

-1 *PLEASE DO NOT REORDER USE CASES* ok.

LBaaS Use Cases

The appendix contains description of the different players in the project / cloud environment.

The use cases below focus on the use cases that a project-user might care about.

Project-user use cases

1. A project-user wants to make his web based application highly available. S/he has n VMs deployed on the private subnet/network. Each VM is with a web server. She has n production load balancing services s/he wants to deploy, and needs a way to guarantee these don't end up on the same physical devices, and doesn't want to have to get the operator involved in this deployment decision.
2. A project-user is paying for a private cloud in which n load balancing appliances have been deployed. She has n-1 production load balancing services she wants to deploy on separate physical appliances, and several QA, staging and test services she wants to deploy, which should all share the same physical host (and shouldn't share hosts with any of the production services). As she may decide to alter the arrangement of the distribution of these load balancing services at any time, she doesn't want the operator to have to be involved in this deployment logic (and neither does the operator).
3. A project-user has an HTTPS application in which some of the back-end servers serving this application are in the same subnet, and others are across the internet, accessible via VPN. He wants this HTTPS application to be available to web clients via a single IP address.
4. A project-user has an HTTP-based application available through a load balanced service on a single IPv4 address. S/he wants to make the same application available via an IPv6 address, too.
5. A project-user has an HTTP-based application available through a load balanced service. He would like to designate some of his back-end application servers as "backup servers" to be accessed only when the primary application servers are unavailable.
6. A project-user has an HTTP-based application available through a load balanced service. He would like a customized "error 503" page displayed whenever all the back-end servers are unavailable (including any backup servers).
7. A project-user has a web-based application in which they'd like the load balancer to speak HTTPS with the web clients, but HTTP with the back-end servers. Back-end servers should be made aware via the X-Proto header whether connections between the web clients and the load balancer service were done via HTTPS.

8. A project-user has an HTTP-based application behind a load balancing service, but needs to make sure no more than n clients connect with each back-end server at a time to ensure they don't get overloaded.
9. A project-user has HTTP-based application behind a load balancing service. S/he needs to make sure that only back-end service that pass a HTTP-based health check are accessible to web clients. However, this health-check is only accessible via a URL that requires HTTP Basic authentication.
10. A project-user has a custom TCP-based service split across many application servers on the same subnet, and would like this service accessible via a single IP address. Further, he has written a custom load check for the back-end servers which powers custom auto-scaling logic. He would like to be able to adjust the weighting of any given back-end server in the load balancing pool on the fly, or add and remove some servers entirely. He needs to be able to do this without interrupting any TCP sessions still in progress.
11. A project-user has an HTTP-based load balanced service. His application has a wide range of completion times for each request, and would like the load balancers to always pick the back-end hosts with the fewest requests in progress for the next client request.
12. A project-user has a very AJAX-intensive web application behind a load balancing service. He would like to speed up the performance for the web client by making sure the load balancer leaves the TCP connection between the web client and the load balancer open for a while after each request (so handshaking doesn't need to happen again). This application may be accessed via HTTP or HTTPS by the web client.
13. A project-user has an HTTP application behind a load balanced service. Some of the requests being processed take a very long time to complete, so she needs to make sure the load balancer doesn't close the connection with the client or server before processing is complete, and the content is returned from the back-end server servicing the request.
14. A project-user has an HTTP application behind a load balanced service. In order to do A/B testing, the "A" back-end servers will be running entirely different code than the "B" servers, yet any web client connecting to the service needs to always be routed either to "A" or "B" hosts depending on which group they first connected to. The project-user would like both the "A" and "B" versions of the site accessible via the same IP address.
15. A project-user has an HTTP application behind a load balanced service. In order to do maintenance on each back-end server, he would like to be able to sequentially remove servers from the load balancing pool, perform the maintenance, and add them back into the pool. He does not want to have to put the site into "maintenance mode" to do this, nor does he want any in progress requests getting interrupted because of this maintenance.
16. A project-user has an HTTP application behind a load balanced service. In order to perform some major maintenance, he would like to be able to put the site into "maintenance mode" for the duration of the maintenance, and then take the site out of maintenance mode when the maintenance is over. In order to test the site before it goes

live again, he would like to have a way to enable the site for specific client IPs only while in maintenance mode.

17. A project-user has an HTTP site behind a load balanced service that powers many different e-commerce sites. She has different groups of back-end servers that should be used for different sites, depending on the HTTP/1.1 hostname that the web client requests when connecting.
18. A project-user has an HTTP application behind a load balanced service. This application has been attacked by malicious 3rd parties in the past coming from a specific block of IP addresses. She would like the load balancers to drop any requests from this block of IP addresses before they reach her back-end servers.
19. A project user wants to troubleshoot his application and needs to download a connection log from the load balancer.
20. A project user wants to host a HTTP website which includes a Java applet. The applet requires an additional TCP port to communicate with its backend application. The Java security policy for applets restricts them to only connecting to the host they were retrieved from.
21. A project-user has a 2048-bit wildcard SSL certificate for "*.example.com" and a 4096-bit SSL certificate for "secure.example.com." She wants to serve all her websites that apply to the "example.com" domain from the same IP (using the SNI standard), and also wants to make sure that any requests for "secure.example.com" use the 4096-bit certificate.
22. A project-user has an old 2048-bit wildcard SSL certificate for "*.example.com" and a new 4096-bit wildcard SSL certificate for "*.example.com" from a new certificate authority. She wants all of the various hostnames that apply to the "example.com" domain served from the same IP (using the SNI standard), and usually using the old certificate-- but she wants requests for "secure.example.com" and "admin.example.com" to use the new certificate.
23. A project-user has a web applicaA project-user has an HTTPS site behind a load balanced service that powers many different e-commerce sites. Each of these e-commerce sites has its own SSL certificate and she would like the load balanced service to use the appropriate one when clients connect, according to the SNI protocol standard.
24. tion that contain unsecured and secured parts (ex: unsecured - <http://www.acme.com/app/catalog> and secured <https://www.acme.com/checkout>). When an application user access the URI that is handled by the secured part via unsecured protocol (ex: <http://www.acme.com/checkout>), she will be redirected to the secured section (ex: <https://www.acme.com/checkout>) and when an application user access the URI that is handled by the unsecured part via a secured protocol (ex: <https://www.acme.com/app/catalog>) she will be redirected to the unsecured part (ex: <http://www.acme.com/app/catalog>)
25. A project-user consistently wishes to classify static content so that she can use the same classifications for all of it's web applications. For this she wishes to classify all picture types as "pictures", classify all streaming types as "videos" and classify all word

office documentation type as “word-documents”. She wants to reuse this definition to direct traffic to the appropriate static content servers. For example browsing to <http://www.acme.com/uri1/piv.gif> will be handled by the static content server hosting images appropriate for www.acme.com and browsing to <http://www.mysite.com/myname/portrait.jpg> will be handled by the static content server hosting images appropriate for www.mysite.com

26. A project user wants to terminate SSL connections so that the load balancing logic will be able to make decision based on layer 7 information. He then wishes the traffic to continue to be encrypted between the load balancer and the application servers. He does not care to provide trusted certificates
 27. A project user wants to specify the allowed cipher suites and the allowed SSL protocols when terminating a connection on the load balancer. He only want to allow TLS1.2 and Elliptic curve ciphers to be used.
 28. A project user wants to manage the allowed protocols and cipher suites in a central way so that if it needs modification dues to new security concerns, it will affect all applications.
 29. A project user has an HTTPS service on the load balancer that has has a back-end pool of servers which should also be accessed over HTTPS. The user is not very sophisticated (ie. doesn't want to have to manage a whole PKI). Each of these back-end servers is configured with its own separate self-signed server certificate, and the user would like the load balancer to authenticate that the self-signed cert is the one configured for that server in order to protect against potential man-in-the-middle attacks.
 30. A project user has several load balanced services and would like to collect rolling data on bytes transferred in and out, number of connections and other aggregate statistics per service. Ideally, this would be data that could be collected via sensu or some other means suitable for generating a dashboard graph.
 31. A project user has a load balanced service and would like to see bytes transferred, connections, and other basic statistics on each pool of servers in the back-end.
 32. L7 content switching
 33. Session persistence
 34. Health Monitoring
 35. IPv4 and IPv6 support
 36. SSL Termination
 37. SSL back end encryption
 38. SNI support
 39. UDP support
 40. Round Robin algorithm support
 41. Backup servers
 42. Spillover
 43. Max Pool/Member Connections
 44. Domain based pool members
- Anycast route injection support (to upstream L3 Devices)
45. Direct server return (DSR) vip support

30-21, 42-45 and 48-56 are taken from

<https://docs.google.com/spreadsheet/ccc?key=0Ar1FuMFYRhgadDVXZ25NM2NfbGtLTkR0TD FNUWJQUWc#gid=0>

Cloud Admin use cases

1. A cloud admin wants to enable IaaS to be used by tenants
2. A cloud admin wants to enable different “flavors/providers” for the LBaaS service
3. A cloud admin wants to remove all load balancers from tenants who don’t need/use the service any longer

Cloud Operator use cases

1. A cloud operator wants to troubleshoot a connectivity issue in which a VIP is not accessible
2. A cloud operator needs to be able to take a physical load balancing appliance out of production at any time without affecting the availability or performance of user load balancing services
3. A cloud operator needs to be able to scale the load balancer to satisfy all its customers needs
4. A cloud operator needs a sophisticated SSL Certificate management which is compliant with the security requirements of an operator (SOC compliant) scalable, robust, and easy to use for the tenants
5. A cloud operator needs sophisticated Metrics Collection to support multiple complex billing scenarios
6. A cloud operator needs a separate admin API for NOC and support operations
7. A cloud operator requires minimal downtime when migrating to newer versions of the load balancer or Neutron
8. A cloud operator might have deployed a mix of different load balancers (hardware, software, HA scheme, etc) to satisfy customers diverse needs. The ability to migrate easily from one load balancing “flavor” to another without noticeable downtime is essential.
9. Resiliency functions like HA and failover in a cloud operator context (take advantage of multiple datacenter, availability zones/regions, etc.)
10. A cloud operator needs load balancer health checks and healthchecks on the rest of the load balancer infrastructure to ensure health of the service and pinpoint problems quickly
11. A cloud operator needs to support multiple, simultaneous drivers for hw/sw/etc. load balancers

Appendix - Players

Project:

- Project User - a person that provisions and uses cloud resources. A user may create VMs and use them. Users may **not** be allowed to create networks, subnets and manage network services.
- Project Admin - a person that in addition of being a user, may create networks, subnets and manage network services.
- Project Operator - a person that monitors project resources and has the capabilities to detect operation failures and gather forensics to pin-point issues.
- Project Operation Automation - code that runs on behalf of the Project Admin/Project Operator, detects issues and failures of the project resources, and automatically reports or remediates them. Such code can run in a “system” provided by the cloud service or on an external system.

Cloud side:

- Cloud Admin - a person that is in charge of installing, and upgrading the cloud Hardware and Software. The Cloud Admin is also managing remediation of critical failures and issues.
- Cloud Operator - a person that monitors the cloud activities, can detect failures collect forensics to troubleshoot issues and can remediate minor “standard” issues.
- Cloud Operation Automation - code that runs on behalf of the Cloud Admin/Cloud Operator, detects issues in the cloud hardware infrastructure, cloud software and cloud services and automatically reports or remediates them.