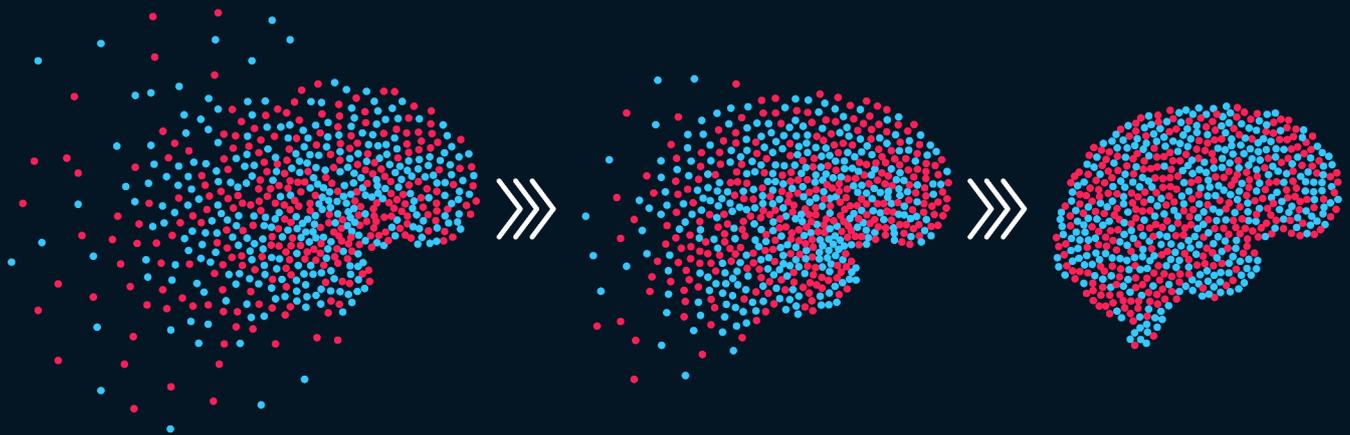


December 2018



Submission to PwC Skills for Australia

# BIG DATA in VET



---

Coder Academy is a part of the Academy of Information Technology

[www.coderacademy.edu.au](http://www.coderacademy.edu.au)

# Coder Academy Submission on Cross Sector Big Data Project

---

## Table of Contents

---

- **Introduction**
- **What is Coder Academy?**
- **Executive Summary**
- **Glossary**
- **Written Responses to Consultation Questions**
  - **General Feedback on Consultations**
  - **Building Capacity and Workers for a Big Data Pipeline**
  - **Cross-Sector Focus of Big Data Units**
  - **The Importance of Business Skills and Understanding**
  - **A Technically Skilled Workforce is Key to Big Data**
- **Proposed Skill Sets for Big Data**
  - **Non-Technical**
  - **Technical**
- **Review of Questions Posed Within the Big Data Consultations**
  - **Analyse and Introduce Big Data**
  - **Manage and Collect Big Data**
  - **Visualise and Use Big Data**
- **Additional Information**
  - **Detail for recommendation 1**

## Introduction

---

Thank you for the opportunity to participate in the consultation process for PwC Skills for Australia's Big Data Cross Sector Project. We enjoyed the opportunity to hear the perspectives of other stakeholders in the VET sector, as well as from the range of experts from other sectors about bringing big data knowledge to all sectors.

We make the following submission to clarify some of the statements we made in the consultations, to provide additional information on implementation with regard to VET units, and propose Skill Sets.

## What is Coder Academy?

---

Coder Academy is a tertiary education provider that works in the VET and higher education spaces by providing agile, industry focused accredited courses in a bootcamp mode. We operate in Sydney, Melbourne and Brisbane, and offer the following courses:

- **Fast Track:** a full-time higher education Graduate Diploma in IT for career changes
- **GenTech:** a part-time higher education Graduate Diploma in IT for school leavers
- **Cyber Security:** a full-time VET Diploma of Software Development, for people with IT experience or qualifications who want a pathway into cyber security

Coder Academy is part of the RedHill Education family, and is a brand of the Academy of Information Technology.

## Executive Summary

---

Coder Academy was pleased to participate in the consultations between the 29th of October and 1st of November. Three of our staff members made contributions: Nick Clark, Dan Adler, and James Holman. All our staff are educators, and work in industry. They collectively have a range of technical expertise and knowledge relating to big data and software development.

We have provided detailed written responses to a number of the consultation questions, and for convenience provide the recommendations from these below:

**Recommendation 1:** The preponderance of big data should be seen as an opportunity for the VET sector to train workers who can build capacity for organisations to have their own big data pipeline.

**Recommendation 2:** Regardless of sector, any organisation seeking insights from big data needs a big data pipeline to collect, process and present big data. Units for big data must address the skills and knowledge required to plan, develop and maintain a big data pipeline. Even if parts of the big data pipeline are outsourced, there must be expertise within a company that can effectively coordinate the outsourcing of particular parts of a big data pipeline.

**Recommendation 3:** Big data units of competency should mandate exposure to different types of data handling strategies and tools, and give workers the skills and knowledge to operate within a big data pipeline so they have the ability and flexibility to collaborate with other skilled workers, and oversee the outsourcing of work to specialists.

**Recommendation 4:** Though units of competency to enable an organisation to utilise big data are non-industry specific, foundational business-skills need to be taught, such that staff utilising big data can effectively communicate its potential to the business, and solve organisational problems.

**Recommendation 5:** An organisation needs to have a technical workforce in place to achieve meaningful outcomes in using big data, and if they do not, units of competency should teach basic technical skills for data storage and software development such that organisations can build out a technology function.

## Glossary

---

We see the difficulties present in defining the concept of "Big Data". Therefore, we provide the following glossary for consideration.

**Big Data** - data that is not in an agreeable format, or is too large to be handled by standard analysis tools, requiring cleaning, preparation, or transformation to obtain useful insights.

Examples of this data include:

- Sensor data collected on a farm that manages livestock and/or crops
- Web data such as comment feeds, Twitter postings or articles
- Mobile physiological sensors collecting data on real-time heart rate, sleep patterns, skin conductance, or other health data
- Financial transactions across markets, or any live feed of financial data

**Big Data pipeline** - a data storage and analysis process. It is usually comprised of different components that deal with some or all of the following:

- a data storage platform, locally or in the cloud, that continuously grows in size as more data is generated
- a method of cleaning data into a usable form
- data visualisations to show the most important features and insights of the clean data
- algorithms that utilise these features and insights for advanced monitoring

Once a pipeline is created, there is usually an automated process to intake new data, and move it through the steps described above. There should also be built in functionality to retrain/redevelop any algorithms to improve predictions over time.

## Written Responses to Consultation Questions

---

### General Feedback on Consultations

We have the following feedback about the consultations:

- The consultations were productive, however, there was an obvious disconnect between the technical and non-technical stakeholders involved, which we provide some thoughts on in the section **Building Capacity and Workers for a Big Data Pipeline**.
- The questions asked during the consultations were somewhat repetitive and did not address practical considerations relating to training and assessment of skills that would assist non-technical individuals' ability to analyse, manage and visualise big data. Our elaboration of focusing training on skills over industry specific knowledge is elaborated on in the section on **Cross-Sector Focus of Big Data Units**.
- We reject the idea that a non-technical person can develop a limited degree of technical knowledge or skills to be able to deal effectively with issues that arise with respect to big data. The culture and processes for dealing with technology in the workplace need to change, and some potential directions to approaching this can be found in the section on **A Technically Skilled Workforce is Key to Big Data**.

## Building Capacity and Workers for a Big Data Pipeline

**Recommendation 1:** The preponderance of big data should be seen as an opportunity for the VET sector to train workers who can build capacity for organisations to have their own big data pipeline.

Organisations may currently use big data for time of event monitoring, but do not have the systems or expertise to collect and analyse a vast time series of past data to make future predictions. To do this, organisations need to create a big data pipeline, which requires the capacity and expertise to maintain. The big data pipeline is described in the glossary up front.

Many organisations do not currently have the capacity or expertise to develop and maintain a big data pipeline, and the focus of big data units in the VET sector must foster skills and knowledge that assist in building capacity for organisations by training workers according to roles that support a big data pipeline.

**Recommendation 2:** Regardless of sector, any organisation seeking insights from big data needs a big data pipeline to collect, process and present big data. Units for big data must address the skills and knowledge required to plan, develop and maintain a big data pipeline. Even if parts of the big data pipeline are outsourced, there must be expertise within a company that can effectively coordinate the outsourcing of particular parts of a big data pipeline.

## Cross-Sector Focus of Big Data Units

Units of competency developed to enable an organisation to utilise big data should not be industry specific because this would lead to workers who have expertise which is too narrowly focused on a particular data set, set of procedures or tools.

The knowledge to create a big data pipeline is not industry or sector specific, and is more likely to be organisation or project specific, based upon the data storage systems and analysis tools that are currently employed at a given organisation. Units of competency should focus on fundamental skills that allow the workforce to make decisions about tools and infrastructure to support their organisation and industry needs.

For example, strategies for cleaning and analysing data can and are tailored to specific industries, organisations and their data sets. However, there are too many different tools and approaches to realistically train a worker in all of them for a given sector - skills and knowledge should focus on common characteristics of tools and approaches in relation to their purpose in the context of a big data pipeline. Developed courses should expose students to a breadth of techniques and data sets to make them versatile across industries.

This approach will work in organisations which employ a range of workers to facilitate a big data pipeline or for organisations who appoint workers trained to outsource specific parts of their big data capacity.

For example:

- a worker who has the training and skills to understand the processes and the tools used in the cleaning and analysis of data will be able to write a brief to have the data visualisation aspect of a project outsourced

**Recommendation 3:** Big data units of competency should mandate exposure to different types of data handling strategies and tools, and give workers the skills and knowledge to operate within a big data pipeline so they have the ability and flexibility to collaborate with other skilled workers, and oversee the outsourcing of work to specialists.

## The Importance of Business Skills and Understanding

Big data exists to solve specific problems for businesses or organisations that harness and utilise big data. Successful big data organisations have direct communication channels between the business and data teams to align the usage of data to these business outcomes. Since organisation-specific problems cannot be covered within units of competency, developed units need to align big data workers with the communication skills necessary to translate data-driven outcomes to the business problems they were intended to solve

**Recommendation 4:** Though units of competency to enable an organisation to utilise big data are non-industry specific, foundational business-skills need to be taught, such that staff utilising big data can effectively communicate its potential to the business, and solve organisational problems.

## A Technically Skilled Workforce is Key to Big Data

**Recommendation 5:** An organisation needs to have a technical workforce in place to achieve meaningful outcomes in using big data, and if they do not, units of competency should teach basic technical skills for data storage and software development such that organisations can build out a technology function.

Our explanation for Recommendation 1 introduced the big data pipeline, which included a mix of infrastructure and analysis tools necessary to utilise big data at an organisation. There is additional work that is required at an organisation to not just introduce the pipeline for a specific project, but make the success of continued business dependent upon the insights gained from that pipeline. For instance, a bank that utilises financial transaction data to predict in-advanced credit default might pilot this tool using big data in isolation, but the ultimate success of such a tool would depend upon a constant monitoring system for customers that automatically collects transaction data and makes real-time predictions. Without creating this automated process, organisations fail to leverage big data towards its full potential.

Yet, building such an automated process is difficult, and requires organisations to have developer teams available that can automate such processes. We understand that not all organisations have this described digital workforce, and thus we believe that units of competency need to enable workers to know at least the terminology and capabilities that developer teams create for organisations to potentially advocate to create or outsource such teams in the future. To bring this point further, we believe the best usage of the created units of competency would be to actually expose workers to developer skills, namely programming, repository usage, application programming interface (API) development and continuous integration/continuous deployment, as this would enable them best to create and/or interface with such teams within their job roles.

## Proposed Skill Sets for Big Data

---

Coder Academy is interested in the development of Skill Sets and units that allow for education providers to offer flexible, relevant and agile training for ICT workers. Currently there is a lack of flexibility in qualifications, and few Skill Sets that enable an already skilled worker to develop competency in using big data.

We provide the following proposals for Skill Sets that would fit within the parameters of current qualifications and roles in the IT sector, and that would be compatible with frameworks for offering qualifications and micro-credentials (Skill Sets) that enable an ICT worker to follow a pathway into a specialised big data role. We also list NEW units of competency that should be created such that the big data Skill Sets are fit for purpose.

### Non-Technical

#### Business Analysis

A Skill Set that provides the skills and knowledge to a non-technical worker to interface between a business and the ICT workforce that enables the collection and analysis of big data. A person with this Skill Set will have the responsibility to confirm that the built technology satisfies the requirements set out by the business.

- **ICTICT511 - Match ICT needs with the strategic direction of the enterprise**
- **ICTICT515 - Verify client business requirements**
- **ICTSAD506 - Produce a feasibility report**
- **ICTICT410 - Conduct post-implementation ICT system reviews**
- **ICTICT402 - Determine project specifications and secure client agreement**
- **ICTICT413 - Relate to clients on a business level**
- **ICTICT509 - Gather data to identify business requirements**

### Technical

#### Data Storage

A Skill Set that provides the skills and knowledge for workers who are looking to understand potential storage options of big data and build the capabilities for this storage.

- **ICTSAS415 - Optimise ICT system performance**
- **ICTSAD501 - Model data objects**
- **ICTDBS407 - Monitor physical database implementation**
- **ICTDBS503 - Create a data warehouse**
- **ICTICT423 - Select cloud storage strategies**
- **ICTPRG412 - Configure and maintain databases**
- **ICTDBS401 - Identify physical database requirements**

- **ICTDBS412 - Build a database**
- **ICTSAD502 - Model data processes**
- **ICTSAS503 - Perform systems tests**
- **ICTSAS416 - Implement maintenance procedures**

## Data Integrity

A Skill Set that provides the skills and knowledge for workers to collect and manipulate big data for storage, and maintains data integrity.

- **ICTSAS415 - Optimise ICT system performance**
- **ICTICT423 - Select cloud storage strategies**
- **ICTSAD502 - Model data processes**
- **ICTPRG405 - Automate processes**
- **ICTNWK403 - Manage network and data integrity**
- **ICTPRG402 - Apply query language**
- **ICTPRG425 - Use structured query language**
- NEW - Build data ingestion and cleaning pipeline
- NEW - Create a data storage computing cluster

## Data Analysis

A Skill Set that provides the skills and knowledge for workers to visualise and build models of big data, and drive insights towards data driven decision making.

- **ICTSAS415 - Optimise ICT system performance**
- **ICTPRG405 - Automate processes**
- **BSBMGT403 - Implement continuous improvement**
- **ICTICT401 - Determine and confirm client business requirements**
- **BSBMKG507 - Interpret market trends and developments**
- **ICTPRG527 - Apply intermediate object-oriented language skills**
- **ICTPRG403 - Develop data-driven applications**
- **ICTPRG406 - Apply introductory object-oriented language skills**
- **ICTPRG407 - Write script for software applications**
- **ICTPRG415 - Apply skills in object-oriented design**
- **ICTPRG404 - Test applications**
- NEW - Utilize statistical concepts to show confidence in an analysis
- NEW - Create a data-backed presentation
- NEW - Apply a mathematical model to solve a business problem
- NEW - Apply machine learning techniques to a business process
- NEW - Implement a model utilizing a cluster-computer framework

# Review of Questions Posed Within the Big Data Consultations

---

## Analyse and Introduce Big Data

1. What workplace tasks require a basic use of Big Data?
2. What are the skills required to perform basic tasks with Big Data? What occupations require these skills?
3. Are these skills being taught within industry or is it a skills gap?
4. What are the steps required to perform tasks that actively engage in the analysis, interpretation, and reporting of data in a workplace context?
5. How do workers make data-driven decisions and what steps are required to do so?

## Manage and Collect Big Data

1. What workplace tasks require the collection and recording of Big Data? What occupations and job roles undertake these tasks?
2. How do workers collect and record Big Data across various industries? What are the skills required to appropriately collect and record this data?
3. What occupations require the management and maintenance of data sets? What tasks are undertaken and what skills are required to perform these tasks?
4. Why is data quality and governance relevant to business and industries? What are the skills required to effectively manage data quality and data governance?
5. Are these skills being taught within industry or is it a skills gap?

## Visualise and Use Big Data

1. What workplace tasks require the visualisation of Big Data? What occupations and job roles undertake these tasks? What are the skills required to visualise data in an effective manner?
2. What workplace tasks require the presentation of Big Data? What occupations and job roles undertake these tasks? What are the skills required to effectively present data?
3. How do workers visualise and/or present Big Data across various industries? Do practices differ by industries? If so, how and why?
4. How do workers use data for advanced operational decision making? What are the skills required to appropriately do this?

## Additional Information

---

### Detail for recommendation 1

In addition, it might be overwhelming for workers to learn the combination of mathematics and computer science skills necessary to create this pipeline. That being said, many these skills have become more relevant to the present-day workforce, and we believe that the units of competency developed for big data could be an enabling tool for workers to gain these modern-day essential skills. We have created more extensive explanations of specific roles and skills relevant to big data in the section **Proposed Skill Sets for Big Data**.