

CLOUD COMPUTING

Issue 5

WORLD

2015

Data Gathering in the Cloud:
Making the Most of Network Analysis

The Transformative
Power of Google

Road to the Cloud
Choosing the Right Companion



Advanced Analytics

Hybrid Cloud Environments

Budget-Based IT

Launch Partners

CBRE



cloudfordata
accelerate analytics



Software

Cloudy skies and rain mean up-
loads are slow and files are hard to
find. Download them, man. It's like

DATA GATHERING IN THE CLOUD

Anton Basil addresses the challenges of data gathering in a dynamic cloud environment

By Anton Basil, Chair of the CEF Analytics Group

Introduction

“Why is the phone connection bad?” “Why is my website down?” “Why is this download taking ages?” The customer does not really want an answer to these questions. On one hand the answer is already known – “these things happen” – on the other hand the customer is not so much asking a question as hinting that they will take their business to another provider if things don’t get better. And that is why the provider really does need to know the answers – and will invest a lot in network analytics to find the answers.

Analytics have become a critical part of the network infrastructure, and not just for troubleshooting purposes. Properly applied, analytics can identify potential bottlenecks and help prevent performance degradation, and can help make the network more efficient, secure and reliable. In a dynamic cloud environment delivering services on demand, it is even more important to manage performance in real time to ensure a good user experience, but at the same time it is becoming much harder to do this. So, what are the challenges to be addressed? As a process, network analytics can be usefully subdivided into three functions: data gathering, data mining and reporting.

Data Gathering in the Cloud

Data gathering is the analytics bit: the more data about actual network behaviour that can be garnered across the entire infrastructure, the greater chance of finding all the answers. But it is subtler than that, because experience teaches that some sorts of data are more useful than others, and it can be better to target essential

parameters rather than have them lost under a mountain of useless data. In addition, you do not want the data collection itself to impact the network operation – avoiding, for example, measuring latency in a way that adds latency.

In a single static network the difficulty of gathering data is proportional to the network’s complexity: a problem that grew steadily as the network expanded and may have kept pace with the IT department’s increasing expertise. Such steady, organic growth is not possible in a dynamic, virtualized network and there is a corresponding need to virtualize the monitoring and test process to keep abreast of the changes in the network and to reduce the cost and labour of physical testing. But there is also a risk that the virtual test software will consume some of the processing resources and itself impact the measured performance. Under controlled laboratory test conditions steps can be taken to compensate for such errors, but it is not so easy to do this in a working environment.

Not only is the data being gathered on a moving target, the problem is compounded when the network spans several provider domains. Even when there is clear agreement about the definition of the various network parameters being measured, different providers might not measure or collate their data in the same way. So, from a data-gathering point of view, cloud services mean a quantum leap in complexity. We no longer have a process that evolves steadily as a single network grows, rather we are forced to correlate inconsistent data across a shifting ecosystem. Trying to gather useful and reliable data in real time will only be possible when there

are global standards defining the key parameters and the way they are measured, allowing providers and different networks to mix and match data in a simple reliable manner.

The Data Mining Challenge

Data mining is literally the analytics bit. The more data that has been gathered, the greater the chance of finding every answer, but it can take much longer to find those answers if too much irrelevant data is included. The key to data mining, however, is to know the right questions to ask – and that does still depend on human experience and judgement. Even if we do know what questions to ask, the answers can only be as good as the data that has been gathered. So, data mining cannot be usefully applied until the data-gathering problem has been sorted. Once sufficient reliable and accurate data is available, and we know what questions to ask, then what is needed is sufficient processing power to perform the analysis fast enough for the resulting output to still be relevant. In today's dynamic network environment that means getting answers in near real time.

Relevant Reporting of Results

Reporting is no longer about churning out reams of statistics. Presenting information about a highly complex system to a human operator requires a more visual format. A topological map of the infrastructure with key data or trouble spots highlighted at the right point on the structure so the operator can immediately see what is happening can provide this network visibility. But reporting can also serve other audiences: an alarm system would require precise co-ordinates rather than a visual map, and an application would be less concerned with locating the source of degradation than knowing its impact on performance parameters.

The CloudEthernet Forum

A dynamic cloud environment becomes a sensitive organism where small local problems can cause ripple effects that lead to disastrous consequences on many levels: poor service performance, loss of critical business for the customer, reputation damage for the provider, and ultimately customer churn. At the same time, it is becoming harder to garner and analyse sufficient data about the inner working of this complex environment. That is why network analytics is one of the five VASPA fundamentals being addressed by the CloudEthernet Forum (CEF): Virtualization, Automation, Security, Programmability and Analytics. If a cloud services consumer experiences degradation in service performance, what can be done? The cloud services provider might offer to boost processing by spinning up further VM resources. The cloud carrier might provide additional bandwidth. Is this a cloud carrier or a cloud provider issue, or has the consumer got unrealistic expectations?

In a competitive business environment this is the sort of situation that can degenerate into

finger pointing and loss of business. What the CEF is doing is recruiting members from all these cloud stakeholder groups to work together on strategies to anticipate and resolve the challenges of a new and disruptive technology.

The cloud stakeholders being recruited by the CEF include major cloud consumers, who can help clarify what is wanted in terms of performance, in order to determine how best to measure it. For example: it is obvious that a cloud consumer subscribing to a streaming video service wants brilliant high-quality video, but the industry has found out that this subjective experience depends on essential parameters like bandwidth latency, jitter and packet loss that can be readily measured and analysed. Hence the CEF needs both cloud carriers and cloud providers onboard to understand the consumer needs, and to find ways to satisfy them.

This is as much about business practice and working relationships as technology: if the CEF can define common standards that will make different provider and carrier systems compatible, would the cloud carrier allow the cloud provider some control of its network? When the consumer complains about service performance, can the provider have access to analyse both network and data centre performance and come up with the optimal balance between processing and bandwidth resources? Or would the carrier be allowed access to the provider's system to do the analysis and resolve the issue if asked? Compare this with the mobile phone industry that has already learned to iron out their differences and make service as seamless as possible. The mobile users expect reasonably consistent performance on their travels, so if they complain that a phone connection has dropped, they do not expect an argument between the various cell services about which network was responsible for the failure.

Conclusion

This is how it should be with cloud services – and so the CEF is working on a defining standard interface and standard APIs between cloud providers and carriers to create an open cloud environment. This is key to enabling consistent data gathering and consistent definitions that will provide sufficient reliable input for data mining across multiple networks. How the data will be mined on sufficient scale is another issue, and the way the results will be reported or presented is another of the key concerns for the CEF – all the more so since analytics output can provide essential data for automation, security and the other VASPA fundamentals.

To achieve these aims, and to create an open cloud environment fast enough to maintain the momentum of cloud migration, the CEF is inviting all types of cloud stakeholders, including systems integrators, NEMs and software developers – as well as the enterprise customers and service providers already referred to – to participate in the standardisation process.