

Hálózatba kapcsolt erőforrás platformok és alkalmazásaik

Simon Csaba, Maliosz Markosz

TMIT

2017

Adminisztratív kérdések

VITMAC03

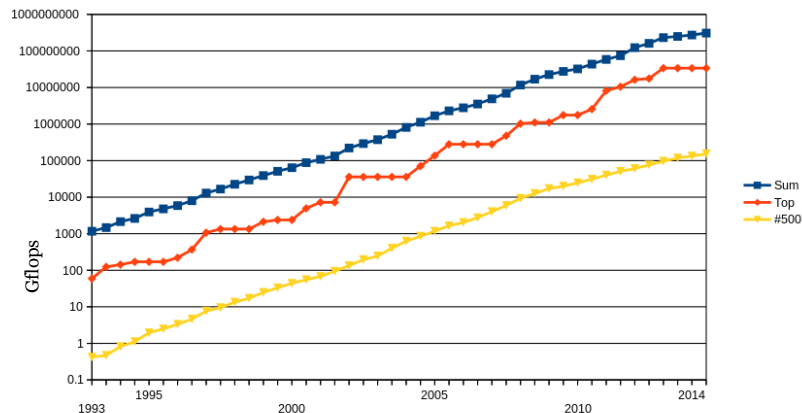
- Hálózatba kapcsolt erőforrás platformok és alkalmazásaik
<http://www.tmit.bme.hu/simon.csaba>
<http://www.tmit.bme.hu/vitmac03>
- TMIT, IE324
- Simon Csaba – simon@tmit.bme.hu
- Maliosz Markosz – maliosz@tmit.bme.hu
- ZH – március 27.
 - Gyakorlatok: jelenlét
 - pótZH
 - Elővizsga – min. 4-es ZH
 - Pótlási héten pót-pót-ZH
- Vizsga – nagy és kis kérdések

Hálózati erőforrásmegosztás

Nagy erőforrás kapacitások elérése

- Nagy erőforrás-igény = nagy gép = drága és exkluzív
 - Szuperszámítógépek
 - Fizikai és adminisztratív akadályok korlátozzák a felhasználást (nem elérhető)
 - Gazdasági szempontból előnytelen (költséges)
- Számítógép hálózat segítségével nagyméretű számítási erőforrás-rendszerek
 - Számítógépfürtök (klaszerek, clusters) -> cluster computing
 - Párhuzamos számítás (parallel computing)
 - Szuperszámítógép-építés lépései ☺

<http://www.wikihow.com/Build-a-Supercomputer>
- Lazán kapcsolt erőforrások kooperációja
 - A 90-es évek végétől, 2000-es évek elejétől kezdve
 - Különálló számítógépek Internet felett kapcsolódva
 - Igény szerinti erőforrás használat



1975: Cray Research Cray-1, Chippewa Falls, WI, USA (160 megaflops)



2011: Fujitsu K computer rack, Kobe, Japan (10 petaflops)

Korai hálózati erőforrás megosztás - 1/3

- Számítógép hálózat segítségével nagyméretű számítási erőforrás-rendszerek
- Számítógépek erőforrásának megosztása „Internet” felett
 - A kezdeti számítógép hálózatok egyik fontos szolgáltatása
 - A kommunikáció még nyilvános telefonhálózatokon történt
 - Kommunikáció költsége alacsonyabb az erőforrások költségeinél
 - Hálózatok tervezése és optimalizálása
 - Gazdasági hatékonyság

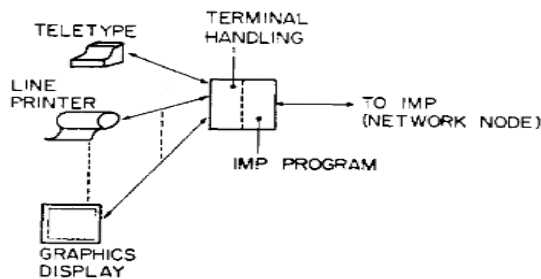


Fig. 2. Terminal IMP.

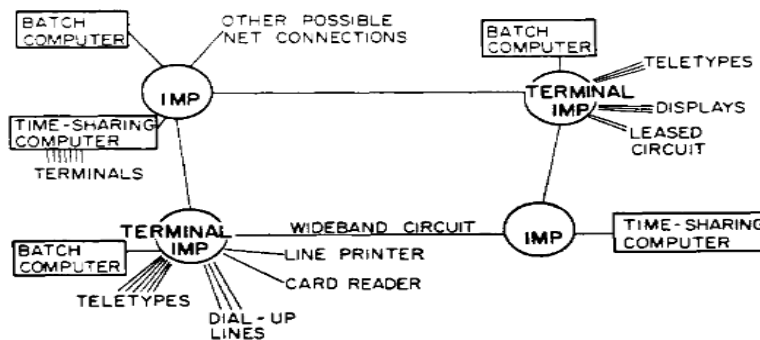


Fig. 3. A simple IMP network.

Kahn, Robert E. "Resource-sharing computer communications networks." *Proceedings of the IEEE* 60.11 (1972): 1397-1407.

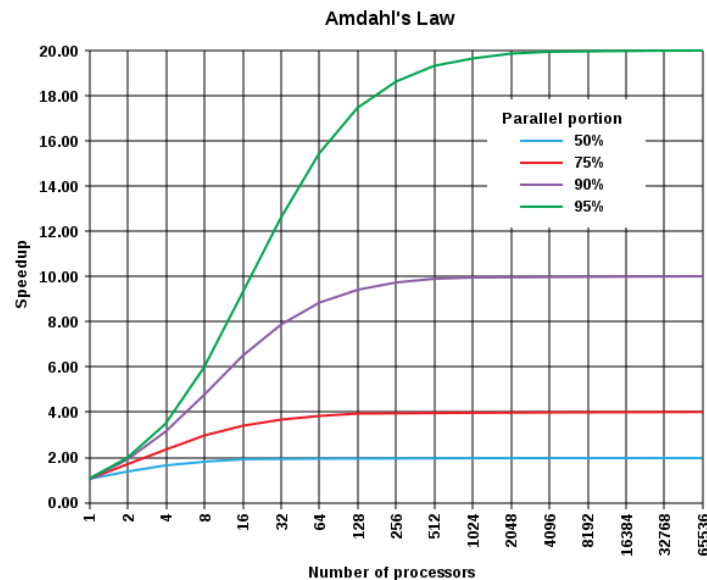
Korai hálózati erőforrás megosztás - 2/3

- „Shared resource”
 - Számítógépek közösen használják az erőforrást
 - Programok, fájlok, nyomtatók
 - Intranet, LAN – még nem Internet
 - Topológiák
 - Workgroup („peer-to-peer”)
 - Kliens szerver
- Korai LAN-ok esetében több alternatíva
 - Apple Filing Protocol / AppleTalk (Apple)
 - SMB / TCP/IP (Microsoft)
 - Network File System (NFS) / TCP/IP (Unix)
 - NetWare Core Protocol (NCP) / SPX (Novell)



Korai hálózati erőforrás megosztás - 3/3

- Számítógép hálózat segítségével nagyméretű számítási erőforrás-rendszerek
 - Számítógépfürtök (klaszterek, clusters) -> cluster computing
 - Nagysebességű helyi hálózaton keresztül összekötött ugyanolyan számítógépek
 - Amdahl törvénye megadja a párhuzamosítással elérhető nyereséget
 - A nem párhuzamosítható része a feladatnak korlátozza a gyorsítás mértékét



Igény szerinti számítógép erőforrások

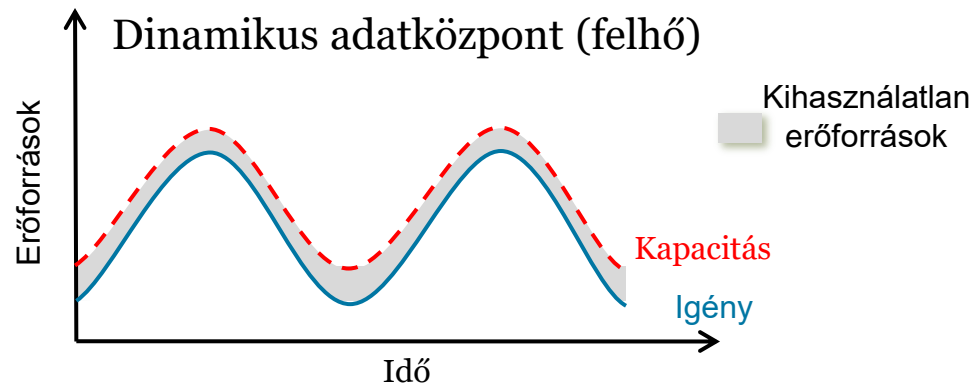
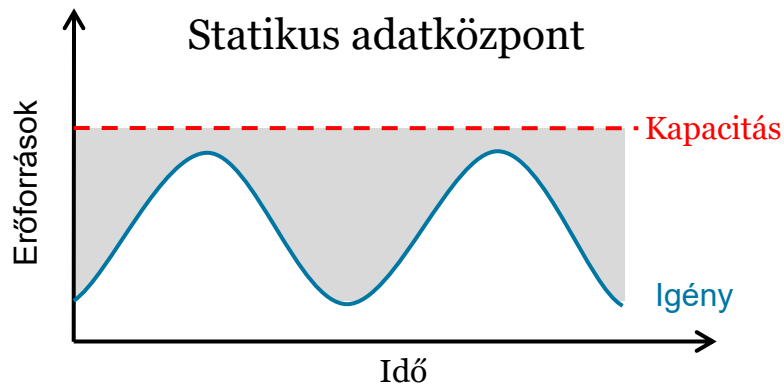
- Resource – as – a – Service
 - John McCarthy 1961-es álma

If computers of the kind I have advocated become the computers of the future, then computing may someday be organized as a public utility just as the telephone system is a public utility... The computer utility could become the basis of a new and important industry.

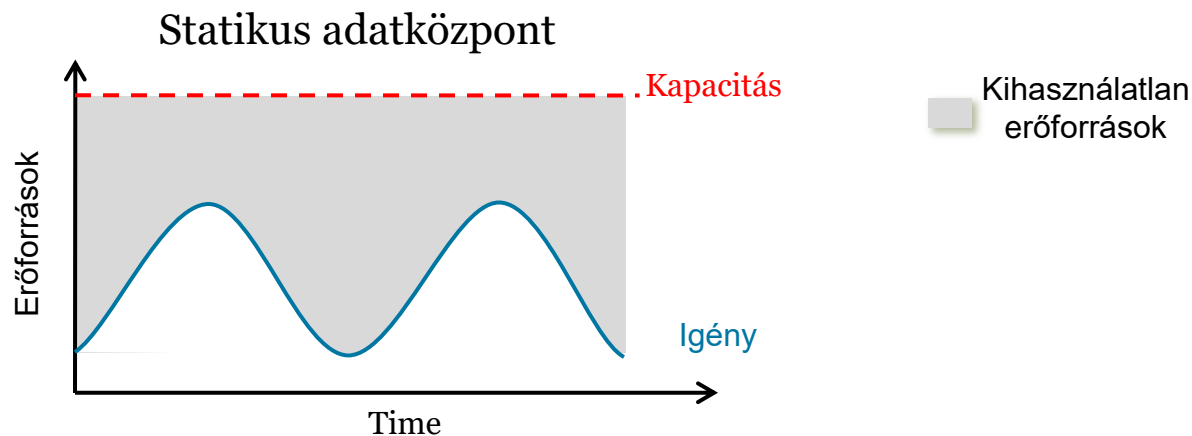
Igény szerinti erőforrások

Resource-as-a-Service

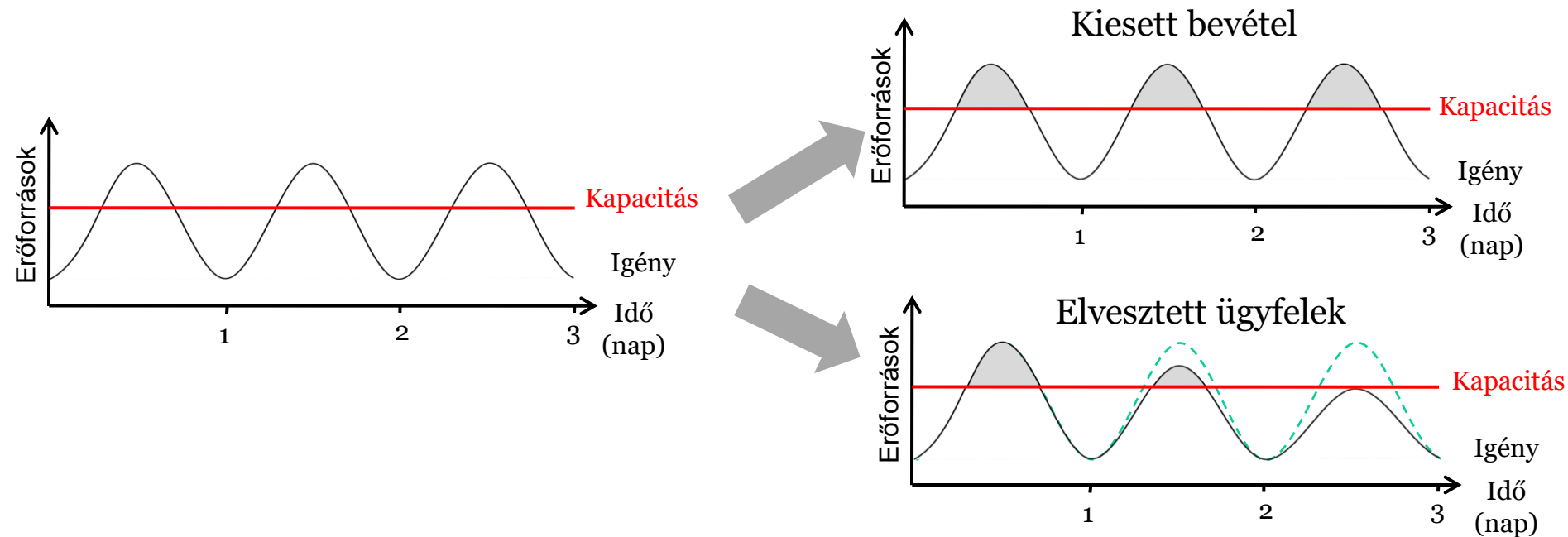
Igény szerinti kiszolgálás és csúcs-igény



Túlbiztosítás = alul kihasználtság



Alultervezés hátrányai



Cloud a változó (növekvő) igények számára

Erőforrásigény és felhasználás nagy adatigényű IT cégeknél

Google 1997



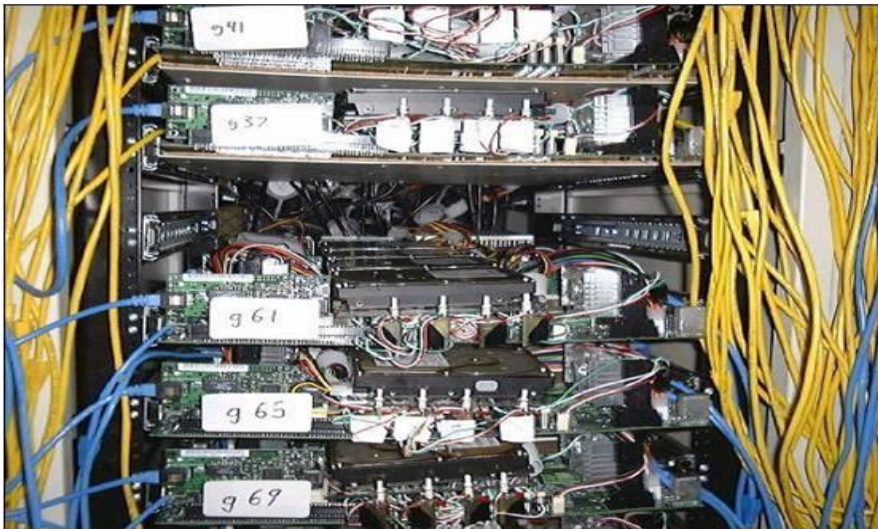
Data, Data, Data

“...**Storage space** must be used efficiently to store indices and, optionally, the documents themselves. The indexing system must process **hundreds of gigabytes** of data efficiently...”

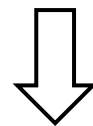
The Anatomy of a Large-Scale Hypertextual Web Search Engine

Sergey Brin and Lawrence Page

Google 2001



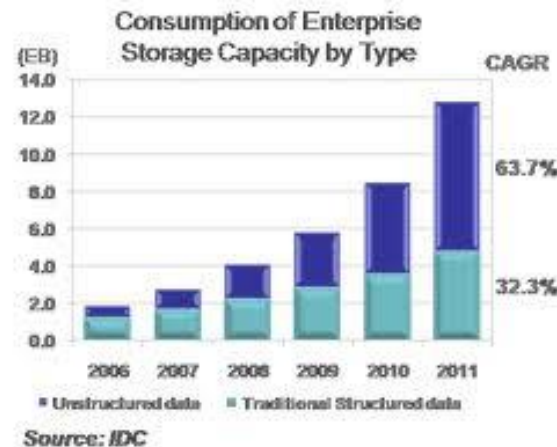
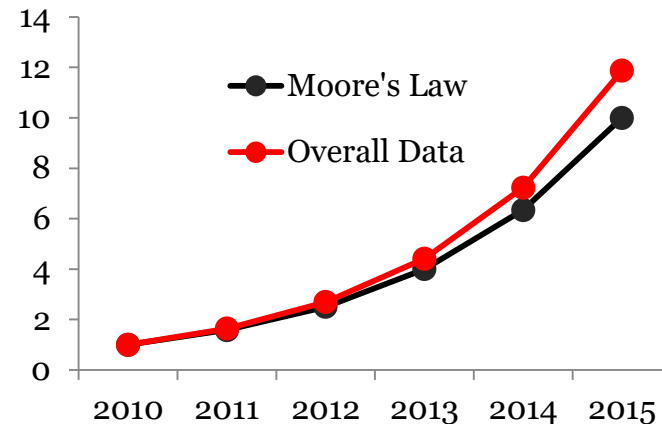
- Commodity CPUs
- Lots of disks
- Low bandwidth network



Cheap !

Datacenter evolution

- Data Volume continuously increases
 - Unconstrained growth
 - Current systems don't scale
- Typical large IT Co. data consumption
 - Facebook's daily logs: 60 TB
 - 1000 genomes project: 200 TB
 - Google web index: 10+ PB



Datacenter Evolution

Capacity:
~10000 machines



Google data centers
The Dalles, OR, USA

Bandwidth:
12-24 disks per node

Latency:
256GB RAM cache

Datacenters → Cloud Computing

Above the Clouds: A Berkeley View of Cloud Computing

Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz,
Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia
(Comments should be addressed to abovetheclouds@cs.berkeley.edu)



UC Berkeley Reliable Adaptive Distributed Systems Laboratory *
<http://radlab.cs.berkeley.edu/>

“...long-held dream of computing as a utility...”

Adatközpontú felhasználói igények

Data, data, data, ...

Our Data-driven World

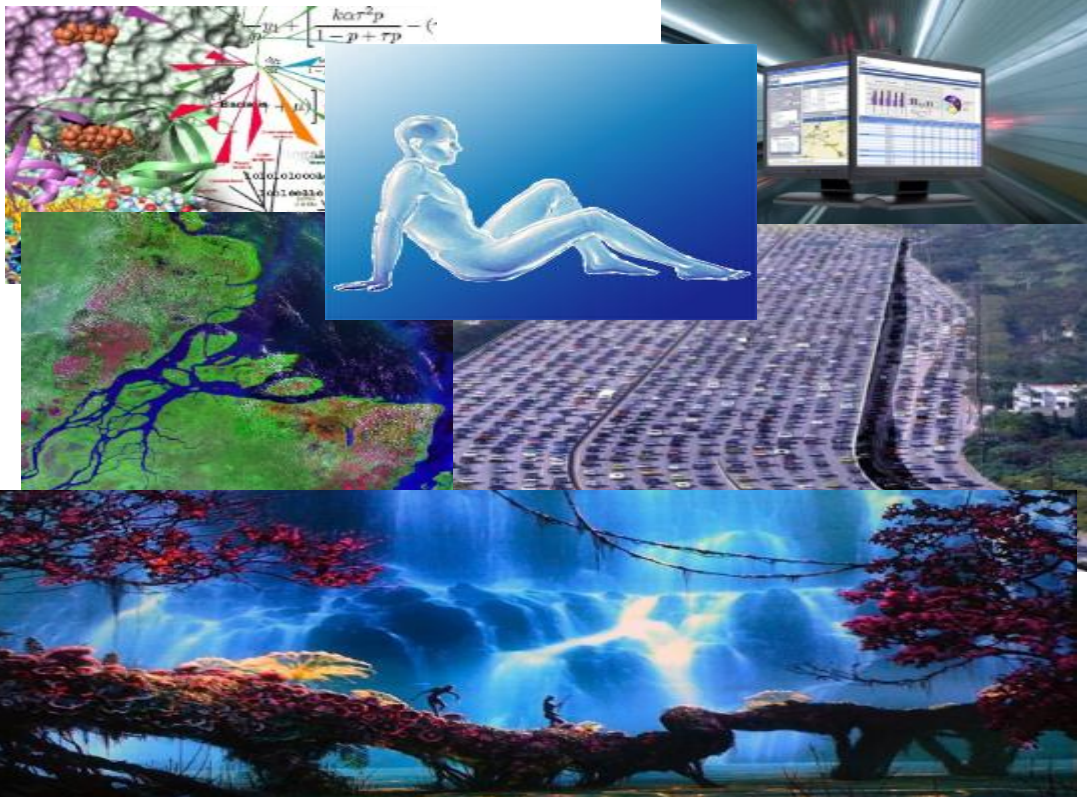
- Science
 - Data bases from astronomy, genomics, environmental data, transportation data, ...
- Humanities and Social Sciences
 - Scanned books, historical documents, social interactions data, ...
- Business & Commerce
 - Corporate sales, stock market transactions, census, airline traffic, ...
- Entertainment
 - Internet images, Hollywood movies, MP3 files, ...
- Medicine
 - MRI & CT scans, patient records, ...

Data-rich World

- Data capture and collection:
 - Highly instrumented environment
 - Sensors and Smart Devices
 - Network
- Data storage:
 - Seagate 1 TB Barracuda @ \$72.95 from Amazon.com (73¢/GB)



What can we do with this wealth?



- What can we do?
 - Scientific breakthroughs
 - Business process efficiencies
 - Realistic special effects
 - Improve quality-of-life: healthcare, transportation, environmental disasters, daily life, ...

- Could We Do More?
 - YES: but need **major advances in our capability to analyze this data**

Big data analytics a felhőben

Cloud and Big Data



“Can we outsource our IT software and hardware infrastructure?”

- Hosted Applications and services
- Pay-as-you-go model
- Scalability, fault-tolerance, elasticity, and self-manageability



“We have terabytes of click-stream data – what can we do with it?”

- Very large data repositories
- Complex analysis
- Distributed and parallel data processing

Data Warehousing, Data Analytics & Decision Support Systems

- Used to manage and control business
- Transactional Data: historical or point-in-time
- Optimized for inquiry rather than update
- Use of the system is loosely defined and can be ad-hoc
- Used by managers and analysts to understand the business and make judgments

Data Analytics in the Web Context

- Data capture at the user interaction level:
 - in contrast to the client transaction level in the Enterprise context
 - Click stream
- As a consequence the amount of data increases significantly
- Greater need to analyze such data to understand user behaviors

Big Data analytics use cases

- Targeted advertising / Clickstream analysis
- Security: anti-virus, fraud detection, image recognition
- Pattern matching / Recommendations
- Data warehousing / BI
- Bio-informatics (Genome analysis)
- Financial simulation (Monte Carlo simulation)
- File processing (resize jpegs, video encoding)
- Web indexing



Data Analytics in the Cloud

- Scalability to large data volumes:
 - Scan 100 TB on 1 node @ 50 MB/sec = 23 days
 - Scan on 1000-node cluster = 33 minutes
 - Divide-And-Conquer (i.e., data partitioning)
- Big Data on static data centers?
 - NO!
- Cost-efficiency:
 - Commodity nodes (cheap, but unreliable)
 - Commodity network
 - Automatic fault-tolerance (fewer administrators)
 - Easy to use (fewer programmers)

Use case: clickstream analysis

razorfish



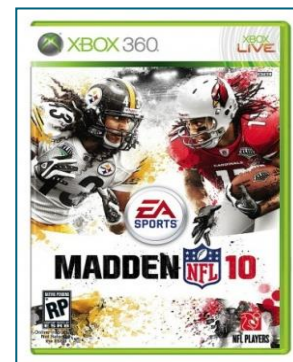
- Big Box Retailer came to Razorfish
 - 3.5 billion records
 - 71 million unique cookies
 - 1.7 million targeted ads required per day
- Goal: Improve Return on Ad Spend (ROAS)



User recently purchased a sports movie and is searching for video games

Targeted Ad

(1.7 Million per day)



Use case: clickstream analysis

razorfish



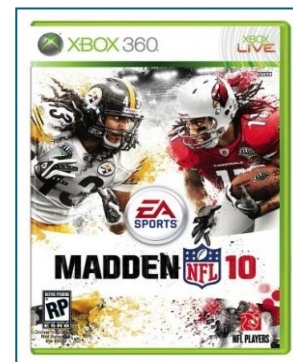
- Lots of experimentation to test the optimal solution
 - 100 nodes on-demand Amazon public cloud service
- Processing time dropped from 2+ days to 8 hours (with lots more data)
 - Increased Return On Ad Spend by 500%



User recently purchased a sports movie and is searching for video games

Targeted Ad

(1.7 Million per day)



Why now (instead of 1961)?

- Experience with very large datacenters
 - Unprecedented economies of scale
 - Transfer of risk
- Technology factors
 - Pervasive broadband Internet
 - Maturity in Virtualization Technology
- Business factors
 - Minimal capital expenditure
 - Pay-as-you-go billing model

The Joys of Real Hardware

Typical first year for a new cluster:

- ~0.5 **overheating** (power down most machines in <5 mins, ~1-2 days to recover)
- ~1 **PDU failure** (~500-1000 machines suddenly disappear, ~6 hours to come back)
- ~1 **rack-move** (plenty of warning, ~500-1000 machines powered down, ~6 hours)
- ~1 **network rewiring** (rolling ~5% of machines down over 2-day span)
- ~20 **rack failures** (40-80 machines instantly disappear, 1-6 hours to get back)
- ~5 **racks go wonky** (40-80 machines see 50% packetloss)
- ~8 **network maintenances** (4 might cause ~30-minute random connectivity losses)
- ~12 **router reloads** (takes out DNS and external vips for a couple minutes)
- ~3 **router failures** (have to immediately pull traffic for an hour)
- ~dozens of minor **30-second blips for dns**
- ~1000 **individual machine failures**
- ~thousands of **hard drive failures**
- slow disks, bad memory, misconfigured machines, flaky machines, etc.**

Long distance links: **wild dogs, sharks, dead horses, drunken hunters, etc.**

Designs, Lessons and Advice from
Building Large Distributed Systems,
Jeff Dean, Google

Összefoglalás

- 1960-as évektől kezdve a számítógép hálózatoknak fontos szerepet szántak az erőforrás-megosztásban
- 1990-as évek közepéig párhuzamos számítás
- 2000-es évektől igény szerinti erőforrás használat
- Adatközpontú analitika
- Cloud, Big Data technológiák