Energy-Efficient Data Replication in Cloud Computing Datacenters

Presented by David Ocejo

Overview

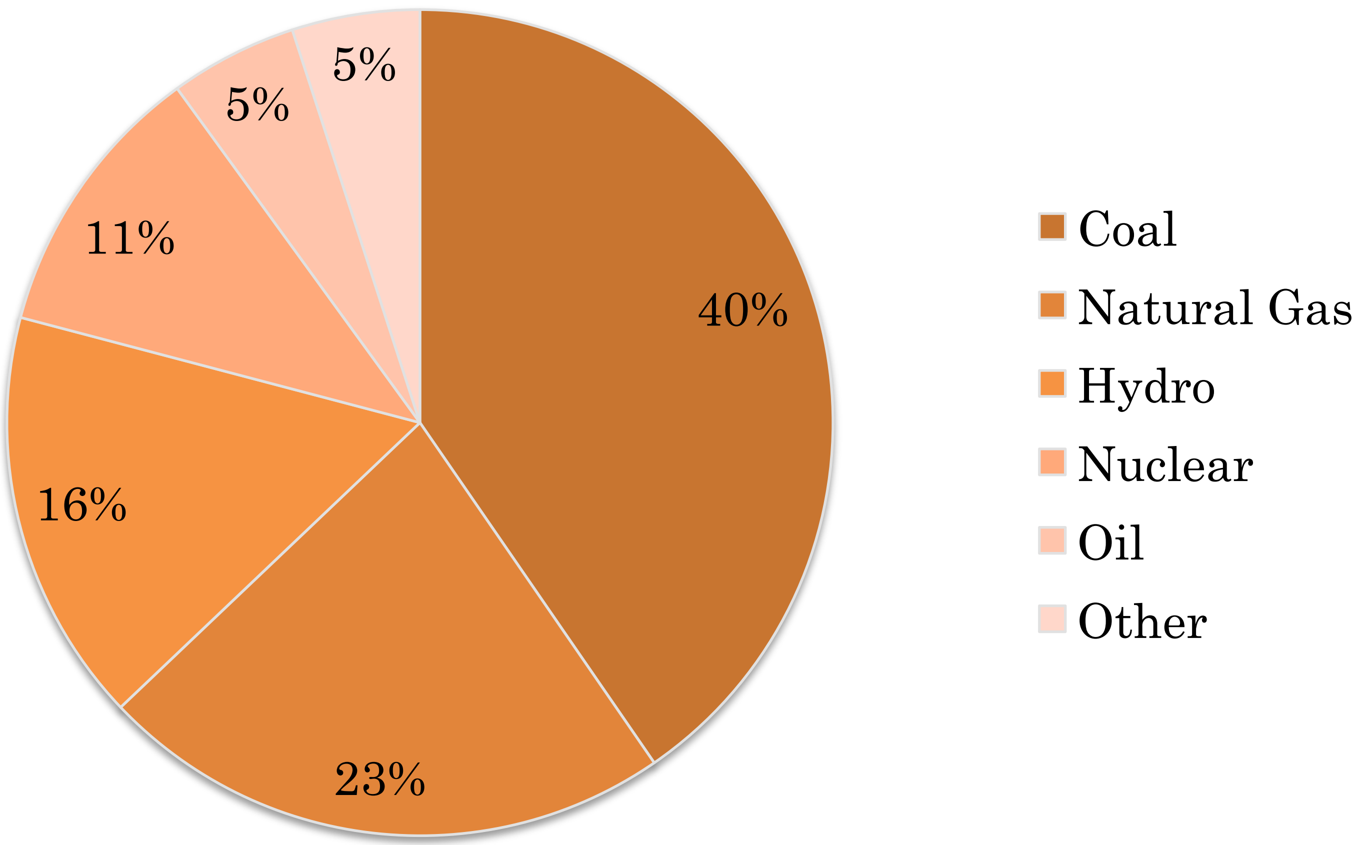
- Problem
- Saving Energy ("Solution")
  - Efficiency
  - Data Center Topology
- Simulation
  - Conditions
  - Results

Problem

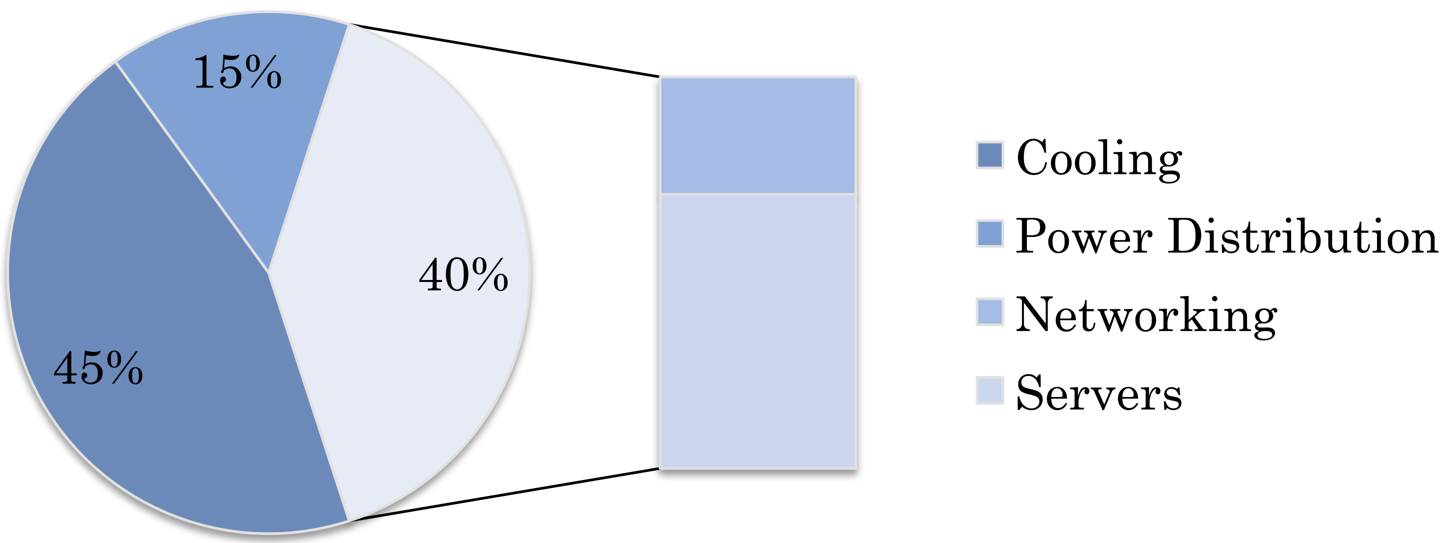
- Increasing energy consumption

- Up to 1.5% of World's Electricity (in 2010)
  - from 1.0% (in 2005)

World's Electricity Generation



DATA CENTER ENERGY CONSUMPTION



Energy Efficiency

- Two approaches:
  - Shutting down components
  - Scaling down performance

Energy Efficiency

- Shutting Down Components
  - Dynamic Power Management (DPM)
  - Dynamic Network Shutdown (DNS)

Energy Efficiency

- Scaling Down Performance
  - Dynamic Voltage and Frequency Scaling (DVFS)
    - Applicable only to CPU
    - Other components still consume at peak rates
  - Dynamic Voltage Scaling (DVS)
    - Links
  - $P = V^2 * f$
    - = (supplied voltage ^{2}) * (operating frequency)

Energy Efficiency

- Virtualization

Our Data Replication Approach

- Joint optimization of energy consumption and bandwidth capacity
- Optimization of communication delays

Data Center

- Three Tier Topology
  - Core Layer
    - Flows going in and out of data center
  - Aggregation Layer
    - Integrates connections and traffic flows from racks
  - Access Layer
    - Where computing servers are arranged into racks

Data Center

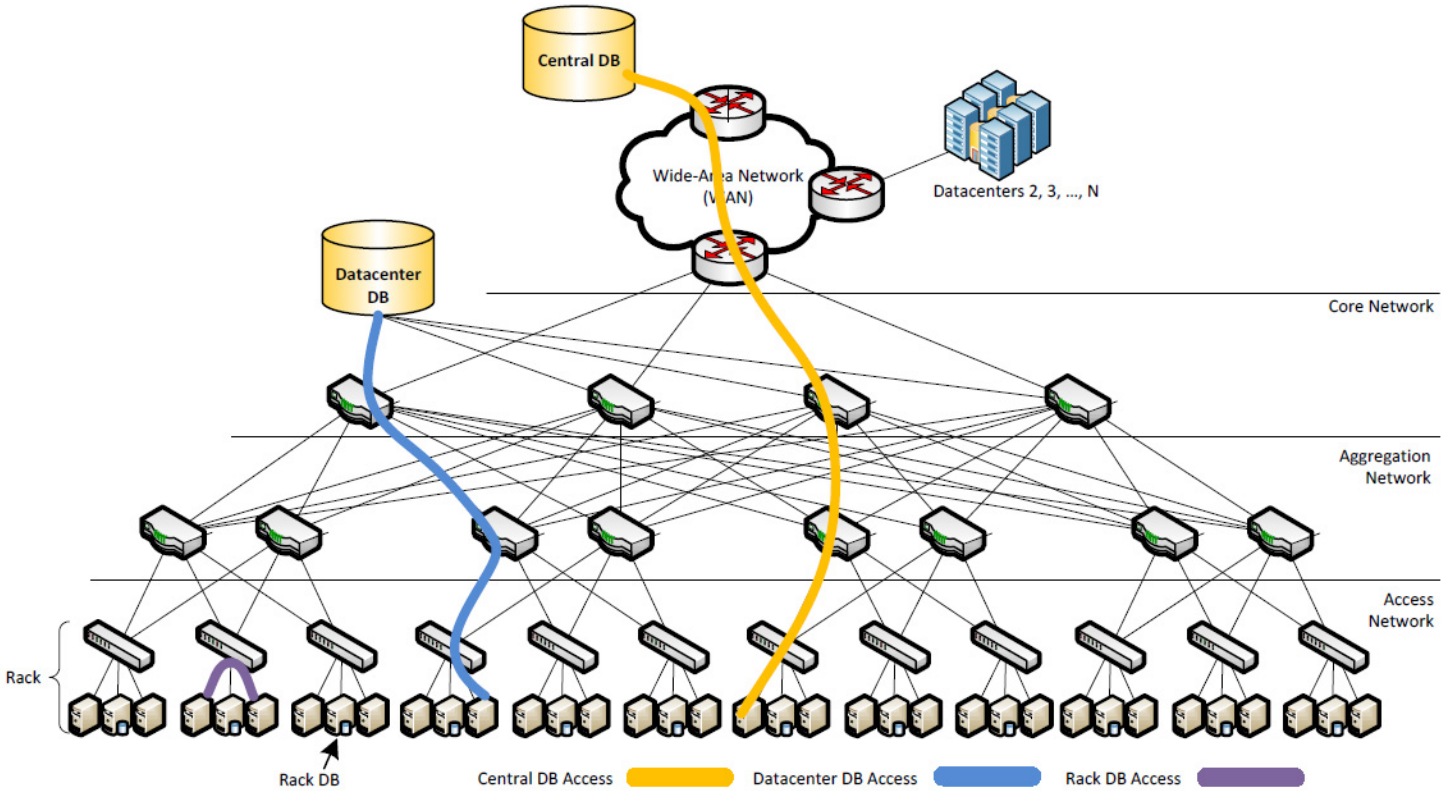
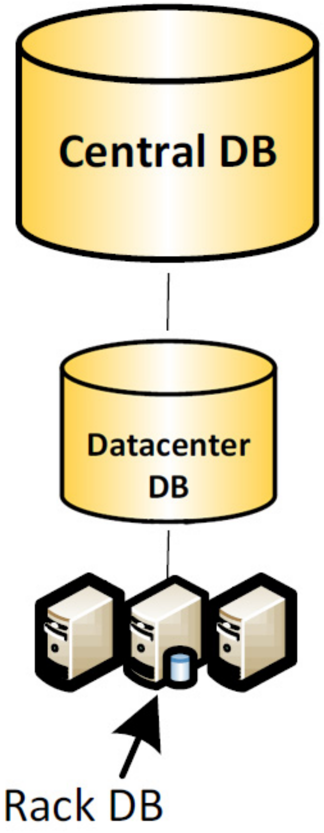


Fig. 1. Three-tier cloud computing data center architecture.

Data Center



DATA CENTER

- External requests directed to Rack DB
  - If necessary, Database DB and Central DB

- Databases maintain and exchange access records
  - Requesting (rack) server and database
  - Number of data item accesses and updates

- Popularity
  - Access rate: number of access events in given time period
  - Decays

Data Center

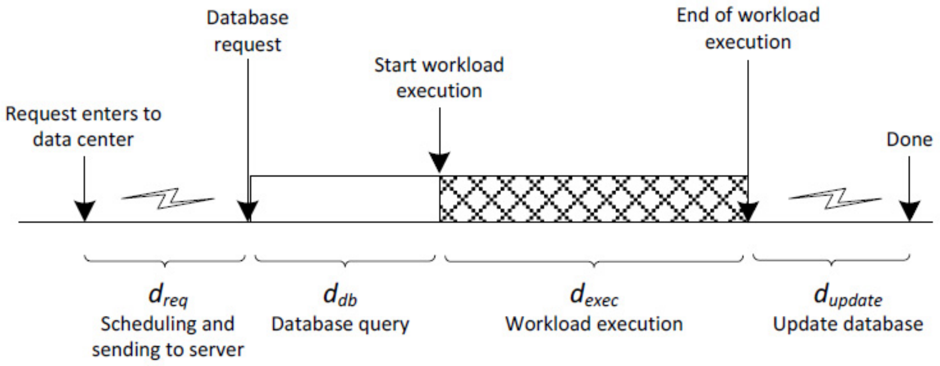
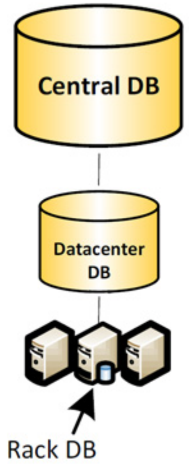


Fig. 2. Workload execution timeline.



Data Center Transmissions

- Uplink – Bandwidth
  - Propagating database requests
  - Updating data items

- Downlink – Bandwidth
  - Delivering workload descriptions
  - Receiving database objects
  - Propagating updates between DB replicas

Power Consumption - Servers

$$= Fixed + (\frac{Peak - Fixed}{2})(1 + load - e^{-\frac{1}{a}})$$

- Servers consume two-thirds when idle
  - Memory modules, disks, I/O, etc. still consuming at peak rate

Power Consumption - Switches

$$= Chassis + (NumberOfLineCards * LineCard) + \sum_{r=1}^{R} (n_p^r * P_p^r * u_p^r)$$

- Power drawn by port running at rate r
- Number of ports running at rate r
- Utilization of ports

- 85-97% fixed energy consumption
- 3-15% consumed by port transceivers

Simulation

- Performed using GreenCloud simulator
  - Cloud computing simulator
  - Packet level communication
- Single data center simulation
  - 60 minutes

Simulation – Conditions

TABLE I. Power Consumption of Datacenter Hardware

Parameter	Power Consumption [W]		
	Chassis	Line cards	Port
Gateway, core, aggregation switches	1558	1212	27
Access switches	146	-	0.42
Computing server	301		

TABLE II. Datacenter Topology

Parameter	Value
Gateway nodes	1
Core switches	4
Aggregation switches	8
Access (rack) switches	32
Computing servers	1024
Gateway link	100 Gb/s, 50 ms
Core network link	10 Gb/s, 3.3 μs
Aggregation network link	10 Gb/s, 3.3 μs
Access network link	1 Gb/s, 3.3 μs

Simulation – Conditions

- DB queries limited to 1500 bytes
  - Fits into single Ethernet packet
- Varying:
  - Data item size
  - Data access and update rates
  - Replication threshold
- DNS power saving enabled

Simulation – Results

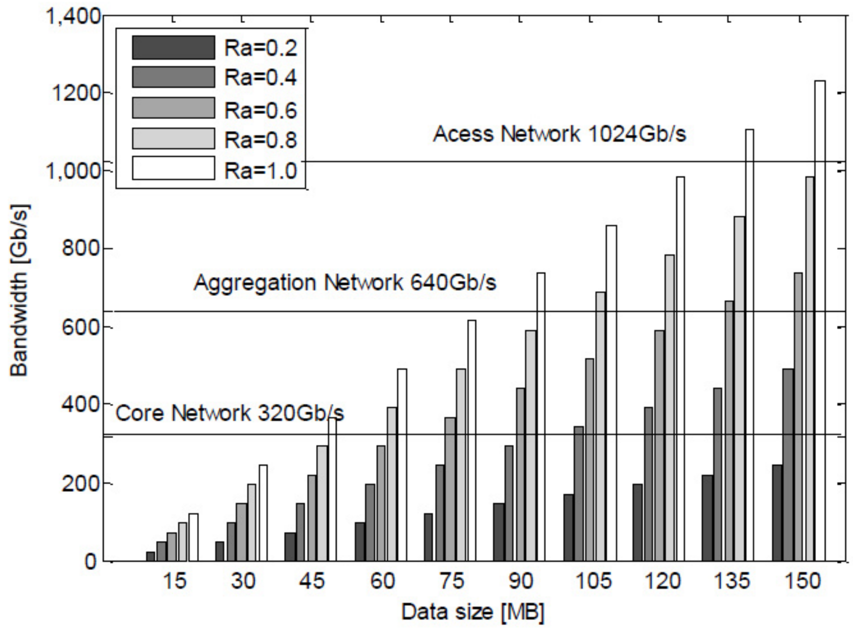


Fig. 3. Downlink bandwidth demand.

SIMULATION – RESULTS

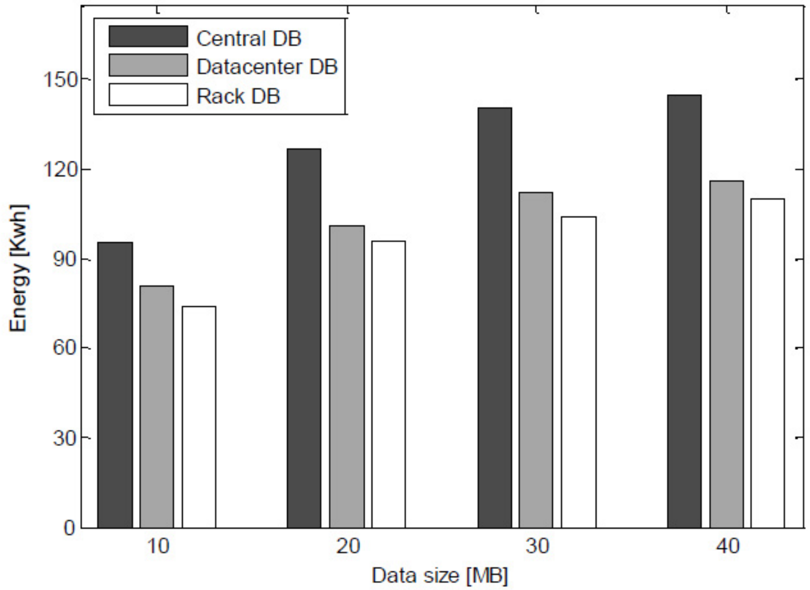


Fig. 4. Energy consumption of servers.

Simulation – Results

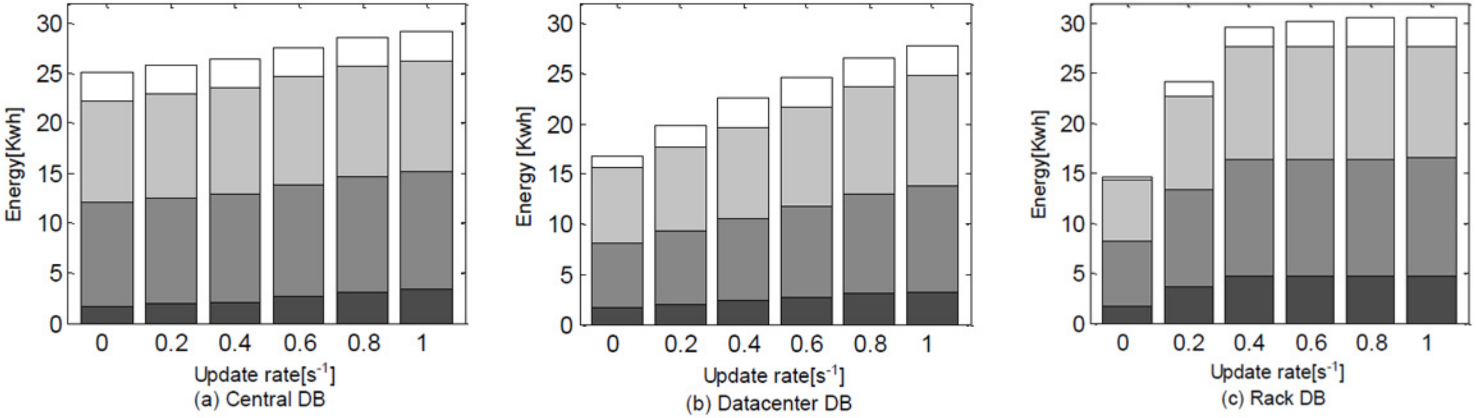
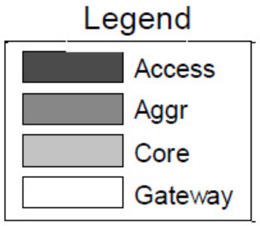


Fig. 5. Energy consumption of network switches.



Simulation – Results

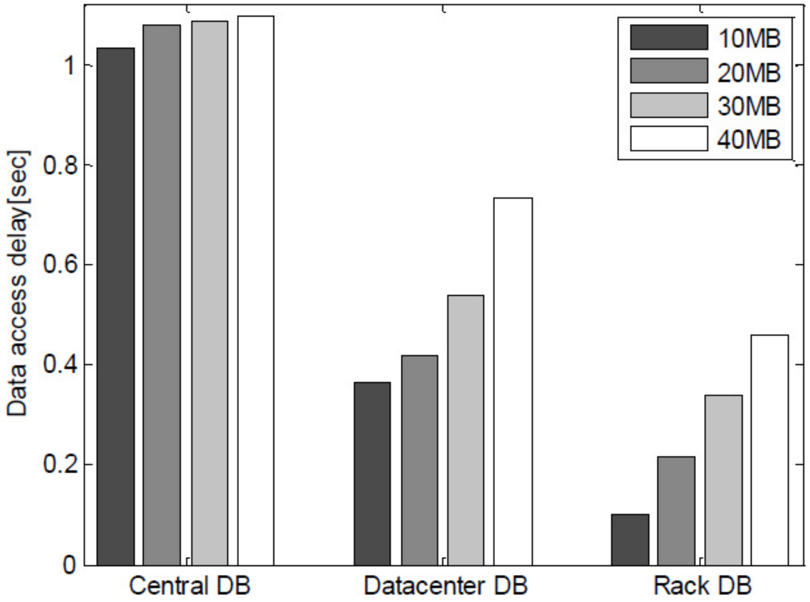


Fig. 6. Data access delay.

Conclusion

- Replicating data closer to data consumers reduces:
  - Energy consumption
  - Bandwidth usage
  - Communication delays
- Degree of reduction dependant on update rate