**NBD(NIST Big Data) Requirements WG Use Case Template Aug 11 2013**

|  |  |  |  |
| --- | --- | --- | --- |
| **Use Case Title** | | Consumption forecasting in Smart Grids | |
| **Vertical (area)** | | Energy Informatics | |
| **Author/Company/Email** | | Yogesh Simmhan, University of Southern California, simmhan@usc.edu | |
| **Actors/Stakeholders and their roles and responsibilities** | | Electric Utilities, Campus MicroGrids, Building Managers, Power Consumers, Energy Markets | |
| **Goals** | | Develop scalable and accurate forecasting models to predict the energy consumption (kWh) within the utility service area under different spatial and temporal granularities to help improve grid reliability and efficiency. | |
| **Use Case Description** | | Deployment of smart meters are making available near-realtime energy usage data (kWh) every 15-mins at the granularity individual consumers within the service area of smart power utilities. This unprecedented and growing access to fine-grained energy consumption information allows novel analytics capabilities to be developed for predicting energy consumption for customers, transformers, sub-stations and the utility service area. Near-term forecast can be used by utilities and microgrid managers to take preventive action before consumption spikes cause brown/blackouts through demand-response optimization by engaging consumers, bringing peaker units online, or purchasing power from the energy markets. These form an OODA feedback loop. Customers can also use them for energy use planning and budgeting. Medium- to long-term predictions can help utilities and building managers plan generation capacity, renewable portfolio, energy purchasing contracts and sustainable building improvements.  Steps involved include *1) Data Collection & Storage*: time-series data from (potentially) millions of smart meters in near-realtime, features on consumers, facilities and regions, weather forecasts, archival of data for training, testing and validating models; *2) Data Cleaning & Normalization*: Spatio-temporal normalization, gap filling/Interpolation, outlier detection, semantic annotation; *3) Training Forecast Models*: Using univariate timeseries models like ARIMA, and data-driven machine learning models like regression tree, ANN, for different spatial (consumer, transformer) and temporal (15-min, 24-hour) granularities; *4) Prediction*: Predict consumption for different spatio-temporal granularities and prediction horizons using near-realtime and historic data fed to the forecast model with thresholds on prediction latencies. | |
| **Current**  **Solutions** | **Compute(System)** | | Many-core servers, Commodity Cluster, Workstations |
| **Storage** | | SQL Databases, CSV Files, HDFS, Meter Data Management |
| **Networking** | | Gigabit Ethernet |
| **Software** | | R/Matlab, Weka, Hadoop |
| **Big Data  Characteristics** | **Data Source (distributed/centralized)** | | Head-end of smart meters (distributed), Utility databases (Customer Information, Network topology; centralized), US Census data (distributed), NOAA weather data (distributed), Microgrid building information system (centralized), Microgrid sensor network (distributed) |
| **Volume (size)** | | 10 GB/day; 4 TB/year *(City scale)* |
| **Velocity**  **(e.g. real time)** | | Once every 15-mins (~100k streams); Once every 8-hours (~1M streams) |
| **Variety**  **(multiple datasets, mashup)** | | Tuple-based: Timeseries, database rows; Graph-based: Network topology, customer connectivity; Some semantic data for normalization. |
| **Variability (rate of change)** | | Meter and weather data change, and are collected/used, on hourly basis. Customer/building/grid topology information is slow changing on a weekly basis |
| **Big Data Science (collection, curation,**  **analysis,**  **action)** | **Veracity (Robustness Issues, semantics)** | | Versioning and reproducibility is necessary to validate/compare past and current models. Resilience of storage and analytics is important for operational needs. Semantic normalization can help with inter-disciplinary analysis (e.g. utility operators, building managers, power engineers, behavioral scientists) |
| **Visualization** | | Map-based visualization of grid service topology, stress; Energy heat-maps; Plots of demand forecasts vs. capacity, what-if analysis; Realtime information display; Apps with push notification of alerts |
| **Data Quality (syntax)** | | Gaps in smart meters and weather data; Quality issues in sensor data; Rigorous checks done for “billing quality” meter data; |
| **Data Types** | | Timeseries (CSV, SQL tuples), Static information (RDF, XML), topology (shape files) |
| **Data Analytics** | | Forecasting models, machine learning models, time series analysis, clustering, motif detection, complex event processing, visual network analysis, |
| **Big Data Specific Challenges (Gaps)** | | Scalable realtime analytics over large data streams  Low-latency analytics for operational needs  Federated analytics at utility and microgrid levels  Robust time series analytics over millions of customer consumption data  Customer behavior modeling, targeted curtailment requests | |
| **Big Data Specific Challenges in Mobility** | | Apps for engaging with customers: Data collection from customers/premises for behavior modeling, feature extraction; Notification of curtailment requests by utility/building managers; Suggestions on energy efficiency; Geo-localized display of energy footprint. | |
| **Security & Privacy**  **Requirements** | | Personally identifiable customer data requires careful handling. Customer energy usage data can reveal behavior patterns. Anonymization of information. Data aggregation to avoid customer identification. Data sharing restrictions by federal and state energy regulators. Surveys by behavioral scientists may have IRB restrictions. | |
| **Highlight issues for generalizing this use case (e.g. for ref. architecture)** | | Realtime data-driven analytics for cyber physical systems | |
| **More Information (URLs)** | | <http://smartgrid.usc.edu>  <http://ganges.usc.edu/wiki/Smart_Grid>  <https://www.ladwp.com/ladwp/faces/ladwp/aboutus/a-power/a-p-smartgridla>  <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6475927> | |
| **Note:** <additional comments> | | | |

**Note: No proprietary or confidential information should be included**