computing
research

Big Data & IoT Review 2017

View from the foothills of the IoT

May 2017

Contents

computing research

1 Executive summary

For several years now, big data technologies have enabled the acquisition and analysis of data that had formerly been out of reach. The Internet of Things (IoT) is now driving big data to the next level with regards to the enabling technologies and also the possibilities. The development paths of technologies such as real-time analytics and machine learning are beginning to converge, and understanding of them and the possibilities they present is increasing. Are we at a crucial tipping point?

For the fifth year running, *Computing* has conducted comprehensive research on the subject of big data. For the first time this year, we have chosen to combine our research on big data technologies and IoT due to the interdependent nature of these two areas. Each requires the other in order for the promise of both to be fulfilled. The *Computing Big Data and IoT Review 2017* explores and analyses the results of this research. Where appropriate, findings are compared with those published by *Computing* in previous years in order to establish trends. The review contains some unique insights from high-ranking IT decision makers into how important they believe IoT is for their organisations and how they are preparing for IoT and initiating projects.

We discuss the challenges IoT poses in terms of the veracity of data, how close we are to genuine data-driven decision making and the technologies that businesses are building their strategies on as well as the degree to which the cloud features in strategy. We also discuss the importance of data scientists to the progression of big data and IoT.

The review concludes with a discussion of what constitutes a leader in big data/IoT and how this relates to the wider picture.

Key highlights from the research include:

- There has been a large increase in the proportion of organisations actively preparing for IoT.

- Although a consensus exists that analysis is the area of IoT offering the greatest benefits, the majority are focusing on the nuts and bolts of deployment and infrastructure.

- Starting small and delivering (or failing) fast is the most popular way of initiating IoT projects. Expectations are high – a majority expected revenue gains of at least 15-20 per cent.

- While the volume of data remains the biggest challenge, there are significant concerns about the veracity of data – and most expect the problems to increase along with the number of devices gathering data.

- Data-driven decision making for the majority of decisions is still some way off, although in three years' time 40 per cent expect this to be the norm in their organisations.

- The majority of those using machine learning were focusing on supervised learning and this is well established in some industries. Unsupervised machine learning has fewer use cases at this stage.

> "The development paths of technologies such as real-time analytics and machine learning are beginning to converge, and understanding of them and the possibilities they present is increasing. Are we at a crucial tipping point?"

◆ Natural language processing (NLP) is the big data technology for which the highest expectations are held – both in terms of what it could deliver and its expected growth.

◆ The IoT technology field is still crowded by competing connectivity protocols.

◆ Blockchain is viewed as offering some potential to overcome some issues of trust and veracity of data posed by IoT.

◆ The volume of data stored or processed in the cloud is expected to almost double in the next three years.

◆ There is little consensus of what the core skills of data scientists are – and specifically what distinguishes them from data analysts. However, some agreement exists that the differences can be reduced to creativity, and an ability to identify unknown unknowns.

◆ Such individuals are thin on the ground and dissatisfaction with the job market is widespread.

◆ Most participants feel that they are 'followers' in terms of their use of data. The distinguishing features of leaders are a clear data strategy, robust governance and a focus on speed to market. While technology is an indicator of advancement and leadership, cultural factors are ultimately deemed of greater importance.

2 Research overview

The key objective of the *Computing Big Data and IoT Review 2017* was to establish how IoT is driving the development of big data technologies and projects, and to identify the challenges organisations are facing from the increased volume of data from exponentially increasing sources.

Key areas of research included:

◆ Whether organisations are more likely to initiate IoT projects than they were a year ago – or the extent to which they are preparing to do so.

◆ The financial expectations organisations have of projects and the approaches most likely to lead to success.

◆ The challenges being posed by increasing volumes, variety and velocity of data generated by IoT, and the likely veracity of that data.

◆ The extent to which data-driven decision making is becoming the norm and the extent to which supervised/unsupervised machine learning is underpinning this.

◆ The most popular technologies for both big data and IoT and the expectations of which technologies will dominate the next three years.

◆ The importance and availability of data scientists and the differences between these and data analysts.

◆ The characteristics of organisations that are leading the field.

2.1 Methodology

The research project was conducted in six phases during March and April 2017, using a combination of qualitative and quantitative methods.

Phase 1 – Telephone interviews with IT decision-makers who participated in the previous big data and IoT research in order to establish key changes that may have occurred in the last 12 months.

Phase 2 – Telephone interviews with IT decision makers representing organisations that have already implemented IoT strategies.

Phase 3 – A focus group consisting of several IT decision makers and analysts was conducted in order to spark debate and gain an understanding of the different motivations for and objectives of big data and IoT projects. Participants were drawn from a range of organisations including banking, media and healthcare.

Phase 4 – A nationwide, online quantitative study completed by more than 340 IT decision makers representing organisations ranging in size from a minimum of 100 employees to enterprises comprising many thousands across multiple industries including retail, telecoms and finance.

Phase 5 – A second focus group consisting of different individuals from the first group was conducted in order to review the quantitative research and gain further insight. Industries represented here included engineering, NGO and education.
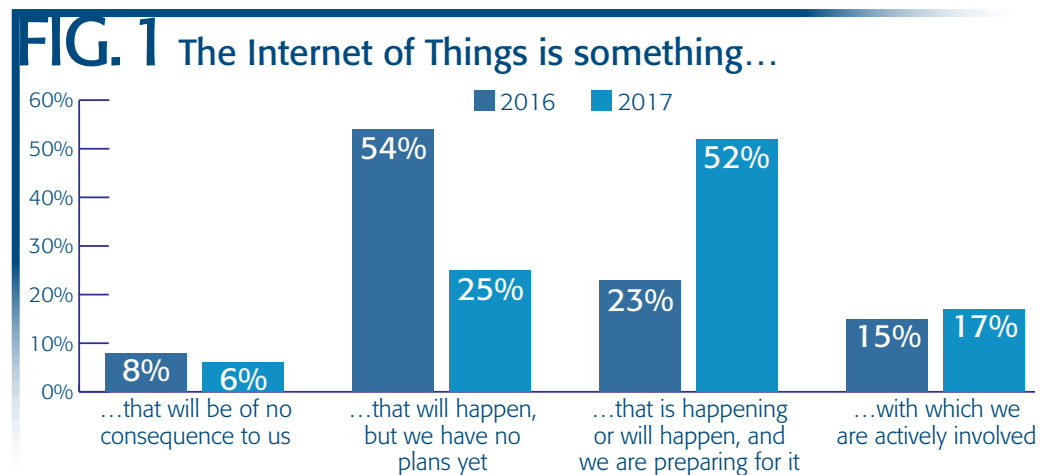
Phase 6 – A final round of in-depth interviews were conducted with several individuals deemed to be experts in the fields of big data and IoT.

> "The low ranking of analytics suggests strongly that organisations are still in the very early stages of IoT, wrestling with how to even gather, store and manage the phenomenal quantities of data that IoT will generate before they even begin to look at the analytical potential"

3 The IoT – preparing to launch

Computing presented its first piece of research on IoT last year. The feeling of a majority of those who participated in the research was that they did not expect IoT to make any significant impact on their particular organisations for between three and five years. While the possibilities for IoT were endless, that was, in itself, part of the problem. What appeared to be lacking was clarity on the eventual objectives of IoT projects. Many organisations didn't really know what benefits they could derive from the sea of IoT data. The idyll of truly data-driven decision making seemed very far away.

As Figure 1 suggests, the last year has seen some subtle changes in thinking. While the proportion of respondents actively involved with IoT has barely perceptibly nudged upwards, the proportion actively preparing for it has more than doubled, leaping from 23 per cent to 52 per cent.

FIG. 1 The Internet of Things is something…

Legend: 2016, 2017

	…that will be of no consequence to us	…that will happen, but we have no plans yet	…that is happening or will happen, and we are preparing for it	…with which we are actively involved
2016	8%	54%	23%	15%
2017	6%	25%	52%	17%

There were also subtle but possibly significant changes in responses to the question "How important is the IoT to your business?". Participants were asked to rate this importance on a scale of one to seven with seven being the most important and one the least. In 2016, 48 per cent of respondents ranked IoT at either one, two or three in terms of importance. This year the proportion rating importance at these lower levels level drops to 35 per cent. However, there was not a corresponding increase in the proportion of respondents rating IoT as very important. Instead, the proportions in the middle of the scale have increased.

The same pattern is apparent when comparing findings with the question about the urgency in preparing for the IoT.

FIG. 2 Regardless of your own involvement in IoT deployment, how urgent is it that you prepare for any impact on your business?

Legend: ■ 2016 ■ 2017

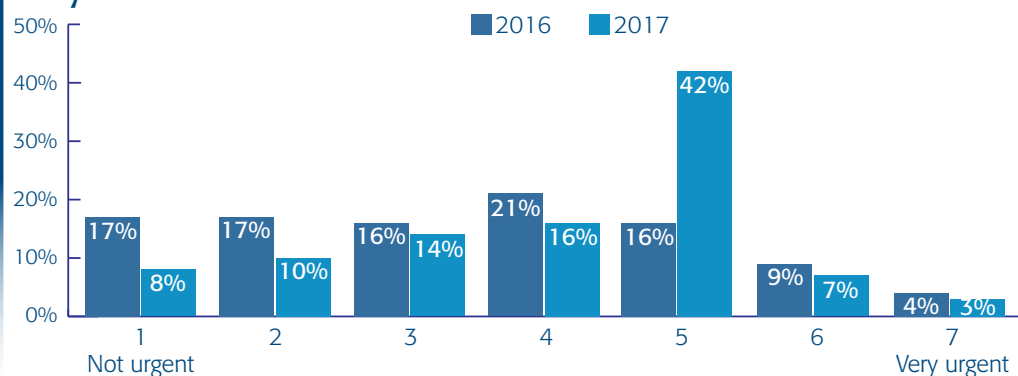| Scale	2016	2017
1 — Not urgent	17%	8%
2	17%	10%
3	16%	14%
4	21%	16%
5	16%	42%
6	9%	7%
7 — Very urgent	4%	3%

Figure 2 shows us that the distribution of answers has changed. In 2016 most participants ranked importance somewhere around two to four. This year we see a big jump in those placing importance at a more important five.

What the combined findings of these questions suggest is that in the last year, the subject of IoT has moved up the agenda. The undefined point in the future where IoT will start to make a difference has moved closer, and businesses are beginning to prepare in larger numbers.

If IoT is moving up the agenda, what aspects of it are organisations focusing on? According to participants in both quantitative and qualitative research, the answer seems to be operational rather than analytical. The most highly ranked perspective was "deployment", followed by "data storage and processing". Analytics was the least important factor.

> " I think it's probably more an operational thing. The use cases around it are very functional in purpose, less analytical… I think the way most people are approaching it is in a more functional way, getting sensors and scanners in place… CTO, Technology

> " Lots of people talk about it, but I think we are still 2 to 5 years away, partly because it is not a quick fix and equipping infrastructure with devices means there is a hell of a lot of stuff that needs to be joined together to be made to work into an effective ecosystem… CIO, Education

The focus on operational issues is consistent with findings from last years' research, which pointed to concerns about the lack of dominant protocols and standards, and also about data protection and privacy. The low ranking of analytics suggests strongly that organisations are still in the very early stages of IoT, wrestling with how to even gather, store and manage the phenomenal quantities of data that IoT will generate before they even begin to look at the analytical potential – a position summed up perfectly by the individual quoted below.

> **The implementation of it is so massive, like putting tags on every item with tons of stock is not easy or sensors in every doorway in every store. It's not an easy thing to do and we are not past the challenge of capturing the data yet, let alone doing something really meaningful with it…**
> CTO, Technology

As this individual suggests, actually having the data is of limited value to business. Only the insights that they can extract from that data are useful. There was some frustration that the market was being driven by hyperbole.

> **The one thing that frustrates me about IoT is that it's always the next big thing, just like RFID was, just like Wi-Fi was and anything else. That's not the business case. The data is out there, it's just looking at it and actually making a sensible analysis, an interpretation of it, to drive business back in** Business consultant

> **What is the use of having the stats, unless you start putting them together and you know what you're looking for? Someone has to pick up what is an interesting statistic to look at and it's not always obvious what is interesting…** IT manager, Automotive

> **It's good to have the immediate information, but unless you're going to do something long term with it, you are losing a lot of value…** IT manager, Automotive

4 Starting small

Despite the analytical potential of IoT still being in its infancy, some businesses are using IoT to useful effect, as the following quotations illustrate. It's also worth noting that organisations connected with oil and gas exploration, manufacturing, distribution, weather forecasting and agriculture have been using sensors for years.

> **For us it's enabled us to see efficiencies we couldn't see before. Now we've got all that data, we can visualise it and track it; it's much easier for us. It's a time-saving exercise; you could do all of this with a clipboard if you wanted to…** IT manager, Automotive

> **Trying to predict where to open a store, but at the same time predict what the demand might be. Finding location, feeding information, forecasting sales and sales drops…** Director digital, analytics and innovation, Professional services

> **We use IoT data; not the actual IoT data itself, but the fact that you have the device. So if we know that you've got a Fitbit, for example, that opens up a whole wealth of valuable information to our advertisers because we know you're health conscious. Our main use of device data is almost always the position, the location of the device – that's why mobile phone data is a terrific source of information…** Head of data strategy, Media

"The 'start small and deliver (or fail) fast' approach seems to be the most popular among those who took part in our research. However, expectations of what IoT can deliver are high"

computing research

Figure 1 above illustrates that 17 per cent of those we surveyed were already actively involved with IoT. A further 52 per cent were actively preparing for it. To add detail to this picture and try to establish how imminent broader IoT adoption is likely to be, we asked survey participants to rate the likeliness that they would begin an IoT project in the next 12 months.

The largest single proportion of respondents (31 per cent) rated themselves at six with a further five per cent at the maximum score of seven (Fig. 3). Overall, 46 per cent of those surveyed were more likely to initiate a project than not. Adding further substance to this hypothesis is the finding that 78 per cent of those who had already implemented IoT were more likely than not to initiate further projects, suggesting that these smaller scale deployments are showing signs of delivering their promised returns.

FIG. 3 How likely is it that you will implement an IoT project in the next 12 months?
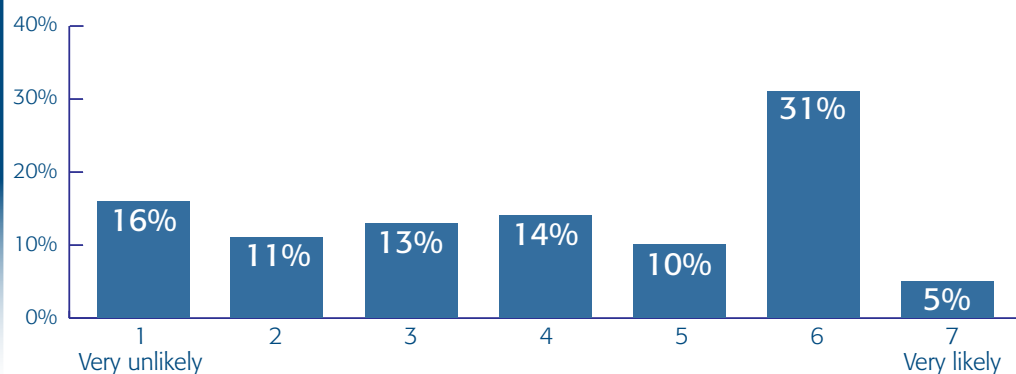
It would seem that one of the key components of success with IoT is to start small. Of those who had initiated a project, 49 per cent stated that they had opted for a small-scale deployment. Certainly, the "start small and deliver (or fail) fast" approach seems to be the most popular among those who took part in our research.

In terms of return on investment (RoI), 59 per cent expected a return in three years or less which is fairly standard for those seeking to make a business case.

> When you get RoI, when should it pay back? It's very hard to say, but unless you do a quick delivery, you probably will never achieve it as you will be delivering wrong. If your RoI is over three years, then the project shouldn't go ahead… IT manager, Automotive

> The thing to watch out for in IoT is it's so fast moving and there are projects after projects, so it's not standing still… IT manager, Automotive

RoI is particularly tricky to quantify for IoT, partly for the reasons discussed in Section 3 and also the *Computing Internet of Things Review 2016*. The objectives of large-scale IoT projects can be a little fuzzy. There is a lack of consensus on where we all want to go with IoT and as a result, it can be difficult to actually quantify how the bottom line improves. Part of the problem is the fact that future insights generated by IoT data are impossible to know – and thus value.

> **Extra revenue from efficiencies that is measurable, but you would struggle to quantify the insights you have gained…**
> IT manager, Automotive

This difficulty in quantifying future returns may be leading to a degree of bullishness. When asked, "How much growth in revenue do you expect to see from IoT in the next three years?", by far the largest single proportion of respondents (a huge 55 per cent of those surveyed) stated 15-20 per cent. A further 18 per cent put their expected growth at more than 20 per cent. Indeed, three per cent put it at in excess of 50 per cent.

This question could have been interpreted as growth in profits as opposed to revenues so it's possible that some of our respondents could have been thinking about the potential of IoT to reduce costs and make efficiencies as well as increasing revenue. Nonetheless, it is difficult to argue that these expected returns are anything other than optimistic.

These findings sit awkwardly with those indicating that the most successful approach to IoT is likely to be starting small. These small deployments allow organisations to quantify benefits to financial outcomes – thus leading to subsequent larger projects. Smaller projects are more likely to succeed, partly because expectations are proportionate and returns faster. IoT proponents would be wise to avoid over-selling benefits that they cannot yet quantify.

"The challenge is as big as the data. When the number of data sources runs into thousands and millions, veracity becomes nigh on impossible to determine"

5 Volume, veracity, velocity and variety

Historically, big data has often been described in terms of the three Vs. Traditional data storage and analytics tools are not up to the job of translating the volume, variety and velocity of data now available to organisations into actionable insight and, ultimately, more personalised goods and services and increased profits. However, as IoT has progressed, a fourth V has become part of the equation. Driven partly by the increase in IoT-enabled devices and sensors, the veracity of data has become a factor – and a challenge.

We asked participants in our survey to tell us which caused them the greatest issues and to rank them accordingly. Volume still came top of the problem list (Fig. 4).

FIG. 4 When looking at these 4 Vs in relation to big data, please rank in order of importance the ones that cause the most issues

Volume

Veracity

Velocity

Variety

> **Volume is still a problem because of storage issues and managing the volumes because they are just so huge and it drives up costs. There also needs to be proper curation of it due to GDPR…** CIO, Education

> **It is understanding what's good and what's garbage because the whole thing with big data is that we keep everything but then you've got so much noise in the dataset it becomes irrelevant, so we use supervised learning to at least get us a decent dataset that we then do analysis on…** Head of Technology

With his reference to "noise in the dataset", the head of technology quoted above leads us to the problem of veracity which was ranked as the next biggest problem. Our panellists explain some of their reservations when it came to trusting data – in particular, IoT data.

> **The validity of the data that comes up from the sensors will depend on whether or not the person that installed it followed the installation procedure. I've had so many cases where there was a monkey installing it, so it's not measuring what it's supposed to. It is a huge issue…** CTO, Oil and gas

> **We don't have any idea really how those things are calibrated. That's not to say that it's useless, that's to say that you've got to be very careful in just assuming that sensor information is a fact…** CIO, Environmental services

It isn't difficult to imagine how incomplete or incorrect data can cause serious difficulties if IoT data is going to be used in the way its evangelists imagine. One of the first ways is to make operational efficiencies. The thought of on-the-fly or automated decisions about pricing, purchasing, logistics and so on being made on the basis of unreliable data is not an attractive prospect. The challenge is as big as the data. When the number of data sources runs into thousands and millions, veracity becomes nigh on impossible to determine.

Sentiment analysis of social media serves as an excellent example of these challenges. Sentiment analysis of Twitter during the UK general election in 2015 would have indicated a win for Labour. We all know how that panned out. The problem was not that the data was in itself inaccurate – it was simply that people intending to vote Conservative tend to be far more shy about posting their views on Twitter, so huge chunks of data went unanalysed.

More nefariously, social media analysis can be distorted by bots or by the planting of fake news. Recent elections and referenda across the world have seen allegations of murky tactics from all sorts of players, and investigations by the Information Commissioners Office and Electoral Commission are ongoing.[1] How to determine good data from bad?
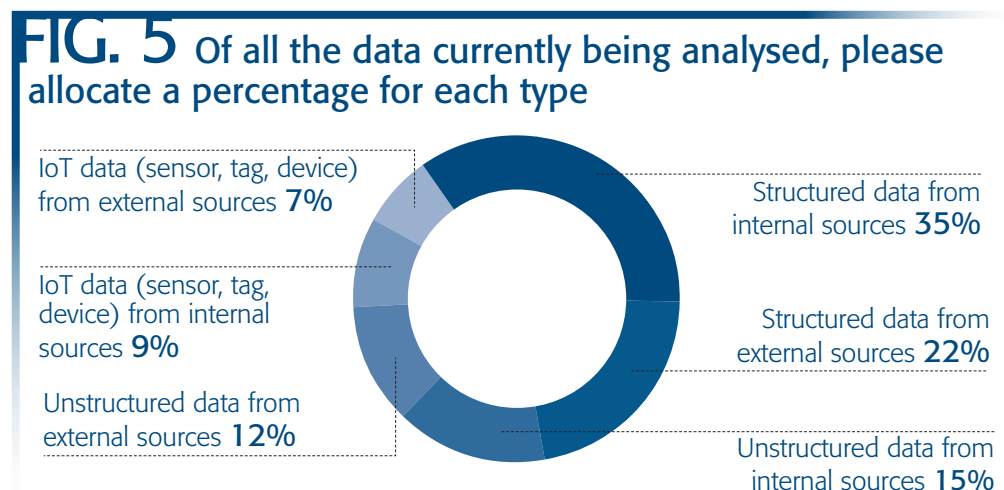
The velocity of data concerned respondents less than in earlier years of big data research. But the importance of real-time analytics has increased.

> We're doing analytic streaming between hospitals and providers. It's real-time analysis. If we notice the red flag is coming, the system is going to crash; then we will stop the patient going in that direction… Business intelligence manager, Health

> Our factories don't know at nine o'clock in the morning if they've got enough raw material for the day. So our game is all about understanding where our raw material is. Is it with us? Is it with our clients? Is it on its way back? Is it on its way out? For us, the game is all about getting more sophisticated about understanding that cycle, so that's about streaming analytics and about processing high volumes of data very quickly… IT director, Textile services

The variety of data that our respondents are both capturing and analysing is wide – although a considerable gap exists between data just being captured or stored and the likely analysis of that data.

Figure 5 shows that currently structured data from internal and external sources is most likely to be analysed. This is not surprising given that traditional relational databases are still dominant in the enterprise supporting data warehouses, CRM and ERP systems, for example.

FIG. 5 Of all the data currently being analysed, please allocate a percentage for each type

IoT data (sensor, tag, device) from external sources **7%**

IoT data (sensor, tag, device) from internal sources **9%**

Unstructured data from external sources **12%**

Structured data from internal sources **35%**

Structured data from external sources **22%**

Unstructured data from internal sources **15%**

[1] https://www.theguardian.com/technology/2017/mar/04/cambridge-analytics-data-brexit-trump

However, Figure 6 illustrates just how our respondents see this picture changing. The largest expected growth areas are the analysis of structured data from external sources, and of unstructured IoT data from sensors, tags and devices deployed within the organisation.

FIG. 6 What types of data are you expecting to analyse more of in the future?

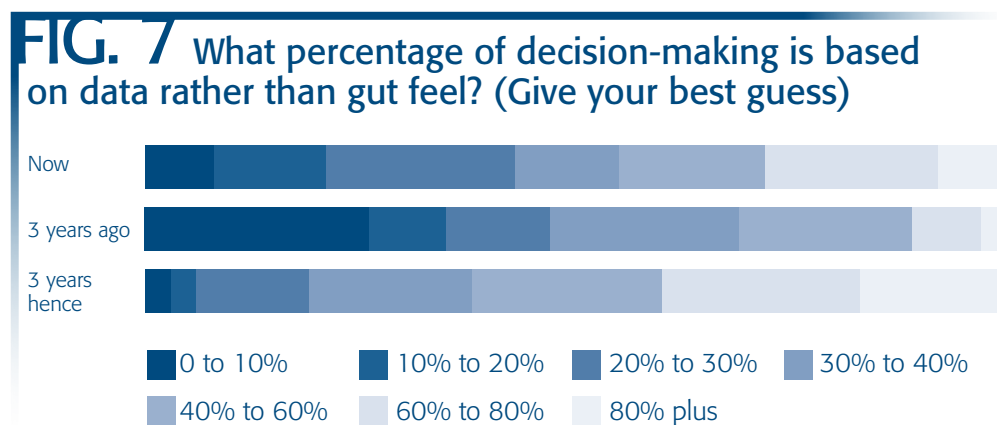|	
Structured data from external sources	61%
IoT data (sensor, tag, device) from internal sources	47%
Unstructured data from external sources	36%
Structured data from internal sources	30%
Unstructured data from internal sources	29%
IoT data (sensor, tag, device) from external sources	16%
Other	1%
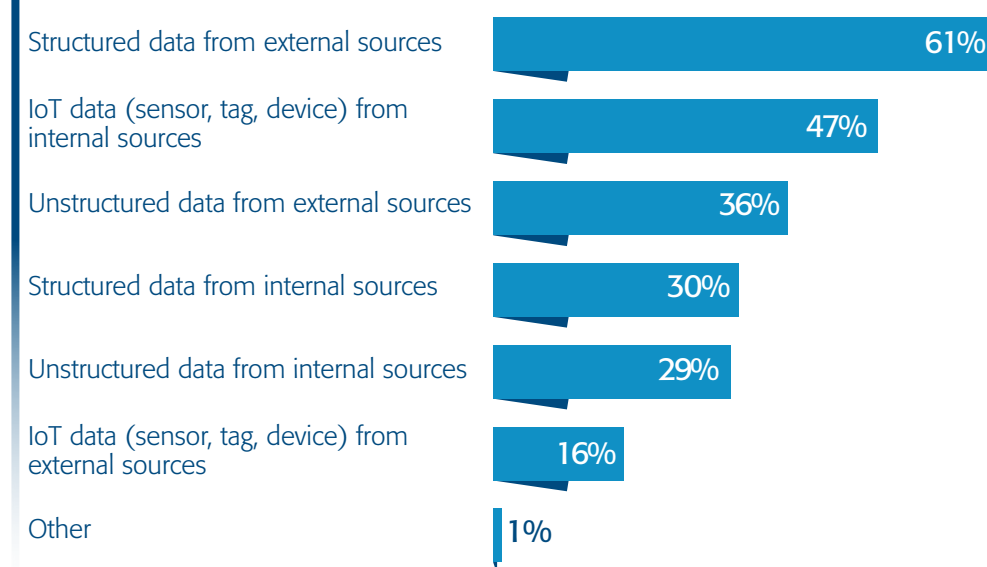
Ranking last, however, was the expectation of greater analysis on IoT data generated by external sources. These findings suggest that organisations are focused on using IoT data to drive internal efficiencies. They may already be facing challenges in trying to close the gap between the capture and analysis of IoT data generated by internal devices and that their focus will be here in the next year or two before they begin to look at externally generated data.

6 Data-driven decision making – how close are we?

Data-driven decision making is the ultimate aim of organisations using big data. To what extent has this goal been reached? The answer, according to those we surveyed seems to be… a little.

When asked to estimate a percentage of decision making that was based on data as opposed to gut feel, only 28 per cent told us that a majority of their decisions (between 60 and 100 per cent of their decisions) were based on data (Fig. 7). However, only 11 per cent stated that this was the case three years ago. In three years' time, 40 per cent expect this to be the case. Cognitive bias notwithstanding (people love a narrative), this represents steady progress.

FIG. 7 What percentage of decision-making is based on data rather than gut feel? (Give your best guess)

Now

3 years ago

3 years hence

- 0 to 10%
- 10% to 20%
- 20% to 30%
- 30% to 40%
- 40% to 60%
- 60% to 80%
- 80% plus

The feeling of many of those we spoke to for this years' research was that, at present, the human aspect of decision making was still prevailing, but that data was increasingly being used to justify those decisions. This is a finding consistent with big data and analytics in general still being relatively young in many organisations.

> " I think there's quite a lot of decision making that is obviously not based on data. It might have some real-world data round the edges but really, that's more often chosen to underpin a decision that is made… Service delivery manager, Banking

> " You're going to make some decisions based on what you know and what your gut feel is. I think the gut feel will diminish a little bit more over time since the data we're getting is actually so much better because it's coming from more sources. You can test it better and you can give yourself a bit more leeway to actually say the data is pointing us in this direction… Business consultant

computing research

It is interesting to note that this finding applies even to fields where one believes instinctively that human decisions would prevail – such as charity, law and education.

> **Working in the not-for-profit sector, a lot is done on gut instinct and relationships. We have now started to analyse funding trends and where we think we're going to have best success with particular types of project, so we're starting to do more with data. Three years ago, our data usage was really low; we had experts and [decisions were] just based on their gut, but that is changing now. But I think we'll never get to a situation where massive volumes of our decision making will start purely on data…** Head of technology, NGO

> **In the US, criminal judgements in courts are now being made by AI systems… It takes into account the likelihood to re-offend… Could it happen in the UK? Well, there are not enough barristers in the UK to read all the documentation to keep the justice system stable…** Change actor/consultant

> **You get data relating to your students from a number of different data points and from that you can identify students who are struggling or at risk of dropping out allowing you to undertake tailored interventions. This learning analytics is becoming much more mainstream…** CIO, Education

The extent of supervised and unsupervised machine learning technology is a fundamental part of the discussion of data-driven decision making. Of the organisations that were using (or planning to use) machine learning for decision making, 51 per cent were focusing primarily on supervised machine learning. Only 13 per cent were focusing on unsupervised learning, with a further 20 per cent unsure. The use of machine learning is driven very much by the type of organisation employing it, and supervised machine learning is reasonably well established in at least some industries.

> **I think of so many useful cases for machine learning and AI in banking, such as with fraud detection. So far it's supervised machine learning, the software learning from humans saying 'this is dodgy; actually that's OK'…** Service delivery manager, Banking

> **If you're scraping data from the web or taking social media data then to get any value out of it you use supervised learning to understand what's relevant and what's just garbage…** Head of technology, NGO

> **We are making use of machine learning and AI. If you've got our latest products then it's a bit like an Amazon thing: 'So you enjoy this programme, what about watching these programmes?'… All of that goes through our machine learning and AI engines…** Head of data strategy, Media

> **We are looking at machine learning, especially when it comes to predictive analytics such as predicting failure. Tracking the events immediately before something happening as it will be helpful for the repair work as well…** IT manager, Automotive

Unsupervised machine learning, in which the software effectively learns from its own experience, is far less commonly deployed but holds some intriguing possibilities. Some of our panellists and interviewees were firmly convinced that this was the future for the organisations they represented.

> **I have absolutely no interest in AI that mimics people, I don't want a machine that thinks like us; I want a machine that finds a solution to a problem in a different way to human beings…** Change agent/consultant

> **With unsupervised machine learning… you're exploring the unknown unknowns…** Big data architect, Technology

7 Technologies to watch

Figure 8 shows some of the big data and analytics technologies that our survey respondents are presently using – along with those they think will be using a few years hence. While real-time analytics was desired by the largest proportion of respondents, the technologies showing the steepest growth curves were Hadoop and NLP. The continuing projected growth of Hadoop is an interesting finding, given that it is already well established and that newer platforms, particularly Spark, have generated quite a buzz.

This is likely to be down to the fact that these platforms are not mutually exclusive. While there is some overlap, each has different strengths. Indeed, the main big data tools are frequently packaged together with Hadoop by distributors.

"The combination of NLP and big data holds some intriguing possibilities for business – particularly those with an eye to using sentiment analysis to mine the rich seam of social media"

> **The market is becoming aware of what Hadoop can and cannot do. They are aware that Hadoop is more mature than Spark but they struggle to find the use cases for Spark. Hadoop is still the starting block, and Hadoop and Spark and not mutually exclusive. A year ago people were saying, 'No, no, don't do Hadoop, do Spark instead', but now people are certainly using the two together…** Analytics consultant, Financial services

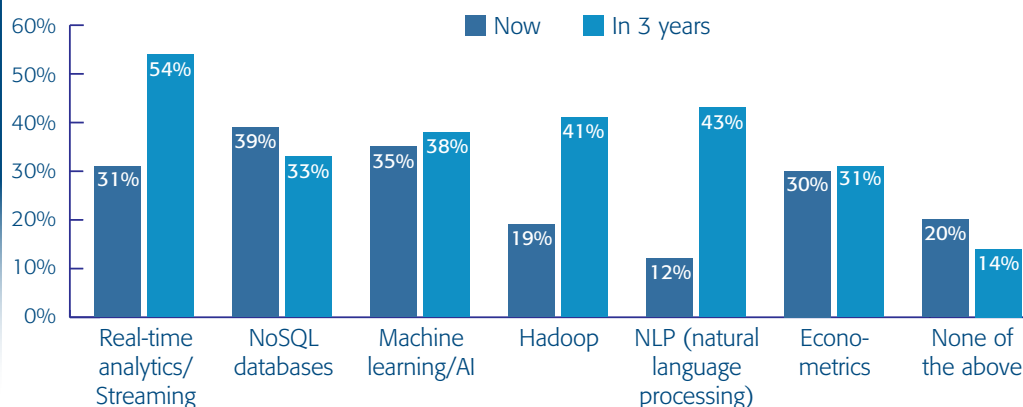> **I see a lot of Spark versus Hadoop, but I find it a funny way to look at it. You have a big ecosystem with a lot of tools that make up Hadoop. Some people look at HDFS, Hive, MapReduce; style jobs are one thing and Spark is different, but it's not really the case, it's just picking the right tools and they all piece together so well in any type of combination…** CTO, Technology

However, some who were beginning projects perhaps later than originally planned were more likely to give Hadoop (or at least 'Hadoop One') a swerve.

> **Partly because of the nature of the industry and in a way it's a bit antiquated on the IT side, so I think at the moment that as we are transitioning to newer technologies, Hadoop will be skipped…** CTO, Oil and gas

> **For it to really be Hadoop, you need HDFS in there. Typically you would have MapReduce, but Spark can replace that MapReduce role; it's still the same type of paradigm. It is really an evolving set of tools that hasn't had a rebrand, but more stuff keeps getting added all the time. We need Hadoop Two which is a different set of tools and technologies or we are accepting that Hadoop is just HDFS and MapReduce. Most people are not using MapReduce anymore…** CTO, Technology

FIG. 8 Which of these big data/analytics technologies are you currently using? What will you be using in the next 3 years?

Legend: ■ Now ■ In 3 years

Technology	Now	In 3 years
Real-time analytics/Streaming	31%	54%
NoSQL databases	39%	33%
Machine learning/AI	35%	38%
Hadoop	19%	41%
NLP (natural language processing)	12%	43%
Econometrics	30%	31%
None of the above	20%	14%

The technology showing the single biggest expected increase in use was Natural Language Processing (NLP). The reason for this perception may have a lot to do with the high-profile arrival of Amazon Echo devices and Alexa personal assistant.

The combination of NLP and big data holds some intriguing possibilities for business – particularly those with an eye to using sentiment analysis to mine the rich seam of social media. The prospect of being able to communicate in "real human" language and use that language to inform marketing and business intelligence and to reduce the gap between humans and machines, as in the case of chat bots and automated customer services, is highly attractive. The expectations of our panellists on the technology were high – particularly of Google.

> **"** The easier that it is for the user to interact with the application, the better it is… I can talk to my car from my Echo, which is weird. But the thing about Google is they've got such a massive solution, engine database behind them. Whereas if I ask my Alexa a question it will say, 'I'm not sure what the answer is'… Business consultant

> **"** Google's offering is coming out I think in June and when that hits, that's going to be a bit of a game changer. It's so much more than just talking to a device and saying 'what time is it?', 'tell me a joke', it's 'switch on the heating'… They've built up their home product offering with all the API hooks; the instant thing will be linking to Google Home… Head of data strategy, Media

However, not everybody was convinced that NLP's moment was nigh. Some viewed it as being good only for specific tasks and in need of greater refinement before it went mainstream.
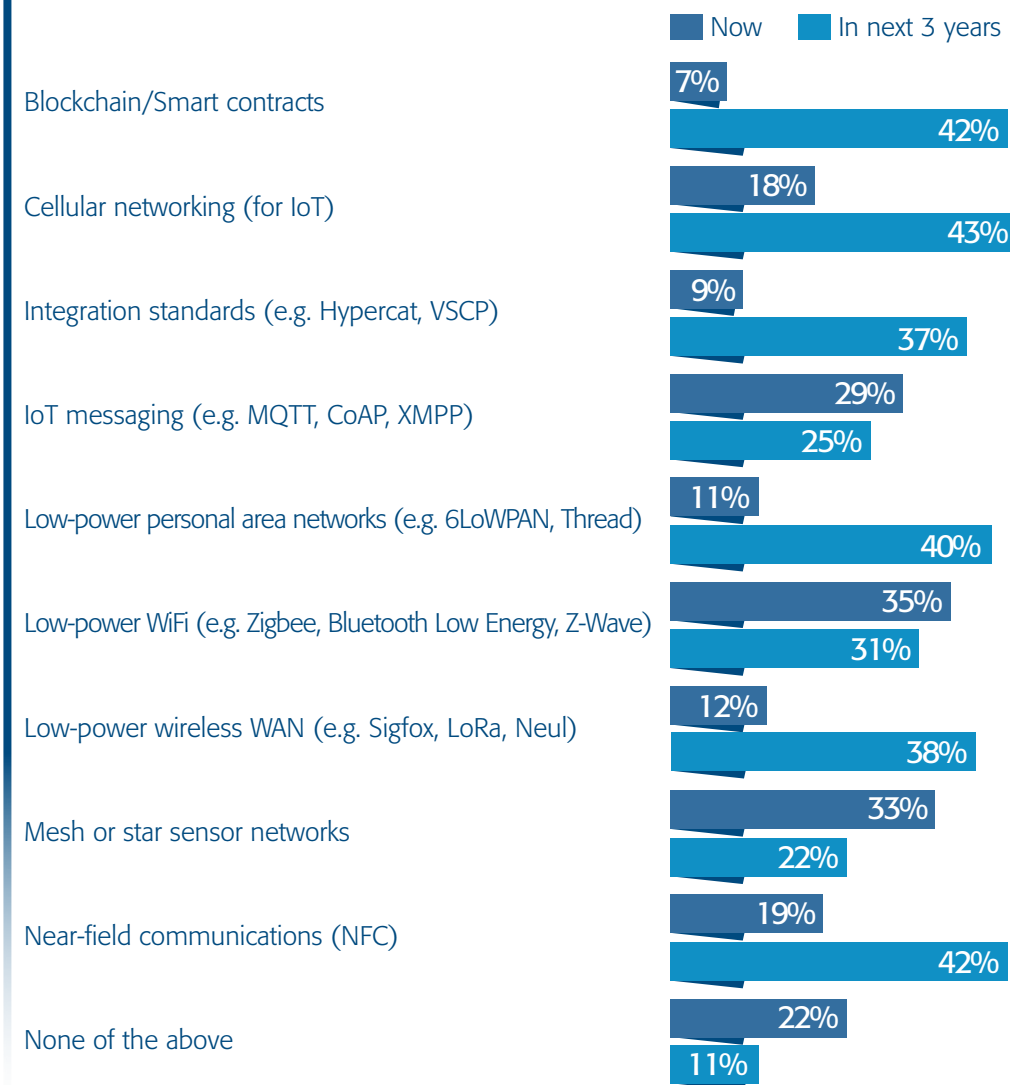
> **"** I can't see in three years NLP being effective across all the many myriad languages that we uncover in our day-to-day lives. We looked at speech-to-text processing for gathering more datasets from particular countries, taking raw audio streams and turning it into usable data, but it's just not quite there at the moment…
> Head of technology, NGO

> **"** NLP: we're actually looking at colleague emails just to make sure that they're not leaking out sensitive data by accident or on purpose… Big data and machine learning SME, Financial services

> **"** I don't see the benefit or the value even in three to five years. I think it's too vague, it's not structured enough, so I don't see real application except customer services, which is basic, so that for me is not really a fully-fledged application across the whole industry…
> Data manager, Research

The IoT technology field is significantly less mature, with the technologies in question currently being more about networking protocols than packaged software, platforms and devices (Fig. 9). Indeed, the lack of any emerging dominant technologies or consensus on standards/protocols is one of the most significant factors in preventing IoT from growing as fast as many envisaged it would. The technology with the biggest expected growth was not really an IoT technology at all, albeit one in which many see potential applications for assuring authenticity – blockchain distributed immutable ledgers and smart contracts.

FIG. 9 Which of these IoT networking technologies are you currently using? What will you be using in the next 3 years?

Legend: ■ Now ■ In next 3 years

Blockchain/Smart contracts
- Now: 7%
- In next 3 years: 42%

Cellular networking (for IoT)
- Now: 18%
- In next 3 years: 43%

Integration standards (e.g. Hypercat, VSCP)
- Now: 9%
- In next 3 years: 37%

IoT messaging (e.g. MQTT, CoAP, XMPP)
- Now: 29%
- In next 3 years: 25%

Low-power personal area networks (e.g. 6LoWPAN, Thread)
- Now: 11%
- In next 3 years: 40%

Low-power WiFi (e.g. Zigbee, Bluetooth Low Energy, Z-Wave)
- Now: 35%
- In next 3 years: 31%

Low-power wireless WAN (e.g. Sigfox, LoRa, Neul)
- Now: 12%
- In next 3 years: 38%

Mesh or star sensor networks
- Now: 33%
- In next 3 years: 22%

Near-field communications (NFC)
- Now: 19%
- In next 3 years: 42%

None of the above
- Now: 22%
- In next 3 years: 11%

> " I don't know how it is going to factor in, but blockchain is going to be massive. Not sure if we are calling it a big data or an IoT thing or its own, it's something else. Blockchain has the potential to transform everything, the financial industry in particular, but also in medicine. Maersk the shipping company has just put in this trail it's doing with custom agencies and companies they deal with… CTO, Technology

> " Smart contracts/ blockchain is interesting, especially with things such as land ownership, which often causes conflict; that's an area where blockchain might be quite interesting, in terms of smart contracts blockchain could be better than older traditional methods… Head of technology, NGO

While not everybody was convinced that blockchain will prove itself useful outside of its cryptocurrency origins over the next three years, the technology offers the possibility of being able to overcome the challenge of data veracity.

> ❝ It's like if you can trace one transaction, you can see the next and the next. You can trace people back through the transactions… Data manager, Research

> ❝ Deploying technologies like blockchain to create a consensus view of the veracity of the data supply chain are kind of areas that we're digging into… CIO, Environmental services

> ❝ Blockchain gives you the potential to lock known calibration characteristics into a secure ledger in a way that the community agrees with, so all of a sudden you see potential to push what currently can only be treated as opinion or herd data, into something that approaches much more of a fact… CIO, Environmental services

Cloud computing is a fundamental part of the technology discussion around big data and the IoT. The proportion of enterprise data being stored or processed in the cloud is 21 per cent today – but that was expected to almost double to 39 per cent in three years' time (Fig. 10). The only surprise among our panellists here was that these figures weren't higher.

FIG. 10 What proportion of your enterprise data do you currently store or process in the public cloud, and what do you predict will be the situation in 3 years' time?

Now	21%
In 3 years	39%

> ❝ At the moment most of it is in our data centres but I'd be very surprised if I ever build another data centre – why would I? The larger data centres are more efficient, actually more secure, more reliable… IT director, Textile services

> ❝ I think a lot hasn't gone yet, but it will be going because 99 per cent of businesses are not in the business of running data services… Business consultant

> ❝ You can do some fantastic things with cloud-based servers. You can script everything, run a script which will set up a service, pull in data, expand the service depending upon how much data there is, process stuff, output some data and then it will collapse and destroy itself. So you're not paying for a service to run 24 hours a day; you're only paying for it for an hour. So that is incredible… the cost of running a data centre is immense… Head of data strategy, Media

> ❝ Twelve years ago I built a system for complex financial instruments in-house and then we did the platform for the parallelisation of it because the algorithm has got to be capable of that, but we were then also limited by what physical kit we had. Now it's like, boom… Service delivery manager, Banking

The major players have all rushed to add IoT services to their offerings. Example include IBM Watson, Oracle IoT Cloud Service, GE Predix, Microsoft Azure IoT Suite and AWS IoT.

The answers of our survey participants to the question, "Do you use any IoT-specific cloud platforms or do you plan to?", reflected the market dominance of Amazon, and perhaps brand recognition, with 39 per cent of respondents using or planning to use AWS IoT; 30 per cent said they use or plan to use IBM Watson.

Our panellists expressed the usual concerns about data governance and security of cloud services but it was also clear that, increasingly, solutions to these traditional challenges are being found.

> " Regulatory and also infrastructure, there are solutions that will allow you to place all your data encrypted in an infrastructure whereby you're disassociating your data from the infrastructure…
>
> Data manager, Research

> " In the health sector, patient information is kept in a very secure environment and obviously those who gain access to that data [are heavily scrutinised]. That data is not available for any other person outside of this environment and if it goes outside it is completely anonymised…
>
> Business intelligence manager, Health

8 Data scientists

Drawing insights from big data is the whole point of the exercise. Crucial to this process is having the right people on board – specifically data analysts and scientists. The difference between the two was an area that *Computing* wanted to explore in order to derive insights on the skills required.

> " I think there is a fundamental difference between the two: a data analyst is somebody that looks at data and derives insights from it. A data scientist is somebody who is able to set up a process to look at data and derive insights in a procedural manner as opposed to personal analysis and personal digestion of the data… CTO, Oil and gas

> " We have a dedicated data science team and they do get paid an awful lot more than our analyst team. Their distinguishing feature is that they nearly all have degrees or doctorates in mathematics. They need to be able to interpret highly complex results and make choices depending upon those results and apply new and interesting maths formulas – interesting to them but impenetrable to anybody else…
>
> Head of data strategy, Media

"The crux of the difference between the two roles in minds of many of our research participants came down to the issue of creativity – the ability to try to identify unknowns"

When asked what the most important skills were for both data analysts and data scientists, our survey respondents gave a strong indication of the differences in the two roles. For a data analyst the most desirable skills were (in ascending order for importance) presentation skills, experience in real-time analytics and an ability to work in multi-disciplinary/collaborative teams.

The equivalent positions in the data scientist rankings were an ability to extract insights from different types of data, the ability to identify problems as well as their solutions and an ability to think creatively using the data. The crux of the difference between the two roles in the minds of many of our research participants came down to the issue of creativity – the ability to try to identify unknowns.

> A data scientist is literally a rocket scientist. They have the ability to think creatively using different points of data, crunch that data and come out with solutions. Analysts are people who understand the business and are able to use existing tools to forecast opportunities. They are not the same thing and are not interchangeable… Director digital, analytics and innovation, Professional services

> The business expectation is that you go to an analyst with a problem, whereas with a data scientist, it is more open-ended: 'we having a problem, go and find it without specifying it'. Finding the solution before specifying the problem. Even the organisation who is asking the data science team might not know what the problem is… Big data architect, Technology

> I think a data scientist is more about innovation, new methods. He's not looking at already present data, he is looking at new sources, new methods, new processes; he is going to invent something… Data manager, Research

The problem is that there really are not enough of these individuals to satisfy demand – which was leading to further creativity with the presentation of candidates.

> If they can get themselves jobs on a hundred grand a year because they convince somebody they're a data scientist, good luck to them. I think data science is a genuine, valid concept but it's what I call a Superman, Superwoman concept. So what it means is you're looking for unicorns and many people are struggling to find the unicorns because there's not many of them out there… CIO, Environmental services

> You put data science on your CV and you take a 20 per cent pay rise pretty much immediately… Director digital, analytics and innovation, Professional services

> People calling themselves data scientists are really analysts and people calling themselves analysts are more report builders. Everyone upgrades themselves one upper notch from what they really are… CTO, Technology

Our research indicates that the general dissatisfaction shown with the job market for data scientists and analysts may well be a factor frustrating the crossing over of big data/the IoT into the mainstream.

9 Conclusions – what makes a leader?

Computing chose to combine its research on big data technologies and IoT this year because it became clear during the course of research last year that it was becoming difficult to disentangle one from the other, at least from a data point of view. While big data as a whole is more mature than IoT, part of the growth of big data technologies has been fueled by IoT and this will continue. In turn, big data technologies such as Hadoop and Spark mean that the vast streams of unstructured data that IoT is generating can be analysed in real-time. Big data and IoT need each other for each to reach its full potential – but much of *Computing*'s research this year suggests that we are at or very near the tipping point.
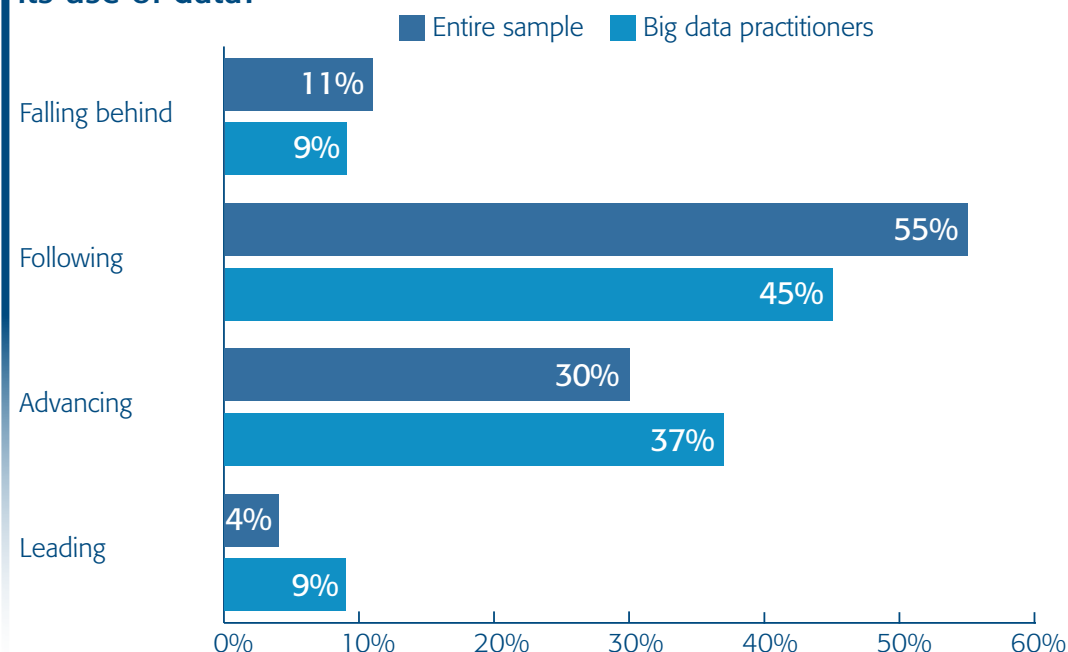
In less than 12 months, the proportion of our survey respondents actively preparing for IoT has more than doubled with significantly greater urgency being accorded to preparations. The rankings of importance of IoT to businesses have also increased.

Businesses may be focusing on the operational aspects of IoT rather than the analytical possibilities, and many are still dealing with the logistics of data gathering and storage before even thinking about analytics. Nonetheless, the changes that our research has recorded in less than a year suggests that this is a revolution in its infancy.

> "All of the stars are aligning – data, digital and devices. All of those things are coming together…"
> (Director digital, analytics and innovation, Professional services)

FIG. 11 Which best describes your organisation in terms of its use of data?

Legend: ■ Entire sample ■ Big data practitioners

Category	Entire sample	Big data practitioners
Falling behind	11%	9%
Following	55%	45%
Advancing	30%	37%
Leading	4%	9%

In terms of their use of data, most organisations would class themselves as 'following' – in other words, they are keeping up but not being particularly innovative. As might be expected, those organisations that classify themselves as big data practitioners are more likely to found towards the 'advanced' end of the spectrum (Fig. 11).

Not all organisations can be at the bleeding edge of technology – there is nothing inherently wrong with being a follower. Also, an overall status of 'following' can conceal pockets of more advanced data analytics within the same organisation – particularly if it is a large, distributed one.

> **In terms of the National Health Service, it's a combination of advanced and trailing as well. Some of the technology is quite old; some of the technology is quite new…** Business intelligence manager, Health

Traditionally conservative banks often have so-called 'internal startups' with free rein to experiment with data, and they often employ the best data scientists that money can buy.

So what makes a leader? More than half of our survey (53 per cent) put a clear data strategy at the top of the list (Fig. 12).

FIG. 12 What are the main attributes/factors that enable an organisation to go from being a follower to a leader in terms of use of data?

Clear data strategy	53%
Robust governance	37%
Focus on speed to market	32%
Analyse IoT data	32%
Focus on innovation	22%
Measurable RoI	21%
Full buy-in by organisation	20%
Making the right technology choice	19%
Doing real-time data analysis/streaming	18%
High proportion of data-based decision making	16%

> **The main one, I would say, is to have a very clear data strategy. Have a published route map of what you're doing with data, how you're storing it, where you're putting it and who is having access to it, how you're anonymising it and how are your governance issues…** Head of data strategy, Media

Not everybody was convinced though. The difficulty of planning and executing a strategy in the face of 'unknown unknowns' was real for some of our panellists.

> **I'm surprised by the strategy because I still get the feeling, at least where I am, that there's still an element of 'there's an answer out there somewhere but we don't quite know where'**… IT director, Textile services

Whether organisations categorise themselves as leading, advancing or following, the proposition that these technologies are at tipping point is strengthened by the fact that 46 per cent of respondents were more likely to initiate an IoT project in the next 12 months than not.

Furthermore, almost 80 per cent of those who had already started small projects were planning to increase their investment. This indicates satisfactory returns, which is just as well because expectations are high. Fifty-five per cent of those surveyed expected projects to increase revenues by 15-20 per cent. Approximately one fifth of the survey had even greater expectations.

When it came to live projects, data volume was still the factor confounding the greatest number of our respondents, suggesting strongly that big data and IoT projects were still at their early stages in a majority of businesses. However, the veracity of data was the second biggest concern. It is likely that as big data matures in organisations, volume will be less of a challenge but the problems of veracity are likely to grow along with the profile of IoT.

Applied analytics was a major factor in the identification of organisations as being 'leading'. Simply gathering huge quantities of data does not a leader make.

> **Leading: People looking at streaming and real-time analytics and who are actually looking to do machine learning. Building genuine machine learning models and using all of that… People who are using all of the latest brand-new technologies that are coming out is another sign of a leader**
> CTO, Technology

> **Event streaming and real-time as the mechanism for movement of data is becoming more and more prevalent and more in terms of being a leader, it's doing real-time analytics on top of that…**
> CTO, Technology

> **Leading with machine learning: organisations that have found a use case for some of the non-standard unsupervised machine learning techniques**… Analytics consultant, Financial services

Our research this year uncovered a significant gap between the amount and type of data being captured and stored and the analysis (or rather non-analysis) of that data. Internally generated data was more likely to be gathered but not at all likely to be analysed. Data from external sources was far less likely to be captured but data that was captured was more likely to be analysed. Our survey detected more of a keenness to analyse data gathered from external sources in the future than internal.

Leaders are also far more likely to be making data-driven decisions – or at least advancing strongly in that direction. Only 28 per cent of our survey said that a majority of their decisions were based on data. However, while the human aspect of decision making is still prevailing, data is increasingly being used to at least inform and justify those decisions.

It is an interesting finding that leaders are not necessarily very technology driven. As Figure 12 shows, technology based attributes such as analytics come fourth on the list of what constitute leaders. Other technology driven attributes such as making the right technology choices and using real-time analytics/streaming come in at eighth and ninth on the list respectively.

Certainly, while analytics was the class of technology desired by the largest proportion of our survey respondents, the technologies showing the steepest growth curves were the reasonably well established Hadoop and the newer NLP. NLP in particular was expected to grow faster than any other big data technology. When it came to technologies pertaining to the IoT, blockchain was that for which expectations were highest – although this may have something to do with the amount of press that immutable ledgers have been getting of late.

The amount of data being stored or processed in the cloud is also expected to increase – from 21 per cent to 39 per cent in three years. These figures to some degree reflect the enormity of Amazon's success in particular. Thirty-nine per cent of respondents were either already using AWS IoT or were planning to do so.

Is the employment of data scientists a differentiator for leaders? Our survey says not in itself, although it's probably fair to say that a business seeking to employ data scientists is likely to have the clear data strategy, robust governance and the other factors that do differentiate. Hiring data scientists is a clear indication that an organisation wishes to use data to explore the unknown unknowns.

> **It has been identified that hiring a data scientist and getting people doing the right roles is really critical along with investing in the right technology…** Director digital, analytics and innovation, Professional services

> **Having a data science team doesn't make you a leader in the market; it's exploring the unknown unknowns that make you a leader…** Big data architect, Technology

The challenge is the scarcity of such individuals. The creativity and innovation embodied in data scientists are thin on the ground. Our research uncovered some cynicism and dissatisfaction with the recruitment of data scientists (and indeed data analysts) and this may well be a factor impeding faster growth of big data/IoT.

It is clear that there are still practical, real-world challenges to the mainstream adoption of big data technologies and, particularly, IoT. However, both have come a long way and our research does suggest that there is a strong desire in many organisations to overcome these challenges, given the potential rewards for using them.

In many organisations, the understanding of the relevant areas such as the platforms, automation and the devices themselves have been developing almost in isolation from each other. Our research this year seems to indicate that these paths are converging – and that the next three years will see these technologies finally become established. The impact on society is rather more difficult to predict.

> **All of the stars are aligning – data, digital and devices. All of those things are coming together and we are now beginning to understand machine learning, we are now understanding automation. Previously all these things were disconnected in different ways; we've now finally, this last year or so, started connecting those things together and we are now waiting for the sensors. We are about to hit that crossroads and we will probably hit it in the next three years…** Director digital, analytics and innovation, Professional services

> **A lot of companies are still gathering data; they are trying to get enough to gain a critical mass so they can get a use case for the data being valuable. It is often within their five-year strategy; therefore they need to position themselves so that they're able to utilise big data and IoT technologies…** Analytics consultant, Financial services

> **For me, if we're not at a tipping point in the technology, we're very close to it. Very large amounts of data are coming in at very high velocity and historically it's been quite hard to put together systems that can analyse it as quick as you need it to be analysed. But I think we're just starting to see people canning the solutions so you can go out and you can buy something that will get you close to where you want to be…** IT director, Textile services

About *Computing*

Computing is published by Incisive Business Media, one of the world's leading B2B information providers and the publisher behind some of the UK's most well respected IT media brands, most notably *CRN*, *The Inquirer*, IT Hound and *V3.co.uk*.

Incisive Business Media's database of hundreds of thousands of IT decision maker subscribers puts it in an unrivalled position to provide valuable research and commentary on the latest business IT purchasing trends and processes.

For more information about our research, surveys and content creation services please call Alan Loader, publisher, on (+44) 207 316 9733 or John Leonard, research director, on (+44) 207 316 9776.

computing
research