

Ensuring Data Integrity and Reliability in Big Data: A Comprehensive Study on Data Quality Management Strategies*

Michal Zrutta

Slovenská technická univerzita v Bratislave
Fakulta informatiky a informačných technológií
xzrutta@stuba.sk

10. October 2023

Abstract

We live in the era of Big Data. By looking at the statistics of globally generated data annually, we can see the increasing trend. If we compare the year 2010 and 2022 it shows us a difference of 95 zettabytes generated in a single year. However, the upward trajectory poses challenges in data quality. Consequently, the demand for realistic data sets, accurate and consistent information, and overall data quality has become a pressing concern. The paper aims to investigate the challenges related to testing in the context of Big Data adoption. It seeks to outline a testing strategy that can effectively validate the high volume, velocity, and variety of information associated with Big Data. [1] [2]

*Semestrálny projekt v predmete Metódy inžinierskej práce, ak. rok 2023/24, vedenie: Mirwais Ahmadzai

1 Úvod

Motivujte čitateľa a vysvetlite, o čom píšete. Úvod sa väčšinou nedelí na časti.

Uveďte explicitne štruktúru článku. Tu je nejaký príklad. Základný problém, ktorý bol naznačený v úvode, je podrobnejšie vysvetlený v časti ?? . Dôležité súvislosti sú uvedené v častiach ?? a ?? . Záverečné poznámky prináša časť ?? .

References

- [1] V. N. Gudivada, R. Baeza-Yates, and V. V. Raghavan. Big data: Promises and problems. *Computer*, 48(03):20–23, 2015.
- [2] A. Mittal. Trustworthiness of big data. *International Journal of Computer Applications*, 80(9), 2013.



Figure 1: Toto je obrazok 1



Figure 2: Toto je obrazok 2