Name of the student:	Tanmay Prashant Rane	**Roll No.**	8031
Assignment Number:	1	**Date of Assignment:**	
Relevant CO's: ITC801.1	**At the end of the course, students will be able to explain characteristics of and trends in big data.**		
Sign here to indicate that you have read all the relevant material provided before attempting this assignment		**Sign:**	

Assignment grading using Rubrics

Indicator	**Very Poor**	**Poor**	**Average**	**Good**	**Excellent**
Timeline (2)	More than a session late (0)	NA (0.5)	NA (1)	NA	Early or on time (2)
Organization of Content (2)	N/A	No sense of organization, Paragraphs lack clear ideas (0.5)	Some paragraphs have clear ideas, support from examples may be missing and transitions are weak (1)	Most paragraphs have clear ideas, are supported with some examples and have transitions. (1.5)	All paragraphs have clear ideas, are supported with examples and have smooth transitions. (2)
Level of content (4)	N/A	Major points omitted or addressed minimally(1)	Content is sound and solid; ideas are present but not particularly developed or supported; some evidence, but usually of a generalized nature.(2)	Well-presented and argued; ideas are detailed, developed and supported with evidence and details, mostly specific. (3)	Exceptionally well-presented and argued; ideas are detailed, well-developed, supported with specific evidence & facts, as well as examples and specific details. (4)
Grammar and Mechanics (2)	N/A	Spelling, punctuation and grammatical errors create distraction, making reading difficult(0.5)	Most spelling, punctuation and grammar correct. Some errors remain(1)	Few spelling, punctuation and grammatical errors allowing reader to follow ideas clearly (1.5)	Assignment is free of distracting spelling, punctuation and grammatical errors(2)

Late submission details (if any)

Reason(s) of late submission	Submission date	Actual submission date	sign of student

Assignment 1
Assignment on Introduction to big data

Course title: Big Data Analytics
Course term: 2019-2020
Instructor name: Saurabh Kulkarni

Q.1.Explain with a case study, a trend of big data in Telecom industry

Ans:

This is a case study about MAPR's guide for big data in telecommunication

The Motivation for Big Data:

With the rapid expansion of smart phones and other connected mobile devices, communications service providers (CSPs) need to rapidly process, store, and derive insights from the diverse volume of data travelling across their networks. Big data analytics can help CSPs improve profitability by optimizing network services/usage, enhancing customer experience, and improving security. According to McKinsey, the potential for Telcos to profit from applying data science effectively is substantial. Examples include:

- Predicting the periods of heaviest network usage, and targeting steps to relieve congestion
- Identifying the customers most likely to defect, and targeting steps to prevent churn
- Identifying the customers most likely to have problems paying bills, and targeting steps to improve the recovery of payments.

Data-Driven Improvement of Services or Product

Telecoms need to share data between cell towers, users and processing centers and due to the sheer volume of this data, it is important to process it near the source and then efficiently transfer it to various data centers for further use. MapR Event Store, a new distributed messaging system, is uniquely effective to transport huge amounts of data and to make this data available with reliable geo-distributed replication across multiple data centers. With MapR Event Store, they can replicate streams in a master-slave, many-to-one, or multi-master configuration between thousands of geographically distributed clusters.

Big Data Use Cases In Telecom

Telecommunication companies collect massive amounts of data from call detail records, mobile phone usage, network equipment, server logs, billing, and social networks, providing lots of information about their customers and network, but how can telecom companies use this data to improve their business?

Most telecom use cases fall into these main categories: **customer acquisition and retention, network services optimization, and security.**

BEFORE

Before, the customer was using FTP to transfer data from antennas to regional data centers and to the HQ Central Data center, but the FTP transfer meant extreme latency throughout the data pipeline.

AFTER

Now data is collected at regional data centers with MapR Event Store and made available in real time to regional dashboards.

MapR Event Store Topics at regional data centers are replicated in a many-to-one configuration to the HQ Central Data Center, making events available in real time to the HQ dashboard. This means they can now monitor global performance and react fast enough to improve customer services.

Being able to process high throughput geo-distributed events in real time enables:

Understanding how and where service issues are trending and how that is affecting customers.
Crowd-based antenna optimization: Monitor quickly changing network usage patterns, and reconfigure network support to handle short-term surges, such as heavy usage near a stadium during a sporting event.

Optimizing Services with Equipment Monitoring, Capacity Planning, and Preventative Maintenance:

- Dropped calls
- Lack of network coverage, resulting in poor customer experience
- Bandwidth issues
- Poor download times
- Inordinate service wait times
- Switching, frequency utilization, capacity use
- Analyzing these events in real time is the key to timely insights on network services in order to improve customer satisfaction.

Threat Detection

Solutionary, a subsidiary of NTT Group, is a leader in Managed Security Services. They provide Threat Intelligence, Incident Response, Compliance and Vulnerability Management as a service to their clients. Their platform collects and correlates vast amounts of data from logs, endpoints, firewalls, and network devices.

They needed to improve scalability as the data volume grew, but it was cost-prohibitive with their existing Oracle database solution. The old solution could not process the unstructured log data at scale and there were also major performance issues.

They replaced their RDBMS solution with the MapR Data Platform to achieve scalability while still meeting reliability requirements. Their new solution combines machine learning algorithms, complex event processing, and predictive analytics to detect real-time security threats.

All of the components of the use case architectures we just discussed can run on the same cluster with the MapR Data Platform, which provides advantages such as:

- Less complexity, fewer moving parts, fewer things to manage: converging multiple clusters for Streams/HBase/Spark/Hadoop into one cluster.
- "Joining" data sources into one core data mediation platform so that applications consume data in an easier way.
- Unified security.
- High reliability and high availability, replication from datacenter to datacenter
- Multi-tenancy: MapR Event Store is able to have essentially unlimited topics for lots of tenants.

The use cases we just went over showed how telecom companies can not only address these requirements, but also profit from the huge amount of information in their data to improve their business.

Q.2 Explain with a case study, a trend of big data in Health Care sector

Ans:

Average human lifespan is increasing along world population, which poses new challenges to today's treatment delivery methods. Health professionals, just like business entrepreneurs, are capable of collecting massive amounts of data and look for best strategies to use these numbers.

What Is Big Data In Healthcare?

The application of big data analytics in healthcare has a lot of positive and also life-saving outcomes. Big data refers to the vast quantities of information created by the digitization of everything, that gets consolidated and analyzed by specific technologies. Applied to healthcare, it will use specific health data of a population (or of a particular individual) and potentially help to prevent epidemics, cure disease, cut down costs, etc.

Now that we live longer, treatment models have changed and many of these changes are namely driven by data. Doctors want to understand as much as they can about a patient and as early in their life as possible, to pick up warning signs of serious illness as they arise – treating any disease at an early stage is far more simple and less expensive. With healthcare data analytics, prevention is better than cure and managing to draw a comprehensive picture of a patient will let insurances provide a tailored package. This is the industry's attempt to tackle the siloes problems a patient's data has: everywhere are collected bits and bites of it and archived in hospitals, clinics, surgeries, etc., with the impossibility to communicate properly.

Indeed, for years gathering huge amounts of data for medical use has been costly and time-consuming. With today's always-improving technologies, it becomes easier not only to collect such data but also to convert it into relevant critical insights, that can then be used to provide better care. This is the purpose of healthcare data analytics: using data-driven findings to predict and solve a problem before it is too late, but also assess methods and treatments faster, keep better track of inventory, involve patients more in their own health and empower them with the tools to do so.

Why We Need Big Data Analytics In Healthcare

There's a huge need for big data in healthcare as well, due to rising costs in nations like the United States. As a McKinsey report states, "After more than 20 years of steady increases, healthcare expenses now represent 17.6 percent of GDP —nearly $600 billion more than the expected benchmark for a nation of the United States's size and wealth."

In other words, costs are much higher than they should be, and they have been rising for the past 20 years. Clearly, we are in need of some smart, data-driven thinking in this area. And current incentives are changing as well: many insurance companies are switching from fee-for-service plans (which reward using expensive and sometimes unnecessary treatments and treating large amounts of patients quickly) to plans that prioritize patient outcomes

As the authors of the popular Freakonomics books have argued, financial incentives matter – and incentives that prioritize patients health over treating large amounts of patients are a good thing. Why does this matter?

Well, in the previous scheme, healthcare providers had no direct incentive to share patient information with one another, which had made it harder to utilize the power of analytics. Now that more of them are getting paid based on patient outcomes, they have a financial incentive to share data that can be used to improve the lives of patients while cutting costs for insurance companies.

Finally, physician decisions are becoming more and more evidence-based, meaning that they rely on large swathes of research and clinical data as opposed to solely their schooling and professional opinion. As in many other industries, data gathering and management is getting bigger, and professionals need help in the matter.

Obstacles To A Widespread Big Data Healthcare

One of the biggest hurdles standing in the way to use big data in medicine is how medical data is spread across many sources governed by different states, hospitals, and administrative departments. Integration of these data sources would require developing a new infrastructure where all data providers collaborate with each other.

Equally important is implementing new online reporting software and business intelligence strategy. Healthcare needs to catch up with other industries that have already moved from standard regression-based methods to more future-oriented like predictive analytics, machine learning, and graph analytics.

Applications of Big data in Healthcare

1) Patients Predictions For An Improved Staffing: One of they key data sets is 10 years' worth of hospital admissions records, which data scientists crunched using "time series analysis" techniques. These analyses allowed the researchers to see relevant patterns in admission rates. Then, they could use machine learning to find the most accurate algorithms that predicted future admissions trends.

2) Electronic Health Records (EHRs): The integrated system has improved outcomes in cardiovascular disease and achieved an estimated $1 billion in savings from reduced office visits and lab tests.

3) Real-Time Alerting: In hospitals, Clinical Decision Support (CDS) software analyzes medical data on the spot, providing health practitioners with advice as they make prescriptive decisions.
For example, if a patient's blood pressure increases alarmingly, the system will send an alert in real time to the doctor who will then take action to reach the patient and administer measures to lower the pressure.
Another example is that of Asthmapolis, which has started to use inhalers with GPS-enabled trackers in order to identify asthma trends both on an individual level and looking at larger populations. This data is being used in conjunction with data from the CDC in order to develop better treatment plans for asthmatics.

4) Using Health Data For Informed Strategic Planning: The use of big data in healthcare allows for strategic planning thanks to better insights into people's motivations. Care mangers can analyze check-up results among people in different demographic groups and identify what factors discourage people from taking up treatment.

University of Florida made use of Google Maps and free public health data to prepare heat maps targeted at multiple issues, such as population growth and chronic diseases. Subsequently, academics compared this data with the availability of medical services in most heated areas. The insights gleaned from this allowed them to review their delivery strategy and add more care units to most problematic areas.

5) Telemedicine
Telemedicine has been present on the market for over 40 years, but only today, with the arrival of online video conferences, smartphones, wireless devices, and wearables, has it been able to come into full bloom. The term refers to delivery of remote clinical services using technology.
Clinicians use telemedicine to provide personalized treatment plans and prevent hospitalization or re-admission. Such use of healthcare data analytics can be linked to the use of predictive analytics as seen previously. It allows clinicians to predict acute medical events in advance and prevent deterioration of patient's conditions.

By keeping patients away from hospitals, telemedicine helps to reduce costs and improve the quality of service. Patients can avoid waiting lines and doctors don't waste time for unnecessary consultations and paperwork. Telemedicine also improves the availability of care as patients' state can be monitored and consulted anywhere and anytime.

Thereby in this case study big data shows tremendous advantages to healthcare industry and improved results are guaranteed.