Scaling Out API

## Introduction

This document covers

## 10,000 Feet View

What is Scaling Out an API?

* Adding nodes

Why Scale Out?

* Increase throughput, concurrent users etc.
* High availability
* Fault tolerance

## Red, Green Deployment

We can use red, green deployment to facilitate rolling updates. Consider the following diagram. This gives us a load balanced, active-active deployment that is highly available. If we want to deploy we take down the green nodes and let the load balancer know. Traffic will then be redirected to blue nodes until we upgrade and bring up the green nodes again. At this point we can do the same with the blue nodes.

Its important to not we use a global load balancer. The global load balancer connects to load balancers in each data centre.

Diagram

Description automatically generated