Performance benchmarking of cloud services

Final Project Proposal for Cloud Computing

Longjie Su 225040158

I. Research Background and Motivation

Cloud computing has evolved into a foundational infrastructure for contemporary information systems. With the large-scale migration of enterprises and research institutions to cloud environments, the performance stability and cost-effectiveness of computing resources, storage systems, and networking services have emerged as critical factors influencing system architecture design and resource provisioning decisions. Public cloud platforms represented by Amazon Web Services (AWS) offer diverse instance types and service models (e.g., Elastic Compute Cloud, Elastic Block Store, Lambda). However, their actual performance is subject to the compound influence of multiple factors, including underlying hardware architectures, network topologies, resource virtualization levels, and service configuration parameters.

Although cloud service providers publish standardized performance specifications, these technical metrics are typically derived from idealized testing conditions and fail to accurately reflect performance characteristics in real-world production environments. Empirical studies indicate that even homogeneous instance types may exhibit significant performance heterogeneity in CPU processing capacity, storage access latency, and network transmission bandwidth across different geographic regions, temporal windows, and workload patterns. Concurrently, the widespread adoption of Serverless computing paradigms (e.g., AWS Lambda) introduces new dimensions of performance uncertainty, including cold start latency and response time fluctuations during concurrent scaling.

Given this context, conducting systematic and reproducible performance benchmarking of cloud services holds substantial practical and academic significance:

1. Quantitatively characterizing performance differentials across service types and configuration schemes to provide empirical evidence for instance selection and architectural design;
2. Revealing dynamic patterns of performance metrics in response to temporal evolution and workload variations, thereby establishing theoretical foundations for deployment strategy optimization;
3. Evaluating performance-to-cost ratios to inform economically optimal cloud resource allocation decisions.

This study proposes to conduct performance benchmarking experiments on multiple computing and storage services using the AWS cloud platform, leveraging an experimental budget allocation of $50 USD (with $45 currently remaining).

II. Research Questions and Objective Formulation

This study systematically investigates performance characterization and cost-effectiveness evaluation of cloud services, with core research questions focusing on three dimensions: first, exploring the differentiated characteristics of various EC2 instance types in terms of CPU computational capacity and memory performance, and examining whether these performance differentials correspond rationally to their pricing strategies; second, conducting comparative analysis of different types of Elastic Block Store (EBS) volumes and local instance storage in terms of critical performance metrics such as I/O throughput and access latency; and third, investigating the patterns of variation in cold start and warm start performance of AWS Lambda functions under different memory configuration schemes and concurrent invocation patterns.

Based on these research questions, this project establishes the following research objectives:

1. To construct a reproducible and extensible performance benchmarking framework for cloud services and implement multi-dimensional performance evaluation experiments targeting computing, storage, and serverless architectures;
2. To systematically collect and analyze key performance indicators across multiple dimensions, including throughput, latency, performance variance, and cost-effectiveness;
3. To establish quantitative performance-to-cost comparative analysis models that provide empirically-grounded guidance for cloud service selection decisions across diverse application scenarios.

III. Technical Approach and Implementation Plan Overview

This study employs a suite of publicly available and extensively validated open-source tools for performance testing and data analysis. The specific technical stack includes: sysbench, fio, iperf3, and Python-based automated testing scripts for systematic assessment of AWS Lambda cold start and concurrent performance. The data processing and visualization workflow utilizes pandas for data cleaning and statistical analysis, and matplotlib for generating visualization charts, with experimental results presented in graphical and tabular formats.

The research implementation is structured into three phases: Phase I focuses on experimental environment preparation, encompassing AWS environment configuration, permission management, and deployment of testing instances and tools; Phase II involves performance testing execution, systematically conducting multi-dimensional performance evaluation experiments and recording results; Phase III encompasses data analysis and report composition, performing statistical analysis and visualization of experimental data to formulate research conclusions and technical reports. The project will initially conduct a feasibility assessment to validate the technical viability and resource requirements of this implementation plan.

To ensure research transparency and result reproducibility, all experimental scripts, configuration files, and raw data will be centrally maintained in a GitHub repository.