\ge

Xiufeng Liu*, Alfred Heller, and Per Sieverts Nielsen xiuli@dtu.dk

Introduction

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.



Fig 1. Using the Cloud to store data generated from different smart city components [1] Smart city data characteristics (5Vs but more Vs are possible):



Methods

The goals:

- Data quality checking and improvement
 - Cleansing data before publishing • Instruction of data quality
- Security and privacy protection

 - Classify data according to different risk levels Using different sharing/publishing strategies 0

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



CITIESData: Towards Cloud Based Big Data Management for Smart Cities

Results

• Data quality checking model

$$f = \sum_{i=0}^{n-1} \omega_i * y_i, \ \omega_0 + \omega_1 + \dots + \omega_{n-1} = 1.0$$
(1)

where f is the overall data quality score, y_i is the data quality of determinate attribute *i*, and ω_i is its weight.

• Anonymization methods and software package

CITIES data management system

- A scalable data processing platform a.
- b. Data cleansing, analytics and visualization

	JODS+	wetadata -	Accounts		
TL-DistrictHeating					
0. Preparation					
%sh					
sudo rm -rf /tmp/* mkdir -p /tmp/data chmod -R 777 /tmp/data					
1. Copy data files from the remote FTP serv	ver (ftp.risoe	e.dk) to staging	g area on the l	cal server (cities.deic.dk)	
<pre>%sh scp -r ftp.risoe.dk:/home/xiuli/DataForUpload/* /tmp for f in \$(ls /tmp/data/*); do sed -i -e 's/,/./g'</pre>	/data \$f; done				
2. Copy data to the database					
<pre>%sh for f in \$(ls /tmp/data/*); do psql cities -c "COPY</pre>	(tbl_sonderborg	g_districtheating_	raw from '\$f' DEL	MITER ';' CSV"; done	
3. In-database data cleansing	indata()"				
A Evenet beyond dolly dots	.ingdata()"				
4. Export hourly and daily data					
<pre>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>></pre>	ur,cons_volume A us_volume) AS vo from readdate)	AS volume, cons_he olume, sum(cons_he as readmonth.sum(atenergy AS heater atenergy) AS heater cons volume) AS vo	ergy FROM tbl_sonderborg_districtheating) T0 '/tmp/allmeters_hourly.csv' DELIMI mergy FROM tbl_sonderborg_districtheating GROUP BY 1,2 ORDER BY 1,2) T0 '/tmp/a lume, sun(coms heatenergy) AS heatenergy FROM tbl sonderbora districtheating GR	TER ',' llmeters
5. Generate data summary					
5. Generate data summary					
5. Generate data summary					
5. Generate data summary %h out=/tmp/netadata.txt echo ^{**-s} out					
5. Generate data summary %sh out=/tnp/netadata.txt echo **>Sout					
5. Generate data summary Ssh out=/tmp/netadata.txt echo **>Sout CITUES Data Admin ubsc Metadata	- Accounts	Internetter			Connected
5. Generate data summary Ssh out=/tmp/netadata.txt echo **>Sout CITIES Data Admin Jobs- Metadata-	✓ Accounts	Interpreter			Connected
5. Generate data summary sch out=/tmp/netadata.txt cho ">Sout CITIES Data Admin Jobs- Metadata- Data-SonderborgDistrictHeating	- Accounts	Interpreter		•	Connected
5. Generate data summary Ssh out-/tnp/netadata.txt echo ">Sout CITIES Data Admin Jobs Metadata- Data-SonderborgDistrictHeating - The data are sagregated into hourd, daily, and monthy, which are saved in the the metadata or deatable the metadata of the data sets, including the the the metadata of the data sets and under the	 Accounts to separate CSV files: number of meters, r 	Interpreter	limeters_dally.csv and all	eters_nonthly.cay;	Connected
5. Generate data summary Ssh out-/top/netadata.txt echo ">Sout CITIES Data Admin Jobs Metadata- Data-SonderborgDistrictHeating The data are aggregated his body, daby, and monthy, which are stand in the financial science data science and a science of the	 Accounts to separate CSV files: te number of meters, u could view the data it 	Interpreter calimeters_hourty.cov, at missing values, min and m in babular and different cha	limeters_dally.csv and all ax date, etc.; rts format, and play with d	eters_monthly.cav; a by DRAG & DROP for fields into the Keys, Groups and Values boxes. In Values, you could also choose the aggreg	Connected gation funct
5. Generate data summary Not out=/op/netadata.txt echo *>Sout= CITIES Data Admin Jobs Metadata Data-SonderborgDistrictHeating The data are aggregated this houry, tably, and monthy, which are saved in 1 You coal data much the data sets from the data sets, including the 1 You coal data much the data sets from the data sets, including the 1 You coal data much the data sets from the data sets, including the 1 You coal data much the data sets from the data sets, including the 1 You coal data much the data sets from the data sets, including the 1 You coal data much the data sets from the data sets, including the 1 You coal data much the data sets from the data sets, including the 1 You coal data much the data sets from the data sets, including the 1 You coal data sets of the data sets from the data sets, including the 1 You coal data much the data sets from the data sets, including the 1 You coal data much the data sets from the data sets, including the 1 You coal data sets of the data sets from the data sets, including the 1 You coal data sets of the data sets from the data sets, including the 1 You coal data sets of the data sets from the data sets from the data sets, including the data sets from the data sets fr	 Accounts to separate CSV files: te number of meters, number of meters, is u could view the data is 	Interpreter salimetera_houry.cov., at missing values, min and m n tabuar and different cha	Invelors, statiy sav and all ax date, etc.; rits format; and play with d	eters_monthly.csv; a by DRAG & DROP the fields into the Keys, Groups and Values boxes. In Values, you could also choose the aggreg	Connected gation funct
5. Generate data summary Sta out=/top/netadata.txt echo =>Sout= CITIES Data Admin Jobs Metadata Data-SonderborgDistrictHeating The data are aggregated this houry, tably, and monthy, which are saved in 1. You can determine the data sets in the data sets, including the 1. You can determine the data sets for the data sets, including the 1. You can determine the data sets for the data sets, including the 1. You can determine the data sets for the data sets, including the 1. You can determine the data sets for the data sets, including the 1. You can determine the data sets for the data sets, including the 1. You can determine the data sets for the data sets, including the 1. You can determine the data sets for the	 Accounts to separate CSV files: to could view the data is a could view the data is 	Interpreter calimeters_bourty.cov, at missing values, min and m in babuse and different cha sesumption	limelers_daily.csv and all ar data, etc.; rts format, and play with d	elers_monthly.csv; a by DRAG & DROP the fields into the Keys, Groups and Values boxes. In Values, you could also choose the aggreg	Connected gation funct
5. Generate data summary Sta out=/top/netadata.txt echo =>Sout= CITIES Data Admin Jobs Metadata Data-SonderborgDistrictHeating The data are aggregated this houry, tably, and monthy, which are saved in the motadata evy decreases the metadata of the data sets, including The data are aggregated this houry, tably, and monthy, which are saved in the motadata evy decreases the metadata of the data sets, including the motadata evy decreases the metadata of the data sets, including the motadata evy decreases the metadata of the data sets, including the motadating sum, say, min, max and count. The transmit At View the consumption distribution of each customer there in Enters	Accounts to separate CSV files: a number of meters, management could view the data	Interpreter calimeters_hourty.csv, al missing values, min and m in tabular and different cha sesumption turne, m3	limeters_dally.cav and all as date, etc.; ns format, and play with d	eters_monthly.cav; a by DRAG & DROP the fields into the Keys, Groups and Values boxes. In Values, you could also choose the aggrey 3. Min value	Connected gation funct
5. Generate data summary Shift out=/tipo/metadata.txt echo **>Sout CITIES Data Admin Jobo Metadata Data-SonderborgDistrictHeating 1. The film etada ask, mickaing in a monthy, which are saved in 2. The data are aggregated into houry, daily, and monthy, which are saved in 3. The film etada ask, mickaing in the metadata of the data sets, including in the data are aggregated into houry, daily, and monthy, which are saved in 3. The film etada ask from the following two chats, where you including aux, are, min, max and court. The true the consumption distribution of each customer Sector 1. Source 1. Sector 1. Sec	Accounts to separate CSV files on could view the data could view the data could view the data could view the data	Interpreter calimeters_hourly.csv, al missing values, min and m in tabular and different cha sesumption sesumption number	Imeters_dally.csv and all ax data, etc.; rts format, and play with d	eters_nonthily.cav; a by DRAG & DROP the fields into the Kays, Groups and Values boxes. In Values, you could also choose the aggrey 3. Min value 3. Min value 5. Start data 4. Start data 5. Start data	Connected gation funct
5. Generate data summary Note:/ tup/metadata.txt echo **>Sout CITIES Data Admin Jobo Metadata CITIES Data Admin Jobo Metadata CITIES Data Admin Jobo Metadata CITIES Data Admin Jobo Metadata Data Sounder Jobo Metadata CITIES Data Admin Jobo Metadata Metadata Metadata Sounder Jobo Metadata Metadata Sounder J	Accounts to separate CSV files a could view the data z.cc v v v s.bb 20	Interpreter calimeters_hourly.csv, al in tabular and different cha sesumption sume, m3 n number 2	Imeters_daily.csv and all ac date, etc.; rla format; and play with d	eters_monthly.csv; a by DRAG & DROP the fields into the Keys, Groups and Values boxes. In Values, you could also choose the aggres a. Min value a. Start data (2014-11-207)	Connected
5. Generate data summary Market Strategy and Strategy an	Accounts to separate CSV files a could view the data	Interpreter calimeters_hourly.csv, al in tabular and different cha instantiation of the second different cha assumption aume, m3 n number	Imeters_daily.csv and all ax data, etc.; rta format; and play with d	eters_monthly.csv; a by DRAG & DROP the fields into the Keys, Groups and Values boxes. In Values, you could also choose the aggres 3. Min value 3. Min value 4. Sert Sale 5. Sert Sale 5. Sert Sale 5. Sert Sale 5. Sert Sale	Connected gation funct
5. Generate data summary	Accounts iso separate CSV files: exceeding of the term of ter	Interpreter salimetera_bourty.cov, al missing values, mir and m in tabular and different cha sessamption same, m3 n number 0	limeters, statily care and all ax date, etc.; rits format; and play with d	eters_monthly.csv; a by DRAG & DROP the fields into the Keys, Groups and Values boxes. In Values, you could also choose the aggres 2. Min value 3. Min val	gation funct
5. Generate data summary S. Generate data summary C. CITIES Data Admin Data Data Data Data Data Data Data Dat	Accounts to separate CSV files to acquire the data i	Interpreter Calimeters, hourly, csv, al inisaigo values, min and inisaigo values, min and different cha sessemption aume, m3 an number b	Imeters_dally.csv and all ax data, etc.; rta format; and play with d	eters_monthly.csv; a by DRAG & DROP the fields into the Keys, Groups and Values boxes. In Values, you could also choose the aggres a. Min value a. Start date 2014.11.207	Connected
5. Generate data summary S. Generate data summary C. CITIES Data Admin Data Control of the data set generated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, which are saved in the data are aggregated into houry, daily and monthly, daily and monthly, daily and monthly, daily and monthly, when you the data are aggregated into houry, daily and monthly, daily and daily and daily and daily and da	Accounts to separate CSV files: uncoded view the data i v v v v v v v coded view the data i	Interpreter Calimeters Jourly cav, al inisisig values, min and in tabular and different cha aume, m3 a	Imsters, dailycesv and all ax date, etc.; rits format; and play with d	eters_monthly.cosy by DRAG & DROP the fields into the Keys, Groups and Values boxes. In Values, you could also choose the aggrey . Min value . So the set 2. Start data 2. Start data	Connected
5. Generate data summary	Accounts	Interpreter calinesters, hourly say, at missing values, min and m in tabular and different cha assumption a number o	Imsters_daily.csv and all ax date, etc.: rits format; and play with d	eters_monthlyces; by DRAG & DROP the fields into the Keys, Groups and Values boxes. In Values, you could also choose the aggrey a. Min value a. So that also c. Start al	Genected gation funct
5. Generate data summary S. Generate data summary sub-reprint the sub-reprint the sub-reprin	Accounts	Interpreter calibrations, bourly cost, ad- miciality solutes, min and m in tabular and different cha sessamption n number o sessamption se	Invelors, daily case and all ax date, etc.: rits format; and play with d	Neters_monthly.car; a by DRAG & DROP the fields into the Kays, Groups and Values boxes. In Values, you could also choose the aggres . Min value . Shore the . Shore t	Gannested gation funct
5. Generate data summary S. Generate data summary sub-reprint the sub-reprint the sub-reprin	Accounts	Interpreter salimating values, min and m missing values, min and m missing values, min and different cha sesamption n studies end different cha sesamption n number o sesamption	Invelors, daily core and all ax date, etc.: rits format; and play with d		Gannested gation funct
5. Generate data summary S. Generate data summary S. dut-forp/netadata.txt cho "> <pre>South cho "><pre>South cho "><pre>South cho "><pre>South cho "><pre>South cho "><pre>South cho "><pre>South cho "></pre> CITIES Data Admin</pre></pre></pre></pre></pre></pre>	Accounts iso separate GSV files: exception of metars, in code view the data code view the data code view the data code code	Interpreter calinetera_boury.cov, al missing values, min and m missing values, min and m missiour and different cha exemption number o ga ga ga	Invetors, daily case and all ax date, etc.: rifs format; and play with d	weers_monthly.cav; aly DRAG & DROP the fields into the Kays, Groups and Values boxes. In Values, you could also choose the aggreg	gation funct
5. Generate data summary	Accounts to separate CSV fless to an unified of meles, is ocoud view the data	Interpreter calimeters_hourty.csv, almissing values, min and m in tabular and different cha securetion number o ge ge	Invetors, daily case and all actions of the second se	eters_monthly.csv; ab y DRAG & DROP the fields into the Keys, Groups and Values boxes. In Values, you could also choose the aggreg	gation funct
5. Generate data summary	Accounts to separate CSV files the number of meters, in could view the data z.ce v v v v v could view the data z.ce could view the data z.ce could view the data could view the data could view the data	Interpreter Calimeters_hourly.csv, al missing values, min and m in solutar and different cha server m	Imelers_daily.csv and all ar data, etc.; etc format; and palay with d	elers_monthly.csx; a by DRAG & DROP the fields into the Keys, Groups and Values boxes. In Values, you could also choose the aggres 3. Min value 3. Min value 3. Sint value 4. Sint data 5. Sint	gation funct
S. Generate data summary S. Generate data summary List C. CITIES Data Admin C. Metadate C. CITIES Data Admin C. Metadate C. CITIES Data Admin C. Metadate C. Control of the second se	Accounts to separate CSV files: a could view the data	Interpreter calimeters_houry.esv, al missing values, min and m in sabular and different cha aurus, m3 aurus, m3 aurus, m3 aurus, m3 aurus, m4 aurus, m3 aurus, m4 auru	Imeters_dally.csv and all ac date, etc.; ns format; and play with d	eters_monthly.cav; a by DRAG & DROP the fields into the Keys, Groups and Values boses. In Values, you could also choose the aggrey 3. Min value a. Sourt data 5. Sourt d	gation funct
S. Generate data summary Stury (thp/netadata.txt echo **>Sout CITTIES Data Admin Job Metadata CITIES Data Admin J	Accounts to separate CSV files to number of neters, a codd view the data i	Interpreter Calineters_houry, esv, ali calineter	Imeters_dally.csv and all ac date, etc.; Its format; and play with d	eters_nonthily.cav; a by DRAG & DROP the fields into the Kays, Groups and Values boses. In Values, you could also choose the aggres a. Min value a. Min va	Connected gation funct

Fig. 4 CITIES data management system





- 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

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

References

- 2.



Conclusions

• We have proposed a smart cities data management

Acknowledgements

1. X. Liu, A. Heller, and PS Nielsen. Research data management for smart cities, In submission to Journal of Information Management. CITIES Data Platform. http://cities.deic.dk