# Weekly Report – 48

Xiufeng Liu

University of Waterloo, CA
`xiufeng.liu@uwaterloo.ca`

## 1 This Week

In this week I continue to study the `ESSEX` data set, and complete the study report. Also, in the group meeting on Wednesday and Thursday, we outlined the route of our research. Before we make a centre data warehouse system for data analysis, we will first continue current work that is to do the benchmark study of energy data on different systems, including Matlab, PostgreSQL+Matlib, KDB+ and other column stores.

## 2 Next Week

I will start to investigate the technologies that could integrate the analysis results available, then consider to build a prototype of an web-based analysis platform. The analytics results from different sources could be easily "added into" the platform and visualized. This is an on-going project that will be completed progressively to the next year. For the benchmark study, we have to outline the work and take steps in soon.

## 3 ESSEX Data

Fig. 1 shows the database schema of the annotated ESSEX data (The tables colored by blue are the original tables, while the ones colored by purple are the annotated tables). To ease understanding, I have classified the tables into different categories.

### 3.1 Energy Production

The data in this table is the metering data from the solar power systems at different sites. A solar power system exploits photovoltaic (PV) technology that converts sunlight into electrical energy. [1] explains the mechanism. That is, PV technology uses solar cells made of semiconductor materials, such as silicon or germanium dosed with small amounts of impurities (typically metals or metalliods). When sunlight strikes a cell, a certain portion of its energy is absorbed with the semiconductor material, the absorbed energy knocks the electrons, and the loosen electrons flow freely under the influence of electric fields. Solar cells have inbuilt electric fields that force the freed electrons to flow in a certain direction. Metal contacts on the top and bottom of the PV cell enable the cell to generate a current in an external circuit. This current, together with the cell's voltage (which is a result of its in-built electric fields), defines the power (or wattage) that a solar cell can produce. This *direct current (DC)* can be used to recharge batteries, and run direct current devices, or

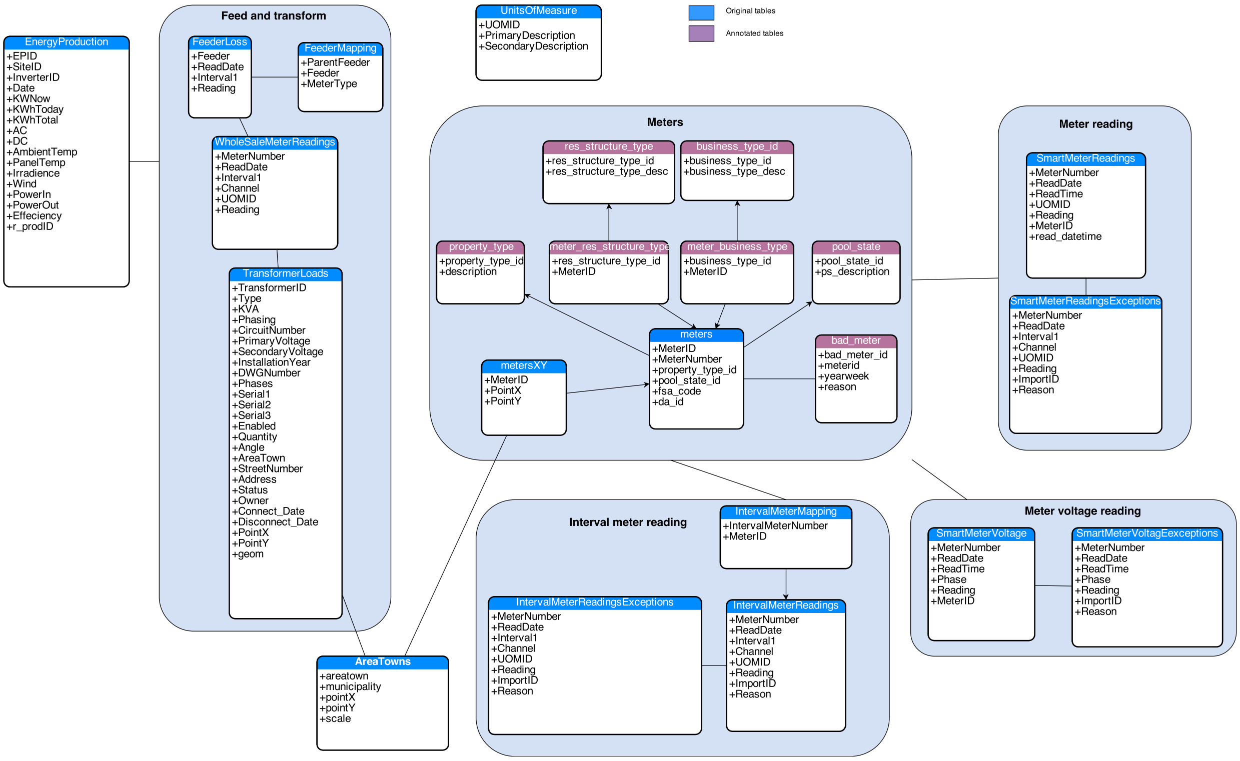


**Fig. 1.** Annotated ESSEX data

**Table 1.** EnergyProduction

| Field | Type | Example | Description |
|---|---|---|---|
| EPID | int(11) | 1 | Sorrogate key of energy production |
| SiteID | int(11) | 1 | The site of solar power system |
| InverterID | int(11) | 1 | The inverter for converting the direct current to alternate current |
| Date | timestamp | 2010-04-26 15:35:49 | The timestamp of metering the data. Here it reads every 5 minutes |
| KWNow | float | 2.593 | The reading at a point of the time |
| KWhToday | float | 20 | The accumulated reading of a site of today |
| KWhTotal | float | 2491 | The accumulated reading of a site |
| AC | float | 256 | The alternative current after convertion |
| DC | float | 325 | The produced direct current of electrons |
| AmbientTemp | float | 18 | The ambient air temperature around solar panel. The efficiency of PV can be affected by its operating temperature, which is primarily a product of the ambient air temperature as well as the level of sunlight. High ambient temperature could lower the energy generation efficiency. |
| PanelTemp | float | 35 | The temperature of panel. The efficiency could also be affected by the panel temperature |
| Irradience | float | 481 | Measure the spectral irradiance in units of watts per meter squared |
| Wind | float | 12 | The wind speed on the site of measure. The wind speed could be corelated to the efficiency |
| PowerIn | float | 2915.25 | The power received from network |
| PowerOut | float | 2554.88 | The power exported to network |
| Effeciency | float | 87.64 | |
| r_prodID | int(11) | 1999693 | Energy production ID |

can be converted via *inverters* into *alternating current (AC)*, the form of electricity most commonly used in homes, offices and industry.

The table `EnergyProduction` contains 50 sites' data, and the data granularity is 5 minutes, 2,929,404 records in total. The measures of this table include the generated energy, i.e., `KWNow, KWhToday, KWhTotal`, the ambient and panel temperatures, irradience, efficiency, and wind. The dimensions include the sites of the solar power systems, the time, and the inverter of converting DC to AC. From this table we could ask some business questions from different perspectives (dimensions), and do the forecasting and correlation analysis. For example, to view the energy production for each of the inverters in time series fashion, study the impact of wind speed, panel and ambient, irradiance to the energy generation efficiency (by correlation), and do the forecasting, etc.

### 3.2 Feed and transform

The power can be traded in the energy market. A utility company can buy the power from the market, and sell to the customers of a region, which is called *wholesale*. The power is distributed through electricity distribution network from transmission systems to the customers. Typically, the distribution network includes medium-voltage power lines, substations and transformers, low-voltage distribution wiring and meters (see Fig. 2 from [2]). A distribution substation receives power from one or more transmission or sub-transmission lines at the corresponding transmission or sub-transmission voltage level, and provides the power to one or more distribution feeders that originate from the substation. Most feeders emanate radially from the substation to supply the load. Some feeding loss occurs when the power is transferred from a substation to another substation. Typically, urban and suburban distribution is done with three-phase systems to serve both residential, commercial, and industrial loads. Distribution in rural areas may be only single-phase if it is not economical to install three-phase power for relatively few and small customers [3].

The `WholesaleMeterReadings` table contains the wholesale data of utility companies. The metering data for each channel is read at one-hour interval per day, and each read was repeated for 30 times (a transformer has 4 isolation transformer units, Channels, which are read separately). The data in `FeederMapping` describes the hierarchy structure of feeders, i.e., a substation and its children. `FeederLoss` collects the feeding loss read at one-hour interval per day. `TransformerLoads` describes the load of transformers, Circuit number, Phases, angle, address, and the geography locations etc. The transformation load is collected at one-hour interval per day. But, note that in this table there are 22 columns without any values (marked by "N/A").

**Table 2.** WholesaleMeterReadings

| Field | Type | Example | Description |
|---|---|---|---|
| MeterNumber | int(11) | 200700539 | Meter number |
| ReadDate | date | 2011-09-01 | The reading date |
| Interval1 | int(11) | 1 | The interval of reading meter, 4 channels in total |
| Channel | int(11) | 1 | Transformer channel |
| UOMID | int(11) | 1 | |
| Reading | int(11) | 1249 | The value of the reading |

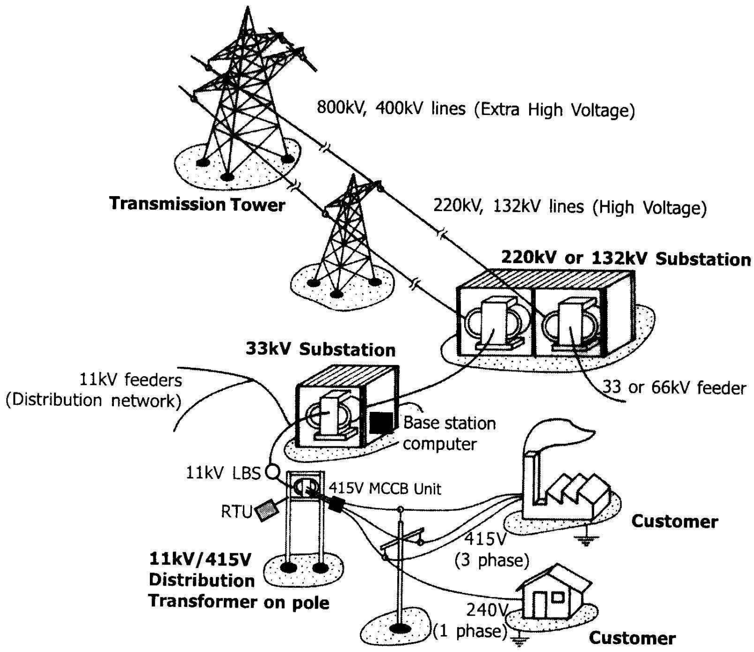


**Fig. 2.** Electricity Distribution Network

**Table 3.** FeederMapping

| Field | Type | Example | Description |
|---|---|---|---|
| ParentFeeder | varchar(250) | 23M4 | Parent feeder ID |
| Feeder | varchar(250) | FSF1 | The feeder ID |
| MeterType | varchar(250) | SM | Meter type |

**Table 4.** FeederLoss

| Field | Type | Example | Description |
|---|---|---|---|
| Feeder | varchar(250) | 23M3 | Feeder ID |
| ReadDate | timestamp | 2012-06-01 00:00:00 | The date of read |
| Interval1 | int(11) | 1 | The interval of reading. It is 1-hour interval from 0 to 23 |
| Reading | float | -0.898 | The value of reading |

**Table 5.** TransformerLoads

| Field | Type | Example | Description |
|---|---|---|---|
| TransformerID | int(11) | 9327 | The transformer ID |
| Type | date | 2011-04-03 | The date of read |
| KVA | float | 0 | The interval of read, 0–23 |
| Phasing | float | 2.05 | The phase value of electricity |
| CircuitNumber | varchar(250) | N/A | |
| PrimaryVoltage | varchar(250) | N/A | |
| SecondaryVoltage | varchar(250) | N/A | |
| InstallationYear | varchar(250) | N/A | |
| DWGNumber | varchar(250) | N/A | |
| Phases | varchar(250) | N/A | |
| Serial1 | varchar(250) | N/A | |
| Serial2 | varchar(250) | N/A | |
| Serial3 | varchar(250) | N/A | |
| Enabled | varchar(250) | N/A | |
| Quantity | varchar(250) | N/A | |
| Angle | varchar(250) | N/A | |
| AreaTown | varchar(250) | N/A | |
| StreetNumber | varchar(250) | N/A | |
| Address | varchar(250) | N/A | |
| Status | varchar(250) | N/A | |
| Owner | varchar(250) | N/A | |
| Connect_Date | date | N/A | |
| Disconnect_Date | date | N/A | |
| PointX | float | N/A | |
| PointY | float | N/A | |
| geom | varchar(250) | N/A | |

### 3.3 Smart Meter Voltage Reading

Transformers transform medium-high voltage power to low voltage before feeding to the customers. The levels of voltage around the world is different. For example, in North America the voltage value is between 100 and 127 volts, while in European countries it is between 220 and 240 volts. The voltage readings beyond this range are regarded as the exception. `SmartMeterVoltage` contains the voltage readings of three phases collected at 15-minute interval per day, and the exceptional readings are added to `SmartMeterVoltageException` for each read. This table also contains the no. of exceptions. The fact is the readings, and the dimensions are the dates, time, phases, and meters.

**Table 6.** SmartMeterVoltage

| Field | Type | Example | Description |
|---|---|---|---|
| MeterNumber | varchar(250) | EP00001280 | Meter number |
| ReadDate | date | 2011-12-21 | |
| ReadTime | time | 00:00:15 | The reading time interval in each day, 15-minute interval |
| Phase | int(11) | 2 | The power phases: 0, 1 and 2 |
| Reading | float | 240.837 | The value of voltage |
| MeterID | int(11) | 19931 | |

**Table 7.** SmartMeterVoltageException

| **Field** | **Type** | **Example** | **Description** |
|---|---|---|---|
| MeterNumber | varchar(250) | EP00018137 | Meter number |
| ReadDate | date | 2011-12-21 | Reading date |
| ReadTime | time | 00:00:00 | The reading time |
| Phase | float | 0 | The power phases: 0, 1 and 2 |
| Reading | float | 608.075 | The value of voltage |
| ImportID | int(11) | 2346 | |
| Reason | int(11) | 2 | |

## 3.4 Smart Meter Reading

The tables in this category hold the power data consumed by customers. `Meters` contains the mappings of `MeterID` to `MeterNumber`. `SmartMeterReadings` is holding the meter reading of *1-minute* interval of each meter, and the exception readings are hold in `SmartMeterReadingsExceptions`. The geography information of meters is in `MeterXY`, including the longitude and latitude on map. The fact is the meter readings, and the dimensions are date, time, and meter ID, and location.

**Table 8.** Meters

| **Field** | **Type** | **Example** | **Description** |
|---|---|---|---|
| MeterID | int(11) | 19419 | Meter ID |
| MeterNumber | int(11) | 1 | Meter number |

**Table 9.** SmartMeterReadings

| **Field** | **Type** | **Example** | **Description** |
|---|---|---|---|
| MeterNumber | varchar(250) | EP00010116 | Meter number |
| ReadDate | date | 2010-10-29 | Reading date |
| ReadTime | time | 00:04:00 | Read time, at 1-minute interval |
| UOMID | int(11) | 1 | |
| Reading | float | 0.74 | The value of reading |
| MeterID | int(11) | 22220 | Meter ID |

**Table 10.** SmartMeterReadingsExceptions

| **Field** | **Type** | **Example** | **Description** |
|---|---|---|---|
| MeterNumber | int(11) | 323 | Meter number |
| ReadDate | timestamp | 2011-03-03 00:00:00 | Reading date |
| Interval1 | int(11) | 500 | |
| Channel | int(11) | 1 | A smart meter typically has several channels that user can program for different tariff rates |
| UOMID | int(11) | 1 | |
| Reading | int(11) | 694 | The value of exceptional reading |
| ImportID | int(11) | 2 | The import ID of the meter that has exceptional reading |
| Reason | int(11) | N/A | The number of reason |

**Table 11.** MetersXY

| **Field** | **Type** | **Example** | **Description** |
|---|---|---|---|
| MeterID | int(11) | 19420 | Meter ID |
| PointX | int(11) | 329888 | Longitude |
| PointY | int(11) | 4678339 | Latitude |

## 3.5 Interval Meter Reading

Interval Meter Readings are at 1-hour interval for each of the channels in a meter. Typically, a meter has multiple channels, each of which show a different tariff. A utility companies can set the rate of power in each channel. The readings are in `IntervalMeterReadings`. The exceptional readings are in `IntervalMeterReadings-Exceptions`. Therefore, the fact is the meter readings that could be viewed from the dimensions, reading date, time, Channels, Meter ID and locations.

**Table 12.** IntervalMeterMapping

| **Field** | **Type** | **Example** | **Description** |
|---|---|---|---|
| IntervalMeterNumber | int(11) | 814110 | Interval meter number |
| MeterID | int(11) | 19894 | Meter ID |

**Table 13.** IntervalMeterReadings

| Field | Type | Example | Description |
|---|---|---|---|
| MeterNumber | int(11) | 816510 | Meter number |
| ReadDate | timestamp | 2011-09-01 00:00:00 | Reading date |
| Interval1 | int(11) | 1 | Read interval, 1hour |
| Channel | int(11) | 1 | A channel showing a tariff where the rate can be set |
| UOMID | int(11) | 1 | |
| Reading | int(11) | 306 | Reading |
| ImportID | int(11) | 29485 | |
| Reason | int(11) | N/A | |

**Table 14.** IntervalMeterReadingsExceptions

| Field | Type | Example | Description |
|---|---|---|---|
| MeterNumber | int(11) | 816510 | Meter number |
| ReadDate | timestamp | 2011-09-01 00:00:00 | Reading date |
| Interval1 | int(11) | 1 | |
| Channel | int(11) | 1 | A channel showing a tariff where the rate can be set |
| UOMID | int(11) | 1 | |
| Reading | int(11) | 306 | Reading |
| ImportID | int(11) | 29485 | |
| Reason | int(11) | N/A | Exceptional reason |

## 3.6 Other Tables

Both are self-explained.

**Table 15.** UnitsOfMeasure

| Field | Type | Example | Description |
|---|---|---|---|
| UOMID | int(11) | 1 | |
| PrimaryDescription | varchar(250) | KWH | |
| SecondaryDescription | varchar(250) | KW | |

**Table 16.** AreaTowns

| Field | Type | Example | Description |
|---|---|---|---|
| areatown | varchar(250) | A1-TEC | |
| municipality | varchar(250) | Tecumseh | |
| pointx | varchar(250) | N/A | |
| pointy | varchar(250) | N/A | |
| scale | varchar(250) | N/A | |

## 4 Paper Reading

The purpose of the weekly paper reading keeps me update the algorithms of data analysis and forecasting in energy. Followings are the papers I read this week.

- **L. Wang, L. Tanm C. Yu, and Z. Wu. "Study and Application of Non-Linear Time Series Prediction in Ground Source Heat Pump System". In *Proc. of CECNet*, pp. 3522–3525, 2012.**
  Lu et al. uses the two different algorithms,(SMOreg) Sequential Minimal Optimization for Regression [4] and M5P(model tree) [5], to do the forecasting of geothermal energy utilization of ground source heat pump system, which is the non-linear prediction. They argued that the selected two algorithms have better accuracy than the other algorithms used in linear prediction, such as Microsoft time series, AR model, Cloud Model, etc. But, the paper does not compared with any of these algorithms. Instead, they test the two selected algorithms using the real data sets, and find that M5P shows a better accuracy comparing with SMOreg.
- **S. Murugesan, J. Zhang, and V. Vittal. "Finite State Markov Chain Model for Wind Generation Forecast: A Data-driven Spatiotemporal Approach". In *Proc. of ISGT*, pp. 1–8, 2012.**
  This paper uses Markov chain model to forecast the aggregate power output from a wind farm. The model considers both spatial and temporal dynamics of wind power output of turbines in the farm. The temporal dynamics of the aggregate wind power is analyzed using auto-regression, while taking into account the diurnal non-stationarity and the seasonality.
- **D. Tsoumakos, and C. Mantas. "The Case for Multi-engine Data Analytics".** `http://www.cslab.ece.ntua.gr/~dtsouma/index_files/MEMS_mhpc2013.pdf`
  This is the position paper that envisions an analytics platform with Multi-Engine Management System (MEMS). The authors first study the state-of-the-art technologies for data analytics today, and the challenges. They found that due to the diversity of the analytics requirements, the data sources, and the hardware conditions, etc., users typically use different systems to do their analytics. They think it is desirable to build a unied analytics system for managing, executing and monitoring multiple, complex jobs. But, inside the system there lies multiple Management Systems that provide adaptive, cost-based and customizable resource management for diverse resources.

## References

1. J. Toothman and S. Aldous. "How Solar Cells Work". `http://science.howstuffworks.com/environmental/energy/solar-cell1.htm` as of Nov 25, 2013.
2. http://www.iitk.ac.in/infocell/Archive/dirimg/power.jpg as of Nov 25, 2013.
3. R. E. Brown, "Electric Power Distribution Reliability". Marcel Dekker, Inc., 2002.
4. C. Li, L. Jiang. "Using Locally weighted learning to improve SMOreg for regression[R]". Hubei: PRICAI 2006,2006:375384.
5. J. Yang, Y. Zhai, D. Xu, et al. "SMO Algorithm applied in time series model building and forecast". In *Proc. of ICMLC*, 2007:2395-2400, 2012.