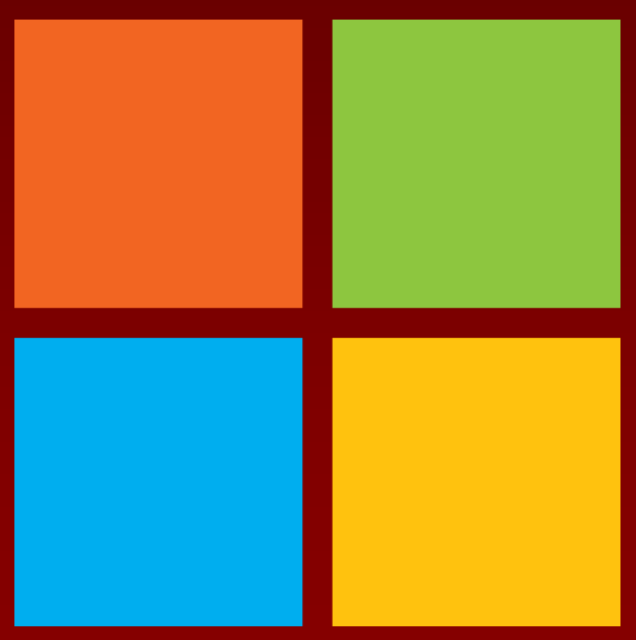


Computation Reuse in Analytics Job Service at Microsoft

Alekh Jindal, Shi Qiao, Hiren Patel, Zhicheng Yin, Jieming Di, Malay Bag,
Marc Friedman, Yifung Lin, Konstantinos Karanasos, Sriram Rao

Microsoft



Motivation

Job Service/ Serverless Analytics/ Analytics-as-a-Service:

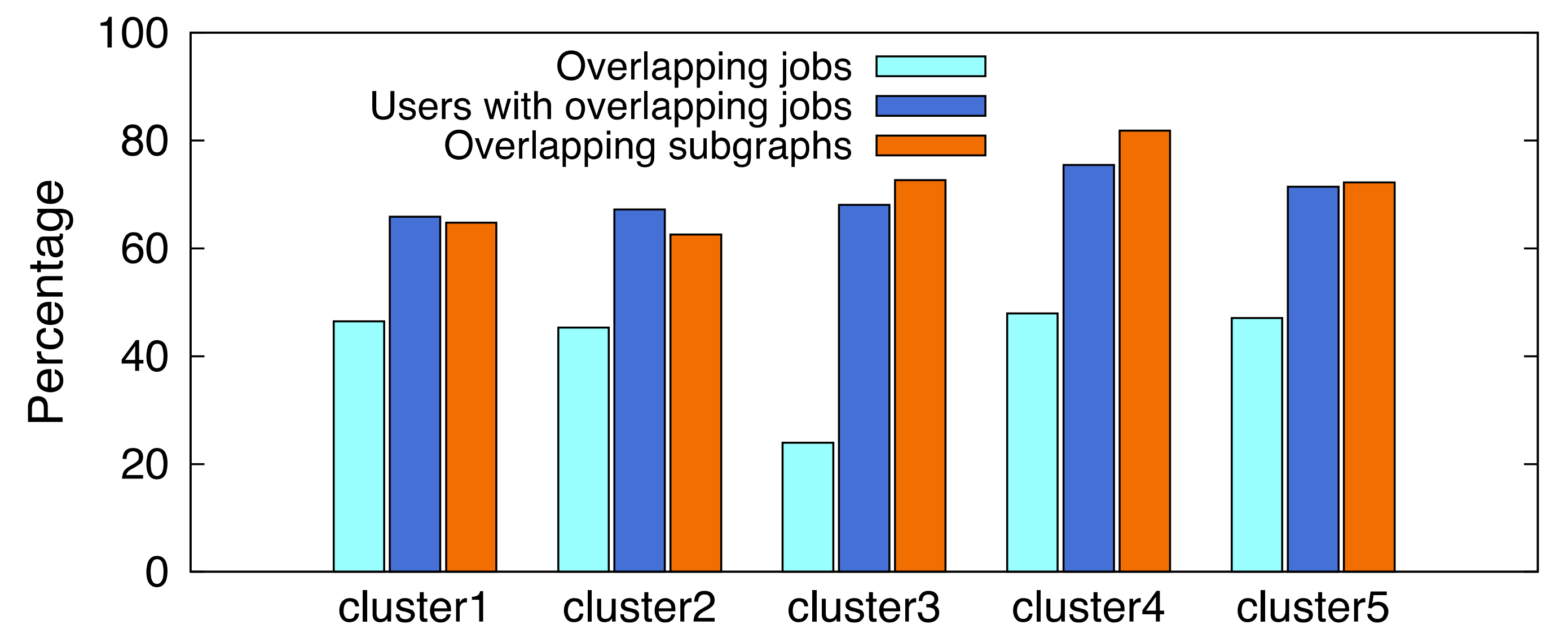
- User Only provide SQL queries over stored data
- Cloud provider takes care of manage h/w, s/w, and execution
- Users only pay for the processing cost

SCOPE Job Service

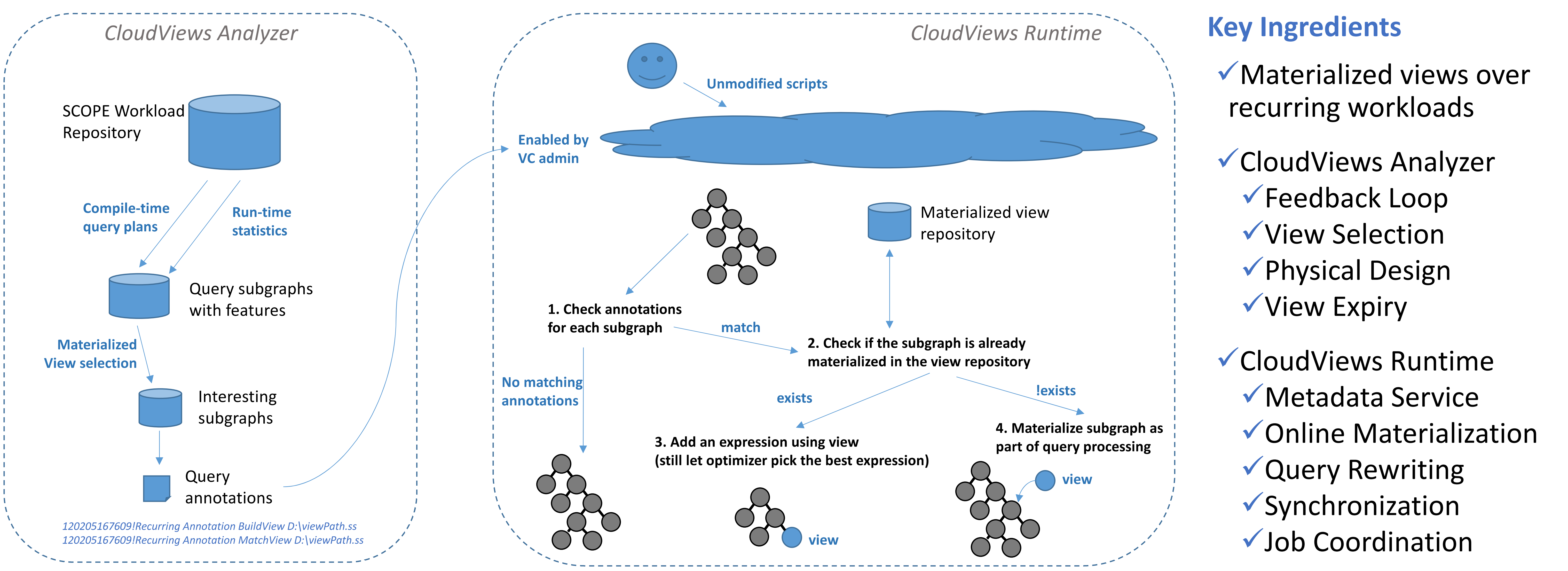
- Hyper scale data processing system for internal data analytics at Microsoft

Computation Reuse Challenge

- Recurring workloads with new inputs/parameters
- Always online with SLA requirements
- Cost estimations very challenging

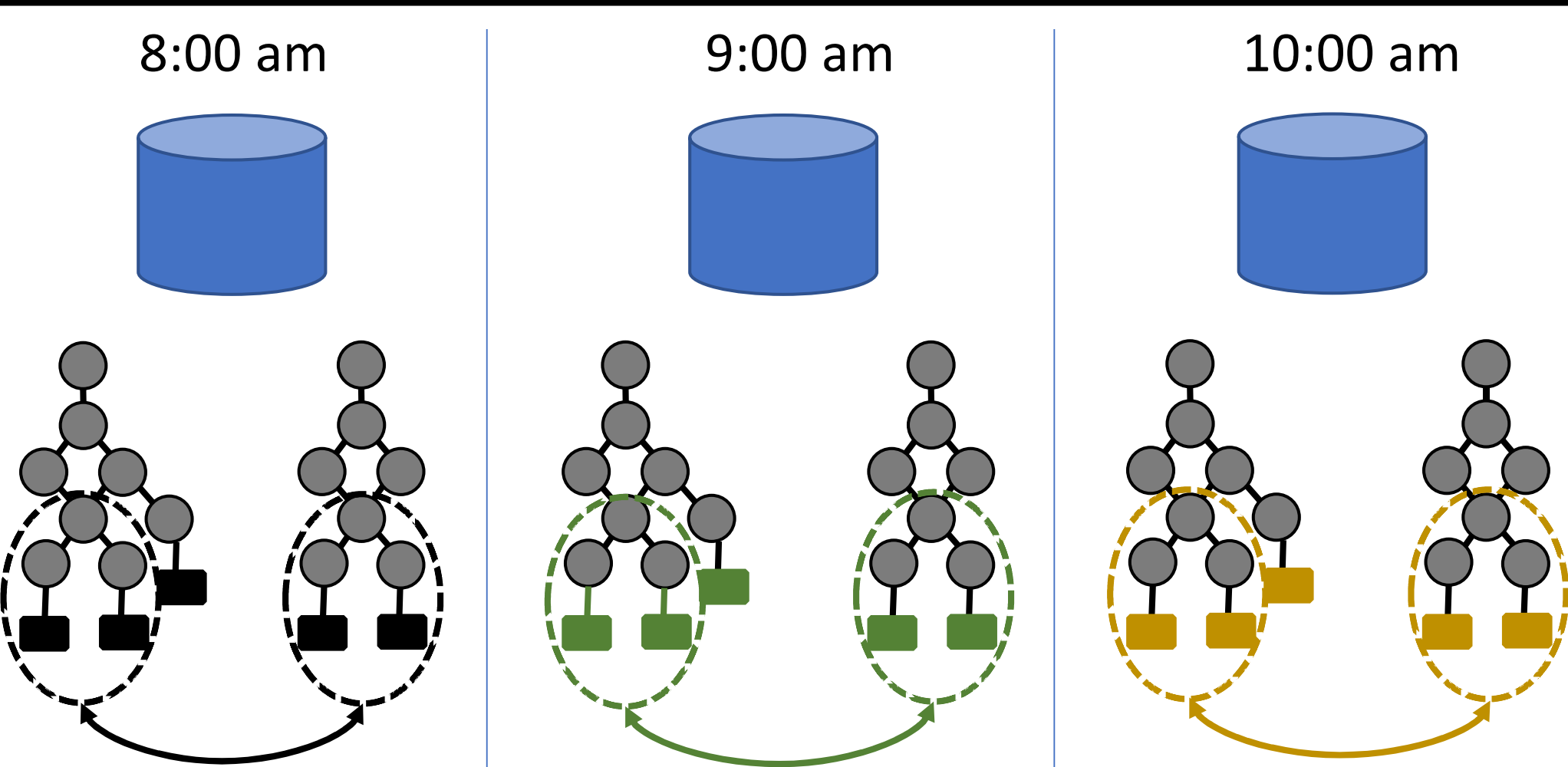


CloudViews Architecture

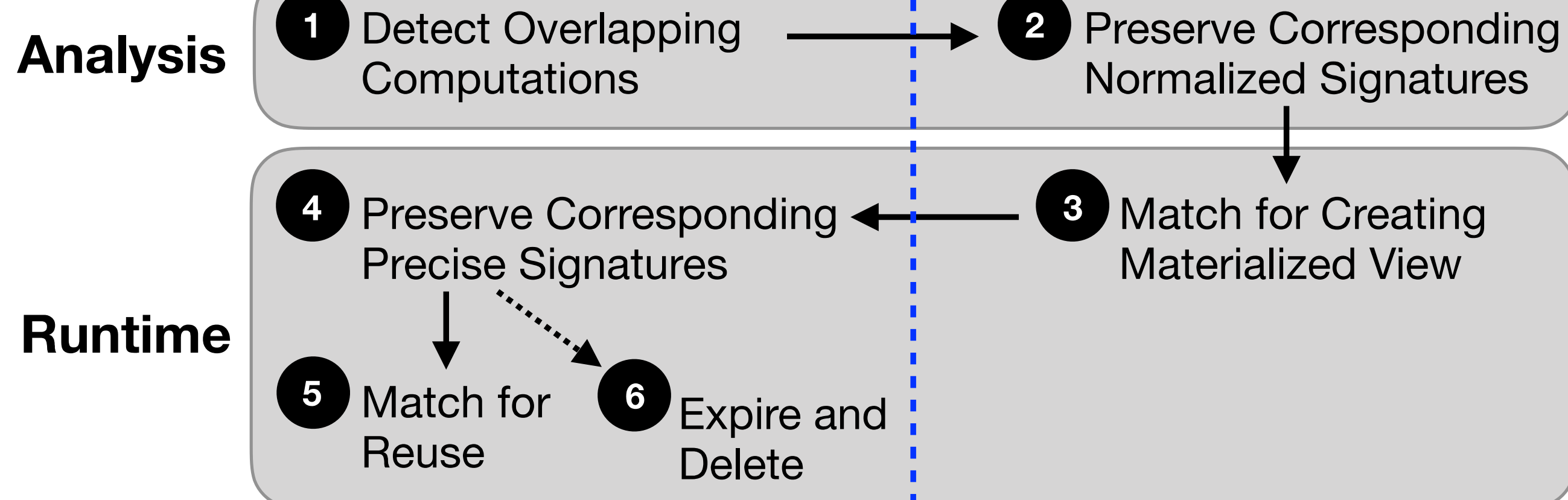


Reuse over Recurring Workloads

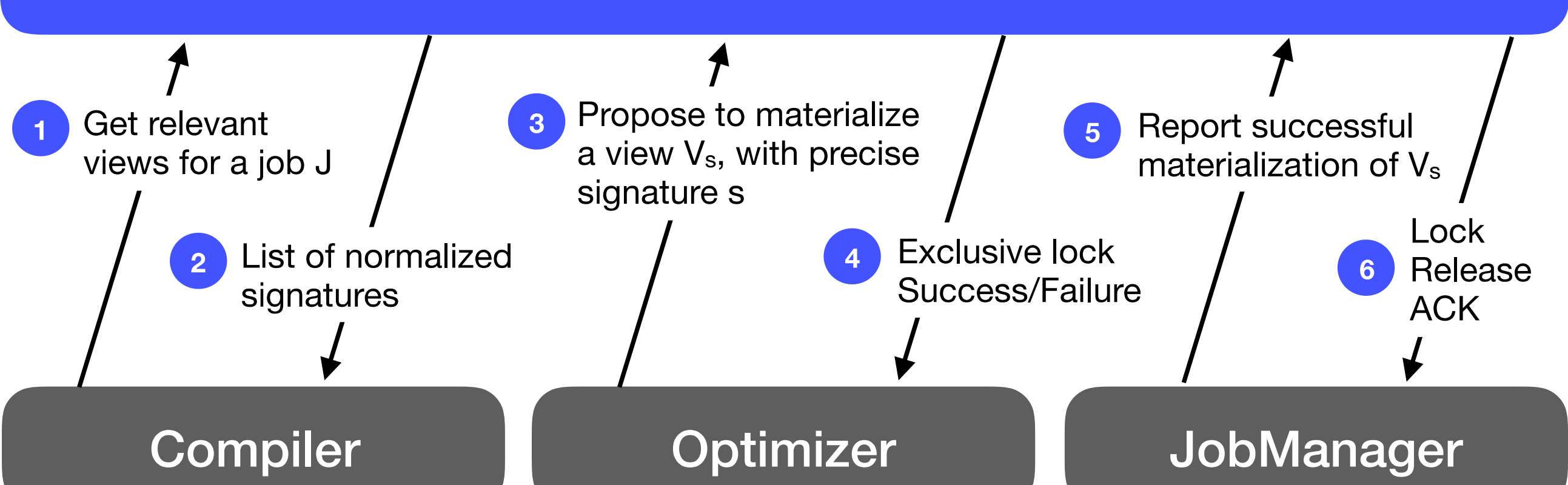
- Periodic jobs
- New inputs/parameters
- Different data formats
- Lots of user code



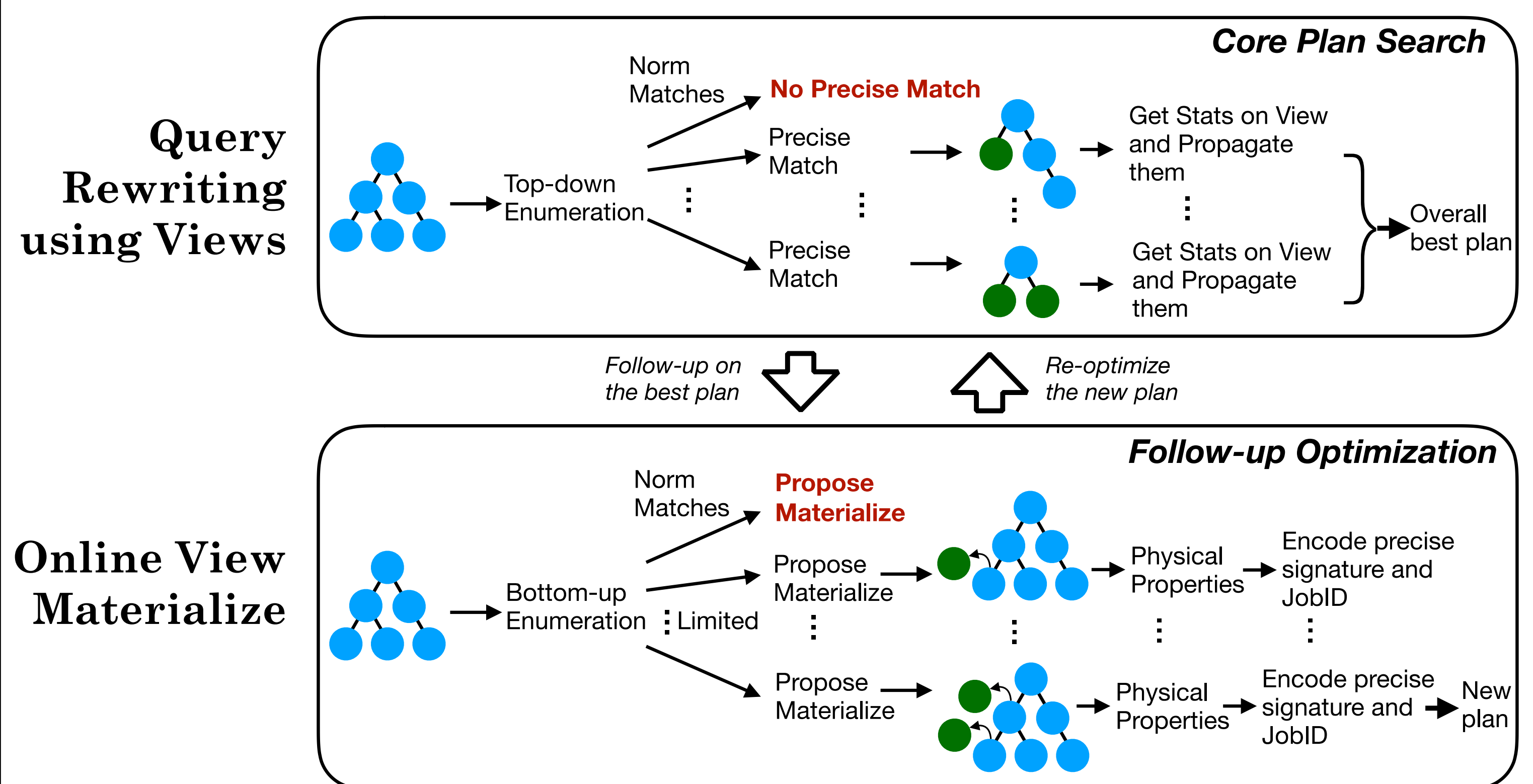
Precise Signature Normalized Signature



CloudViews Metadata Service

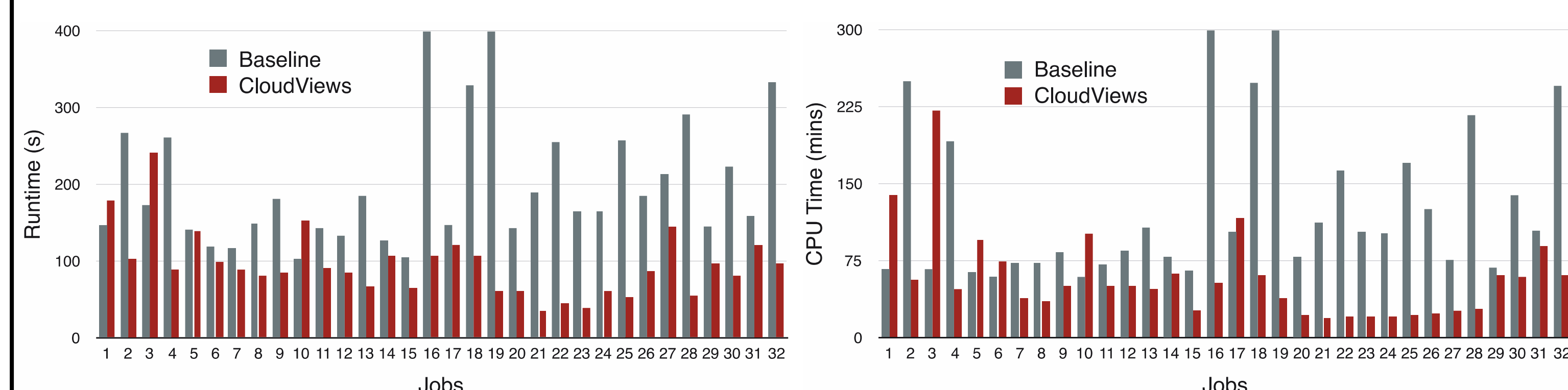


Query Optimization



Performance Impact

Improved Latency and Processing time.



★ We Are Hiring! ★