Role of Big Data and Analytics in Smart Cities

Sanjeev Kumar[1], Anand Prakash[2]

[1]Senior Program Manager, HCL Technologies

[2]Senior Test Lead, Tech Mahindra

Abstract: *The aim of this paper is to study the real potential of using Big Data Analytics in Smart Cities. In this work, we studied cases across the globe where decision maker are using Big Data Analytics as a tool for making Smart City. The paper covers how Internet of Things, Machine to machine, Big Data and Smart Cities Linkages can help in doing predictive analytics which can be helpful to human wellbeing. This paper focused on two main areas – Smart Grid and Traffic Congestion Management where Big Data Analytics can be useful for decision makers and city planner. The report includes various pilot project currently undergoing for making a city smarter along with benefits to human wellbeing. The report also considered various challenges that can be encountered while implementing Big Data solution in making Smart Cities.*

Keywords: Big Data, IoT, Smart City, Smart Grid, Traffic Congestion

1. Introduction

Smart City are going to use digital information and communication to improve the urban services w.r.t. quality, performance, human wellbeing. Smart City can contribute to reduction of cost and resource consumption and in effectively and actively engagement with the citizens. Smart City applications are developed with the goal of improving the management of urban flows and allowing for real time responses to challenges.

Number of people living in cities is likely to double by 2050. Many experts believe six billion people will live in cities by 2050, as compared to the 3.6 billion now and this increase is likely to put enormous pressure on the available resources.

Currently, 31% of India's population lives in cities; these cities also generate 63% of the nation's economic activity. These numbers are rapidly increasing, with almost half of India's population projected to live in its cities by 2030. Smart Cities focus on the most pressing needs and on the greatest opportunities to improve quality of life for residents today and in the future.

The Smart Cities Mission is a new initiative by the Government of India to drive economic growth and improve the quality of life of people by enabling local development and harnessing technology as a means to create smart outcomes for citizens.

2. Big Data, Internet of Things and Smart Cities

The term **Big Data** is used to describe a huge volume of both structured and unstructured data which is so large and complex that it is difficult to manage and process using traditional database and software tools. As per Gartner, ―Big Data" is high-volume, -velocity and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making.

Characteristics of Big Data:
- Volume: Volume refers to amount of data.
- Variety: Variety refers to the data from multiple sources in the form of both structured and unstructured format. Data source is diverse. Gartner mentioned it as dark data. Similar to dark matter in physics, dark data cannot be seen directly, yet it is the bulk of the organizational universe.
- Velocity: Velocity deals with the rate at which data flows in from various sources. Big Data flow is massive and continuous.
- Veracity: Veracity refers to uncertainty of data. It deals to quality, trustworthiness, accuracy of data.

The **Internet of Things (IoT)** is the network of physical objects or "things" embedded with electronics, software, sensors, and network connectivity, which enables these objects to collect and exchange data. The IoT allows objects to be sensed and controlled remotely across existing network infrastructure, creating opportunities for more direct integration between the physical world and computer-based systems, and resulting in improved efficiency, accuracy and economic benefit.

The IoT is a computing concept that describes a future where every day physical objects will be connected to the Internet and be able to identify themselves to other devices. Cisco estimates the IoT will consist of 50 billion devices connected to the Internet by 2020.

Smart city means different thing to different people. Its conceptualization varies from city to city, country to country depending on the level of development, willingness to change and reform, resources and aspirations of the city residents. A smart city would have a different connotation in India than, say, Europe. Even in India, there is no one way of defining a Smart city.

In the imagination of any city dweller in India, the picture of a smart city contains a wish list of infrastructure and services that describes his or her level of aspiration. Urban planners ideally aim at developing the entire urban eco-system, which is represented by the four pillars of comprehensive development-institutional, physical, social and economic infrastructure.

12

In Smart Cities, data comes from sensors, devices, video/audio, networks, log files, transactional applications, web, and social media - much of it generated in real time and in a very large scale.
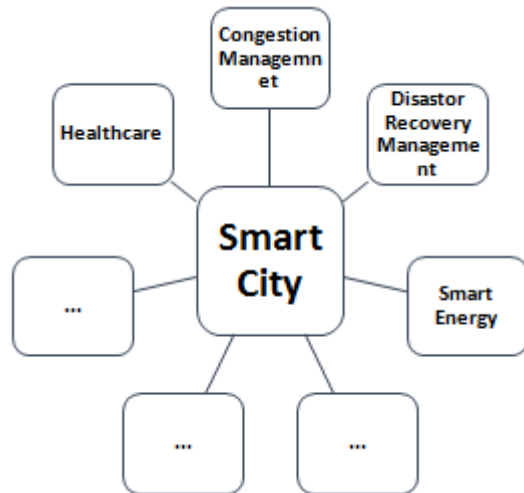

Figure 1: Illustration of Smart City

Smart Cities are cities that use smart infrastructure, the IoT, Machine to Machine connectivity to improve and manage power, resources, and urban planning.

Data integration, Aggregation, Validation, Cleansing

In Smart cities, volumes of data will grow exponentially in all realms from personal to enterprise and global. It will comes from numerous sources like Enterprise systems, Website logs, Web Services, Social Media, Mobile devices, Sensors, GPS, etc. and in various forms - Structured, Semi-structured and Unstructured. It became extremely important to understand data sets and organize them.

Data integration involves combining data from several disparate sources, which are stored using various technologies and provide anamalgamated view of the data. A complete data integration solution involves discovery, cleansing, monitoring, transforming and delivery of data from a variety of sources to integrated location.

In Data aggregation process, information is searched, gathered and presented in a summarized – report, dashboard etc. form which can help in analysis.

Data validation process is required to prevent inaccurate or irrelevant data from entering a program. Data validation ensures that the data complies with the requirements and quality benchmarks. This ensuring that the data sent to connected applications is complete, accurate, secure and consistent.

Data cleansing/cleaning/scrubbing is a process of fixing or removing corrupt or inaccurate data. This process is needed because incorrect data can lead to wrong conclusions and bad analysis, especially when there is huge quantities of Big Data.

Poor data quality is anobstruction to effective and high-quality decision-making.

Data anonymization and visualization
While processing and analyzing the data, it is important that information should be protected so that it cannot be misused. The objective is to be right information to right person at right place. Confidentiality, Integrity and availability should not be compromised.

Data anonymization or masking is information sanitization whose intent is privacy protection. This technique eitherencrypt, hide or remove personally identifiable information from data sets, so that the people whom the data describe remain anonymous.

Data anonymization enables the transfer of information across a boundary, such as between two departments within an agency or between two agencies, while reducing the risk or threat of unintended disclosure. There are a various data anonymization techniques available. Some are mentioned below

Data Encryption
In encryption is about encoding the data. The techniqueshelpnot only in effective data confidentiality, but also transforms data into an unreadable format. For example, once data encryption is applied to the fields containing usernames, "SmithTao" can be ―@ek1ls%&$". The challenges with encryption is sometimes when data input validation or testing require a specific data typelike cost, dates or salary , on encrypted data it may appear to be the wrong data type to the system trying to use it. One needs to decode the data first and then do the validation

Substitution
Substitution is replacing the contents of a data with predefined list of characters or symbols. These characters can be easily traced to the original values whenever required.

Shuffling
Shuffling is similar to substitution, except the anonymized data is derived from the column itself. Essentially the data in a column is randomly moved between rows until there is no longer any reasonable correlation with the remaining information in the row. Here challenge is if the shuffling algorithm is determined, then the data can be easily ―n-shuffled".This technique is not effective when applied on small amounts of data sets. For example, if there are only less rows in a table it probably will not be too difficult to dig out which of the shuffled data really belongs to particular row.

Number & date variance
In this technique,algorithm modifies each value in a column by some random percentage of its real value so that the exact data can be reached to an untraceable point. For e.g., a column having salary details might have a random variance of ±n% placed on it. Some values can be higher, some are lower but all would be not too far from their original range. Similarly dates could be varied with in an arbitrary range of ± n days.

Paper ID: NOV161007

13

Nulling or Truncating

Nulling or truncating consists of eradicating sensitive data by deleting it or replacing with NULL values. This technique mat not be used if vital information needs to be searched with sensitive data. For e.g. it is be difficult to query customer accounts if important information such as customer name, address and other contact details have null values.

Masking Data

In masking data technique, Sensitive information is mask with character (e.g. X). For e.g. In credit or debit card, numbers may look like:

5467 6454 0020 6778
After the masking the information would appear as:
5467 XXXX XXXX 6778

Look and feel would still the same while important information is masked to avoid from stolen.Effective visualization helps users in analyzing and reasoning about data and evidence. It makes complex data more accessible, understandable and usable. Even when data volumes are very large, patterns can be spotted quickly and easily.
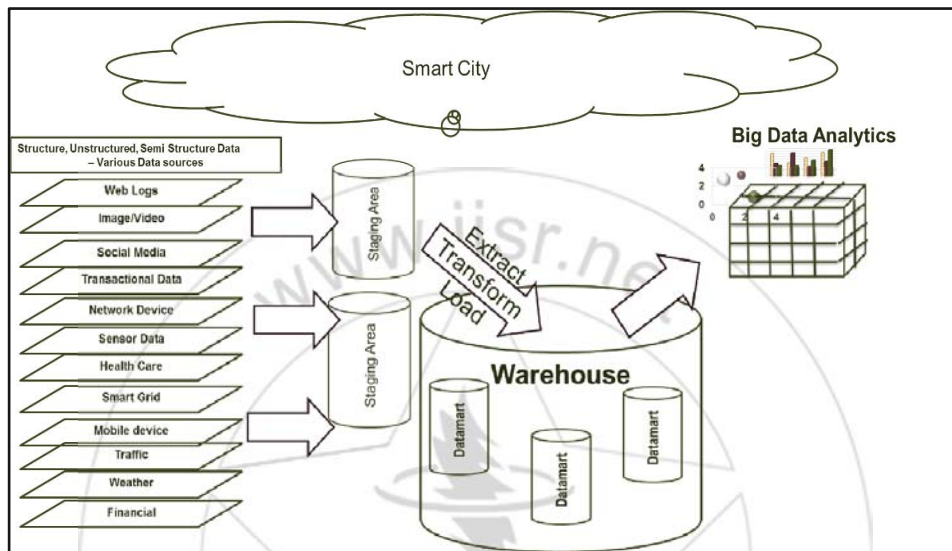

Figure 2: Data Source, Warehouse and Business Analytics – Pictorial View

3. Big Data Analytics

Big data analytics is the process of probing big data set to reveal hidden patterns, unknown correlations and other important information that can be used to make decisions. Big data analytics uses advance techniques likepredictive modeling, text analytics, machine learning, forecasting and statistical analysis. It will helps to identify trends, weak spots or determine conditions for making better and faster decisions about the future which can be very significant to the concerned area.

By using BigData analytics one can extract only the relevant information from terabytesto exabytes, and analyze it to transform business decisions for the future. It helps in giving better insight and knowledge about the datasets.

4. Use of Big Data Analytics

4.1 Smart Grid

Grid refer to the network that hold electricity from the plant where it is generated to consumption. It includes substation, switches, wires, transformer etc.

LikeSmart" Phone meansa phone with a computer in it, Smart Grid means ―computerizing" the electric utility grid. A Smart Grid is an electricity network based on digital technology that is use to supply electricity to consumers via two-way digital communication. This system allows for monitoring, analysis, control and communication within the supply chain to help improve efficiency, reduce energy consumption, reducing cost, and maximize the transparency and reliability of the energy supply chain. The Smart Gridis introduced with the aim of overcoming the weaknesses of conventional electrical grids by using smart meters where each device on the network can have sensors to gather data - power meters, voltage sensors, fault detectors, etc.
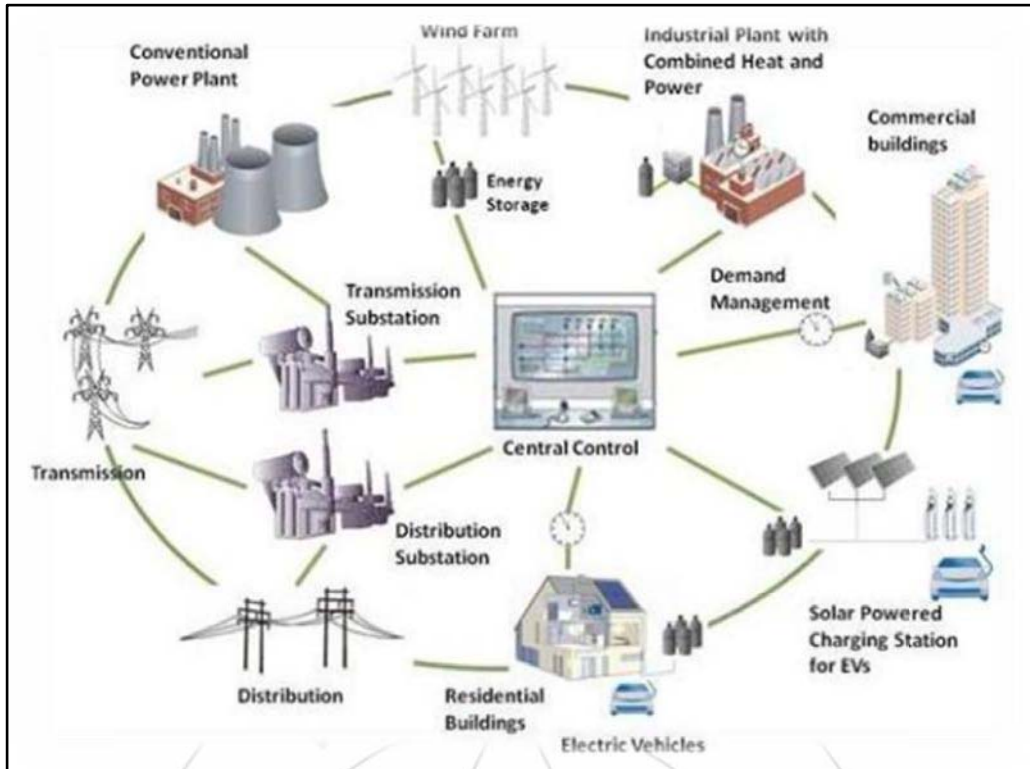
Smart Grid view:

Paper ID: NOV161007
14

Figure 3: Source: http://www.cepl.gatech.edu/

It is estimated that by 2020
- number of installed smart meters in Europe will reach 240 million
- North America will have 150 million smart meters in use.
- China is forecasted to install about 400 million smart meters.
- Japan would deploy about 60 million smart electricity meters
- South Korea would plan to deploy between 500,000 and 1.5 million smart meters per year in homes.

Smart energy is one of the eight pillar of India's Smart City program. Three crucial dimensions of smart energy systems are:
1. Smart Grid
   - Smart Grid electrification of all households with power available for at least 8 hours per day by 2017
   - Establish Smart Grid test bed by 2014 and Smart Grid knowledge center by 2015
   - Implementation of 8 Smart Grid pilot projects in India with an investment of US$10 million
2. Energy Storage:
   - Addition of 88,000 MW of power generation capacity in the twelfth five year plan (2012-17)
   - India needs to add at least 250-400 GW of new power generation capacity by 2030
   - The Power Grid Corporation of India has planned to invest US$26 billion in the next five years
3. Smart Meters
   - India to install 130 million smart meters by 2021

Table 1, shows the electricity Generated (from Utilities), Distributed, Sold and Lost in India:

						(in Giga Watt hour) = (10^6 x Kilo Watt hour)		
Year	Gross Electricity Generated from Utilities	Consum-ption in Power Station Auxiliaries	Net Electricity Generated from Utilities	Purchases from Non-Utilities + Imported from Other Countries	Net Electricity Available for Supply	Sold to Ultimate Consumers & Other Countries	Loss in transm-ission	Loss in transm-ission (%)
1	2	3	4=2-3	5	6=4+5	7	8=6-7	9
2005-06	623,819	41,970	581,849	10,345	592,194	411,887	180,145	30.42
2006-07	670,654	43,577	627,077	11,931	639,008	455,749	183,012	28.64
2007-08	722,626	45,531	677,095	12,685	689,780	501,977	187,620	27.20
2008-09	746,626	47,573	699,053	14,181	713,234	553,995	178,420	25.02
2009-10	796,281	49,706	746,576	14,391	760,967	612,645	193,455	25.42
2010-11	844,846	52,952	791,894	19,839	811,733	694,392	194,537	23.97
2011-12	922,451	56,499	865,952	15,516	811,506	685,194	208,400	25.68
2012-13	963,722	59,799	903,923	20,577	924,500	824,301	226,395	24.49
2013-14(p)	1,022,614	62,250	960,364	20,577	980,941	882,592	226,009	23.04

Paper ID: NOV161007

15

Table 1, Source– Energy Statistics 2015, www. mospi.gov.in, (p) – Provisional figure
In year 2013-14, the provisional figure of transmission loss is 23.04%.

Table 2shows Aggregate Transmission and Commercial Loss (AT & C) projection drawn in consultation with states up to year 2021-22.

State	Source: PFC 2012-13	Year wise projection with 2012-13 as base								
		2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22
Andhra Pradesh	13.63	13.38	13.13	12.88	12.58	12.28	12.00	12.00	12.00	12.00
Arunachal Pradesh	60.26	56.76	53.26	49.76	46.26	42.76	39.26	36.00	33.00	30.00
Assam	31.85	29.85	28.35	26.85	25.35	23.85	22.00	20.00	18.50	17.00
Bihar	54.63	50.63	46.63	42.63	38.13	34.00	30.00	27.00	24.00	21.00
Chhattisgarh	25.12	24.48	22.37	20.66	19.07	17.49	16.17	15.11	14.52	14.00
Delhi	15.22	14.72	14.22	13.97	13.72	13.22	13.00	12.50	12.25	12.00
Goa	14.14	13.89	13.64	13.39	13.14	12.64	12.00	12.00	12.00	12.00
Gujarat	19.87	18.87	17.87	16.87	15.87	15.37	15.00	14.50	14.25	14.00
Haryana	32.55	30.05	27.55	24.55	21.55	19.55	18.00	17.00	16.00	15.00
Himachal Pr.	9.53	14.50	13.50	12.50	11.50	10.50	10.00	10.00	10.00	10.00
J & K	60.87	55.87	51.87	47.87	43.87	39.87	35.00	30.00	26.00	22.00
Jharkhand	47.49	43.49	39.49	36.49	32.49	29.49	27.00	24.00	21.00	18.00
Karnataka	20.78	18.25	17.90	17.43	16.86	16.29	15.72	15.15	14.57	14.00
Kerala	10.53	11.15	10.80	10.50	10.25	10.00	10.00	10.00	10.00	10.00
Madhya Pradesh	31.15	27.65	25.15	23.15	21.15	19.15	17.00	16.00	15.00	15.00
Maharashtra	21.95	20.45	18.95	17.45	16.45	15.45	15.00	14.50	14.25	14.00
Manipur	85.49	78.49	71.49	64.49	56.49	48.00	40.00	34.00	28.00	22.00
Meghalaya	26.60	33.11	31.29	29.79	28.29	26.79	25.29	23.79	22.29	20.79
Mizoram	27.55	27.02	26.14	25.77	24.59	23.49	22.13	21.13	19.75	18.62
Nagaland	75.30	67.21	64.21	59.21	53.21	47.21	41.21	35.21	29.21	24.21
Odisha	42.94	38.94	37.00	35.00	33.00	30.50	28.00	25.50	23.00	20.50
Puducherry	9.13	19.00	18.00	17.00	16.00	15.00	14.00	13.00	12.00	11.00
Punjab	17.66	17.16	16.66	16.16	15.66	15.16	15.00	14.50	14.25	14.00
Rajasthan	20.00	19.00	18.00	17.00	16.00	15.50	15.00	15.00	15.00	15.00
Sikkim	53.51	49.51	45.51	41.51	37.51	33.51	29.00	26.00	23.00	20.00
Tamil Nadu	20.72	20.22	19.72	19.22	18.97	18.72	18.47	18.22	18.00	18.00
Telangana			13.13	12.88	12.58	12.28	12.00	12.00	12.00	12.00
Tripura	33.85	29.85	27.35	24.85	22.35	20.85	20.00	18.00	17.00	16.00
Uttar Pradesh	42.85	38.85	34.85	31.85	28.85	25.85	22.85	20.00	17.00	15.00
Uttarakhand	23.18	21.68	20.18	18.68	17.68	16.68	16.00	15.00	14.50	14.00
West Bengal	34.43	30.51	29.00	28.00	26.00	24.00	23.00	22.00	21.50	21.00
	25.38	23.70	22.17	20.85	19.52	18.31	17.27	16.36	15.57	14.95

Table 2, Source – Government of India, Ministry of Power

Indian power system has faced high AT&C Losses around 25% in year 2012-13, poor distribution network, wide demand – supply gap of energy, poor asset management etc. In some of the sates the losses are more than 50%.

The power supply position during 2009-10 to 2015-16 in the country is mentioned table below:

Year	Energy				Peak			
	Requirement	Availability	Surplus(+)/Deficits(-)		Peak Demand	Peak Met	Surplus(+) / Deficits(-)	
	(MU)	(MU)	(MU)	(%)	(MW)	(MW)	(MW)	(%)
2009-10	8,30,594	7,46,644	-83,950	-10.1	1,19,166	1,04,009	-15,157	-12.7
2010-11	8,61,591	7,88,355	-73,236	-8.5	1,22,287	1,10,256	-12,031	-9.8
2011-12	9,37,199	8,57,886	-79,313	-8.5	1,30,006	1,16,191	-13,815	-10.6
2012-13	9,95,557	9,08,652	-86,905	-8.7	1,35,453	1,23,294	-12,159	-9.0
2013-14	10,02,257	9,59,829	-42,428	-4.2	1,35,918	1,29,815	-6,103	-4.5
2014-15	10,68,923	10,30,785	-38,138	-3.6	1,48,166	1,41,160	-7,006	-4.7
2015-16*	8,37,958	8,19,225	-18,733	-2.2	1,53,366	1,48,463	-4,903	-3.2

Table 3, *Provisional up to Dec 2015, source http://powermin.nic.in/power-sector-glance-all-india

In year 2014-15, the deficit in energy demand was 3.6% and during peak time it was 4.7%.

Around 300 million people in India do not have access to electricity. High AT&C losses leads to underinvestment in distribution sector and unreliable demand. Due to traditional outage management, restoration time is very long. Minimization in AT&C loses can help in removing deficit.

Following smart grid project are initiated in India to see the potential of Smart grid which can be replicated to other cities and states:

Smart Grid Pilot Projects	Benefits/ Objectives
• Panipat City Subdivision (Haryana State) • Electrical Division No.1, of Agartala town • Electricity Department, Government of Puducherry • AP CPDCL, Andhra Pradesh - Jeedimetla Industrial Area • PSPCL, Punjab - Tech-II Sub-division, SAS Nagar • WBSEDCL, West Bengal - Siliguri Town in Darjeeling District	• Reduced AT&C Losses • Reduced Peak Load Consumption • Reduced Cost of Billing
Additional City Area Division (ACAD), Mysore	• Reduction in AT&C losses • Shifting of load in industrial and domestic consumer during peak hours • Reduction in number of transformer failure • Reduction in Meter Reading cost • Reduction in unforeseen outages and also recovery time for unforeseen outages
KSEB, Kerala - Selected Distribution Section offices spread over the geographical area of Kerala State	• Reduction in AT&C losses through reduction in loss due to manual error, tampers, thefts, short assessment etc., • Savings on employee and travel cost for meter reading • Introducing incremental tariff for peak hours through Tariff on Demand
UGVCL, Gujarat Project proposes covering consumers in Naroda and agricultural unmetered consumers in Deesa-II	• Reduction in AT&C losses • Savings in Peak Power Purchase cost by reduction of peak load • Reduction in Transformer failure rate • Reduction in number of outages • Reduction in Meter Reading cost, Cost of payment collection etc.
APDCL, Assam - Guwahati distribution region	• Increased available energy during peak time • Revenue increase through Power Quality measurements and power factor penalty • Reduction in AT&C Losses • Reduction in interest payments due to deferred Capital Investment in sub-transmission networks • Improvement of availability (reduction of Customer Minutes Lost) • Improved management of power procurement options • Unscheduled Interchange using Short Term Load Forecasts
CSPDCL, Chhattisgarh - Siltara – Urla area of Raipur District (Chhattisgarh State)	• Reducing Distribution AT&C losses • Reducing Peak load consumption through shifting of Peak Load demand to a non-peak time thereby saving UI charges • Reducing cost of billing
Himachal Pradesh Location Industrial town of KalaAmb	• Shifting peak load • Reduction in penalties • Reduction in outages
JVVNL, Rajasthan - Sanganer Sub Division, Jaipur City	• Reduced AT&C Losses • Reduced Peak Load Consumption • Reduced Line Outages and DT Failures • Increased energy sales due to reduced failures/increased availability
Power Grid Corporation of India with Govt. of Puducherry is developing Smart grid pilot	• Indigenization of Technology • Common Information sharing platform • Scalable and replicable at other places • Shall help in demonstration of effectiveness of each functionality • Evolving policy advocacy, Regulations, etc. for successful implementation • Evolving commercial mechanism
Smart Meter by Tata Power Ltd – Delhi Distribution - Project covers over 100sqkm of industrial and commercial belt in North and North West Delhi The components include AMI, ADR infrastructure, smart meters, radio frequency mesh-based communication, meter data management system and integration with other operational technology and IT systems like outage management system (OMS), SAP, etc.	• Analysing of consumption patterns closely and optimize equipment energy consumption. • Through customer portal, consumers have options to set threshold values for key parameters, like load, power factor and receive alerts once these thresholds are breached through SMS or email. • To improve the reliability of power and manage peak power demand more optimally • To Permit Tata Power Delhi Distribution to effectively manage grid emergencies

Paper ID: NOV161007
17

Smart Grid Pilot Projects	Benefits/ Objectives
Innovari and reliance Infra Project – Automated demand side management expansion project 9 million target in future, currently raghuleela mall, ITC, Blue Dart and Marriott- Consumers	• Real time visibility • Service Reliability • Control of cost and electricity usage • Shift in peak load • Capacity requirement of the grid

Source:http://www.indiasmartgrid.org/   http://www.tata.com/
www.powergridindia.com/

Selected Smart Grid projects acrossthe globe

Projects	Benefits using Analytics
Ontario (Canada) Smart Metering deployment project 4.8 million smart meters , 4.5 million customers Business case – return of $1.6 billion	• Demand management • Accurate meter reading • Timely consumption information • Proactive consumer services • Reduction in number of crew visits to read and service meters • Reduces tempering and theft of electricity • Operational benefits
Denmark, Power matching city 240K Customers	• Decentralisation distribution • Peak load management • Better service and efficiency • Customer can decide when they can buy electricity • Smart App: control indoor climate and save energy • Cost saving
Reken(Germany), Uppsala(Sweden), Castellon(Spain), Forli-Cesena(Italy), Virchlabi(Czech Republic) and Carros(France) GRID4EU - Large scale distribution of European smart Distribution Networks	• Efficient electricity supply • Cost Reduction • Reduction in load shedding • Voltage control • Short recovery time after grid failure • Remote control level of the grid • Higher reliability • Increasedsurveillance
France – Smart electric Lyon 2500 Customers	• Consumer Analysis : Tariffs and Segmentations • Sectorial Organisation Analysis • Real time information about consumptions • Getting personalised advice • Identification of upstream and downstream services and their behaviours • Energy and environment analysis – amount of carbon dioxide emissions • Information on technology price. Production, network, quality, continuity on demand
Korea Electric Power Corp – Use of Big Data	• In 2014, state-run electric utility Korea Electric Power Corp (KEPCO) launched two pilot projects on ways to use big data to improve demand management and risk forecasting. • The first pilot set up an energy consulting business based on AMI data while the second established a risk forecasting system analyzing social networking service data. AMI provides extra metering points for spatial awareness of planning requirements
Pacific Northwest Smart Grid – This project will be a unique demonstration of unprecedented geographic breadth across five Pacific Northwest states-Idaho, Montana, Oregon, Washington, and Wyoming. It will involve about 60,000 metered customers, and contain many key functions of the future Smart Grid, ultimately moving the nation closer to establishing a more efficient and effective electric grid	• validate new Smart Grid technologies and business models • provide two-way communication between distributed generation, storage, and demand assets and the existing grid infrastructure • quantify Smart Grid costs and benefits • advance standards for interoperability (the smooth, seamless integration of all elements of the electric system) and cyber security approaches

Source: //www.indiasmartgrid.org:Smart grid Project a global Snapshot

Smart Grid enables real time monitoring and control of power system as well as helps in reduction of AT&C losses, demand response and demand side management, power quality management, outage management, smart home energy system etc. It will be one of the backbone infrastructure to enable new business models inSmart City, electric vehicles, and smart communities apart from more resilient and efficient energy system and tariff structures.

Smart Grids will generate unprecedented amounts of data — from energy production all the way to consumption — and connects with various devices and systems, empower all involved to enhance their energy efficiency.

The data is generated from different intelligent components such as smart meters, automated distribution systems, and sensing & measurement devices; this is then transmitted to the utility station for further predictive analysis.
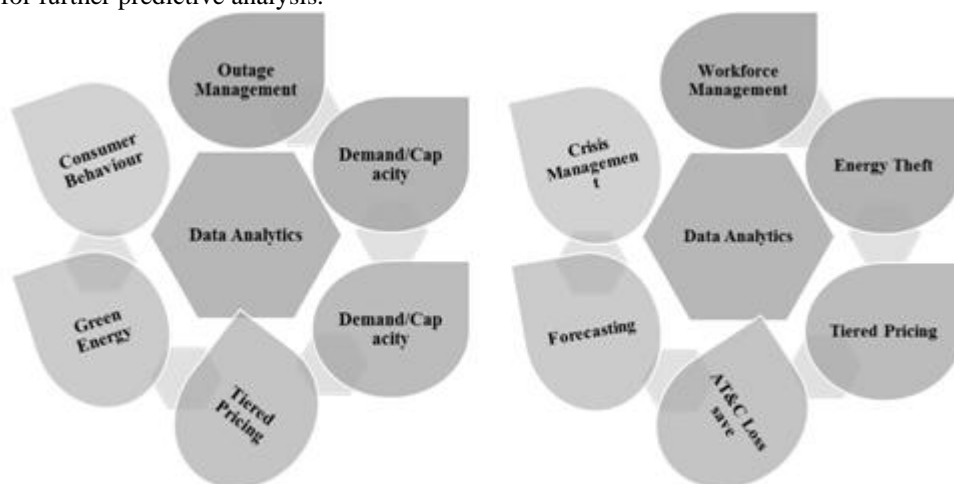
Big Data Analytics –Smart Grid

Figure 4: Smart Grid – Data Analytics

Based on some of the pilot projects running and implemented, Big Data Analytics whose main source of data is from IoT (Machine to Machine) can help planners, decision makers and designer of Smart City in following ways:

- **Outage management** - Rather than waiting for customers to report outages, companies can receive reports from the grid itself.The Smart Grid can provide the report on outage, document its recovery in real-time and isolate locations of physical damage.
- **Demand and Capacity Management**- Data gather from both producer and consumer side can help in analyzing demand and capacity. It will help in understanding the peak/non hours of usage.
- **Better understanding on AT & C Losses** –Data collected from each sensor on regular interval from production to consumption will give insight on where the actual loss is happening. Companies can look for how these losses can be reduced based on deep data analysis.
- **Forecasting based on seasonality** – Utilities data collected on regular interval during the years can help analyst in understanding the pattern of usage of the energy. Seasonality pattern can help in looking for when and in which region the demand will be high or low. Companies can act in advance to tackle these situations.
- **Asset Management**–With Big Data Analytics companies can predict how millions of distribution grid assets are about to fail and helps in determining how costly or dangerous those breakdowns are going to be. They can decide which of a thousand different options for asset maintenance, replacement, and procurement strategies is the best. Asset health analytics can help in justify all manner of grid-edge investments -- real-time distribution automation and demand response, strategic deployment of energy storage or grid resiliency projects, and can also help in managing future challenges like the increased threat of storms and natural disasters to grid resiliency.
- **Crisis Management -** Big Data can help in making visual analytics systems which can be place in Smart Grid to help decision makers to perceive and analyze the crisis situations.

- **Mobile Workforce Management**– Decision makers canbetter mobilize their workforce by doing real time monitoring of the systems in place on the Smart Grid. Managers can get accurate information on Fault location, real time load. People can be mobilize accordingly to fix the issue.
- **Energy Theft** - Use real-time metering data can help in discovering unaccounted consumption when energy is being diverted and stolen
- **Behavioral Analytics of consumers -** With so much data available on the consumer utilization of energy at various time interval duration – morning to night, different seasons, across different region, different utility pricing plan - Big Data analytics can help in understanding the Consumer Behavior.
- **Tiered pricing**– Companies can opt for tiered pricing based on the demand and capacity of the energy. Based on large dataset available, Big Data Analytics can help in defining new utility retails rate policy.
- **360 Degree View -** City Planner and organization can get the 360 view of full energy supply chain which enables then to do predictive analysis on various aspect. It can help in improving end user experience, customer loyalty and Demand strategy.
- **Green Energy -** Through predictive analytics, producer can generate electricity from renewable sources with small plant set up and in incremental way depending upon the needs of a city. Large dataset can help in doing analysis on sources of carbon emission which can be reduced.

Traffic Congestion Management

The urbanisation and metropolitan cities trend is fast growing. About 70% of the global population expected to be living in cities by 2050, the execution of smart city technologies will be a focal point for governments in both developed and developing regions. The number of vehicles on the world's roads is expected to double to around 2.5 billion by 2050.

According to national highway authority of India, Indian road carries almost 67% of the freight and 80% of the passengers annually. National highways make up only 2% of the overall road network by length, they are accounted for 40% of the total road traffic. Number of vehicles grows in 2014-15 is 8.68%.

Category	2009-10	2010-11	2011-12	2012-13	2013-14	2014-15
Passenger	23,57,411	29,82,772	31,46,069	32,31,058	30,87,973	32,20,172
Commercial	5,67,556	7,60,735	9,29,136	8,32,649	6,99,035	7,083
Three Wheelers	6,19,194	7,99,553	8,79,289	8,39,748	8,30,108	9,49,021
Two Wheelers	1,05,12,903	1,33,49,349	1,54,27,532	1,57,44,156	1,68,83,049	1,84,99,970
Grand Total	**1,40,57,064**	**1,78,92,409**	**2,03,82,026**	**2,06,47,611**	**2,15,00,165**	**2,33,66,246**
Percentage Increase		27.28%	13.97%	1.25%	4.13%	8.68%

Table 4, Automobile Production Trends - Source http://www.siamindia.com

There is no clear statistic as to how many auto-rickshaws (a three-wheeled vehicle for hire) and two-wheelers (scooters and motorcycles) operate in India's densest urban areas. Some cities have six-passenger motorized vehicles called fat-fat (which describes their exhaust sound). As for two-wheelers, estimates of motorcycles, scooters and bicycles are up to 50 million in India.

The accident rate among cars in India is the highest in the world. India has about 1% of the world's cars (some 4.5 million) manages to kill over 100,000 people in traffic accidents each year. This amounts to 10% of the entire world's traffic fatalities. The U.S., with more than 40% of the world's cars, creates just 43,000 fatalities.

Study suggests China and India will contribute more than one third of the global urban population increase between 2014 and 2050.Between 2014 and 2050, the urban areas are expected to grow by 404million people in India. Rapid urbanisation is concern in India as it is causing traffic congestion, air pollution due to heavy motorisation. Smart cities can help in tackling the rapid urbanisation issue if systematic planning is done along with the use of effective technology and infrastructure.

Smarter cities can new technologies and insights to transform their systems, operations and services. One of the key elements of the plan of Smart Cities is the ability to implement Intelligent Transport Systems (ITS) to deliver city wide mobility services.

Selected Congestion Management Pilot Programsacross the globe

Cities where Big Data Analytics is helping in Traffic and Congestion management are:

Project	Benefits using Analytics
The Dutch city of Eindhoven (Holland), for instance, worked with IBM to pilot a traffic management solution that collects and merges braking, acceleration and location data from in-vehicle sensors with traffic data gathered from the road.	Using analytics, the disparate data from thousands of sensors enabled officials to respond to dangerous road conditions, accidents or growing traffic density in near real-time. The solution also alerts drivers of traffic incidents through smartphones and built-in navigation devices, allowing them to find alternate routes
The city of Toulouse, France deployed a solution that collects and analyzes its citizens' social media comments about city planning and traffic issues so authorities can be more responsive to pressing needs	Toulouse has accelerated its average response time to road-maintenance issues by 93 percent, from 15 days to just one day.
AGT International along with Singapore government, initiated Safe SmartCity test bed pilot project as part of a consortium with O'Connor, and Hitachi Asia, a regional subsidiary of Hitachi, Ltd. Pilot have integrated multi-agency solution and named CityMIND. The CityMIND platform uses advanced analytics in addition to web intelligence, powerful sensors and big data to connect various interfaces and outputs, from video and facial recognition cameras to smart phones, in order to create a comprehensive and reliable view of the city.	Using Analytics from CityMIND, a number of key smart cities technologies were implemented successfully in Singapore like • identify traffic congestion and crowding, • improve cleanliness by using sensors in trash bins to automatically signal when they need to be emptied, • increase communications between first responders, authorities and citizens with specially-designed mobile device applications and raise awareness of relevant social media trends
Zhejiang China – The City has 1000 digital monitoring checkpoint systems which capture images and video data continuously. All the structured data - time, place, and vehicle information and semi structured data - pictures and videos are stored in a centralized data center	The traffic management division uses this data to do real-time analysis of traffic conditions and statistical analysis of traffic accidents as well as to analyses any traffic violations and driver records. City is witnessing following benefits: • With traffic violation image information stored in the system, traffic police departments easily retrieve vehicle information such as the color, model, and license plate in real time along with other relevant information such as historical behavior, driving routes, the vehicle's operating company, and the identity of the driver. • Traffic police can now easily retrieve plate numbers and the driving track of a passing vehicle from the over 2.4 billion records in the system.

	• Investigating traffic cases that require complex inquiries, such as data from multiple checkpoints or multiple vehicles, now takes just 10 seconds
Los Angeles, the Merge Platform tool – combine data from 6000 on street parking space, parks and other data source to provide integrated data for analysis	• The analytics are used to observe parking maintenance, revenue and occupancy and by applying a dynamic pricing algorithm, a more well-adjusted use of parking is achieved. • This has resulted in a 10% reduction in parking related congestion caused by drivers circling looking for a free space

It is estimated that reducing congestion by 25 percent equals $185 billion in savings in the United States and can increase retail sales by six percent.

A congestion management plan must take into consideration all relevant factors like the ever-exploding vehicle population on the road, geometry of the city roads, travel needs of citizens, and the needs of various administering authorities having jurisdiction in parts or whole of the city.

In Smart Cities, data is the fuel that drives intelligent transportation systems and it will be gather from everywhere – Sensors on signals, GPS trackers, social media posts, mobile phones and cameras. Cities can use predictive analytics from the data gathered to understand how congestion can be reduce. Drivers use social media to detect and avoid gridlock. City planners analyze data to pinpoint where new bus routes are most needed.

Overall the Big Data in the transportation sector, is still in its early stages to play a significant role in improving its public services and the quality of people's lives.

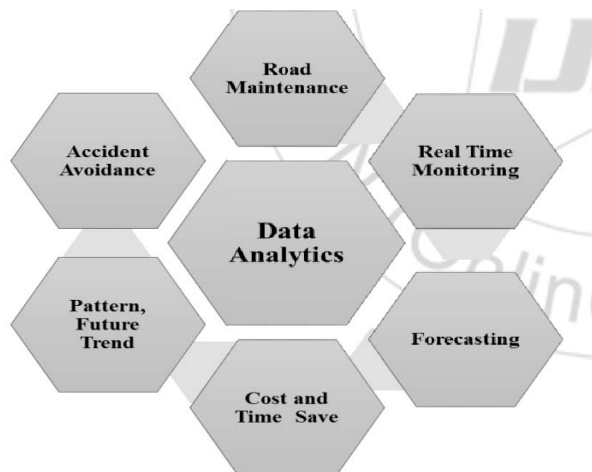**Big Data Analytics - Role in Congestion Management**

Figure 5: Congestion Management - Data Analytics

- **Monitoring traffic performance** - By monitoring traffic performance and patterns over time, cities can make significant progress in cutting congestion, emissions and noise; determine where to place buses and build mass transit stations to ensure they operate at full capacity; and improve emergency vehicle response times.
- **Reduce accidents** – With Big Data analytics – Traffic department have beforehand knowledge on the traffic situation at particular location which they can share with commuters and advice them to take the detour or avoid congested area which can help in reducing accidents occur due to congestion.

- **Fuel cost and time** - Big Data Analytics can help in analyzing the fuel cost incurred and precious time people lost due to congestion.
- **Future trends -**Based on historical and current real-time traffic data, Big Data Analytics can help in understanding the future trends. Big data analytics can help in understanding the seasonality of the traffic patterns, motion of traffic during the day and night, location.
- **Predict speed of the traffic and volume** - Data aggregates from multiple devices - cameras, detectors, bluetooth, mobile and social media can help in identify and measure traffic speed and volume on city roads using predictive analytics.
- **Pattern of Traffic and behaviour** - Big Data Analytics can help in analysis of the historical data to gain performance insights and understanding of patterns of behaviour of traffic and road incidents.
- **Road Maintenance**: Through sensor on the roads, video camera, Analyst can analyse which road demands repairing.

1. Challenges in Implementing Big data
Smart cities are considered as dynamic and evolving environments. Following are the challenges organisation, planner may face while designing and implementing the Smart City Solutions Using Big Data and Analytics.
- **Data sources and size**
  In Smart City Data will be generated from multiple sources and in multiple format – Structured, Unstructured, and Semi Structured. Storing and processing such data is not possible using traditionalsoftware. Also future anticipation about the data formats and sources needs to be considered and factoredinto the solution while designing the smart city applications. Due to evolving environment it is difficult to predict the growth of data – 3Vs.
- **Data sharing and information**
  The smart city vision holds out the promise of integrating data from multiple organizations, diverse environments, and a wide variety of intelligent devices. Data integration even within organizations is one of the hardest challenges in the IT world, Data integration from various department would be challenge. There are challenges in sharing of information and data across various departments. Every department have its own data warehouse which works in silos, sharing information is difficult. Also there are legal mandatory privacy and security requirement like DPA (Data Protection Act) to ensure citizen rights of privacy. Smart City solution should be design keeping in mind how to prevent the privacy of the citizen.
- **Quality of Data**
  Since data in Smart city will be from various heterogeneous system and in different formats, maintaining data quality could be challenge. Transactional

Paper ID: NOV161007

21

system may require data in one format but when it comes to Big Data warehouse where data from different sources are gathered may require specific format without compromising the meaning and quality. If data is coming from third party, it needs to be ensured that they have proper mechanism in place to fetch the quality data.

- **Security of Data**
  Big Data breaches will be big with potential of more serious damages to reputation and legal repercussion. Movement of data across various sources in a secure manner is a biggest challenge in Smart city application implementation. Security here means right information to a right person at right time and at right place. While designing smart city, it is important to look at how data can be masked and secured so that it cannot be reached to unwanted persons.

- **Technology Advancement**
  In Today's environment, keeping the pace with technology is very difficult and costly effort. The smart city solution should be design in a manner that with advancement of the technology, upgrades to the existing solution can be achieved easily or with less effort.

- **Skill Gaps**
  A lack of data skills may be the barrier to the effective use of big data for city management. Managing and analysing large data sets and developing insights for effective policy making or operational improvement requires skills that are in short supply, particularly in the public sector.

- **Lack of Standards**
  With respect to Smart City Development, Standards will play important role in adhering to common platform. Currently there are lack of standards – Data integration, Policies, Procedures, formats which needs to be addressed.

5. Conclusion

The role Big Data is fundamental in building smart cities. Reliable technology and infrastructure which can tackle machine to machine, machine to human, human to equipment is required form managing public amenities and services in such cities. Big Data Analytics will help in analyzing and predicting information generated by smart devices connected through various channel in Smart Cities. Since data will come from various heterogeneous sources, this is challenging, as big data analytics and applications are not developed enough for real time processing of large data sets. Instead of targeting many sources, emphasis must be on a number of imperative sources of information, which are more crucial to public life and human wellbeing.Amenities that can be induced by IoT and Big Data are numerous in Smart Cities. However, the technology must be utilized with proper insight. Authorities must be open to implement innovative ideas and judge them with priority to public wellbeing. Only an accurate juxtaposition of physical and technological infrastructure can divulge a successful creation of such habitats. Technology infrastructure in smart cities is all about IoT, Thus, leveraging and managing Big Data is crucial to transition of cities to smart cities.

References

[1] Smart City Mission Statement and Guidelines http://smartcities.gov.in/writereaddata/SmartCityGuidelines.pdf
[2] "Big Data, smart cities and city planning" www.complexcity.info/files/2013/12/BATTY-DHG-2013.pdf
[3] Cities Get Smart with Big Data: by Ellis Booker http://data-informed.com/cities-get-smart-big-data/
[4] The Uses of Big Data in Cities: By Luis M. A. Bettencourt
[5] ―The Urban Internet of Things: By Laura Adler" http://datasmart.ash.harvard.edu/news/article/the-urban-internet-of-things-727
[6] Big Data: All roads lead to smart cities By Julián García Barbosa, Sales Executive at Telefónica http://en.blogthinkbig.com/2014/05/21/big-data-smart-cities/
[7] How Big Data and the Internet of Things Create Smart Cities by Bernard Marr http://www.datasciencecentral.com/profiles/blogs/how-big-data-and-the-internet-of-things-create-smart-cities
[8] Smart Cities Big Data - deloitte_smart-cities-big-data_en_0115
[9] http://cmrindia.com/india-energy-vision-2015-the-future-lies-in-smart-grids/
[10] The real-time city? Big Data and smart urbanism: By Rob Kitchin
[11] Smart Cities - Ranking of European medium-sized cities www.smart-cities.eu
[12] Big Data Analytics ―Itel IT Center Peer Research | Big Data Analytics"
[13] http://smartcitieschallenge.in/what-is-a-smart-city
[14] https://en.wikipedia.org/wiki/Smart_city
[15] http://www.informationweek.com/big-data/software-platforms/verizon-iot-strategy-smart-cities-thingspace-for-developers/d/d-id/1322893
[16] http://www.businesstoday.org/articles/2015/03/a-new-way-of-living-smart-cities-and-the-internet-of-things-2/
[17] Soft_Grid_2013_2020_Big_Data_Utility_Analytics_Smart_Grid.pdf
[18] http://www.pnwsmartgrid.org/about.asp
[19] http://smartgrid.ieee.org/newsletters/october-2015/big-data-analytics-for-smart-grid
[20] http://www.powergridindia.com/
[21] http://www.greenbiz.com/blog/2014/08/08/big-data-transform-smart-grids-worldwide
[22] http://energy.gov/oe/services/technology-development/smart-grid
[23] https://www.powergridindia.com/_layouts/PowerGrid/User/ContentPage.aspx?PId=154&LangID=English
[24] http://www-935.ibm.com/services/multimedia/Managing_big_data_for_smart_grids_and_smart_meters.pdf
[25] https://itu4u.wordpress.com/2015/10/13/how-big-data-will-help-fight-global-epidemics/
[26] https://www.datainnovation.org/2015/07/big-data-means-big-relief-for-disaster-victims/
[27] http://www.mckinsey.com/insights/infrastructure/big_data_versus_big_congestion_using_information_to_improve_transport

Paper ID: NOV161007

22

[28] http://www.intel.in/content/dam/www/public/us/en/documents/case-studies/big-data-xeon-e5-trustway-case-study.pdf

[29] http://www.ibm.com/smarterplanet/us/en/traffic_congestion/article/traffic-management-and-prediction.html

[30] http://esa.un.org/unpd/wup/Highlights/WUP2014-Highlights.pdf

Author Profile

Sanjeev Kumar is MBA (Executive), Faculty of Management Studies, University of Delhi. Sanjeev is currently working as Senior Program Manager in HCL Technologies. He has rich experience in managing complex program and multiple projects in the field of Software Industry. His main interest of research area are Big Data, IoT, Analytics, and Lean.

Anand Prakash is MBA (Executive), Faculty of Management Studies, University of Delhi. Anand is currently working in Tech Mahindra as Senior Test Lead. He has extensive experience in managing testing projects across various domains. His main research area of interest are Big Data, Business Intelligence Analytics.