

Finding new methods to power analytics

James Lawson, Contributing Editor

PROCESSING POWER

The digital world is creating once unthinkable volumes of data, and with demands for super-fast analysis of this data, software vendors and service providers are having to find new and innovative processing methods using cloud-based infrastructure, though concerns linger over data security in the cloud.

Contributors:



**JON
CANO-LOPEZ**

Chief Executive, REaD Group



**MURDO
ROSS**

Head of Solution Design, DST
Customer Communications



**JONATHAN
TAYLOR**

Chief Technology Officer,
Smartfocus



**JAN
WIERSMA**

VP Global Cloud Services, SDL

As digital marketing reigns supreme, so the nature of marketing data and analytics is changing fast. Data volumes are ballooning while analytics must deliver answers in near real time. To cope, software vendors and service providers alike are turning to distributed computing platforms and cloud infrastructure.

"Our site content data is growing steadily but data volumes for social media analysis are increasing very rapidly," says Jan Wiersma, VP Global Cloud Services at SDL. "There, we're seeing between 200% and 400% growth year on year."

SDL has a diverse technology portfolio but its database choices split roughly into two camps. It retains relational platforms like Oracle and SQL Server for conventional single customer view databases but has moved to NoSQL (Not only SQL) platforms to handle data from social and web analytics.

NoSQL encompasses a wide range of database types. Hadoop with its various associated standards is the best known, offering an open-source software framework for storing data and running applications in parallel across clusters of server nodes, but there are many others, from Pivotal Greenplum

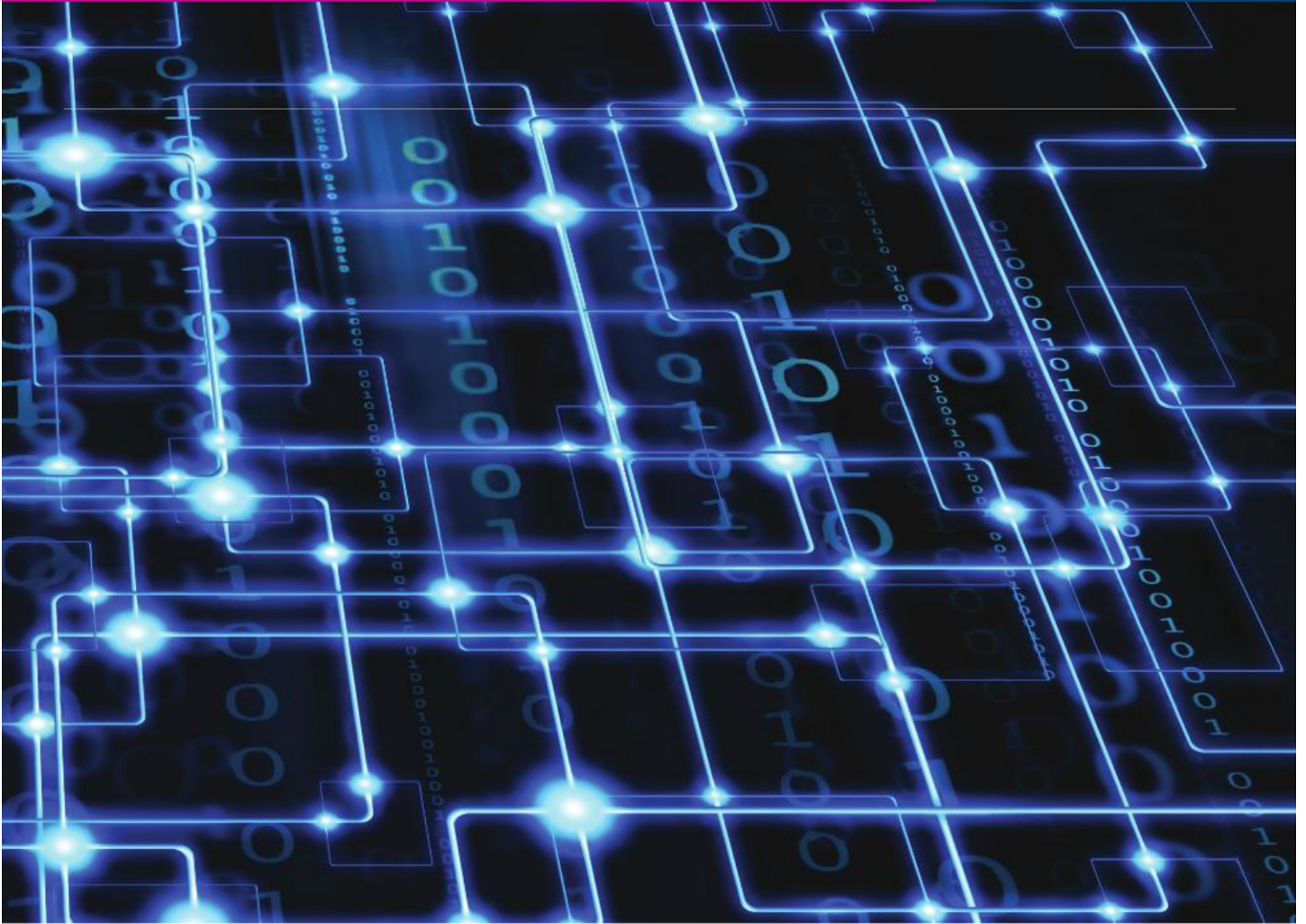
to Cassandra, MongoDB and Redshift.

Massively Parallel Processing (MPP) and rapidly scalable storage let Hadoop and its brethren store, access and process data very quickly. Where the data model in relational databases must be tuned to support specific future queries, NoSQL platforms are ideal for storing semi- or unstructured data for which no fixed schema exists.

PROCESSING POWER

"The volume we can cope with, it's the variety of the data we receive that can be tough to deal with," says Gordon Weston, Director of Technology at Indicia. According to Weston, this is where Hadoop or its Azure equivalent HD Insight are invaluable, able to rapidly store terabytes of heterogeneous data in near-raw form. Analysts can then pre-process it into a suitable form at their leisure, stripping out extraneous data, aggregating and building cubes.

REaD Group is a good example of this shift, previously running SQL Server but shifting to a MPP Neo4j Graph database platform three years ago. That helps it handle the 800 million data items it updates every month as part of its development of a new multi-channel, multi-



device UK consumer database.

"That will rise to one billion monthly next year," says Jon Cano-Lopez, Chief Executive at REaD Group. "We need to be able to access the data in many different ways, rather than having a fixed structure based on name and address."

MPP frameworks like Hadoop can scale up from single servers to thousands of machines. Able to detect and cope with hardware failures on individual servers in a cluster, high availability is designed in. "With MPP using technologies like Hadoop and MapReduce, you can now process terabytes at the speed of gigabytes," says Murdo Ross, Head of Solution Design at DST Customer Communications. "That means you can deal with vast volumes and get answers very quickly."

Another big advantage of NoSQL platforms is their open-source licence model. That's one reason that Smartfocus has moved away from relational databases entirely.

"Most of them are expensive licence-based solutions," says Jonathan Taylor, Chief Technology Officer, Smartfocus. "They just aren't flexible enough to cope with the volumes and types of new data we acquire."



If I were a hacker searching for customer data, I would look for a marketing agency with an in-house server. For hosting, you need to actively assure security and apply the same parameters and policies that you would to an in-house server.



Finding new methods to power analytics**James Lawson, Contributing Editor**

Database volumes on Smartfocus's cloud platform range between 1Tb and 10Tb, with each of its 2,000 clients adding over 100 million transactions per week on average. To cope, it employs a Cloudera Hadoop-based solution. This and a wide range of other applications showcase the state of the art in real-time data management and analysis.

For example, one of the most important tasks is to quickly "cut" data into different structures or views for analysis. Smartfocus analysts use Hive to create multiple relational-style structures on top of large Hadoop data sets and can query the data using a SQL-like language called HiveQL. Impala is another option, an MPP analytic database engine that also runs on top of Hadoop and enables high-speed SQL queries.

Tools like these let analysts transform and view the data as they wish, while truly exploratory data mining applications can look for unseen relationships between variables in the raw data. Spark, a high-speed cluster computing framework well suited to applications like machine learning, is another option for very large data sets while Redis powers Smartfocus's web recommendation engine.

"Redis uses in-memory cache-style processing," says Taylor. "We prepopulate its data store with customer data and, when we recognise a customer, that lets us serve a relevant offer in milliseconds."

Smartfocus is also moving to real time data streaming rather than traditional batch loading. Able to cope with gigabytes of data per second, its Kafka messaging system stores and direct incoming data streams (web behaviour, beacon data, social posts) to the relevant place for processing in real time.

"That lets us send personalised location-based, in-store offers to apps based on a map of the store, persona type, dwell time, weather and many other variables," explains Taylor.

DATABASE DEPLOYMENT

With these huge, high-speed platforms becoming the default choice, the advent of the cloud's service-based infrastructure has driven equally seismic changes in database deployment. Instead of in-house server farms, the choice is now between a private cloud hosted at an outsourced data centre or public clouds like Amazon Web Services (AWS) that deliver instant, flexible processing power,

“

The volume we can cope with, it's the variety of the data we receive that can be tough to deal with.

”

storage and applications.

Services like Amazon's Elastic Compute Cloud (EC2) offer ready-to-go virtual machines complete with operating systems and application programmes. This "bare-bones" processing and storage hardware still has to be configured but there are also numerous other hardware and pre-configured platform services to choose from, from web hosting and business intelligence to machine learning and Glacier archiving.

Flexibility for storage (bursting) and computing power is huge. For example, EC2 capacity can be increased or decreased in real time to more than 1,000 virtual machines simultaneously, with costs based on the computing and network resources consumed.

"Services like EC2 are certainly agile but not really cost-effective because you have to do most of the set-up and management yourself," Wiersma argues. "Services like Amazon RDS (Relational Database Service) are better value and just as flexible."

Hybrid models are common. For example, SDL has its own data centres and servers but also uses public cloud services. Indicia similarly has in-house, private and public cloud options.

"We use the Azure and AWS public cloud to process a lot of our data because it's so

cost efficient," says Weston. "It's very hard to predict resource demands and also to get sign-off for capex. You might buy a new server or two but if you can't forecast demand accurately, you'll soon run out of space."

But high performance at the lowest cost isn't always the province of the public cloud. Taylor says that virtualised Amazon machines struggle to offer ultimate performance while public cloud access to the high-end, high memory physical servers Smartfocus prefers tends to be expensive.

"For us, the Total Cost of Ownership is lower for a private cloud," says Taylor. "One of our learnings has been that running many commodity machines or virtualised servers doesn't really work for us. But because it is so flexible, we do use Amazon for tasks like web recommendation and to cope with varying demand and data volumes."

CLOUD SECURITY CONCERS

Beyond lingering doubts over bandwidth or the financial stability of cloud providers, security is often cited as the biggest barrier to public cloud adoption. Very strict customer guidelines at some businesses can place tight limits on how far vendors and MSPs can go with public cloud adoption, with the preference often to be able to physically inspect and certify a data centre here in the UK.

"I think it would be irresponsible for a company in our industry to put customer data on the public cloud," says Cano Lopez. "Until I can be persuaded otherwise, that's my decision. We need to know where the servers are, what the security is and to be able to guarantee that in a contract."

Following the recent European Court of Justice decision that rendered the Safe Harbor agreement with the USA invalid, a fixed physical location is seen as even more important. There's much confusion in this area however and the advice of SDL's Privacy Officer is that there's little problem with transferring data to the USA.

"We're having a lot of conversations with clients about this at the moment," states Wiersma. "If you are only processing and not storing the data, then you can transfer it between the EU and the USA. As long as the two parties to a data transfer agreement are satisfied, there's no problem."

DST's Ross argues that many IT policies that demand licensed software or on-premise

servers are completely outdated. He cites the Digital Marketplace – a Cabinet Office list of approved cloud providers that includes MS Azure – as an example of how secure the public sector thinks cloud services are.

CLOUD-BASED COMPROMISES

"The compromise is usually a local cloud solution hosted at a regional data centre so you know where everything is at all times," he says. That's the route The Read Group has taken, leasing servers in one data centre with another backup data centre 80 miles away. But is the public cloud being unfairly maligned?

"The AWS platform has matured so there's less and less 'public' about it," says Wiersma. "A new EC2 instance [virtual machine] sits on a closed network inaccessible to anyone else, plus all the data is encrypted and only you have the key."

In the case of EC2, it's also possible to specify where a new instance deployed

geographically and Amazon will also guarantee that the data won't leave the regional borders. Though many marketers might fear a TalkTalk-style hacking episode, that particular breach had nothing to do with server hosting.

"If I were a hacker searching for customer data, I would look for a marketing agency with an in-house server," says Weston. "For hosting, you need to actively assure security and apply the same parameters and policies that you would to an in-house server."

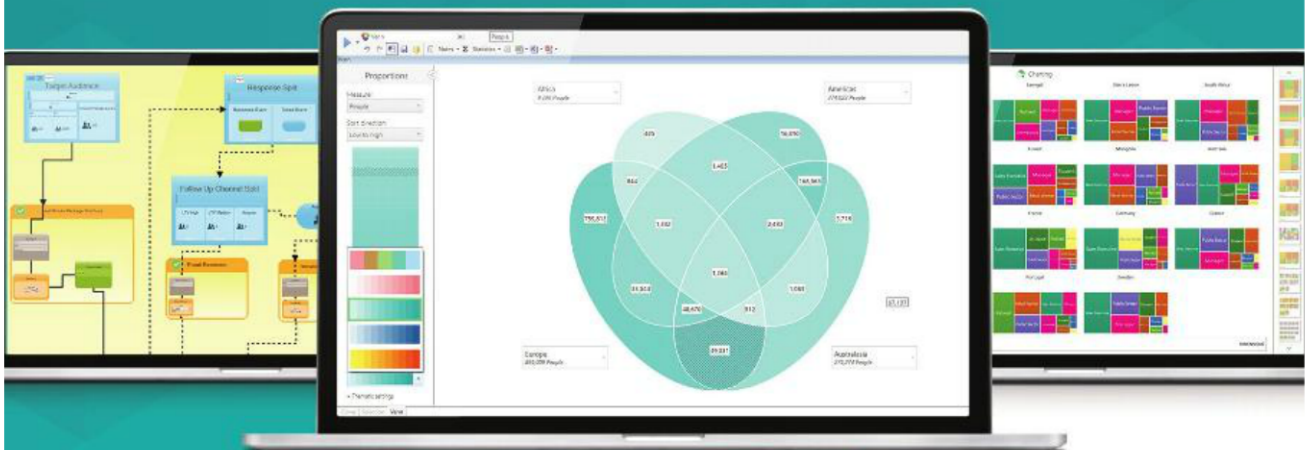
"You need to make sure all the end points are secure, that all data is encrypted and your staff are well trained in security processes," he continues. "Obtain independent security reviews of your infrastructure. Even if a hacker did manage to access a server, your other tiers of logical security should prevent them getting hold of any data."

Whether private or public, or supporting Hadoop or SQL Server, it's clear that cloud infrastructure is becoming the de facto

standard for marketing technology. Platforms like Amazon RDS offer massive agility, letting new businesses start up immediately with no capital investment. Private clouds may not be as flexible, particularly when it comes to scaling down, but they still beat most in-house installations on cost.

The classic reasons for outsourcing also support the shift to the cloud: a focus on core competencies and added competitive advantage. Data-driven marketing is in a transition stage, slowly moving from building and managing everything in-house to a service-based set-up. If your company is great at creating software or analysing customer data, why still build and manage your own commodity hardware? "The days of hosting your data in your own data centre should be well behind us," concludes Weston. "Options like MS Azure, AWS and Rackspace offer a safe, available service 365 days a year – that's their business." ■

With the unique power and automation of FastStats® we can help your business achieve more efficient and effective marketing campaigns.



The FastStats® software suite provides a comprehensive range of data exploration and visualisation tools. Multi-channel, event-driven campaign management software offers powerful marketing automation that will save you time and simplify management and reporting.

Find us online at www.apteco.net



FastStats®
Marketing Analysis & Automation Software

Address Management

PostcodeAnywhere
postcodeanywhere.com/dbmmag

postcoder
web

240+
countries
in one api

allies

01508 494488
www.alliescomputing.com

the Address Experts

HOPEWISER

Address & Postcode Look-up
Cloud Based Pay Per Click
Enterprise Solutions
CRM Integrations
Find Your Nearest Software
Bank Account Validation
Bureau Services
Additional Datasets

www.hopewiser.com/dbm
enquiries@hopewiser.com
Tel: 0845 600 9231

Analysis & Predictive Marketing

UNLOCK INTELLIGENCE
BY BREAKING DOWN DATA SILOS

<MetaMetrics>
integrated marketing analytics

- DATA ANALYSIS & DASHBOARDING
- PREDICTIVE ANALYTICS
- MULTI-CHANNEL MEDIA MODELLING

AGILE ANALYTICS
for a fast-moving multi-channel era

www.metametrics.co.uk

datatech
analytics

THE ANALYTICAL RECRUITMENT AGENCY
FOR ANALYTICAL, STATISTICAL AND DATA JOBS...

datatech.org.uk

Senior Consultant - London or the Midlands

Negotiable Salary - Plus Benefits

This international company works across multiple industries introducing their products and expertise for risk management. This role will afford the opportunity of displaying your problem solving skills. You will engage and build relationships with clients and prospects to understand their needs. You will understand client's business, industry context and market pressures and help plan, execute and manage targeted proofs of concept. You have experience in Optimizer or Optimization Modeller; SAS Marketing Optimization, Experian Marketswitch; GAMS; Risk Solver. You are an excellent communicator and have strong written language, verbal, and presentation skills.

Senior Insight Analyst - Southern Home Counties

Circa 50K - Plus Benefits

This service provider is searching for senior people to assist and encourage their growth. They are well respected in their field and this will afford you the scope to progress as they grow. You will produce KPIs and Reporting (Dashboards, trends, new optimisation and opportunity models with tools like Qlikview). You have excellent communication skills allowing you to create high level summaries for the team to help make better priorities. Your analysis and insights will drive business problem solving and/or make strategic business decisions. You are highly accurate in your analysis and presentation skills are excellent. Ideally you hold proven experience of applying multivariate statistical techniques (e.g. Regression (GLMs), Clustering, CHAID) to business problems.

CRM Marketing Executive - London

to 30K - Plus Benefits

A service provider, whose name will soon be on everyone's lips, is investing in their staff to become a leader in their field. This is an international company and you will be joining a newly created team affording the opportunity to "shine". Using the Email Service Provider and Faststats to analyse performance of campaign results on a daily basis. You will implement changes based on findings and lead on all post campaign analysis across email and direct mail campaigns. Working closely with the internal teams to develop innovative, integrated CRM campaigns, you will be required to work autonomously. Day-to-day management of the CRM programmes including setting up and managing email marketing templates across brands.

Customer Insight Manager - Home Counties

Negotiable Salary - plus Benefits

This unique service provider has an extensive market reach across continents with a vision to transform business. They are proud to boost an unrivalled and differing client base which is constantly growing. This is a role, ensuring that all product development and product management priorities are based on the needs of clients. You will be part of a team charged with developing world-class products. Building a deep understanding of internal and external customers and their objectives. Working with the global digital team, and development partners to ensure focus on the right priorities. Identifying the right amount of customer insights needed will be critical. Expert ability to analyse & synthesize large volumes of information. You hold a great ability to take data and explain so that the whole organisation can understand and work with it

Call us on 01256 314660 email: analytics@datatech.org.uk

