How did Netflix Beat blockbuster? Database Systems

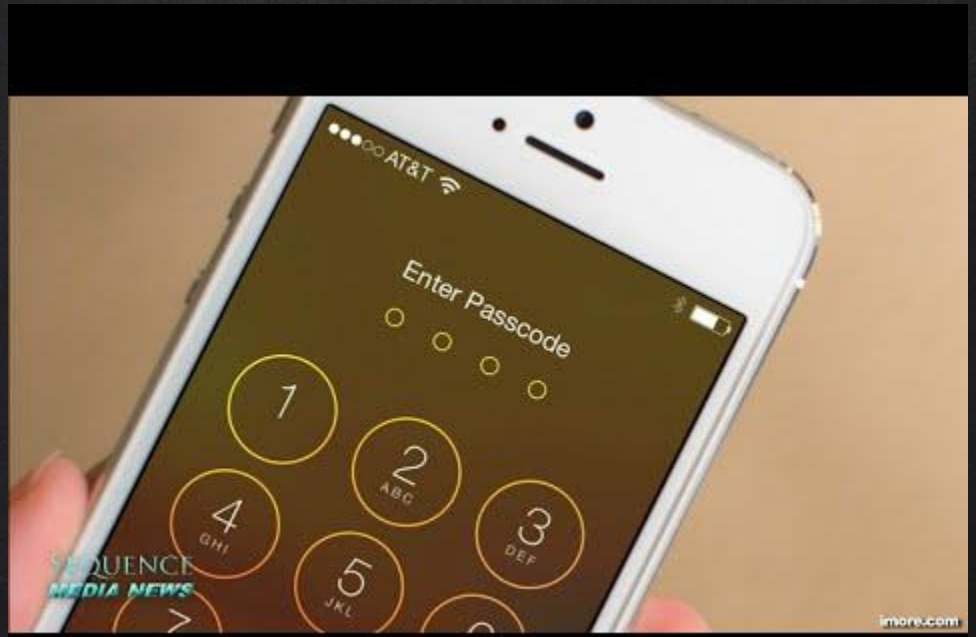
Announcement

- Quiz 5: No laptop/calculator needed/allowed
- Bonus 4 Due on 3/31

Exciting Tech News now and in the past week

Security, Twitter, AI

FBI hacked into Iphone without apple's help



"The FBI has now successfully retrieved the data stored on the San Bernardino terrorist's iPhone and therefore no longer requires the assistance from Apple required by this Court Order," DOJ spokeswoman Melanie Newman said in a statement, March 29, 2016.

Microsoft's new Chatbot TAY

https://www.youtube.com/watch?v=YN4wvTeVHxs

Gerry
@geraldmellor

Follow

"Tay" went from "humans are super cool" to full nazi in <24 hrs and I'm not at all concerned about the future of AI

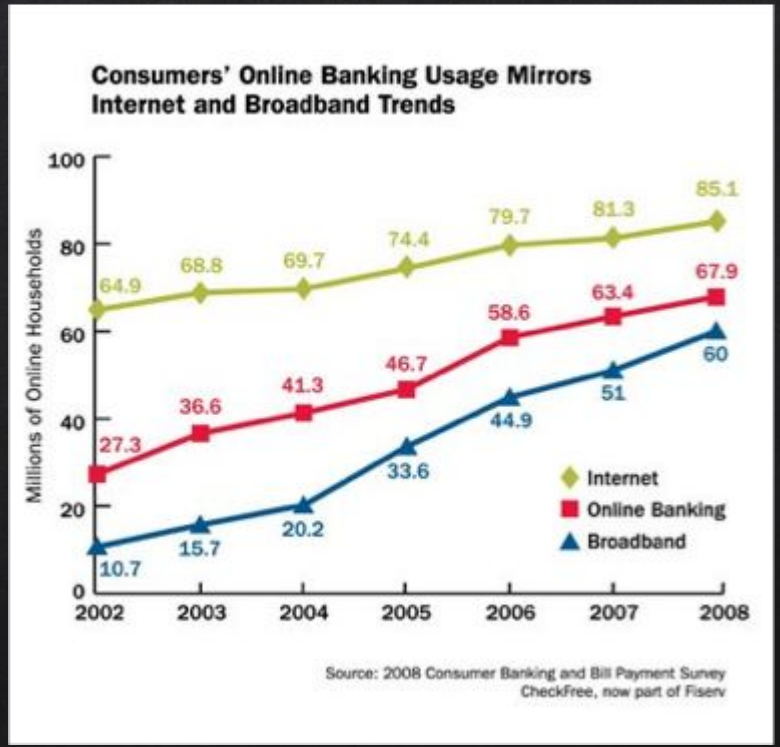
11:56 PM - 23 Mar 2016

13,324 9,872

Database Systems

Think About Online Banking

What computer technologies make this possible?



Online Banking

- Internet
- Secure communications
- Databases
  - Storage
    - Reliable
    - Efficient

Where are Databases Used?

- Medical records
- Student records
- Banking
- Websites
- Consumer, Home, Personal
  - E-commerce
  - Movie streaming services

Database management Systems

- Database
  - Organized collection of data
  - Support modifying and querying the data
- Database management system
  - DBMS
  - Control the creation, maintenance and use of a database
  - Allows diff users to concurrently access the same database

Data Structure

Can't we just sort the data?

- Binary search finds items quickly

Data Structure

Can't we just sort the data?

- Binary search finds items quickly

Want to view data in different ways

- Online goods presented by
  - Low to high price
  - High ratings (movies, etc.)
  - Just arrived
- Would need to maintain different sorted lists for each desired view
- Look at something like Newegg laptop search

The Relational Model

The data items and their relationships are organized into tables:

MovieId	Title	Genre	Rating
101	Sixth Sense, The	thriller horror	PG-13
102	Back to the Future	comedy adventure	PG
103	Monsters, Inc.	animation comedy	G
104	Field of Dreams	fantasy drama	PG
105	Alien	sci-fi horror	R
106	Unbreakable	thriller	PG-13
107	X-Men	action sci-fi	PG-13
5022	Elizabeth	drama period	R
5793	Independence Day	action sci-fi	PG-13
7442	Platoon	action drama war	R

Database Schema

- Schema
  - Describes the structure of the database
  - Specifies what data can enter the database

```
Movie (MovieId:key, Title, Genre, Rating)
```

Database Schema

Table

- A collection of entries (records) in a database
- Follows the form of the schema

Database Schema

- Record
  - A set of related data
    - A person's name and phone #
  - A row in a database table
- Field
  - A single piece of data in the table
  - Sometimes means the whole column
- A key field is a unique identifier for a record in a table
  - Why would movies need a number ID as a key field?

What would be a good schema for a movie rental business?

Database Reliability

Database Reliability

- Database must be extremely reliable
  - Banks
- Deal with
  - Computer crashes
  - Hard drive failure
  - Power outages
  - Data

Transactions

A database transaction is a collection of actions

- Must all be completed for the database to remain consistent
- Considered a reliable unit of work
  - An atomic action

Database Change Example

- David's bank account
  - Checking: $500
  - Saving: $200
- Transfer $100 from checking to savings
  - Two actions needed
    - Subtract $100 from checking
    - Add $100 to savings

Database Change Example

- Start state
  - Checking: $500
  - Saving: $200
- Action: subtract $100 from checking
  - Checking: $400
  - Saving: $200
- Action: add $100 to saving
  - Checking: $400
  - Saving: $300

Alternative Change Example

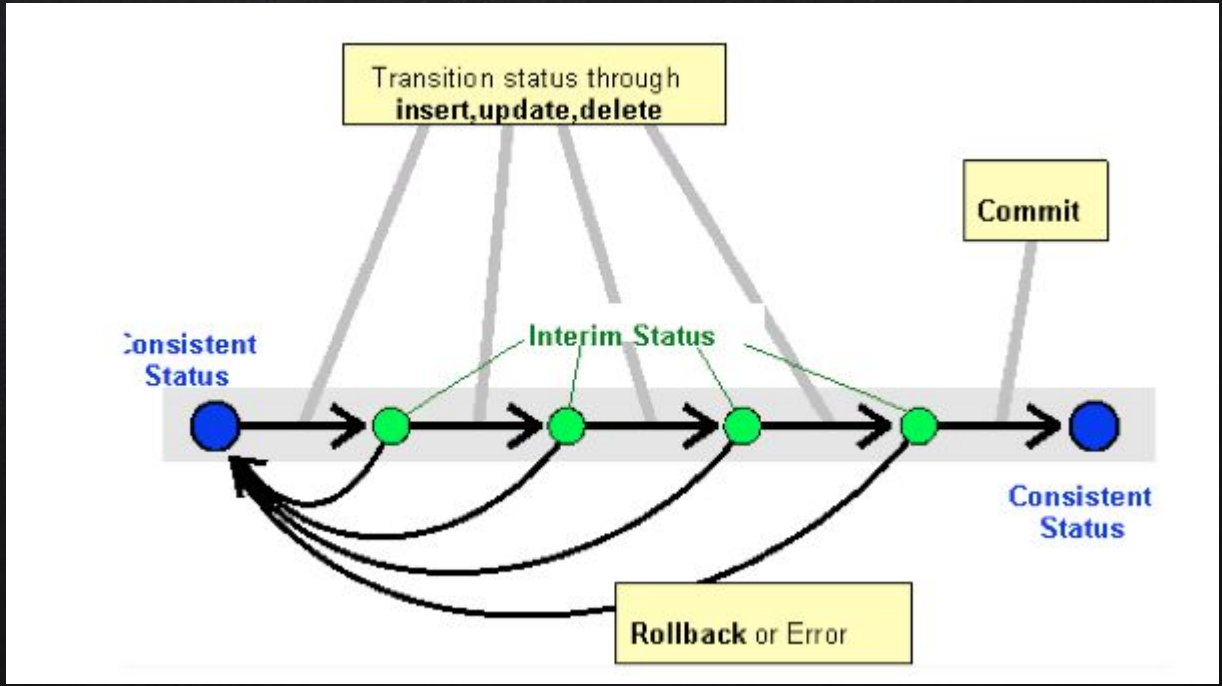
- Start state
  - Checking: $500
  - Saving: $200
- Action: subtract $100 from checking
  - Checking: $400
  - Saving: $200
- Power failure and reboot!
  - Checking: $400
  - Saving: $200 --- Incorrect final state

Transactions

- Provide 2 safeguards
  - Changes specify start and end state
  - All elements in a transaction must complete or it is tried again
- Start Transaction
  - Change checking from $500 to $400
  - Change savings from $200 to $300
- End Transaction

Transactions and Failures

- Parts of a transaction can repeat with no harm
  - Change checking from $500 to $400
- A failed transaction can roll-back to the original state to start again – needed if error



Database Replication

- A database on one hard drive provides little reliability
- Replicated databases
  - Multiple synchronized copies
    - Geographically distributed (why?)
  - Not just backed up
    - That just provides a copy current to when the backup occurred

Viewing Data in a DB

- Database allow different views of the data
  - The user asks questions of the DB and gets specific answers
  - How is this different from web search?
- Rather than having to program specific views of data and maintain those programs, the DB generates these views as needed

Database Query Language

DB Query Language

- DBMS use query languages to retrieve data from a database
- An example is
  - Structured Query Language (SQL)
    - SQL or sequel
  - A small programming language
  - Also used to create and modify databases
- We will focus on queries

SQL

- The main query statement is SELECT
  - Select records from a table
  - Has the form:

```
select attribute-list from table-list where conditions
```

  - Attribute-list: which columns are returned
  - Table-list: which tables are searched
  - Conditions: restricts the returned data
    - For example, some range of dates

SQL Example

Movie

MovieId	Title	Genre	Rating
101	Sixth Sense, The	thriller horror	PG-13
102	Back to the Future	comedy adventure	PG
103	Monsters, Inc.	animation comedy	G
104	Field of Dreams	fantasy drama	PG
105	Alien	sci-fi horror	R
106	Unbreakable	thriller	PG-13
107	X-Men	action sci-fi	PG-13
5022	Elizabeth	drama period	R
5793	Independence Day	action sci-fi	PG-13
7442	Platoon	action drama war	R

select Title from Movie where Rating = 'PG'

The result is a list of all titles that have a PG rating...

SQL Example

Customer

CustomerId	Name	Address	CreditCardNumber
101	Dennis Cook	123 Main Street	2736 2371 2344 0382
102	Doug Nickle	456 Second Ave	7362 7486 5957 3638
103	Randy Wolf	789 Elm Street	4253 4773 6252 4436
104	Amy Stevens	321 Yellow Brick Road	9876 5432 1234 5678
105	Robert Person	654 Lois Lane	1122 3344 5566 7788
106	David Coggin	987 Broadway	8473 9687 4847 3784
107	Susan Klaton	345 Easy Street	2435 4332 1567 3232

select Name, Address from Customer

The result is the name and address of all customers...

The where clause can be omitted if there are no special restrictions.

SQL Example

Movie

MovieId	Title	Genre	Rating
101	Sixth Sense, The	thriller horror	PG-13
102	Back to the Future	comedy adventure	PG
103	Monsters, Inc.	animation comedy	G
104	Field of Dreams	fantasy drama	PG
105	Alien	sci-fi horror	R
106	Unbreakable	thriller	PG-13
107	X-Men	action sci-fi	PG-13
5022	Elizabeth	drama period	R
5793	Independence Day	action sci-fi	PG-13
7442	Platoon	action drama war	R

Select * from Movie where Genre like '%action%'

The result is all attributes of records in which the genre contains the word "action".

Use * to denote that all attributes should be returned. % matches any words

SQL Example

Movie

MovieId	Title	Genre	Rating
101	Sixth Sense, The	thriller horror	PG-13
102	Back to the Future	comedy adventure	PG
103	Monsters, Inc.	animation comedy	G
104	Field of Dreams	fantasy drama	PG
105	Alien	sci-fi horror	R
106	Unbreakable	thriller	PG-13
107	X-Men	action sci-fi	PG-13
5022	Elizabeth	drama period	R
5793	Independence Day	action sci-fi	PG-13
7442	Platoon	action drama war	R

select * from Movie where Rating = 'R' order by Title

The result is all attributes of R-rated movies sorted by title.

Discussion

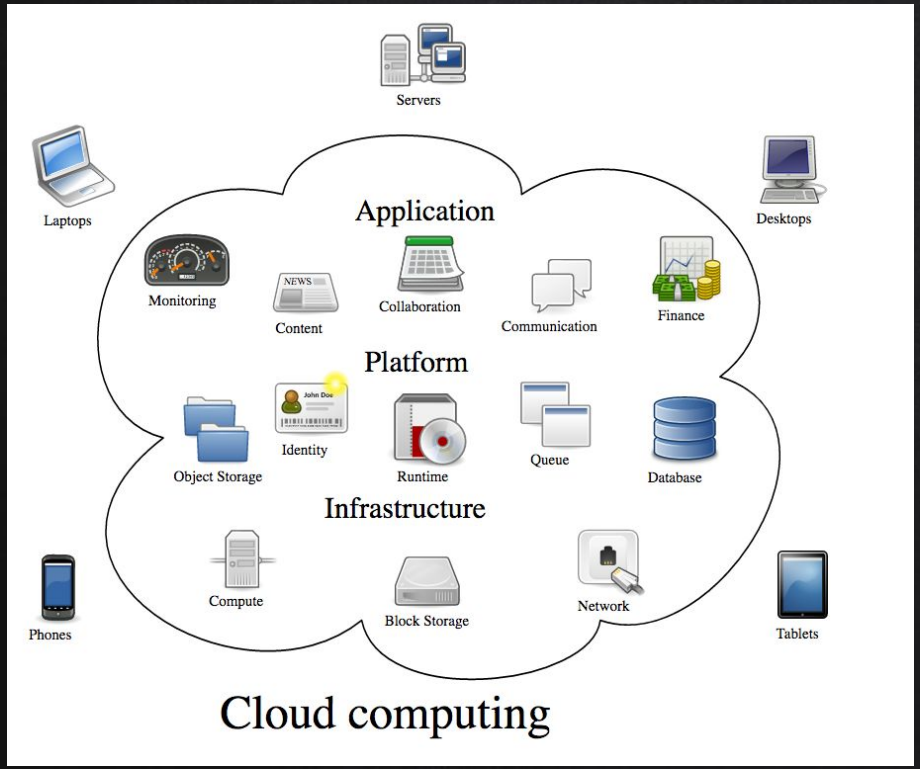
- These are pretty simple queries
  - Look at complicated ones next time
- Learn by example
  - Easy to make queries if you can find a similar one that works
- Working with databases is a whole subfield of computer science
  - Needs specialized training / skills

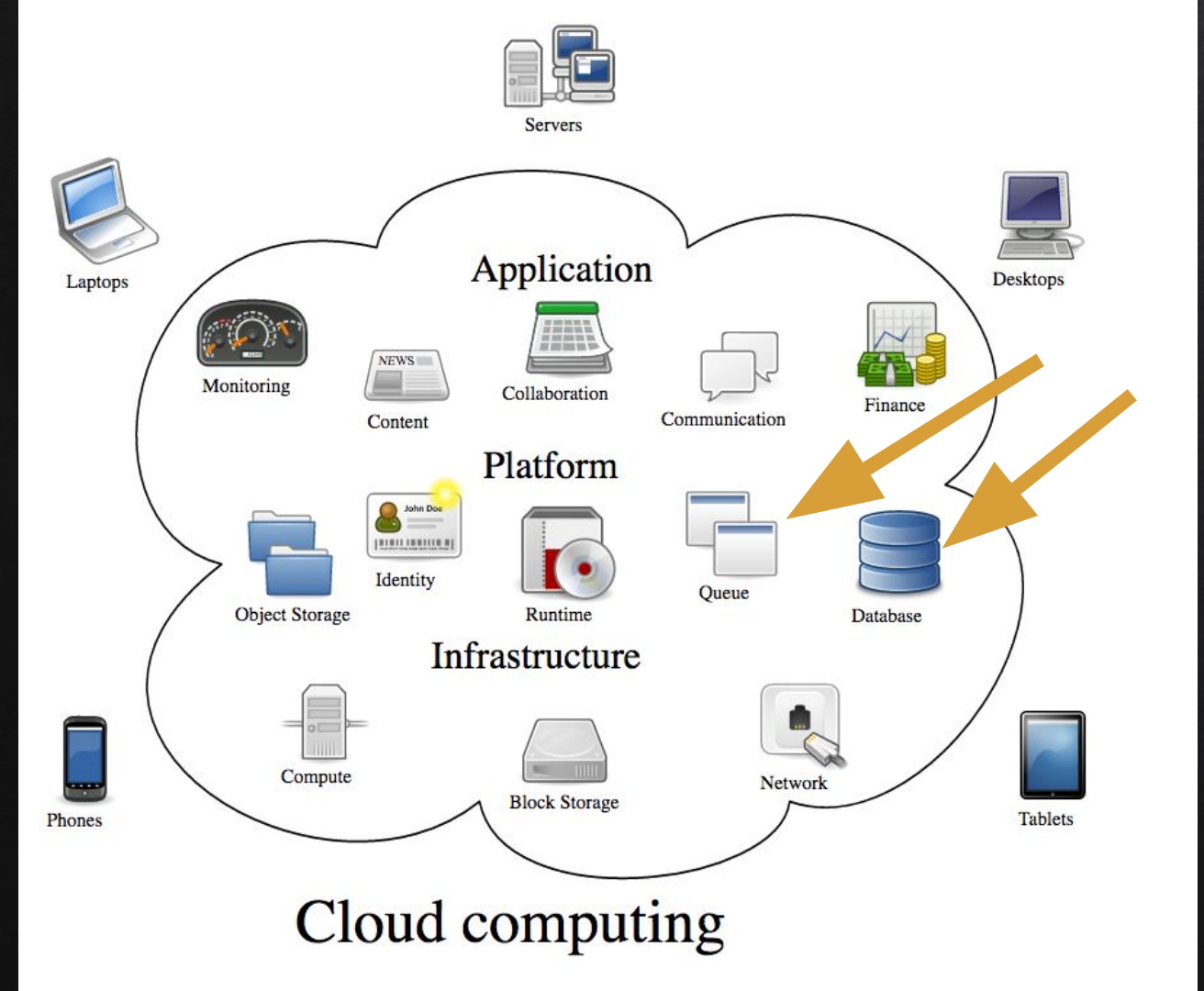
Netflix
Cloud computing
Cloud storage

https://www.youtube.com/watch?v=lQGHsBOZJBw

Cloud Computing

- For a user, the network elements representing the provider-rendered services are invisible, as if obscured by a cloud.

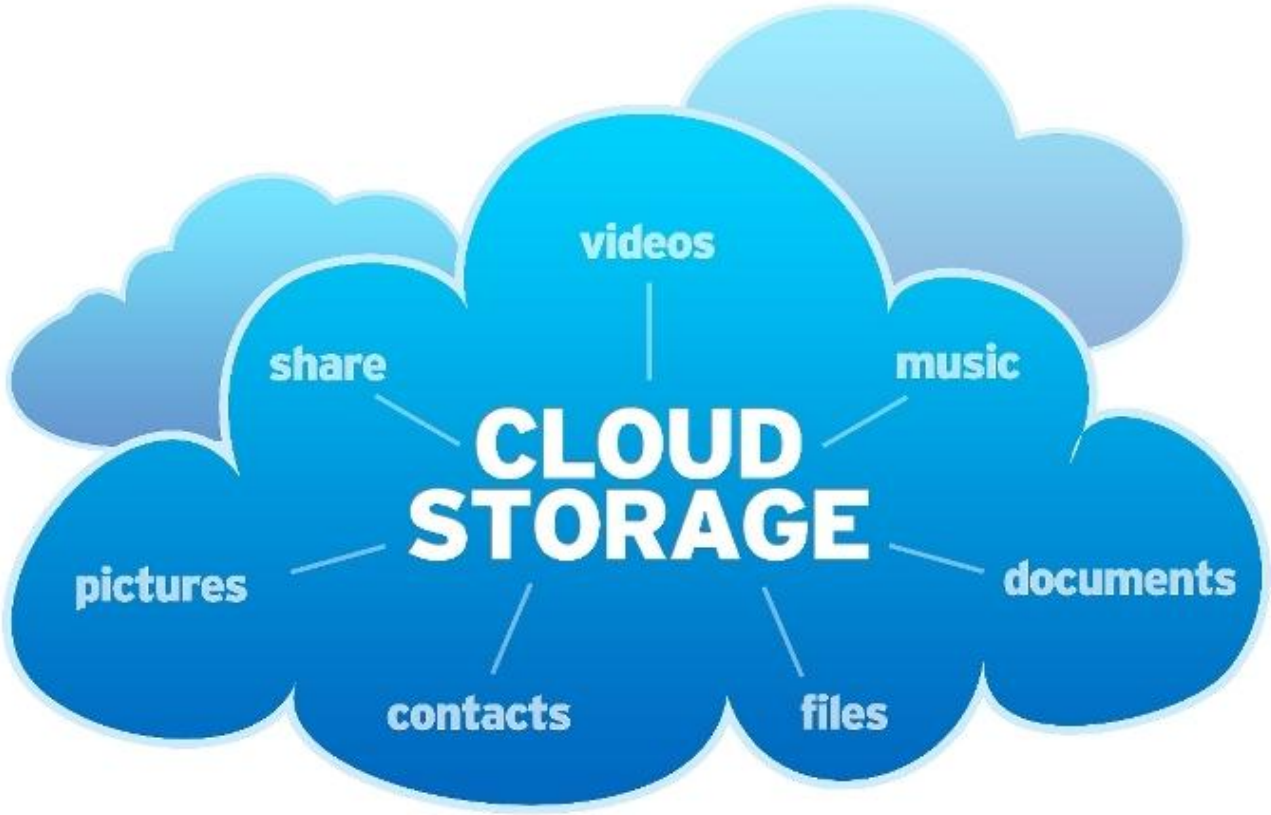




By Sam Johnston [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0)], via Wikimedia Commons

Cloud Computing

- **On-demand computing**
- Internet-based computing
- Provide shared processing resources and data to computers and other devices on demand.
- Enabling ubiquitous, on-demand access to a shared pool of configurable computing resources
- Minimal management effort
- https://en.wikipedia.org/wiki/Cloud_computing



http://www.mixturecloud.com/wp-content/uploads/2016/01/cloud1.jpg

http://www.multipelife.com/wp-content/uploads/2015/09/cloud-storage.jpg
https://smidcloud.com/docs/how-to-choose-a-safe-Cloud_Storage-service.pdf

Cloud Storage

A model of data storage in which the digital data is stored in logical pools, the physical storage spans multiple servers and often locations, and the physical environment is typically owned and managed by a hosting company.

Read more: https://en.wikipedia.org/wiki/Cloud_storage

NoSQL

- "non SQL", "non relational", "Not only SQL"
- DB provides a mechanism for storage and retrieval of data that is beyond tabular relations used in relational databases
- “Beyond tables”
- Web 2.0 companies: Facebook, Google and Amazon.com
- Increasingly used in big data and real-time web applications

NoSQL and NetFlix

- Netflix + cloud
- persist and query data within highly distributed infrastructure.
- build fast, fault tolerant systems at Internet scale.
- move beyond the constraints of the traditional relational model.
- high availability (a.k.a. better customer experience) usually trumps strong consistency.
- little room for vertical scalability or single points of failure.

Read more: http://techblog.netflix.com/2011/01/nosql-at-netflix.html

“And while it is not easy to re-architect your systems to not run join queries, or not rely on read-after-write consistency (hey, just cache the value in your app!), we have found ourselves braving the new frontier of NoSQL distributed databases.”

2011, NetFlix, Yury Izrailevsky

Quiz 5

THANKS!

Any questions?

You can find me at
beiwang@sci.utah.edu

http://www.sci.utah.edu/~beiwang/teaching/cs1060.html

Credits

Special thanks to all the people who made and released these awesome resources for free:

- Presentation template by SlidesCarnival
- Photographs by Unsplash