

# DATA ANALYSIS ABOUT GOOGLE

Sandeep PVN

B.Tech (CSE), PES University, Bangalore

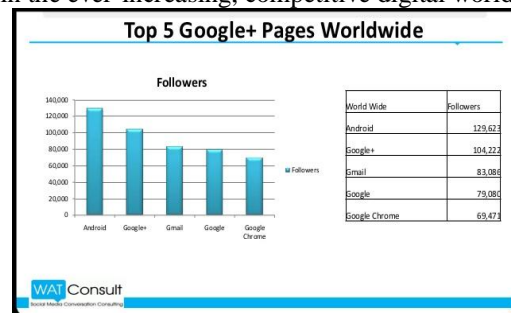
**Abstract—** Google is a very vast company that basically gets its revenue from AdWorks. Google over the past years has acquired and made many subsidiaries such as YouTube, Drive, Picasa, Gmail, Maps and many more. It has become the most important leading IT giant all around the globe. The Google search being the most important of them all. Any information any individual need, First thing they do is open the Google home page and search from it. Just imagine the tremendous amount data Google has to handle all the time! Well, supposedly you are searching for something on Google, It is an astonishing fact to know that at the very instant Google also provides the necessary information to over 2 million users at the same time.

Let's take another example of YouTube. The humongous amount of video clippings on it. Google has to stream them every second. Did you also know that even if an individual was to watch all the videos on YouTube continuously he would not be able to watch all the videos in not two one but even two lifetime?

## I. INTRODUCTION

Another big achievement of Google is Gmail. The world's largest email coup. With millions of emails sent every day and thousands of new users signing up every day. There are about 500 million active gmail accounts. Apart from these Google drive, maps, google+ also contribute equally. Also another major role of google is the android. Over 52% of smartphones used today are android. The google play store is a huge library of apps, news feeds etc. with almost 200 million applications. Google Trends provides an index of the volume of Google queries by geographic location and category. Google Trends data does not report the raw level of queries for a given search term. Rather, it reports a query index. The query index starts with the query share: the total query volume for search term in a given geographic region divided by the total number of queries in that region at a point in time. The query share numbers are then normalized so that they start at 0 in January 1, 2004. Numbers at later dates indicated the percentage deviation from the query share on January 1, 2004. This query index data is available at country and state level for the United States and several other countries. There are two front ends for Google Trends data, but the most useful for our purposes is <http://www.google.com/insights/search> which allows the user to download the query index data as a CSV file. As people and companies rely on basic infrastructures to function, businesses also rely on an information systems infrastructure (consisting of hardware, software, networks, data, facilities, human resources, and

services) to support their decision making, business processes, and competitive strategy. Business processes are the activities that organizations perform in order to reach their business goals and consist of core processes and supporting processes. The core processes make up the primary activities in the value chain; these are all the processes that are needed to manufacture goods, sell the products, and provide service, and so on. The supporting processes are all the processes that are needed to perform the value chain's supporting activities, such as accounting, human resources management, and so on. Almost all of an organization's business processes depend on the underlying information systems infrastructure, albeit to different degrees. For example, an organization's management needs an infrastructure to support a variety of activities, including reliable communication networks to support collaboration between suppliers and customers, accurate and timely data and knowledge to gain business intelligence, and information systems to aid decision making and support business processes. In sum, organizations rely on a complex, interrelated information systems infrastructure to effectively thrive in the ever-increasing, competitive digital world.



Google is a "BIG DATA HANDLING" System and it has almost 1 million servers in all its data centres. That is typically enough storage to gather almost every possible piece of information on this planet and beyond. Google has also been into various other smart fields. It obtains huge amounts of data from various sources. The other products of google are

- Google chrome
- Google goggles
- Google translate
- Google maps
- Picasa
- Google news
- Google earth
- Panoramic
- googledocs

- googlesheets
- googlekeep
- google+
- Hangouts
- googlecode

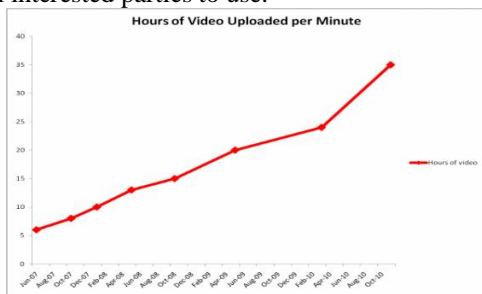
**Gmail**

There are almost 500 million active gmail users. Gmail along with drive has provided every user with a default storage space of 15GB. Based on statistics we can derive that the average gmail user is a male of the age of about 31 and is single. From further data we can estimate that this user uses about 12GB of data. That would make it about “1 EXABYTES” of data to be handled per month.

Many Gmail users receive tens or hundreds of mails per day. The Priority Inbox attempts to alleviate such information overload by learning a peruser statistical model of importance, and ranking mail by how likely the user is to act on that mail. This is not a new problem, however to do this at scale, performing real-time ranking and near online updating of millions of models per day significantly complicates the problem. The challenges include inferring the importance of mail without explicit user labelling finding learning methods that deal with no stationary and noisy training data constructing models that reduce training data requirements; storing and processing terabytes of per-user feature data; and finally, predicting in a distributed and fault tolerant way. While ideas were borrowed from the application of ML in Gmail spam detection, importance ranking is harder as users disagree on what is important, requiring a high degree of personalization. The result is one of the largest and most user facing applications of ML at Google.

**YouTube**

YouTube, the video hosting service, offers students, teachers, and practitioners of qualitative researchers a unique reservoir of video clips introducing basic qualitative research concepts, sharing qualitative data from interviews and field observations, and presenting completed research studies. This web-based site also affords qualitative researchers the potential avenue to share their reusable learning resources for all interested parties to use.



YouTube is a video-sharing website headquartered in San Bruno, California. In October 2006, Google.Inc announced that it had acquired YouTube for \$1.65 billion in Google stock,

and the deal was finalized on November 13, 2006. All YouTube users can upload videos up to 15 minutes each in duration. These videos are also streamed at different quality resolutions varying from 140 pi to 1080 pi (full HD) and even 4K resolution. More than 1 billion unique users visit YouTube each month. Over 6 billion hours of video are watched each month on YouTube. That’s almost an hour for every person on Earth 100 hours of video are uploaded to YouTube every minute Millions of subscriptions happen each day.

The number of people subscribing daily is up more than 3x since last year, and the number of daily subscriptions is up more than 4x since last year Mobile makes up almost 40% of YouTube's global watch time this contributes to hundreds of petabytes of data every month.

	YouTube Users	Users (M)	% Users
Age	All	72.0	-
	<18	13.3	18%
	18-34	13.2	18%
	35-44	14.2	20%
	45-54	15.5	22%
Gender	55+	15.8	22%
	Male	37.0	51%
	Female	35.0	49%

**Android**

Android is a free and open source operating system. It was initially developed by Android Inc. but later in 2005 it was purchased by Google. In 2007 a group of 78 companies formed a group called Open Handset Alliance (OHA) to develop and distribute Android. It is an operating system for low powered devices, those runs on battery and is full of hardware. Android applications uses hardware features through abstraction and provide a defined environment to run applications. An Android application is written in java and run in a virtual machine which is called Dalvik virtual machine which executes its own byte codes.

Android is a mobile operating system (OS) based on the Linux-Kernel and currently developed by Google. As of July 2013 the Google Play store has had over 1 million Android apps published, and over 50 billion apps downloaded. One main feature of android development is google play. Google play is the main interface where all the android users download their apps from a range of over 1 million choices. Android is the leading OS in smartphones with over 1 billion active users! There have been many updates for android. Starting with Froyo, Gingerbread, Honeycomb, Ice-cream, Sandwich, Jelly bean, Kit Kat and the latest being Lolipop.

Mainly an android application consists of four components:

- Activity: An activity is a single screen of an application with which user can interact like click a photo, dial a number etc. Intents are used for transition between different activities and each application can have multiple activities.

- **Service:** A service is a component that does not provide any user interface but runs in the background to perform long running tasks.

- **Content Provider:** To exchange data between different applications, a component is used. It is called content provider which handles retrieval of data and stores data in database files or on a network.

- **Broadcast Receiver**

The broadcast messages from other application or from system are called intents and the component which responds to these intents is called broadcast receiver.

Android has emerged as a strong competitor in mobile sector as it is supported by large companies especially by Google. Manufacturers can modify the system as per their needs due to Android's openness and extensibility. Today Android operating system is not only used in Mobiles and tablets, its implementation in electronic devices is increasing rapidly. Smart TV and Smart Camera are examples of new implementation and in future android will be in many household devices like washing machine, Oven and many more.

## II. CONCLUSION

Hence we derive at the conclusion that google with its vast resources of over a million servers spread over 6 data centres across the globe is definitely. BIG DATA handling giant whose incoming data and as well as managing capacity is increasing over three-fold times every single year! Thus making google one of the biggest and the best companies in the world.

## III. REFERENCES

- [1] Barritt, C., & Alderman, F. L., Jr. (2004). Creating a reusable learning objects strategy: Leveraging information and learning in a knowledge economy. San Francisco: Pfeiffer.
- [2] Chenail, R. J. (2004). When Disney meets the research park: Metaphors and models for engineering an online learning community of tomorrow. *Internet and Higher Education*, 7(2), 107-121.