techUK representing the future

Understanding, Demystifying and Addressing the UK's Big Data Skills Gap

October 2016

Contents

Introduction

Big data and data analytics are vital to the UK's digital growth, expected to be worth £241 billion to the UK economy by 2020 creating 157,000 additional jobs[1]. However this will only be possible if we have people with the right skills to fill key big data roles. Without the right talent pool, the UK's ability to capitalise on the big data revolution will be constrained.

The UK's digital skills gap is already costing the UK £2bn per year[2] due to the inability of employers to fill key digital skills roles. The big data and data analytics sector stands to lose the most given it is expected to count for the largest proportion of UK digital vacancies following a survey of techUK members which found that 62% will require more big data capabilities over the next five year[3].

techUK Big Data Hero, Alison Lowndes, Deep Learning Solutions Architect, Nvidia Ltd:

"We need more girls to study STEM subjects and to understand how much they can offer industry from an engineering and computer science standpoint."

"Code should be taught as prolifically as reading to open the doors to the digital age for everyone".

techUK's Big Data Hero, Michael Comerford, Data Scientist, Agile Solutions:

"I came to my current position following a fairly unconventional route. I have an academic background in sociology and a PhD in computing science, but I also made a short detour into Russian history along the way too.

As a self-taught programmer and data wrangler I picked up technical skills working on various side projects before completing my studies, and then working for the Urban Big Data Centre. There I led on initiatives including the integrated multimedia city data project (iMCD) which looked at how citizens interact with their environment using lifelogging, GPS tracking and survey data.

We need to encourage people that don't see themselves as 'IT' to get involved more. Of course developers and programmers are vitally important but if we want to move the big data agenda forward, in our own organisations and for our clients, we need big data to be something everyone can have a stake in."

To ensure the UK seizes the opportunity to lead the world in data driven economic growth we need to understand the specific skills and capabilities that are needed. Requirements go beyond the simple definition of data scientists.

This paper identifies the skills required across a big data supply chain and highlights areas where there is the biggest skills deficit. It then makes a number of recommendations relating both to the domestic skills pipeline and the UK's continued ability to attract some of the best talent from around the world.

1   http://www.sas.com/content/dam/SAS/en_gb/doc/analystreport/cebr-value-of-big-data.pdf
2   O2 and Development Economics research (2013). 'Three quarters of a million digitally-skilled workers needed to power UK economy by 2017. Retrieved from http://news.o2.co.uk/wp-content/uploads/2013/09/The-Future-Digital-Skills-Needs-of-the-UK-Economy1.pdf
3   Data drawn from private techUK Members Survey, October 2014

Roles and Skills required across the Big Data supply chain

Who is involved in the Big Data supply chain?

A wide range of skills and specialisations are required to deliver a successful big data strategy. We have provided a definition of big data and data analytics in appendix A. These requirements go beyond just data scientists. techUK has identified eight key roles involved in putting in place a big data strategy.

1. Chief Data Officer
2. Data Infrastructure Engineer
3. Data Integration Engineer
4. Big Data Developer
5. Solutions Architect
6. Data Scientist
7. Data Analyst
8. Visualisation Expert

While this paper explores the specific roles and skills sets involved in each of these roles, it is important to remember that all big data strategies are different, as are their implementation. This is not a definitive list of the individuals that every organisation must hire. Sometimes not all of these roles will be necessary. For example, one individual may perform more than one of these roles.

This list, and the job descriptions outlined below, are designed to help illustrate the depth and breadth of the skills needed in the delivery of a big data project. The way these skills are integrated into an organisation will be unique in each case. This outline simply provides a way to consider and determine where there is a big data skills gap which needs to be addressed.

Data Protection Officers (DPOs)

It should be noted that under a new requirement in the European Commission's General Data Protection Regulation (GDPR)[3] many organisations will need to appoint a Data Protection Officer (DPO). The DPO will require relevant skills to ensure compliance with data protection legislation.[4] While this will be crucial to the development of big data strategies, the DPO will not require technical big data skills and therefore has not been included in our list of big data skills and roles. techUK would recommend that any organisation implementing big data strategies ensures that their DPO is involved in the design of a project to ensure compliance. The expected responsibilities of DPOs is still developing. Similarly all those listed above should be familiar will responsibilities under the GDPR to ensure compliance.

---

3   http://www.techuk.org/insights/news/item/6842-how-will-new-eu-data-rules-impact-my-tech-business

4   http://www.allenovery.com/SiteCollectionDocuments/Radical%20changes%20to%20European%20data%20protection%20legislation.pdf

What skills are required across these roles?

More important than the roles identified above are the different skills which contribute to the big data and data analytics ecosystem.

Below we identify the skills required across these different roles in order to fully understand the skills needed across the big data supply chain. We can then attempt to assess which skills may be more urgently needed than others.

Chief Data Officer

The job role of Chief Data Officer is relatively new and is still evolving. CDOs have at least some responsibility for information governance including security and privacy, and most also for the analytics functions. A small minority have only governance as their responsibility.

Similar themes across existing Chief Data Officer roles involve creating and staffing new information governance and data quality management functions, managing data as a corporate business asset, from a foundational level (master data management) and bridging data silos for enterprise wide use.

They are responsible for implementing a strategy around information governance, enterprise analytics, information architecture and data assets as well as supporting strategic corporate business goals by improving data quality for broader corporate use and addressing customer needs (internal and external). They should report to the board of the business.

Typical background: Business

Solutions Architect

Solutions Architects are required to bridge the gap between the business problem and the big data solution which meets the client's needs. They must work with the whole big data team to ensure the right architectural style for the system being developed, select the right data architecture and analytical approach, select the right technologies to build it from and define the key decisions and architectural principles to guide its development and operation. Solution architects are highly technical engineers, with strong knowledge across a range of technologies and practices while also having deep knowledge of one specialisation. They typically, but not exclusively, come from a software development background. They quite often also have a specific domain specialisation to understand the client problem effectively.

Typical background: Software and Engineering

Data Infrastructure Engineer

The infrastructure engineer builds reliable, scalable, usable big data platforms and install and manage tools such as Hadoop, Storm, and Spark.

This role requires expertise with new (and sometimes relatively unstable) data infrastructure such as Cassandra, Hadoop and Spark, as well as more recognisable technologies like Ab Initio, MS SSAS, Linux and networking. These people are involved in building reliable, scalable, usable platforms for the others to work on.

Given the rapid pace of change in big data technology, Infrastructure Engineers must constantly adapt and update their skills.

Typical background: Computer Science and High Performance Computing

Data Scientist

Data Scientists need to have programming, analytical, statistical, mathematical, predictive modelling skills as well as business strategy skills to build algorithms which answer business needs.

A Big Data Scientist understands how to integrate multiple systems and data sets. They need to be able to link and mash up distinctive data sets to discover new insights. This often requires connecting different types of data sets in different forms as well as being able to work with potentially incomplete data sources and cleaning data sets to be able to use them.

Data Scientists need to be able to program in different programming languages such as Python, R, Java, Clojure, Matlab, Scala, Pig or SQL. They need to have an understanding of Hadoop, Hive and/or Spark. They need to be familiar with disciplines such as

Natural Language Processing
Machine learning
Conceptual modelling
Statistical analysis
Predictive modelling
Hypothesis testing

Data Scientists also need to understand how businesses operate and be able to communicate their findings, orally and visually, working as part of a wider domain team.

Typical background: STEM

Big Data Developer

Big Data Developers must be highly analytical with a structured way of thinking and be able to apply Hadoop, Cassandra, HBase, MapReduce, Pig, Hive, Storm and other appropriate big data/NoSQL technology to solve a wide variety of different big data problems. They must be the Hadoop ecosystem expert on the team and have the ability to identify problems, debug issues and implement solutions. They should be a proactive thinker with a proven ability to research and implement high performance data management solutions.

This role requires a deep understanding of Java MapReduce implementation and other Hadoop ecosystem tools as well as a broad understanding of big data and NoSQL technologies. They should have strong expertise in using relational technologies in highly concurrent applications and experience with the production of Hadoop based implementations is required. The developer will be familiar with the Hadoop ecosystem of products like HBase, Hive and Pig and must be aware of ETL patterns as well as having strong programming skills.

Typical background: Computer Science and High Performance Computing

Data Analyst

Data Analysts complete a range of highly analytical tasks such as identifying trends, investigating correlations, understanding drivers of customer experience, creating management account and information, study key metrics, and provide and maintain analytic dashboards. On occasion they are also responsible for maintaining data quality, and quality assurance. Most importantly, they are the main link between the rest of the organisation and the data team, and they perform all the reporting within the organisation.

This role needs to be able to articulate complicated concepts and results to senior stakeholders, and they need to be consultative in terms of interpreting the requirements of the business into the analytics team.

Typical background: Business, Economics, Accountancy, Marketing

Data Integration Engineer

Data Integration Engineers analyse external data sets and identify the most appropriate mechanism to import them. They must liaise with Data Scientists and Data Infrastructure Engineers so the big data strategy works as a whole. They design metadata models, data quality measures, test and support data feed software. Once these systems are designed there needs to be ongoing data quality monitoring and the routine reporting of quality and trends to Data Scientists.

This role requires a deep understanding of big data platform technology and big data oriented storage technologies as well as strong understandings of relational storage technologies and ETL technologies. Software development skills across a range of relevant technologies are necessary along with strong software engineering skills into analysis, design, automated testing and continuous delivery.

Typical background: Computer Science and High Performance Computing

Data Visualisation Expert

The Data Visualisation Expert needs to be versed in programming and data analysis, but does not have to be an expert. As more and more data is collected and analysed, decision markers at all levels welcome data visualisation software (such as Tableau, Qlik and Alteryx) that enables them to see analytical results represented visually, in pictorial or graphical form, finding relevance among the millions of variables and communicate concepts and hypotheses in an easy to understand manner. Visualisation is crucial within big data as it presents data in a way that decision makers can easily interpret and allows for accelerated insight.

Data storytellers have also been abundant in journalism, where communicating numbers and data has traditionally been an important task. Today, more companies are seeking individuals with these skills, to support the translation of the results of the data team to the board, the rest of the organisation and to external stakeholders.

Typical background: Depends on use-case, diverse

As previously mentioned, these role descriptions are not designed as a prescriptive list of individuals that every organisation implementing a big data strategy must hire. Many of these roles are senior and individuals will not begin their careers with these job titles. The purpose of these definitions is to identify the specific skills involved in the delivery of a big data project.

It should also be noted that along with the technical skills outlined above, domain expertise and knowledge of the organisation implementing a big data project will be required across those involved in the project.

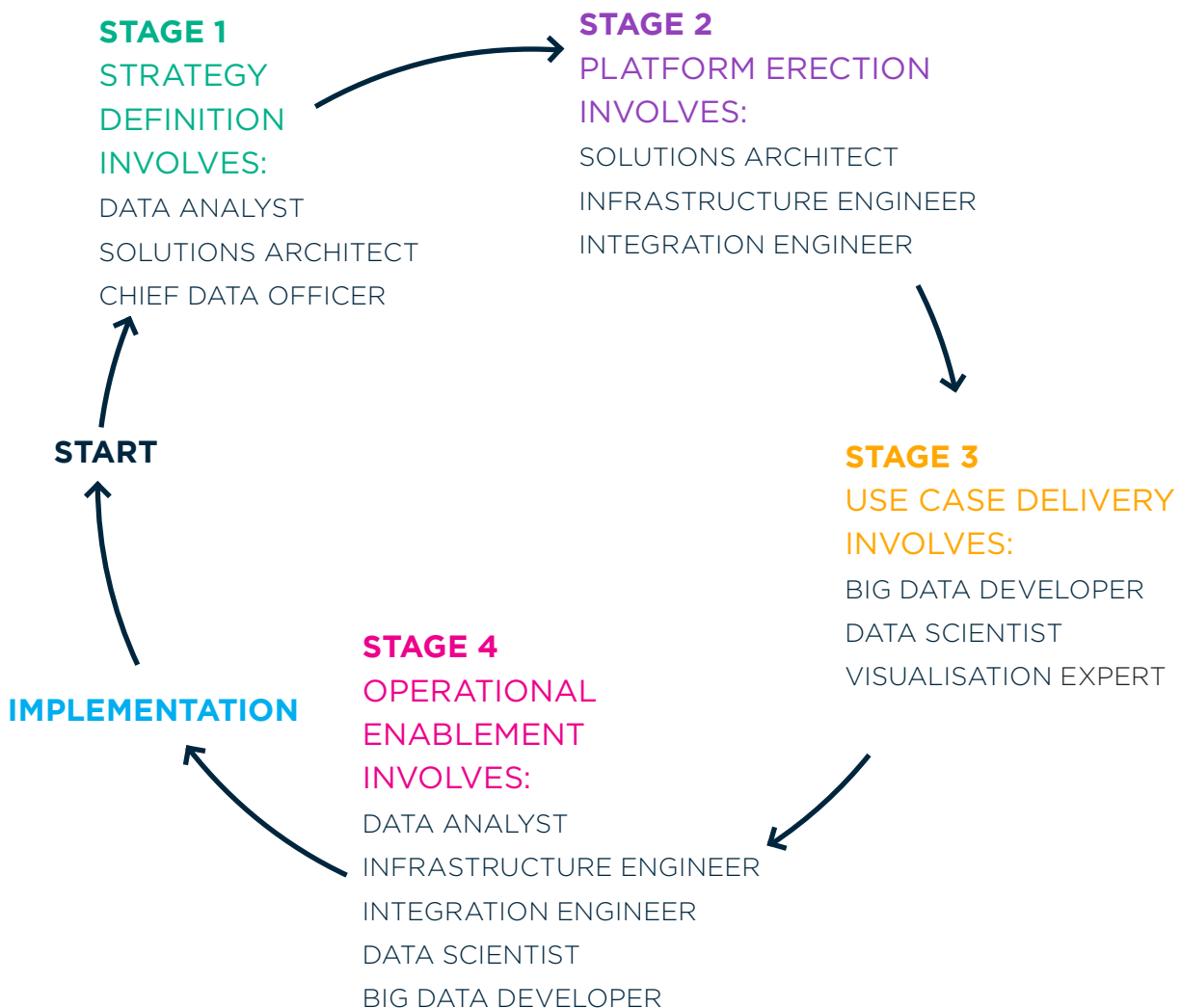## Where do these roles fit into a Big Data project?

The image below maps how all these different roles work together from when the decision is taken to implement a big data project or strategy, to its successful implementation. It highlights how all the roles we have defined above contribute at different stages of the project for the successful delivery of a big data solution.

techUK Big Data Hero, Hal Bertram, Data Visualisation Expert, ITO World.

"Coming from film and animation myself, I think we need to recognise the diversity of skills in the sector and the possibilities the future could hold."

techUK Big Data Hero, James Hodge, Principle Product Manager, Splunk

"The big data landscape at the moment can feel like a bewildering plethora of technologies, integrations, software vendors, algorithms and use cases. I think we need more work in the industry to demystify what a career in big data is. This is where organisations like techUK are doing some great work to listen to industry and then engage with people trying to understand wat happens within a career in big data."

STAGE 1
STRATEGY
DEFINITION
INVOLVES:
DATA ANALYST
SOLUTIONS ARCHITECT
CHIEF DATA OFFICER

STAGE 2
PLATFORM ERECTION
INVOLVES:
SOLUTIONS ARCHITECT
INFRASTRUCTURE ENGINEER
INTEGRATION ENGINEER

START

STAGE 3
USE CASE DELIVERY
INVOLVES:
BIG DATA DEVELOPER
DATA SCIENTIST
VISUALISATION EXPERT

STAGE 4
OPERATIONAL
ENABLEMENT
INVOLVES:
DATA ANALYST
INFRASTRUCTURE ENGINEER
INTEGRATION ENGINEER
DATA SCIENTIST
BIG DATA DEVELOPER

IMPLEMENTATION

Where does the Big Data skills gap lie today?

Now we understand the varying skills necessary for the UK to capitalise on the big data revolution we can determine which of those skill sets we are currently lacking.

This is a difficult task and there may be no completely comprehensive way of achieving a complete picture of the skill sets available in the UK. However, job search website Adzuna has compiled information from almost 5000 unfilled vacancies in May 2016.

This snapshot from May 2016 shows where the most vacancies existed across the eight key roles identified by techUK, in order of most in-demand skill set:

techUK Big Data Hero, Katie Russell, Head of Data Science, ONZO

"I think a lot of people are excited about pursuing a career in data science….but there's a gap in some big data careers which are adjacent e.g. dedicated Scala Engineers, Data Engineers. They're genuinely interesting and exciting roles, but people need to talk about why they're exciting."

Role	Number of advertised vacancies
Data Analyst	2,177
Data Infrastructure Engineer	927
Solutions Architect	926
Data Scientist	485
Big Data Developer	147
Data Integration Engineer	124
Visualisation Expert	14
Chief Data Officer	14

This information provides some interesting insights. While we have been led to believe that it is Data Scientists that are most in demand in the UK, we can see that it is actually Data Analysts, Data Infrastructure Engineers and Solutions Architects which are in higher demand.

Brexit

The UK's big data market has thrived on its ability to attract the best data skills talent from across Europe. Following the UK's EU Referendum vote we must ensure international talent in the UK continues to feel welcome and valued. If the UK can no longer benefit from free movement, a new 'smart migration' system will be needed that prioritises the needs of high growth sectors, such as big data and data analytics. Following Brexit, action must be taken on skills as otherwise the big data and data analytics industry will struggle to achieve its potential.

How to address the Big Data skills gap

It is now clear that action must be taken to address the UK's big data skills gap to fulfil the skills requirements across the big data supply chain, beyond that of Data Scientists. This is both a short and long term issue which needs to be tackled now. For the UK's digital economy to thrive it is essential that businesses are able to recruit Data Analysts, Data Infrastructure Engineers, Solutions Architects and Data Scientists.

Immediate action is required to tackle the shortage we face today if the UK is to capitalise on the immediate opportunities of the big data revolution. Where there is a gap in national capabilities, businesses must be able to recruit internationally which will require a new smart approach to skilled immigration.

techUK has repeatedly called on the UK to adopt a 'smart migration' approach which would allow firms to access the international talent they need to grow. However, as seen in the approach taken to the new rules for Tier 2 skilled migration, the introduction of extra restrictions such as the new skills charge, increased salary thresholds and limitations on intra-company transfers, will not make it easier for companies to access the talent they need to grow[5].

techUK's was pleased to see the Government's independent Migration Advisory Committee (MAC) recommend to Government that Data Scientists be added to the Government's Shortage Occupation List. However, given that the big data roles needed are Data Analysts, Data Infrastructure Engineers and Solution Architects these roles should feature prominently on any future Government preferred shortage occupation list. Adding these big data specialists to the Shortage Occupation list alone will not solve the talent shortages faced by tech companies.

If the UK is to reap the rewards of the big data revolution in the future then the long term issue of a domestic talent pipeline must also be addressed. This requires an approach that puts young people on a trajectory towards these roles from a young age.

techUK Big Data Hero, James Hodge, Principle Product Manager, Splunk

"The big data landscape at the moment can feel like a bewildering plethora of technologies, integrations, software vendors, algorithms and use cases. I think we need more work in the industry to demystify what a career in big data is. This is where organisations like techUK are doing some great work to listen to industry and then engage with people trying to understand wat happens within a career in big data."

The new computing curriculum in England was welcomed by the industry as a way of engaging children with tech from a younger age. For the UK to guarantee a future pipeline of talent in big data and data analytics both programming and analytics should be introduced even earlier in the school curriculum. Analytics skills are currently introduced at Key Stage 4 which is non-compulsory. Introducing it earlier in the curriculum, and at a point where pupils are guaranteed to encounter analytics skills, will help get school pupils interested in these skills and provide opportunities to develop them from an early age.

Existing schemes such as the Apprenticeship Levy must also be made to work for big data analytics. In its current form, it is not an effective route to re-skill existing individuals in the workforce into data analytics roles, but it can be a great stepping stone for first-time employees.

In order for employers to use the Apprenticeship Levy to reskill existing employees in data analytics, it would require a large commitment from both the employer and the individual trained. The process would take 24 months, and training alone would cost £18,000. This excludes the cost of administration and the wage of the employee. Therefore, it is an expensive and time-intensive method for employers who are considering re-skilling staff into data analytics roles. Employers may seek alternative methods which are less expensive and time-consuming.

However, the Apprenticeship Levy can work for school-leavers or first-time employees. The Level 4 Data Analytics standards would effectively train first-time employees in the basics of the role, including training in collecting, organising and studying data to provide business insight. The apprenticeship standard would be a great stepping stone into more technical work in the area.

In order for the data analytics sector to best benefit from the Apprenticeship Levy, the policy must work to ensure there is more uptake of higher-level apprenticeships from a younger age. Most of the apprenticeship growth in the last four years came from those aged over 24 and in apprenticeships at Level 2. While the sector looks forward to working with the Government on the continued development of the Levy, there is a consensus that the there is considerable work to be done to ensure the Levy works for, not against, the needs of the most dynamic and innovative sectors of the UK economy.

---

5   http://www.techuk.org/insights/news/item/8161-reforms-to-tier-2-visa-announced-by-government

There is also work for industry to help develop big data and data analytics skills by up-skilling existing workforces. While there is a need for more STEM graduates to fill future big data roles, there are plenty of skills already developed by individuals that with some tweaking and re-training could be used within big data and data analytics.

There are a number of options for up-skilling within industry.

- There should be opportunities for reverse-mentoring, which has proven to be successful in other areas[6].
- Within organisations it is possible to bring non-analytical employees together with analytics specialists to work on projects. This would allow analytics skills to be developed by others in the organisation as well as providing the domain expertise needed in big data projects.
- There are also opportunities for individuals within an organisation to learn additional analytics skills from each other, developing expertise in different areas. BUPA has created a global community of 300 data analysts to share ideas[7], which will also contribute to upskilling.
- A significant number of the existing workforce will have graduated with advanced STEM knowledge but followed different career paths. Individuals with an interest in data science or analytics, and an appropriate understanding of statistics and mathematics can be re-trained to become data analysts or scientists. For larger organisations this could be done through internal training schemes, or for smaller companies an external training provider might be more appropriate. Where appropriate, organisations can use their Apprenticeship Levy funds and retrain existing employees within the "data analyst" apprenticeship standard, however the organisation will need to ensure that the specifications of the standard meet the skills needs of their organisation. However, the skills needs of some organisations may fall out of the scope of existing apprenticeship standards. If this is the case, organisations who need these skills can become "trailblazers" and create a new standard aimed at re-training people in highly technical data skills.

A softer-touch approach to help guarantee future big data and data analytics skills is by highlighting potential career opportunities and positive examples of current industry leaders. Both Government and Industry have a role to play here. If the positive uses of big data to the UK economy and society are highlighted then a career in big data could be a more attractive prospect for young people when they are deciding which skills they should be developing. Additionally if they see role models, leading the way in big data, there will be greater inspiration to follow in their footsteps. techUK have been highlighting industry leaders through the Big Data Hero campaign. Some of those heroes have been quoted throughout this paper.

6   https://blogs.cisco.com/diversity/the-results-how-reverse-mentoring-can-enhance-diversity-and-inclusion
7   http://www.computerweekly.com/news/2240242165/CIOs-are-ill-prepared-for-data-driven-business

techUK Recommendations

This paper has identified the eight key job roles and skills needed to implement a big data project, where the current skills gap exists and potential threats to combating the gap.

For the UK to fully benefit from the big data revolution techUK makes the following recommendations:

1) **Data Analysts, Data Infrastructure Engineers and Solution Architects should feature prominently on any future Government preferred shortage occupation list, joining Data Scientists.**
   o In light of the UK's decision to leave the European Union, the Government must determine the best method to amend the migration system to ensure the UK still attracts talent it lacks domestically.
   o A new 'Smart Migration' policy is needed which allows firms to access the international big data talent they need to grow.
   o Big data and data analytics experts must feel they are welcomed and valued in the UK.

2) **The UK's existing tech workforce should be upskilled.**
   o techUK suggests a model of reverse mentoring in industry, whereby young skilled IT professionals teach existing staff new skills.
   o Where re-training requires skills that fall out of the scope of existing apprenticeship standards, organisations can become "trailblazers" and create a new standard aimed at re-training people in highly technical data skills.

3) **The Department for Education must work to ensure there is more uptake of higher-level apprenticeships from a younger age with the forthcoming Apprenticeship Levy.**
   o In its current form, the Levy is not an effective route to re-skill existing individuals in the workforce into data analytics roles, but it can be a great stepping stone for first-time employees. Therefore, more must be done to focus on these higher-level skills.

4) **Programming and Analytics should be introduced earlier in the school curriculum.**
   o As well as being able to attract international talent, the UK must grow its domestic talent pipeline to ensure we have a future generation with the right skills to continue the innovation necessary to maintain the UK's reputation as a big data and data analytics leader.

5) **Promote the value & importance of Big Data and Data Analytics.**
   o Industry and Government should do more to explain the contribution of big data and data analytics to the UK's economy and society in order to inspire future industry leaders.
   o techUK's Big Data Hero campaign is an example of the positives stories that can be told.

Conclusion

The UK has an amazing opportunity to be a world leader in the field of big data and data analytics. However this will only be possible if action is taken to address both short and long term skills requirements. Through a combination of attracting the best international talent, and growing a domestic talent pipeline, the UK will be able to ensure it has the skills required across the big data supply chain. Doing this will allow the UK to benefit from the big data revolution today and well into the future.

Appendix A

Definition of Big Data:

'Big data' is a term used to describe very large data sets used and analysed to reveal trends and patterns. Big data has five defining characteristics: volume, velocity, variety, variability and complexity[8]. It includes both 'structured' (such as traditional databases or spreadsheets) and 'unstructured' data (such as photos, videos, and social media updates). While personal identifiable information can be involved in big data, not all big data will be personal data. Big data will also involve technical data, such as traffic or 'meta' data as well as anonymised data.

Definition of Data Analytics:

Data analytics is the examination of raw data through qualitative and quantitative techniques and processes to uncover hidden patterns, unknown correlations and other useful information and draw conclusions and drive decisions and actions to make better business decisions. According to IBM there are five types of data analytics tools; predictive, prescriptive, descriptive, diagnostic and cognitive[9][10].

---

8   SAS (accessed August 2015). Big Data – What it is & Why it Matters. Retrieved from http://www.sas.com/en_us/insights/big-data/what-is-big-data.html

9   IBM (accessed August 2015). Analytics Technology – overview. Retrieved from http://www.ibm.com/analytics/us/en/analytics-technology/

10   SAS (accessed August 2015). Big Data Analytics: What it is and Why it Matters. Retrieved from http://www.sas.com/en_us/insights/analytics/big-data-analytics.html

techUK represents the companies and technologies that are defining today the world that we will live in tomorrow.

The tech industry is creating jobs and growth across the UK. In 2015 the internet economy contributed 10% of the UK's GDP. 900 companies are members of techUK. Collectively they employ more than 800,000 people, about half of all tech sector jobs in the UK. These companies range from leading FTSE 100 companies to new innovative start-ups. The majority of our members are small and medium sized businesses.