Pump

Description:

It's an AI based cost management tool for AWS Cloud Services. It can be used to save cost on different AWS services and it's free to use.

Working:

It works by collecting AWS usage data from various sources, such as CloudWatch Logs, CloudTrail Logs, and the CUR API. This data is then processed and analyzed to identify areas where cost can be minimized. It can identify instances that are not being used to their full potential and recommend that you resize them to a smaller size. It can also identify unused resources that are contributing to the cost such as EBS Volumes and Snapshots. It tracks usage patterns for different services so that it can suggest in future if you want to upscale that particular resource. All the recommendations can be applied automatically using Pump which in the end utilizes AWS CloudFormation service. Here are the supported AWS Services by Pump.

AWS Services	Before ($/mo.)	After ($/mo.)
EC2 (Linux/Unix)	$100	$41
Sagemaker	$100	$42
RedShift	$100	$44
EC2 Data Transfer*	$100	$44
ECS	$100	$51
Lambda	$100	$51
EBS	$100	$51
ElasticCache	$100	$52
OpenSearch	$100	$52
RDS	$100	$58
S3*	$100	$91
MediaLive	$100	$41

Pricing and Integration:

It is free of cost and can be currently integrated with AWS Cloud only.

Cast AI

Description:

It is also an AI based cost optimization tool with a free trial. It is used for optimizing Kubernetes workloads on cloud services. It automatically scales instances up and down based on workload demands, preventing over-provisioning and reducing costs. It also provides accurate cost forecasting for future periods. It can also be used for root cause analysis of increasing cloud costs.

Working:

It employs ML models to predict future workload demands based on historical data and current trends. This predictive capability enables proactive resource allocation, ensuring that resources are available when needed and preventing over-provisioning when demand is low. The structure of Cast AI is as follows;

- Ingestion Layer:

This layer collects and aggregates data from various sources, including Kubernetes clusters, cloud providers, and external monitoring tools. It processes and normalizes the data into a consistent format for analysis.

- Analysis Layer:

This layer employs a suite of ML algorithms to analyze the ingested data, extracting insights and patterns related to workload patterns, resource utilization, cost trends, and anomalies.

- Action Layer:

This layer translates the insights and recommendations generated by the analysis layer into actionable commands and configurations. It interacts with cloud providers and Kubernetes clusters to implement resource optimization, cost reduction and automation tasks.

Pricing and Integration:

It is integrated with Kubernetes clusters to gather information about pods, deployments, containers, and resource utilization. It can also be integrated with external monitoring tools like Prometheus, Grafana and New Relic to gather additional data for comprehensive analysis.

Monitoring and Cost Insights are free of cost. If you want to automate the scaling of CPUs then the base cost is $200 USD with $5 per CPU up to 500 CPUs per month and $1000 USD with $5 USD per CPU up to 2000 CPU per month.

References:

https://www.tryfondo.com/blog/pump-launches

https://medium.com/cast-ai