

# ***Big Data Meets Complex Event Processing: AccelOps Delivers a Better Architecture to Attack the Data Center Monitoring and Analytics Problem***

*Transcript of a BriefingsDirect podcast on how enterprises can benefit from capturing and analyzing real-time data to improve IT management.*

**Listen to the [podcast](#). Find it on [iTunes/iPod](#). Download the transcript. Sponsor: [AccelOps](#)**

**Connect with AccelOps: [Linkedin](#), [Twitter](#), [Facebook](#), [RSS](#).**

**Dana Gardner:** Hi. This is [Dana Gardner](#), Principal Analyst at [Interarbor Solutions](#), and you're listening to [BriefingsDirect](#).



Today, we present a sponsored podcast discussion on how new data and analysis approaches are providing significantly improved IT operations monitoring, as well as stronger [security](#). We'll examine how advances in big [data analytics](#) and [complex events processing \(CEP\)](#) can come together to provide deep and real-time pattern-based insight into large scale IT operations.

[AccelOps](#) has developed the technology to correlate events with relevant data across IT systems, so that operators can gain much better insights faster, and then learn as they go to better predict future problems before they emerge. [Disclosure: [AccelOps](#) is a sponsor of [BriefingsDirect podcasts](#).]

With us now to explain how these new solutions can drive better IT monitoring and remediation response and keep those critical systems performing at their best is our guest, [Mahesh Kumar](#), Vice President of Marketing at AccelOps. Welcome to BriefingsDirect, Mahesh.

**Mahesh Kumar:** Dana, glad to be here.

**Gardner:** It's always been difficult to gain and maintain comprehensive and accurate analysis of large-scale IT operations, but it seems, Mahesh, that this is getting more difficult. I think there have been some shifts in computing in general in these environments that makes getting a comprehensive view of what's going on perhaps more difficult than ever. Is that fair in your estimation?

**Kumar:** Absolutely, Dana. There are several trends that are fundamentally questioning existing and traditional ways of monitoring a data center.

**Gardner:** Of course we're seeing lots of [virtualization](#). People are getting into higher levels of density, and so forth. How does that impact the issue about monitoring and knowing what's going on with your systems? How is virtualization a complexity factor?

**Kumar:** If you look at trends, there are on average about 10 [virtual machines \(VMs\)](#) to a physical [server](#). Predictions are that this is going to increase to about 50 to 1, maybe higher, with advances in hardware and virtualization technologies. So that's one trend, the increase in density of VMs is a complicating factor for capacity planning, capacity management, performance management, and security.



Corresponding to this is just the sheer number of VMs being added in the enterprise. Analysts estimate that just in the last few years, we have added as many VMs as there were physical machines. In a very short period of time, you have in effect seen a doubling of the size of the IT management problem. So there are a huge number of VMs to manage and that introduces complexity and a lot of data that is created.

Moreover, your workloads are constantly changing. [vMotion](#) and [DRS](#) are causing changes to happen in hours, minutes, or even seconds, whereas in the past, it would take a week or two for a new server to be introduced, or a server to be moved from one segment of the network to the other.

So change is happening much more quickly and rapidly than ever before. At the very least, you need monitoring and management that can keep pace with today's rate of change.

## *Cloud computing*

[Cloud computing](#) is another big trend. All analyst research and customer feedback suggests that we're moving to a hybrid model, where you have some workloads on a public cloud, some in a private cloud, and some running in a traditional [data center](#). For this, monitoring has to work in a distributed environment, across multiple controlling parties.



Last but certainly not the least, in a hybrid environment, there is absolutely no clear perimeter that you need to defend from a security perspective. Security has to be pervasive.

Given these new realities, it's no longer possible to separate performance monitoring aspects from security monitoring aspects, because of the distributed nature of the problem. You can't have two different sets of eyes looking at multiple points of presence, from different angles and then try to piece that together.

Those are some of the trends that are causing a fundamental rethink in how IT monitoring and management systems have to be architected.

**Gardner:** And even as we're seeing complexity ramp up in these data centers, many organizations are bringing these data centers together and consolidating them. At the same time, we're seeing more spread of IT into remote locations and offices. And we're seeing more use of mobile and distributed activities for data and application. So we're not only talking about complexity, but we're talking about scale here.

**Kumar:** And very geographically distributed scale. To give you an example, every office with [voice over IP \(VoIP\)](#) phones needs some servers and network equipment in their office, and those servers and network equipment have to be secured and their uptime guaranteed.

So what was typically thought of as a remote office now has a mini data center, or at least some elements of a data center, in it. You need your monitoring and management systems to have the reach and can easily and flexibly bring those under management and ensure their availability and security.

**Gardner:** What are some of the ways that you can think about this differently? I know it's sort of at a vision level, but typically in the past, people thought about a system and then the management of that system. Now, we have to think about really clouds and fabrics. We're just using a different vocabulary to describe IT. I suppose we need to have a different vocabulary to describe how we manage and monitor it as well.

**Kumar:** The basic problem you need to address is one of analysis. Why is that? As we discussed earlier, the scale of systems is really high. The pace of change is very high. The sheer number of configurations that need to be managed is very large. So there's data explosion here.

Since you have a plethora of information coming at you, the challenge is no longer collection of that information. It's how you analyze that information in a holistic manner and provide consumable and actionable data to your business, so that you're able to actually then prevent problems in the future or respond to any issues in real-time or in near real-time.

You need to nail the real-time analytics problem and this has to be the centerpiece of any monitoring or management platform going forward.

### ***Fire hose of data***

**Gardner:** In the past, this fire hose of data was often brought into a repository, perhaps indexed and analyzed, and then over time, reports and analysis would be derived from it. That's the way that all data was managed at some point.

But we really can't take the time to do that, especially when we have to think about real-time management. Is there a fundamental change in how we approach the data that's coming from IT systems in order to get a better monitoring and analysis capability?

**Kumar:** The data has to be analyzed in real-time. By real-time I mean in streaming mode before the data hits the disk. You need to be able to analyze it and make decisions. That's actually a very

efficient way of analyzing information. Because you avoid a lot of data sync issues and duplicate data, you can react immediately in real time to remediate systems or provide very early warnings in terms of what is going wrong.

The challenges in doing this streaming-mode analysis are scale and speed. The traditional approaches with pure relational databases alone are not equipped to analyze data in this manner. You need new thinking and new approaches to tackle this analysis problem.

**Gardner:** Also for issues of security, you don't want to find out about security by going back and analyzing a bunch of data in a repository. You want to be able to look and find correlations about what's going on, where attacks might be originating, how that might be cutting across different aspects of your infrastructure.

People are trying different types of attacks. So this needs to be in real-time as well. It strikes me that if you want to solve security as well as monitoring, that that is also something that has to be in real-time and not something that you go back to every week or months.

**Kumar:** You might be familiar with [advanced persistent threats \(APTs\)](#). These are attacks where the attacker tries their best to be invisible. These are not the brute-force attacks that we have witnessed in the past. Attackers may hijack an account or gain access to a server, and then over time, stealthily, be able to collect or capture the information that they are after.

These kinds of threats cannot be effectively handled only by looking at data historically, because these are activities that are happening in real-time, and there are very, very weak signals that need to be interpreted and there is a time element of what else is happening at that time. What seems like disparate sets of activity have to be brought together to be able to provide a level of defense or a defense mechanism against these APTs. This too calls for streaming-mode analysis.

If you notice, for example, someone accessing a server, a database administrator accessing a server for which they have an admin account, it gives you a certain amount of feedback around that activity. But if on the other hand, you learn that a user is accessing a database server for which they don't have the right level of privileges, it may be a red flag.

You need to be able to connect this red flag that you identify in one instance with the same user trying to do other activity in different kinds of systems. And you need to do that over long periods of time in order to defend yourself against APTs.

## *Advances in IT*

**Gardner:** So we have the modern data center, we have issues of complexity and virtualization, we have scale, we have data as a deluge, and we need to do something fast in real-time and consistently to learn and relearn and derive correlations.

It turns out that there are some advances in IT over the past several years that have been applied to solve problems outside of IT, that is to say the operations part of IT, that can be brought to bear here.

This is one of the things that really jumped out at me when I did my initial briefing with AccelOps. You've looked at what's being done with big data and in-memory architectures, and you've also looked at some of the great work that's been done in [services-oriented architecture \(SOA\)](#) and CEP, and you've put these together in an interesting way.

Let's talk about what the architecture needs to be in order to start doing for IT what we have been doing with retail data or looking at complex events in a financial environment to derive inference into what's going on in the real world. What is the right architecture, now that we need to move to for this higher level of operations and monitoring?

**Kumar:** Excellent point, Dana. Clearly, based on what we've discussed, there is a [big-data](#) angle to this. And, I want to clarify here that big data is not just about volume.

[Doug Laney](#), a META and a [Gartner](#) analyst, probably put it best when he highlighted that big data is about volume, the velocity or the speed with which the data comes in and out, and the variety or the number of different data types and sources that are being indexed and managed. I would add to this a fourth V, which is verdicts, or decisions, that are made. How many decisions are actually impacted or potentially impacted by a slight change in data?

For example, in an IT management paradigm, a single configuration setting can have a security implication, a performance implication, an availability implication, and even a capacity implication in some cases. Just a small change in data has multiple decision points that are affected by it. From our angle, all these different types of criteria affect the big data problem.

When you look at all these different aspects of IT management and how it impacts what essentially presents itself as a big data challenge or a big data problem, that's an important angle that all IT management and monitoring products need to incorporate in their thinking and in their architectures, because the problem is only going to get worse.

**Gardner:** Understanding that big data is the issue, and we know what's been done with managing big data in this most comprehensive definition, how can we apply that realistically and practically to IT systems?

It seems to me that you are going to have to do more with the data, cleansing it, discovering it, and making it manageable. Tell me how we can apply the concepts of big data that people have been using in retail and these other applications, and now point that at the IT operations issues and make it applicable and productive.

## *Couple of approaches*

**Kumar:** I mentioned the analytics ability as central to monitoring systems – big-data analytics to be specific. There are a couple of approaches. Some companies are doing some really interesting work around big-data analysis for IT operations.

They primarily focus on gathering the data, heavily indexing it, and making it available for search, thereby derive analytical results. It allows you to do forensic analysis that you were not easily able to with traditional monitoring systems.

The challenge with that approach is that it swings the pendulum all the way to the other end. Previously we had a very rigid, well-defined relational data-models or data structures, and the index and search approach is much more of a free form. So the pure [index-and-search](#) type of an approach is sort of the other end of the spectrum.

What you really need is something that incorporates the best of both worlds and puts that together, and I can explain to you how that can be accomplished with a more modern architecture. To start with, we can't do away with this whole concept of a model or a relationship diagram or entity relationship map. It's really critical for us to maintain that.

I'll give you an example why - when you say that a server is part of a network segment, and a server is connected to a switch in a particular way, it conveys certain meaning. And because of that meaning, you can now automatically apply policies, rules, patterns and automatically exploit the meaning that you capture purely from that relationship. You can automate a lot of things just by knowing that.

If you stick to a pure index-and-search approach, you basically zero out a lot of this meaning and you lose information in the process. Then it's the operators who have to handcraft these queries to have to then reestablish this meaning that's already out there. That can get very, very expensive pretty quickly.

Even at a fairly small scale, you'll find more and more people having to do things, and a pure index and search approach really scales with people, not as much with technology and automation. Index and search certainly adds a positive dimension to traditional monitoring tools but that alone is not the answer for the future.

Our approach to this big-data analytics problem is to take a hybrid approach. You need a flexible and extensible model that you start with as a foundation, that allows you to then apply meaning on top of that model to all the extended data that you capture and that can be kept in flat files and searched and indexed. You need that hybrid approach in order to get a handle on this problem.

**Gardner:** I suppose you also have to have your own architecture that can scale. So you're going to concepts like [virtual appliances](#) and scaling on-demand vis-à-vis [clustering](#), and taking advantage of in-memory and streaming capabilities to manage this. Tell me why you need to

think about the architecture that supports this big data capability in order for it to actually work in practical terms?

**Kumar:** You start with a fully virtualized architecture, because it allows you not only to scale easily. From a reach standpoint, with a virtualized architecture, you're able to reach into these multiple disparate environments and capture and analyze and bring that information in. So, virtualized architecture is absolutely essentially for you to start with.

### *Auto correlate*

**M**aybe more important is the ability for you to auto-correlate and analyze data, and that analysis has to be distributed analysis. Because whenever you have a big data problem, especially in something like IT management, you're not really sure of the scale of data that you need to analyze and you can never plan for it.

Let me put it another way. There is no server big enough to be able to analyze all of that. You'll always fall short of compute capacity because analysis requirements keep growing. Fundamentally, the architecture has to be one where the analysis is done in a distributed manner. It's easy to add compute capacity by scaling horizontally. Your architecture fits how computing models are evolving over the long run. So there are a lot of synergies to be exploited here by having a distributed analytics framework.

Think of it as applying a [MapReduce](#) type of algorithm to IT management problems, so that you can do distributed analysis, and the analysis is highly granular or specific. In IT management problems, it's always about the specificity with which you analyze and detect a problem that makes all the difference between whether that product or the solution is useful for a customer or not.

**Gardner:** In order for us to meet our requirements around scale and speed, we really have to think about the support underneath these capabilities in a new way. It seems like, in a sense, architecture is destiny, when it comes to the support and monitoring for these large volumes in this velocity of data.

Let's look at the other part of this. We talked about the big data, but in order for the solution to work, we're looking at CEP capabilities in an engine that can take that data and then work with it and analyze it for these events and these programmable events and looking for certain patterns.

Now that we understand the architecture and why it's important, tell me why this engine brings you to a higher level and differentiates you in the field around the monitoring?

**Kumar:** A major advantage of distributed analytics is that you're freed from the scale-versus-richness trade-off, from the limits on the type of events you can process. If I wanted to do more complex events and process more complex events, it's a lot easier to add compute capacity by just simply adding VMs and scaling horizontally. That's a big aspect of automating deep forensic analysis into the data that you're receiving.

I want to add a little bit more about the richness of CEP. It's not just around capturing data and massaging it or looking at it from different angles and events. When we say CEP, we mean it is advanced to the point where it starts to capture how people would actually rationalize and analyze a problem.

For example, the ability for people in a simple visual snapshot to connect three different data points or three events together and say that they're interrelated and they point to a specific problem.

The only way you can automate your monitoring systems end to end and get more of the human element out of it is when your CEP system is able to capture those nuances that people in the [NOC](#) & [SOC](#) would normally use to rationalize when they look at events. You not only look at a stream of events, you ask further questions and then determine the remedy.

### ***No hard limits***

**T**o do this, you should have a rich data set to analyze, i.e. there shouldn't be any hard limits placed on what data can participate in the analysis and you should have the flexibility to easily add new data sources or types of data. So it's very important for the architecture to be able to not only event on data that are is stored in in traditional models or well-defined relational models, but also event against data that's typically serialized and indexed in [flat file databases](#).

This hybrid approach basically breaks the logjam in terms of creating these systems that are very smart and that could substitute for people in terms of how they think and how they react to events that are manifested in the NOC. You are not bound to data in an inflexible vendor defined model. You can also bring in the more free-form data into the analytics domain and do deep and specific analysis with it.

Cloud and virtualization are also making this possible. Although they've introduced more complexity due to change frequency, distributed workloads etc. they've also introduced some structure into IT environments. An example here is the use of converged infrastructure ([Cisco UCS](#), [HP Blade Matrix](#)) to build private-cloud environments. At least at the infrastructure level it introduces some order and predictability.

**Gardner:** All right, Mahesh, we've talked about the problem in the market, we have talked about high-level look at the solution and why you need to do things differently, and why having the right architecture to support that is important, but let's get into the results.

If you do this properly, if you leverage and exploit these newer methods in IT, like big data, analytics, CEP, virtual appliances and clustered instances of workload and support and when you apply all those to this problem about the fire hose of data coming out of IT systems, a comprehensive look at IT in this fashion, what do you get? What's the payoff if you do this properly?

**Kumar:** I want to answer this question from a customer standpoint. It is no surprise that our customers don't come to us saying we have a big data problem, help us solve a big data problem, or we have a complex event problem.

Their needs are really around managing security, performance and configurations. These are three interconnected metrics in a virtualized cloud environment. You can't separate one from the other. And customers say they are so interconnected that they want these managed on a common platform. So they're really coming at it from a business-level or outcome focused perspective.

What AccelOps does under the covers, is apply techniques such as big-data analysis, complex driven processing, etc., to then solve those problems for the customer. That is the key payoff -- that customer's key concerns that I just mentioned are addressed in a unified and scalable manner.

An important factor for customer productivity and adoption is the product user-interface. It is not of much use if a product leverages these advanced techniques but makes the user interface complicated - you end up with the same result as before. So we've designed a [UI](#) that's very easy to use, requires one or two clicks to get the information you need, UI driven ability to compose rich events and event patterns. Our customers find this very valuable, as they do not need super-specialized skills to work with our product.

**Gardner:** What's important to think about when we mention your customers is not just applying this value to an enterprise environment, but increasingly the cloud, the virtualization, the importance of managing performance to very high standards, is impacting the cloud providers, [managed service providers \(MSPs\)](#), and [software-as-a-service \(SaaS\)](#) providers.

## *Up and running*

**T**his sounds like an architecture, an approach, and a solution that's going to really benefit them, because their bread and butter is about keeping all of the systems up and running and making sure that all their [service level agreements \(SLAs\)](#) and contracts are being managed and adhered to.

Just to be clear, we're talking about an approach or fairly large cross-section of the modern computing world, enterprises, and many stripes of what we consider service providers.

**Kumar:** Service providers are a very significant market segment for us and they are some of our largest customers. The reason they like the architecture that we have, very clearly, is that it's scalable. They know that the architecture scales as their business scales.

They also know that they get both the performance management and the security management aspects in a single platform. They're actually able to differentiate their customer offerings compared to other MSPs that may not have both, because security becomes really critical.

For anyone wanting to outsource to an MSP, the first question or one of the first questions that they are going to ask, in addition to the SLAs, are how are you going to ensure security? So to have both of those options is absolutely critical.

The third piece really is the fact that our architecture is [multi-tenant](#) from day one. We're able to bring customers on board with a one-touch mechanism, where they can bring the customer online, apply the right types of policies, whether it's SLA policies or security policies, automatically in our product and completely segment the data from one customer to the other.

All of that capability was built into our products from day one. So we didn't have to retrofit any of that. That's something our cloud-service providers and managed service provider customers find very appealing in terms of adopting AccelOps products.

Subscription based licensing, which we offer in addition to perpetual licensing, fits well with the CSP/MSP business model.

**Gardner:** All right. Let's introduce your products in a little bit more detail. We understand you have created a platform, an architecture, for doing these issues or solving these issues for these very intense types of environments, for these large customers, enterprises, and service providers, but you are separate. You have a product for security and a product for performance. Tell us a little bit about your portfolio.

### ***Key metrics***

**Kumar:** What we've built is a platform that monitors data center performance, security, and configurations. The three key interconnected metrics in virtualized cloud environments. Most of our customers really want that combined and integrated platform. Some of them might choose to start with addressing security, but they soon bring in the performance management aspects into it also. And vice versa.

And we take a holistic cross-domain perspective -- we span server, storage, network, virtualization and applications.

What we've really built is a common consistent platform that addresses these problems of performance, security, and configurations, in a holistic manner and that's the main thing that our customers buy from us today.

**Gardner:** Tell us a little bit about the vision, before we go into some more detail and close out. Are we really, Mahesh, getting to the point when we start employing solutions like yours, and take that comprehensive view in real-time? It sounds as if we're doing business intelligence for IT.

We really are getting to the point where we can have dashboards and we are not just making inferences and guesses. We're not just doing Boolean searches on old or even faulty data. We're really looking at the true data, the true picture in real-time, and therefore starting to do the analysis that I think can start driving productivity to even newer heights than we have been accustomed to. So is that the vision, [business intelligence \(BI\)](#) for IT?

**Kumar:** I guess you could say that. To break it down, from an IT management and monitoring standpoint, it is on an ongoing basis to continuously reducing the per capita management costs. As you add the number of VMs or devices, you simply cannot scale the management cost, in a linear fashion. You want to have continuously reducing management cost for every new VM added or new device introduced.

The way you do that is obviously through automation and through a self-learning process, whereby as you continue to learn more and more about the behavior of your applications and infrastructure, you're able to start to easily codify more and more of those patterns and rules in the system, thereby taking sort of the human element out of it bit by bit.

What we have as a product and a platform is the ability for you to increase the [return on investment \(ROI\)](#) on the platform as you continue to use that platform day-to-day. You add more information and enrich the platform with more rules, more patterns, and complex events that you can detect and potentially take automated actions on in the future.

So we create a virtuous cycle, with our product returning higher and higher return on your investment with time. Whereas, in traditional products, scale and longevity have the opposite effect.

So that's really our vision. How do you reduce the per capita management cost as the scale of the enterprises start to increase, and how do you increase more automation as one of the elements of reducing the management cost within IT?

**Gardner:** You have given people a path to start in on this, sort of a crawl-walk-run approach. Tell me how that works. I believe you have a trial download, an opportunity for people to try this out for free.

### ***Free trial download***

**Kumar:** Most of our customers start off with the free trial download. It's a very simple process. Visit [www.accelops.com/download](http://www.accelops.com/download) and download a virtual appliance trial that you can install in your data center within your firewall very quickly and easily.

Getting started with the AccelOps product is pretty simple. You fire up the product and enter the credentials needed to access the devices to be monitored. We do most of it agentlessly, and so you just enter the credentials, the range that you want to discover and monitor, and that's it. You get started that way and you hit Go.

The product then uses this information to determine what's in the environment. It automatically establishes relationships between them, automatically applies the rules and policies that come out of the box with the product, and some basic thresholds that are already in the product that you can actually start measuring the results. Within a few hours of getting started, you'll have measurable results and trends and graphs and charts to look at and gain benefits from it.

That's a very simple process, and I encourage all our listeners and readers to download our free trial software and try AccelOps.

**Gardner:** I also have to imagine that your comments a few moments ago about not being able to continue on the same trajectory when it comes to management is only going to accelerate the need to automate and find the intelligent rather than the hard or laborious way to solve this when we go to things like cloud and increased mobility of workers and distributed computing.

So the trends are really in your favor. It seems that as we move towards cloud and mobile that at some point or another organizations will hit the wall and look for the automation alternative.

**Kumar:** It's about automation and distributed analytics and about getting very specific with the information that you have, so that you can make absolutely more predictable, 99.9 percent correct of decisions and do that in an automated manner. The only way you can do that is if you have a platform that's rich enough and scalable and that allows you to then reach that ultimate goal of automating most of the management of these diverse and disparate environments.

That's something that's sorely lacking in products today. As you said, it's all brute-force today. What we have built is a very elegant, easy-to-use way of managing your IT problems, whether it's from a security standpoint, performance management standpoint, or configuration standpoint, in a single integrated platform. That's extremely appealing for our customers, both enterprise and cloud-service providers.

I also want to take this opportunity to encourage those of your listening or reading this podcast to [come meet our team at the 2011 Gartner Data Center Conference, Dec. 5-9, at Booth 49](#) and learn more. AccelOps is a silver sponsor of the conference.

**Gardner:** I am afraid we will have to leave it there. You've been listening to a sponsored BriefingsDirect podcast. We've been talking about how new data and analysis approaches from AccelOps are attaining significantly improved IT operations monitoring as well as stronger security.

I'd like to thank our guest. We have been here with Mahesh Kumar, Vice President of Marketing at AccelOps. Thank so much, Mahesh.

**Kumar:** Thank you, Dana.

**Gardner:** This is Dana Gardner, Principal Analyst at Interarbor Solutions. Thanks again for listening and come back next time.

**[Listen](#) to the [podcast](#). Find it on [iTunes/iPod](#). Download the transcript. Sponsor: [AccelOps](#)**

**Connect with AccelOps: [LinkedIn](#), [Twitter](#), [Facebook](#), [RSS](#).**

*Transcript of a BriefingsDirect podcast on how enterprises can benefit from capturing and analyzing real-time data to improve IT management. Copyright Interarbor Solutions, LLC, 2005-2011. All rights reserved.*

**You may also be interested in:**

- [A Technical Look at How Parallel Processing Brings Vast New Capabilities to Large-Scale Data Analysis](#)
- [Why Data and Information Management Remain Elusive After Decades of Deployments and How to Fix It](#)
- [Cloud and SaaS Force a Rethinking of Integration and Middleware as Services for Services](#)
- [Delivering Data Analytics Through Workday SaaS ERP Applications Empowers Business Managers at Actual Decision Points](#)