



red violet Reduces Compute Costs by 50% Using Amazon EC2 Spot Instances

Software and services company red violet significantly reduces compute costs for large parallel-processing clusters using Amazon EC2 Spot Instances. The company is an innovator in the data fusion space, using cloud computing, proprietary algorithms, and big data to resolve complex data problems. The company uses Amazon EC2 for data processing, relies on Amazon EC2 Spot Instances to keep costs manageable, and counts on Amazon EBS storage to ensure data can persist when Spot Instances are reclaimed.

Creating Trust in a Complex World

Data is everywhere, but that doesn't mean things are always clear. With trillions of transactions occurring every day in the United States alone, businesses need ways to identify who they are doing business with. Software and services company [red violet](#) uses proprietary algorithms, a data repository containing holistic profiles on more than 95 percent of the U.S. adult population, and modern data science techniques to provide real-time solutions to industries including financial services, insurance, law enforcement, government, and collections. The company's solutions enable identity authentication, due diligence, risk mitigation, legislative compliance, and asset recovery.

"We help solve what's known as the entity-resolution problem," says Angus Macnab, senior vice president of data science and engineering at red violet. "Simply put, if you're looking for information about a person named John Smith, there might be millions of records associated with that common name. It becomes a large probabilistic challenge to decipher individuals based on unstructured and even incomplete sets of identifying data. We use proprietary algorithms and cloud-based computing resources to positively identify individuals and businesses, uncover the relevance of disparate data points to identify linkages that are otherwise not obvious, help our clients meet legal requirements, increase efficiency, and drive results."

Competing in the Cloud

Prior to starting red violet, the company's principals built and sold companies focused on similar solutions using technology built using traditional data-center infrastructure. With red violet, they decided to develop a cloud-native version on Amazon Web Services (AWS) to deliver new levels of scalability and performance and massively reduce the company's capital expenditure.

This approach enables red violet to run large data-processing jobs on daily, weekly, and monthly cadences as various types of data are ingested into the system—without committing to a large infrastructure purchase. Achieving maximum cost efficiency without compromising on analytics speed and performance is critical to maintaining the company's competitive edge.

"Building the data sets our customers rely on requires an enormous amount of processing power," says Jeff Dell, chief information officer at red violet. "In the past, we built out traditional data centers with a thousand or more nodes to handle the compute load. The

Company: red violet
Industry: Big data and analytics
Country: United States
Employees: 100
Website: www.redviolet.com

About red violet

red violet is a software and services company specializing in big data analysis, providing cloud-based, mission-critical solutions to enterprises in a variety of industries. The company uses leading-edge, proprietary data fusion technology and a massive data repository to provide information about people, businesses, assets, and their interrelationships.

Benefits

- Cuts more than \$10 million upfront cost
- Reduces compute costs by 50–70%
- Frees budget for data science
- Runs hundreds of instances daily

AWS Services Used

- [Amazon Elastic Compute Cloud](#)
- [Amazon EC2 Spot Instances](#)
- [Amazon Elastic Block Store](#)

Spins up hundreds of large Amazon EC2 Spot Instances daily.



“Using Amazon EC2 Spot Instances frees up budget we can use to improve our data science algorithms and increase throughput.”

Dan MacLachlan, Chief Financial Officer, red violet

initial investment alone runs into the tens of millions of dollars—yet ultimately those machines were only running a few hours a day at the most, which is an enormous waste of money.” The AWS Cloud enables red violet to perform these kinds of processing-intensive tasks using [Amazon Elastic Compute Cloud](#) (Amazon EC2) instances, which incur costs only when they are running.

“Our competitors are running on old-school data-center architecture,” says Dan MacLachlan, chief financial officer at red violet. “Using AWS gives us the advantage of having the latest and greatest technology, and the capacity and power we need. We can spin up quickly when we have high data volumes. Within the industry, it makes us light years ahead of where competitors are, even those with a longer operating history.”

Efficiency Beyond the Basics

In its quest to provide the best services at competitive price points, the company sought to optimize its compute costs even further. Using [Amazon EC2 Spot Instances](#), red violet can access spare compute capacity available in the AWS Cloud at steep discounts. Using Amazon EC2 Spot Instances reduces red violet’s compute costs between 50 and 70 percent compared to using On-Demand instances. This empowers red violet to increase compute capacity without increasing its budget.

red violet’s analytical workloads run on clusters of many 16xlarge instances. These workloads are massively parallel and must run continuously. Amazon EC2 Spot Instances are based on excess capacity. In some circumstances, that capacity can be reclaimed by the system, which is why these instances are available at much lower cost than on-demand instances.

red violet’s workloads are sensitive to interruption, so it decided to implement an

additional layer that would enable it to take advantage of Amazon EC2 Spot Instances. The company uses a proprietary process to enable a technique known as checkpointing, in which the job state is saved periodically. “Using checkpointing, if a node goes down, we can restart from an interim step,” says Macnab. To enable this solution, the company uses [Amazon Elastic Block Store](#) (Amazon EBS), which can persist beyond the life of a compute instance, as opposed to instance-attached storage.

A Homegrown Solution

With extensive in-house development expertise, red violet created its own checkpointing solution from code. “We use C++ wherever possible because we want to be able to access the computer at a low level and realize maximum performance,” says Macnab. “The notice of a node going down is handled by our automation system. This system requests another node, brings the node into the placement group, attaches the Amazon EBS volumes, and the process continues where it left off.”

With this approach, the company can use Amazon EC2 Spot Instances even for its sensitive, highly parallel compute jobs. “We spin up hundreds of large instances per day using Amazon EC2 Spot Instances,” says Macnab.

This has resulted in significant savings for red violet, which translates to more processing power for the same budget. Or better data-fusion solutions than competitors while keeping costs low for customers. “Our clients’ businesses depend on getting data they can trust,” says MacLachlan. “Without it, they are wasting time and can even face legal and regulatory consequences for identifying the wrong individual and acting upon bad information. Using Amazon EC2 Spot Instances frees up budget we can use to improve our data-science algorithms and increase throughput. That gives our customers the most up-to-date information whenever they need it.”

To learn more, visit aws.amazon.com/ec2/spot.