Maximizing Efficiency By Trading Storage for Computation

Ian F. Adams, Darrell D.E. Long, Ethan L. Miller

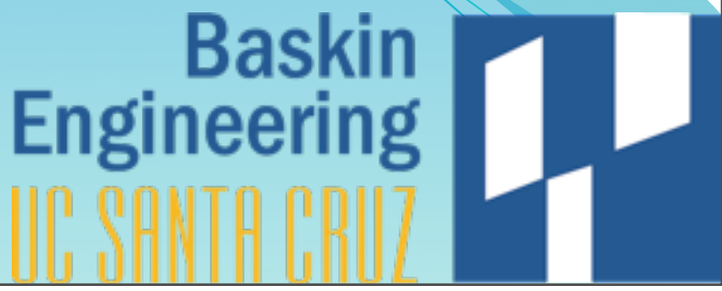
University of California, Santa Cruz

Shankar Pasupathy

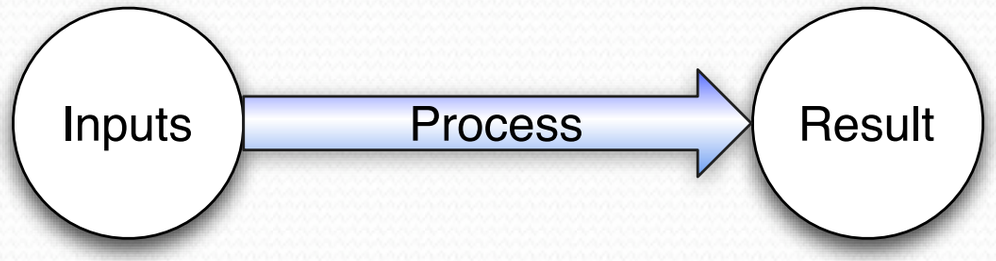
NetApp

Mark W. Storer

Pergamum Systems



Trade Storage for Computation

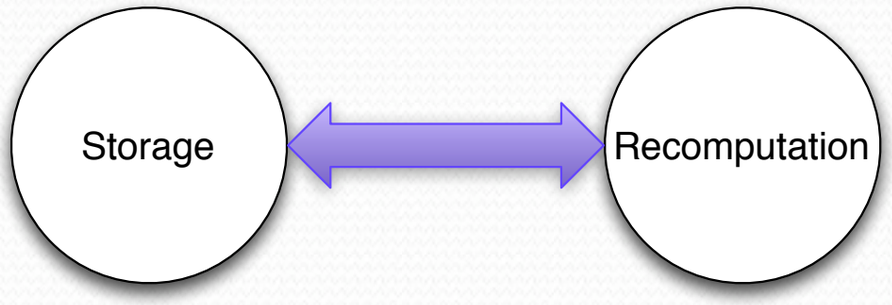


- Storing rarely used intermediate or final results is wasteful
  - If results are still used, even infrequently, they may not be discarded

- In the cloud, computing a result is often cheaper than storing it
  - store the inputs and process, recompute on demand

- **Challenges:**
  - Decide when re-computation is cheaper than storage
  - Provenance: what must be known in order to reproduce a result?
  - Result integrity: must a result be identical to the original?

Old Idea, New Potential



- Balancing storage and computation is common
  - Data de-duplication, file compression, dynamic programming…

- The cloud allows opportunities for large scale tradeoffs
  - Quickly allocate resources to compute on demand
  - No need for over provisioning to prepare for rare events
    - purchase on demand computation when needed

Example

- Archive of 100,000 photos
  - Provide bmp, jpeg, tiff, Adobe, png
  - Use Amazon Web Services

- **Store Everything**
  - 1600x1200 resolution images in 5 formats= 2.2TB
  - 100 GB of requests
  - C_s =***$347.00 per month***

- **Recompute Formats on Demand**
  - Raw BMP=550 GB
  - 720 on demand "small" linux instance hours
  - 100 GB out from S3 to EC2+100 GB EC2 Out
  - C_r =***$224.00 per month***

AWS Prices- June 12 2009	
S3 Storage Cost	$0.18 per GB/ Month
S3 Data In	$0.03 per GB
S3 Data Out	$0.17 per GB
EC2 Sm. Machine Instance	$0.10 per hour
EC2 Data In	$0.10 per GB
EC2 Data Out	$0.17 per hour

C_r = Cost of Re-computation
C_s = Cost of Storage

Challenges

- Calculating C_r and C_s can be difficult
  - Issues include: Likelihood of reuse, lead time, penalties for unavailable data, adaptation to cloud providers, pricing forecasts....
  - Requires knowledge from both cloud providers and users
    - e.g. user has "miss penalty" data, provider has infrastructure understanding

- Provenance: What is necessary to recompute a result?
  - Provenance aware systems (e.g. PASS) can aid analysis

- Result Integrity: Can the original result be recomputed?
  - Can a similar result suffice? Or must it be identical?

Questions? Comments?

Thanks to our sponsors:

- SSRC industrial sponsors
- Petascale Data Storage Institute

Thanks to:

- Mark W. Storer (Pergamum Systems)
- Darrell D.E. Long (UC Santa Cruz)
- Ethan L. Miller (UC Santa Cruz)
- Shankar Pasupathy (NetApp)
- Ahmed Ahmer (University of Pittsburgh)