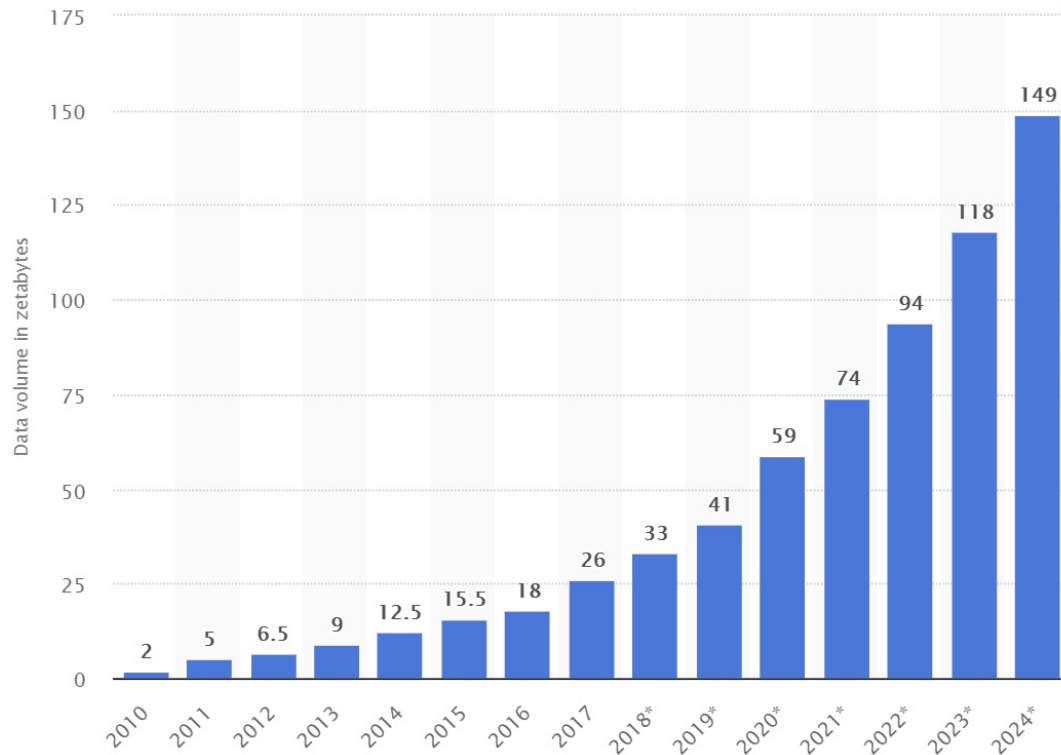


**What is Big
Data?**

Volume

- Size of the data being stored
- 40,000 queries to Google each second
- 300 hours of new video on YouTube each minute
- More than 100 tb of data on Facebook daily
- Smart devices produce 5 quintillion bytes of data daily

Painting a Picture – in Zettabytes!



© Statista 2020

[Additional Information](#)

[Show source](#)

What are we doing with all that data?

- Nothing!
- 99.5% of collected data never gets used or analyzed
- Less than 50% of IoT data is used in decision making

Variety

- Types of Data and how they differ
 - Video
 - Audio
 - Text
- The composition each company has of the types of data

Velocity

- Speed of data generation and storage
- Real-time analysis or consumed later?

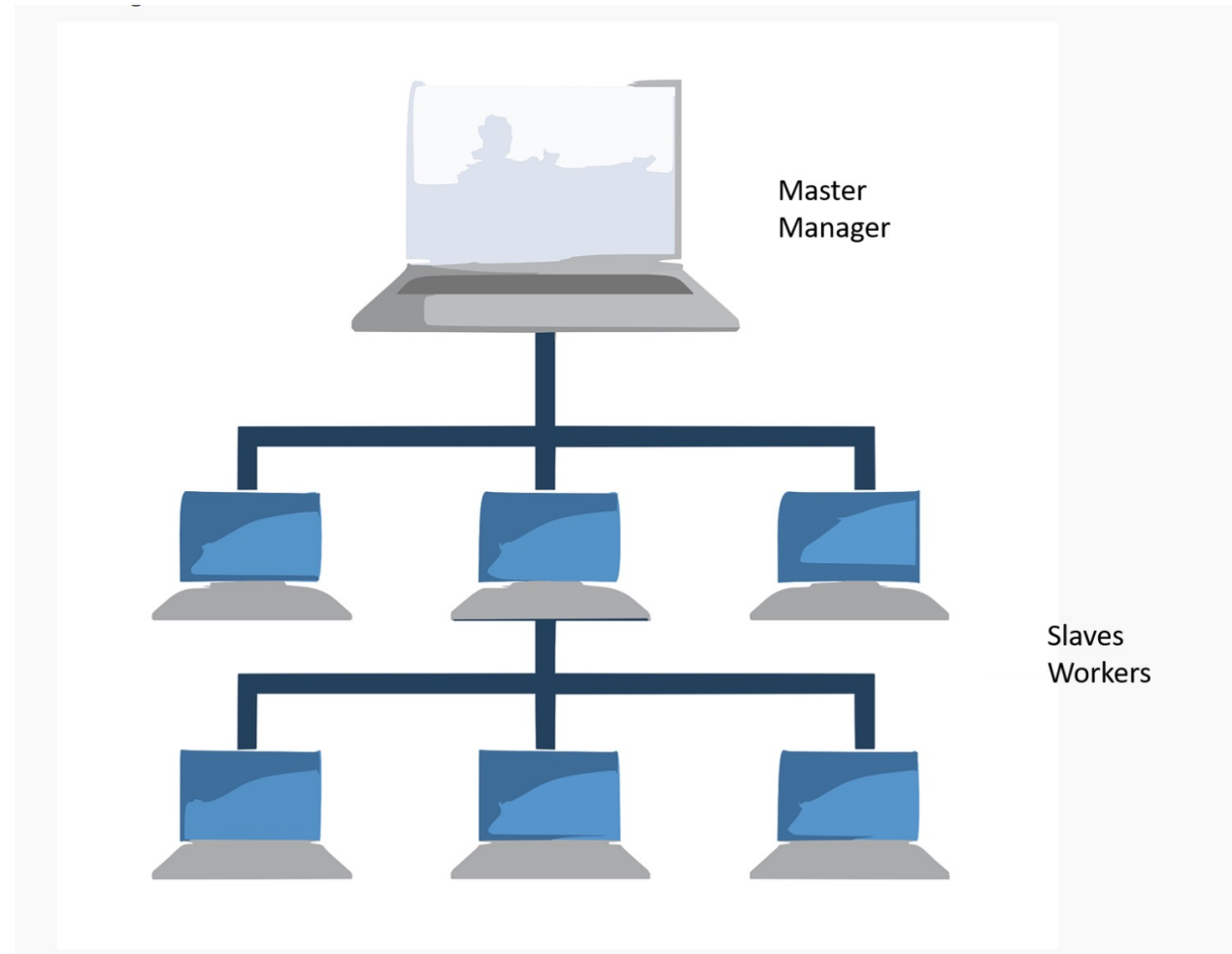
Veracity

- How trustworthy is your data?
- Are there discrepancies? Duplicate records?
- How confident are you in making a decision with your data?

Value

- Does it provide useful information and insights?
- How can you use data to change the world?

How Distributed Processing Works



About Big Data...

- It's finicky
- Companies have their own proprietary software for it
- Sampling means you don't have to use ALL the data

Have a backup!

- Redundancy – data stored on a backup
- Fault tolerance – resilience

Interacting with Big Data

- Hadoop
- Google Co-Lab
- Amazon Web Services
- Cloud-based R & Python
 - Saturn Cloud
 - DASK

Questions?