The need for a predictive approach to service design

The case of avoidance v/s fulfillment

The challenge that CIOs today face is to approximate 'new age' digital business IT services using hybrid IT infrastructure estates, consisting of various vintages of investments and varying levels of technical debt. The goal is the 'look-&-feel' of Facebook, with architecture stacks that carry the burden of legacy.

To add to these challenges, is the challenge of a legacy approach to designing 'day 2' operational services. A new approach is required for the hybrid estate, and for the aspirational 'new age' estate (it's always going to be an asymptotical journey).

The design for operational IT services has always been driven by evident demand, leading to the approach of designing for fulfillment. To elaborate, traditionally, IT services have been sized to respond to the volumes of 'tickets'. The conventional wisdom is to size for faults, not anomalies. We differentiate here between the word 'fault' as something that is almost random, versus 'anomaly', that conveys as sense of being predictable in a systemic sense.

The current approach to incident management has consequently been stateless in some sense, where we expend resources on fulfilling demands one at a time, without actively learning how to do it better in the future (or, indeed, not doing it at all!). We can almost hear the ITIL proponents protesting here, citing such proactive threads such as 'problem management', 'capacity planning' et al as steps in that very direction. Suffices to say that many of these threads are forensic at best, with minimal impact on actual service design, by taking the approach of fixing root causes after they occur. While there is impact on design, there is little focus on being predictive (and consequently, avoidance).

The new approach to service design

As our IT systems get more complex, dynamic and organically connected to our businesses, the call is for a far more predictive and 'stateful' approach to service design and demand fulfillment, based on real-time detection of anomalies, avoiding tactical response. What we mean by the term 'stateful' is that the fabric retains its understanding of 'good states' for the system as whole and detects any deviations thereof. In other words, the key is being aware of the state at the systemic level, and using that awareness as the basis for anomaly detection and incident avoidance. We are not suggesting that demand management is not important, we are suggesting that there is place for a more elegant and predictive approach.

To summarize, the 'what' of the predictive, detection-based approach are as follows -

- Look for business-critical IT functionalities provided by IT (applications, features, what have you...)
- Look for the 'heart beats' that represent the health of these functional units, not monitoring information or tickets, but real transactional events that represent the functioning of the system
- Start monitoring those 'heart beats' for anomalies, based on real-time event streams from the environment (wire data, being aware of 'good state')
- Respond only to 'anomalies' in those 'heart beats', 'cut a ticket' when you see an anomaly, not when you see a fault. Incidentally, the response to the ticket could be automated!
- If you miss an anomaly, 'learn', incorporate, repeat

While this set of steps outlines the 'what', we need to ask and answer the question of 'how'.

The 'how' is not about tools or software, as much as it is about changing the design paradigm and dealing with the glide path into a profoundly different mindset that is required to operationalize this new thinking, focused on responding to real issues, avoiding unnecessary work.

The reality is that we, as IT operations professionals, have taken refuge in large delivery organizations, based on (medieval) military principles – light infantry, supported by artillery, followed by armored divisions, and so on. To continue the military analogy, the 'battles' now are more asymmetrical, calling for a different approach.

Avoidance v/s fulfillment

Going back to the core approach here, the question to be answered is how to reform from staffing for fulfillment to staffing for anomalies. We use the word 'staffing' as a placeholder for service design.
There as some core ideas that need to be internalized to achieve this.

The first of these is the real-time visualization of events in the IT landscape. There needs to be a way to understand what happens in real time when an anomaly occurs, just before and just after. The objective is to fit a deterministic model to the environment viz. to be truly predictive. The hard part is to do this in real time, against forensically, as has been the norm. The approach should be to move away from using traditional monitoring thresholds to cut tickets (only), to using model or patterns to dynamically detect anomalies as the basis of orchestrating response (and 'root cause identification').

The next idea is to design operations to react only to the anomalies, as against faults. This, in itself will reduce the need for staff on standby. The trick here is to figure out what the reduction in demand will be, and hence, time the ramp-down. An analytics approach will provide the first clues. The challenge is to buy in to that approach without undermining the inherent stability and reliability of the environment that is currently underwritten by having lots of people waiting around for a failure. How do you go from that protection to truly lean staffing, with the belief that it is sufficient?

The approach must be inductive and based on (machine) learning.

• First, to enable the inductive approach, start small viz. start with one business function and the support IT landscape. Size the support for that part of the environment on the principles described herein – reduce staff there. Try and test. If that works, add more functions as you go along, and commit to take out the cost there and then. The accruing benefit is to move from one paradigm to another gradually, ensuring the advertised availability and reliability parameters through the period of change.

• The second (and inherently interconnected) idea is that of learning – viz. the ideal of improving detection using machine-based methods. The reality is that you will start with a limited set of filters that do not detect all anomalies (especially in highly interconnected, dynamic environments). Those undetected anomalies (that will be caught later by conventional means – monitor → correlated → ticket), need to form the basis to formulate new filters. And, this needs to be done using machine learning models, not manually (concept assisted machine learning).

Conclusion

The result of the new approach will be a two-toned operations landscape for some time: one part managed for anomalies, the other managed for fulfillment. The idea is for the former to replace the latter in quick time – quick defined by each situation.

None of these ideas forego the need for conventional instrumentation of the IT environment viz. The monitor → correlate → ticket. This is the scaffolding that provides the last line of defense and the basis of intelligently adjusting anomaly detection algorithms.

The ideas here are simple. Look for trouble in real time, and as perceived by the business, find the most elegant and parsimonious way to detect, mitigate and remediate.

About the authors

Murthy Malapaka
Head of Transformation Services - North America, Wipro Ltd.

Murthy has 28 years of experience as a technology innovator and change agent. Over these years, he has assumed various technology leadership roles across application and infrastructure architecture domains, specializing in availability and reliability. He has been providing consulting services to CIOs and CTOs in their journey from client server to on demand infrastructure services.
Murthy can be reached at murthy.malapaka@wipro.com.

Ashutosh Mankar
General Manager – Client Engagements, Wipro Ltd.

Ashutosh has over 3 decades of industry experience in the area of cloud and infrastructure services, including solution architecture, service delivery and business development. As the client engagements leader for clients in the US, he is responsible for ensuring Wipro's continuing relevance for clients.
Ashutosh can be reached at Ashutosh.mankar@wipro.com.

Wipro Limited
Doddakannelli, Sarjapur Road,
Bangalore-560 035, India

Tel: +91 (80) 2844 0011
Fax: +91 (80) 2844 0256
wipro.com

Wipro Limited (NYSE: WIT, BSE: 507685, NSE: WIPRO) is a leading global information technology, consulting and business process services company. We harness the power of cognitive computing, hyper-automation, robotics, cloud, analytics and emerging technologies to help our clients adapt to the digital world and make them successful. A company recognized globally for its comprehensive portfolio of services, strong commitment to sustainability and good corporate citizenship, we have over 175,000 dedicated employees serving clients across six continents. Together, we discover ideas and connect the dots to build a better and a bold new future.

For more information, please write to us at **info@wipro.com**