

# Load Balancers in the Cloud-Research Strategy applied in SAP Cloud

Giridhar Kankanala<sup>1</sup>, Sudheer Amgothu<sup>2</sup>

<sup>1</sup>Independent Researcher, Department of Computer Science, Illinois, USA

<sup>2</sup>Independent Researcher, Department of Computer Science, Boston, USA

**Abstract:** SAP load balancers are used to distribute network traffic across multiple servers in an SAP landscape environment. Load Balancer ensure no single server is over-loaded with requests, thereby harmonizing performance and reliability of the SAP system. Load balancers can be hardware appliances or software-based solutions that are specifically designed to work with SAP applications. They help to optimize resource utilization, improve scalability, and enhance the overall user experience. Moreover, load balancers can provide high availability and fault tolerance by automatically routing traffic away from failed or underperforming servers. This paper elaborates the fundamental concepts of SAP systems are one of large enterprise systems with complex architectures, there are workload load balancers in SAP at every level to distribute traffic balancing workloads at SAP Application Level, Network level, to optimize traffic and reduce latency across intranet/internet, Cloud, In-House/On-Premise, local and Wide Area Networks, Cloud Providers also provide Load Balancers to make sure systems are HIGHLY Available for businesses avoiding outages and make sure systems are available most of the time close to 100% or almost all Cloud Providers are promising above 95% availability.

**Keywords:** SAP Load Balancer, load balancer on Cloud Providers, load balancers on internal/external traffic

## 1. Methodology & Materials

### a) What is SAP?

SAP stands for “Systems, Applications & Products in data processing” – this is German Software Company that provides enterprise application software (ERP) help businesses manager their operations, Supply Chain, Manufacturing, several horizontal alliances of a company including Finance, Human Resources, and Customer Relationships using SAP ERP software.85% of fortune 500 companies rely on SAP for their core business processes, Customers using SAP have reported 20% increase in operational efficiency

### b) What is Hyperscaler?

Hyperscalers is a type of large-scale data center that offers massive computing resources, typically in the form of an elastic cloud platform, Organizations use them to deploy and manage large scale applications and services. In SAP it is basically offered as Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) over these hyperscale cloud platforms for SAP applications considering their scalability, Network connectivity for their mass/critical distributed environments across multiple datacenters across global/regions

### c) What are Computing resources in SAP

SAP can be implemented in customers own hardware or Cloud environments – considering proven best practices hardware recommendations, Fast time to value, Continuous Innovation, Open & extensible, Security, Compliance and scalability, Automatic and continual updates – these are possible based on cost/money on the table to choose between below snip to show SAP trend on top Cloud Vendors market share between Q1 2021 to Q2 2024

### d) SAP Cloud service provider

A CSP (cloud service provider) is a third-party company that provides scalable computing resources that businesses can access on demand over a network, including cloud-based compute, storage, platform, and application services, Cloud Computing is becoming go-to model for accelerating digital transformation and offers flexibility, scalability, resiliency with Security protocols in place that business demands without being concerned with physical limitations, easy to manage & maintain plannings instead of building a own infrastructure for your business applications, you can rent any CSP which can be shared with other individuals or other organizations.

### e) Load Balancer

Main purpose of load balancing is reliability, this is to prevent server overload and maintaining high availability, performance free, balancing/distributing workload incoming requests across available resources, this creates highly available, scalability and reliable web applications, databases, Network, Server Managements/Operating Systems.

## 2. SAP Load Balancer Common Guidelines

SAP architecture carries message server which must be configured for Load Balancing, there are several blogs, helpful documents, URLs available but shifting of architectural solutions, more tools around this area more enhanced processes, this is at bottom of all the architecture layer at Application level workload is distributed optimized to improve and achieve maximum performance by SAP work processes, these settings accept incoming inbound processing of master data, forecasting jobs, it will break down and distribute operations into many smaller tasks decompresses into small tasks owning by each dialog work process until available maximum number of work processes whatever percentage it was assigned/allocated to use by each application sever settings.

- Allow system to perform/process tasks in parallel to workloads
- Distribute workload equally within the system
- It can limit the scope to local to application server & particular client
- Configurations on how many numbers of work processes based on available CPUs & Memory

It is recommended to not allow more than 75% of DIALOG work processes to regulate the number of work processes available on application, if customer's application is on Cloud Provider sometime these CPUs/RAM are assigned but distributed across the Node underlying architecture takes care of based on assigned policy configurations.

### 3. SAP ABAP & Java stack Load Balancer Concept

SAP have both SAP NetWeaver ABAP application & SAP NetWeaver Java consists of multiple instances based on the anticipated workloads systems are designed/architected such a way to have multiple application instances with equal number of resources, sometimes organizations consider shutting non-used applications, or turn-them ON during workloads or dedicated to certain critical activities for load balancing distributions like for HTTP based traffic, RFC based traffic, Gateway based traffic, Non-Dialog workloads for Jobs, during month-end workload traffic, often consist of multiple instances based on scalability workload is balanced between these instances.

Corresponding message server stacks the requests to individual ABAP or Java instance stacks, there are solutions to have both ABAP & Java called as dual-stack instance likely for Solution Manager or optimized solutions they can be installed on same system.

- In ABAP Stack – message server assigns work to Dialog Processes of the instances
- In Java Stack – message server assigns work to Java server processes

### 4. SMLG & Logon Load Balancing steps

Logon Groups are place where you can allocate available application servers for each corresponding functional/technical workloads distribution (FI/SD, MM, based on necessary), this will update message servers to balance the load among the servers in the logon group, if all the available work processes are busy it'll not allow stops new traffic/ no new logons are encouraged, this **prevents systems crash situations**

RFCs – when you have multiple systems in the environment – it is recommended to configure RFCs communication with enabling Load Balancer which is by RFC configurations. Message server delegates the incoming requests

There are several best practice recommendations from SAP what values to limit, what to be for parallel processing's explicit recommendations based on initial configurations and based on performance optimization tuning recommendations, not same values applicable in Non-Production & Production

environments based on transactional data and workload traffic.

### Monitoring & Tracing traffic

Based on load distribution between work processes (RFC/HTTP/other traffic) sometimes hard to influence and obtain the culprit during mass workload analysis, SAP provided monitoring or 3<sup>rd</sup> party monitoring tools are available to alert when these assigned monitoring thresholds reach 60% or 75% to alert – this is again to avoid/prevent outages technical team decision making either to kill the processes or restart the processes.

Trace logs or SM21 logs, Gateway logs, performance trace logs several means are available to back-track the cause.

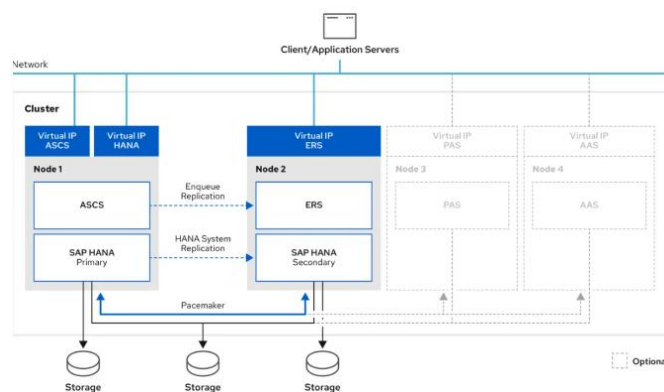
### 5. SAP Web Dispatcher

This is one of greatest tool provided by SAP a self-configuring tool for HTTP/HTTPS load balancing in SAP NetWeaver application Server landscape, SAP Web Dispatcher can be hosted separately and setup to distribute workload for each system or setup multiple systems/instances for DEVELOPMENT or QUALITY or PRODUCTION landscapes based on individual customers systems sizes.

Backend this connects to SAP Application Message Server (msg\_server) which communicates to its internal SAP systems stacks the requests to their own processors.

### Web Dispatcher Host & Port setup:

communicates to Message Server – corresponding recommended Ports should be open between these hosts between Web Dispatchers & SAP Hosts, these are updated during initial configurations or modified based on what ports are assigned between, its updated through configuration parameters in ABAP/Java profiles and Web Dispatchers profiles. (icm/server\_port\_<xx>, rdisp/mshost, ms/http\_port)



### 6. Server Management Load Balancer

This is very critical layer in SAP world, based on customers Primary Landscape & Secondary Landscape for High Availability & Disaster Recovery situations, both side servers are hosted in Region 1 (Primary) & Region 2 (Secondary), Customer servers are hooked as Primary with Region 1 and continuous data sync happens between Primary Region to Secondary Regions servers, this is additional setup required based on Customers budgets, other insights if this feature

should be enabled for only Production environments or Non-Productions too.

Operating Systems should have same versions, same kernels on both sides Primary & Secondary regions in case failover happens SAP applications talks to same servers as they will have same FQDNs.

High Availability/Failover solutions at Database Layer or Application corresponding settings should be configured at OS level as 'Cold Standby'

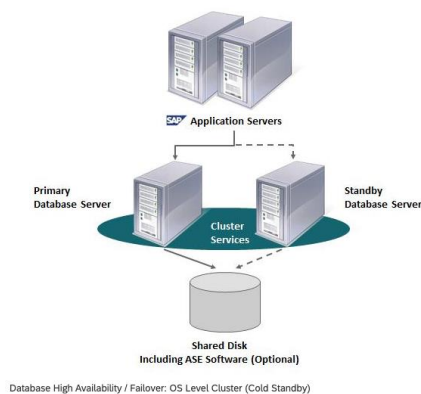
Certain challenges like

- OS/Application/DB Maintenances requires downtime for both primary & Secondary
- Risk of I/O corruption
- Real-time outage might cause transactional data loss for that duration
- Based on RPO/RTO timelines-SAP DB & SAP Applications are not available during failover & database recovery

Load balancer high availability software's to be installed and configurations to be done to accomplish these Load Balancing and High Availability, HANA System Replications solutions between these Primary/Secondary Nodes.

Software's like pacemaker can be used for these large-scale environments. SAP can help with automation mechanism based on their certified hardware vendors Installations/Collaborations.

Not all OS vendors clusters support's these solutions these have to certified Hardware platforms to setup, to name few like RedHat Enterprise Linux, SuSe Enterprise Linux, IBM PowerHA on AIX, Microsoft Cluster Server (MSCS) there are few others who are support for these Failover Load Balancer solutions.



## 7. Network Load Balancer (Cloud Provider or Internal)

Typically, inbound/external load balancer are setup with encrypt/decrypt protocols along with Web Application Firewalls, how customers want to access these SAP systems when they're on SAP Cloud over the internet by enabling inbound/outbound load balancer by using SSL/TSL protocols, corresponding entries to be maintained in both side DNS/firewall respective signed certificates to be registered/installed in public DNS. Each cloud provider has their own recommendations however keeping security

constraints on top priority. Customer Single-Sign-On requires these Client Certification authentication to be in place to pass HTTP request authenticated. Client certification authentication process must be followed between Customers Network and Cloud Provider's network inbound/outbound load balancer configurations.

FQDN usage is recommended instead of IPs or Hostnames, Cloud providers provide physical hostnames and Virtual Hostnames and have their own local IPs, Virtual IPs which is not obtained in public/external sites. Corresponding FQDNs to be updated in DNS and other 3<sup>rd</sup> party vendor or integration hostnames/FQDNs to be updated for smoother operations across customer – SAP – Cloud Providers Gateways, DNS resolutions have to pass the pings/nslookup tests.

Network level multiple tools like F5 Load Balancer, Citrix NetScaler, NGINX Plus, ADCs are few proven robust and most compatible with most of cloud providers between SAP to Cloud Providers and Customer networks Application Delivery Controllers. They will use Proxy Network setup, HTTP/HTTPS, TCP or UDP protocols based traffics.

## 8. Conclusion

Having architecture involved with Load Balancer is bare minimum to attain good performances, even workload distributions, underlying Security which will minimize Server level or application-level vulnerabilities. Load Balancers are one of cost-effective solutions when compared to other infrastructure costs. These are one-time setup not frequent changes however updates are often performed for improved security. Overall, this help increasing improved user experiences based on today's fast pacing AI world. Server-side load balancers have evolved a lot these last few years however embedding with Server & Network layers in almost every Cloud Provider. Load Balancers for SAP environments at Application Level, Database Level, Server Management Level, Network Level as much as relevant brings enhanced operations and systems availability to businesses.

## References

- [1] **SAP Load Balancing** - <https://help.sap.com/docs/CARAB/e95c8443f589486bbfec99331049704a/d34da3514b835172e10000000a423f68.html>
- [2] **SAP ABAP Stack/Java Stack** - [https://help.sap.com/doc/saphelp\\_nw73ehp1/7.31.19/en-US/67/f2a88355de4690a429a25ef83c89b2/content.htm?no\\_cache=true#:~:Concept,each%20server%20\(see%20below\)](https://help.sap.com/doc/saphelp_nw73ehp1/7.31.19/en-US/67/f2a88355de4690a429a25ef83c89b2/content.htm?no_cache=true#:~:Concept,each%20server%20(see%20below))
- [3] **SAP Migration Strategies** - <https://www.ijsr.net/archive/v12i12/SR23128151813.pdf>
- [4] **SAP Web Dispatcher** - [https://help.sap.com/saphelp\\_nwce711/helpdata/en/48/8fe37933114e6fe10000000a421937/frameset.htm](https://help.sap.com/saphelp_nwce711/helpdata/en/48/8fe37933114e6fe10000000a421937/frameset.htm)
- [5] **Choosing Right Computing Resources** - <https://espieta.org/Volume4-Issue2/JETA-V4I2P122.pdf>
- [6] **Server Management** - <https://documentation.suse.com/sles-sap/sap-ha-support/html/sap-ha-support/index.html>
- [7] **O'Reilly Load Balancing Book** - <https://www.oreilly.com/library/view/load-balancing-in/9781492038009/ch01.html>