Somaiya TRUST

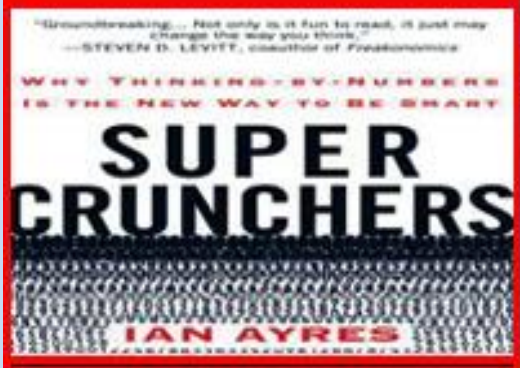
SIMSR

QURIOSITY

THE MONTHLY NEWSLETTER FROM QUANTINUUM

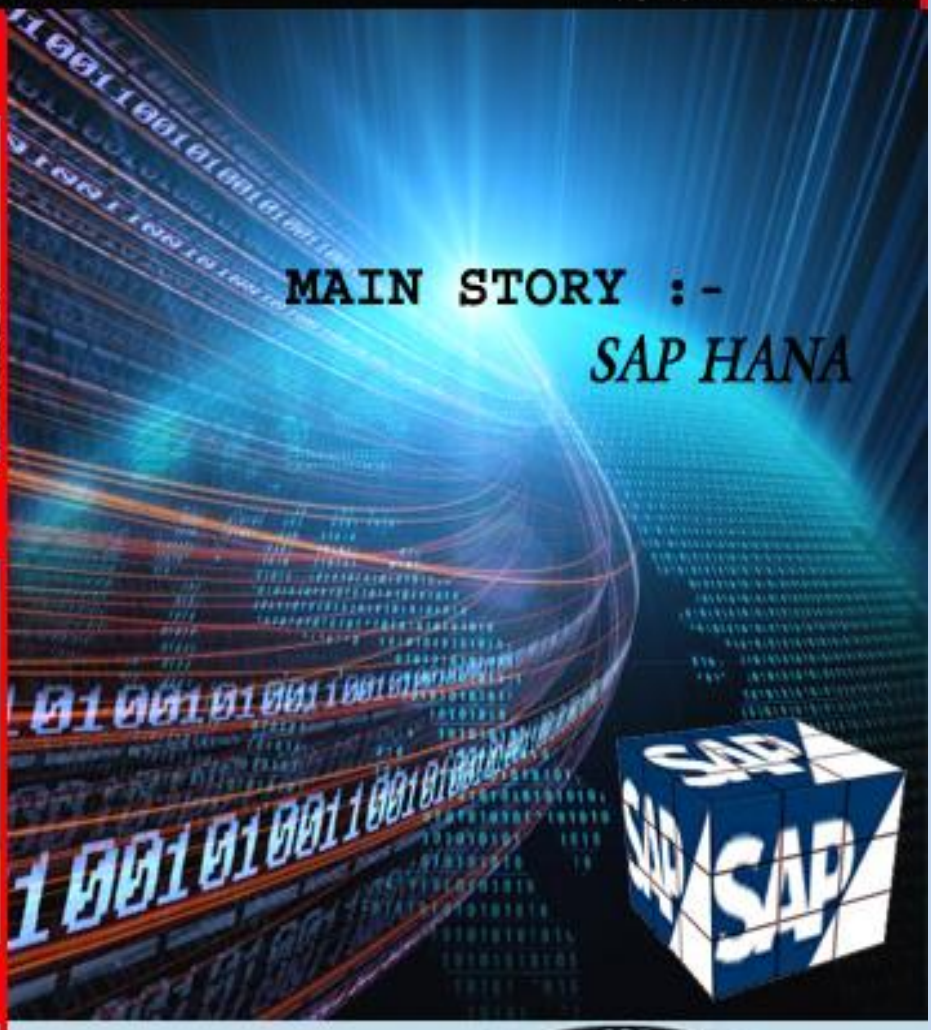
FEBRUARY 2013 VOLUME 4 : ISSUE 1

BOOK REVIEW



QUANT GURU
K.S.S NAMBOORIPAD

QUANT FUN



K.J Somaiya Institute of Management Studies & [illegible], Vidyavihar,Mumbai

Quantinuum - The Quants Forum of SIMSR

Its more than NUMBERS

THE

Quriosity

FEBRUARY 2013 VOLUME 4: ISSUE1

From the Mentor's Desk...

Dear All,
We are pleased to issue our first newsletter for 2013, right after the annual day which was held in Feb, 2013.

We are planning to introduce new features as we go along for which we need your wholehearted support. Please give any useful inputs by writing to us.

This issue covers SAP Analytics, 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.

The issue covers a main story on "SAP HANA" which will help demystify the intriguing concept of SAP and unravel it in the simplest way possible.

Quant digest furnishes interesting reads in the form of articles such as –"Factoring in the deadly math of cancer" and "Statistics help clear fog for better climate change picture".

The book review covers a highly enlightening subject– "Super Crunchers: Why Thinking-By-Numbers is the New Way to be Smart"

Quant guru is a tribute to a highly revered mathematician of current times '**K.S.S.Nambooripad'.**

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

SAP HANA – The Analytical Beast

SAP is undoubtedly the world leader in Enterprise Resource Planning (ERP) and is here to stay. With unmatched technical sophistication, SAP AG has released a new in-memory database technological architecture, named SAP HANA.**SAP HANA** often called as High Performance Analytic Appliance is a one-stop shop for the all your analytical operations. Taking a diversified attempt from its core competence in database management systems towards a high performance analytical and transactional application running on low cost RAM's, multi core processors and the fast access of solid state drives. They have ended up with a State-of-the--Art Architecture Analytical Softwares.

While it was initially envisioned as an analytics engine that would simply help customers perform queries much faster because they were conducted in memory, HANA is now more than that. "SAP HANA evolved very quickly from a high-performance analytical appliance to a 'platform' for real-time analytics and applications. The HANA appliance is at the core of everything; the trick is getting it to connect to whatever resources you want to speed up, and to some sort of client that enables the whole thing to be useful to end users. That requires either a direct interface to HANA or some sort of intermediary layer.

There are four components within the software group:

- SAP HANA DB (or HANA DB) refers to the database technology itself,
- SAP HANA Studio refers to the suite of tools provided by SAP for modelling,
- SAP HANA Appliance refers to HANA DB as delivered on partner certified hardware (see below) as an appliance. It also includes the modelling tools from HANA Studio as well as replication and data transformation tools to move data into HANA DB,
- SAP HANA Application Cloud refers to the cloud based infrastructure for delivery of applications (typically existing SAP applications rewritten to run on HANA).

It started as a research project in 2008, for SAP AG along with Hasso Plattner Institute & Stanford University on the basis of an idea of a possible new architecture. It is one that enables real-time complex analytics and aggregation, up to date with every transaction, in a way never thought possible in financial applications".

In 2009 a development initiative was launched at SAP to integrate the three technologies namely, TREX (Text Retrieval and Extraction), P*Time (Online Transaction Processing) & MaxDB (for Persistence of a traditional database) to provide a more comprehensive feature set. The resulting product was named internally and externally as NewDB until the change to HANA DB was finalized in 2011.

SAP HANA is not SAP's first in-memory product. Business Warehouse Accelerator (BWA, formerly termed BIA) was designed to accelerate queries by storing BW infocubes in memory. This was followed in 2009 by Explorer Accelerated where SAP combined the Explorer BI tool with BWA as a tool for performing ad-hoc analyses. Other SAP products using in-memory technology were CRM Segmentation, By Design (for analytics) and Enterprise Search (for role based search on structured and unstructured data). All of these were based on the TREX engine.

It is designed as an entire package containing libraries of business application functions, namely, Business Function Library & Predictive Analysis Library. It also includes the open source programming language for statistical analysis tool, "R".

The Persistency Layer is responsible for the durability and atomicity of transactions. It manages data and log volumes on disk and provides interfaces for writing and reading data that are leveraged by all storage engines. This layer is based on the proven persistency layer of MaxDB, SAP's commercialized disk-centric relational database. The persistency layer ensures that the database is restored to the most recent committed state after a restart and that transactions are either completely executed or completely undone. To achieve this efficiently, it uses a combination of write-ahead logs, shadow paging, and savepoints.

One implication of HANA's ability to work with a full database in memory is that computationally intensive KPI (Key Performance Indicators) calculations can be completed rapidly when compared to disk based databases. Pre-aggregation of data in cubes or storage of results in materialized views is no longer necessary.

SAP HANA Information Composer is a web based tool which allows users to upload data to a HANA database and manipulate that data by creating Information Views. In the data acquisition portion, data can be uploaded, previewed and cleansed. In the data manipulation portion objects can be selected, combined and placed in Information Views which can be used by SAP Business Objects tools.

Security and role based permissions are managed by the Authorization Manager in HANA DB. Besides standard database privileges such as create, update or delete HANA DB also supports analytical privileges that represent filters or drill-down limitations on queries as well as access control access privileges to values with certain attributes.

HANA DB components invoke the Authorization Manager whenever they need to check on user privileges. The authentication can then be done either by the database itself or be further delegated to an external authentication provider, such as an LDAP directory.

Varun S

PGDM-IB (2012-14)

QUANTS NEWS DIGEST

"Factoring in the deadly math of cancer"

Two Duke Researchers are focusing on the deadly mathematics behind the mutated genes and damaged cells that drive cancer.

Cancer is the end result of an accumulation of genetic mutations," says Rick Durrett, a professor of mathematics at Duke. "It can be boiled down into a series of probabilities of whether or not a cell will become mutated, whether the cell will get the correct combination of mutations to become cancerous, and at what rate the cancerous cells continue to divide."

Cervical cancer illustrates his point. Tumors on a woman's cervix develop from a series of mutations associated with chronic infection from human papillomavirus (HPV), which the Center for Disease Control lists as the world's most common sexually transmitted infection.

Epidemiologist Evan Myers, a professor of obstetrics and gynecology at Duke, has created models showing how screening or vaccinating for HPV affects the likelihood of an individual or group getting cervical cancer.

What's needed, Myers says, are better models of the disease's underlying biology. That's where mathematicians such as Durrett and Marc Ryser, a visiting assistant professor in Duke's math department, enter into the cancer equation (ref: http://www.biomedcentral.com/1471-2407/10/3). Mathematical models can complement clinical and biological data of the tissue-level effects of HPV, Ryser says. He explains that the mutations and cellular dynamics of the virus are hard to observe and track in real patients, but mathematical models can simulate an infection's progress without sampling a single cell.

"With our model, we can calculate the probability of infected cells continuing to divide and mutate. We can run simulations to see how the disease spreads in an individual and how it could spread to a person's sexual partners," Ryser says.

In the future, he would like to combine these kinds of tissue-level HPV models with the models Myers is developing for entire populations. The models together may let clinicians see "what is possible biologically and if it consistent with what we see clinically," Myers says. "If the models match reality, we could start to use them to make predictions about what transmission or treatments look like in the real world."

“**Statistics Help Clear Fog for Better Climate Change Picture**”

Statistics is an important tool in sorting through information on how human activities are affecting the climate system, as well as how climate change affects natural and human systems, according to a Penn State statistician.

"One key aspect of climate change is risk," said Murali Haran, associate professor of statistics. "Without the language of statistics and probability, you can't talk about risk."

As more research is conducted and more data are gathered, Haran said that scientists are gaining a better understanding of current and future climate conditions, as well as predicting the risk of the dramatic and costly affects of this change.

"We have a better understanding of the climate now than we have ever had before," said Haran. "With greater availability of data and more sophisticated climate models, our knowledge continues to increase."

In addition to realizing how the climate may change, risk managers need to assess what affect any changes -- even low probability ones -- can have on the economy and society, according to Haran.

"Recognizing that some low probability events can create high impact outcomes is also important," said Haran.

One of the analytical tools that Haran and other statisticians use to make assessments on the future of the environment is **Bayesian statistics**(ref : http://en.wikipedia.org/wiki/Bayesian_statistics), which is a formal system of statistical inference that uses all available current information to estimate the probability of future events.

Haran said that some people misinterpret uncertainty to mean not knowing about climate change, but quantifying uncertainty actually refers to expressing a range of sureness on assessments or predictions, which is central to careful science. Researchers from across disciplines must work together to more efficiently study climate change and other problems that society faces, he said

ADITI PALIWAL
PGDM COMMUNICATION (2012-2014)

BOOK REVIEW: SUPER CRUNCHERS

Super Crunchers: Why Thinking-By-Number is the New Way to be Smart by Ian Ayres, is a popular science book about the mathematical and statistical analysis of large datasets. It addresses the way our culture is changing in the present world, by concentrating on Data Mining from a science perspective.

The author defines super crunching as "*It is statistical analysis that impacts real-world decisions. Super Crunching decisions usually bring together some combination of size, speed and scale. The sizes of the datasets are really big — both in the number of observations and in the number of variables…And the scale of the impact is sometimes truly huge. This isn't a bunch of egghead academics cranking out provocative journal articles. Super Crunching is done by or for decision makers who are looking for a better way to do things*." (Pg 10) In other words, it is data mining. The book concentrates more on the applications and implications of these powerful tools, rather than the process of mathematical analysis. The core subject of this book is how to make decision in this super crunching world – evidence or intuitive? Evidence triumphs every time.

Interestingly, the upsurge in data mining is largely attributed to the drastic decrease in storage costs in last few years, much more than any rise in processing power. On the other hand, neural network technology has helped evolve better techniques.

The book explains different applications in various industries and shows how data mining has affected in each of these industries – basically proving that the evidence by statistics overruns human intuition every time. It explains the unique nature of human expertise: what does it really take to be a human expert, when math experts can convert large data sets into much more accurate predictions about human behavior? Of course, the humans play an important role of deciding what data is to be collected, what questions to be asked, whom to ask and how to apply the results of the analysis.

To the credit of the author, he does not seem to miss out on the hard issues in such scenarios – privacy concerns, dangers of over-reliance on programmed creativity and other similar areas. The author explains that data mining is a very powerful technology, and while balance is essentially a safe-play practice, understanding it completely is much more preferable for the change.

Instead of a Luddite rejection of this powerful new technology, it is better to become a knowledgeable participant in the revolution. Instead of sticking your head in the sands of innumeracy, I recommend filling your head with the basic tools of Super Crunching. (p.191)

He has had a clear way to introduce the new technology and analyzing its impact from a 360-degree perspective, with pros and cons. The tool can be used for many purposes and it is important to be passionate about the needs to inculcate the basic knowledge of statistics in the public. Quoting the author - "*We have to get students to learn this stuff…We have to get over this phobia and we have to get over this view that somehow statistics is illiberal. There is this crazy view out there that statistics are right-wing*"…One can crunch numbers and still have a passionate and caring soul. You can still be creative. You just have to be willing to put your creativity and your passions to the test to see if they really work. (p. 215)

This book summarizes the way technology of data mining has changed our culture and its importance in the current scenario from a science and business perspective.

Manish Murthy
PGDM 2012-14

Editor's note:
This book is published by Bantam Books and available from Amazon also.

QUANT GURU of the MONTH

Nambooripad was born on 6 April 1935 in a middle class namboodiri family in Puttumanoor near Cochin in central Kerala. He received traditional vedic education up to the age of fifteen after which he joined a modern school offering formal education. He obtained the B.Sc.(Hons) degree of University of Kerala from Maharaja's College, Ernakulam, in 1956. He spent a few years teaching mathematics in some privately managed colleges before joining the newly started Department of Mathematics, University of Kerala, as a research scholar in mathematics in 1965.

He was initially under the supervision of Prof. M. R. Parameswaran. A year later he came under the guidance of Prof. B. R. Srinivasan. About two years later, consequent on the departure of Prof. B. R. Srinivasan from University of Kerala, Nambooripad became a student of Prof. Y. Sitaraman. He was awarded the Ph D degree in 1974. Nambooripad's basic contributions are in the **structure theory of regular semigroups**. A semigroup is a set S together with an associative binary operation in S. A semigroup S is said to be regular if for every a in S there is an element b in S such that aba = a.

Nambooripad axiomatically characterized the structure of the set of idempotents in a regular semigroup. He called a set having this structure a biordered set.

"The axioms defining a biordered set are quite complicated. However, considering the general nature of semigroups, it is rather surprising that such a finite axiomatization is even possible." A full treatment of the theory was published as a single paper number of the Memoirs of American mathematical Society in 1979. "In the mid 70's A. H. Clifford became very much excited by the work of Nambooripad on the structure of regular semigroups in terms of idempotent ordering and sandwich matrices and wrote several expository papers on Nambooripad structure theorem for regular semigroups".

He later developed an alternative approach to describe the structure of regular semigroups. This particular work utilizes the abstract theory of cross-connections to provide a useful framework for studying various classes of regular semigroups.

TeX (a popular typesetting for complex mathematical formulae) was introduced into Kerala by Nambooripad. After a visit to the United States in early 1990s, he brought the TeX programme to Kerala in fourteen floppy disks. Nambooripad encouraged his students to learn and use TeX, especially for preparing their theses. One of his students was E. Krishnan, one of the authors of the LaTeX primer2 published as an electronic book by the Indian TeX User Group. Krishnan also played an important role in establishing the Free Software Foundation of India. Another person inspired by Nambooripad was C.V.Radhakrishnan who is running a company called River Valley Technologies since 1995 for typesetting of scientific journals and books. Nambooripad was the prime catalyst for the formation of Indian TeX Users Group in 1998. He was the inaugural Chairman of the Group.

He was with the Department of Mathematics, University of Kerala, since 1976. He served the Department as its Head from 1983 until his retirement from University service in 1995. After retirement, he is associated with the academic and research activities of the Center for Mathematical Sciences, Thiruvananthapuram in various capacities

DHARA KATKORIA
MMS 2012-14

QUANTIZ of the MONTH

Q1) There are 100 employees in a conference room in New York City. You note that 99% of them are managers. How many managers would need to leave the conference in order to reduce the percentage of managers in the hall to 98?

Q2) There are 25 horses, each of which runs at a constant speed that is different from the other horses. Since track only has 5 lanes, each race can have at most 5 horses. If you need to find the 3 fastest horses, what is the minimum number of races needed to identify them?

Q3) There are two whole numbers, difference of their squares is a cube and the difference of their cubes is a square. These are the smallest possible numbers Can you find the numbers?

Q4) Four people A, B, C and D need to cross a river. The only way to cross the river is by an old bridge, which holds at most 2 people at a time. Being dark, they can't cross the bridge without a torch, of which they only have one. So each pair can only walk at the speed of slower person. They need to get all of them across to the other side as quickly as possible. A is the slowest and takes 10 minutes to cross, B takes 5 minutes, C takes 2 minutes and D takes 1 minute.

What is the minimum time to get all of them across to the other side?

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.

Solutions to last issue's Quiz of the month

1. 2:3

2. 100

3. 25 m/sec

4. 46 %

5. 6

Keeping the Grey matter Alive!

QUANTIZ TEAM

QUANT FUN

Sudoku of the Month.

							2	
		2	8	4			1	6
7	6				2		8	
	5	3			7			
9	4						7	2
			4			3	5	
	3		2				4	7
4	2			9	5	1		
	7							

Please send us the answers at simsr.quantinuum@gmail.com.

Answer and name of the winner will be published in the next issue.

Solution of the Last Month Sudoku:

9	6	3	5	7	8	4	1	2
1	2	8	3	6	4	7	5	9
4	7	5	1	2	9	3	8	6
8	9	1	4	3	2	5	6	7
7	3	4	6	5	1	9	2	8
2	5	6	9	8	7	1	4	3
3	4	9	2	1	6	8	7	5
6	1	7	8	9	5	2	3	4
5	8	2	7	4	3	6	9	1

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 the hundred zeros to be exact)"

Quant Connect

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@simsr.somaiya.edu

Team Leaders

Satyadev Kalra (8291687568)
Abin Abraham (9594374903)
Somjeet Dutta (9769513003)
Vaibhav Goel (9769456493)

Editorial Team

Aditi Paliwal (9819116068)
Abhishek Kumar (9819099671)
Manish Murthy (9167679676)
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.