Dayu: Fast and Low-interference Data Recovery in Very-large Storage Systems

Zhufan Wang*, Guangyan Zhang*, Yang Wang†,
Qinglin Yang*, and Jiaji Zhu‡

*Tsinghua University
†The Ohio State University
‡Alibaba Cloud

THE OHIO STATE UNIVERSITY
COLLEGE OF ENGINEERING

Overview of Data Recovery Protocol

- Basic idea: replicate data chunks and re-replicate lost chunks

- Responsibility of the recovery protocol:
  - Schedule the source, the destination, and the bandwidth for re-replicating each lost chunk

- Goals of scheduling: high quality and high speed
  - High quality: Achieve fast and low-interference recovery
  - High speed: The scheduling algorithm should not become the bottleneck

Observations from a Production System

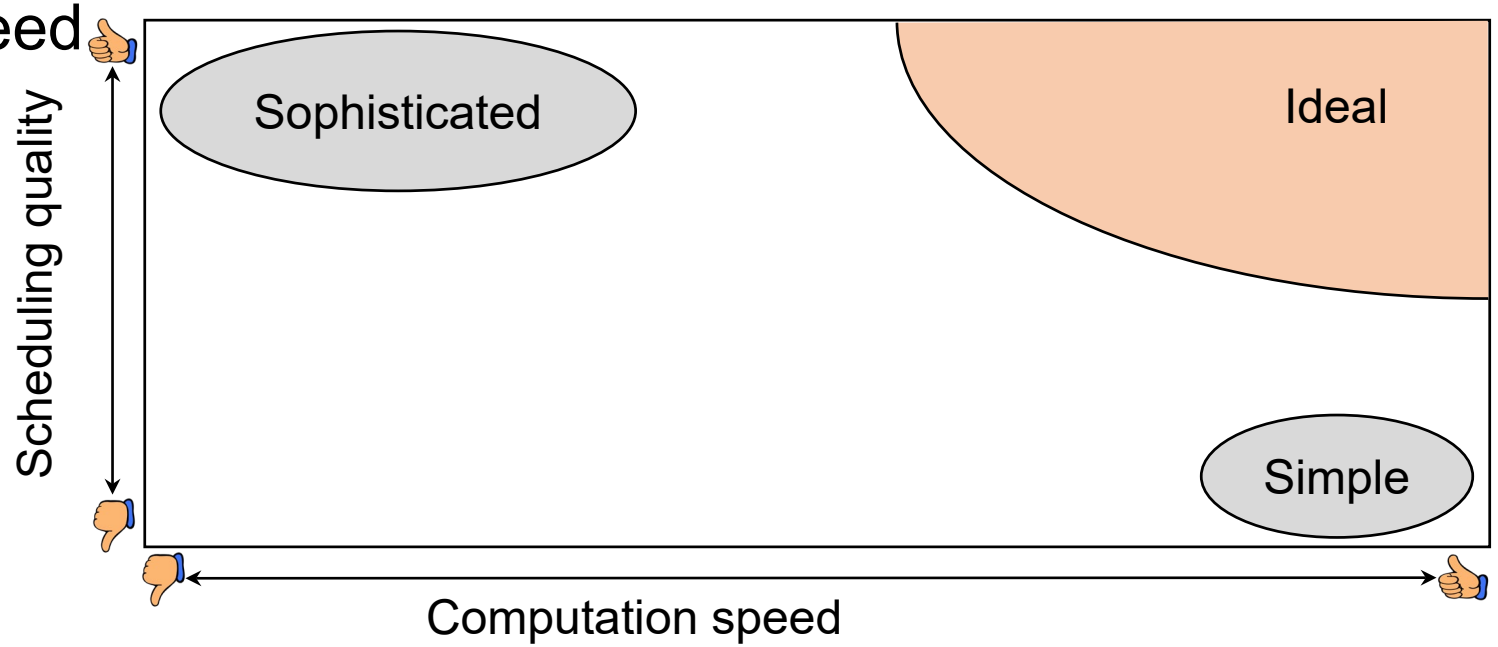
- Target system:
  - AliCloud's distributed storage system: Pangu
  - Deployed on a datacenter with approximately 3500 nodes

- Observations:
  - Very-large scale
  - Tight time constraint
  - Imbalanced resources
  - Dynamic foreground traffic

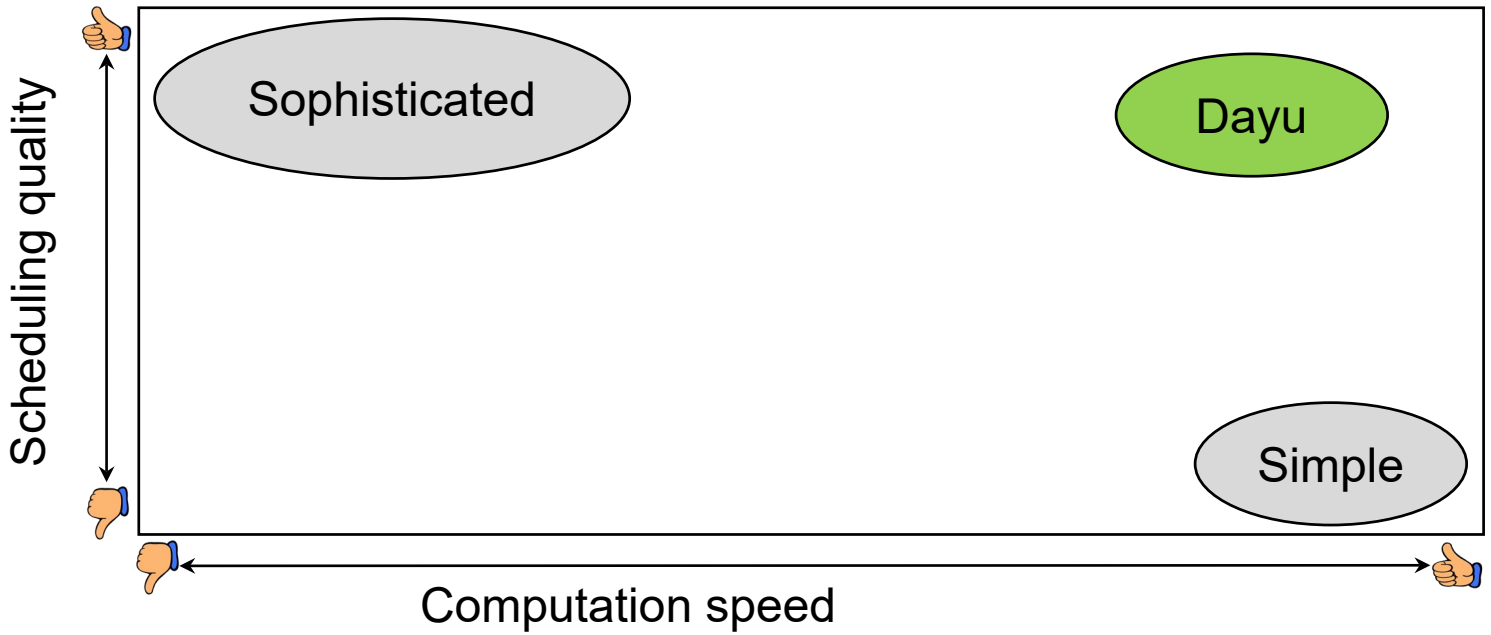
- Challenge: the scheduling algorithm needs to compute a large and complex problem within seconds

Existing Approaches

- Simple and decentralized scheduler
  - E.g. GFS, HDFS, Azure, RAMCloud, Sparrow, etc
  - High speed but low quality
- Sophisticated and centralized scheduler
  - E.g. CAR, PPR, Mirador, DH-HDFS, Firmament, etc
  - High quality but low speed



Dayu: High-quality and high-speed Recovery



- Evaluation result:
  - 2.96x recovery speed with only 3.7% increase in tail latency
  - Can scale to the cluster of 25K nodes

Dayu: Fast and Low-interference Data Recovery in Very-large Storage Systems

Coming soon.

ATC 2019

Friday, July 12, 12:30 am

Track: Storage Failure & Recovery

THE OHIO STATE UNIVERSITY
COLLEGE OF ENGINEERING