authentications are supported for the customers based on Google, Facebook and Twitter authentications. Only security-token based authentication mechanisms are supported for internal employees.

The user profiles and roles are stored inside PostgreSQL database. The user credentials are encrypted before storing inside database.

### Catalog Service

The catalog service serves two purposes:

* Managing the catalog of items – some items though available at business-level may not be available at individual store-level.
  + The business owner sets the catalog of items at the holistic business level.
  + The store manager may decide at store level to offer only a subset of items catalog specific to region.
* Managing the catalog of recipes for preparing those items.
  + The business owner sets the standardized recipes including ingredients for preparing items.
  + A kitchen manager browses through the catalog of recipes as standardized.

Mongo DB is used to store both catalogs of items and recipes.

### Shopping Cart Service

This service manages all the user activities and business logic with respect to managing the shopping cart. This includes all activities just until placing an order. There are all sorts of combinations of:

* Pizza sizes (small, medium, large), pizza kinds (veg, non-veg), pizza range (value, luxury, exotic), pizza crust type (thin, hand-tossed, cheese-burst), pizza base (Maida, Wheat), pizza toppings (leaves, olives, jalapenos, etc.)
* Sides, desserts and beverages
* Items prices, taxes and delivery charges

Since this service deals with transient data, it uses Redis cache. Persistence is turned on to restore the state in case of service crashes or failures.

### Order Service

Order service comprises of two major functionalities:

* Order management
  + Persist order details, retrieve them when the customer asks about past orders (customer-to-order mapping), when the delivery person asks about delivery location (order-to-customer mapping).
  + The order details are stored inside a Mongo DB database, with sharding enabled for easier scaling purposes – the state identifier of the store location is used as the sharding key.
* Order fulfillment
  + Facilitate approval of order (by store front desk operator) and place approved orders in kitchen queue.
  + Maintain and notify all status updates for placed orders – order-placed, order-confirmed, in-preparation, ready, out-for-delivery, delivered
  + The order details along with status updates are stored inside Redis cache with persistence enabled.

### Payment Service

The payment service supports various payment modes: Credit/Debit cards, UPI, Digital wallets, one-time payment URLs. At present, these payment modes are supported using one payment gateway provider. In future, it may engage services from more than one payment gateway provider.

This service persists details about payment transactions and respective states in Mongo DB database.

### Store Service

This service manages all the store related functionalities:

* Managing serving radius and deliverable locations
* Adding, updating or deleting stores
* Managing store operational details (address, contact, business hours)
* Managing employees belonging to stores

The store related details are stored inside PostgreSQL database. The employee related details are stored inside separate PostgreSQL database.

## Executive Gateway / BFFs

The executive gateway organizes all the APIs required to fulfill internal employees’ / executives’ needs related business operation functionalities – like inventory management, business insights and sentiments insights.

### Inventory Service

This service attempts to manage the inventory levels of all ingredients at individual store level for the next “n” days, as defined by the store manager. It uses the combination of ingredients required for preparing various items and the expected amount of sales for the items in order to calculate the inventory levels for next “n” days.

* The store manager places purchase orders for the required ingredients at the right time – based on reorder levels.
* The store manager reports details about the inventory that went waste.

The inventory related details are stored inside a PostgreSQL database.

### Insights Service

This service offers various analytics to aid and improve the growth of business. Various metrics of analytics are calculated:

* Profitability – to start with, it is initially calculated as: Sales – Inventory – Wastages – Operating expenses (like salary, etc.)
* Trends of Sales, Inventory and Wastages – simple trend curves and heatmaps across days in a month

The calculated details are stored inside a PostgreSQL database.

### Sentiments Service

This service works on 2 kinds of events – direct customer feedback and social media ratings:

* Places a watch over social media looking for sentiments as expressed by various people, for example, likes, dislikes, comments – Facebook and twitter at present.
* Gathers the incoming sentiments events stream and looks for interested patterns and transforms them into ratings.
* Correlates the ratings with respect to past customer orders or other customer engagements

To start with, this service initially employs a simple brute-force pattern matching mechanism and thus it may not reflect accurate customer sentiments. In future, this service can employ any sophisticated AI based (like ML) correlation and pattern recognition algorithm for sensing the moods precisely.

The sentiments from event streams and ratings calculated from those events are stored inside a MongoDB database with sharding based on the store location enabled.

# Sequence Diagrams

This section describes the physiology or major functional workflows happening in the system.

## Order workflow

The order workflow is realized through various collaborating services publishing and consuming the events from RabbitMQ (or EventBus) designated exchanges.

It starts from “Order Fulfillment” background service receiving a subscribed event “OrderStatusOrderPlacedEvent”. This service carries on the order workflow in collaboration with few other microservices as mentioned below.

Text, letter

Description automatically generated

## Insights Analytics workflow

The Insights service right now performs simple calculations and analytics tasks as of now:

* Profitability = Sales – [Inventory + Wastages]
* Dashboard showing the observed trend metrics:
  + Profitability in a store – across various periods (days, weeks, months, years)
  + Profitability in a region – across various stores
  + Profitability in the whole business – across various regions, stores and periods
  + Like Profitability, calculate trend metrics on other dimensions: sales, inventory and wastages

Later, once we have enough data gathered over a period, this can be moved to architectures based on “time-series analytics” or even “big data analytics”.

## Sentiments workflow

The Sentiments service subscribes to the streaming APIs of “Twitter” and “Facebook”.

A simple pattern-matching transformation of “incoming mood messages” to “rating” is used in the first stage of development. It can be moved later to “Kafka stream processing” based architecture.

# Design Diagrams

## User Authentication and Authorization

## Order Capture (Shopping)

## Order Fulfillment

## Inventory Management

## Insights and Analytics

# Deployment

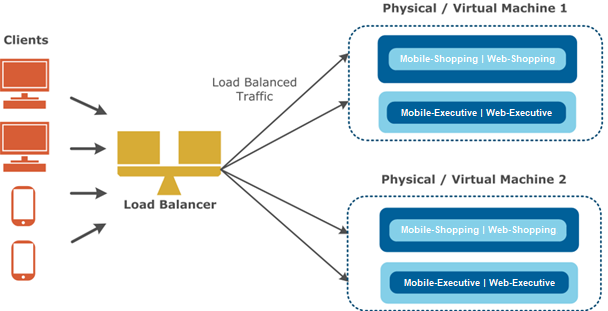
## API Gateways

Gateways (BFFs) are deployed as mentioned below in 4 different containers:

* Mobile–Shopping
* Web–Shopping
* Mobile–Executive
* Web–Executive

To avoid API gateways potentially becoming single-point-of-failure, it is proposed to place them behind “Load Balancer”.

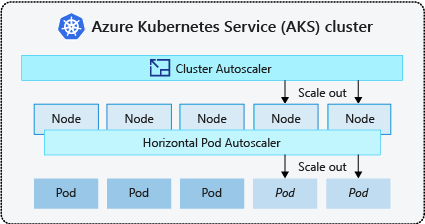
Note: the below image shows “Mobile-Shopping | Web-Shopping” as being deployed together in the same container. But they are deployed separately into their respective containers in production.



The “Load Balancer” here serves dual purpose:

* “load balancing” – round-robin
* “high availability” – forwarding traffic only to healthy nodes

## Auto-scaling on demand



Clusters can scale-out on two levels:

* Horizontal pod auto-scaler – uses “Metrics Server” in a Kubernetes cluster to monitor the *resource demand of pods*. When an application needs more resources, the *number of pods* is automatically increased to meet the demand.
* Cluster auto-scaler – watches for pods that *cannot be scheduled on nodes* because of resource constraints. The cluster then automatically increases the *number of nodes*.

### Scaling-out

When certain metrics cross designated thresholds, it can alert Horizontal pod auto-scaler to scale out more pods. E.g.:

* Number of requests waiting on “API gateways” or “Load balancer” – this might indicate a prolonged spike in number of active users
* Length of “message queues” belonging to certain RabbitMQ exchanges – this might indicate a heavy load on certain backbone business functionalities, for e.g. like “order fulfillment” queues

The increased number of pods may sometimes add pressure on Cluster auto-scaler to scale out more nodes.

### Scaling-in

When the occasional or momentary spike conditions revert to normalcy, the Horizontal pod auto-scaler may find a few pods being idle and subsequently can decide to wind them up to a minimum designated number of replicas.

This situation of reduced pods will eventually trigger winding down few nodes in the cluster to its minimum designated.

In a nutshell:

* We shall use manual scaling-out strategy for API gateway and API gateway facing services – based on perceived or predictable increase in load – e.g. on expanding number of outlets or on regular increase in the number of customers
* We shall use “auto scaling on demand” strategy for event-driven services that are facing event bus – based on spike in business functionalities scenarios

Important Note: in our context, it looks like we may not need auto-scaling on demand, because the demand is very low, as described in detail inside the section named “Deployment Landscape 🡪 Compute instances for event bus”. We can fairly achieve desired scaling for our use-cases through manual scaling-out approaches alone.

## Blue / Green Deployment

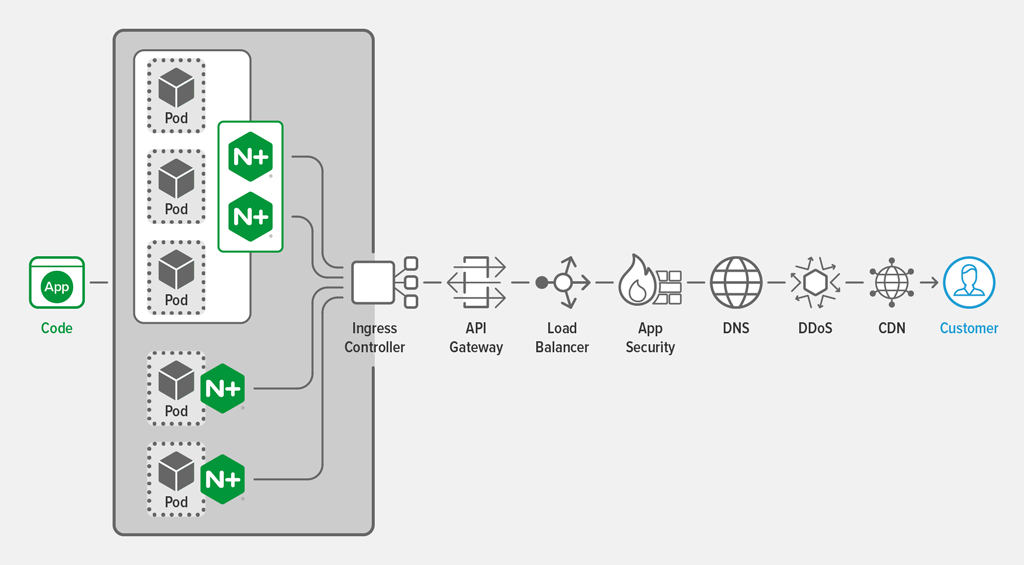
To fulfill the need of non-disruptive deployment, we shall employ the Blue / Green Deployment approach.

We shall use the combined capabilities of “Load Balancer” and “Kubernetes Cluster” to perform a rolling update over two designated environments – blue and green. This is achieved through “configuration parameterization of load balancer” and “Kubernetes helm charts”.

Alternatively, we can use infrastructure-as-code through the capabilities of “terraform” and “Kubernetes helm charts” to achieve the same.

## Deployment Landscape

This section describes the deployment landscape of proposed architecture in its motion. It would loosely look like:



### Compute instances for event bus

* Queues in event bus – we can arrive at below rough numbers based on the requirements:
  + Number of message queues – around 150
  + Almost all the services publish or subscribe to the queue for various events:
    - User added / changed / session-expire
    - Catalog added / updated / deleted
    - Order placed from shopping cart
    - Order fulfillment – order handling, payment and delivery events
    - Store alerts for inventory levels, profitability, wastages
    - Business insights and Sentiments alerts
  + Max 1 KB message sizes
  + 10 millisecond processing time per message
* Customers serving:
  + Total – 500,000
  + Per day – 30,000 (6% of total)
  + Ultra-peak hours – 6 hours (between 12:00-15:00 and 19:00-22:00) – 21,000 (70% of per day)
  + Peak hours – 3 hours (between 11:00-12:00, 18:00-19:00 and 22:00-23:00) – 4,500 (15% of per day)
  + Non-peak hours – 15 hours – 4,500 (15% of per day)
* Message rates:
  + 6 ultra-peak hours – 100 messages / second (6000 messages / minute from 60 users)
  + 3 peak hours – 40 messages / second (2500 messages / minute from 25 users)
  + 15 non-peak hours – 8 messages / second (500 messages / minute from 5 users)

**Configuration suggested:**

* Instance set 1 – 2 nodes cluster, 4 vCPUs, 4GB RAM – to serve a load of 50,400 messages of 1 KB each per day, with a peak load of 100 messages of 1 KB each per second

### Compute and storage instances for application services

* Instance set 2 – instances 2.1 & 2.2 across 2 availability zones – load-balancing pairs or replicas for high availability – these run a combination of large number of business lightweight functionalities
  + Application Gateways – 4 containers are stored
  + Identity service
  + Store service
  + Catalog service
  + Payment service
  + Delivery service
  + Identity database server – PostgreSQL database – replication and automatic failover between instances 1 & 2
  + Store database server – PostgreSQL database – replication and automatic failover between instances 1 & 2
  + Catalog database server – MongoDB database – replication and automatic failover between instances 1 & 2
  + Payment database server – MongoDB database – replication and automatic failover between instances 1 & 2
  + Delivery database server – Redis cache – distributed caching with persistence enabled between instances 1 & 2
* Instance set 3 – instances 3.1 & 3.2 across 2 availability zones – load-balancing pairs or replicas for high availability – these run the most critical and most resource hogging functionalities – these would be hit by the spike load
  + Shopping Cart service
  + Order management service
  + Order fulfillment service
  + Inventory service
  + Shopping cart data – Redis cache – with persistence enabled
  + Order data – MongoDB database – replication and automatic failover between instances 1 & 2, sharding (??)
  + Order fulfillment cache – Redis cache – with persistence enabled
  + Inventory data – PostgreSQL database – replication between instances 1 & 2
* Instance set 4 – instances 4.1 & 4.2 across 2 availability zones – for analytics jobs
  + Insights service
  + Sentiments service
  + Insights data – PostgreSQL database – replication between instances 1 & 2
  + Sentiments data – MongoDB database – replication between instances 1 & 2, sharding (??)
* Instance set 5 – instances 5.1 & 5.2 across 2 availability zones – for web server and observability kind
  + Web SPA
  + Mobile app
  + Health check monitor service (Grafana kind or .Net Core health-check client)
  + Logging service
  + Logging data – files storage – no replication