Moneyball Your Network with Big Data Analytics

BY ALEX HENTHORN-IWANE, VP MARKETING · OCT 19, 2015

Turning data analytics into significant competitive advantage

Recently, the team at Kentik tweeted the following: "#Moneyball your network with deeper, real-time insights from #BigData NetFlow, SNMP & BGP in an easy to use #SaaS." There are a lot of concepts packed into that statement, so I thought it would be worth unpacking for a closer look. We'll use the Moneyball analogy as a way of exploring the vast, untapped potential of the telemetry data that's being generated by telecom, cable, web enterprise, hosting, and other IP networks.

Most of you are probably familiar with the concept of Moneyball, which was popularized by Michael Lewis' book (and later movie) of the same name. Lewis chronicled how Billy Beane, the GM of the Oakland Athletics, upended baseball operations with a new approach built on applying analytics to statistical data on player performance. As Lewis tells it, baseball was a field where tons of statistical data has been gathered and analyzed on an amateur basis for many decades but most of that data was ignored by the professionals running baseball operations. By mining data that was copiously available but previously unleveraged, Beane gained competitive advantage for his team, and GMs across baseball began seeing the value hidden in arcane stats.

Like baseball, network operations is a field in which a huge volume of data is available. IP networks emit a ton of statistical information — flow data (NetFlow, sFlow, IPFIX, etc.), SNMP, and routing data from BGP — that is routinely collected. And though the underlying reasons differ, network operations, like pre-Moneyball baseball operations, has yet to fully leverage this wealth of available data. The problem hasn't been that the data has been discounted or ignored, but rather that traditional approaches available for handling the data are obsolete and ineffective, making it difficult to extract actionable insight.

The key realization here is that network telemetry data is big data. Any decent-size network, especially one that is multi-homed to the Internet, can easily generate millions or billions of telemetry records per day. Unfortunately the techniques typically applied to ingest, process, and analyze that data are about 15 years out

of date. Neither traditional on-premises enterprise software nor appliance-based architectures were built to handle the scale of network data happening today. In fact, they have so little capacity that in most cases you have to guess well ahead of time what kinds of questions you may need to ask in the future, so that the desired details can be summarized into pre-canned aggregates and reports. Most of the rest of the raw data is simply thrown away. But real life usually doesn't fit neatly into pre-canned reports. By the time you realize that you need to ask different questions, it's too late: the raw data is long gone.

Big Data and cloud architecture change this picture. The trick is to build a system that truly addresses the need for both real-time operational insights and longer-term business planning. It's easy to skew to one side or the other. If you lean towards answering longer-term questions, you'll typically collect data from different sources discretely, and hold off on processing until query run-time. That approach preserves flexibility but isn't fast enough to be operationally useful. On the other hand, you can design for real-time performance by limiting the questions that can be asked at any given moment, but then you don't have much flexibility.

Architectures without big data scalability make you guess in advance what questions you'll need answered in the future.

The beauty of Kentik is that we've developed a third way, building a big data approach that covers both ends of the short vs. long spectrum. Kentik Data Engine (KDE), the datastore behind Kentik Detect, is architected for real-time ingest of full-resolution network data at Terabit scale. And as it ingests it also brings diverse data types into a single, unified schema. That gives you a rich, deep set of detailed data on which to run real-time queries that ask any question you want. Your short-term "monitoring" questions might be something like "Is there an anomaly occurring right now?" or "Is anyone on our network on the receiving end of a DDoS attack?" The answers help you to improve operational performance and efficiency and to deliver better application and service performance. As for longer-term questions, you might want to know, "Who can I peer with to reduce my transit costs?" or "Are there organizations with significant traffic transiting my network that I can potentially convert into customers?"

This is where we come back to Moneyball. You can only fully leverage the data emanating from your network if you're able to access all of the raw detail in an environment that enables fast, flexible querying across both short- and long-term time-spans. Once you've got that, you can move aggressively to convert the answers to your questions into a boost in performance and ROI. And that's what ultimately gives you the Moneyball effect: turning data analytics into significant competitive advantage.

If you're not sure that you have that kind of advantage going for your network I encourage you to check out www.kentik.com to get an idea of what I'm talking about. I'd also love to hear what you think. Hit me up via my LinkedIn or on twitter at @heniwa. Thanks!

Ready for more information?
Please email us at info@kentik.com or visit us at www.kentik.com.