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):
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
- 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
- Generate the values of 51 attribute, address
- Compute the two measures, usageValue and heatEnergy
- Compute the quality score
- Data quality checking model

\[ f = \sum_{i=1}^{n} \omega_i \gamma_i + \omega_{n+1} + \cdots + \omega_{n+k} = 1.0 \] (1)

where \( f \) is the overall data quality score, \( \gamma_i \) is the data quality of determine attribute \( i \), and \( \omega_i \) is its weight.

Conclusions
- We have proposed a smart cities data management framework
- Proposed the method of publishing/sharing data according to different data sensitivity levels.
- Proposed linear regression based data quality checking method
- Implemented a smart cities data platform for streamlining the data management process
- The cities data platform has good performance supporting big data management towards the Cloud

References