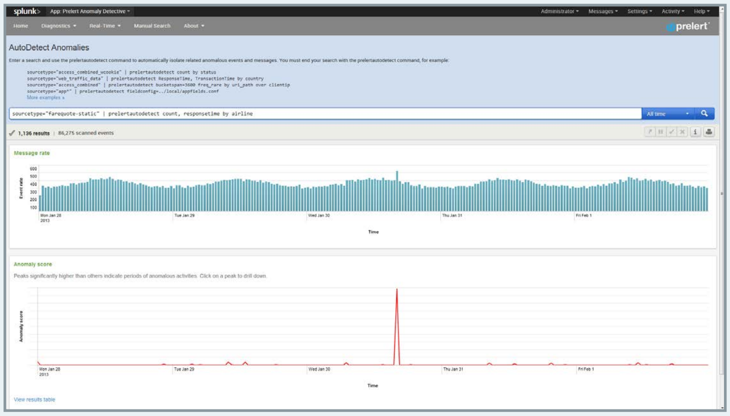
Why Choose Anomaly Detective

IT is a now a big data world. With so much data coming from so many different sources, as a modern IT, IT security, or DevOps professional, you likely have more data to drive your everyday decisions than was ever available before. On the other hand, unless you have the right tools to unlock the potential information it contains, all that data may just be a frustrating nuisance. Enter Anomaly Detective.

What is Anomaly Detective?

Anomaly Detective is a powerful machine learning analytics product that layers on top of data aggregation technologies like Splunk. Anomaly Detective automatically establishes models of normal behaviors it observes in huge volumes of data. It then uses highly accurate statistical analysis to identify anomalies in those behaviors and provides correlated results back to the user that explain the rarity, severity or impact of the outlier data. In IT environments, for instance, Anomaly Detective rapidly identifies the outlier behaviors that could indicate performance problems or security threats. Anomaly Detective makes the forensic analyses, troubleshooting and proactive monitoring efforts vastly more efficient by eliminating the time-consuming need for humans to mine the data through search or configure monitoring thresholds for alerts.



AutoDetect Anomalies

Enter a search and use the prelertautodetect command to automatically isolate related anomalous events and messages. You must end your search with the prelertautodetect command, for example:

```
sourcetype="access_combined_wcookie" | prelertautodetect count by status
sourcetype="web_traffic_data" | prelertautodetect ResponseTime, TransactionTime by country
sourcetype="access_combined" | prelertautodetect bucketspan=3600 freq_rare by uri_path over clientip
sourcetype="app*" | prelertautodetect fieldconfig=../local/appfields.conf
```

More examples »

```
sourcetype="farequote-static" | prelertautodetect count, responsetime by airline
```

✓ 1,136 results | 86,275 scanned events

Peaks significantly higher than others indicate periods of anomalous activities. Click on a peak to drill down.

Anomaly Detective's AutoDetect feature takes the response time data above (in blue) and maps out the anomalous behavior it detects below (in red).

WE LIVE IN A DATA-DRIVEN WORLD

Enterprises and government bodies around the world have come to understand that the vast amounts of data generated in the execution of their mission is a potential goldmine of information. This has driven the explosive growth of big data technologies that allow this data to be aggregated, indexed and mined through products like Splunk, Hadoop and Elastic Search. Mining this data through search and basic statistical analysis is limited in that it only allows us to find patterns or behaviors that we know to look for. More advanced analytics techniques, such as machine learning predictive analytics, are required to uncover the "unknown" behaviors that provide insights into yet unrealized business opportunities and risks.

IT organizations are also large generators of machine data. In this case the behaviors hidden in the data can describe business opportunities, application or service performance issues or security threats. And so, identifying behavior inconsistencies (or anomalies) in that data has become an important business function. But even the most advanced of the data aggregation and search technologies, like Splunk, are limited in their ability to uncover these anomalies. And those limitations can have a huge impact in environments that are large in scale, limited in resources, complex, or that are in a constant state of change.

- To perform an effective search, you have to know what you're searching for. In many environments, that may not always be obvious, especially when you're dealing with new issues as opposed to those you are already familiar with.
- In environments with too many moving parts setting thresholds for alerts just doesn't scale. And setting thresholds on a small portion of your environment to alert you when a major problem occurs will provide limited value in the troubleshooting or forensic analysis process.
- In environments that change rapidly, like DevOps, the high rate of new code implementations are much more likely to generate new problems that arise from mistakes or unforeseen consequences. New problems are the most difficult to spot or diagnose using standard processes.
- In all but the simplest environments, the ability to cross-correlate multiple data sources is a paramount need. It is only through cross-correlation that complex relationships can be uncovered that often hide the root cause of intermittent and chronic problems or severe outages. Cross-correlation is a difficult proposition without advanced analytics.

Increasingly, IT professionals are realizing that the methods and processes developed to monitor environments in the last century just don't scale to the complexity of today's IT shops. Anomaly Detective complements these environments to help them bridge that gap.

HOW DOES ANOMALY DETECTIVE WORK?

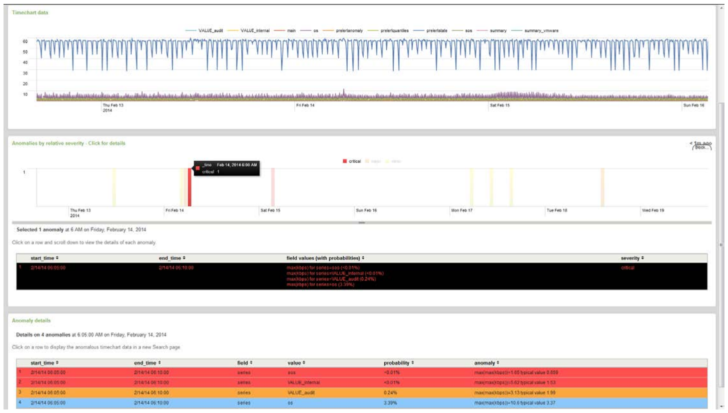
Anomaly Detective uses powerful machine learning analytics to self-learn the behaviors hidden in large data streams. It uses highly accurate statistical analysis to identify the anomalies that are behind performance problems or security threats. What does that mean? That means that, much like the human brain, Anomaly Detective can "learn" to recognize patterns and things that fall outside of those patterns.

But Anomaly Detective has an advantage over us. For over 50 years, it has been an **accepted principle of neuroscience** that the human brain is only able to make decisions based on somewhere between 5 and 9 factors at any one time. That makes it pretty difficult for us to fully comprehend the inner relationships, interactions and dependencies inherent in today's IT, Internet or security environments.

Computers don't have this limitation. Hence, a form of artificial intelligence known as machine learning has evolved to help us decode complex systems like the human genome, terrorist communications networks and modern IT environments. With machine learning, computers can find behavior patterns and anomalies in huge sets of data more accurately, more consistently and a lot faster than humans.

Machine learning is not new. It has been in use in various forms for decades. But previous implementations required significant human configuration, tuning and guidance and were only usable to uncover "known" behaviors or relationships in data. These earlier supervised forms of machine learning required humans to define models that pre-defined a foundation upon which the system was supposed to add what it learned from the data.

Some existing implementations of this form of machine learning often used in IT management require the user to define important relationships between things like hosts, VMs, OSes and networks. This requirement made supervised machine learning impractical in situations where the structure of the observed systems changes over time or the user cannot adequately define the relationships. Many early machine learning products for IT monitoring purposes, for instance, take months to implement and need remodeling every year or so.



Selected 1 anomaly at 6 AM on Friday, February 14, 2014

Click on a row and scroll down to view the details of each anomaly.

	start_time	end_time	field values (with probabilities)	severity

1	2/14/14 06:05:00	2/14/14 06:10:00	max(kbps) for series=sos (<0.01%)
			max(kbps) for series=VALUE_internal (<0.01%)
			max(kbps) for series=VALUE_audit (0.24%)
			max(kbps) for series=os (3.39%)
				critical

Anomaly details

Details on 4 anomalies at 6:05:00 AM on Friday, February 14, 2014

Click on a row to display the anomalous timechart data in a new Search page

	start_time	end_time	field	value	probability	anomaly
1	2/14/14 06:05:00	2/14/14 06:10:00	series	sos	<0.01%	max(max(kbps))=1.65 typical value 0.859
2	2/14/14 06:05:00	2/14/14 06:10:00	series	VALUE_internal	<0.01%	max(max(kbps))=5.62 typical value 1.53
3	2/14/14 06:05:00	2/14/14 06:10:00	series	VALUE_audit	0.24%	max(max(kbps))=3.13 typical value 1.99
4	2/14/14 06:05:00	2/14/14 06:10:00	series	os	3.39%	max(max(kbps))=10.6 typical value 3.37

Hover over an anomaly for more detailed information

With the challenges inherent in the big data pursuits of the 21st century, data scientists realized that many of the systems they were striving to understand, like the human genome, could not be analyzed with the current machine learning technologies. Simply put, they required humans to define things in the systems that they didn't sufficiently understand. This led to the development of unsupervised machine learning, the technology on which Anomaly Detective is built. This newer, cutting edge method gives computers the ability to recognize patterns in data—any data—without human-defined models. In the IT environment, for instance, unsupervised machine learning technologies can be implemented in minutes and require very low, if any, ongoing tuning.

In addition to unsupervised machine learning, Anomaly Detective leverages a form of predictive analytics to assure the accuracy of its results. By understanding behavior patterns, Anomaly Detective can get very good at predicting the likely range of values that should occur next in a stream of data. It calculates the probability of the actual value that occurs, and if the probability is significantly low, it will label that occurrence as an anomaly.

In Prelert's approach, anomalies occur when the event observed is sufficiently rare (say the occurrence of a new process on a machine), represents a severe change in state (like email rates per user that jump from 10 an hour to 30,000) or is correlated with a number of other anomalous events (like a change in response time that is related to an error in shopping cart software and a queue overflow condition in backend software). In the latter case, the correlated anomalies would be presented together as a super anomaly providing tremendous value in the diagnostic process.

Don't you need a supercomputer to handle this kind of complex analysis? No. Modern advancements in computer science have resulted in some very clever ways of giving commodity hardware the ability to do sophisticated analytics on large streams of data in real-time. From a time to value perspective, the Anomaly Detective app for Splunk Enterprise, for instance, can be downloaded, installed and provide answers in minutes. If you are ready to give it a try, you can **download a free trial version** right now. It's easy to install and it will be able to start looking for anomalies in your data in minutes.

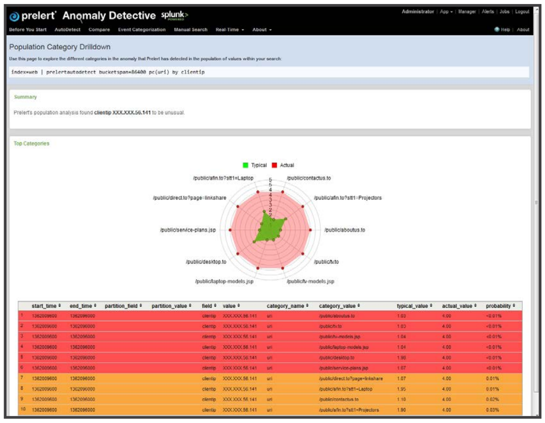
If you would like to learn more about what Anomaly Detective can do, read on for how Anomaly Detective brings benefits to three separate use cases:

- IT security
- IT operations/APM
- DevOps

ANOMALY DETECTIVE: THREE USE CASES

1. IT Security

IT security teams are more than familiar with the difficulty of finding unusual activities in large sets of data. The tools they use to keep their organizations' networks secure (intrusion detection and prevention systems, security appliances and the like) generate mountains of data. IT security teams realize that although their perimeter defenses are effective in blocking the vast majority of attacks, those that might get through are the very ones that can severely impact customers, profitability and reputation.


prelert Anomaly Detective splunk>
Administrator | App | Manager | Alerts | Jobs | Logout
Before You Start
AutoDetect
Compare
Event Categorization
Manual Search
Real-Time
About
Help | About

Population Category Drilldown

Use this page to explore the different categories in the anomaly that Prelert has detected in the population of values within your search:

```
index=web | prelertautodetect bucketspan=86400 pc(uri) by clientip
```

Summary

Prelert's population analysis found **clientip XXX.XXX.56.141** to be unusual.

Top Categories

Typical
Actual
/public/afin.to?stt1=Laptop
/public/contactus.to
/public/direct.to?page=linkshare
/public/afin.to?stt1=Projectors
/public/service-plans.jsp
/public/aboutus.to
/public/desktop.to
/public/tv.to
/public/laptop-models.jsp
/public/tv-models.jsp

	start_time	end_time	partition_field	partition_value	field	value	category_name	category_value	typical_value	actual_value	probability
1	1362009600	1362096000			clientip	XXX.XXX.56.141	uri	/public/aboutus.to	1.03	4.00	<0.01%
2	1362009600	1362096000			clientip	XXX.XXX.56.141	uri	/public/tv.to	1.03	4.00	<0.01%
3	1362009600	1362096000			clientip	XXX.XXX.56.141	uri	/public/tv-models.jsp	1.04	4.00	<0.01%
4	1362009600	1362096000			clientip	XXX.XXX.56.141	uri	/public/laptop-models.jsp	1.04	4.00	<0.01%
5	1362009600	1362096000			clientip	XXX.XXX.56.141	uri	/public/desktop.to	1.98	4.00	<0.01%
6	1362009600	1362096000			clientip	XXX.XXX.56.141	uri	/public/service-plans.jsp	1.07	4.00	<0.01%
7	1362009600	1362096000			clientip	XXX.XXX.56.141	uri	/public/direct.to?page=linkshare	1.07	4.00	0.01%
8	1362009600	1362096000			clientip	XXX.XXX.56.141	uri	/public/afin.to?stt1=Laptop	1.95	4.00	0.01%
9	1362009600	1362096000			clientip	XXX.XXX.56.141	uri	/public/contactus.to	1.10	4.00	0.02%
10	1362009600	1362096000			clientip	XXX.XXX.56.141	uri	/public/afin.to?stt1=Projectors	1.90	4.00	0.03%

Use Population Category Drilldown to check for anomalies in the behavior of system users

Besides storing this data for compliance reasons, most teams attempt to analyze it to detect successful attacks, rogue users or advanced threats. Unfortunately, the techniques used to analyze this data include:

- Forensic searches for known indicators that an attack has been successful sometimes weeks or months after the data has been logged; or
- The "college stats book" approach. Driven by the need to detect threats as soon as possible after they occur (or even before), some IT professionals try to go outside their area of specialty by training themselves to write increasingly complicated mathematical algorithms.

The problem with these approaches is they are often ineffective and are predominantly designed to catch breaches days, weeks or months after they have been successful. Further, relying on the search for "known" attack profiles and signatures does little toward detecting the advanced threats that arise as attackers refine their techniques. Many companies are looking for ways to respond to threats more effectively and nearer to real time. Anomaly Detective gives you that ability.

How Anomaly Detective Can Help IT Security Professionals

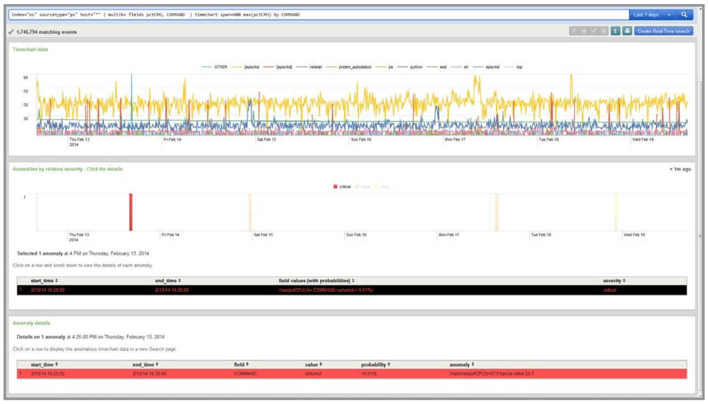
Because it can spot patterns and anomalies in and across multiple data streams much faster than human users, Anomaly Detective not only vastly speeds up threat detection to near real-time, but it gives IT security professionals the ability to identify threats they didn't even know to look for. In addition to known exploits, Anomaly Detective finds "zero-day" and advanced threats never encountered previously without the need for signatures or attack profiles.

Anomaly Detective provides real-time insights to the behaviors hidden in the huge volumes of security data you are already collecting.

Example: Anomaly Detective in Action

The IT security team of a government agency was struggling to leverage Blue Coat proxy data to find the telltale signs of advanced attacks. They wanted to find a way to cut through the noise without missing the important information that would allow them to identify real threats.

They fed the Blue Coat data into Anomaly Detective and it set to work right away defining normal communication patterns for users and looking for events that were outside of the norm. Without having to apply any rules to the data, they quickly found a very interesting case where one user hit a Microsoft IIS server with over 20,000 requests in an hour!



```
index="os" sourcetype="ps" host="*" | multikv fields pctCPU, COMMAND  | timechart span=600 max(pctCPU) by COMMAND
```

Selected 1 anomaly at 4 PM on Thursday, February 13, 2014

Click on a row and scroll down to view the details of each anomaly.

	start_time ⇕	end_time ⇕	field values (with probabilities) ⇕	severity ⇕
1	2/13/14 16:25:00	2/13/14 16:30:00	max(pctCPU) for COMMAND=splunkd (<0.01%)	critical

Anomaly details

Details on 1 anomaly at 4:25:00 PM on Thursday, February 13, 2014

Click on a row to display the anomalous timechart data in a new Search page.

	start_time ⇕	end_time ⇕	field ⇕	value ⇕	probability ⇕	anomaly ⇕
1	2/13/14 16:25:00	2/13/14 16:30:00	COMMAND	splunkd	<0.01%	max(max(pctCPU))=57.5 typical value 22.7

Take data from multiple sources (top chart) and cross-correlate it to see what systems were impacted together to cause such a big anomaly score

2. IT Operations/Application Performance Management (APM)

IT operations and application performance management (APM) professionals are responsible for monitoring large (and constantly changing) IT systems of networks, computers, servers, applications, middleware, databases, web servers, cloud servers…the list goes on. Most of these components generate a huge volume of data that describes their behavior and performance and the behavior of those who use them.

While this data can be extremely useful to IT operations teams, there are two problems with it:

1. The sheer scale of the environment makes traditional monitoring techniques cumbersome at best and impossible at worst. Some companies have upwards of 20,000 servers. That's a lot of data to sort through to understand what normal is and set thresholds and alert rules. Considering how many components they're working with, being just slightly off in their calculations can cause thousands of unwanted alerts.

2. Just because every component is working "normally" doesn't mean they're working normally together. IT problems often arise because of the interaction of different components, many of which, on their own, may be behaving properly. Seeing through hundreds of thousands of data points to find the complex chains of events that can lead from a small misconfiguration to a major outage is something the human brain isn't suited for. One way to solve this problem is to bring in more human brains—large teams of experts to analyze data and look for problems—but not every company has the resources for this approach.

How Anomaly Detective Can Help IT Operations/APM Professionals

With Anomaly Detective, IT operations/APM teams can skip the agonizing process of setting and maintaining effective thresholds altogether. Anomaly Detective allows them to perform real-time troubleshooting, discovering and responding to problems as they happen, before they can snowball out of control and significantly impede a company's operations. Anomaly Detective also gives IT operations teams insight into problems that occur across systems, through the interaction of otherwise normally functioning components.

Example: Anomaly Detective in Action

A loan authorization vendor for car dealers was using Splunk to monitor logs from Internet Information Services (IIS), internal applications, Windows Performance Monitor, and Windows events. A problem somewhere in a production application had resulted in unacceptably long wait times for dealer credit managers and bad loan rates. Even though the data had been aggregated with Splunk, it took a team of two developers and one operations support person 78 man hours over 12 days to identify and fix the problem.

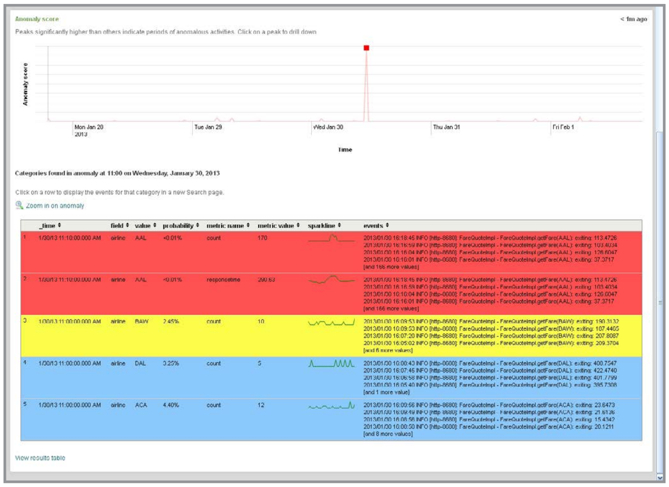
When the same team retroactively ran their Splunk data through Anomaly Detective, they discovered that not only would they have been alerted to the underlying issue earlier, but they would have been able to immediately find the causal application problem. It would have only taken two engineers three hours to find and fix the problem!

prelert®

3. DevOps Professionals

The priorities in DevOps are speed and flexibility. Because DevOps teams make changes to their products much more frequently than with the traditional software development approach, they can't rely on older data modeling methodologies (supervised machine learning, for example) to detect and respond to problems.

How can you set a baseline when you're deploying new code 30 times day?


Anomaly score
< 1m ago
Peaks significantly higher than others indicate periods of anomalous activities. Click on a peak to drill down
Anomaly score
Mon Jan 28 2013
Tue Jan 29
Wed Jan 30
Thu Jan 31
Fri Feb 1
Time

Categories found in anomaly at 11:00 on Wednesday, January 30, 2013

Click on a row to display the events for that category in a new Search page.

Zoom in on anomaly

	_time	field	value	probability	metric name	metric value	sparkline	events
1	1/30/13 11:10:00.000 AM	airline	AAL	<0.01%	count	170		2013/01/30 16:18:45 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(AAL): exiting: 113.4726 2013/01/30 16:16:59 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(AAL): exiting: 103.4034 2013/01/30 16:16:04 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(AAL): exiting: 126.6047 2013/01/30 16:16:01 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(AAL): exiting: 37.3717 [and 166 more values]
2	1/30/13 11:10:00.000 AM	airline	AAL	<0.01%	responsetime	290.63		2013/01/30 16:18:45 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(AAL): exiting: 113.4726 2013/01/30 16:16:59 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(AAL): exiting: 103.4034 2013/01/30 16:16:04 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(AAL): exiting: 126.6047 2013/01/30 16:16:01 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(AAL): exiting: 37.3717 [and 166 more values]
3	1/30/13 11:00:00.000 AM	airline	BAW	2.45%	count	10		2013/01/30 16:09:53 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(BAW): exiting: 190.3132 2013/01/30 16:09:53 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(BAW): exiting: 107.4465 2013/01/30 16:07:20 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(BAW): exiting: 207.8087 2013/01/30 16:05:02 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(BAW): exiting: 209.3704 [and 6 more values]
4	1/30/13 11:00:00.000 AM	airline	DAL	3.25%	count	5		2013/01/30 16:00:43 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(DAL): exiting: 400.7547 2013/01/30 16:07:45 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(DAL): exiting: 422.4740 2013/01/30 16:06:58 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(DAL): exiting: 401.7799 2013/01/30 16:05:40 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(DAL): exiting: 395.7308 [and 1 more value]
5	1/30/13 11:00:00.000 AM	airline	ACA	4.40%	count	12		2013/01/30 16:09:56 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(ACA): exiting: 23.6473 2013/01/30 16:09:49 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(ACA): exiting: 21.6136 2013/01/30 16:08:56 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(ACA): exiting: 15.4342 2013/01/30 16:00:50 INFO [http-8680]: FareQuoteImpl - FareQuoteImpl.getFare(ACA): exiting: 20.1211 [and 8 more values]

View results table

Click on an anomaly to drill down to show what events triggered it so you can find root cause faster.

How Anomaly Detective Can Help DevOps professionals

Anomaly Detective enables the continuous development mode that DevOps team favor by providing continuous monitoring of their data that constantly evolves to changes without requiring time intensive thresholds and rules. Because Anomaly Detective is "hands off," requiring little human involvement to define baselines and set thresholds, it allows them to focus on what they do best—development, not troubleshooting.

Example: Anomaly Detective in Action

A social platform developer employs only a small team of gifted engineers but is constantly on the forefront of innovation thanks to their commitment to the continuous deployment DevOps method. They sometimes go through 50 changes or more on production days and so they were finding the traditional method of monitoring design and management impractical. Anomaly Detective has become a key part of their process because it allows them to see how their frequent changes affect their data in real time, allowing them to adjust on the fly, rapidly improving their products to better serve their users.

NEXT STEPS: TAKE A TEST DRIVE

One of the most attractive features of Anomaly Detective is how simple it is to install and use. If you're intrigued by the information in this paper and have been experiencing the frustration of setting useful thresholds, performing accurate searches, and basically making sense of mountains of data, **download a free trial** for yourself and your company. This will give you the opportunity to witness firsthand how efficiently and intelligently Anomaly Detective works on even the most massive data sets.

ANOMALY DETECTIVE
Free Download