QURIOSITY

THE MONTHLY NEWSLETTER FROM QUANTINUUM

BOOK REVIEW

BIG DATA

A REVOLUTION
THAT WILL TRANSFORM HOW
WE LIVE, WORK, AND THINK

VIKTOR MAYER-SCHÖNBERGER
KENNETH CUKIER

QUANT GURU SHAKUNTALA DEVI

QUANT FUN

READY SET GO!
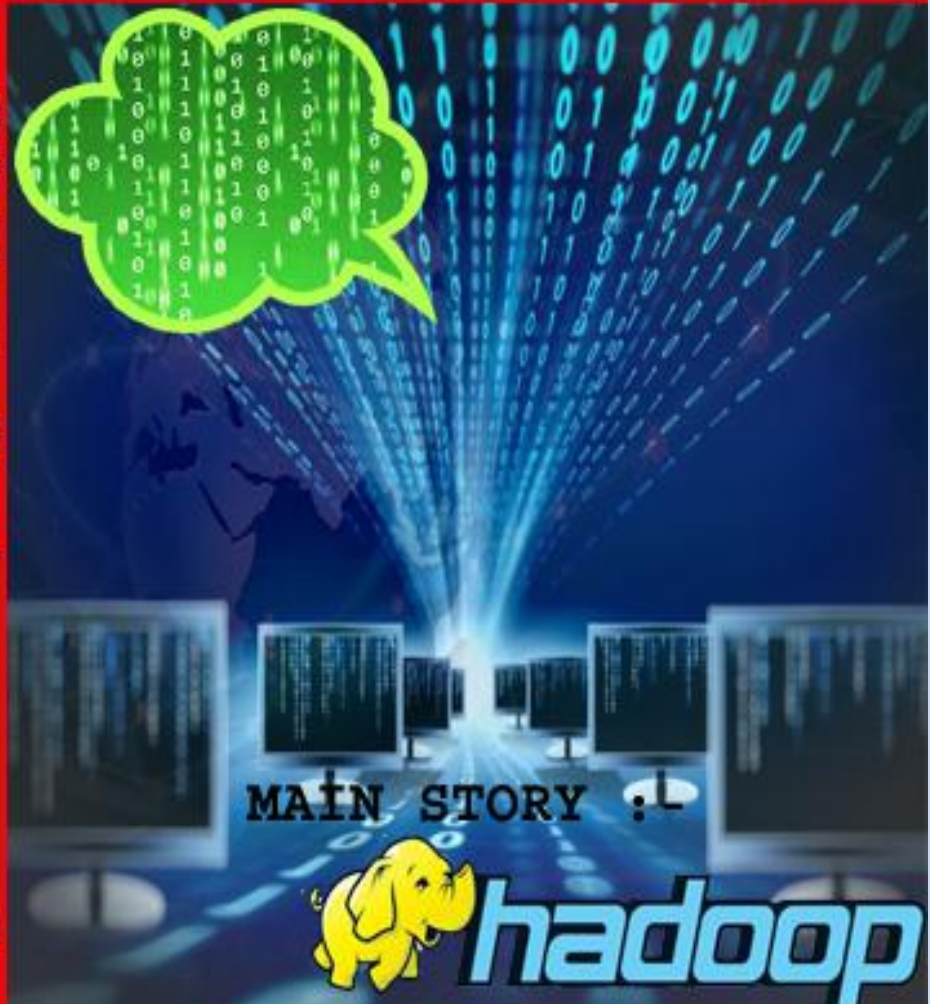
MAIN STORY :

hadoop

From the Mentor's Desk...

Dear All

The July issue of Quriosity is in your hands.
In this issue, I am pleased to observe that the new members who were inducted during July , have already started with their contributions.
You will observe that most of the articles have been written by them.
For this, I say keep it up and let us get going.
The issue cover other regular features like Book review, Quant Guru and other interesting features.

We hope you will like it all.

Happy Reading

Regards
Prof N.S.Nilakantan
Mentor- QUANTINUUM

From The Editor's Desk

Greetings to all,

We are extremely delighted to come out with yet another captivating and informative issue of "Quriosity" : the quant magazine @ SIMSR for you.

This issue covers the main story on "**HADOOP**" which will facilitate a further understanding in the sphere of Big Data Analytics.

The feature article on "Analytics" encapsulates the growing importance of analytics in the world of business.

Quant APP furnishes a detailed statistical analysis on the aggravating problem of rising complaints against auto – rickshaw drivers in Mumbai.

The book review covers a highly enlightening read – "**Big Data – A revolution that will transform how we live , work , and think**" by Kenneth Cukier and Victor Mayer – Schonberger .

Quant guru is a tribute to a highly revered mathematician '**Shakuntala Devi'.**

Further to stimulate the grey matter in your brain; we bring you the wonders of our regular features- **quant trivia, quantiz and quant fun**

Happy reading!!

EDITOR

HADOOP – Delivering Big Data Analytics

Ever wondered what enables search engines to deliver lightening fast results or Facebook to search for your long forgotten friend in seconds? The answer is not another tech heavy word rather it is as cute as a name that a child gave to his toy elephant – 'Hadoop'.

Hadoop (or Apache Hadoop as it is technically known) is an open-source software framework that allows processing of large data sets across clusters of servers. It helps in processing and analyzing a large amount of data and enables you to run analytics that are deep and computationally extensive. Technically, Hadoop has been derived from Google's MapReduce technology. Doug Cutting and Mike Cafarella are credited with development of Hadoop in 2005 (Doug named it after his son's toy elephant). Yahoo has played a key role in evolution of Hadoop over the years.

So what really constitutes Hadoop platform? Let's have a look at the building blocks of Hadoop architecture:

1. **Hadoop Common**: The common utilities that support other Hadoop modules.
2. **Hadoop MapReduce**: A system for parallel processing of large data sets.
3. **Hadoop Distributed File System (HDFS)**: A common file system that spans all the nodes in a Hadoop cluster for data storage. It links together the file systems on many local nodes to make them into one big storage system.

The secret behind Hadoop's storage efficiency for large data is that Hadoop spreads it out i.e. HDFS assumes nodes will fail, so it achieves reliability by replicating data across multiple nodes. And as far as its computational efficiency and intensity is concerned it can be attributed to its ability to harness the power of multiple processors running in parallel.

The advantages that Hadoop offers to businesses are:

1. **Scalable:** New nodes can be added as needed and that too without the need to change the data formats, the way data is loaded or applications on top.
2. **Cost Effective:** It offers an affordable way to model all your business data. It brings in the advantage of parallel processing to commodity servers, bringing down the cost of data storage drastically.

3. **Flexible:** Hadoop isn't restricted to a single schema of data storage. It can be used to store and analyze any type of data be it structured or complex, from a single source or multiple sources. Data from multiple sources can be aggregated enabling deeper and complex analysis.
4. **Fault tolerant:** This is one striking feature that every business will love to have. Imagine a situation where you lose some business data (that is indeed a nasty time for management) here in Hadoop comes to your rescue by redirecting your work to another location of data and you are good to go.

A wide variety of companies use Hadoop for research and production Some of these include Facebook, Yahoo, Amazon.com, American Airlines, Apple, eBay, Electronic Arts, Microsoft, Google, SAP AG, NetFlix, Twitter.

An interesting case at hand is Facebook, considering its fan following. Facebook claims to have the largest Hadoop cluster in the world with an estimated storage of 100 PB (Petabyte) that grows at the rate of roughly half a PB per day. Facebook has used Hadoop for integrating its messaging service with SMS, e-mail and chat. Storing of these messages required high write throughput with cheap and elastic storage, all of this was made possible through Hadoop. Other than this real time analytics related to Facebook activities across websites with Facebook plugins, ads and pages (like click through rates, website visits) were derived using Hadoop platform. Such systems were required to be highly fault tolerant.

As per an estimate about 80% of the world's data is unstructured and most businesses don't even attempt to use this data to their advantage. Imagine a way to store and analyze all this data, Hadoop shows the way forward, it definitely is a competitive advantage for businesses that feed on data.

PRATIK SHARMA

PGDM(2013-2015)

ANALYTICS – Revolutionized The World

Analytics, the word that runs through the nerves of today's business world. Every business existing today is in the need to process data in quintillions. The data, which the business handling today is proliferated in a gargantuan scale. It started in the range of petabytes in 1980's and now has reached a few hundred exabytes. Today data analytics has become a great boon for the researchers to get clear idea about the business flow. The quench for big data has further aggravated the application of analytics across all businesses which are handling mammoth of datasets. It uses predictive and descriptive models to gain valuable knowledge from the data given and to derive an insight for further action that tiles the way for effective decision making.

Analytics plays a great role in market optimization where the effect of campaigns and demographic studies help in achieving consumer targeting. It also helps in portfolio management for risk assessment from which the banks and investment firms aims at minimising the risk and maximising the returns. Predictive models aims at assessing individual's delinquency behaviour and credit worthiness which has become essential fragment in finance industry.

Data analytics has fascinating applications in the field of scientific research. The data obtained from large particle physics experiments are processed to get the relevant information of interest. It is capable of processing the data that runs parallel across thousands of servers. It has also revolutionised the cloud and social space where happens the large commercialisation of the world we live in.

The emergence of analytics has created revolutions in social networking space. Facebook has evolved with variant analytic techniques to keep track of the clicks in the external plugins. It helps even the small advertisers to keep trailing the targeted audience based on their age, gender and interests.

Recently, Twitter has acquired BackType whose integration helps website owners understand the traffic they receive from Twitter and to measure its effectiveness on their businesses. Google analytics helps in tracking visitors from referrers, including search engines, social networks and direct visits. The crave for big data has deepen its roots across multitude fields.

Analytics has wide variety of business application in maintaining customer retention, direct marketing, behaviour based advertising, email targeting and so on. It provides better insights to the business users with the operational data stored in transactional systems, thereby helps in integrating analytics with action and measurement. Diverse scenarios with a myriad of locations and services need analytic solution that can provide decision makers with comprehensive intelligence to take sound and effective decisions.

PRIYATHARSINI V

PGDM (2013-15)

QUANT APP – A Powerful Tool

CASE:

As per the report in Hindustan Times, dated 24th February, 2013, the number of complaints against auto-taxi drivers has reduced on the RTO (Regional Transport Office) helpline by 50% in Mumbai, especially in the western suburbs.

The current statistics say that complaints from the Andheri region have reduced from 2014 in April, 2011 to 971 in April, 2012. The reason cited was the installation of e-meters and stringent checks on auto drivers.

Commuters say RTO helpline is not publicised properly. Drivers should publish the helpline number on their vehicle. They further said that they are discouraged to complain as they are not sure whether action is ever taken against the accused drivers.

OBJECTIVE OF STUDY:

We want to see whether there has been a significant change in the proportion of complaints against the auto drivers in Mumbai, from 2011 to 2012. By significant change we mean statistically significant change.

For this analysis we use a Z-test, which will compare the proportion of complaints of two populations of the year 2011 and year 2012. The sample taken will be data related to the Andheri region. From the official data from the RTO website, we found the total number of autos plying in the Western part of Mumbai to be 65000, during the year 2010-2011. Thus it is safe to say, out of 65000 autos, 15000 autos ply in the Andheri region. With the help of the sample statistics we will conclude whether the population parameter, that is, the proportion of complaints against drivers in Mumbai has really decreased.

Notation:

Sample statistics:

x1: Number of complaints against drivers from Andheri in 2011
n1: Number of autos plying in Andheri in 2011
p1: x1/n1 = 2014/15000

x2: Number of complaints against drivers from Andheri in 2012
n2: Number of autos plying in Andheri in 2012
p2: x2/n2 = 971/15000

Population parameter:

$\pi 1$: Population proportion of complaints in 2011
$\pi 2$: Population proportion of complaints in 2012

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

HYPOTHESIS TESTING:

Ho: $\pi 1 = \pi 2$, i.e. there is no change in the proportion of complaints from 2011 to 2012.
H1: $\pi 1 > \pi 2$, i.e. there has been a decrease in the proportion of complaints from 2011 to 2012.

Using Megastat in Excel, we get the following results:

Hypothesis test for two independent proportions				
p1	p2	Z-Value	P-Value	Decision taken
0.1343	0.0647	20.12	0.00	Reject Null Hypothesis

CONCLUSION:

The P-value is coming out to be 0 which is less than 0.05, the level of significance. Hence, the null hypothesis is rejected as the p-value is significant enough at 95% level of confidence. Thus we can conclude that the proportion of complaints for the year 2012 is less than that of 2011.
This is one practical use of Statistical Analysis & testing in real-life data.

SATARUPA BANERJEE
PGDM - FS (2012-2014)

BOOK REVIEW: Big Data

Big Data – A revolution that will transform how we live, work, and think is the book by two authors, Kenneth Cukier and Viktor Mayer-Schönberger. Cukier is the Japan business correspondent of The Economist in Tokyo and is also the author of a 14-page cover story in 2010 called "The Data Deluge." Earlier, he was the paper's technology correspondent in London, focusing on intellectual property and Internet governance. , Schönberger is Professor of Internet Governance and Regulation at the Oxford Internet Institute, University of Oxford. He conducts research into the network economy. Earlier he spent ten years on the faculty of Harvard's John F. Kennedy School of Government.

So what is Big Data? "It is a collection of data sets so large and complex that it becomes difficult to process using available database management tools or traditional data processing applications". Big Data is an excellent concept for general masses to what has become a topic of debate everywhere, faster than any other technology driven E-word in recent era. Mayer Schonberger and Kenneth Cukier give us a complete, exhaustive and highly readable overview of the benefits and risks that are associated with big data concept, which they feel is "the ability of society to harness information in novel ways to produce useful insights or goods and services of significant value."

The Crux of the book is when the authors discuss about the potential risks and possible ways to address them, providing a platform to a much needed interaction regarding what and how to be done about big data or go about big data. Apart from privacy, the authors point out that perfect anonymization (Hidden Identity) is impossible in the age of big data.

The chapters on "Risk and Control" take the concepts to further realistic approaches and scenarios. These chapters cover the topic of privacy with the latest perspective and are mostly actionable. Mayer Schonberger and Kenneth Cukier give us a definition of profiling vs. selecting suitable predictors of customer behaviour in the book. But they take the right step in outlining societal complications, such as "penalties based on propensities", which they call "nauseating." The authors also note the rise of the algorithmists – professionals with math, science and computer science backgrounds to help assure accountability for the very systems we create and value.

The book depicts the impediments to revolution in mighty strokes. The authors claim the impact of data driven models and innovations will advance the march of humankind in the right direction.

The book correctly points out that data is rapidly becoming object like raw material to business community. Many organisations will tap into the new data affluence, the outcome of a long historical process that includes 'datification' and the diffusion of technologies that have tremendously reduced the costs involved in data production, storage and processing.

So, where's the revolution? The book argues for three rather simplistic shifts.

The first shift – the new world is characterised by even more data than ever before. The authors say that just as a movie emerges from a series of photographs, increasing amounts of data are as important because quantitative changes bring about qualitative changes in every aspect. The technical equation in big data is the ability to survey a whole population instead of just sampling random portions of it.

The second shift is that looking at vastness of data also permits us to loosen up our desire for exactness. Apparently, in big data, with less error from sampling we can accept more measurement error. According to the authors, science is obsessed with sampling and measurement error as a consequence of coping in a 'small data' world.

The third and most radical shift implies "we won't have to be fixated on causality [...] the idea of understanding the reasons behind all that happens." This is a straw man argument. The traditional image of science the authors discuss (fixated with causality, paranoid about exactitude) conflates principles with practices. Correlational thinking has been driving a lot of processes and institutional behaviours in the real world.

By analysing huge amounts of information, it's possible to discover patterns and relationships that up to now have been invisible to us. In this way, we can find new solutions to tough problems, and opportunities we'd never otherwise have suspected.

This book summarizes the way Big Data has changed the problem solving methods both scientifically and arbitrarily. It has helped many industries to find solution to long term problems using the right data with right techniques.

BHAVIN SATRA & SONIA NIKAM
PGDM – FS (2013-15)

QUANT GURU of the MONTH

Shakuntala Devi (November 4, 1929 – April 21, 2013), popularly known as the "Human Computer", was a child prodigy and mental calculator. Her talents earned her a place in the 1982 edition of The Guinness Book of World Records.

Shakuntala Devi was born in Bangalore, India. Her father discovered about her ability to memorize numbers while teaching her a card trick when she was about three years old. Devi, who had no formal education and who simply picked up reading and writing, had the ingenious ability to tell the day of the week of any given date in the last century in a jiffy. By age six she demonstrated her calculation and memorization abilities at the University of Mysore.

Devi travelled across the globe demonstrating her arithmetic talents, including a tour of Europe and performance in New York in 1976. Her charismatic mathematical abilities were studied by Arthur Jensen, a professor of psychology at the University of California, Berkeley.

In 1977 in USA, she competed with a computer to see who gives the cube root of 188138517 faster, where she won. That same year, at the Southern Methodist University she was asked to give the 23rd root of a 201-digit number; she answered in 50 seconds.

Her answer—546,372,891—was confirmed by calculations done at the U.S. Bureau of Standards by the Univac 1101 computer, for which a special program had to be written to perform such a large calculation. On June 18, 1980, Devi cast a spell demonstrating the multiplication of two 13-digit numbers 7,686,369,774,870 x 2,465,099,745,779 picked at random by the Computer Department of Imperial College, London almost instantaneously. She correctly answered 18,947,668,177,995,426,462,773,730 in 28 seconds. This event is mentioned in the 1982 Guinness Book of Records.

Shakuntala Devi wrote a number of books on mathematics and astrology including "Fun with Numbers", "Astrology for You", "Puzzles to Puzzle You" and "Mathablit". Shakuntala Devi was honoured with the Lifetime Achievement Award in Mumbai a few weeks before her death in April, 2013.

VAISHNAVI K
PGDM (2013-15)

QUANTIZ of the MONTH

1) If the eight-digit number 5668x25y is divisible by 48, find the least value of x+y.

2) Find a four-digit whole number n, such that the last four digits of n^2 are in fact the original number n.

3) A student attempted to find the average of 10 two-digit numbers. Due to some mistake, he interchanged the digits of one of the numbers. As a result, the average was changed by 8.1. Find the number whose digits have been interchanged.

4) What is the missing number in Triangle Four?
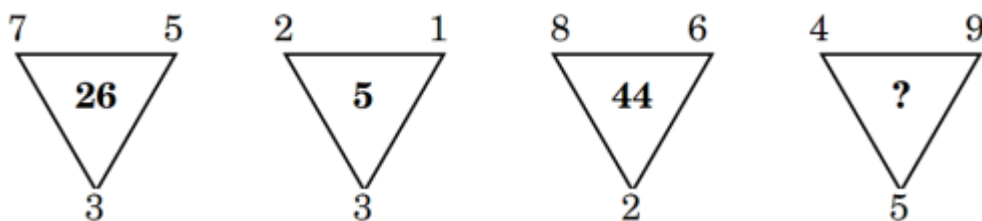
5) There are 20 boys and 30 girls in a class. Boys have Rs. 2 coins each and girls have Rs. 3 coins each. Find the minimum number of transactions required for both of them to have the same amount.

Please send us the answers at simsr.quantinuum@gmail.com. Answers and Name of the winner (first all correct /most correct entry) will be published in the next issue.

KIRAN KUMAR REDDY
PGDM - FS (2013-2015)

QUANT FUN

Sudoku of the Month.

**Please send us the answers at <u>simsr.quantinuum@gmail.com</u>.
Answer and name of the winner will be published in the next issue**

QUANT TRIVIA

"The name of the popular search engine 'Google' came from a misspelling of the word 'googol' which is a very large number- the number one followed by hundred zeros to be exact"

QuantConnect

Quantinuum, the Quant's forum of KJ Somaiya Institute of Management Studies and Research is formed with two objectives. Firstly to remove the common myth from the students mind that mathematics is difficult. Secondly to give students an exposure on how to make decisions in real life business problems using quantitative techniques. This helps to bridge the gap between theory and the practical application.

For any further queries and feedback, please contact the following address
KJ Somaiya Institute of Management Studies and Research
Vidya Nagar, VidyaVihar, Ghatkopar East
Mumbai -400077

Mentor
Prof. N.S.Nilakantan (9820680741) – email – nilakantan@.somaiya.edu

Team Leaders
Manish Murthy (9167679676)
Varun S (9819104859)
Aditi Paliwal (9819116068)
Abhishek Kumar (9819099671)

Editorial Team
Aditi Paliwal (9819116068)
Abhishek Kumar (9819099671)

For any queries, drop us a mail at simsr.quantinuum@gmail.com
For more details: http://quantinuum.weebly.com/

Like our newsletter? Want to contribute and see your name in print?

Rush your articles, concepts, trivia, facts and news about the

Wonderful World of Numbers to us by email to simsr.quantinuum@gmail.com.