Ask a Question Write a Blog Post Login

Personal Insights

Follow

👍 Like

RSS Feed

Savan Kharod

October 5, 2021 | 6 minute read

7 Best Practices for Creating a High Availability Infrastructure

0 0 2,176

Cloud computing has revolutionized the way an IT company works in the present age. Without cloud power, many small and large organizations would have massive trouble conducting their functions and serving customers.

The cloud makes all hardware and software resources flexible, accessible, and efficient through its interface and features. But to avail this service productively, it needs to be working at all times.
With increasing demands for cloud computing services, companies rely on them fully to be stable whenever they are needed. Thus, any downtime for applications or services creates issues for the business and its customers. This is

Assigned Tags

Cloud Operations

Cloud Integration

SAP HANA

high availability

SAP HANA High Availability

Similar Blog Posts

Cloud Operations Best Practices for SAP SuccessFactors Run Operations

By Robin Haettich Oct 15, 2021

Cloud Operations Best Practices for SAP Ariba Run Operations

By Robin Haettich Jul 08, 2022

Preventative controls - using

where high availability infrastructure comes in.

What Is High Availability?

High availability is a system or component that remains operational for long periods of time without any interruption. It is generally measured as a percentage, with a 100% system meaning that the service has experienced zero downtime.

However, it is difficult to maintain a complex system that never fails. A 99% and above the rate of uptime is considered excellent performance. Companies like Google, Amazon, and Microsoft keep their service availability at 99.9%.

Cloud computing available is measured in several factors, such as recovery time, unanticipated loads, increased usage, scheduled maintenance, and unscheduled maintenance periods. Hence, it is a percentage defined by service level agreements (SLAs).

High availability and uptime, though they sound similar, are not the same concepts. Availability denotes how accessible the server is or the number of connections it allows, whereas uptime refers to whether the server is reachable by even one service. Hence, uptime is an infrastructure level capacity, and HA is application level.

How to Achieve High Availability

There are various steps you can take to achieve high availability in your system, like having more components for a safety buffer, performing regular checks, and replacing failed servers. Here are seven practices that can help you create a high availability infrastructure.

organizational policies to provide guardrails for SAP's public cloud accounts

By Jay Thoden van Velzen Apr 05, 2021

Related Questions

How would I use leonardo-iot ODATA services in Business Application Studio?

By Rewati Raman Sep 22, 2021

SAP BTP Certification for Operations / Architect

By Matthias Kiekens May 28, 2021

How to create SAP CAP Service from Parameterized Calculation view?

By Rewati Raman Jun 10, 2022

1. Use Autoscaling

Autoscaling is the automatic adjustment of cloud computing services, in which the cloud computing dynamically adjusts the number of computing resources in a server, depending on the load and active servers. SAP's Application Autoscaler lets you do this by automatically increasing or decreasing the number of application instances based on the policies defined.

Through autoscaling, you can ensure that no additional resources are utilized when not in need and also that there is no dearth of active instances when load increases. You can scale up or out, which is either increasing the size of the machine to increase RAM for processing or combining multiple independent computers into a system.

2. Choose Between Complexity and Simplicity

Choosing between complexity and simplicity is a major factor in creating a high availability infrastructure. Think about the difference between a staircase and an elevator. A staircase is simple and effective, it does not need "downtime" unless there is a massive calamity. On the other hand, an elevator has added features and benefits, but needs maintenance and can break down more easily.
Web services behave the same way.

Adding complexity to it may lead to major benefits and additional features but it may also decrease availability percentage. On the other hand, a service with simple functions may not need downtime as much.

When you are building for high availability, you must think of this factor and trade off some features. Microservices solve this problem somewhat, by adding different functions as individual systems that are integrated into one unit. In case of downtime, all instances will not be affected together.

3. Implement Multiple Application Servers

Multiple application servers are very helpful for large data centers. If you overburden a server, it is more likely to crash and cause an outage. Instead, you should deploy applications on multiple servers so that all of them continue running with minimal downtime.

This eliminates the possibility of having a single point of failure. In case of an outage, you can direct traffic to another server and achieve high availability.

4. Enable Monitoring

You can use a monitoring tool that integrates into your application and provides reports on its performance, real-time functions, and error rates. These tools also give warnings if errors exceed a predefined threshold and helps you to detect ongoing or upcoming problems easily.

Once you know that there is an unusual failure rate, you can set self-healing tasks in motion or deploy resources to ensure the downtime is restricted to a minimum.

5. Implement Load Balancers

A load balancer acts as a reverse proxy and distributes traffic among a number of servers to reduce excessive load on a single one. This increases the capacity of the application and reduces downtime. High availability load balancing is crucial in preventing component failures and massive outages.

Primary and secondary balancers can be used to distribute workload across data centers and assure uninterrupted application delivery. You can use different algorithms for stateful and stateless applications like hashing or hash buckets. This is how most teams strive for high availability.

There are numerous SaaS companies out there that provide load balancer as a service. You can evaluate them to see if they fulfill your requirements.

6. Configure Failover Setup

If you keep more than one instance in the same location, there is a high risk of a single point of failure (SPoF). In case any component breaks in that chain, the whole server will face a risk. To avoid that, you can introduce network redundancy.

Host auto-failover is a local solution for fault recovery that can be considered an alternative or supplemental measure to system replication. In the SAP HANA system, one or more standby hosts are added and configured to work in standby mode. In case of failure of an active host, a standby takes its place and keeps the connection active.

HANA can be deployed on-premises or on the cloud from various cloud service providers. It also runs on multiple operating systems like Red Hat Enterprise Linux and SUSE Linux Enterprise Server. It is designed to support horizontal and vertical scaling.

7. Try Clustering Techniques

High availability server clusters are basically groups of servers that support services or applications with high uptime requirements and rendering the same set of functionalities. It prevents high downtime by detecting failures, performing failover, and restoring operations.

Clustering techniques can improve performance, scalability, and availability. High availability clusters make sure the critical server applications keep running all the time. In case one server collapses, the important services are immediately switched to another server.

Characteristics of a High Availability Infrastructure

A high availability infrastructure is important for cloud services to maintain regular functions and prevent mission-critical services from crashing. With an HA system, you can ensure your uptime remains optimum and your consumers do not face errors or issues. There are several characteristics of a high availability infrastructure.

No Single Points of Failure

With a high availability system, you will not find single points of failure. Failover setup is ready to detect any impending errors or occurred failures. In such a system, if one instance fails or one location is affected, there are over-failover options that take its place and keep the functions going. This helps in maintaining uptime and prevents the stoppage of critical services.

Hardware Redundancy

Hardware redundancy can be achieved by building two or more computing systems or physical copies of a hardware component. A system might consist of redundant servers, power supplies, memories, etc. If you apply redundancy to high-level components, you can reduce the risks of outage in case of higher loads.

Reliable Crossover

Reliable crossover must be set up to ensure that the failover systems take over when a server crashes or stops responding. This is another aspect of building redundancy into high availability systems. It lets a backup component take the place of a failed one. The act of successfully switching from one component to another without losing any data or reducing performance is known as reliable crossover.

Software and Application Redundancy

Software and application redundancy works in the same way as hardware redundancy, as in it performs the same functionality by executing different elements, in case another instance is affected. It is a crucial concept in high availability, fault tolerance, and reliability engineering. It also encompasses self-healing programs. Applying redundancy ensures you achieve reliability goals subject to techno-economic constraints.

Data Redundancy

A high availability system ensures data redundancy, which means the same data exists in multiple places. This reduces the risk of loss and ensures the data can be recovered if one memory location or server has failed. It also permits the correction of errors in transmitted or stored data.

Self-monitoring for Failure

HA systems contain self-healing and self-monitoring functionalities that can detect unusual failure rates or affected instances. It ensures that the error is detected and fixed in record time, without causing any massive harm to the functionalities. The more efficient the self-monitoring feature you have, the higher your uptime will go.

Conclusion

A high availability infrastructure is a combination of hardware, software, and applications that are designed to recover quickly in case of outages and maintain functionalities in a way that assures minimum downtime and more than 99% availability. Solutions such as SAP HANA and other failover systems can perform this task, thereby reducing failure risks and maintaining operations smoothly.

Be the first to leave a comment

You must be Logged on to comment or reply to a post.