Public Spot Instance Dataset Archive Service

Kyunghwan Kim*, Subin Park*, Jail Hwang, Hyeonyoung Lee, Seokhyeon Kang, Kyungyong Lee

{bryan9801, subean, jaeil, gusdud0122, kh3654p, leeky}@kookmin.ac.kr

ARTIFACT AVAILABLE

What Is Spot Instance Dataset

Spot instances use spare computing capacity.

Provide computing resources at lower price.

It cloud be interrupted or not-fulfilled.

Cloud Vendors provide their own dataset

- Spot Price Dataset (AWS, Azure, GCP)
- Spot Interruption Dataset (AWS, Azure)
- Spot Availability Dataset (AWS)

Difficulty of Collecting Dataset

Accessing various spot datasets is challenging.

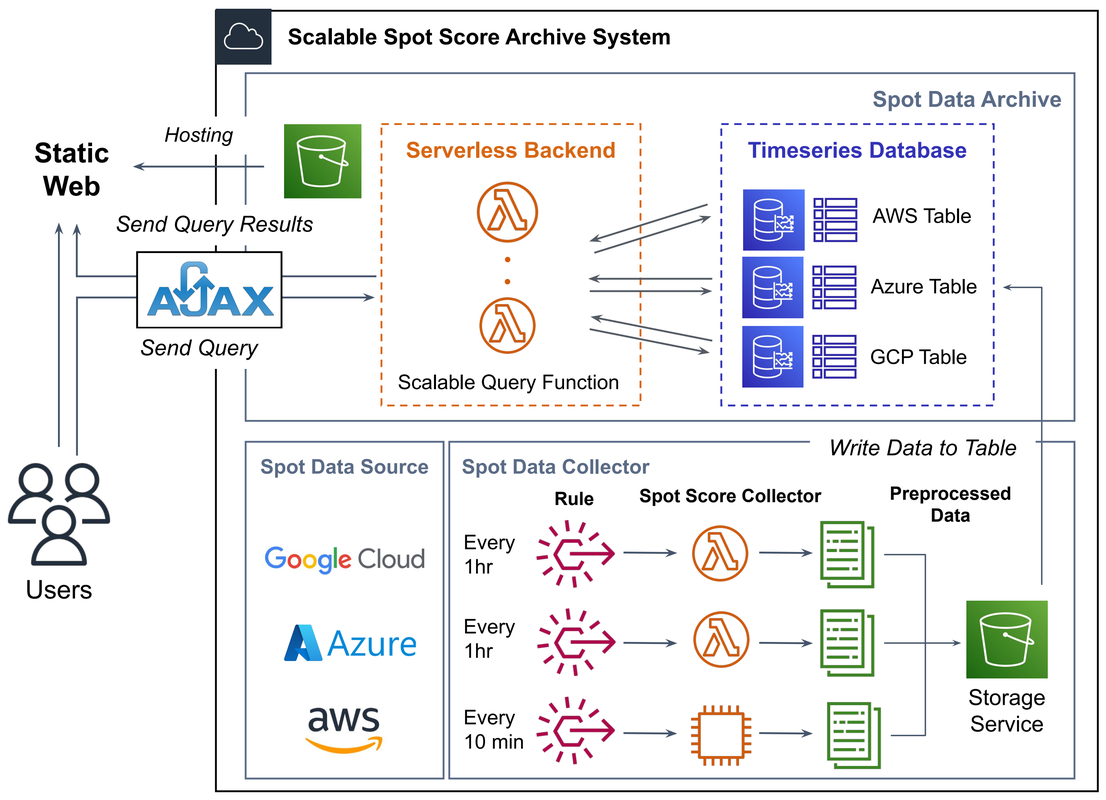
Cloud vendors provide their own information in

- distinct locations
  - AWS spot advisor, Azure price API …
- different access mediums
  - programmatic access, web interface …

Our Goal

- Collect spot dataset
- Empirically analyze dataset
- Make public web service

Spot Data Collector Architecture



[The architecture of data collector implementation]

Collecting dataset from AWS, Azure, and GCP.

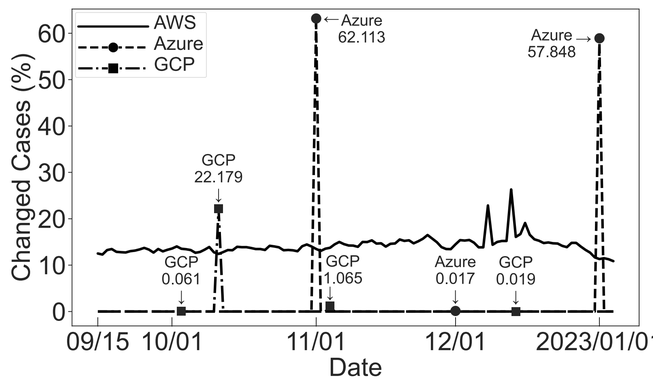
Implementation using a serverless architecture.

Spot Instance Dataset

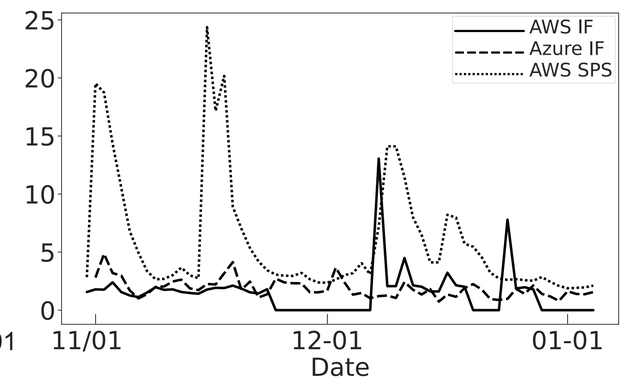
Spot Savings(Savings)	spot savings % compared to on-demand
Interruption Score(IFS)	spot interruption rate score in the past month
Availability Score(SPS)	the probability score of successful spot request

[Collected Dataset features]

Changed Frequency of Spot Instance Datasets



[Savings changed frequency]



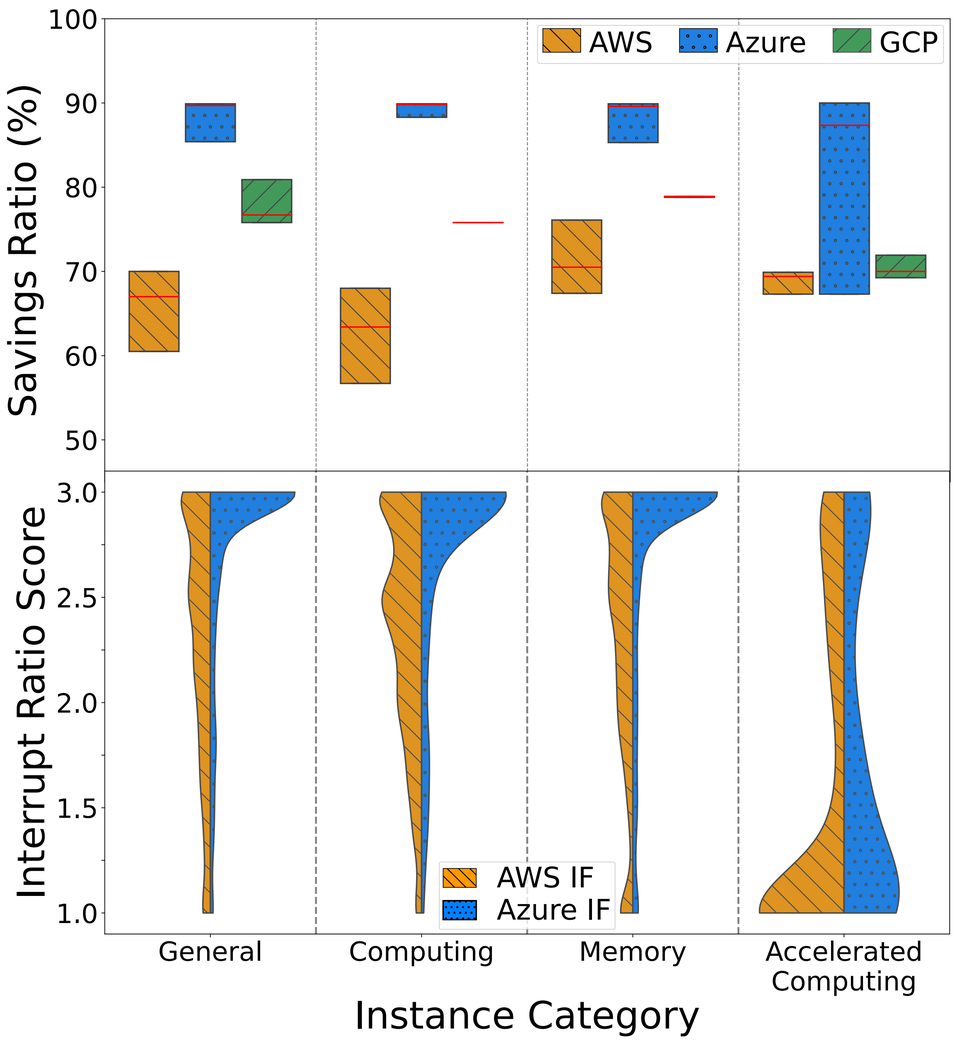
[IFS-SPS Score changed frequency]

Savings : Regular changes of AWS, not Azure and GCP.

IFS: Regular change of Azure, not AWS.

SPS : Regular change pattern, occasionally high change.

Value Distribution of Spot Instance Datasets



[Savings ratio & Interruption score distribution]

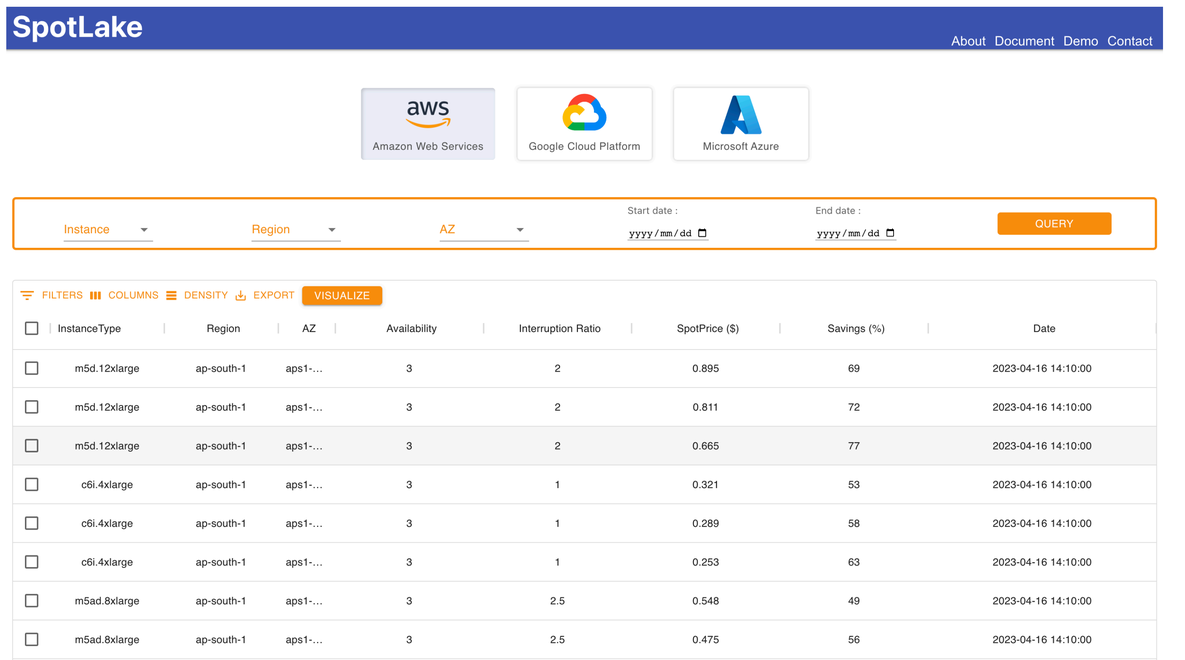
Savings ratio in the order of Azure > GCP > AWS.

Higher interruption-free score in Azure than AWS.

Spot instances in Accelerated Computing in distinct pattern.

This indicates price is not a reliable source for interruptions.

Spot Instance Dataset Archive Service Online



Public Web Service Link

https://spotlake.ddps.cloud

KOOKMIN UNIVERSITY

Distributed Data Processing System Lab