

Northeastern University - Seattle



CS6650 Building Scalable Distributed Systems
Professor Ian Gorton

Building Scalable Distributed Systems

Week 1 – Introduction to Scalable Systems

(Part 2)

Modern web sites and scale



21ST CENTURY APPLICATIONS

Some (out-of-date) numbers

Dropbox – [exascale storage system](#), millions of new files per hour, from 2012-16 seen over 12x growth.

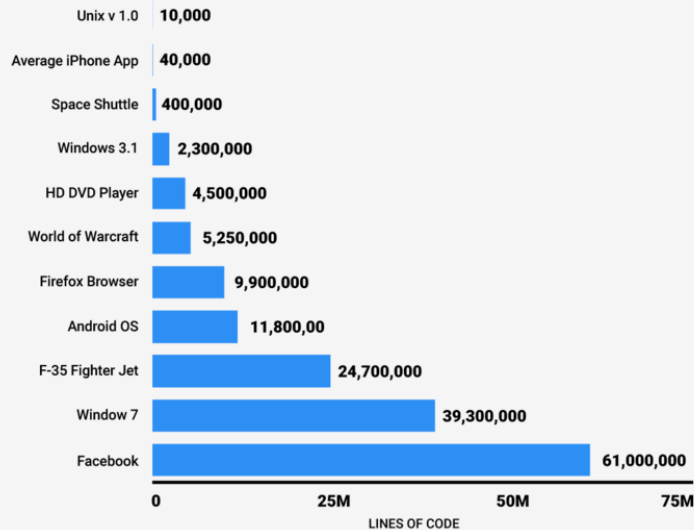
Gmail – [1.2 billion users](#), blocks 10M spam messages every minute

Netflix – runs on AWS, [took 7 years to migrate](#), uses 15% of global downstream internet traffic

Facebook - [More than 300 million photos](#) uploaded per day and 510,000 comments posted and 293,000 statuses updated per minute

Youtube - Users watch [4,146,600 YouTube videos](#) every minute

HOW MANY LINES OF CODE MAKE UP THESE POPULAR TECHNOLOGIES



SOURCE: NASA, Quora, Ohloh, Wired

BUSINESS INSIDER

Google repository statistics, January 2015.

Total number of files	1 billion
Number of source files	9 million
Lines of source code	2 billion
Depth of history	35 million commits
Size of content	86TB
Commits per workday	40,000

<https://www.businessinsider.com/google-runs-on-5000-times-more-code-than-the-space-shuttle-2016-7?international=true&r=US&IR=T>

- **Anatomy**
- **Of a**
- **21st Century**
- **Software Systems**

**Massive
Cloud Scale
Processing**



**Distributed
Globally**

CAP Theorem



GLOBAL USER BASE



- COMPLEX BIG DATA ANALYTICS
- BUSINESS INSIGHTS
- SYSTEMS OPERATIONS



Building Massive Scale Systems is difficult



**Software Architecture
Quality Attributes**





- Only one large internet site publish detailed usage statistics every year



APOLOGIES FOR ANY INCONVENIENCE

by
Jennifer Milne

- So let's look at some [usage numbers](#)

Pornhub

==2018==

33.5
BILLION

VISITS TO PORNHUB

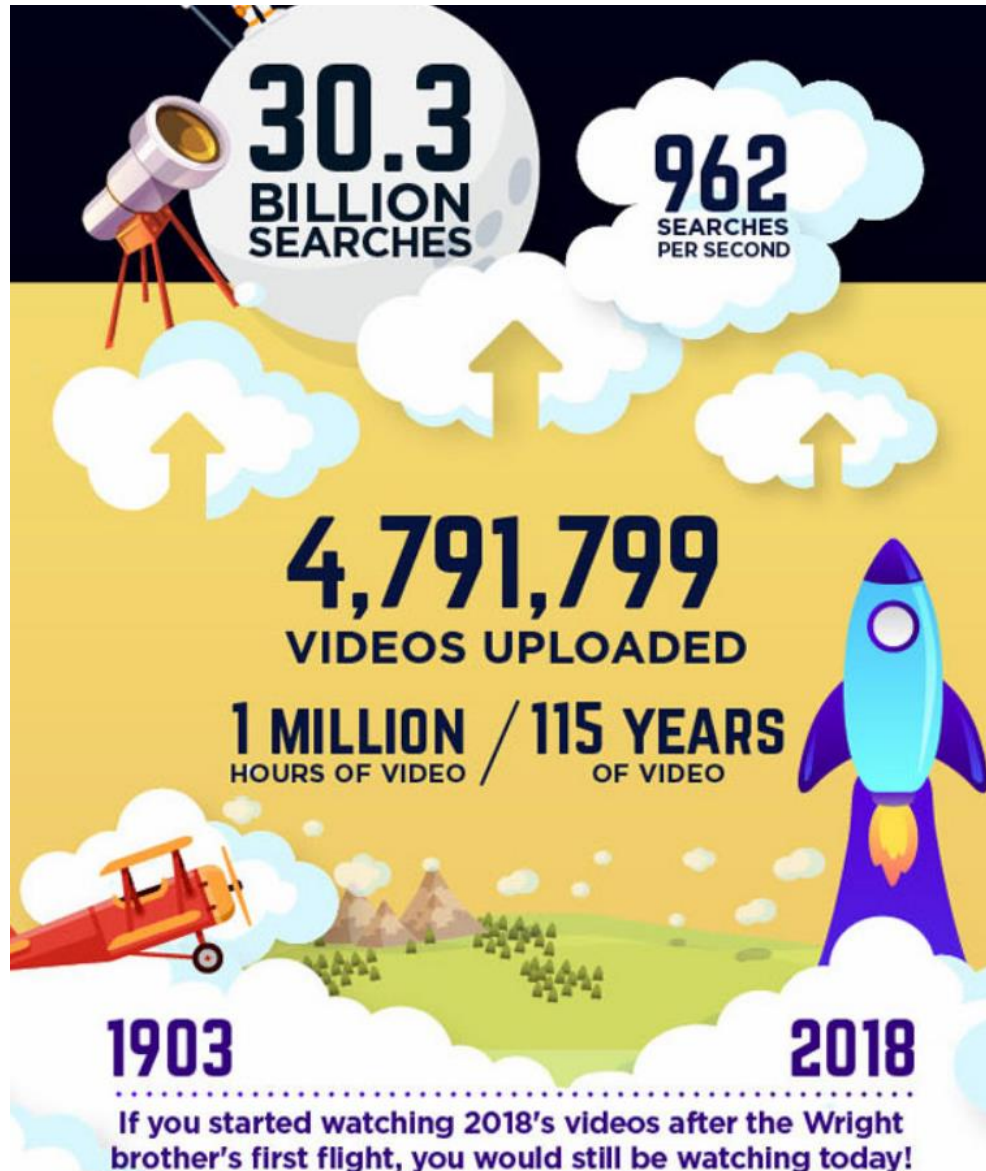
92 MILLION

DAILY AVERAGE VISITS

=

THE POPULATIONS OF
CANADA, POLAND AND AUSTRALIA
EVERY DAY





4403
PETABYTES
OF DATA TRANSFERRED

That's 574 MB of data
for every person on
earth, or about 283
photos worth

147GB
PER SECOND

529,200 GB
PER HOUR

12,700,800 GB
PER DAY

That's more
bandwidth than
the *entire internet*
consumed in 2002

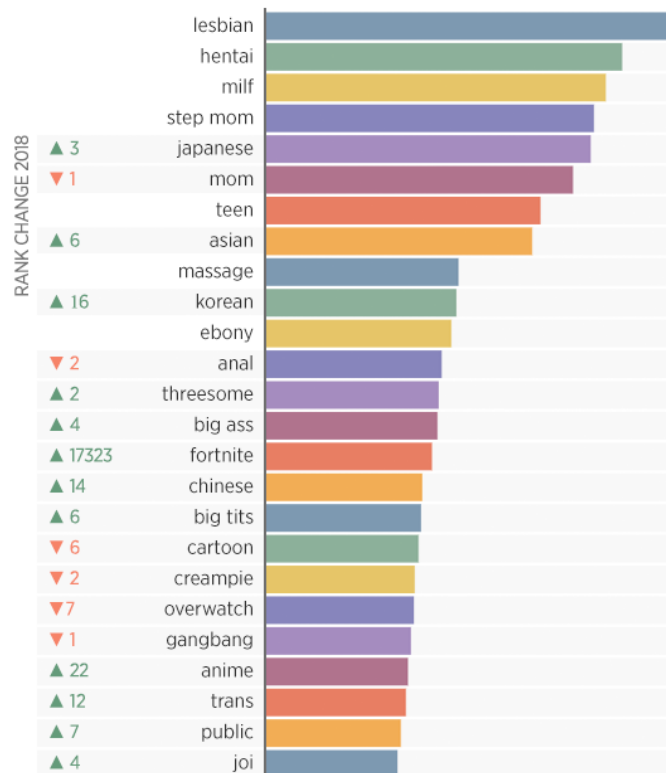




2018
YEAR IN
REVIEW

Pornhub

Most Searched for Terms of 2018

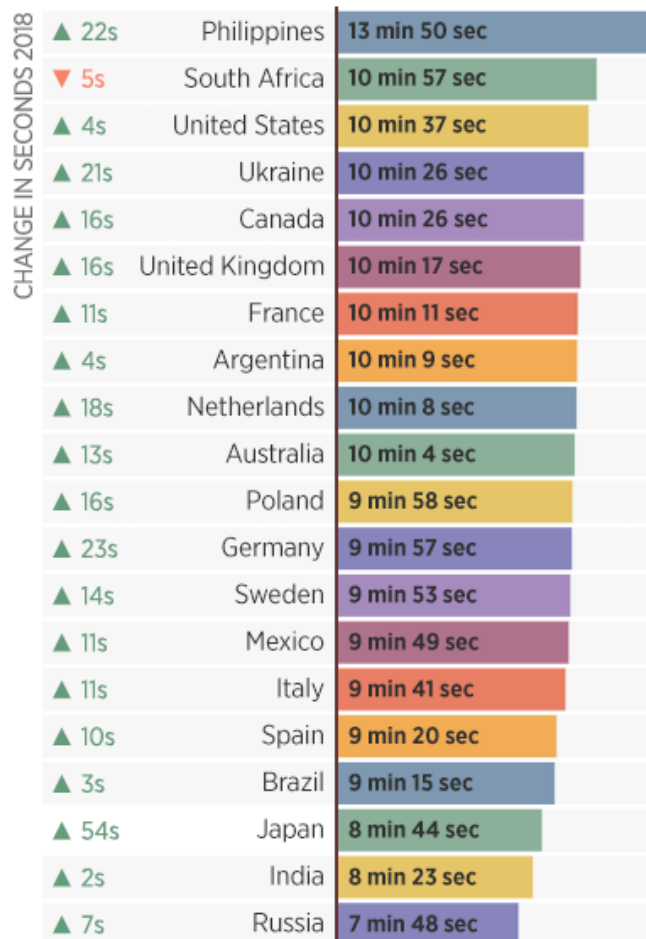


Top 20 Countries by Traffic



Largest Rank Changes of 2018:
 ▲33 Kazakhstan
 ▲26 Cuba
 ▲15 Czechia
 ▲31 Syria
 ▲17 Colombia
 ▲12 Turkey

Time Spent Per Visit



10 min 13 sec

Average Visit Duration
Worldwide

▲ 14 sec

Increase in 2018

Largest Duration
Increases of 2018

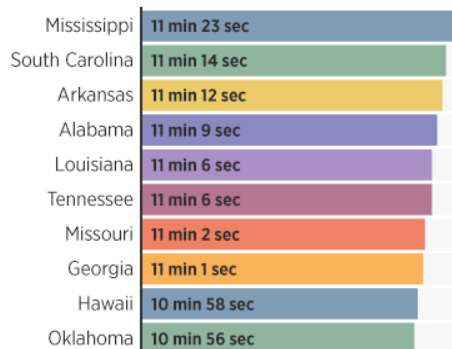
Czech Republic	▲ 112s
Algeria	▲ 80s
Morocco	▲ 56s
Sri Lanka	▲ 55s
Japan	▲ 54s
Latvia	▲ 50s
Fiji	▲ 47s
Ethiopia	▲ 52s
Moldova	▲ 40s
Hong Kong	▲ 37s

2018
YEAR IN REVIEW

Pornhub

Time Spent Per Visit - United States

States That Last the Longest



10 min 37 sec

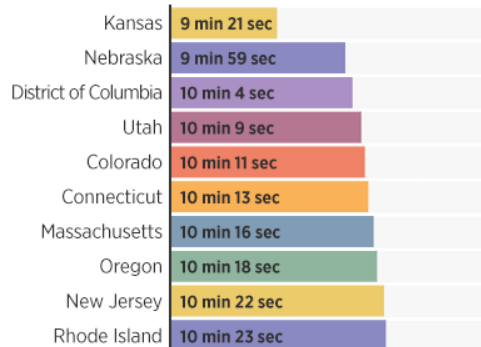
Average Visit Duration
in the United States

▲ 4 sec

Increase in 2018

Largest Duration
Changes of 2018

States That Last the Shortest



Idaho	▲ 54s
Oregon	▲ 23s
District Columbia	▼ 20s
Nebraska	▼ 19s
Kansas	▲ 16s
New Hampshire	▲ 14s
Maine	▲ 13s
Montana	▲ 12s
South Dakota	▲ 11s
Vermont	▲ 11s
Mississippi	▼ 11s





Next ...

What do we mean by
scalability?