

Data Integrity/Consistency with Big Data

ITC6000 - Database Management Systems
Adrian An & Dhruv Wadkar

TABLE OF CONTENTS

01

Five Parts of Data
Integrity

02

Why does Data
Integrity Matter?

03

Types of Data
Integrity

04

What is Big Data?

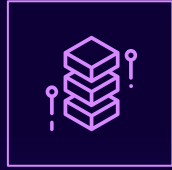
05

Issues w/ Big Data

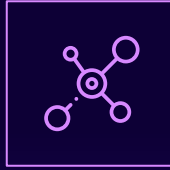
Five Parts of Data Integrity



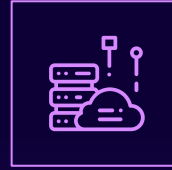
Data
Accuracy



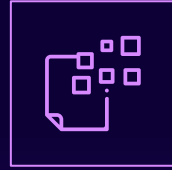
Completeness



Validity



Consistency



Security

Importance of Data Integrity



Economic Benefit

Understand target
markets



Aids Operations

Tacking of internal
activities



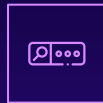
Privacy Protection

Avoid risk of leaks and
increase trust



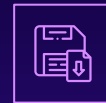
Personalized Offerings

Competitive
advantage



Strategy

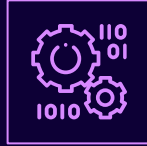
Enables reliable
analysis



Performance

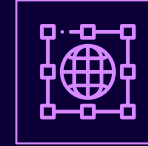
Minimizes redundancy
and Increases accuracy

Types of Data integrity



Physical data integrity

- Protecting wholesomeness and accuracy.
- Affected by hardware mishandling or interference.
- Associated with human activities.
- Mostly affected by the power outages, storage errors and hackers targeting the system.



Logical data integrity

- Data Form and Use.
- Appears in other integrity forms:
 - a. Relational
 - b. Referential
 - c. Entity
 - d. Domain
 - e. User-defined integrity

What is Big Data?



Complex Information

Extensive Dataset
that need complex
Databases



Rows & Layers

High attributes and
columns



Transferability

Has ease with large
amount of data



Multi-Usage

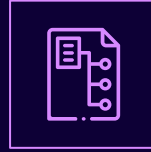
Analysis,
Manipulation, or
Efficient storage

Issues w/ Big data



Data Integrity

- High Volume
- High Velocity
- Time gated processing
- High floor for Users



Data Intrusion

- Hacking or Cyber attacks
- Illegal access



Negative Externality

- Subject to Bias
- Lack of Understanding
- Incorrect Judgement

References

- Fortinet. (2022). What is data integrity? <https://www.fortinet.com/resources/cyberglossary/data-integrity>
- Martin, K. E. (2020). Ethical issues in the big data industry. In Strategic Information Management (pp. 450-471). Routledge.
- Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2020). Factors influencing effective use of big data: A research framework. Information & Management, 57(1), 103146.
- Shi, P., Cui, Y., Xu, K., Zhang, M., & Ding, L. (2019). Data consistency theory and case study for scientific big data. Information, 10(4), 137.
- Wang, L. (2017). Big Data in intrusion detection systems and intrusion prevention systems. Journal of Computer Networks, 4(1), 48-55.
- Lebdaoui, L., El Hajji, S., & Orhanou, G. (2016, September). Managing big data integrity. In 2016 International Conference on Engineering & MIS (ICEMIS) (pp. 1-6). IEEE.
- Tang, Y. (2020, February). Negative Externality of the Data Collection and Its Governing Methods. In The International Conference on Cyber Security Intelligence and Analytics (pp. 494-500). Springer, Cham.