Youtube Lecture Slides

Slide: s2.png

« A video search engine is a web-based search engine which crawls the web primarily for video content.

? YouTube is not strictly a video search engine as it does not crawl the web looking for video content

* The indexing of video content is normally done by acquiring meta-data associated with the video, e.g.

? Author, title, creation date, duration, coding quality, tags, description

? Other aspects of video recognition are subtitles (using formats STR or SUB) and transcription (using format TTXT)

* The ranking of videos under a query is generally done using:

? Relevance: using metadata and user preferences

? Ordered by date of upload

? Ordered by number of views So indexing and ranking

? Ordered by duration are a lot simpler than

for document search

? Ordered by user rating engines

Slide: s3.png

* Those no longer existing

? CastTV was a Web-wide video search engine that was founded in 2006

* No longer active

? Munax released their first version all-content search engine in 2005 and powers both nationwide and worldwide search engines with video search

° http://www.munax.com/ no longer active

? ScienceStage is an integrated universal search engine for science-oriented videos. All videos are also semantically matched to millions of research documents from open-access databases.

? No longer active

* A few remain

? Bing does crawl for videos, see https://www.bing.com/videos/

? blinkx (renamed as RhythmOne) was launched in 2004 and uses speech recognition and visual analysis to process downloaded video rather than rely on metadata alone

© Attp://www.blinkxtv.com/ now redirects to 3 60Drily.com

Slide: s4.png

* Largely because of the large file sizes involved, video hosting is highly concentrated on a fairly small number of websites

? vimeo.com, first to support HD video, focuses on short, arty, films

? vevo.com, a joint venture of Universal Music Group, Sony Music Entertainment and Warner Music Group

? dailymotion.com, owned by Vivendi, hosts high quality videos

* Most of these websites which host video allow their videos to,be embedded on other websites

* YouTube.com has become the defacto site for uploading videos

* ? It is legal to crawl YouTube, see their Terms of Service, www. youtube.com/static?template=terms

* ?3, You are not allowed to access the Service using any automated means (such as robots, botnets or scrapers) except (a) in the case of public search engines,.in accordance with YouTube?s robots.txt file; or (b) with YouTube?s prior written permission;?

Slide: s5.png

¢ Hulu is an America subscription video on demand service jointly owned by Walt Disney, 21? Century Fox, Comcast, and Time Warner

? In December 2017, Disney acquired Fox's partial ownership, giving it a majority stake; other owners include Comcast

* Netflix is an American subscription video on demand service, that originally delivered DVDs;

? They develop their own content as well as offering content from major film distributors

¢« Amazon Prime is an American subscription video on demand service offering television and file shows for rent or purchase

¢ Disney+ a recent entry

* There are many others: XtremeHD, Sling TV, Apple TV+, HBO Max, Acorn TV, etc

¢ Entertainment

Slide: s6.png

+ Subtitles: there are two formats, one for subtitles and one for transcripts

? There are three main types of video subtitling services:

1. open caption: burned into the video

2. closed caption: can be turned on/off, generally at the bottom of the screen

3. SDH (Subtitles for the Deaf and Hard of Hearing): similar to closed-caption, but includes words describing actions or moods

? SRT or SUB for subtitles

* SRT. srt) stands for ?SubRip Subtitle? file, and it's the most common subtitle/caption file format. It is a text format

? TYXT for transcripts

+ Speech Recognition, used to extract phrases from audio transcripts for better indexing

Youtube Lecture Slides

? Gaudi, Google Audio Indexing uses voice recognition to locate the exact spot where words are spoken

?_ https://www.searchenginejournal.com/google-audio-search-will-it-ever-be-possible/397129/

? Text Recognition: uses OCR on video slides to detect words,

? e.g. TalkMiner System, see https://www.youtube.com/watch?v=7N6L_m9LywM.

Slide: s7.png

* YouTube is an American video
hosting website headquartered in San
Bruno, California, created by three
former PayPal employees: Chad Hurley,
Steve Chen, Jawed Karim in February
2005.
* In November 2006, it was bought by ' 9 cen one a
Google for US$1.65 billion ey eens -
* In 2020 Google announced that
YouTube generated revenue of $19.8 sem at | a
billion
* The site allows users to upload, view, rate,
share, add to favorites, report and comment on
videos sama 2 12:13
. In January 2022, the website was ranked as the cow ccm once
second most popular site by Alexa Internet, ~? ? _?] =
web traffic analysis company (now owned by
Amazon)
??See also > For details see Related Articles page, Mar 2020
https://en.wikipedia.org/wiki/List_of_mo
st_popular_websites

Slide: s8.png

YouTube as a Search Engine
* YouTube - The 2nd vstibe Te
Largest Search
Engine (cite:Infographic)
mushroomnetworks.com) a
it}
* YouTube processes more YouTube is Transforming
wa the Way We Discover It processes more than
than 3 billion searches a Nrfor ciation Vie the K 4 billi

month ws illion

. searches a month

to he :

¢ It's bigger than Bing, ME Founded in 2005 by

Yahoo!, Ask and AOL 2005 | tte ? PayPal Bought in

: employees

combined! me Google 2006

OOBIE py Google for

. http://www.mushroomnetworks.com/ $1.65 billion

infographics/youtube---the-2nd- 5

largest-search-engine-infographic boing Aol.

ey /o-o

Bigger than Bing, Yahoo, Fastest growing video sharing

Ask and AOL combined website in the world at the moment

Slide: s9.png

WOU Tube Traffic - Some Facts

¢ As of 2021:

? 60 hours of video are uploaded every minute, or one hour of

video is uploaded to YouTube every second.

Over 4 billion videos are viewed a day

Over 800 million unique users visit YouTube each month

? Over 3 billion hours of video are watched each month on

YouTube

? More video is uploaded to YouTube in one month than the 3

major US networks created in 60 years

? 70% of YouTube traffic comes from outside the US

? YouTube is localized in 39 countries and across 54 languages

It is estimated that YouTube holds 1 sextillion gigabytes of data

? _ https://www.quora.com/What-is-the-total-size-storage-capacity-of-YouTube-and-at-what-rate-is-it-

increasing-How-is-Google-keeping-up-with-the-increasing-demands-of-Y outube%E2%80%99s-capacity-

given-that-thousands-of-videos-are-uploaded-every-day

Slide: s10.png

¢ Since crawling, indexing and ranking are not big challenges for

YouTube, what are the major hurdles

1. What video formats are acceptable

? For uploading

? For downloading

2. How are videos to be displayed on: desktops, iPhones, iPads, Android

devices, etc

3. How does YouTube distribute videos worldwide

A content distribution network (CDN)

4. How does YouTube monetize its website?

? YouTube? s ContentID system

5. How does YouTube keep users watching

? The YouTube Recommendation System

Slide: s11.png

* In order to upload a video you must be a registered user

* Inaddition YouTube offers a special type of account called a channel; channels

include

thumbnails of videos you've uploaded,

members to whom you?ve subscribed,

videos from other members you've picked as favorites,

lists of b hi friend:

Isis oF members Who are your iriend?s, With 1 million subscribers, a YouTuber

your subscribers, and will make between $300,000 ? $2 million

* Biggest YouTube Channels as of 2021 To be in the top 1000 YouTubers you

must haye ~1.8 million subscribers

As of 09/2020, there are more than

2000 YouTubers with over a million

subscribers

https://www.statista.com/statistics/277758/most-popular-youtube-

channels-ranked-by-subscribers/

~ vy yeeput Ellis Horowitz, 2011-2022

Slide: s12.png

© 1of 1 uploaded YouTube x

@ Secure _https://www.youtube.com;

pps [| CSCI572 Home Page [| CSCI571 Home Page ['| CSCI 351 Home Page [\| Ellis Horowitz? Hom.

Youtgli3 Search

YouTube

captures:

PROCESSING DONE

FEB cick Pubs to make your video ive

Name

ae LennonBithday

Description Upload status:

ogy Description

Tags Your veo wile est

hips/youu ber S61

Note: YouTube

Tour videos wiTP

: ; faster oo

immediately tore formation vat oor

: Mp Corte,

assigns a UR " p>) voc Tenants ©

Note: YouTube we

suggests possible _ =

thumbnails

Video on How to Upload a Video

Bi computer science... [} DynONS -- Host Se

Ellis

Bother Bookmarks

* a@

Public ,

Ao sr on goo

++ Add o playlist

?custom thumbnail

Video Manager + Add more videos

https://support.google.com/youtube/answer/57407

Slide: s13.png

@ O&O Bi 1oft uploaded- YouTube x Ellis

CO @ Secure _https://wwwyoutube.com

pps [) CSCI572 Home Page [) CSCI571 Home Page [) CSCI351 Home Page [) Ellis Horowitz? dm... J Computer Science... [} Dy

= You Search

PROCESSING DONE

EB click ?Publish? to make youssideo live. Draft saved

Basie info Translations ?Advanced settings

Translate into (0)

Upload status: Select language Select language .

brtpsi/youtube/kX014x5G1RW

LennonBirthday

Video / Audio quality:

fasterifyou encode into a

streamable le format. For

?more information, vist ou

Help Center.

Get professional translation @

Video Manager + Add more videos

Slide: s14.png

iversity of Southern Californi

© © er tot uploaded - Youtube

> Q & Secure https:/jwww.youtube.com,

p= () CSCI872 Home Page) CSCI871 Home Page [CSCI381 Home Page (Elis Horomit? Hom.

YoulD

YouTube allows

the creator to A

specify: icine Comme

ation comments,

eras show AL >

Sort by Ten comments ~

io quality

License and ownership

Syndication =?

Caption auID

Embedding Sytem

Age restrictions

Categories

[Ziusers can view retings for this video

License and rights ownership O

Caption certiation ©

Distribution options

low embedsing @

[Znotity subscribers ©

Enable age restriction ©

[computer Science... [) DynONS ? Host Se. ote: Bookmark

+ 00

Category

Video location

Video language

?Community contributions

?Allow viewers to contribute translated tle, descriptions,

and subtiiesioc ©

Recording date

Today

Video statistics

Youtube Lecture Slides

[ZiMake video statistics onthe watch page publicly vsble ©

30 video

This videos 30

Content decaration

?This vdeo contains pad promotion suchas paid product

placement, sponsorships or endorsement ©

Slide: s15.png

Business Model: Ads, Ads, Ads

Sample YouTube Search Results for Katy Perry

= Yuu katy perry * a

Katy Perry

Music

4 . Lipstick By Katy Pe

First result is an Ad yeoman

2"4 and 4" results are

stored at Vevo

Katy Perry - Chained To The Rhythm (Lyric Video)

ft. Skip Marley Top Tracks Albu

34 and 5" results are Dirac ? vow ?

links to a Katy Perry

channel with 106 videos

Last Friday Night (6.1)

KatyPerryVEVO Few 7

: 7 Hot n Cold 444

To the right is a mix ?? :

of Katy Perry songs he One That Got Away ?

and some ?related?

artists

Katy Perry - Topic

?Subecrbe

This Is How We Do

View all

Related Artists

© We

aylor Swift LadyGaga_?_Russell Brand

Slide: s16.png

YouTube Search Results

PT ute) ?computer science

Begins with an ad ? Computer science eduction: why does it suck so

much and what if t didn? | Ashley Gavin |

Technology For Students

The next 4 results are \ i

ordered by the number of ry

views: 420,004,

369,979, 228,004

Lec 1 | MIT 6,00 Introduction to Computer Science

?and Programming, Fall 2008

Lecture 0 - Introduction to Computer Science |

Subsequent listings are

a mixture of highly

viewed videos, but older, q] Question: How Important is Math in a Computer

e.g. Lec 1 MIT has BLM Science Degree?

a: . . Computer Science a good major?

3 million+ views but is .

7 years old

=| Computer Science Explained in less then 3

minutes

It is not obvious how the Computer science is for everyone | Hadi Partov | |

ii i ?TEDxRainie :

ranking was determined Daan

Computer Science Tutor

\COMPUTERY

SCIENCE vo

?Vlog: What to expect in @ Computer Science

course

Computer science education: why does it suck so

much and what if it didnt? | Ashley Gavin |

[ACS Degree - What to Expect

Slide: s17.png

University of Southern Ca

@Boulube algorithms Qe

* During a search

YouTube provides once

filters for users to

refine their search:

UPLOAD DATE

TYPE

Youtube Lecture Slides

DURATION ; sotsoioine cut coecunp one

FEATURES ALGORITHMS.

SORT BY i

MIT 6.006 Introduction to Algorithms, Fall 2011

John MacCormick?s Nine Algorithms That Changed the Future

Slide: s18.png

* YouTube uses the following metrics for ranking search results:

1. Meta Data

? video titles, descriptions and tags are core ranking factors

? include links to a website and social profiles

2. Video Quality

? HD ranks higher than low quality videos

3. Number of views, likes, shares and links

Subtitles and Closed Captions

? captions are crawled by the YouTube search engine and used for ranking

* What is not known is how YouTube weights the individual

factors to make up their final ranking

Slide: s19.png

* YouTube Upload Characteristics

YouTube supports 8 video formats for uploading: MOV, MP4 (MPEG4), AVI,
WMV, FLV, 3GP, MPEGPS, WebM

Aspect Ratio: the standard aspect ratios are: 4:3 or 16:9. When the video is
uploaded to the site, YouTube will either leave it as-is (for 16:9) or add vertical
black bars (for 4:3)

The maximum file size you can upload to YouTube is 128GB.

By default, you can upload videos that are up to 15 minutes long, though
that can be extended

Many videos have a short life cycle, e.g. a recent tennis match that is soon
forgotten, however, there is no time limit for videos to remain on YouTube,
unless

* You delete the video.

* You delete your account.

* You violate copyright or community guidelines

Slide: s20.png

* Desktops/laptops

? Videos are played in your browser assuming it supports HTMLS

? This avoided the need to use Adobe Flash Player

¢ Smartphones

? YouTube apps exist for Android and iPhone devices

¢ There is no native support for running YouTube videos so a separate app is

required

? For YouTube's videos to run on Apple products YouTube's content had to be

transcoded into Apple's preferred video standard, H.264

* Other Devices

? Apple TV, Fire TV, iPod Touch,

? TiVo, PlayStation, Wii Game consoles,

? Xbox Live, Roku Players

? Google Chromecast

Slide: s21.png

YouTube Makes

School of Engineer oegmmendations to Retain Viewers

OB)oulut ?computer algorithms

* YouTube Search ?ents = o

Results Example About 521,000 results = FILTER

for query LGORITHMS,

?computer eo

. ? an

algorithms MIT 6.006 introduction to Algorithms, Fal 2011

« Assume we

choose the first

result

What's an algorithm? - David J. Malan

N=6 |

[=e aa

aa |

TIDES SIRIRNGSIT Data Sructues and Algorithms Complete torial Computer

ae Education for AI

to maximize watch time Modes

?Advanced Algorithms (COMPSCI 224), Lecture 1

Recommendations are made

K xs _S& __Concepts of Algorithm, Flow Chart & C Programming

https://www.nbcnews.com/tech/social-media/algorithms-take-over-youtube-s-recommendations-highlight-human-problem-n8675

Slide: s22.png

vies YouTube Recommendation Algorithm

Youtube Lecture Slides

Binds computer a = e

* Given the query

?computer algorithms?

followed by a selection,

YouTube makes

recommendations for

subsequent videos

FUNCTION eaawueen SUPRSEPTEWEER

TOYS & HOBBIES

¢ Recommendations account

for 60% of all video clicks

Slide: s23.png

¢ Association Rule Mining

? For each pair of videos v; v; compute co-visitation counts, i.e. they count how

often they were co-watched; if c;; is the co-visitation count, then relatedness is

defined as .

Cij

f (vi, U5)

where c; and c; are the total occurrence counts across all sessions for videos

vy, and v;. f(v; vj) is a normalization function that takes the global

popularity of both the seed video and the candidate video into account; e.g.

Si vp) = 64 * Gj

The set of related videos, R; for a given seed video v; is determined by taking the

top N candidate videos ranked by their scores r(v,, vj)

r(vi, v5) =

Related videos induce a directed graph over the set of videos, namely:

For each pair of videos (v;, v;), there is an edge e, from v; to v; iff vis in R;

For details see: The YouTube Recommendation System

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.434.9301 &rep=rep1 &type=pdf

Slide: s25.png

University of Southern Ca

A video rich snippet means that when nas

someone searches for something on Google, kB dian

you can have a small tiny video show up 2 : vo"

next to your result to let the user know that

particular result (yours) has a video to help =WISTIA pices Ereequent Vitec aeuep es Dormelae

Et Apps Yk Bookmarks [} CSCI571Home Page [} CSCIS72Home Page » (5) Other bookmarks

Google weeded out the video competition in

Web search by predominantly displaying

only video-rich snippets for YouTube videos

back in 2014.

Here is a graph outlining the percentage

share of video-rich snippets in Google; 7

91% are from YouTube

see

https://wistia.com/blog/where-did-my- iam ©

video-snippets-go

e.g. try ?tutorial on bitcoin?

Slide: s26.png

* Acontent distribution network (CDN) consists of a large set of content

servers and a means for dynamically selecting servers based on

knowledge of the location of the user and possibly the content being

requested

* Some sights operate their own CDN, e.g. Google, YouTube

¢ There are third party companies that offer CDN services such as

Akamai, Limelight and Level 3 Communications (now part of Century

Link)

* See the Akamai video for 5 minutes (Tom Leighton, start at 0:44-5:00),

* https://www.youtube.com/watch?v=Ni_60cbMydg

Slide: s27.png

* Two Critical Technology Challenges for YouTube:

? how to identify billions of videos

? How to efficiently deliver the video to the desktop/mobile device

* The Solutions:

* Identification: YouTube assigns a fixed-length, 11 character string, base

64, unique identifier to each video, see

* https://www.youtube.com/watch?v=gocwRvLhDf8 (5 min)

* Efficient Delivery: YouTube makes use of Google's data centers using

them as a content distribution network

? https://www.youtube.com/watch?v=6yrijdhvAtI (2 mins)

Slide: s28.png

University of Southern C.

+ A map of Google? s data centers, see

https://www.google.com/about/datacenters/inside/locations/index.html

Nae ? Se.

@ Mica % te

Pacific amen Indian

Ocean ean stralia

¢ a. Se

? @&

Figure 4: Geographical distribution of YouTube
Video Cache Locations.

Slide: s29.png

1. videos are uploaded from a desktop to a
central Data Center
2. the video is then transcoded into multiple
formats ?
3. transcoded copies are sent to the A
Content Distribution Network me
\ APL servers |
Video transcoding is a technique of
converting a video into multiple different
formats and resolutions to make it playable
across different devices and bandwidths.
The technique is also known as video
encoding. This enables YouTube to
stream videos in different resolutions such as
144p, 240p, 360p, 480p, 720p, 1080p & 4K.

Slide: s30.png

A local DNS server resolves
www.youtube.com and is re-
directed to a YouTube server
which downloads the page
information and a pointer to
a YouTube server that can
i i DNS resolutions for YouTube
deliver the video, e.g. feeb
v23 lscache5 Cc youtube com youtube.com, and other youtube
. o~ . server hostnames.
The request to v23.lscache5
YouTube Video Cloud
User

Youtube Lecture Slides

?+ ?viesatronrena 4 HTTP GET request to [2 HTTP Reply containing the basic HTML

may be further resolved TF vieeotont ond servers

download the video page, and URLs to download other

components including flash video file.

Front end web server 3 HTTP GET request to 4 HTTP Reply with flash video file.

(Hostname www.youtube.com) "download the flash video file.

4 steps describing the delivery of a YouTube video

http://www-users.cs.umn.edu/~zhzhang/Papers/youtube-tech-report.pdf

Slide: s31.png

¢ The design of the

YouTube video delivery

. Video ID Space

system consists of three P

components: onsen sng

1. a ?flat? video id space, z Finer ted

2. a multi-layered logical ?Secondary

server organization ea

consisting of five ?GS nomespocs)

anycast namespaces

(and two unicast

namespaces), and

3. a 3-tiered physical Figure 3: YouTube Architectural Design.

cache hierarchy with

(at least) 38 primary

locations, 8 secondary

and 5 tertiary

locations.

3-Tier Physical

Cache Hierarchy

https://www-users.cse.umn.edu/~zhang089/Papers/youtube-tech-report. pdf

Slide: s32.png

References to YouTube? s CDN

¢ There are four research papers that investigated and discussed the YouTube

CDN, they are:

Vivisecting YouTube: An Active Measurement Study, 2012, cited by Jefay

Dissecting Video Server Selection Strategies in the YouTube CDN, 2011, cited

by Jefay

YouTube Traffic Dynamics and Its InterPlay with a Tier-1 ISP, 2010

https://www-users.cse.umn.edu/~zhang089/Papers/youtube-tech-report.pdf

¢ All of the papers describe a complicated re-direction scheme to find the nearest data center to serve the video; they attempt to minimize Round Trip Time or RTT

¢ For rarely-called-for videos the ?Dissecting? paper did a study requesting in California a rare video and observed that the first request came from the Netherlands, but future requests were served from California

? Conclusion: videos are constantly being moved around to be closer to the place that is requesting then,

right Ellis Horowitz, 2011-2022

Slide: s33.png

* YouTube challenges in the early days

YouTube had no way of making money and its infrastructure is very expensive

YouTube was being sued by content creators as many of YouTube? s videos were uploaded illegally

YouTube solved both problems at once, by

* Developing a system for spotting copyrighted content

+ Allowing the copyright owner to decide if he wants to keep the content on the site and let ads appear, splitting the revenue with YouTube or taking the content down

Here is a video that describes how YouTubers make money

https://www.youtube.com/watch? v=v8F4jrtZtNE

(8 min)

Slide: s34.png

* YouTube?s solution was to create a fingerprint database of copyrighted content, called Content ID

* YouTube solicited cooperation from content owners asking them to submit copies of their content so YouTube could fingerprint them

? There are millions of reference files in YouTube? s Content ID database.

* When a new video is uploaded, it is immediately checked against the database, and the video is flagged as a copyright violation ifa match is found.

¢ When this occurs, the content owner has the choice of

1. blocking the video to make it unviewable,

2. tracking the viewing statistics of the video, or

3. adding advertisements to the video

https://arstechnica.com/tech-policy-policy/2014/10/youtube-has-paid-1-billion-to-rights-holders-via-content-id-since-2007/

Youtube Lecture Slides

Slide: s35.png

Content ID is based off audio and video samples that rights holders have uploaded to YouTube

User uploads a video.

YouTube then queues up the video to be processed i.e. it is transcoded into multiple formats

including:

? HTML5, H.264, WebM VP8, HD, non-HD, and others

4. If the video contains audio, a hash is then calculated based off a time frequency graph called a

spectrogram.

? Target zones (peak points in the spectrogram) are marked, then the target area between

them is also taken and hashed

5. For the video portion, a sample section of frames of the video is taken.

? A hash is created from those sampled frames of the video

* Note recently YouTube has introduced a new version of ContentID it calls Copyright Match

* See the following videos for details, (2 min). https://youtu.be/5-2R-IZITZ8

Slide: s36.png

_ Creating an

Acoustic Fingerprint

¢ The audio signal is digitized and converted to a spectrogram ? a time-

frequency graph

The graph below plots three dimensions of audio: frequency versus amplitude

versus time

? Acommon format is a graph with two dimensions: one axis represents time, and

the other axis represents frequency; a third dimension indicating the

amplitude of a particular frequency at a particular time is represented by the

intensity or color of each point in the image.

frequency

Time axis

Slide: s37.png

* According to stats released by YouTube 99.5 percent of all copyright issues

specifically related to sound recordings are automatically resolved by Content ID

¢ In addition to music, Content ID also identifies 98% of copyright claims, including

those tied to film, TV, gaming

¢ The actual hashing algorithm used by YouTube remains proprietary, but it has been

suggested that YouTube uses finite-state transducers to compute the hash function,

e.g. see

« Eugene Weinstein, Pedro J. Moreno; Music Identification with Weighted Finite-State

Transducers, Proceedings of the International Conference in Acoustics, Speech and

Youtube Lecture Slides

Signal Processing (ICASSP), 2007

Slide: s38.png

* The storage you can buy with $100 Gigabytes of storage you can buy with $100

has grown exponentially ? or 5000

equivalently, the cost of storing 1GB 4500

of videos has decreased exponentially 4000

Kryder?s Law

https://en. wikipedia.org/wiki/Mark_Kryder#Kryder%27s_law_projection

Slide: s39.png

* An answer by Rasty Turek EE ene me

from Quora 3} 640x360 mp4

* There is roughly 24TB ofnew & 64oxs60 webm

videos uploaded daily 3 400%240 -

* Each video is re-encoded based

on pre-selected profiles and See sia

each is stored as a separate file [© 176x144 Sap

¢ Here is his computation: ©} 4k (no audio) mp4

24TB * 4x (for profiles) * 365 days = 35PB/year

So YouTube needs to store roughly 35PB of new data every year.

& download

& download

+ download

& download -

& download -

& download

& download

59.01 MB

15.34 MB

19.07 MB

8.51 MB

5.94 MB

- 2.12 MB

- 297.69 MB

From multiple sources we know that there is roughly 1 Billion videos that have been uploaded

to YouTube to date. Assuming each video has on average size of 86MB we can compute their

total storage needs as:

86MB * 4 (for profiles) * 1,000,000,000 = 320PB

So it is estimated that YouTube needs to have at least 320PB of storage currently and that

the storage needs are growing each year by 35 PetaBytes

Slide: YT_ss_design_45W.gif

Youtube System Design
FAAS
mmm -->- Sik
me ?. = Content Delivery
Mobile Client Mobile Client Web Client Network (CDN)
ee een ee == Media Cache
b--------->
-_?e?en
Web Server
Media Storage (S3)
(Video aimee)
Thumbnail Storage
User Database Metadata Cache
Metadata Database

Slide: YT_ss_design_45W.png

YouTube System Design! &
Discover the intricate architecture that powers the world's leading video-sharing
platform:
1. Content Delivery Network (CDN): The backbone of YouTube's rapid content
delivery, ensuring seamless streaming for users worldwide.
2. Load Balancer: Efficiently distributes incoming requests across multiple servers,
optimizing performance and preventing bottlenecks.
3. Application Servers: The engine behind YouTube's functionality, handling user
requests, interactions, and serving content dynamically.
4. User Database: The repository storing user information, preferences, and
interactions, ensuring personalized experiences.
5. Transcoding Servers: Vital for converting and optimizing video files into various
formats, accommodating diverse user devices and network conditions.
6. Thumbnail Storage: A dedicated space for storing video thumbnails, enhancing
visual appeal and facilitating quick content recognition.
7. Web Server: Facilitates user interaction, serving web pages and ensuring a
seamless browsing experience.
8. Metadata Database: Stores crucial metadata associated with videos, enabling
efficient content organization and retrieval.
9. Metadata Cache: Optimizes data retrieval speed by storing frequently accessed
metadata, enhancing overall system efficiency.

Youtube Lecture Slides

10. Media Storage ($3): Robust and scalable storage solution for housing the vast
library of YouTube videos, ensuring accessibility and reliabili
yar Reais uTube, TBE EG UBIHIS Bocessed Videos ty.