

→ Technology evolution to fulfill emerging needs.

landline phones → smart phones.

floppy drives → HDD → server → cloud.

- slow transfer speed.
- low storage

terabytes of data on the cloud without being bothered about size constraint.

* Drivers that contribute to the generation of data.

① IoT → connects your physical device to the Internet & makes it smarter

Smart Air conditioner
Smart TV.

— constantly monitors room temperature along with the outside temperature & accordingly decides what should be the temperature of the room.

— Imagine how much data would be generated in a year by Smart air conditioners installed in tens of thousands of houses.

— IoT is contributing a major share to Big Data.

② Social Media:

— info about user's behaviour.

- ③ online video streaming app'n and/or Netflix, root etc
- ④ online shopping app'n
- ⑤ Virtual personal Assistant
- ⑥ Credit/Debit transactions

every minute

meta - 1736,111 Instagram pics
4,166,667 likes & 200,000 photos.
347,222 tweets
300 hours of video uploaded - youtube.
204,000,000 - emails.

→ Large amount of data - $\&$ how to process?
not able to processed with single system.

so, solve the problem with multiple users.

But, they all share the common central storage.

→ with central storage -

- └ difficult for parallel processing
- └ if central storage goes down, (C.P.e single point failure).

→ define the hierarchy.

→ In hadoop, the data is stored in a distributed fashion with replications, to provide fault tolerance.

→ for parallel processing, first data is processed by the slaves where it is stored for some intermediate results & then those intermediate results are merged by master node to send the final results.

③ Major challenges with Big data:

└ ① storing the colossal amount of data -
- storing huge data in

└ ② storing heterogeneous data.

└ ③ processing speed.

* Drivers that contribute to the Big Data Market :-

BUSINESS & ORGANIZATION

① Opportunity to enable innovative new business models.

- reducing traffic congestion.

- collect data from

in unified form.

- Cameras

- temperature

- GPS data from Uber & other taxis

- loc. data collected from users & driver's

- Can be build real time traffic system that improve traffic flow.

- Reduce fuel emissions

- public transport planning.

② Potential for new insights that drive competitive advantage:

- Retail market - Improving customer relations

enhance 360° view of customer

→ better engagement

→ revenue

→ long term loyalty.

- Growing no. of high tech crime

L Improving security - by processing and analyzing new data types such as social media, emails, and hours & hours of video footage.

L Analyzing data in motion and at rest can help find new associations or uncover patterns.

TECHNICAL

- ① Data collected & stored continues to grow exponentially.
- ② Data is increasingly everywhere and in many formats.
- ③ Traditional solutions are failing under new requirements.

FINANCIAL

- ① Cost of data system, as a % of IT spend, continues to grow.
- ② Cost advantages of commodity hardware & open source software.
 - Computer or PC that has no outstanding features & widely available for purchase
 - considered as disposable & replaced rather than repaired.

* Big Data challenges & Solutions:

① Quality of Data:

- Big data must be cleaned, prepared, verified, reviewed for compliance and constantly maintained.
- The issue is information comes in so quick that organizations think that it's hard to play out the majority of the data preparation activities to guarantee ideal data quality.
- Companies basically store the majority of their approaching big data without doing a lot to it.
- This makes data contaminated.
- Also, inaccurate information can raise the danger of business decisions being founded on wrong data.

Solution:

- characterize your business rules for data cleaning and preparation.
- figure out which data you completely not require and set up data cleansing automation to discard this data before it ever hit your system.

② Data Security:

- The data made available from numerous sources, and therefore has potential security problems.
- you may never know which channel of data is compromised.

- Thus compromising the security of the data available in the organization, f giving hackers a chance to move in.

Solution:

- Introduce Data Security best practices for secure data collection, storage and retrieval.

③ Insufficient understanding and acceptance of big data

- Without clear understanding, a big data adoption project risks to be doomed a failure.

- Companies may waste lots of time and resources on things they don't even know how to use.

Solution:

- Big data, being a huge change for a company.

- For acceptance of big data, the implementation and use of the new big data solution need to be monitored and controlled.

④ Confusing Variety of Big Data technologies

- It can be easy to get lost in variety of big data technologies now available on the market.

Solution:

- Choose the technology based on your application.

⑤ Troubles of upscaling:

- The most typical feature of big data is its dramatic ability to grow.
- And one of the most serious challenges of big data is associated exactly with this.
- Solution's design may adjusted to upscaling with no extra efforts.
- But, the actual problem lies in complexity of scaling up so that system's performance doesn't decline and you stay within budget.

Solution:

- Audits can help identify weak spots and timely address them.

⑥ Integration of platform:

- There are plenty of backend dispersed data store.
- Some of these dispersed data stores are not locally supported by the platform.
- Depending on the data store, one should ~~not~~ utilize an alternate API.

Solution:

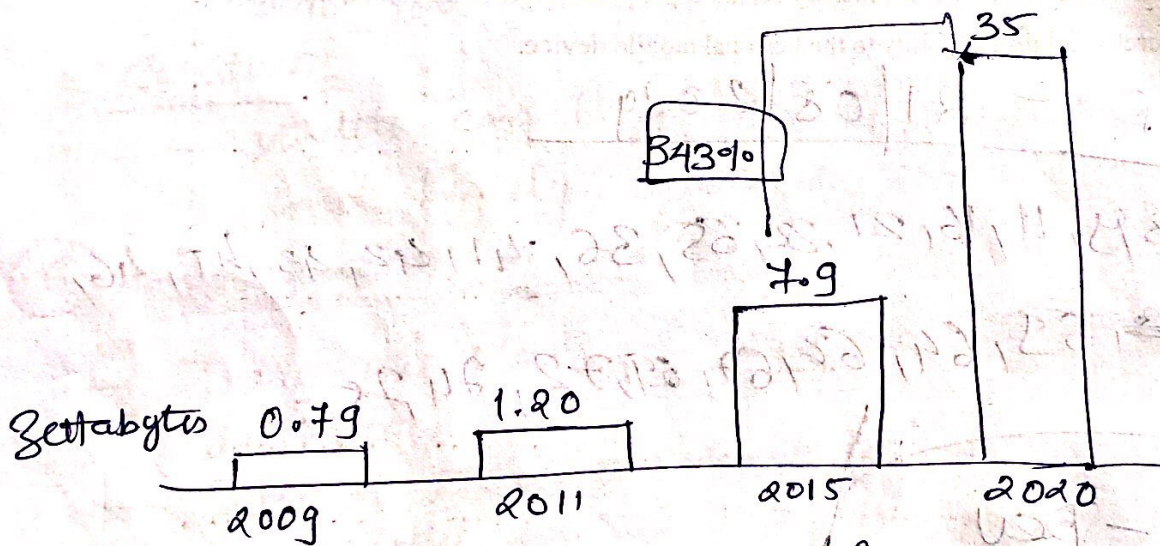
- There are software automation tools accessible with several pre-created APIs for a wide range of data.

→ More data has been created in the past 2 years than in the entire previous history of the human kind.

→ By 2020, about 1.7 megabytes of new information will be created every second for every human being in the world.

→ By 2020, the data we create and copy will reach around 35 zettabytes, up from only 7.9 zettabytes today.

Growth in global data



→ How Big is zettabytes?

- 1 bit ⇒ 1 or 0
 - 8 bit ⇒ 1 byte
 - 1024 bytes ⇒ 1 KB
 - 1024 KB ⇒ 1 MB
 - 1024 MB ⇒ 1 GB
 - 1024 GB ⇒ 1 TB
 - 1024 TB ⇒ 1 PB
 - 1024 PB ⇒ 1 Exabyte
 - 1024 Exabyte ⇒ 1 zettabyte
- word document, song.
 - single ppt or pictures,
 - large videos & DVDs
 - computers & hard drives
 → 3.4 years of 24x7 full HD video recording
 → movie Avatar require 1 PB to render those graphics.

1024 PB \Rightarrow

1 Exabyte

\rightarrow

Big urban city data
Busy international airport
data that includes:

- \rightarrow airplanes are capturing & transmitting data
- \rightarrow all people in those airports have mobile device
- \rightarrow security cameras.
- \rightarrow staff data.

\rightarrow It is also predicted that the amount of data produced will increasingly outpace available storage.

\rightarrow It is not convenient for any enterprise to have such a huge storage capability not even in a distributed form. \because it is costly.

\rightarrow Emerging cloud computing services has resolved this issue.

\Rightarrow Advances in cloud computing have contributed to the increasing potential of Big Data.

\rightarrow Emergence of cloud computing has highly contributed to the launch of the big data era.

\rightarrow cloud computing allows users to access highly scalable computing and storage resources through the internet.

→ By using cloud computing, companies can use server capacity as needed and expand it rapidly to the large scale required to process big data sets and run complicated mathematical models.

→ cloud computing lowers the price to analyze big data as the resources are shared across many users, who pay only for the capacity they actually utilize.