Available online at www.sciencedirect.com

ScienceDirect

Procedia Technology 21 (2015) 49 – 53

Procedia Technology

SMART GRID Technologies, August 6-8, 2015

Stream Computing: Opportunities and Challenges in Smart Grid

Shibily Joseph[a], Jasmin E A[b*], Soumya Chandran[c]

[a]*Research Scholar, Govt.Engineering College ,Thrissur- 680009,Kerala, India*
[b]*Associate Professor, Govt.Engineering College ,Thrissur- 680009,Kerala, India*
[c]*Asst. Professor, Govt.Engineering College ,Thrissur- 680009,Kerala, India*

Abstract

Traditional Power Grid is transformed to Smart Grid by utilizing the technological advancements of Information and Communication Technology. In Smart Grid data is flowing between different components of the system. Advanced and online analytic on this massive data is required to trigger instantaneous action for grid operation and management. Traditional Big Data handling techniques store and analyze high volume data with variety, but failed to handle high velocity data. Stream computing can analyze high velocity data with variety which is essentially required in online data analytics. This paper compared different Big Data analysis techniques and reveals the importance of stream computing and its opportunities in Smart Grid data analytics.

Peer-review under responsibility of Amrita School of Engineering, Amrita Vishwa Vidyapeetham University

Keywords: Smart Grid; Advanced metering infrastructure; Stream computing; Data analytics; Big data; Outage detection

1. Introduction

In our daily life, many dramatic changes are being witnessed, which are mainly influenced by the advancements of information and communication technology (ICT). Power grid is not an exception for this, which is also transformed to a new grid termed as Smart Grid. Thus Smart Grid is a modernized electrical grid that uses information and communications technology to gather and act on information, such as information about the behaviours of suppliers and consumers, in an automated fashion to improve the efficiency, reliability, economics, and sustainability of the production and distribution of electricity [1].

* Corresponding author. Tel.: +91-9495465409
E-mail address: eajasmin@gmail.com

Peer-review under responsibility of Amrita School of Engineering, Amrita Vishwa Vidyapeetham University
doi:10.1016/j.protcy.2015.10.008

The major components of smart grid are power stations, power transmission lines, pluggable electric vehicles (PEV) charging stations, residential subdivision installed with solar panels, residential complex with advanced metering infrastructure (AMI) and energy smart houses with electric appliances connected to the smart grid. One of the major characteristic of smart grid is the seamless interaction between these components using networks. All these components generate data with high velocity, variety and finally lead to huge volume of data .Advanced data analytics leverages this data to achieve the goals of smart grid like efficiency, reliability, fault tolerance etc. Stored processing of these data will not be a feasible solution for all situations. Fraud detection, system monitoring, outage detection are some typical examples in smart grid which require on line data analytics. Among the smart grid components, penetration of smart meter deployment is very less compared to its benefits. Direct operational benefits of advanced Metering Infrastructure(AMI) implementation are meter reading automation, operational efficiencies in field and meter services, reduction in unaccounted energy, operational efficiencies in billing and customer management, improvement in capital spend efficiency and improvement in outage management efficiency.

Data storage and analysis is challenging in smart grid due to the volume, velocity and variety of data generated by different components of smart grid. This paper will give an introduction to big data followed by overview of big data analytics, big data architecture and big data tools. Another important section is the introduction to stream computing, stream computing architecture and frame works. A table comparing the features of Data warehouses (DWH), Hadoop and stream computing is presented. Literature reviews on different research works in stream computing applied in smart grid and conclusion which specify the importance of stream computing to be applied in smart grid are also included.

2. Big Data

Big data is defined as a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a “massive scale” [2]. Big data is now characterized by four V’s which are volume, velocity, variety and value. Including all four V’s we can say that big data extract deep value from high velocity, volume and variety data using advanced analytics. In smart grid phasor measurement units (PMU) data, smart meter data, sensor data, whether data are some examples of big data[5].

2.1 Big Data Analytics and tools

Big data analytics is the use of advanced analytical techniques against very large, diverse data sets that include different types such as structured/unstructured and sizes varying from terra bytes to zeta bytes in different processing modes (streaming/ batch) [3]. Data analysis will be performed in batch wise or in real time. Batch processing analytics commonly based on Map Reduce system. Such systems envisage the volume and variety of data. Real time systems analyze the data on the fly and perform event pattern analysis or continuous queries. Such systems process continuous unbounded stream of data. This type of processing is called stream computing, especially required for certain application areas like trading, fraud detection, system monitoring, outage detection and management , demand response programs in smart grid.

A good big data architecture combines, Data warehouses (DWH), Hadoop and real time processing. Big data architecture comprises tools for storage, complex computing, immediate reaction, and monitoring of streaming data in real time [6]. Data warehouses (DWH) are used for storage. Apache Hadoop is an open source software library, is a frame work that allows for the distributed processing of large data sets across clusters of commodity hardware using simple programming models. It is designed to scale up from single servers to thousands of machine each offering local computation and storage. This follows a transaction based architecture, where events are stored on main frame or database, analyzed and action performed. At the same time stream computing tools monitor millions of events in a specific time window to react proactively, they are behavior based architecture where events are analyzed in real time and action performed and then stored in databases for further analytics.