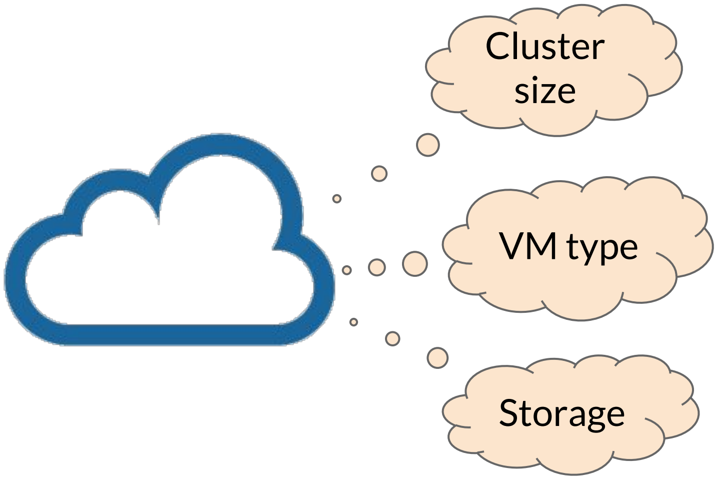
Selecta: Heterogeneous Cloud Storage Configuration for Data Analytics

Ana Klimovic[*], Heiner Litz[+], Christos Kozyrakis[*]

[*] Stanford University

[+] University of California Santa Cruz

Configuring analytics in the cloud



# of VMs in cluster?

CPU cores, GB of DRAM, network bandwidth, accelerators?

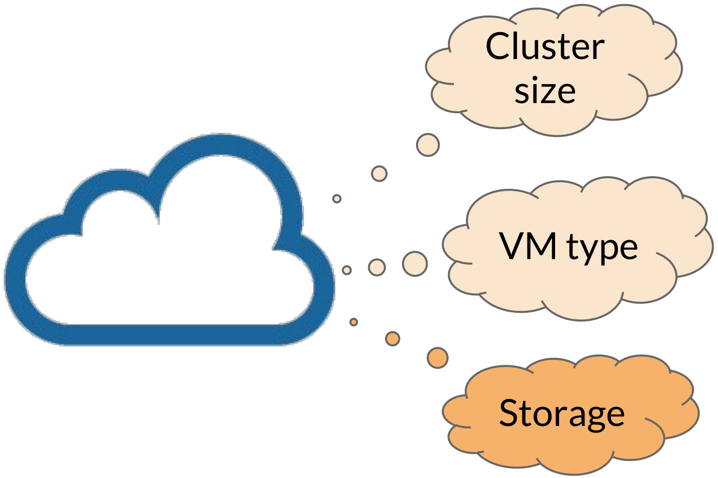
Block, file, object, key-value storage?

Directly attached to VM or remote?

Storage media: HDD, Flash, DRAM?

Cloud cluster configuration is difficult yet critical for performance & cost.

Configuring analytics in the cloud



# of VMs in cluster?

CPU cores, GB of DRAM, network bandwidth, accelerators?

Block, file, object, key-value storage?

Directly attached to VM or remote?

Storage media: HDD, Flash, DRAM?

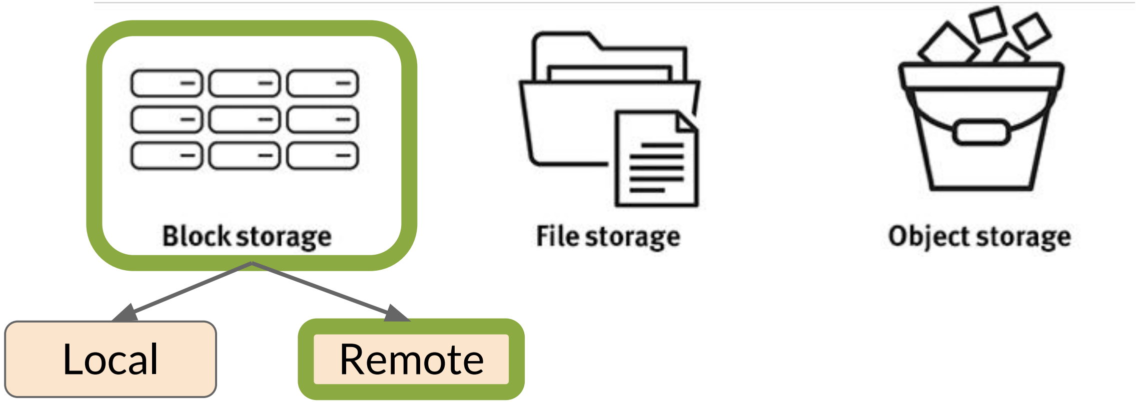
Cloud cluster configuration is difficult yet critical for performance & cost.

Configuring storage for analytics

- Storage configuration is particularly critical for data analytics

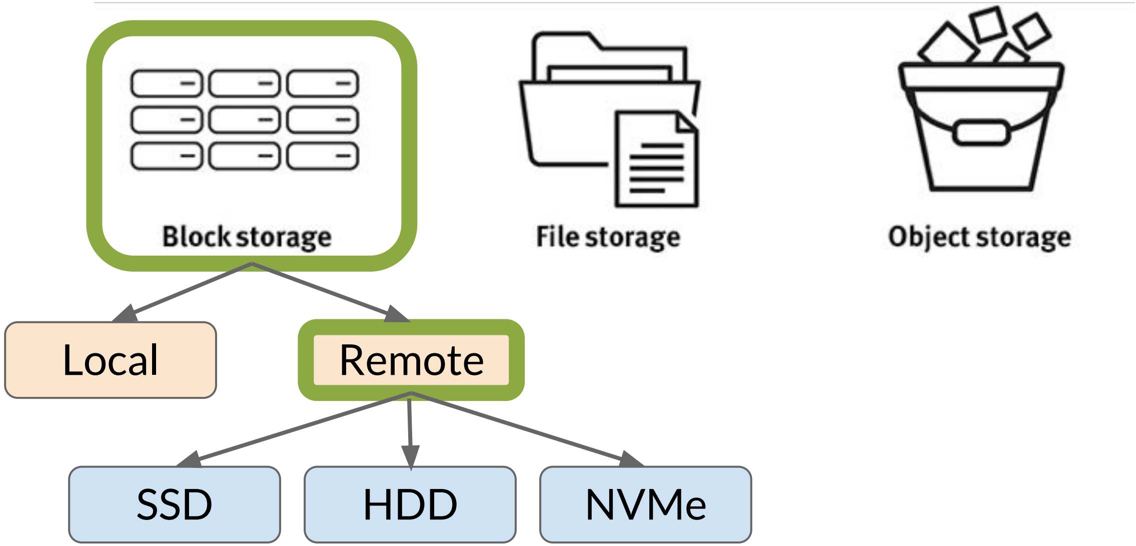
Configuring storage for analytics

- Storage configuration is particularly critical for data analytics



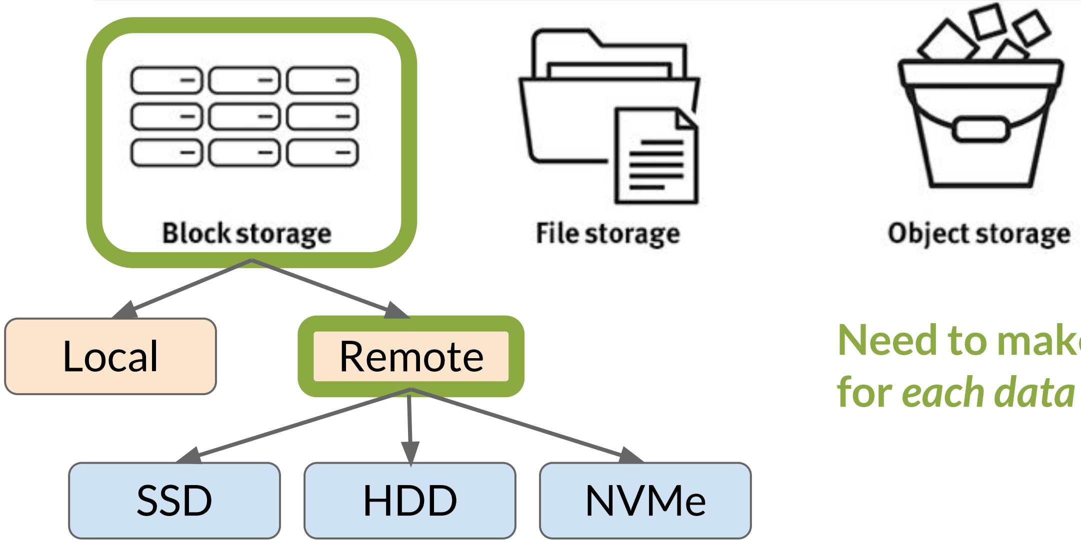
Configuring storage for analytics

- Storage configuration is particularly critical for data analytics



Configuring storage for analytics

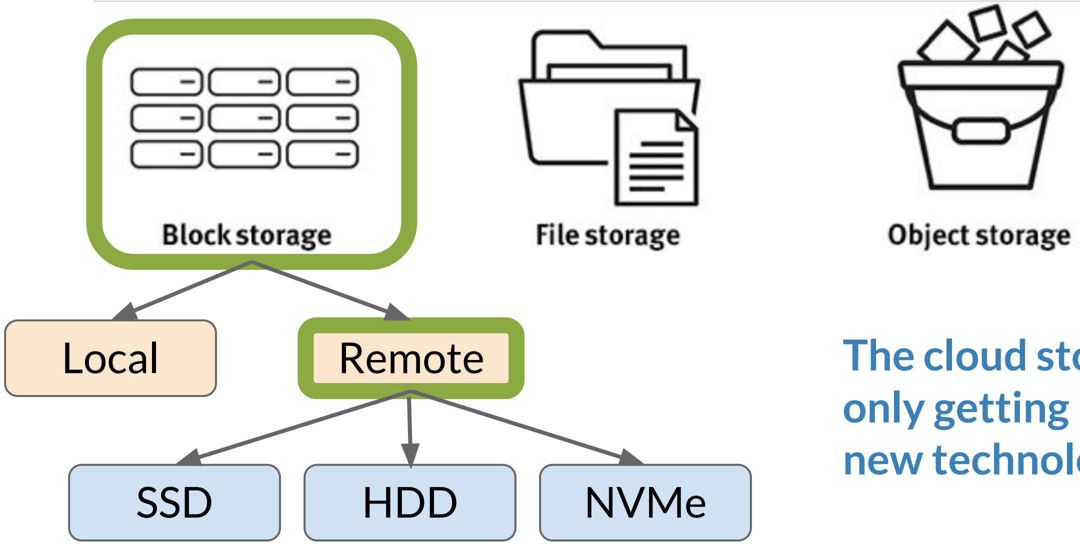
- Storage configuration is particularly critical for data analytics
- Jobs often have multiple data streams (e.g., shuffle, input/output data) with diverse I/O characteristics, making them suitable for different storage options



Need to make storage decisions for *each data stream* in a job

Configuring storage for analytics

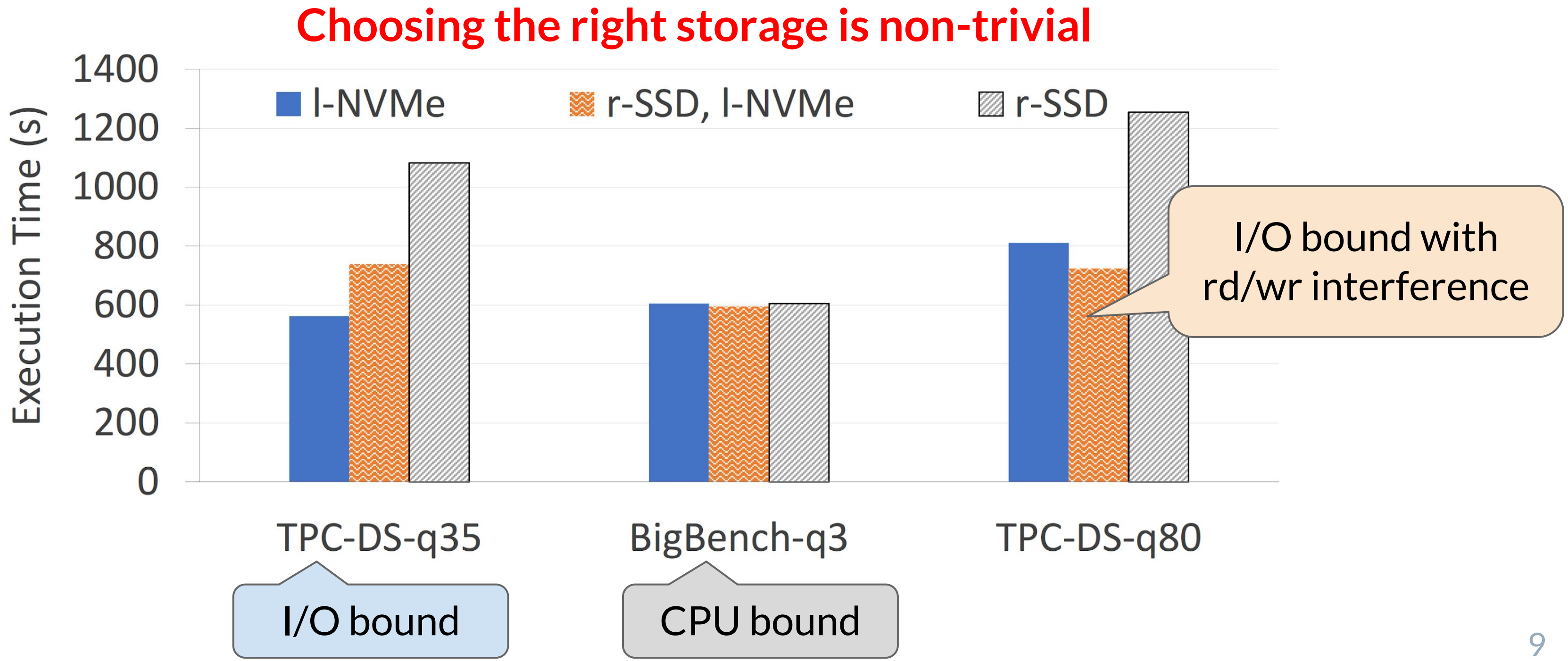
- Storage configuration is particularly critical for data analytics
- Jobs often have multiple data streams (e.g., shuffle, input/output data) with diverse I/O characteristics, making them suitable for different storage options



The cloud storage landscape is only getting more diverse with new technology (e.g., 3D X-point).

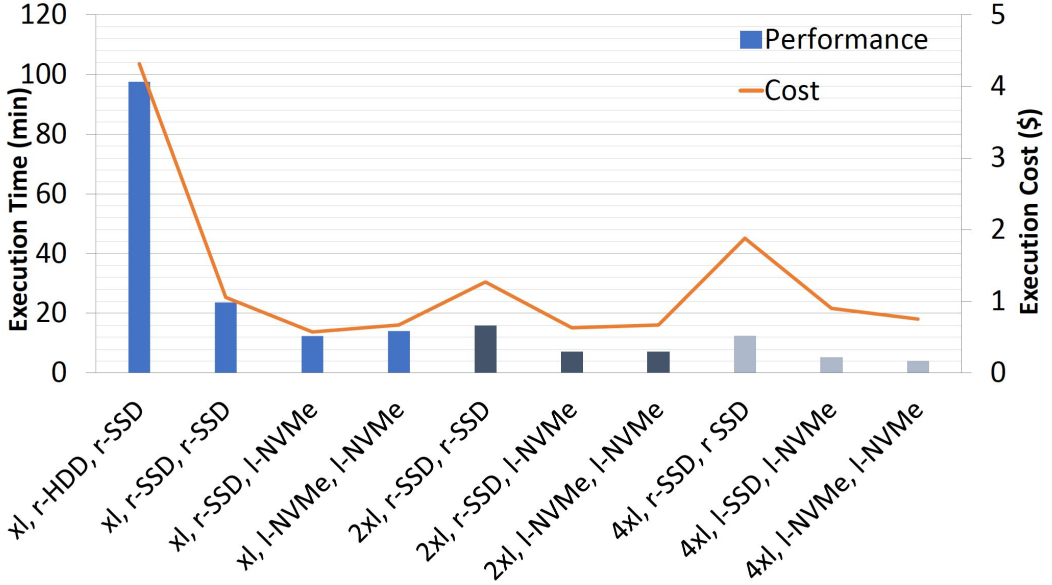
Storage configuration is challenging

- Example: selecting between 3 storage options — all other parameters constant



Performance and cost impact

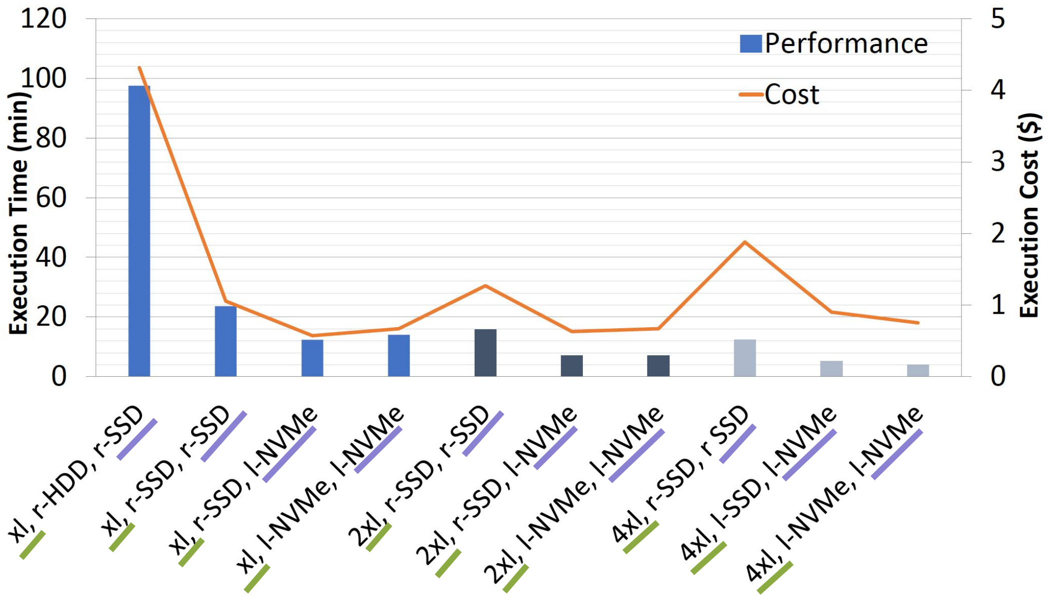
- Compare the performance and cost of TPC-DS query 64 on 10 configurations



<VM size, input/output storage, intermediate storage>

Performance and cost impact

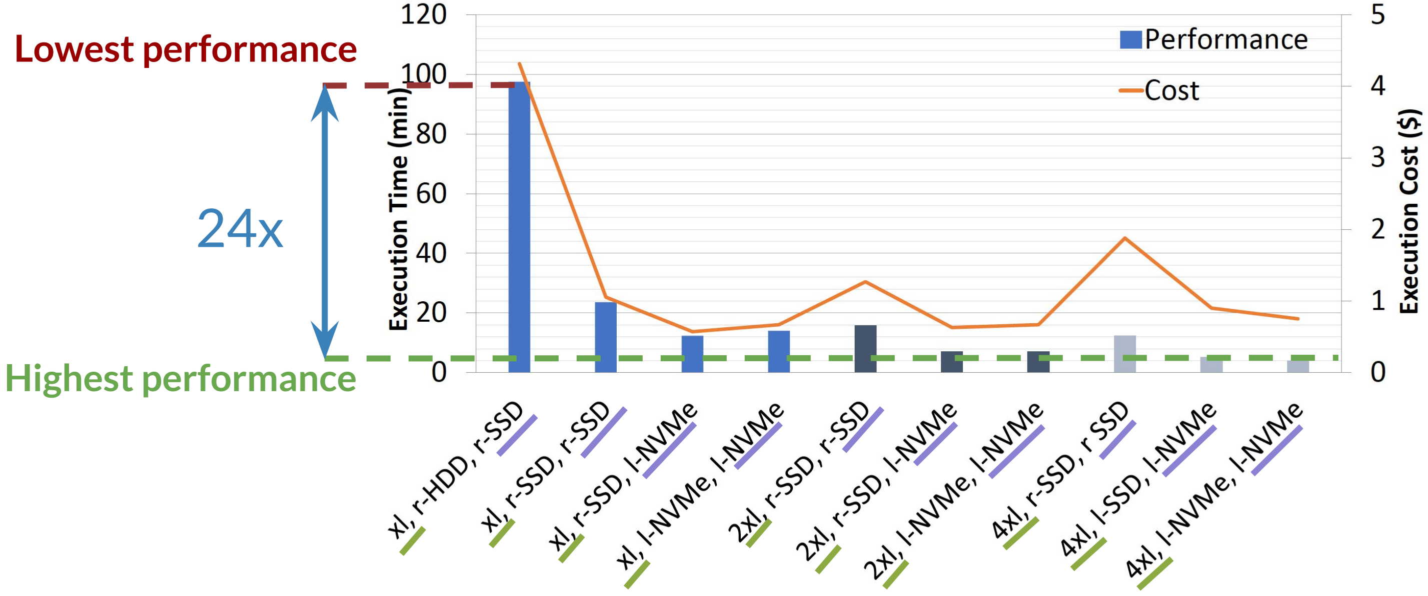
- Compare the performance and cost of TPC-DS query 64 on 10 configurations



<VM size, input/output storage, intermediate storage>

Performance and cost impact

- Compare the **performance** and cost of TPC-DS query 64 on 10 configurations



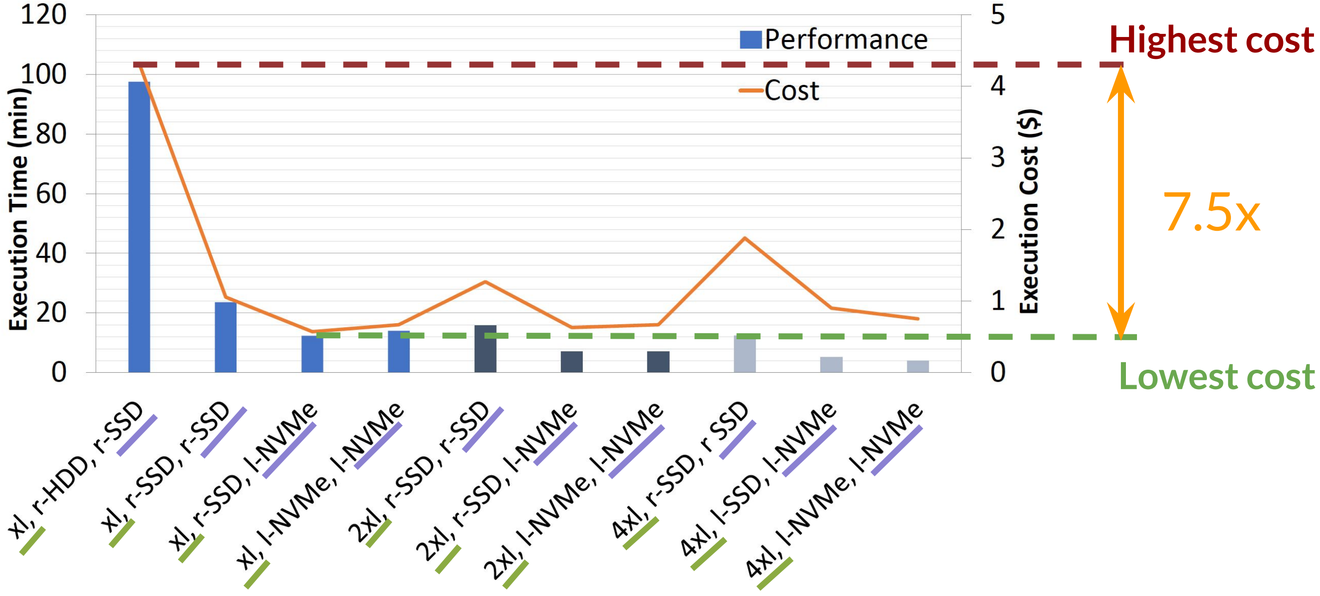
<VM size, input/output storage, intermediate storage>

Contrary to a previous study [1] that showed optimizing storage improves Spark performance by only 19%.

[1] Ousterhout, K., et al. *Making Sense of Performance in Data Analytics Frameworks.* NSDI'15.

Performance and cost impact

- Compare the performance and **cost** of TPC-DS query 64 on 10 configurations



<VM size, input/output storage, intermediate storage>

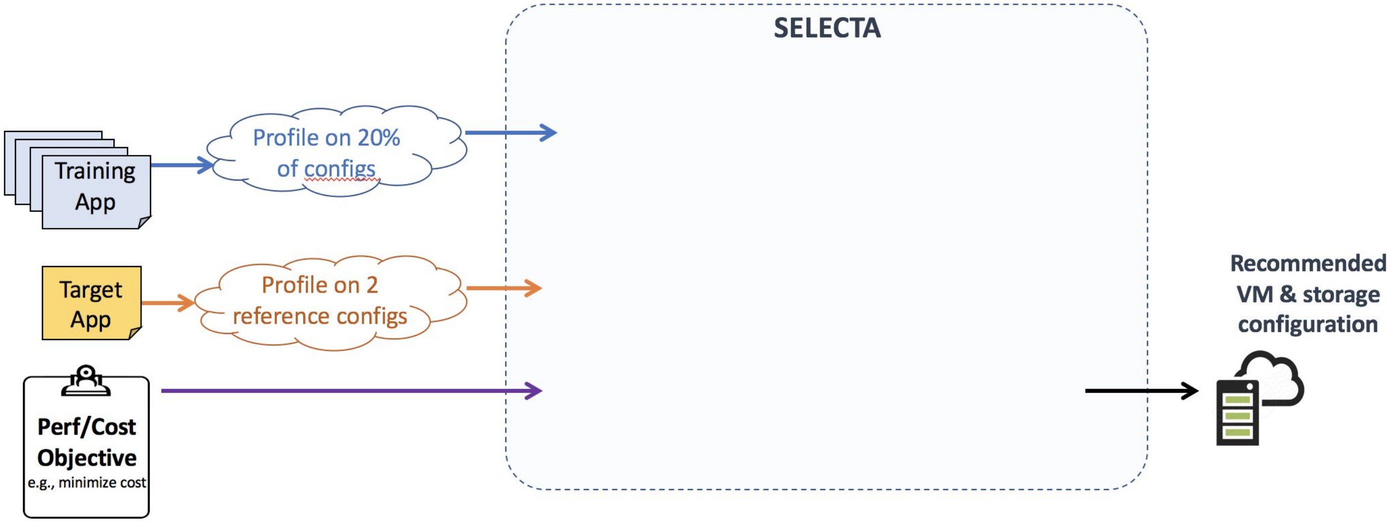
Contributions

1. ***Selecta***, a tool that recommends near-optimal cloud VM and storage configurations for target applications based on sparse training data

2. Analysis of data analytics performance with different storage options:
    - Which storage options are good fit and for different data streams?
    - What lessons do we learn for the design of future cloud storage systems?

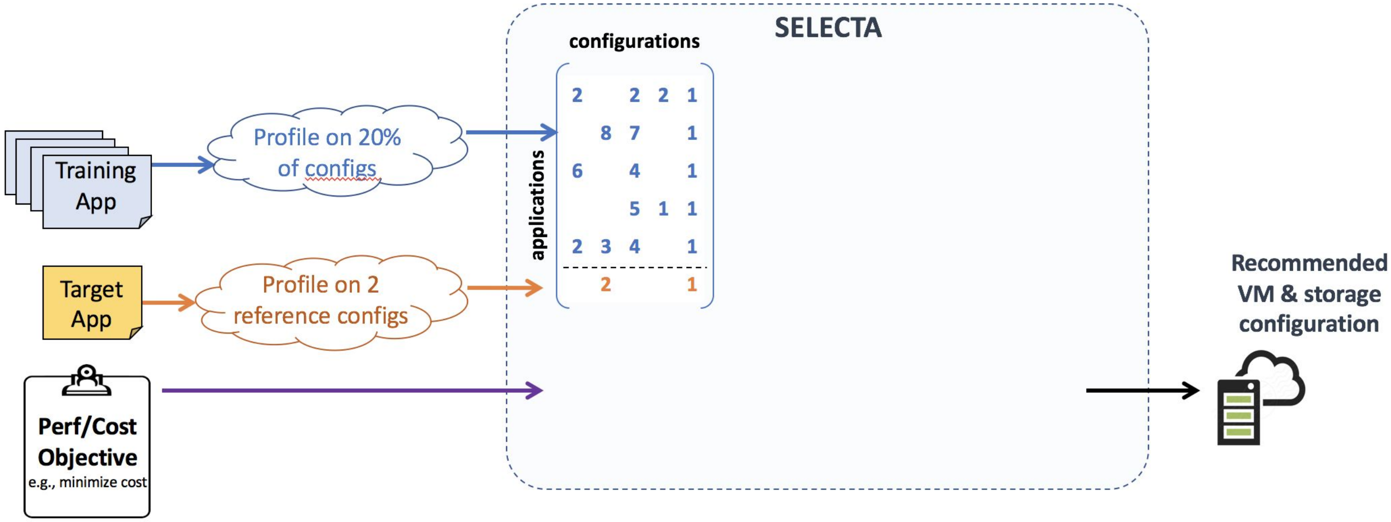
Selecta

- A system that predicts the performance of a target application on candidate configurations using *sparse* training data across jobs → recommend the right config



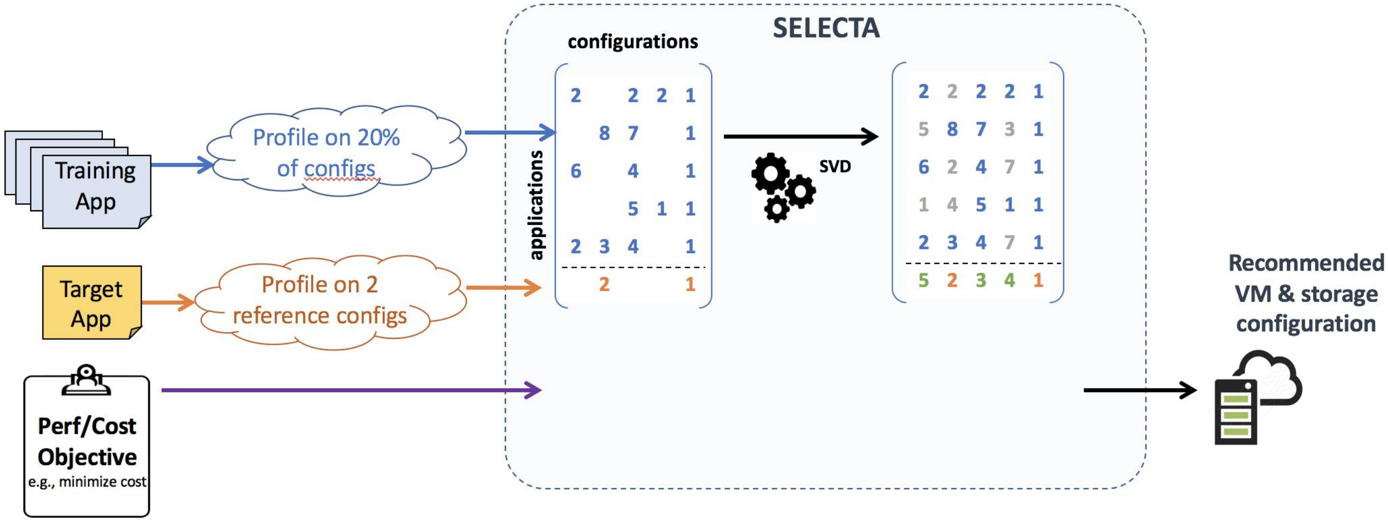
Selecta

- A system that predicts the performance of a target application on candidate configurations using *sparse* training data across jobs → recommend the right config



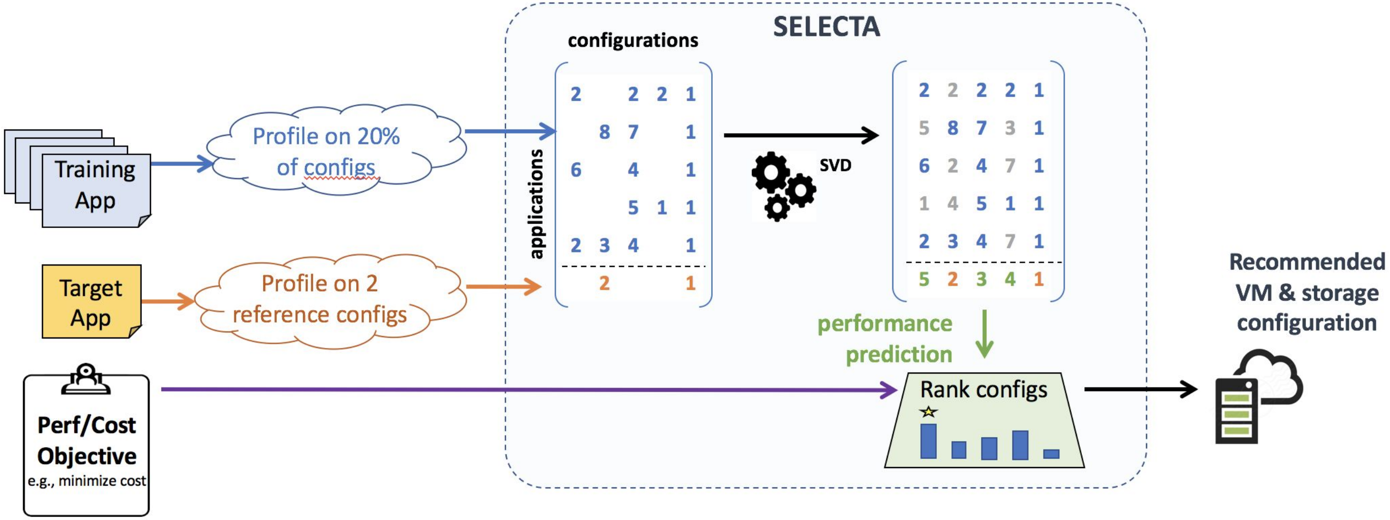
Selecta

- A system that predicts the performance of a target application on candidate configurations using *sparse* training data across jobs → recommend the right config



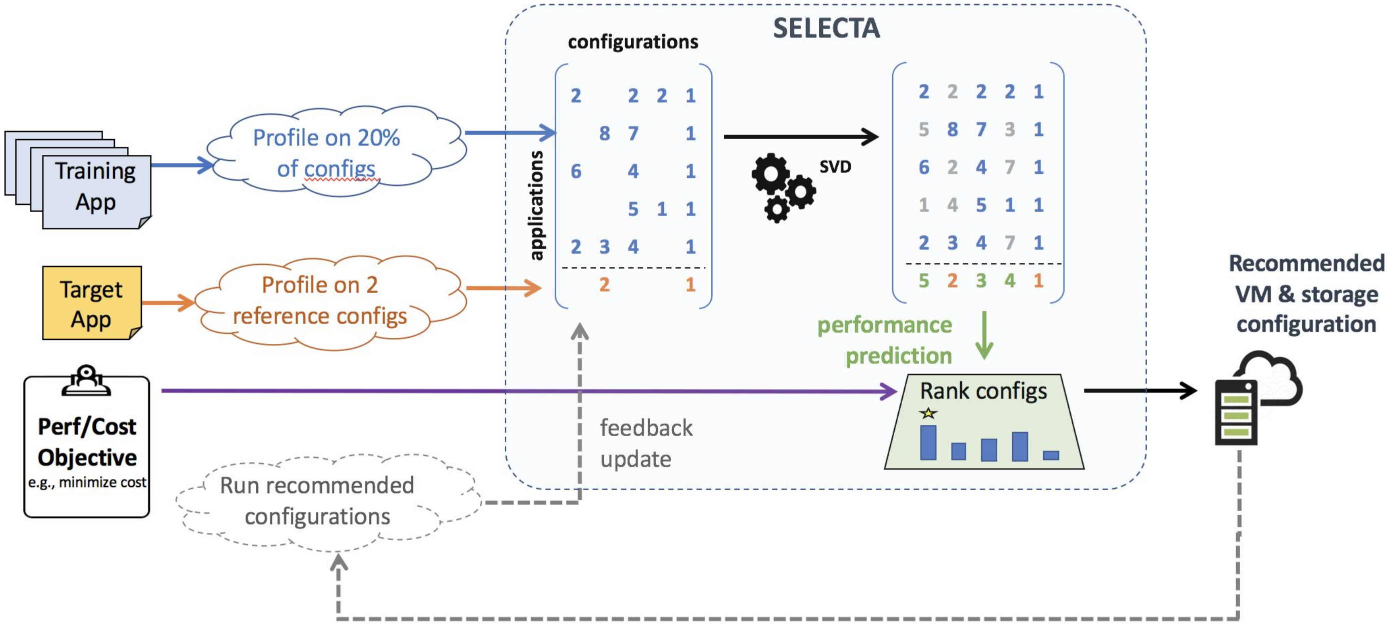
Selecta

- A system that predicts the performance of a target application on candidate configurations using *sparse* training data across jobs → recommend the right config



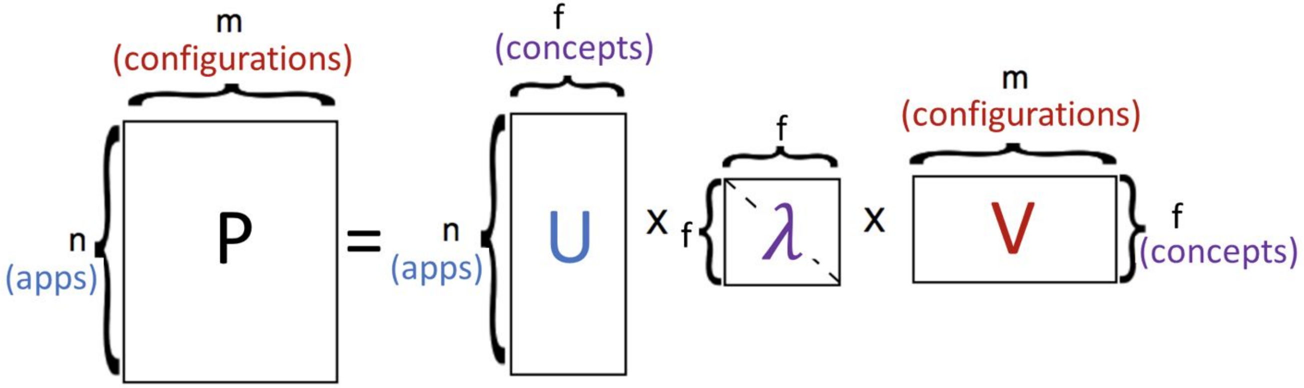
Selecta

- A system that predicts the performance of a target application on candidate configurations using *sparse* training data across jobs → recommend the right config



Collaborative Filtering

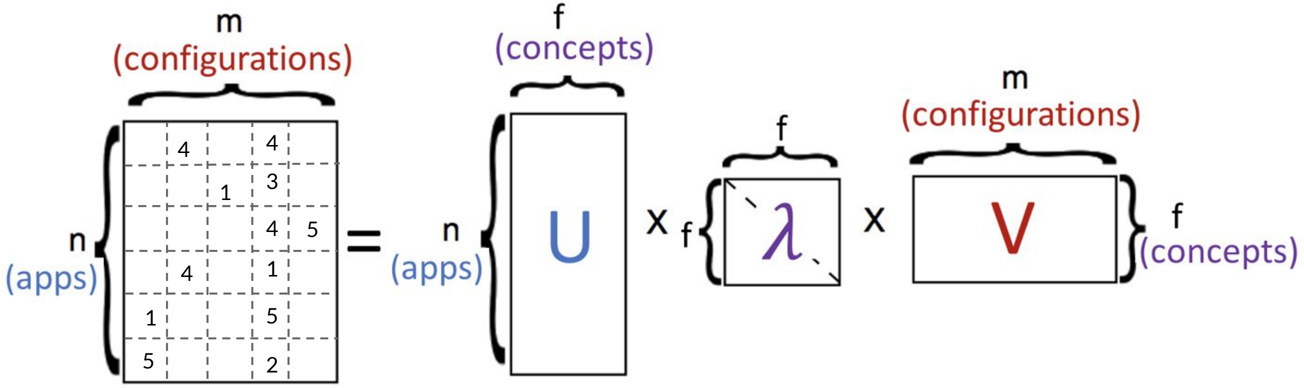
- Collaborative filtering approach: use singular value decomposition (SVD) to decompose app-config matrix **P** to uncover latent ("hidden") similarity concepts



Collaborative Filtering

- Collaborative filtering approach: use singular value decomposition (SVD) to decompose app-config matrix **P** to uncover latent ("hidden") similarity concepts

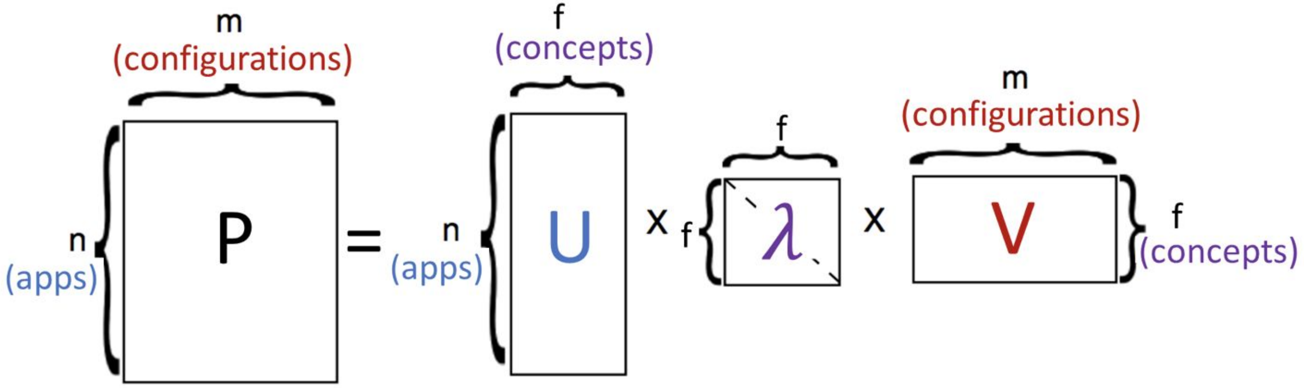
- P is sparse and SVD requires dense matrix → use stochastic gradient descent to update unknown entries; objective function minimizes error on known entries



Collaborative Filtering

- Collaborative filtering approach: use singular value decomposition (SVD) to decompose app-config matrix **P** to uncover latent (“hidden”) similarity concepts

✔ Automatically infers (latent) features
✔ Works well with sparse training set
✔ Agnostic to the applications and configurations used

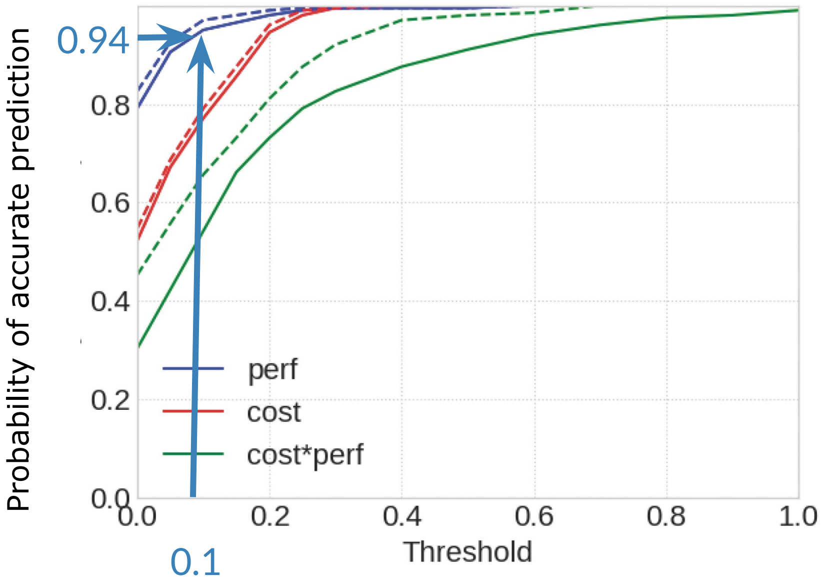


Evaluation Methodology

- Run >100 different Spark SQL/ML applications on 17 different configurations
- Two dataset sizes for each application
- Our candidate configuration space (in Amazon EC2):
  - 8-node clusters of 3 different VM sizes (vary CPU cores & DRAM per node)
  - Storage options:
    - Remote block storage (EBS) HDD
    - Remote block storage (EBS) SSD
    - Local block storage NVMe
    - S3 object storage

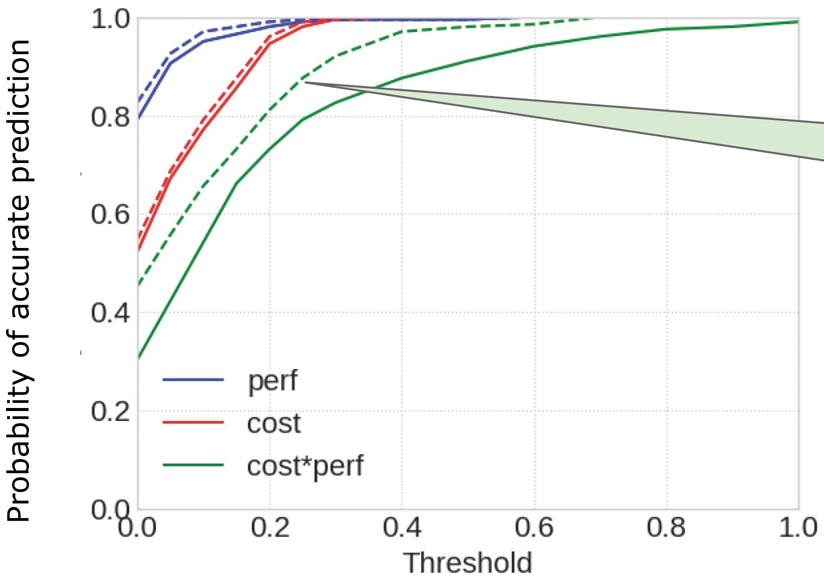
Selecta's Accuracy

- What is the probability of predicting a configuration that is near-optimal?



Selecta's Accuracy

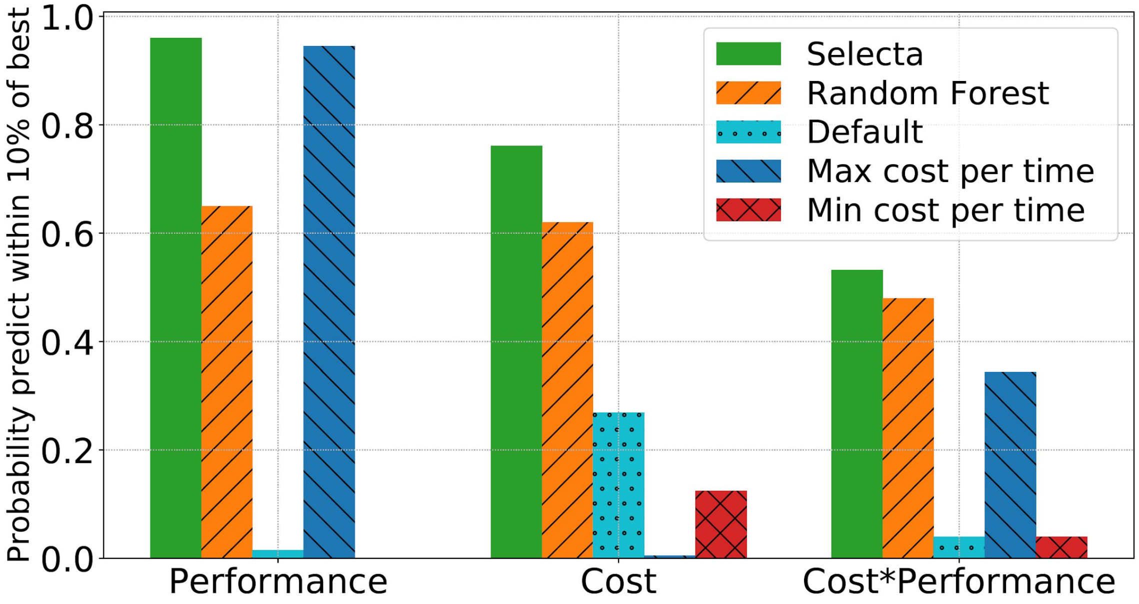
- Recommend near-optimal (T = 0.1) config for best perf with 94% probability
- Recommend near-optimal (T= 0.1) config for best cost with 80% probability



Dotted line shows improvement with one feedback round

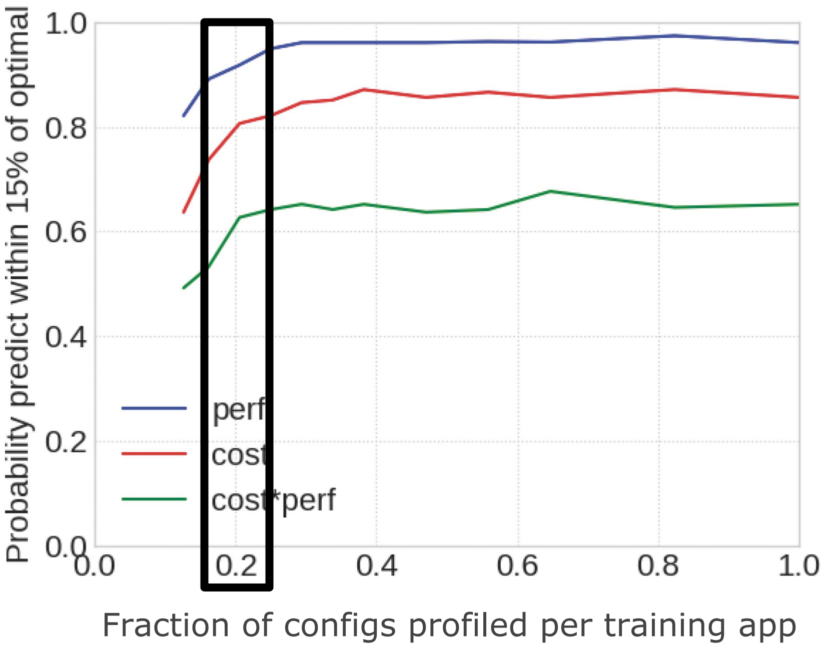
Comparison to alternative approaches

- Selecta's collaborative filtering learns best from the sparse training data even though it does not leverage as many features as the random forest predictor



Sensitivity analysis

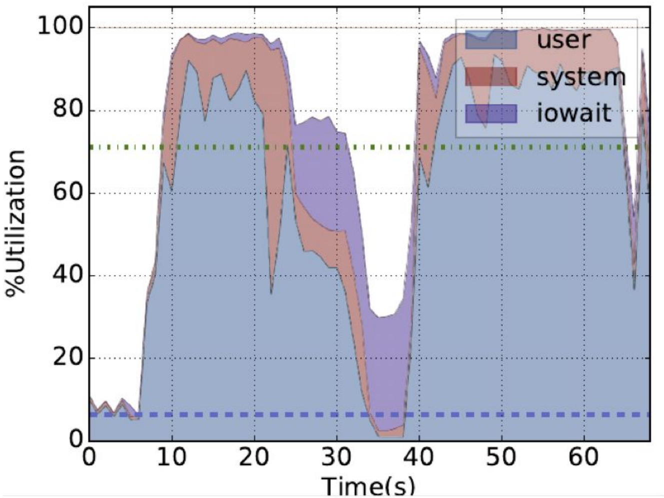
- Training matrix should be ~20% dense in steady state for good accuracy



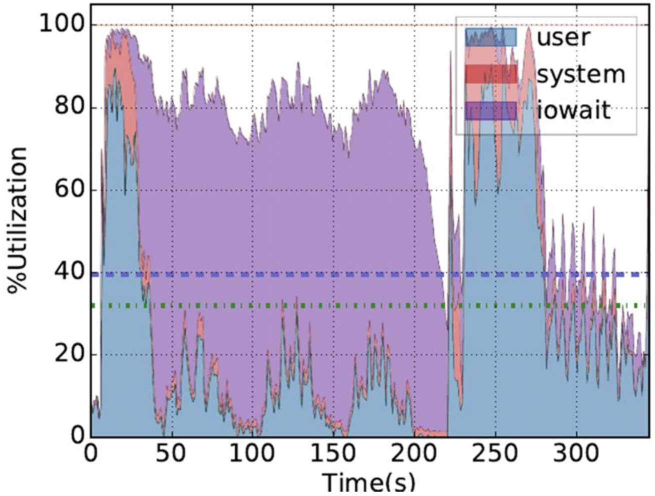
Dealing with application changes

- Changes in the input dataset can alter the CPU vs. I/O intensity of the job and influence the choice of optimal configuration
- When CPU utilization varies beyond a threshold, treat the job as a new application

300 GB dataset



1TB dataset



Lessons for storage system design

- NVMe storage is performance *and* cost efficient for data analytics
  - Great fit for intermediate data (shuffle, broadcast, etc.)
  - Good performance for input/output data but can get expensive to store the data long-term (use S3 instead)
- Fine-grain allocation of storage capacity and bandwidth -- disaggregated from compute resources -- is desired for better utilization
- There is a need to optimize across layers (apps, frameworks, OS) as many configurations fail to achieve their potential due to software inefficiencies

Conclusion

- Cloud cluster configuration is difficult yet critical for performance and cost
- Selecta is a tool that uses collaborative filtering to make near-optimal configuration recommendations for a user's performance-cost objective
    - 94% probability of predicting configuration with near-optimal performance
    - 80% probability of predicting configuration with near-optimal cost
- We use Selecta to explore the cloud storage landscape in the context of data analytics to guide the design of future storage systems