**Purposed Architecture**

**Assumptions**

1. Two backend pools are designed, one for App services and other for vmss.
2. Multiple applications are running in one vmsspool at different ports and urls.
3. Routing will happen based on those paths and port numbers.
4. Custom probes are set for each application running in vmsspool.
5. Health probe is based on health page generated for app, api or app service individually and response code 200 will be considered successful.
6. If any of the application running on vm is unhealthy, the whole vm will be considered as unhealthy and in turn will be replaced.
7. Logs and metrics will be collected on log analytics workspace and Application insights as per Microsoft best practices and security recommendations.

**Architecture Diagram**

**Diagram

Description automatically generated**

**Approach**

Steps to be followed, in order to create infrastructure: -

* Create new repository in Azure DevOps organization
* Multi-stage Pipeline setup
  + Provision AppGW or WAF
  + Provision App service and VMSS
  + Provision custom health Probe
  + Check connectivity

**Reference**

[azure-docs/application-gateway-limits.md at main · MicrosoftDocs/azure-docs (github.com)](https://github.com/MicrosoftDocs/azure-docs/blob/main/includes/application-gateway-limits.md)

[How an application gateway works | Microsoft Docs](https://docs.microsoft.com/en-us/azure/application-gateway/how-application-gateway-works)

**Limitations**

**Application Gateway limits**

The following table applies to v1, v2, Standard, and WAF SKUs unless otherwise stated.

| **Resource** | **Limit** | **Note** |
| --- | --- | --- |
| Azure Application Gateway | 1,000 per subscription |  |
| Front-end IP configurations | 2 | 1 public and 1 private |
| Front-end ports | 1001 |  |
| **Back-end address pools** | **1001** |  |
| Back-end servers per pool | 1,200 |  |
| HTTP listeners | 2001 | Limited to 100 active listeners that are routing traffic. Active listeners = total number of listeners - listeners not active. If a default configuration inside a routing rule is set to route traffic (for example, it has a listener, a backend pool, and HTTP settings) then that also counts as a listener. See [Frequently asked questions about Application Gateway](https://docs.microsoft.com/en-us/azure/application-gateway/application-gateway-faq#what-is-considered-an-active-listener-versus-inactive-listener) for additional details. |
| HTTP load-balancing rules | 4001 |  |
| **Back-end HTTP settings** | **1001** |  |
| Instances per gateway | V1 SKU - 32 V2 SKU - 125 |  |
| SSL certificates | 1001 | 1 per HTTP listener |
| Maximum SSL certificate size | V1 SKU - 10 KB V2 SKU - 16 KB |  |
| Authentication certificates | 100 |  |
| Trusted root certificates | 100 |  |
| Request timeout minimum | 1 second |  |
| Request timeout maximum to private backend | 24 hours |  |
| Request timeout maximum to external backend | 4 minutes |  |
| Number of sites | 1001 | 1 per HTTP listener |
| URL maps per listener | 1 |  |
| Maximum path-based rules per URL map | 100 |  |
| Redirect configurations | 1001 |  |
| Number of rewrite rule sets | 400 |  |
| Number of Header or URL configuration per rewrite rule set | 40 |  |
| Number of conditions per rewrite rule set | 40 |  |
| Concurrent WebSocket connections | Medium gateways 20k2 Large gateways 50k2 |  |
| Maximum URL length | 32 KB |  |
| Maximum header size | 32 KB |  |
| Maximum header field size for HTTP/2 | 8 KB |  |
| Maximum header size for HTTP/2 | 16 KB |  |
| Maximum file upload size (Standard SKU) | V2 - 4 GB V1 - 2 GB |  |
| Maximum file upload size (WAF SKU) | V1 Medium - 100 MB V1 Large - 500 MB V2 - 750 MB V2 (with CRS 3.2 or newer) - 4 GB3 |  |
| WAF body size limit (without files) | V1 or V2 (with CRS 3.1 and older) - 128 KB V2 (with CRS 3.2 or newer) - 2 MB3 |  |
| Maximum Private Link Configurations | 2 | 1 for public IP, 1 for private IP |
| Maximum Private Link IP Configurations | 8 |  |
| Maximum WAF custom rules | 100 |  |
| Maximum WAF exclusions per Application Gateway | 40 |  |

1 In case of WAF-enabled SKUs, you must limit the number of resources to 40.

2 Limit is per Application Gateway instance not per Application Gateway resource.

3 Must define the value via WAF Policy for Application Gateway