A FROST & SULLIVAN EXECUTIVE SUMMARY

Don't Fear Big Data. Understand It. Manage It.
Put It to Work for YOU

In Collaboration With:

FROST & SULLIVAN

SCOTTISH DEVELOPMENT INTERNATIONAL

THE UNIVERSITY of EDINBURGH informatics

Autonomy an HP company

View the onDemand version of the eBroadcast: **www.frost.com/bigdata**

Data is everywhere. Data is growing by the day, overwhelming databases and related systems, yet while some may see it as a nightmare, Big Data may be the key to unlocking incredible business potential. There is immense value for those who are able to extract information from data in a manner that enables them to engage with customers more effectively, to drive optimal business outcomes, and to mitigate risk.

Stratecast | Frost & Sullivan and Scottish Development International presented this eBroadcast in order to transcend the market buzzwords and hype that surround Big Data, to effectively define it, to discuss its structures (and its challenging lack of structure), and to show how Big Data meshes with analytics and Stratecast's Business Intelligence model as well as how researchers in Scotland are creating new solutions for harnessing the potential of Big Data.

In short: Don't fear Big Data. Understand it. Manage it, and put it to work for you.

What Is Big Data and How Did We Get Here?

The Big Data issue was inevitable. From the first databases in the 1960s, through the popularization of the internet, to the rise of the first internet service providers (ISPs) in the 1980s and '90s, through the desktop computing boom, and now to the increasing adoption of smartphones, tablets, and other mobile devices—Big Data is growing more complex with each passing year.

To cut through the mass of nebulous definitions in the market for Big Data, Frost & Sullivan defines it as: Millions to billions of records that are physically housed in a single database, utilizing the grid computing model across the database. Big Data can be physical (on-site), virtual (the Cloud), or both.

Big Data can be as small as 1 Binary Digit to as large as 1 Geopbyte (equal to 1 quintillion Gigabytes), which is the largest measurement of storage. The typical discussion of Big Data today refers to the Yottabyte (equal to 1 quadrillion Gigabytes).

"Another definition for Big Data is that if your enterprise requires 100 Terabytes of storage or more—you're dealing with the Big Data issue," said Jeff Cotrupe, Global Program Director, Stratecast | Frost & Sullivan.

All of this data can be broken into Structured and Unstructured Data.

Structured Data | Relational Databases (RDBs)

- Information that exists in a database: columns and rows
- Phone numbers or ZIP Codes
- Readily access via Structured Query Language (SQL) requests

Unstructured Data

- No identifiable or consistent structure
- Not easily captured in structured databases
- Examples of unstructured data include:
  - Web content including both site and social
  - Email
  - Business productivity/word processing documents
  - Images and other electronic objects
  - Books
  - Multimedia/rich media: audio and video

"Unstructured data is what really makes up what everyone is talking about with Big Data. There is no identifiable or consistent structure. There are a lot of examples of it, most of it coming from the web, but not all of it," said Cotrupe.

Utilizing NoSQL Databases to Sort Unstructured Data

While structured data is growing, unstructured data is exploding. The problem with Big (Unstructured) Data, however, is in determining how to handle it. The solution lies in augmenting RDBs with NoSQL databases, said Frost & Sullivan's Cotrupe. NoSQL databases are:

- Next-generation, non-relational, distributed, open-source databases
- Scalable to handle massive volume of unstructured data generated on the web
- Simplified application programming interfaces (APIs) so users can quickly modify them to meet their needs

NoSQL databases were developed as companies like Google and Amazon needed new tools to deal with the volumes of data that they were creating. Many other companies have since stood on their shoulders over the years (such as Facebook) to build other versions and other types of NoSQL databases.

These sites also create an excess of Big Data through saved searches, shopping carts, likes/comments/feedback, and on-the-fly personalization. Many of these items do not need to be recorded to a traditional database. This is where Business Intelligence enters the equation.

The Massive Potentials of Business Intelligence and Big Data

In 1958, IBM researcher Hans Peter Luhn defined Business Intelligence (BI) as: "The ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal."

It has taken nearly fifty years for much of this definition to be realized. Today, there is a massive market opportunity in BI. Online analytics in 2011 were a $1.369 billion market with a 22 percent CAGR. Frost & Sullivan expects the market to grow to $2.5 billion by 2014. Video Quality of Experience (QoE), meanwhile was a $900 million market in 2012 with a 12 percent CAGR that Frost & Sullivan forecasts to reach $1.25 billion by 2015.

But online analytics and video QoE are only small pieces of the larger puzzle. "The overall BI and analytics market is in the tens of billions," said Cotrupe.

Scotland, in particular, is at the forefront of some substantial developments in the understanding and use of Big Data. "I'm very excited about the opportunities for software to make this all come together for you with Big Data and to be doing so in Scotland," said Cotrupe.

The Scottish Informatics & Computer Science Alliance (SICSA)

The Scottish Informatics & Computer Science Alliance (SICSA) (http://www.sicsa.ac.uk) is at the forefront of the Big Data issue. SICSA is a research pool consisting of 14 Scottish Universities, Informatics, and CS Departments that is funded by the Scottish Funding Council. SICSA includes the Edinburgh University School of Informatics, which is the largest in Europe (in the informatics and computer science space).

"We have a large number of academic institutions all within a relatively small geographic area. And, within this area, Scotland tends to over perform," said Colin Adams, Director of Commercialisation, School of Informatics, the University of Edinburgh. "We have twice the number of world-rated scholars you would expect for the size of the country we are, twice as many research grants, and we've got twice as many PhD students as we would normally have in this type of a country population size."

Approximately 2,500 people graduate from the universities that form SICSA each year. There are more than 80 full-time researchers at SICSA at any given time, and the organization has produced a large number of startups and spinouts.

Scotland's Strength in Big Data

Life Sciences Companies	Analytics Companies	Analytics (cont'd)
eCAT	Memex (SAS Company)	Think Analytics
Formedix	Sumerian	Agile Knowledge Management
Blackford Analysis	Winterwell Associates	VisionWare
Kelvin Connect	SiteTagger	Social Artisan
Aridhia	Vixo	DataPa
CliniSys Group	Ciqual	Swirrl
Fios Genomics	Data2Text	Actual Analytics

Source: Scottish Enterprise

SICSA offers a single point of contact to allow enterprises to "get some sort of traction on what their major skill needs are so the local academic grouping can react to that," said Adams. Additionally SICSA has created tailored Masters courses to examine topics such as:

- Design informatics, bringing design and the technical area together
- Cybersecurity making sure there is a skilled group going into the workforce from academia

SICSA's organizational structure creates one touch-point rather than, say, fourteen different touch-points at 14 different universities.

SICSA's Research Into the Big Data Issue

The universities that comprise SICSA are each involved in different areas of research that fall under the Big Data umbrella. These include:

Smart Data

- Machine Learning in many application areas
  - Satellite tracks from Astrophysical images
  - Monitoring babies in incubators
  - Compiler Optimization
- Information retrieval
  - A number of optimized search techniques/applications
  - Structured/unstructured Data
- Language understanding/Text Mining
- Text Search and video annotation
  - Several applications in Social Media

Edinburgh, Scotland is also one of the main areas of the world doing research into automated translation—a major solution to sorting through Big Data. One such example is the Moses statistical machine translation system, which is open-sourced and used by the U.N.

Other agencies use it as the basis for their machine translation approaches, not just for text but also for image analysis (e.g., closed circuit TV) or for archiving large video archives (e.g., BBC) to zero in on particular areas of content and put metadata down on exactly what is going on in particular elements.

Is Big Data a Minefield or a Goldmine?

When considering the Big Data issue enterprises need to ask: Is Big Data a minefield or a goldmine? To Dr. Richard Brierton, Senior Lead EMEA Marketing, HP Autonomy Software Management, HP Software, the answer is that Big Data is probably both an asset and a liability. "There's a huge amount of business value and risk but the only way to uncover that business value or risk is to understand the underlying information," he said.

Autonomy enables enterprises to understand all of their Big Data. "We power business operations through enterprise search, business process management automation. We protect enterprises from falling foul of complex laws and regulations around data handling so by understanding the meaning they are able to implement effective information governance policies to apply accurate data and retention and destruction policies," said Brierton.

Autonomy's solutions and toolkits are provided to over 65,000 customers globally. Its technology is embedded in over 400 products through OEM licensing. Meanwhile, ten out of the top twelve banks use Autonomy technology to power their businesses, and eight out of the ten top global firms also use it to protect and discover information. Its software is used to power over 20,000 websites worldwide.

Return on INFORMATION

Organizations measure value through their return on investment (ROI). But with Big Data, ROI is developing a new meaning—Return on Information. All CIOs, CMOs, CFOs, and line of business executives are judged on their ability to meet this new metric. So how does an executive typically measure Return on Information? "They look at the value of the information exposed to the enterprise and they look at the cost of extracting that value," said Brierton.

The factors that determine the magnitude of the value of ROI include:

- How much data can be analyzed in terms of the number of sources and the volume of information we can handle.
- How deep you can analyze the data to extract meaning. How easy is it for C-level executives to understand and act on that data?
- The time taken to deliver that value. The quicker the better.

Autonomy has broken the Return on Information concept into an equation.

Return On Information

$$\text{ROI}_{\text{Return on information}} = \frac{\text{Data sources} \times \text{Data volumes} \times \text{Depth of analytics} \times \text{\# / level of users}}{\text{Time}} = \text{Data value}$$

$$\text{Hardware} + \text{License} + \text{Support} = \text{Total cost}$$

"In a nutshell what this equation tells us is the more sources we can look at, the higher the volume of information we can analyze, the greater the depth of the analysis, and the easier we make it to understand data to make business decisions (and the quicker we can do this), the greater the value of the information," said Brierton.

The Vitality of Human Information

As mentioned earlier in this summary, when enterprises look to see what their customers think about a new product or service they've just launched, they look to Twitter, blogs, and Facebook. There is a term for this kind of data: Human Information.

While understanding the meaning of structured data is relatively simple because it's factual—big numbers, dates, amounts, addresses, etc.—understanding the meaning of Human Information is not so easy. It contains ideas, emotions, nuances, and ambiguity.

"The language, sentiments, and styles of language used in these online channels can be difficult for researchers to parse if they are looking to determine, for example, customer reactions to a new product launch. For this need, automated translation can solve many problems in the harnessing of Big Data," said the University of Edinburgh's Adams.

This presents a challenge that traditional competing methods are not built to handle. Databases are built on structured information. Automated translation can derive keywords in the text of emails, video and audio; however, doing so can lose the subtleties and meaning of the information, noted Autonomy's Brierton.

3 Key Facets of Human Information

1. Human information is diverse. It comes from many sources and many different file types and formats, from tweets to video and audio—all needs to be understood.
2. The concept that two pieces of human information are never exactly the same, they have an idea of conceptual distance or closeness (e.g., a front page news story in a newspaper vs. an opinion story on the same topic toward the back of a newspaper).
3. The meaning of words used to express a concept can change depending on the context they're used in (e.g., today the Cloud refers to virtual data storage, whereas "15 to 20 years ago a cloud meant a white fluffy thing in the sky," said Brierton).

IDOL: Meaning-Based Computing

There is a need for real-time answers to Human Information and in those situations Meaning-Based Computing can meet the needs of an enterprise.

Meaning-Based Computing moves away from the restrictive approach of tags and keywords to an understanding of the ideas and concepts in information. It fits the needs of the information as opposed to the other way around.

Autonomy HP's meaning based computing platform is called IDOL or the Intelligent Data Operating Layer.

IDOL allows software applications to access and understand both structured and unstructured data in their native forms. IDOL provides a single processing layer that connects to all sources and types using 400 connectors and understands the meanings and concepts and sentiments they contain. It essentially enables enterprises to leverage 100 percent of their Big Data to power their business.

"It's the only platform that understands all types of data automatically in real time," said Brierton.

For Autonomy, "the approach we take is mathematically based; it doesn't require keywords or tagging," said Brierton. "It's language independent, it's over 15 years in the making, and we invested about $280 million in R&D to build this meaning-based computing engine. It contains over 170 patents."

Meaning-based computing and IDOL allow organizations to move away from blanket-restricted access of information, locking everything down to a more controlled method of access.

"We allow enterprises to promote themselves more effectively by providing them with greater situational awareness—the ability to understand the customer better, to serve them better, and to be clever about customers acquired and retained," said Brierton. "This allows organizations to capitalize on emerging opportunities to drive growth and profit.

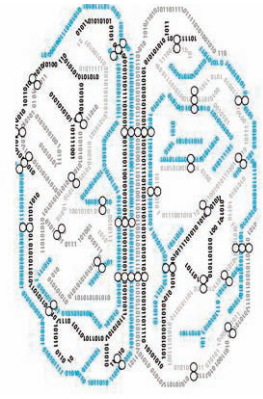
Final Thought

"Stratecast | Frost & Sullivan valued the opportunity to lead this webcast for three main reasons," said Cotrupe. "First, while many people are now talking (or worrying, or both) about Big Data, many are unclear about what it is really all about. We appreciated the chance to demystify Big Data and break it down into manageable concepts: to talk about what Big Data actually is; to reveal its 'DNA;' and to show organizations strategies and technologies they can use to stop 'fearing the data' and instead put it to work for them."

The second major reason to do the eBroadcast, said Cotrupe, was to help spotlight the great work being done by commercial providers such as HP, academia, such as the University of Edinburgh, and by government, in this case Scottish Development/Scottish Enterprise: creating private-academic-public-partnerships to tackle Big Data.

"The third reason we so valued this opportunity," said Cotrupe, "is that it is clear to us that Scotland offers both the resources and support systems to encourage and work closely with organizations that are, or wish to become, providers of Big Data solutions. In today's global economy—and with the flood of Big Data only rising, not subsiding—we welcome this positive development."

Computers Can Now Process Human Information



ABOUT SCOTTISH DEVELOPMENT INTERNATIONAL

SDI aims to assist in the growth of the Scottish economy, by encouraging inward investment and helping Scottish-based companies develop international trade. A joint venture between the Scottish Government, Scottish Enterprise and Highlands and Islands Enterprise Scottish Development International brings together the resources of these organizations to deliver support for companies investing in Scotland. Learn more at www.sdi.co.uk.

ABOUT AUTONOMY

Autonomy, an HP company, is a market-leading software company that helps organizations all over the world understand the meaning in information. A pioneer in its industry, Autonomy's unique meaning-based technology is able to make sense of and process unstructured, "human information," and draw real business value from that meaning. Learn more at www.autonomy.com.

ABOUT THE UNIVERSITY OF EDINBURGH, SCHOOL OF INFORMATICS

The Social Informatics Cluster in Edinburgh works closely with members of ISSTI and with the Edinburgh e-Health Interdisciplinary group. Currently they are working on Information Infrastructures and their role in analysing long-lived, large-scale systems. Learn more at www.ed.ac.uk/schools-departments/informatics

ABOUT FROST & SULLIVAN

Frost & Sullivan, the Growth Partnership Company, enables clients to accelerate growth and achieve best-in-class positions in growth, innovation and leadership. The company's Growth PartnershipService provides the CEO and the CEO's Growth Team with disciplined research and best practice models to drive the generation, evaluation, and implementation of powerful growth strategies. Frost & Sullivan leverages 50 years of experience in partnering with Global 1000 companies, emerging businesses and the investment community from more than 40 offices on six continents. Learn more at www.frost.com.

DISCLAIMER

This Executive Summary discusses key insights and excerpts from a live presentation and panel discussion by Frost & Sullivan, Scottish Development International, University of Edinburgh, School of Informatics, and Autonomy on December 6, 2012. This summary presents industry insights, best practices, and case studies discussed by the presenters, in the context of the live presentation and panel discussion. For more details, visit http://www.frost.com/bigdata. Frost & Sullivan is not responsible for the loss of original context or the accuracy of the information presented by the participating companies.