



Yearbooks and Newsletters

Spring 2015

ACM Chapter Volume XV Issue I

Touro College Student Chapter of the Association of Computing Machinery

Follow this and additional works at: https://touro scholar.touro.edu/archives_books



Part of the [Higher Education Commons](#), and the [Science and Technology Studies Commons](#)

Recommended Citation

Touro College Student Chapter of the Association of Computing Machinery. (2015). ACM Chapter Volume XV Issue I. Retrieved from https://touro scholar.touro.edu/archives_books/253

This Book is brought to you for free and open access by Touro Scholar. It has been accepted for inclusion in Yearbooks and Newsletters by an authorized administrator of Touro Scholar. For more information, please contact touro.scholar@touro.edu.

TOURO COLLEGE

STUDENT CHAPTER OF THE ACM

PRESENTS

Big Data and Data Science

Spring 2015

VOLUME XV : ISSUE I

Welcome to the Touro College ACM Student Chapter

Welcome to the Spring 2015 edition of the Touro College Student Chapter of the ACM (Association of Computing Machinery) newsletter published by the Flatbush Campus Women's Division. The purpose of the student chapter of the ACM, as described in the bylaws, is "to promote an increased knowledge of, and greater interest in, the science, design, development, construction, languages, management, and applications of modern computing." This newsletter certainly helps accomplish that goal.

The theme of this newsletter is big data and data science. Big data is a relatively new concept in computing. The term was only added to the Oxford English Dictionary less than two years ago and to Merriam-Webster's Collegiate Dictionary within the last twelve months. These are fascinating topics which have evolved significantly over the past few years and will, undoubtedly, continue to develop and play an increasingly more central role in the field of computer science in the years to come.

As the quantity of the data being produced annually continues to increase exponentially, so do the challenges inherent in analyzing, interpreting, and utilizing that data. The field of data science, while not exclusively limited to big data, addresses these challenges. Reading the articles in this newsletter, as well as taking advantage of other opportunities to learn about big data and data science, will certainly prove to be beneficial.

I would like to encourage all Computer Science and MIS majors to join the chapter of the ACM. Membership in the student chapter affords you the opportunity to meet and observe accomplished professionals from various areas of the computing industry. In addition, participation in a professional organization beyond the required course work serves as evidence of serious interest and dedication to advancing in the computer industry, something employers in today's job market are always looking for.

Given the competitive nature of the application process, extracurricular involvement such as this might give you the needed edge over other job applicants. Keep in mind, however, that you may be asked on an interview what you did as a member of the Student Chapter of the ACM. The more actively you have participated in chapter events and publications, the more impressive your answer will be. Joining the ACM Student Chapter is free. All you have to do is fill out the form and return it to one of the lab technicians. Your level of participation, however, is up to you.

Our current officers are: Devora Green – President, Rachel Aziza Bensoussan – Vice President, Esti Sendrovitz – Secretary, and Chayala Glazer – Treasurer. I would like to thank Devora for all of her hard work organizing and pulling together this newsletter, as well as this semester's meeting. I would also like to thank the other contributors to this newsletter.

Please consider writing an article for the next edition of the newsletter. Writing an article will give you the opportunity to participate in a meaningful way, while at the same time research and learn about a cutting edge facet of technology. Please approach me, or one of the officers, if you would like to volunteer or if you have any ideas for future chapter speakers or events.

I would also like to thank Dr. Issac Herskowitz and the faculty of the Department of Computer Science for the phenomenal job they have done preparing our students for this ever-changing career. They expend countless hours learning new technologies to ensure that our graduates will have state-of-the-art skills. Without our faculty's hard work and willingness to go the extra mile, our graduates would not be where they are today.

Shmuel Fink
Deputy Chairman
Department of Computer Science, Touro College

BIG DATA

DEVORA GREEN

Big data is the term used to reference the collection of enormous amounts of data. The data can be in a structured, semi- structured or unstructured format which makes it hard to input into relational databases. In addition to this, the data sets are so large that they cannot be processed using standard procedures. Companies have to learn how to deal with the data and use it to their advantage. Many new companies were built around big data as they came into existence while old companies have to integrate it with their existing systems for data collection and analysis.

Big data is usually classified by the three V's; Volume, Velocity, and Variety. Volume refers to the amount of data, which has increased exponentially over the years. In the past, the storing of data was an issue in terms of the amount of storage required, but today, that is no longer the case as there are new forms of storage

that are cheaper and larger. Now the issue is how to determine which parts of the data are important and how to analyze it properly in order to gain from it. Velocity is the speed at which they are receiving the data and the speed that it needs to be processed in. This has also changed due to technological advances. Now data needs to be dealt with in almost real time. Variety of data is referring to the format it comes in. Having the ability to deal with structured numeric data, unstructured documents, emails, videos and many other forms of data is a struggle. However, according to a survey conducted by the



New Vantage Partners in 2012, this is one of the most enticing factors for businesses. They (cont. on pg. 5)

Table of Contents:

Big Data	
Devora Green	1
Data Analytics	
Chava Shapiro	1
Data Science	
Nechama Spira	2
Data Mining	
Ariella Feuchtwanger	3
Finance	
Chayala Glazer	6
Hadoop	
Atara Ney	7
SAS	
Ahuva Neuhoff	10
Data Journalism	
Rena Kerstein	11
Health Care	
Malkie Golub	12
Hour of Code	
Rachel Bensoussan	13
Job Tips	
Prof. B. Sheinfeld	15

Layout: Esti Sendrovitz

Data Analytics

CHAVA SHAPIRO

Data analytics, also called data analysis, is the science of studying data and using it to draw conclusions and find patterns in the data for the purpose of making decisions. By using special software tools and applications, data analytics allows data scientists to analyze incredibly large amounts of data that regular analytics cannot process, thereby aiding in finding solutions to the most challenging business problems.

Most big analytics projects begin with a question a business has. Data analytics is a way for the business to get an answer. The first step of data analysis is finding and gathering data from different sources. Data is collected from structured databases and from unstructured systems, such as emails, books, and articles. The collection of data from unstructured sources is unique to data analytics. After all of the data is collected, special software tools and (cont. on pg. 3)

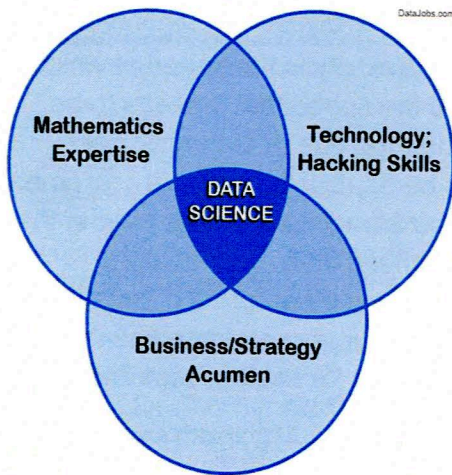
What is Data Science?

NECHAMA SPIRA

In general terms, Data Science is the extraction of knowledge from data. But what really is data science and why is it such a big hype now? What do data scientist do, and why are they so important? There is a massive amount of data on the web that is essential to every one of us. Our daily lives include shopping, reading the news, listening to music, searching for information, expressing our opinions and communicating with each other. All of the above is internet data. In addition, every business, such as, the medical industry, pharmaceuticals, bioinformatics, social welfare, government, education, and retail have their own important data. The influence that data is having on every industry is growing at a rapid rate. At the same time, an abundance of low-cost computing power is available and new technologies are emerging to organize and make sense of the avalanche of data.

Simply using all the data that is available to us is not actually what we mean by “data science”. A data application attains its value from the data itself, and generates more data as a result. It’s not just an application with data; it’s a data product. Data science allows the production of data products.

There are many different aspects and skills that are involved in data science. The three main expertise that must lie in a data scientist are Mathematics, technology; hacking skills and business/strategy acumen. The first one being mathematics, we must understand that at the core of deriving insight from data is the capability to view the data through a numerical lens. Also, many business problems solution often include building analytic models that are deeply grounded in the hard math theory, and comprehending how models work is as essential as knowing the process to build them. Additionally, data



science is all about statistics and therefore statistics must be well understood by data scientists.

The second important skill a data scientist must have is technology and hacking skills. When we talk about hacking skills we are referring to originality in using technical skills to build things and to find ingenious solutions to problems. You can think of a hacker as a solid algorithmic thinker – that is, being able to take a messy problem and reconstruct them in to more decipherable ways. The last most important skill is strong business acumen. It is very important to realize that data scientist are strategy consultants. Since they can learn from data the way no one else can they can offer businesses extremely valuable resources. A data scientist with all three components would be an incredible asset to any company.

Now that we understand what a data scientist’s ability and expertise are, we can better understand what he actually does. Data scientist use the advanced technology to increase awareness, simplicity and direction for those who are working with the data. The data scientist must keep into account the rapid changes that take place in our modern-day environment. It’s the job of a data scientist to find patterns in large amounts of data and connect them to real world decisions.

These data scientist are the ones to help us keep up with the fast pace world of data. In the past few years there has been a huge explosion with the amount of data that is available to us. By now, pretty much everything is data related. The internet is full of data driven apps, almost any e-commerce application is data-driven, and there is a database behind any web front end. The world of data science continues to amaze us all the time and will continue to be a great benefit to our business and our personal lives.

Data Analytics (cont. from page 1)

applications are used to process and analyze the data to find a pattern or come to a conclusion that will help the business resolve their question.

There are many types of data analytics which fall into one of two categories: reactive or proactive. Reactive analytics provides business reports by looking at the past. This kind of analytics is more limited than proactive analytics, which is looking towards the future and making decisions, predictions, and optimizations.

A few types of data analytics are EDA, CDA, and QDA. Exploratory data analysis (EDA) is used to discover new patterns and information in the data. Confirmatory data analysis (CDA) proves existing hypotheses to be true or not. Qualitative data analysis (QDA) is used mostly in social sciences and uses non-numerical data like pictures, words, and videos.

Data analytics also plays a large part in our daily lives. So much can be predicted and so many decisions can be made through data analysis. For example, when used in the health fields, data analysis can decode DNA quickly and figure out which gene is responsible for a disease. When used by a business, it can predict which ads internet users will click on when they are online. Banks and credit card companies use data analysis to prevent fraud. For example, a few years ago, Citigroup hired "Watson", an IBM computer which is famous for beating champion contestants on "Jeopardy". Watson uses data analysis to process unstructured data and help the company cut down on fraud.

Data analytics is an important tool which can be applied to virtually any field or business. As technology and internet usage further develop, it will become increasingly important in our world.¹

Data Mining

ARIELLA FEUCHTWANGER

What is Data Mining?

Data mining is the process of going through pieces of data (generally large amounts of it, for businesses or markets) to find patterns and/or relationships between them in order to predict what will happen in the future. The methods of finding these patterns come at the crossroads of artificial

intelligence, machine learning, statistics, and database systems.

Stages of Data Mining

There are three stages in the process of data mining: exploration, building a model/finding the pattern, and applying the model to future predictions. (cont.. on pg. 4)

¹ <http://searchdatamanagement.techtarget.com/definition/dataanalytics>
http://www.sas.com/en_us/insights/analytics/big-data-analytics.html
http://www.webopedia.com/TERM/B/big_data_analytics.html
<http://www.economist.com/node/21554743>

Data Mining (cont. from page 3)

- **Exploration:** This stage generally begins with preparing the data. It includes cleaning the data, transforming the data, selecting subsets of the data, and, in the case of large amounts of data, selecting information based on specific factors. This stage can involve anything from a simple selection of data to be used to make the predictions, to an elaborate exploration using graphs and statistical methods to identify which data is most relevant and to determine how complex and/or the nature of the models, which will be taken into account in the next stage.
- **Model Building and Validation:** This stage involves the consideration and selection of the best model for the data involved. While it may sound simple, it is occasionally a very intricate process. Many techniques have been developed to reach the goal of finding the perfect model, many of which are based on “competitive evaluation of models.” This method utilizes many models and compares their performance. The final model is the one which performs the best. These techniques are known as the “core of predictive data mining” and include:
 - **Bagging (Voting, Averaging)**
- This applies to predictive data mining, which combines the predictions from many models or from a single model for different data. An example of this is a tree classifier, which, in practice, will result in different trees for
 - **Boosting-** This technique involves developing a number of models and using the combined predictions from all of them to make a single prediction.
 - **Meta-Learning-** Similar to boosting, this technique utilizes the predictions of many models, specifically in cases which the models are very different. This is also referred to as Stacking, or Stacked Generalizations.
- **Deployment:** This final stage uses the model selected in the previous stage and applies it to new data to make predictions or approximations of a new outcome.

different samples. This shows how models can be unstable even with small data sets.

What is Data Mining Used For?

Data mining has been popularized in business, where it is used to reveal structures—which guide businessmen in making better business choices. While there was recently an interest in finding new analytic techniques designated for business issues, such as Classification Trees, Data Mining is still based on statistical principles, such as the Exploratory Data Analysis, the traditional method, and modeling. It can also be used in medicine by utilizing the electronic records, which leads to a conflict of legality. The United States Congress passed the Health Insurance Portability and Accountability Act, which requires patients to provide “informed consent” for the information they have provided and will provide in the future. (cont. on pg. 5)

Data Mining (cont. from page 4)

Data mining is a relatively new field, and was not previously of interest to statisticians. After applying it to business, however, this field has become an important area aiding in the discovery of major

theories. Despite the tangles of medical controversies due to privacy concerns, data mining has become an important aspect of research today.

Big Data (cont. from pg. 1)

now have the ability to analyze new and diverse data types. For example, a health insurance company has the ability to more accurately predict customer dissatisfaction by analyzing speech-to-text data from call center recordings. Another example would be a retail bank using this factor to get control on its multi-channel customer interactions by analyzing log files. By merging structured and unstructured data these companies are able to obtain a clearer picture of their customers and the companies' own operations.

The big data industry is expected to grow from US \$10.2 billion in 2013 to about US\$54.3 billion by 2017.

The goal of big data is not just to collect it. The data is there to be used in order to help businesses reduce costs and time, develop new products, and make smarter business decisions. UPS's project ORION, On-Road Integration Optimization and Navigation, tracks their 46,000 vehicles. In 2011 UPS saved more than 8.4 million gallons of fuel. By cutting off one mile daily per driver they believe they can save \$30 million. Macy's was able to reduce its time for pricing 73 million items for sale from 27 hours to 1 hour. This allows them to change prices more easily and therefore that can change them more frequently according to changes in the retail market. In addition, processing data in real time allows companies to send recommendations to mobile devices while customers are able to take advantage of them due to their proximity.

Big data is now exploding. Organizations in all industries are combining traditional analysis with big data. This is probably the biggest revolution in extracting value from data since the 1980's. It is important to realize that value of big data is only in what we can gain through its processing and analysis. The advances in management of data will lead to changes in skills, leadership, organizational structures, and technologies.²



"By folding in our HR analytics into our other data, we find the people we retain longest are young, college educated females with two grams of saturated fat, soy lecithin, and artificial vanilla flavors."

² StatSoft. (n.d.). What is Data Mining (Predictive Analytics, Big Data). Retrieved from Dell Software: <http://www.statsoft.com/Textbook/Data-Mining-Techniques>

U.S. Newswire. (2008, June 16). HIPAA Privacy Rule Impedes Biomedical Research. Retrieved from <http://search.proquest.com/docview/451304164?accountid=1437>

WHAT IS BIG DATA AND HOW IS IT USED IN THE FINANCE WORLD?

CHAYALA GLAZER

Big Data is the term used for data sets that get so big that the traditional data processing tools become inadequate. Big Data is used by many corporations and businesses especially in the finance world. In fact, there was a 97% increase in the use of big data in the last two years! The goal of big data is to help organizations get usable, real-time, and easy-to-understand insights to enable their organizations to thrive. Indeed, when big data is used in the best way, it not only has the potential to improve finance systems but also to create new businesses and industries.

In the finance world, big data is used in two major ways. Firstly, it is used to analyze investments, econometrics, trading activity and other data analysis. Secondly, it is used to understand customers, competitors, and their behavior. It uses many available sensors to create a computational and quantitative mindset to make sense of the busy world. In fact, LinkedIn, a large social network company, went public because of the data it captured to identify and build products. The company continues to use data science tools and techniques to build products that attract its clients.

There are three major types of data analytics in the field of finance: descriptive analytics, predictive analytics and prescriptive analytics. Descriptive analytics deal with understanding and analyzing historical data; they help companies understand past activity. Using a variety of

methods like data mining, machine learning, modeling and game theory, predictive analytics can estimate the likelihood of a future outcome. In finance, its applications often include predictions such as predicting the price of a given stock, determining the credit score of a customer, and determining a risk involved for an investment. Then, for example, once a person's credit score is determined, the company can determine how much the person will be able to pay in the future. Prescriptive analytics use complex algorithms like the Monte Carlo Simulation Model to inform the user about various possible steps that can be carried out and suggests the best one. The Monte Carlo Simulation Model is one of the most popular prescriptive models in finance and is used by many companies and institutions for financial planning, making decisions and limiting financial risks.

In short, data science in the finance industry can not only help create a very customer-driven enterprise, but it can also help in enhancing risk-management, making intelligent decisions and streamlining the operations for any financial institution. They also focus on many areas including fraud detection, customer segmentation and targeting, pricing securities and products, and competition analysis.³

With the world of big data getting larger each year, use of analytics is becoming a necessity in the financial services industry, and using it properly is the underlying differentiator between
(cont. on pg. 7)

³ <http://radar.oreilly.com/2011/08/finance-data-science-big-data.html>
https://datasciencecmu.wordpress.com/category/33_data-science-in-finance/

successful and unsuccessful firms. Additionally, as 71% of the banking and financial firms have reported, the use of analytics is creating good competition over other organizations. It's worth using professional analytics, because from 64% of companies that deploy Data Analytics projects, 56% of them struggle to get value

from their data. Don't let your organization fall behind! Use data analytics to gather information to help your organization thrive.

Retailers could increase their profit margins by more than 60% through the full exploitation of big data analytics

Hadoop- Make Way for the Yellow Elephant

ATARA NEY

Hadoop, which started off as a stuffed, yellow, toy elephant's name, has become the name of a successful software framework that stores and processes big data.

Hadoop History

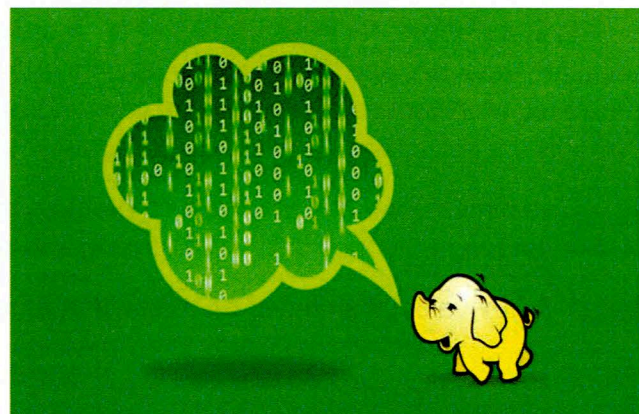
Hadoop's main creators were Doug Cutting and Mike Cafarella. The two originally headed a project called Nutch, which was supposed to be an open source web engine that not only used a web crawler but that also could return pertinent answers quickly. They planned on accomplishing this by distributing data and processes across a cluster of computers (this will be explained in greater detail below). In 2006, the Nutch project was split into two parts – the web crawler and the distributed computing. Cutting renamed the distributed computing portion after his son's stuffed elephant Hadoop, while the web crawler remained with the name Nutch. In 2008, Hadoop was released by Yahoo as an open source project. Hadoop's framework and companion utilities are now managed by the Apache Software Foundation.

How Hadoop Works in a Nutshell (probably a peanut shell)

Today we have access to a lot of data, more than we can physically deal with. Hadoop technology helps us manage this data in two major ways- 1) it stores the data and 2) it processes the data so it can be used. Of course you ask "What is so major about that - isn't that what databases are supposed to do?" But what do you do when you are dealing with an overwhelming amount of data? How can you possibly analyze all of it? Or, what if the data consists of different types of data, like complex and structured data, that all need to be processed in an organized fashion? And how long will this all take? Hadoop can store and process for a colossal amount of data, various types of data, and in a short amount of time.

Hadoop accomplishes this by using clusters of computers. Instead of relying on one computer's hard drive, Hadoop uses groups of computers. The Hadoop software takes incoming data and spreads it across the servers. A specific part of the software keeps track of the different

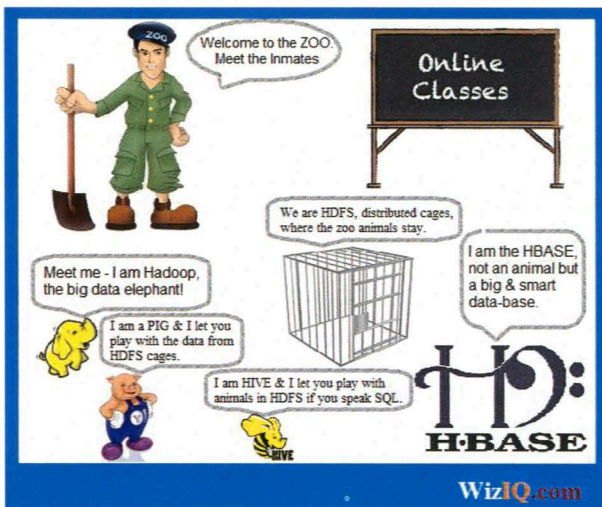
(cont. on pg. 8)



Hadoop (cont. from page 7)

locations of the data. The data is copied to several locations, so that, in the event a computer malfunctions, the data can still be accessed.

Besides the greater capacity, another advantage of using multiple computers is the increase in processing power and speed. Since each computer has a bunch of CPU's, the work can be divided up between them all, and multiple processes run at once.



Hadoop Components

The Hadoop project has different modules for running it. Below we'll discuss some of them.

Hadoop Distributed File System (HDFS) – This is the default distributed file system used to hold the data that is entered onto the clusters. It is the part of Hadoop that does the file storing. It itself is distributed amongst the computers, and it does the data replicating. As mentioned above, it stores many different types of data and does not need to be organized before being

processed. It is the default one because you could use a different distributed file system, such as Cassandra FS amongst others.

MapReduce – Since Hadoop is not a database, it cannot use the typical SQL (Structured Query Language) when analyzing and processing data. The MapReduce programming framework takes care of this. Like HDFS it runs on all the servers, which allows parallel processing. It doesn't use queries, since as we said it is not SQL. Instead, it processes requests using two components – JobTracker and TaskTrackers. JobTracker divides the computing jobs it is being told to do into defined jobs and shifts those defined pieces over to TaskTrackers that are on the same machine as the data. Once the job is run, the data is sent back. Since there are many clusters, data can be processed in parallel. However, MapReduce is a complicated language, causing making up tasks to be complex, and it is not ideal for all analytic tasks. Therefore, some other components were created in addition to it like Pig, Hive and other analytical components.

Pig – Pig is a platform for manipulating data. It specifically uses a language called Pig Latin to manipulate, analyze, and create new data sets with the data in HDFS. It bypasses the complicated MapReduce commands, because it consists of a compiler for MapReduce. You technically are still calling MapReduce jobs, but you are using a simpler language to do it.

Hive – Hive circumvents the MapReduce problem by using a language called QL which is a SQL-like language. It does data warehousing and presents data in the form of tables.

HBase – A key-value, distributed database that runs on top of Hadoop. (cont. on pg. 9)

Hadoop (cont. from page 8)

Zookeeper – An application that takes care of all the distributed processes.

Mahout – A data mining library.

There are many other components that are used in sync with Hadoop. However, due to limited time, we will pause here.

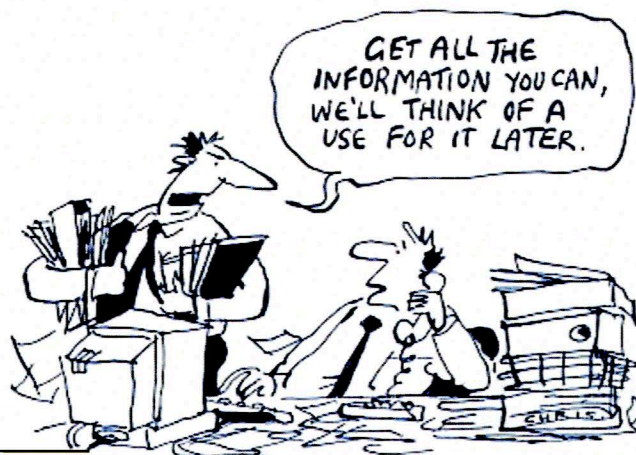
So Who Uses Hadoop?

Many companies will find themselves using Hadoop for a number of reasons. They could have limited storage space or need to constantly backup their data. Hadoop's computer clusters which self-replicate help avoid these problems. Hadoop is also fast working when it comes to computations because of its parallel processing feature. It also organizes large sets of data. For these reasons, markets such as finance, online retail, and even traffic management use Hadoop. Companies such as Amazon, Netflix, and Microsoft find Hadoop indispensable. It has even been theorized that Hadoop may take the place of the EDW (Enterprise Data Warehouse).

Hadoop is definitely something programmers should look into, especially if you know Java. For starters, Hadoop is written in Java, so it is easier to learn and work with. With so many companies using Hadoop, it will be a bonus on your resume if you show you know Hadoop. "According to Dice's Open Web, a portal that specializes in analyzing the hiring trends, "Java Hadoop" combination is the most sought after professional skill in the IT industry."

All in all, you might want to make some room in your mind for this mighty computing, yellow elephant.⁴

The value of the Hadoop market is expected to soar from \$2 billion in 2013 to \$50 billion by 2020, according to market research firm Allied Market Research.



⁴http://www.sas.com/en_us/insights/big-data/hadoop.html#

<http://hadoop.apache.org/#What+Is+Apache+Hadoop%3F>

<http://radar.oreilly.com/2011/01/what-is-hadoop.html>

<http://readwrite.com/2013/05/23/hadoop-what-it-is-and-how-it-works>

<http://www.orzota.com/hadoop-as-the-source-of-truth/>

<http://www.dezyre.com/article/5-reasons-why-java-professionals-should-learn-hadoop/75#.VQH08HnwsdU>

SAS

AHUVA NEUHOFF

Not to be confused with the Scandinavian Air System abbreviation, "SAS" in the computer world refers to the Statistical Analytical System software. This software is an integrated system which can mine, alter, manage and retrieve data from a variety of sources and perform statistical analysis on it. The suite was originally developed by the SAS Institute to be used for advanced analytics, business intelligence, and data management. Data used for SAS software must be in excel or database format. The software has the capacity to link and set up formulas connecting data variables, and has the capacity to store thousands of data entries which allows it to be exceedingly more efficient than Microsoft Access. In addition, a unique function of SAS is its ability to project analysis. Based on the information garnered from thousands of data values, SAS creates a formula that will connect the data and even project a future outcome. Voting polls are calculated with the help of SAS software. Eligible voters are linked to state, party, age, nationality, and hundreds of other such variables, and based on past voting trends of these people, SAS will generate a predictive analysis of what their party leanings will be in the coming election.

SAS is a major contributing factor to the success of many industries, including banking, consumer products, government, health care, education, and weather. HSBC has used SAS to reduce incidences of fraud in credit card companies, lower costs, and increase customer confidence. Nestle takes advantage of SAS's planning product to ensure the right amount of product reaches the retailers' shelves and the consumer's

hands. How do you suppose the analysis incorporated in the US census bureau is achieved? SAS supplies the bureau with daily progress reports, budget updates, and ad hoc analysis through the information gathered from the decennial census maintained in their business analytics product. Perhaps the most significant achievement of the software suite is the advancements that it has provided for the medicine world. Not only does SAS aid insurance organizations, but it is also used by many researchers in the field. Researchers gather information on an individual's age, health, weight, and other medical influences, and with the formulas provided through the analytical projections, remedial conclusions are deduced, such as a linkage of tobacco inhalation to cancerous activity in the body. Meteorologists gain from SAS's predictive function, as they use it to predict weather forecasts, although they are not always 100% accurate. Even SAS itself takes advantage of its software to engage with data more easily, eliminate the need to use IT resources, and to find errors and anomalies that cost time and money.

SAS software has contributed to the industrial productivity. Without SAS software, employees would be compelled to spend extra hours mentally analyzing data, whereas with the software, all that was required was reduced to 15 – 20 minutes or less. The efficiency of the workplace has been greatly affected, and the software is extremely simplistic to use. As one pleased SAS consumer said, "You don't have to be a programmer. You don't have to understand the code. The whole process is simple, uncomplicated, and seamless."⁵

⁵ www.sas.com

Data Journalism

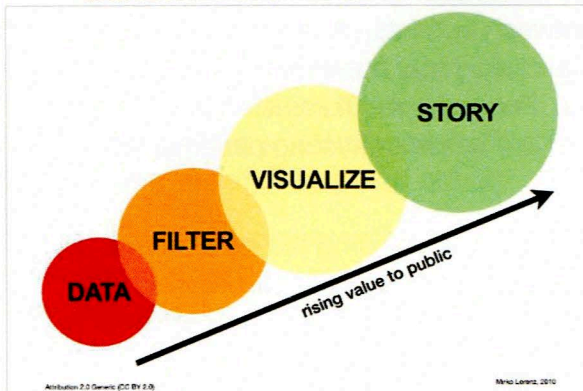
RENA KERSTEIN

What is Data Journalism?

Both 'data' and 'journalism' can be troublesome terms. Some people think of 'data' as any collection of numbers, most likely gathered on a spreadsheet. Twenty years ago, that was pretty much the only sort of data that journalists dealt with. However, we now live in a digital world, a world in which almost anything can be — and almost everything is — described in numbers.

Data journalism is a journalism specialty reflecting the increased role that numerical data is used in the production and distribution of information in the digital era. It reflects the increased interaction between content producers and several other fields such as design, computer science, and statistics. From a journalist point of view, it represents "an overlapping set of competencies drawn from disparate fields." Data journalism has been widely used to unite several concepts and link them to journalism. Some see these as stages leading from the simpler uses of new technologies to more complex ones in the journalistic process.

DATA-DRIVEN JOURNALISM = PROCESS



Why is Data Journalism important?

Data Journalism filters the flow of data. When information was scarce, most of the efforts of journalists were devoted to hunting and gathering. Now that information is abundant, processing is more important. Data is processed at two levels: (1) analysis to bring sense and structure out of the never-ending flow of data and (2) presentation to get the important and relevant information into the consumer's hands. Like science, data journalism discloses its methods and presents its findings in a way that can be verified by replication.

Data journalism is a new set of skills for searching, understanding and visualizing digital sources in a time that basic skills from traditional journalism are simply not enough. -It is not a replacement of traditional journalism, but an addition to it. In a time where sources go digital, journalists can be and have to be closer to those sources. The Internet opened up possibilities beyond our current understanding.

"Data journalism is 80% perspiration, 10% great idea, 10% output." The Journalists spend hours making datasets work, reformatting pdfs, and mashing datasets together. We can see from this how much they go through before they get the data out to the public. Mostly, they act as the bridge between the data (and those who are pretty much hopeless at explaining it) and the people out there in the real world who want to understand what that story is really about.

Data-driven journalism is the future. Journalists need to be data-savvy. Stories used to be obtained through chatting to people in bars and such. That still may be done, however, now it is also be about poring over data and equipping oneself with the tools to analyze it and pick out what is interesting and relevant. We have to let the world know this so that we can be fully equipped with the tools to handle this future data-driven journalism.⁶

⁶ <https://datajournalismhandbook.org/>

Data Science in the Health Care Industry

MALKIE GOLUB

In 2002 Google AdWorks revolutionized the internet by introducing the “cost per click” (CPC) model. This CPC model outperformed the “cost per thousand impressions” (CPM) model because Google was able to bill based on the success of a link by tracking each time a user clicked on it. The collected data Google now stored about each user, enabled them to manipulate the success of each link by using targeted ads that the user was likely to click. Since then, predictive analytics have studied user behavior using data from large companies like Google, Twitter, Facebook, and LinkedIn and were able to transform many industries.

In the medical industry when doctors prescribe a plan of treatment, they follow a standard protocol. These treatments certainly have a success rate, whether it be 60%, 80%, or 95% effective. However, doctors cannot really predict which 60%, 80%, or 95% of the patients the medication will work on because the relationship between treatments and outputs can be difficult to predict. Treatments are considered effective or ineffective based on double-blind studies that rarely take into account the differences between patients. A drug that is effective on only 1% of patients would be rejected by any traditional clinical trial however, it would be a valuable treatment if we can determine who that 1% is. For example, Tomaxifen used to treat breast cancer patients was considered 80% effective. Now by differentiating patients, we can make it 100% effective on 80% of the patients.

Big data can help determine for which patients any particular drug will be effective. By collecting enough data about medical treatments from clinical studies, hospital records, insurance data, and doctor’s offices, we will be able to predict accurately which treatments will be effective for which patients. In addition, we also have access to biological data, clinical data and health outcomes data in electronic health records (EHRs) and by combining all this data too, we can observe which treatments work for which people. Linking all electronic health records so the data is immediately stored in a data center isn’t an unrealistic dream; it’s technology we have now. Also included, can be the health social networks that have popped up over the internet like PatientsLikeMe, ACOR, and 23andMe. These health social networks have thousands of users who generate invaluable pools of data. This can create an ongoing evaluation of whether a treatment is effective, and under what conditions. We can store which drugs combinations had synergistic effects and whether there was a good outcome.

This will not only save patients from undergoing treatments unnecessarily but also save billions of dollars annually. The United States spends over \$2.6 trillion on health care every year. Over \$600 billion are spent on treatments that don’t work or cause iatrogenic side effects. We are on the verge of a newer and smarter system where treatments will be administered effectively, based on each individual’s personal biology.⁷

⁷ How Data Science is Transforming Health Care, *Solving the Wanamaker Dilemma*. Tim O’Reilly, Julie Steele, Mike Loukides, Colin Hill

Hour of Code

RACHEL AZIZA BENSOUSSAN

A global movement. A worldwide effort. An international project.

The Hour of Code.

December 8-14, 2014 was a momentous week for the world of computer science. Over the course of that week, over one hundred million people, from 77,000 different locations, learned one hour of code in a movement fittingly known as the Hour of Code (HOC).

I first heard of the HOC from a senior in Touro College who sent out an email to all computer science majors informing us of the upcoming event. Various Bais Yaakov schools around Brooklyn were willing – and eager – to have a computer science student instruct the HOC in their school. And so I signed up. There were several resources I used to plan my lesson. These include videos available online introducing the HOC and links to online tutorials on the HOC website that would demonstrate to the students what coding is all about.

I began the class with the introductory video, in which teachers and students who had participated in a previous HOC discussed their experience. Many well-known individuals– both in the computer world and out – were also featured speaking about this international event. Seeing President Obama, the CEO of YouTube, and the CEO of Microsoft on the screen made the students feel that they were part of a bigger whole, that the class they were about to have was no ordinary class; rather, it was an introduction to a vast world they didn't know existed.

I discovered, within the first 5 minutes of the class, that 98% of high school students don't understand what a computer program is, how it works, or what a computer programmer does. So I explained; I taught them that a program is a series of instructions given to the computer to help it accomplish a specific task. Programmers write the instructions in a language that the computer understands, in code, to tell the computer what to do – and they pray that it works successfully! I told them what sort of skills are needed for the job: A logical mind, good problem solving skills, a mathematical inclination, and a whole lot of patience. We discussed the hours, and the option of working from home. (cont. on pg. 14)

What is the Hour of Code?

- 1** **A self-guided activity for students of all ages**, kindergarten and up. Tutorials will work on a browser, tablet, smartphone, or even with no computer at all.
- 2** **No experience needed** from teachers and students
- 3** **Tutorials will feature** lectures from Mark Zuckerberg and Bill Gates, and artwork from popular games Angry Birds and Plants vs. Zombies.
- 4** **An effort to demystify code**, and change the conversation around computer science.



AFTER THE SUCCESS WITH MARKETING ANALYTICS ON HER WEBSITE, JANE HOPED TO LEARN SOMETHING BY ATTACHING COOKIES TO CUSTOMERS WHO VISITED HER STORE.

Fun Facts

- ❖ Every 2 days we create as much information as we did from the beginning of time until 2003.
- ❖ Over 90% of all the data in the world was created in the past 2 years.
- ❖ It is expected that by 2020 the amount of digital information in existence will have grown from 3.2 zettabytes today to 40 zettabytes.
- ❖ The total amount of data being captured and stored by industry doubles every 1.2 years.
- ❖ Every minute we send 204 million emails, generate 1.8 million Facebook likes, send 278 thousand Tweets, and upload 200 thousand photos to Facebook.

Hour of Code (cont. from page 13)

After we clarified the basics, we went on to do an online tutorial together. The tutorial was a simplistic GPS simulation, in which the students had to give instructions to the little cartoon man on the screen in order to direct him towards his destination. The first two or three levels were simple: go straight, turn left, go straight; or, go straight until endpoint; or, go straight, turn right, go straight until the endpoint. All the students were involved; they yelled out instructions for me to put on the screen, and they argued with one another as they noticed flaws in the other's logic.

At level 4, "if" statements were introduced, and the students were pushed to think in a way they rarely do. They had to instruct the little man to go straight for a while, but turn if there was a turn available, and keep on going until the endpoint. After that level, the man had two directions to turn to. As the levels progressed, the instructions became more and more complex. The girls thought out loud, argued some more, and helped each other reach a conclusion to get the man to his endpoint. They were stimulated, they were alive, and they were having a total blast.

When we arrived at level 10, the bell rang, announcing the end of the class. I told the girls where to find more tutorials online, should they want to try it out at home. They walked away energized and simply amazed at how much work goes in to telling a little man to move across a screen, stating emphatically how complicated it is to follow the logic of a "nested if" statement.

Just wait until they hear about Objects.

Job Search: Tips for Success

PROFESSOR B. SHEINFELD

Searching for a job is a job in its own right. You should invest appropriate time, effort, and focus as you would for a real job. Your skills are the primary factor in getting a job—but there are many tangential items that can lose the job for you (on your own or in comparison to other candidates). My objectives here are to make sure that that you not only present yourself in the most optimal light, but that you also don't sabotage yourself.

Much of this is common sense— This column is based on what I have found to be helpful from being both an interviewee and an interviewer as a Sr. Executive in the business/financial world for over 20 years. I've broken down the tasks into several steps – from Resume through Job Offer- in order to focus on optimum performance in each area.

STEP 1- YOUR RESUME:

This is important because it is your first impression to your potential employer.

- Identify your accomplishments- List accomplishments that are pertinent to your professional goals.
- Know your audience and keep your accomplishments appropriate for that audience.
- Your resume will probably get only a 1 minute scan (if that); it must make as much impact as possible.

Things to include:

- ✓ Use lots of key words- for searching online and for underlining by the reviewer
- ✓ Use **power words**—while the initial reviewer will rarely initially read the whole resume from beginning to end (they just search for key words)—this will be important when the resume is handed off to a hiring manager and the manager will want to make sure that the applicant can write/speak professionally.
- ✓ **Be truthful**- do not exaggerate, but make sure that you give yourself credit for your accomplishments- the job market is not the place for extreme modesty.

Make it obvious that you have something to offer the company.

Things to leave out:

- × **Objectives**- Who cares what you want? At this point you want a job. (After you get an offer you can decide if it meets your objectives, but don't put yourself in a position where you eliminate any opportunities before they even exist)
- × **Personal**- Who really cares what you do in your spare time?
- × Anything that is **religious** in nature
- A resume must look professional- Use a nice, professional font (sans serif is preferred, lately). For the paper version, good quality paper and printing is essential. (If your printer isn't good, then go to a "Kinkos" to have it done).
- Don't do anything that is strange/out of the ordinary- that is not the type of attention that you want.
- You may need several different versions to have your resume tailored for a specific position.
- Make sure to check your spelling and grammar.
- Include:
 - Your name and address

- Keep in mind how others will react to your name. Do not write something “cutesy”.
- Do not write something that may confuse the reader. There are many different nationalities in the marketplace, so people are starting to get used to foreign-sounding names. Nevertheless, there are still prejudices in the workplace, and people may be uncomfortable with names that they can’t pronounce or that may present an image of someone that they may not feel comfortable working with.
- An email address
 - Make sure it is an appropriate email address. You may want to have an account just for job search or job-related correspondence.

Digital considerations:

Remember that we are living in a digital world. It is important that you demonstrate that you are current with technology.

- You will most likely be googled by a potential employer—to check your information to find out more about yourself.
- Make sure that you are on professional sites like LinkedIn.
- Make sure that there is nothing on Facebook (or other sites) that will reflect negatively on you.
- Make sure that all information about you on the web is truthful and consistent with your resume.
- If necessary, you should work on correcting any negative information.

- You should be able to correct, defend, or repair any negative or inconsistent information if you are asked about it.

STEP 2- SENDING YOUR RESUME

Directly- Companies are looking to save agent fees, so they first look for resumes submitted directly to them. Research companies where you want to work and submit your resume directly to Human Resources.

- Statistics show that if you are recommended by an employee of that company, you have a much better chance of an interview and a job offer.
- If that employee gives your resume directly to the hiring manager, your chances are even better.
- Send your resume with a cover letter or cover email. You may need different versions to fine tune it for each position.
- Job-search websites- these job-search sites are also great sources of job-search information
 - Dice.com
 - Monster.com, Careerbuilder.com, etc.
 - LinkedIn.com (make sure you are on LinkedIn.com)
 - TheLadders.com (also a great source of information/advice)
 - Craigslist.com

Who can help you:

- Headhunters/Agents- It may be difficult to find an agent as an entry level applicant, but you still should try.
- Anyone you know may be a contact- “Don’t leave home without it” (your resume).

STEP 3- BEFORE YOUR INTERVIEW

Do your homework

- On the company- its products, news, job requirements, etc. There is a lot of public information. (Google, Vault.com, the company's corporate site)
- On the interviewers- Facebook, LinkedIn, Google
- Keep up to date with current events- (Bloomberg west Cable TV, CNET, ZDNET, etc.)
- REMEMBER- Information on the internet is a 2-way street. Just as you will research your potential employer, they will research you. Be careful of what can be found out about you on the Internet.
- Have an Elevator Speech- and practice saying it.
- Never turn down any interview:
 - It is good practice
 - There may be a spot available in the future
 - You may be recommended for another, more suitable spot

What to wear?

It won't get you the job but it can make you lose getting a job offer

- Dress conservatively- appropriate to the business. Wall Street may be more conservative than small companies, startups, and advertising.
- Even if the dress code is business casual, for an interview you may still want to wear a suit (or at least a jacket- no dresses) since it conveys the respect and importance you place on this interview.
- Whatever the standard, lean toward the conservative- conservative clothing, shoes, jewelry. Do not wear cutesy or religious jewelry. Wear

conservative makeup, and very light perfume (if any).

- You may want to ask the headhunter or the HR contact. 'I want to make sure I understand your company's culture and dress appropriately.'
- Remember- You want to be seen as a professional and someone who they will want on their team. Although an attractive person will always be more highly regarded than an unattractive person, the goal is to get a job based on your skills.
- Bring samples of your work with you
- Bring a few extra copies of your resume
- Prepare for questions- Do some research- there are many sites with suggested questions and answers.
- Think of technical questions you will be asked.
- Questions that you may be asked but shouldn't:
 - Religion (e.g., Shabbos)
 - Female (e.g., family plans)
 - Country of origin (they do need to make sure you are allowed to work in the US, but it is illegal to ask- in fact, most large companies have training for managers who interview.) However, they will often ask anyhow, or sneak a similar question into the interview.
- Trick questions- "What is your worst fault"- never say anything negative about yourself.

Write down what you will say and practice saying it

STEP 4- AT YOUR INTERVIEW

- Turn off your cell phone.
- Have a phrase to get you going- for example, "Its show time!" puts it in perspective for me. This is important, and I must put my best foot forward. It takes a bit of showmanship, takes you out of your comfort zone, and gives the interview a bit of excitement and enthusiasm. Come up with a phrase that works for you.
- When you walk into the interview, it helps to put yourself in the head of the interviewer.
 - Remember, the purpose of the interview is for the interviewer wants to find out what you can do for the company, and if you are right for this job.
 - Keep in mind that hiring staff is one of the most important things a manager

Fun Facts

- ❖ Google alone processes on average over 40 thousand search queries per second, making it over 3.5 billion in a single day.
- ❖ Around 100 hours of video are uploaded to YouTube every minute and it would take you around 15 years to watch every video uploaded by users in one day.
- ❖ Facebook users share 30 billion pieces of content between them every day.
- ❖ If you burned all of the data created in just one day onto DVDs, you could stack them on top of each other and reach the moon – twice.

does- it impacts the delivery and success of their projects; a good hire is very positive. A bad hire negatively impacts success of the company's projects, and reflects badly on the manager.

- The interviewer also knows that someone else, probably their boss, will do the follow-up interview. Sending a bad candidate to the boss is embarrassing.
- Most interviews are straightforward, but be aware of stress interviews, multi-person interviews, and "nice-guy" interviews.
- There are some "bad apples" who don't play fair, or others who are afraid of interviewing. The best you can do is react professionally and gracefully.
- Don't waste time and effort trying to "read" your interviewer; they are not your friends. Interviewers have more experience at this than you; don't be lulled into comfort or be frightened by them.
- Arrive 5-10 minutes early- never arrive late, but too early isn't good either.
- Be nice, friendly, and polite to everyone. This means the secretary, receptionist, other employees, etc. too!
- Smile! Be enthusiastic!
- Don't act timid, nervous, or arrogant.
- Handshake- you need to think about how to handle this in advance.
- Sit comfortably in your chair- not too close, not too far, and don't slouch.
- Mirror your interviewer- sit the same way, and place your hands similarly.
- If your resume was submitted by a headhunter, ask to see it- it may have been re-written and have incorrect information, typos, etc. Make sure you give the interviewer a copy of your

resume which has your address, phone number, and email address (the headhunter resume will leave this out).

- Listen to the questions. If necessary, rephrase the question or ask them to repeat it. Think through your answers
- **'For example'** are the two most important words in your interview.
- "Off the record" does not exist. This is not a chat between friends. This is a very serious, professional discussion, and "anything you say can and will be held against you."
- Always present the best possible image that is consistent with the truth.
- Questions regarding: Shabbos, family, etc. Remember, you need to think about these in advance and have a good, rehearsed answer.
- You will be asked if you have any questions- have one going in. Any questions that you ask should be aimed at how you can contribute, not at what you want.
- Never discuss salary- you can only lose. I'm most interested in the opportunity, and I'm sure that we can come to an agreement within the salary range for this job.
- Don't say anything negative about anyone. There is always the thought that you will do the same to the interviewer if you are hired.

STEP 5- AFTER YOUR INTERVIEW

- Write down all the questions you were asked and what you answered. Think about how you would improve your answer the next time you are asked this question, since you will probably be asked this question again by someone else.

- Write down the names of everyone you spoke to- including interviewers, secretaries, and other staff.
- Write down anything pertinent to the interview.

All this will be helpful if you have a follow-up interview

- While waiting (or on your way out) look around. See other employees, who they are- racial/gender/religious, what they wear. Ask yourself: 'does it look like a nice place to work, do people look happy.' Note how the employees are interacting with one other.

WRITE A THANK YOU NOTE (1-5 from Careerbuilder.com/TheLadders.com)

1. A thank-you letter creates an opportunity to re-connect with employers.

Chances are you aren't the only candidate being interviewed for an open position. Writing a follow-up letter allows you to build a relationship with the interviewer and develop rapport. By expressing your gratitude for the interview and recapping the highlights of the meeting, you revisit the reasons why you are the best fit for the position.

2. Following up keeps your candidacy top of mind.

A candidate often makes the mistake of putting too much control in the interviewer's hands. He believes that if he's the best candidate, the interviewer will remember him, and keep him in the loop regarding the selection process. Unfortunately, this doesn't often happen. It's critical that a candidate remind prospective employers of his interest in a position. The thank-you letter is the perfect vehicle for communicating this interest.

3. Written correspondence gives you another chance to sell your strengths.

While the thank-you letter expresses gratitude for the meeting, it also serves a much more strategic purpose. It provides an opportunity for the candidate to present her skills and accomplishments in another format and market the value she'll add to the employer.

4. The document enables you to address points you neglected to discuss during the interview.

Many candidates, after leaving the interview, think of other things they could have said during the meeting. Don't label this a liability; turn it into an asset by discussing these points in the thank-you letter. Remind the reader of your ability to produce similar results for their organization.

5. A letter helps develop rapport and increases the employer's comfort level in your candidacy.

A good strategy is to recap a part of the conversation where you and the interviewer shared similar views on a job-related topic. The thank-you letter can also be a forum for demonstrating your consultative problem-solving skills. By addressing current issues the employer is facing and proposing solutions, you are contributing to the company's success even before you are onboard.

6. It demonstrates that you know how to act in a business environment

STEP 6- THE JOB OFFER

- Consider all offers.
- You can use one offer leverage for other offers.
- Be gracious to companies that you reject. You may want to work for the turned down company in the future.

HATZLACHA!

Fun Facts

- ❖ AT&T is thought to hold the world's largest volume of data in one unique database – its phone records database is 312 terabytes in size, and contains almost 2 trillion rows.
- ❖ 570 new websites spring into existence every minute of every day.
- ❖ 1.9 million IT jobs will be created in the US by 2015 to carry out big data projects. Each of those will be supported by 3 new jobs created outside of IT – meaning a total of 6 million new jobs thanks to big data.
- ❖ Today's data centres occupy an area of land equal in size to almost 6,000 football fields.
- ❖ Between them, companies monitoring Twitter to measure “sentiment” analyze 12 terabytes of tweets every day.
- ❖ The amount of data transferred over mobile networks increased by 81% to 1.5 exabytes (1.5 billion gigabytes) per month between 2012 and 2014. Video accounts for 53% of that total.

PUBLICATION
COMPUTING MACHINERY
ASSOCIATION FOR
OF THE
A STUDENT CHAPTER

TOURO COLLEGE

