ZIF

WHITE PAPER

BUSINESS SERVICE RELIABILITY WITH AIOPS

Business services are a set of business activities delivered to an outside party, such as a customer or a partner. Successful delivery of business services often depends on one or more IT services. For example, an IT business service that would support "order to cash", as an example could be "supply chain service." The supply chain service could be delivered by an application such as SAP, with the customer of that service being an employee in finance/accounting using the application to perform a customer facing service such as accounts receivable, or the collection of cash from an outside party. A business service is not simply the application that the end user sees – it is the entire chain that supports delivery of the service, including physical and virtualized servers, databases, middleware, storage and networks.

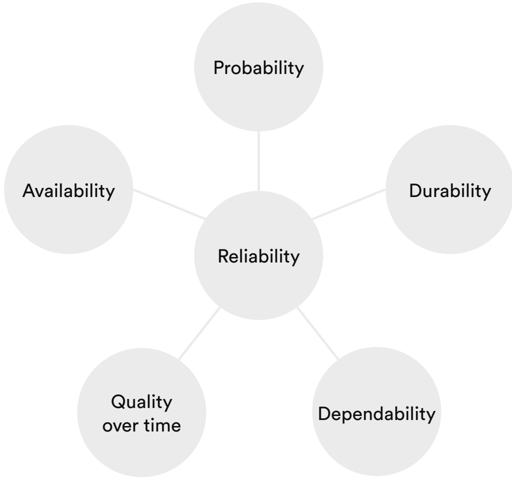
A failure in any of these can affect the service – and so it is crucial that IT organizations have an integrated, accurate and up-to-date view of all of these components and of how they work together to provide the service.

The technologies for Social Networking, Mobile Applications, Analytics, Cloud (SMAC) and Artificial Intelligence (AI) are redefining the business and the services that businesses provide. Their widespread usage is changing the business landscape, increasing reliability and availability to levels that were unimaginable even a few years ago.

Availability versus Reliability

At first glance, it might seem that if a service has a high availability then it should also have a high reliability. However, this is not necessarily the case.

Availability and Reliability have different meanings, serve different purposes and require different strategies to maintain desired standards of service levels. Reliability is the measure of how long a business service performs its intended function, whereas availability is the measure of the percentage of time a business service is operable. For example, a business service may be available 90% of the time, but reliable only 75% of the time from a performance standpoint.



Service reliability can be seen as:

* Probability of success
* Durability
* Dependability
* Quality over time
* Availability to perform a function

Merely having a service available isn't sufficient. When a business service is available, it should actually serve the intended purpose under varying and unexpected conditions. One way to measure this performance is to evaluate the reliability of the service that is available to consume. The performance of a business service is now rated not by its availability, but by how consistently reliable it is. Take the example of mobile services - 4 bars of signal strength on your smart phone does not guarantee that the quality of the call you received or going to make. Organizations need to measure how well the service fulfils the necessary business performance needs.

Recognizing the importance of reliability, Google initiated Site Reliability Engineering (SRE) practices with a mission to protect, provide for, and progress the software and systems behind all of Google's public services - Google Search, Ads, Gmail, Android, YouTube, and App Engine, to name just a few - with an ever-watchful eye on their availability, latency, performance, and capacity.

Drivers for Service Reliability

End-User Focus

Business services are becoming end-user focused. The modern-day sophisticated consumer of business services demands always-on services and instantaneous response times. Delivering exceptional user experience has become paramount. The user has become the driving force behind the continuous evolution of products and services. Therefore, organizations are adopting the highly productive, agile development practices of Continuous Integration & Continuous Delivery (CI/CD) as part of digital transformation. Digital transformation means using digital technologies to do something better or to create something that did not exist previously. Digital Transformation has been in place since digital computers came into existence. For example, when mechanical cash registers were replaced with computerized cash registers, that was a digital transformation start. But the technologies today catapult such transformation from evolution to revolution -a revolution to provide the most satisfying experience to the end user of the service.

Google stands as a shining example. Just think how many products and services we use from Google today – ostensibly a search engine company!

Applications that support business services are being Re-architected

Microservices and Serverless are examples of new modular application architectures. A monolith application is split into smaller services called microservices, each of which typically caters to one capability of the application. Microservices are stand-alone, and follow their own build/deploy cycles, enabling rapid development and scaling. They run inside containers or VMs that provide their execution environment. Container solutions like Docker, CoreOS Rkt and container orchestration solutions like Kubernetes provide rapid resolution to issues/performance lags in microservices & their containers. Serverless computing platforms like AWS's Lambda are event-driven and based on the premise that the application is split into functions that get executed based on events. Serverless provides complete abstraction of the OS, server, and infrastructure, so that application developers have no administrative overhead and can focus on adding value to their applications.

These modular architectures are gaining a lot of traction due to the many benefits: they enable agility in iterative delivery of new features & services; allow reuse of existing services that provide required functionality; help mold the business service in a way that best-fits usage patterns & so on. But they introduce multiple new IT monitoring challenges for the ITOps team, due to the exponential increase in the number of objects and their interplay, that need to be monitored for each application.

The breakdown of monolithic applications into hundreds or even thousands of smaller, cohesive, functional microservices has resulted in significantly reduced code footprint inside each service. However, these microservices now need to interact a lot with each other. Function calls within the code in monoliths have been replaced by calls going over the network in microservices.

The state of every request has to be transferred from one service to another to build a response. The result is an explosion of chatter such as API calls, RPCs, database calls, memory caching calls, etc.

In production, the critical piece that DevOps need to monitor is no longer the code inside a microservice, rather the interactions between various microservices. Most issues, such as hotspots, chokepoints or cascading failures that arise in production are due to the complex interplay between services. Continuous deployments are the norm and new dependencies between services may emerge after deploys. Whenever there is an issue, precious time is spent in chasing service dependencies, either by looking up (outdated) documentation or consulting other developers.

Increasing Complexity of Services

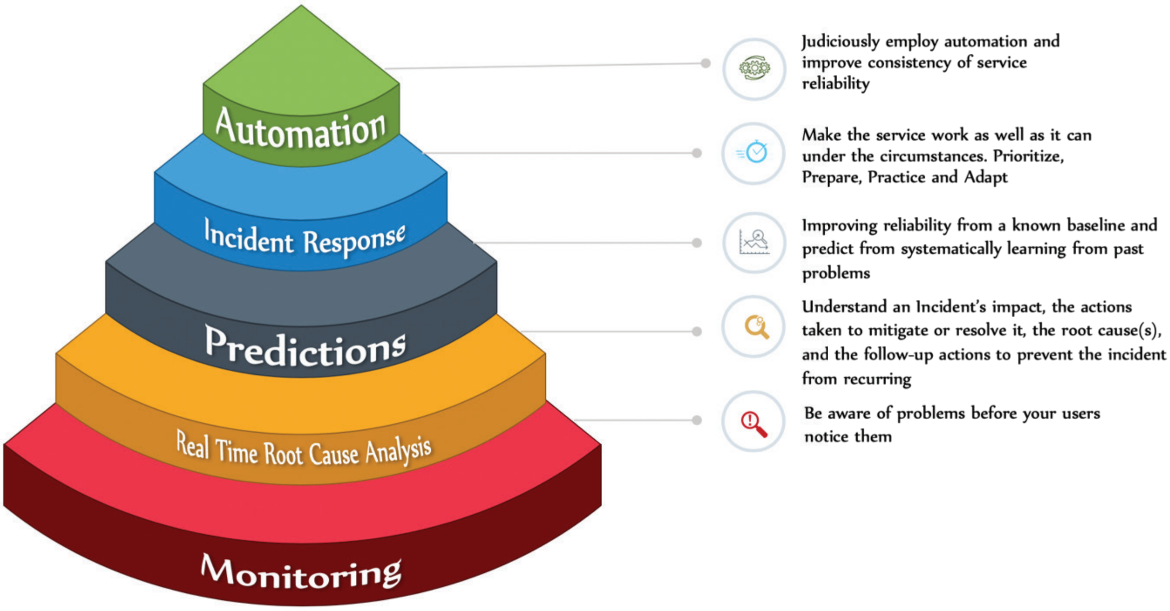
Today, on an average, a single transaction uses 80-100 different technologies like mobile computing, cloud computing, edge servers, IoT, big data, VMs, containers, serverless, to name a few. From the management standpoint, this increases the complexity many-fold. The technology overload introduces multiple points of failure and needs careful coordination & handover of execution control between involved parties. Ensuring the smooth running of such business services becomes a challenge due to the complexity of the transactions and the number of players who need to be doing their part well and working together seamlessly with the other moving parts.

Hierarchy of Service Reliability

While there might be many definitions applied to service reliability, the important elements in Business Service Reliability from basic to advanced are:

- Monitoring
- Real Time Root Cause Analysis
- Predictions
- Incident Response
- Automation

The following picture depicts the objectives for each of the element in the hierarchy.



Four Golden Signals for Monitoring Service Reliability

Championed by the Google SRE team and the larger web-scale SRE community as the most fundamental metrics for tracking service health and performance are the Four Golden Signals. While a team could always monitor more metrics or logs across the system, the four golden signals are the basic, essential building blocks for any effective monitoring strategy as define what it means for the system to be “healthy” - as seen by the actors interacting with that service, either if they are final users or another service in your microservice application.

Here is a brief description of these four golden signals:

Latency

The time it takes to service a request, with a focus on distinguishing between the latency of successful requests and the latency of failed requests.

Traffic

A measure of how much demand is being placed on the service. This is measured using a highlevel service-specific metric, like HTTP requests per second in the case of an HTTP REST API.

Errors

The rate of requests that fail. The failures can be explicit (e.g., HTTP 500 errors) or implicit (e.g., an HTTP 200 OK response with a response body having too few items).

Saturation/Contention

How "full" is the service. This is a measure of the system utilization, emphasizing the resources that are most constrained (e.g., memory, I/O or CPU). Services degrade in performance as they approach high saturation.

AIOps

AIOps — a term coined by Gartner and short for "artificial intelligence for IT operations" —refers to the use of artificial intelligence (AI) and machine learning (ML) to automate data correlation, enable root cause analysis, and deliver predictive insights for both IT teams and businesses. AIOps solutions leverage ML to not only automate routine tasks, but also gather and interpret large volumes of historical data to identify potential problems before they manifest themselves in IT environments.

The common set of features of any AIOps platform to provide insights into data are:

Machine Learning and AI

The core feature of Artificial Intelligence for IT Operations Systems, machine learning (ML) uses predictive and intelligent analysis to supplement and enhance a system's decision-making ability.

Real-Time Processing

AIOps platforms need to be able to analyze and process large amounts of data at speed. Real-time processing allows enterprise IT organizations to respond immediately to issues like anomalies and security breaches.

Deep Reinforcement Learning

AIOps platforms leverage deep reinforcement learning (DRL), which converts observed patterns and learned responses into ever more refined algorithmic behavior. With DRL, algorithmic output is used as a new or additional input to alter existing input values.

Pattern Recognition

AIOps platform must recognize and follow complex rules and patterns, in order to accurately detect and assess events, and respond appropriately.

Domain Algorithms

Domain algorithms define the precise operations and decision-making processes that the AI will prioritize. These are specific to an IT organization's goals and data in a certain industry or environment.

Automation

This is one of the key reasons why AIOps is receiving such enthusiasm from the industry. Effective AIOps solutions and systems reduce IT operators' workloads by automating menial or repetitive tasks, increasing efficiency on the human side of the enterprise.

Data Aggregation

Many Artificial Intelligence for IT Operations platforms carry out the collection and statistical synthesis of varying types of data from an eclectic range of sources.

However, it's not enough to just turn data into insights. Insights aren't actionable on their own, nor are they effective in a vacuum. In order for insights to truly be actionable and effective, they must be integrated directly into workflows that support business services.

Therefore, Business Service Reliability with AI is not limited to alerts management and reducing the noise level of monitoring data. For ensuring service reliability, AIOps platform is:

- Required to ingesting, aggregating and normalizing structured and unstructured data like logs, events, change requests, known errors, configurable items, physical and logical topologies
- Correlating the data, reducing noise, Root Cause Analysis (RCA) and actionable insights
- Analyzing patterns and predicting incidents based on the patterns

If AIOps platform is designed to uncover insights more efficiently and integrate them into workflows for actioning, it can help the business provide reliable services and enhance the end user experience.

ZIF (Zero Incident Framework™)

ZIF.ai – a business unit of GAVS Technologies developed AIOps based TechOps platform - **Zero Incident Framework ™** (**ZIF**) that enables proactive detection and remediation of incidents.

ZIF Platform is available in three versions for our customers to evaluate and experience the power of AI driven Business Service Reliability:

ZIF Business Xpress

ZIF Business Xpress has been engineered for enterprises to evaluate AIOps before adoption. 10 to 40 devices can be connected to ZIF Business Xpress, to experiment with the value proposition.

ZIF Lite

For small and medium enterprises

ZIF Business

Targeted for enterprise wide adoption.

For more details, please visit www.zif.ai