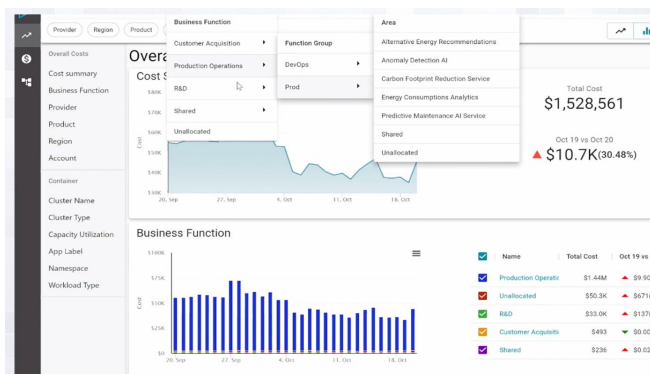


## Overview

A startup that provides an AI-powered cloud management platform for enterprises. The platform automates the process of tagging cloud resources, such as virtual machines, databases, and storage, to help enterprises manage their cloud infrastructure efficiently.



## CUSTOMER

**Anomaly Detection for Cloud Cost Management**  
 Designed and founded by engineers in 2015, optimizes the world's cloud computing spend, making cloud computing profitable and sustainable, for every organization. It creates cloud cost visibility and enables resource transparency by empowering engineering teams.

**Country:** USA

**Industry:** Private Sector

**Customer Size:** 10 - 50

**Publish Date:** 24/02/2023

## Problem Statement

The problem at hand revolves around cloud cost and usage visibility, forecasting, and anomaly detection. The first problem pertains to detecting anomalies in usage and cost, and learning usage patterns through the use of AI/ML to better forecast and understand the anomalies. Anomaly detection is crucial in ensuring that any irregularities in usage or cost are immediately identified and addressed. In this regard, sub-problem root cause detection becomes equally important to understand the cause of such anomalies, thereby enabling the team to take necessary corrective measures. The second problem is related to visibility and attribution, specifically identifying the teams within the organization that are incurring how much cost. This requires automated tagging for attribution through the use of AI/ML, which is essential in gaining visibility and transparency into the overall cloud cost structure. It helps in identifying cost drivers and providing cost accountability, thereby enabling the team to optimize cloud usage while maintaining budget control. The key challenges are centered around detecting anomalies in usage and cost, understanding usage patterns, identifying root causes, and gaining visibility into the overall cost structure. Solving these problems would require advanced technologies like AI/ML, which can be used to predict and mitigate potential issues, leading to efficient cloud cost management and optimization.

## Technical Solution

Combination of AI/ML technologies to solve the challenge. For time series forecasting, the platform leverages LSTM neural networks that can learn the patterns in the cloud usage data and predict the future usage. To detect anomalies in real-time, the platform uses the Isolation Forest algorithm that can identify the unusual usage patterns and alert the administrators.

For automatic tag generation, this company uses NLP techniques, such as Word2Vec and NER, to extract the relevant keywords and entities from the cloud resource names and descriptions. The platform also leverages machine learning models to suggest tags based on the past usage patterns and the existing tags.

To handle the large volume of data, they use big data processing technologies, such as Apache Hadoop, Apache Spark, and Apache Kafka. These technologies help the platform to store, process, and analyze the data at scale. Leverages cloud infrastructure, such as AWS, Azure, and GCP, to deploy the platform and integrate with the cloud providers' APIs.

Technologies	Domain
Python, Azure Cloud, AWS EMR, AWS S3, AWS SQS, Scale AI, Flask, Docker, ClickHouse, PostgreSQL, MySQL, Microsoft SQL Server, Kubeflow, Kubernetes, PySpark (Apache Spark)	Data Science, Big Data, Anomaly Detection, Batch Processing, ML Ops

## Results

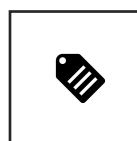
AI-powered cloud management platform has helped enterprises to reduce the manual effort and cost of managing their cloud infrastructure. The platform has achieved the following results:

- 1. Improved tagging accuracy by 80%:** The platform's automatic tag generation has improved the tagging accuracy by 80%, resulting in better visibility and control over the cloud resources.
- 2. Real-time anomaly detection:** The platform's real-time anomaly detection has helped the enterprises to identify and fix the issues before they impact the business.
- 3. Efficient resource utilization:** The platform's time series forecasting has helped the enterprises to optimize their resource utilization and reduce the cloud cost.



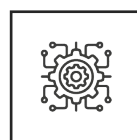
### Source Data Step

Load, Combine and prepare lense allocation and cost tables



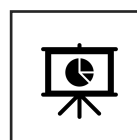
### Retagging Detection

Determine retagging cost across the lense



### Anomaly Model

Calculate thresholds  
Find minimum cost difference  
Label anomalies



### Root Cause Analysis

Identify top 3 contributing factors  
Validate