CITIESData: Towards Cloud Based Big Data Management for Smart Cities

Liu, Xiufeng; Heller, Alfred; Nielsen, Per Sieverts

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
With the increasing presence of Internet of Things (IoT) and future internet technologies in smart cities, a large amount of data are generated. The data need to be properly managed and analyzed for various application using integrated ICT approach. The ICT technologies for a smart city will deal with the data from different domains, including environmental, energy, transportation and many others. We present a cloud-based ICT platform that can collect, store, share/publish, analyze, and visualize scalable data from city environment.

The goals:
- Data quality checking and improvement
  - Cleansing data before publishing
- Security and privacy protection
  - Classify data according to different risk levels
  - Using different sharing/publishing strategies

The architecture:
- Apply the virtual machine (VM) based secured environment for using highly sensitive data
- Use the cloud-based data management system, OwnCloud, for semi-sensitive data sharing
- Use the open data platform, Zenodo, for indexing, and sharing

The process:
- Retrieve duplicate rows
- Do the anonymization:
  - Replace household ID with surrogate key
  - Suppress the values of ID attribute, none
  - Generalize the values of QI attribute, address
- Compute the two measures, usageValue and heat/Energy
- In missing values

The architecture of big data management system

Data quality checking model
\[ f = \frac{1}{n} \sum_{i=1}^{n} (\omega_i \times y_i) \]  
where \( f \) is the overall data quality score, \( y_i \) is the data quality of determine attribute \( i \), and \( \omega_i \) is its weight.

Anonymization methods and software package
- CITIES data management system
  a. A scalable data processing platform
  b. Data cleaning, analytics and visualization

References

Acknowledgements
This research was supported by the CITIES research project (NO. 1035-00027B) funded by Innovation Fund Denmark.