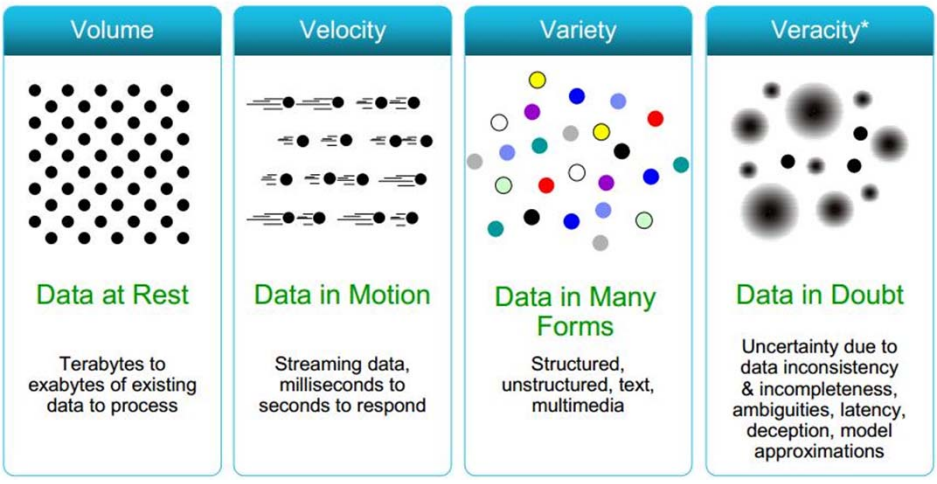
University of South Australia

Big Data Security

1

Big Data is complex



https://www.datasciencecentral.com/profiles/blogs/data-veracity

2

What is Big Data security?

- All **measures** and **tools** used to guard both the data and analytics processes from attacks, theft, or other malicious activities that could harm or negatively affect them.

3

Big Data analytics run on sensitive data

- Disease patterns
- Customer buying habits
- Tax fraud or criminal activity
- Social media analysis
- Students at risk
- Healthcare records
- Location services and GPS tracking

4

Big Data security challenges

Distributed frameworks

Hadoop originally had no built-in security

Non-relational data stores

Security provided via middleware

Storage on multiple tiers

High-priority 'hot' data usually stored on flash media

Logs from endpoints

Need to validate authenticity

5

Big Data security challenges

Real-time security

Huge amounts of information, need to deal with false positives

Analytics solutions

Secure against external threats, internal abuses of privileges

Access controls

Encrypted authentication and validation of users

Granular auditing

Missed attacks and their consequences, a lot of data in itself to be enabled and protected

6

Addressing Big Data security threats

- Encryption (data in-transit and at-rest)
- User access control
- Intrusion detection and prevention
- Centralised key management
- Physical security

7

Security Information and Event Management (SIEM)

- **Big data technologies**, e.g. Hadoop or Elasticsearch, are now leveraged by most SIEM solutions.
- SIEM vendors have also embraced security orchestration, automation and response (**SOAR**).

- Gartner predicts that by 2022:
  - **50%** of all SIEM tools will be **cloud-native** and delivered as a service from the vendor.
  - **75%** of all SIEM vendors in the Gartner Magic Quadrant will offer **advanced analytics features**, as well as orchestration and automation features.

8

Industry leaders

- Gartner report ID G0038109, published on 18 February 2020

Source: Gartner (February 2020)