

Response Time Service Level Agreements for Cloud-hosted Web Applications



Hiranya Jayathilaka, Chandra Krintz and Rich Wolski
University of California, Santa Barbara



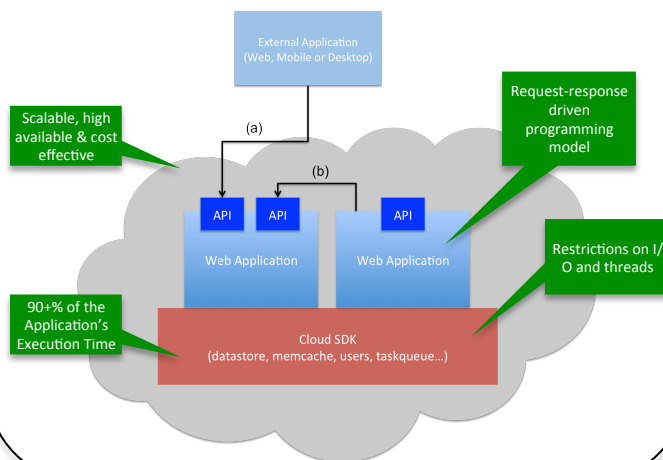
Motivation

- Many applications rely on the web APIs deployed in the cloud.
- But modern cloud platforms do not provide any guarantees regarding the performance of deployed web APIs and applications.
- There is a lack of tools for reasoning about the performance of cloud-hosted applications.

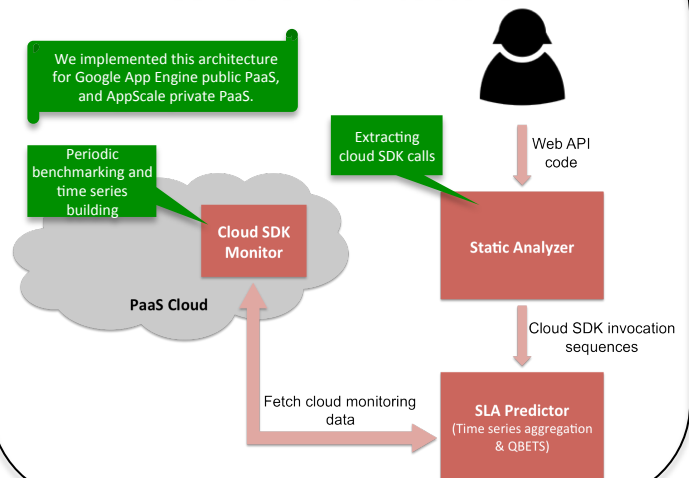
Our Solution: Cerebro

- Computes statistical bounds on the response time performance of web applications/APIs deployed in PaaS clouds.
- Combines static analysis of PaaS applications with runtime monitoring of the cloud platform to automatically predict performance SLAs for cloud-hosted web applications.

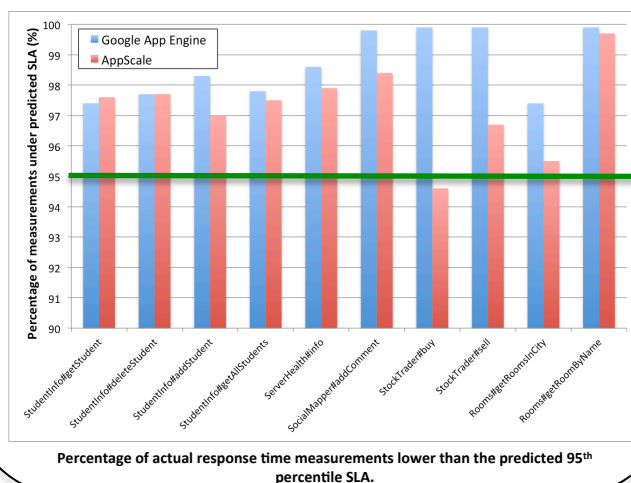
PaaS-hosted Applications



Cerebro Architecture



Prediction Accuracy



Prediction Tightness

