**CORE SERVICE DESIGN:**

**Azure Load Balancer**

atabricks

|  |  |
| --- | --- |
| IT Owner Details | |
| **Department** | DTS |
| **Contact Name** | Dominic Panzera |
| **Address** | 375 Manningham Road, Doncaster, Victoria 3108 |

|  |  |
| --- | --- |
| Document Control | |
| Title | Ambulance Victoria – Azure Load Balancer Core Service Design |
| **File Name** | Ambulance Victoria – Azure Load Balancer Core Service Design v2.0.docx |
| **Version** | 2.0 |
| **Status** | Updated |
| **Release Date** | 14/02/2024 |

|  |  |  |  |
| --- | --- | --- | --- |
| Preparation |  |  |  |
| **Prepared** | Daniela Nikolic |  |  |
| **Authorised** | Dileep Pradeep |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Version Tracking | | | |  |
| Version | Remarks | Change Requested | Pages Affected | Release Date |
| 1.0 | Initial Release | N/A | All | 8/02/2024 |
| 2.0 | Updated | Minor updates | 13-18 | 14/02/2024 |

Table of Contents

[1. Overview 5](#_Toc158798426)

[1.1 Purpose and Audience 5](#_Toc158798427)

[1.2 Scope and Key Deliverables 5](#_Toc158798428)

[1.3 Glossary and Definitions 6](#_Toc158798429)

[2. Executive Summary 7](#_Toc158798430)

[3. Resource Cost 8](#_Toc158798431)

[4. WAF and Security Control Alignment 8](#_Toc158798432)

[4.1 Reliability 9](#_Toc158798433)

[4.1.1 Overview 9](#_Toc158798434)

[4.1.2 Azure Load Balancer Reliability Checklist 9](#_Toc158798435)

[4.2 Cost Optimisation 9](#_Toc158798436)

[4.2.1 Overview 9](#_Toc158798437)

[4.2.2 Azure Load Balancer Cost Optimisation Checklist 9](#_Toc158798438)

[4.3 Operational Excellence 10](#_Toc158798439)

[4.3.1 Overview 10](#_Toc158798440)

[4.3.2 Azure Load Balancer Operational Excellence Checklist 10](#_Toc158798441)

[4.4 Performance Efficiency 10](#_Toc158798442)

[4.4.1 Overview 10](#_Toc158798443)

[4.4.2 Azure Load Balancer Performance Efficiency Checklist 10](#_Toc158798444)

[4.5 Security 11](#_Toc158798445)

[4.5.1 Overview 11](#_Toc158798446)

[5. Architecture Summary 12](#_Toc158798447)

[5.1 Resource Overview 12](#_Toc158798448)

[5.2 RBAC 12](#_Toc158798449)

[5.3 Design Decisions and Justifications 12](#_Toc158798450)

[5.3.1 SKU 12](#_Toc158798451)

[5.3.2 Load Balancer Deployments 13](#_Toc158798452)

[5.3.3 Logging and Monitoring 13](#_Toc158798453)

[5.3.4 Dedicated Subnet 13](#_Toc158798454)

[5.3.5 Persistence 14](#_Toc158798455)

[5.3.6 Floating IP 14](#_Toc158798456)

[6. Azure Policies 14](#_Toc158798457)

[7. Configuration Templates 15](#_Toc158798458)

[7.1 Primary Region Load Balancer 15](#_Toc158798459)

[7.2 Primary Region Load Balancer Inbound Rule 16](#_Toc158798460)

[7.3 Secondary Region Load Balancer 17](#_Toc158798461)

[7.4 Secondary Region Load Balancer Inbound Rule 18](#_Toc158798462)

[8. Acceptance 19](#_Toc158798463)

# Overview

This document covers the baseline design for the Azure Load Balancer core service. The intention of this document is to define the overall resource design in isolation from a specific application. It is aimed to highlight the general process and requirements for building a Azure Load Balancer in a repeatable fashion with consistent configurations. Design decisions and justifications have been included in the Architecture section, and this document can be used as a reference for new builds that require a Azure Load Balancer.

This design caters to a Level 2 design which covers both Microsoft’s WAF (Well Architected Framework)[[1]](#footnote-2) and the Department of Health Control list.

Any deviations required to the standards defined in this document will require separate exemption and approval from the Cloud Governance Forum if they are required for any reason for a specific build.

## Purpose and Audience

This document will outline the standard design and configuration of this Azure service in Ambulance Victoria’s Azure tenancy as a baseline for any application infrastructure deployments.

This design is intended to:

* Meet Microsoft WAF standards.
* Meet the controls stipulated by the Department of Health.
* Define the baseline required for the deployment of the resource.

The audience for this document is those involved in the planning, designing, and implementing of the Application/Data infrastructure. This includes:

* + Ambulance Victoria IT staff

It is assumed that the reader knows and is familiar with Azure Cloud concepts and related topics.

## Scope and Key Deliverables

The scope of this core service design is to define the baseline deployment requirements and standards for the Azure Load Balancer core service.

The key deliverables for this are:

* This design to outline the service definition Level 2 baseline standards.
* A technical configuration document that defines the deployment of this resource for each of the Service Tiers, or for any other logical standard such as size
* IaC templates for repeatable deployment of this core service

## Glossary and Definitions

|  |  |
| --- | --- |
| **Term** | **Definition** |
| **AV** | Ambulance Victoria |
| **WAF** | Well Architected Framework |
| **CAF** | Cloud Adoption Framework |
| **Level 1** | Refers to a resource that has been designed to a CAF standard |
| **Level 2** | Refers to a resource that has been designed to a WAF standard with Department of Health controls overlayed |
| **AZ 2** | Refers to Ambulance Victoria’s legacy Azure Landing Zone still in use in some regards |
| **AZ 3** | Refers to Ambulance Victoria’s current Azure Landing Zone, also referred to as the Enterprise landing zone. This is the target state for migrations. |
| **SLA** | Service Level Agreement as defined by Microsoft |
| **DH** | Department of Health |
| **IaC** | Infrastructure as Code |
| **NSG** | Network Security Groups |
| **ALB** | Azure Load Balancer |
| **LB** | Load Balancer |

Table 1: Glossary and definitions

# Executive Summary

This design covers the baseline standards for the Azure Load Balancer Core Service. This service has been assessed against the five pillars of WAF as well as the Department of Health Security Controls.

This section contains a summary of the major design decisions that have been made for defining the baseline of this resource as an outcome of the WAF and Security analysis detailed throughout this document.

Of the five WAF Pillars, it was found that Reliability and Security were relevant.

For this service the main baseline configurations include:

There are some notable differences across the service tier configurations for this service:

* The Standard SKU must be used for all Production workloads – Platinum, Gold, and Silver.
* The Basic SKU may be used for Non-Production, Development, or Test workloads if a Microsoft SLA is not required, and noting that it is not secure as it is open to inbound connections by default.
* Health probes will be used to monitor health of backend pools.
* Diagnostic settings will be sent to the central log analytics workspace for that region.

# Resource Cost

The costs for Azure Load Balancer in the Standard SKU depend on whether the tier chosen is Global or Regional[[2]](#footnote-3),[[3]](#footnote-4):

|  |  |  |
| --- | --- | --- |
| Standard Load Balancer | Regional Tier Price | Global Tier Price |
| First 5 rules | **$0.038**/hour | **$0.038**/hour |
| Additional rules | **$0.016**/rule/hour | **$0.016**/rule/hour |
| Inbound NAT rules | Free | Free |
| Data processed (GB) | **$0.008** per GB | No additional charge\* |

Table 2: Pricing Construct for Standard SKU Load Balancer

|  |  |
| --- | --- |
| Gateway Load Balancer | Price |
| Gateway hour | **$0.019**/hour |
| Chain hour | **$0.016**/hour |
| Data processed (GB) | **$0.007** per GB |

Table 3: Pricing construct for Gateway Load Balancer

# WAF and Security Control Alignment

The following are the five pillars of the Microsoft Well Architected Framework:

* [Reliability](https://learn.microsoft.com/en-us/azure/well-architected/#reliability)
* [Cost optimization](https://learn.microsoft.com/en-us/azure/well-architected/#cost-optimization)
* [Operational excellence](https://learn.microsoft.com/en-us/azure/well-architected/#operational-excellence)
* [Performance efficiency](https://learn.microsoft.com/en-us/azure/well-architected/#performance-efficiency)
* [Security](https://learn.microsoft.com/en-us/azure/well-architected/#security)

For this design, the security section will also cover the Department of Health Controls in addition with any Microsoft Security Best Practices. Each of these sections will detail relevant controls or baseline requirements for this core service that will be put in place.

## Reliability

### Overview

The term reliability refers to the availability of the system and its ability to recover from failure[[4]](#footnote-5). Resiliency strategies must be built into each element of the architecture. The pillars of reliability include:

* Design for business requirements
* Design for failure
* Observe application health
* Drive Automation

### Azure Load Balancer Reliability Checklist

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Checklist Item | Applicable to AV | Built Into Design | Enforcement Option | Applicability |
| **R1** | For production workloads, use the Standard Stock Keeping Units (SKU). | Yes | Yes | IaC | At deployment |

Table 4: WAF Reliability checklist summary

## Cost Optimisation

### Overview

The cost optimisation pillar is structured to support creating cost-effective workloads in the cloud[[5]](#footnote-6). It looks at removal of unnecessary spend and improving operational efficiency. The principles of cost optimisation revolve around:

* Choosing the correct resources
* Setting up budgets and maintaining cost constraints
* Dynamically allocate and deallocate resources
* Optimising workloads whilst aiming for scalable costs
* Continuously monitoring and cost managing

### Azure Load Balancer Cost Optimisation Checklist

There is no guidance for Cost Optimisation for Azure Load Balancer.

## Operational Excellence

### Overview

Operational Excellence aims to ensure that once the architecture is built, the ongoing operations are flawless. This includes repeatable and reliable deployments, automating to eliminate human error. To do this the following must be considered:

* Optimise the build and release process (including CI/CD and IaC)
* Understand Operational Health
* Test recovery and failure
* Focus on continuous improvement
* Use loosely coupled architecture

### Azure Load Balancer Operational Excellence Checklist

The guidance for Operational Excellence was identical to the Reliability guidance, and so this has not been repeated.

## Performance Efficiency

### Overview

Performance Efficiency refers to the ability of your systems and applications to meet user demands without breaking or creating a negative user experience[[6]](#footnote-7). This covers capacity and scalability:

* Design for horizontal scaling
* Run stress and performance tests
* Continuously monitor performances, particularly in Production systems

### Azure Load Balancer Performance Efficiency Checklist

There is no guidance for Azure Load Balancer under Performance Efficiency.

## Security

### Overview

Security refers to the ability of the environment to resist and manage threats.

This section covers both Microsoft Best Practices as well as relevant security controls provided by the Department of Health. With respect to the Microsoft WAF, Security is underpinned by the following[[7]](#footnote-8):

* Plan resources and how to harden them
* Automate and use least privilege
* Classify and encrypt data
* Monitor system security, plan incident response
* Identify and protect endpoints
* Protect against code-level vulnerabilities
* Model and test against potential threats

In addition to the Microsoft controls, the Department of Health has mandated security posture to Ambulance Victoria. Note there may be duplication between the Microsoft Security Best Practices and the Department of Health controls.

The following Microsoft Security Benchmark controls are applicable to this service:

* NS-1 Establish network segmentation boundaries
* LT-4 Enable network logging for security investigation

There are no Department of Health controls above and beyond the Azure Security Benchmark for Azure Load Balancer.

# Architecture Summary

## Resource Overview

Azure Load Balancer is an Azure-native service that provides Layer 4 Load Balancing capabilities[[8]](#footnote-9). As traffic arrives at the frontend of the Load Balancer, it pushes traffic to backend pool instances as defined by load-balancing rules implemented. Backend pools consist of Virtual Machines, or an instance of a Virtual Machine in a Virtual Machine Scale Set.

Azure Load Balancers can be created as Internal Load Balancers to service and route purely internal traffic with Private IPs for the frontend only. They can also be created for Public Load balancing to balance internet traffic incoming to Virtual Machines.

Standard SKU Load Balancers are inherently secure and built on the zero-trust network security model. They are closed to inbound connections unless specifically opened by an NSG.

## RBAC

For this resource, the following RBAC can be applied:

|  |  |
| --- | --- |
| Role Name | Description |
| Network Contributor | Lets you manage networks, but not access to them. |

Table 5: RBAC roles relevant for this core service

## Design Decisions and Justifications

This section covers the design decisions and justifications that reflect the findings of the WAF and Security alignment. This will form the baseline requirements for the Azure Load Balancer core service and will be captured in the accompanying Configuration Template with a set of pre-approved deployment settings for this resource. Any changes, modifications or removals to the pre-approved deployments must have specific approval from the Cloud Governance Forum prior to deployment.

### SKU

**Design Reference:** Table 2 – [R1](#_Azure_Load_Balancer)

**Design Decision**: Use the Standard SKU for Production workloads.

**Design Justification**: Basic Load Balancers do not have any SLA[[9]](#footnote-10). This is like a service being in public preview, and it comes with no Microsoft SLA. As such it is not recommended to use the Basic SKU for Production workloads. It can be used for Non-Production workloads if an SLA is not required.

### Load Balancer Deployments

**Design Reference:** Microsoft Security Benchmark [NS-1](#_Azure_Load_Balancer_1)

**Design Decision:** Deployment of Standard SKU Load Balancers meet network segmentation requirements by default.

**Design Justification:** Azure Load Balancers can support network segmentation by deploying as an Internal Load Balancer which will only allow traffic to backend resources from within certain Virtual Networks or peered Virtual Networks without exposure to the internet. It can also be deployed with SNAT to mask the IP addresses of backend resources for protection from direct internet exposure.

### Logging and Monitoring

**Design Reference:** Microsoft Security Benchmark [LT-4](#_Overview)

**Design Decision:** Diagnostics will be deployed.For backend pools, health probes will be used by the Load Balancer to check backend health.

**Design Justification:** Diagnostics will be configured for AllMetrics and sent to the Log Analytics Workspace that is central to that region.

For health probes, to check the backend pool health, these should be configured for each backend pool so that any failures will be detected.

### Dedicated Subnet

**Design Reference:** N/A

**Design Decision:** Azure Load Balancer will be placed in its own dedicated Subnet, in the same Virtual Network as the backend it is balancing.

**Design Justification:** Though it is not a technical requirement to place Azure Load Balancer in its own subnet, it can make it simpler to control traffic from the Firewall, and allows NSGs to control traffic between the Load Balancer subnet and the subnet where the Virtual Machines it is balancing are, so overall can be a more secure and controlled architecture this way.

### Persistence

**Design Reference:** N/A

**Design Decision:** Session Persistence will be set to None as the default.

**Design Reference:** Session Persistence ensures that the sessions initiated by a client will go through to the same backend instance. This can solve incompatibility issues between Azure Load Balancer and RDP as a specific example. It ensures consistency in the connection between the Client and the backend pool. It is not required for all applications, so should be enabled if required by the application stack during the design phase.

### Floating IP

**Design Reference:** N/A

**Design Decision:** Floating IP can be used but will not be a default deployment.

**Design Justification:** Floating IP does not need to be a standard requirement for Azure Load Balancer deployments, it should be defined by the application requirements[[10]](#footnote-11). It should be used in situations where the application stack requires the use of the same port by multiple application instances on a single VM in the backend pool. Some examples of this are high availability clusters and network virtual appliances.

# Azure Policies

There are no Azure Policies relevant for this resource.

# Configuration Templates

## Primary Region Load Balancer

|  |  |
| --- | --- |
| Configuration Item | Configuration Value |
| **Name** | Internal: lbi-[env]-ause-[appname]-01  Or  Public/External: lbp-lbe-[env]-ause-[appname]-01 |
| **Subscription** | AV ALZ [Subscription Name] |
| **Resource Group** | rg-[env]-ause-[appname]-[workload]-01 |
| **Region** | Australia Southeast |
| **SKU** | Standard |
| **Type** | Public/Internal |
| **Tier** | Regional/Global |
| ***Frontend IP Configuration Settings*** |  |
| **Name** | Internal: fip-lbi-[env]-ause-[appname]-01  Or  External/Public: fip-lbe-[env]-ause-[appname]-01 |
| **Virtual Network** | vnet-[env]-ause-[appname]-01 |
| **Subnet** | snet-[env]-ause-[appname]-[workload]-01 |
| **Assignment** | Static |
| ***Backend Pools Configuration Settings*** |  |
| **Name** | Internal: bep-lbi-[env]-ause-[appname]-01  Or  External/ Public: bep-lbe-[env]-ause-[appname]-01 |
| **Backend Pool Configuration** | NIC/IP Address |
| ***Health Probe Configuration Settings*** |  |
| **Name** | hp-lbi-[env]-ause-[appname]-01  Or  hp-lbe-[env]-ause-[appname]-01 |
| **Protocol** | TCP/HTTP/HTTPS |
| **Port** | Port Number (e.g. 80,443) |
| **Interval (seconds)** | 5 |

## Primary Region Load Balancer Inbound Rule

|  |  |
| --- | --- |
| Configuration Item | Configuration Setting |
| **Inbound Load Balancing Rule Name** | rule-[protocol]-[portnumber]-[appname]-01 |
| **Frontend IP Address** | Internal: fip-lbi-[env]-ause-[appname]-01  Or  External/Public: fip-lbe-[env]-ause-[appname]-01 |
| **Backend Pool** | Internal: bep-lbi-[env]-ause-[appname]-01  Or  External/Public: bep-lbe-[env]-ause-[appname]-01 |
| **Protocol** | TCP/UDP |
| **Port** | Port Number |
| **Backend Port** | Port Number |
| **Health Probe** | Internal: hp-lbi-[env]-ause-[appname]-01  Or  External/Public: hp-lbe-[env]-ause-[appname]-01 |
| **Session Persistence** | None (default, can be set to Client IP/Client IP and Protocol) |
| **Enable TCP Reset** | Enable/Not enabled |
| **Enable Floating IP** | Enable/Not enabled |

## Secondary Region Load Balancer

|  |  |
| --- | --- |
| Configuration Item | Configuration Value |
| **Name** | Internal: lbi-dr-auea-[appname]-01  Or  Public/External: lbp-lbe-dr-auea-[appname]-01 |
| **Subscription** | AV ALZ [Subscription Name] |
| **Resource Group** | rg-dr-auea-[appname]-[workload]-01 |
| **Region** | Australia East |
| **SKU** | Standard |
| **Type** | Public/Internal |
| **Tier** | Regional/Global |
| ***Frontend IP Configuration Settings*** |  |
| **Name** | Internal: fip-lbi-dr-auea-[appname]-01  Or  External/Public: fip-lbe-dr-auea-[appname]-01 |
| **Virtual Network** | vnet-dr-auea-[appname]-01 |
| **Subnet** | snet-dr-auea-[appname]-[workload]-01 |
| **Assignment** | Static |
| ***Backend Pools Configuration Settings*** |  |
| **Name** | Internal: bep-lbi-dr-auea-[appname]-01  Or  External/ Public: bep-lbe-dr-auea-[appname]-01 |
| **Backend Pool Configuration** | NIC/IP Address |
| ***Health Probe Configuration Settings*** |  |
| **Name** | hp-lbi-dr-auea-[appname]-01  Or  hp-lbe-dr-auea-[appname]-01 |
| **Protocol** | TCP/HTTP/HTTPS |
| **Port** | Port Number (e.g. 80,443) |
| **Interval (seconds)** | 5 |

## Secondary Region Load Balancer Inbound Rule

|  |  |
| --- | --- |
| Configuration Item | Configuration Setting |
| **Inbound Load Balancing Rule Name** | rule-[protocol]-[portnumber]-[appname]-01 |
| **Frontend IP Address** | Internal: fip-lbi-dr-auea-[appname]-01  Or  External/Public: fip-lbe-dr-auea-[appname]-01 |
| **Backend Pool** | Internal: bep-lbi-dr-auea-[appname]-01  Or  External/Public: bep-lbe-dr-auea-[appname]-01 |
| **Protocol** | TCP/UDP |
| **Port** | Port Number |
| **Backend Port** | Port Number |
| **Health Probe** | Internal: hp-lbi-dr-auea-[appname]-01  Or  External/Public: hp-lbe-dr-auea-[appname]-01 |
| **Session Persistence** | None (default, can be set to Client IP/Client IP and Protocol) |
| **Enable TCP Reset** | Enable/Not enabled |
| **Enable Floating IP** | Enable/Not enabled |

# Acceptance

Signature of this page by appropriately delegated representatives of ​Ambulance Victoria​ signifies acceptance of this design document.

Logicalis will commence build and implementation work once it receives a signed copy of this design document.

Signature of this page by appropriately delegated representatives of Ambulance Victoria signifies acceptance of this design document.

Logicalis will commence build and implementation work once it receives a signed copy of this design document.

|  |  |
| --- | --- |
| Project | Core Services |
| Document Version | 1.0 |

**Signed on behalf of Ambulance Victoria**

|  |  |
| --- | --- |
| Name | Dan Howarth |
| Position |  |
| Signature |  |
| Date signed |  |

**Signed on behalf of Logicalis Australia**

|  |  |
| --- | --- |
| Name | Daniela Nikolic |
| Position | Senior Cloud Engineer |
| Signature |  |
| Date signed |  |

1. https://learn.microsoft.com/en-us/azure/well-architected/ [↑](#footnote-ref-2)
2. https://learn.microsoft.com/en-us/azure/load-balancer/manage [↑](#footnote-ref-3)
3. https://azure.microsoft.com/en-us/pricing/details/load-balancer/ [↑](#footnote-ref-4)
4. https://learn.microsoft.com/en-us/azure/well-architected/resiliency/overview [↑](#footnote-ref-5)
5. https://learn.microsoft.com/en-us/azure/well-architected/cost/overview [↑](#footnote-ref-6)
6. https://learn.microsoft.com/en-us/azure/well-architected/scalability/overview [↑](#footnote-ref-7)
7. https://learn.microsoft.com/en-us/azure/well-architected/security/security-principles [↑](#footnote-ref-8)
8. https://learn.microsoft.com/en-us/azure/load-balancer/load-balancer-overview [↑](#footnote-ref-9)
9. https://learn.microsoft.com/en-us/azure/well-architected/service-guides/azure-load-balancer/reliability [↑](#footnote-ref-10)
10. https://learn.microsoft.com/en-us/azure/load-balancer/load-balancer-floating-ip [↑](#footnote-ref-11)