

# Detect and Mitigate Fraud using Machine Learning with SparkML on EMR



partner  
network

## Customer Need

Blackhawk Network, a leader in gift and payment card offerings, needed to build a cloud-based Data Lake to run advanced analytics and machine-learning for fraud detection

## Cloudwick solution

Cloudwick team, data engineers and scientists, built data pipelines on AWS using data lake and advanced analytics. This included a data cleansing exercise followed by transformation and finally saving the data on Amazon S3. Cloudwick data scientist established a machine learning model for batch fraud detection, one of the 17 identified use cases that Blackhawk envisages working on. Blackhawk plans to extend the use of the Data Lake to drive settlements, merchant scoring, rebate forecasting, inventory forecasting and other uses.

For its fraud detection solution, Cloudwick team built data pipelines on AWS using data lake and Spark machine learning algorithms; historical data extracted from the previous five years of the company's on-premises Netezza data warehouse, and cleansed, transformed and saved to Amazon S3. Cloudwick engineers build a batch fraud detection use case using relevant features from a previously identified Random Forest (approach to classify a wide range of data) machine learning model and tested for accuracy of prediction using Spark ML on EMR because Spark is an efficient, scalable, and fault tolerant.

Once the model achieved accuracy, Cloudwick integrated it into the Spark ML pipeline, where it is being reused on different data sets for ongoing fraud detection. With the new solution, Blackhawk has replaced its complex, ineffective rule-based framework with an efficient machine learning algorithm that detects fraud quickly. In addition, the scalable and flexible cloud-based data pipelines are extremely cost effective. Finally, identifying and investigating fraud with machine learning translates into huge savings - both monetary and reputation - for Blackhawk.

## Results and Benefits

- Machine learning algorithms detect fraud quickly and replacing the complex, inefficient rule-based framework.
- Scalable, flexible cloud-based data pipelines are extremely cost effective.
- Loosely coupled data lake architecture allows reuse of pipelines with updated data.
- Identifying and investigating fraud with machine learning translates into huge savings - both monetary and reputation.

## About Blackhawk Networks



**Blackhawk Network delivers branded payment programs to help meet today's most challenging business objectives. Blackhawk Networks offer a full portfolio of comprehensive services to ensure the success of branded payment programs, including design, distribution, fulfillment and marketing. With a presence in over 26 countries, they execute programs worldwide.**

## About Cloudwick

Cloudwick is the leading provider of enterprise business and technology modernization services and solutions to the Global 1000 and helps enterprises gain competitive advantage from open source, data lake, big data, cloud and advanced analytics. Cloudwick is an advanced AWS partner with Big data and machine learning competency.

Cloudwick