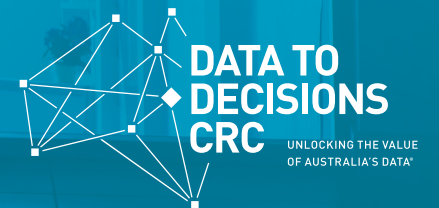# DATA SCIENCE COMPETENCY FRAMEWORK

DATA TO
DECISIONS
CRC
UNLOCKING THE VALUE
OF AUSTRALIA'S DATA®

CONTENTS

INTRODUCTION

While various competency frameworks exist in relation to Information Technology, the needs of the evolving Data Science and Analytics field were previously undefined. After identifying this gap, Data to Decisions CRC (D2D CRC) embarked on a project to create a generic data science competency framework. By describing the skills, knowledge, experience and personal attributes relevant to working in data science/analytics, including Big Data, the Data Science Competency Framework aims to support the development of the Big Data workforce. In particular, the framework can be utilised to support workforce planning, recruit, develop individuals and teams, highlight career pathways and enable competency recognition.

DEFINING COMPETENCY AND COMPETENCY FRAMEWORK

The terms **'competency'** and **'competencies'** focus on the inputs of an individual into the completion of a task. They can be defined as the behaviours (and technical attributes where appropriate) that individuals must have, or must acquire, to perform effectively at work. Competencies generally include:

• Knowledge – the cognisance of concepts, theories, models and principles gained from formal training and/or experience

• A skill is a developed proficiency or dexterity in mental operations or physical process that is often acquired through specialised training

• Experience is the accumulated application of skills and knowledge in practice

• Individual attributes and properties, qualities or characteristics of individuals that reflect one's unique personal makeup

A **'competency framework'** is a structure that sets out and defines each individual competency (such as problem-solving or people management) required by individuals working in an organisation or part of an organisation.

The competency framework does not attempt to describe everything that people within the field are required to do, for example, it does not include descriptions of software tools or technology-specific skills or knowledge, industry experience or qualifications as these are often organisation specific. Similarly, the framework doesn't identify the mix of competencies an individual, team or organisation may require – these are often specific to the context. However, it does offer a suggestion as to which competencies are most likely to be found in different data science related job families and to what extent they are expected to be demonstrated.

(adapted from: The definition of competencies and competency frameworks, UK Chartered Institute of Personnel and Development and The definition of competencies and their application at UN, University of Nebraska-Lincoln)

INTRODUCTION

THE DATA SCIENCE COMPETENCY FRAMEWORK

The structure of the Data Science Competency Framework is captured in the below diagram. It contains three key competency areas, each holding a number of relevant competencies.

- **Data analytics solution life cycle –** organises those competencies related to processing and managing data projects;

- **Technical –** includes competencies relating specifically to Big Data, technologies and tools;

- **Core –** identifies the data science related aspects of competencies that often have organisational relevance such as project management.
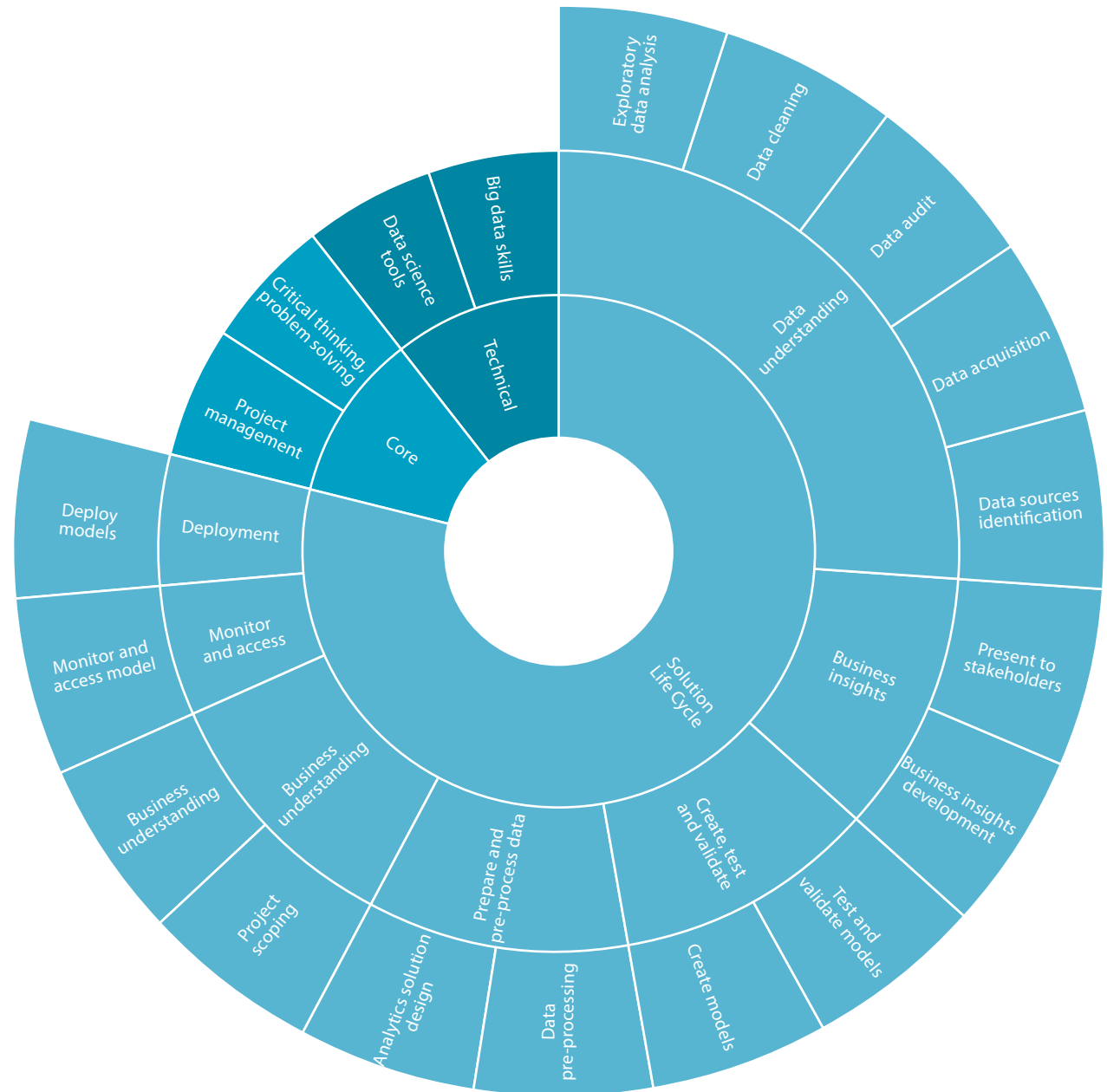
INTRODUCTION

DATA RELATED JOB FAMILIES

After detailed research, workshops and meetings, three major job families were identified within the scope of the data science workforce. Covering the breadth of the data analytics solution life cycle and stack, the job families are identified as Data Scientist, Data Engineer and Data Analyst.

The job family descriptions provided below are not intended to be an exact match to organisation's role structures as a wide variety of factors and organisational nuances impact role design. The descriptions are intended to be a guide that assist organisations to identify relevant competencies and the extent to which they might be expected to apply to their roles.

Additional job families may be added to the Data Science Competency Framework through future reviews, which will also incorporate any evolutionary changes or new trends impacting the data workforce and required competencies.

	BUSINESS UNDERSTANDING	DATA UNDERSTANDING	DATA PREPARATION	MODELLING	TEST & VALIDATE	DEPLOYMENT	COMMUNICATION OF INSIGHTS	ONGOING ASSESMENT
DATA SCIENTIST	■	■	■	■	■		■	
DATA ANALYST	■	■					■	■
DATA ENGINEER		■	■		■	■		

The above diagram is a visual guide showing which job families have strengths over the data solution life cycle. The job families will also require some capability in the non identified areas however, these are areas which are less critical.

INTRODUCTION

DATA SCIENTIST

A Data Scientist uses their knowledge of data mining techniques (statistics, machine learning, AI), their programming skills and business knowledge to extract insights from large sets of data. This includes cleaning, transforming and combining data sources, using mathematical models, machine learning and visualisations to analyse the data and communicating findings to business stakeholders. They are often required to produce answers in days, rather than months, and typically work via exploratory analysis using a variety of tools and languages. A Data Scientist uses their domain knowledge to interpret raw data and results.

A Data Scientist has an understanding of data storage and processing technologies, allowing them to perform analysis and develop models that will run efficiently and reliably within the technology constraints.

DATA ENGINEER

A Data Engineer designs, develops and maintains software architectures to collect and analyse large data sets. This includes installing, testing and configuring scalable databases and data processing systems and creating software components to collect, parse, manage, analyse and visualise data. They often tackle problems associated with data integration and unstructured data sets, employing a variety of programming languages and tools to combine data and systems and to improve data quality.

Data Engineers work with business stakeholders to determine the required data sets and analysis tools. Their goal is to provide clean, usable data and analytic output to other roles in the organisation.

There is an overlap between a Data Engineer and a Software Engineer, with the Data Engineer having a stronger emphasis on the data storage and processing layers, and a deeper understanding of data modelling and data lifecycle management.

DATA ANALYST

A Data Analyst interprets data and analytic output, using their domain knowledge to draw conclusions and support data driven decision making. This includes accessing, manipulating, querying and analysing data using a range of software and tools, applying statistical analysis techniques, and presenting the output of analysis to business stakeholders.

The goals of a Data Analyst are to identify and interpret trends or patterns in complex data sets, to recognise and define process improvement opportunities and to effectively summarise conclusions and recommendations.

There is an overlap between a Data Analyst and a Data Scientist, with the Data Analyst having more emphasis on interpreting results, presenting conclusions and making business recommendations.

INTRODUCTION

MATURITY LEVELS

To help support implementation, the Data Science Competency Framework includes four broad groupings that reference the maturity levels within the Australian Public Service Work Level Standards (WLS) & Integrated Leadership System (ILS) and Skills Framework for the Information Age (SFIA). The following table maps the various maturity levels within APS WLS & ILS and SFIA against those within the Data Science Competency Framework. However, like the competencies themselves, the organisational and team context is likely to influence the maturity levels of roles in particular organisations and is therefore just a guide.

The framework also includes an Awareness level to describe the level of competency for those who are not data science or data analytics practitioners but need to know about data science/analytics in order to maximise their effectiveness (for example, IT, marketing, service delivery, HR and intelligence analysts).

APS WLS & ILS	SFIA	D2D CRC
Level 4&5	Level 3 & 4	Practitioner
Level 6	Level 5	Senior
Exec. Level 1	Level 6	Lead
Exec. Level 2	Level 7	Director

INTRODUCTION

The competency descriptions should be read in correlation with the below general maturity level information. The job characteristics at each level offer additional information about the level of autonomy, influence, accountability and complexity expected.

JOB CHARACTERISTICS	**PRACTITIONER** (EQUIVALENT TO APS 4&5 OR SFIA 3/4)	**SENIOR** (EQUIVALENT TO APS L6 OR SFIA 5)	**LEAD** (EQUIVALENT TO APS EL1 OR SFIA 6)	**DIRECTOR** (EQUIVALENT TO APS EL2 OR SFIA 7)
Autonomy	Works under general direction. Uses some discretion in in less complex tasks. Work reviewed at frequent milestones.	Works under general direction within a framework of accountability. Plans own work to meet given objectives and processes.	Establishes team objectives and assigns responsibilities.	Strategy and policy formation and application on the organisational level.
Influence	Interacts with and influences colleagues. Responsible for components of projects.	Some responsibility for the work of others and allocation of resources. Influence the success of projects and team objectives.	Influences policy and strategy formation. Initiates relationships with internal and external partners at senior management level.	Makes decisions critical to organisational success. Inspires and influences the organisation, and the industry at executive levels. Develops long-term strategic relationships with customers, partners, industry leaders and government.
Complexity	Performs a variety of work with some level of complexity.	Broad range of complex technical or professional activities. Resolves complex issues.	Deep understanding of complex technical and business issues. Performs highly complex work activities. Contributes to the implementation of policy and strategy.	Leads formulation and implementation of policy and strategy. Deep understanding of the industry and the organisation including implications of emerging technologies.
Leadership and Accountability	Analytical and systematic approach to work within deadlines. Effective communication and teamwork skills. Appreciates the business context.	Selects applicable standards, methods, tools and applications. Communicates fluently and can present complex information. Facilitates collaboration. Plans, schedules and monitors work to meet time and quality targets. May coach other staff members.	Authority and accountability for a significant area of work, including technical, financial and quality aspects. Absorbs complex information and communicates effectively at all levels. Manages and mitigates risk. Understands the implications of new technologies and industry changes. Demonstrates clear leadership. Understands applicable legislation and codes of practice promoting compliance. Mentors and coaches other staff members.	Authority and accountability at an organisational level. High level strategic management and leadership skills. Understands and communicates complex ideas to all levels in a persuasive and convincing manner. Has a broad and deep business knowledge, including emerging technology the activities of other organisations. Assesses the impact of legislation, and actively promotes compliance. Mentors and coaches other staff members.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

BUSINESS UNDERSTANDING – PROBLEM IDENTIFICATION
Ability to establish the problem and whether it can be solved by analytics.

AWARENESS

Being aware of what kind of business problems can be addressed by data-driven solutions.

Understanding the principles and general ideas of creating a data science/analytics solution.

PRACTITIONER

Utilises communication and collaboration skills to gain an understanding of the organisational problem from stakeholders.

Utilises general knowledge of data science/analytics and experience of use cases and solutions to identify when an issue may be amenable to a data-driven solution and to assist with setting data science/analytics goals and deliverables.

SENIOR

Utilises sound communication and collaboration skills to gain an understanding of the organisational problem from stakeholders.

Establishes whether the problem is amenable to a data driven solution using experience and knowledge of a variety of use cases and potential solutions.

Ability to set data science/analytics goals and deliverables based on the established success criteria and to define key metrics of the solution's success.

LEAD

Utilises significant communication and collaboration skills to gain an understanding of complex organisational problems, from multiple stakeholders.

Knowledge of a variety of data science /analytics use cases that outline solutions to large-scale business issues.

Experience in designing complicated, multi-staged data-driven solutions to real-world problems including collecting business requirements and determining if the issue is amenable to a data-driven solution or finding the key components that are amenable to that.

Ability to set data science/analytics goals and deliverables for complex projects and to identify key success metrics.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

BUSINESS UNDERSTANDING – BUSINESS AND DATA UNDERSTANDING

Ability to utilise understanding of the orgnisation and it's data

AWARENESS

Awareness of organisational data sources and the implications of utilising data to inform organisational strategy, decision making and service delivery.

PRACTITIONER

General understanding of the organisational context, direction and key strategies.

Understanding of a specific domain, the data applicable to the domain and the meaning of the data and its implications for the organisation and decision making.

SENIOR

Understands the organisational context, functions and key strategies in-depth.

Ability to apply detailed understanding of a specific domain, the data applicable to the domain and the meaning of the data for the organisation to enhance success, improve decision making and deliver insights.

LEAD

Extensive understanding of the complexities of the organisation context, functions and strategy.

Utilises organisational understanding and detailed knowledge of a range of data sets across a variety of domains to create insights impacting strategy, decision making and service delivery.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

DATA UNDERSTANDING – DATA SOURCES IDENTIFICATION
Ability to establish the availability and accessibility of data.

AWARENESS

Appreciate data requirements and issues of availability and accessibility. Appreciate the role of data wrangling/hacking in data science/analytics

PRACTITIONER

Ability to assist in establishing the key required internal and external data sources as well as data availability and accessibility.

Utilises knowledge of data sources including how they are collected, where and how they are stored, within and external to the organisation, to verify relevance of potential data sources.

SENIOR

Ability to establish the key required internal and external data sources as well as data availability and accessibility for moderate-scale projects.

Utilises in-depth knowledge of a range of data sources including how they are collected, where and how they are stored, and interrelationships, both within and external to the organisation, to verify relevance of potential data sources.

LEAD

Ability to lead the team to establish the key required internal and external data sources as well as data availability and accessibility for large-scale projects.

Utilises extensive knowledge of a range of data sources both internal and external to the organisation, including how they are collected, where and how they are stored, and their interrelationships, to verify relevance of potential data sources.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

DATA UNDERSTANDING – DATA ACQUISITION
Ability to acquire and wrangle data.

AWARENESS	PRACTITIONER	SENIOR	LEAD
Awareness of privacy and ethical implications of sourcing internal and external data	Ability to conduct data acquisition from relational databases and flat files. Ability to hack/wrangle low complexity data selecting appropriate techniques, such as parsing, or an alogorithm, to create a data structure relevant to the problem.	Uses experience and knowledge of appropriate techniques and their strengths, such as ETL batch processing, streaming ingestion, scrapers, APIs and crawlers, to acquire open source data. Ability to conduct data acquisition from relational databases and flat files. Ability to hack/wrangle complex data, selecting appropriate techniques, such as parsing, and algorithims into a data structure relevant to the problem.	Applies a wide range of advanced data wrangling techniques, such as parsing, and algorithms for complex, multi-source data and large or complex projects. Ability to apply advanced methods of data acquisition for multiple types of data sources including ETL batch processing, streaming ingestion, scrapers or APIs for open source data.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

DATA UNDERSTANDING — DATA AUDIT
Ability to conduct data quality assessment.

AWARENESS	PRACTITIONER	SENIOR	LEAD
Awareness of the necessity of data audit.	Knowledge of basic data audit techniques and approaches. Ability to assist the team with data quality assessment using experience of relevant tools and programming languages and general understanding of the data, potential issues such as missing values, duplicate data, and the implications of data quality for the data science/analytics process.	Experience with a range of data audit techniques and approaches. Utilises experience to design, review and monitor optimal approach for data quality assessment for complex projects and to conduct data quality verification utilising a detailed understanding of the data, potential issues, such as missing values, duplicates, inconsistencies and the implications for the data science/analytics process.	Extensive and/or in-depth knowledge of data audit techniques and approaches and experience in applying them in complex settings. Ability to design, review and monitor optimal approach for data quality assessment for complex or large projects utilising extensive knowledge of the data, potential issues such as missing values, duplicates and inconsistent formats, and the implications for the data science/analytics process.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

DATA UNDERSTANDING – DATA CLEANING
Ability to identify and resolve established data issues.

AWARENESS

Awareness of the necessity of data cleaning and potential data quality issues.

PRACTITIONER

Knowledge of some basic data cleaning techniques and approaches such as data wrangling, batch processing, data mining, data enhancement, data harmonisation and data standardisation.

Ability to assist the team with data cleaning of noisy and incomplete data using relevant tools and programming languages.

Develops understanding of why data requires cleaning, including the organisational context, and the implications of this for data science/analysis processes.

SENIOR

Experience in utilising a number of data cleaning techniques and approaches for structured and unstructured data such as data wrangling, batch processing, data mining, data enhancement, data harmonisation and data standardisation.

Ability to conduct data cleaning of noisy, incomplete data or data with established data quality issues using experience of relevant tools and programming languages.

Utilises knowledge of how the interaction of multiple data issues, such as missing data, outliers, multiple values and meaning of data, impacts analysis and identifies an appropriate cleaning approach.

LEAD

Extensive and/or in-depth knowledge of best-practice data cleaning techniques and approaches for a variety of data types such as data wrangling, batch processing, data mining, data enhancement, data harmonisation and data standardisation.

Extensive experience in utilising these for cleaning complex, large, incomplete data or data with established quality issues.

Ability to design and implement data cleaning approach for complex data and projects.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

DATA UNDERSTANDING – EXPLORATORY DATA ANALYSIS
Ability to establish data specifics and relationships.

AWARENESS	PRACTITIONER	SENIOR	LEAD
Awareness of the necessity of exploratory data analysis.	Knowledge of basic descriptive analysis techniques including data visualisations to establish variable distributions and inter-variable relationships. Ability to assist in exploratory data analysis tasks.	Knowledge of a variety of exploratory data analysis methods and techniques, such as box plots, histrograms, scatter plots and Pareto charts, suitable for various data types. Ability to conduct exploratory data analysis activities including data visualisations to establish variable distributions and inter-variable relationships for moderate-scale projects. Ability to conduct missing data analysis and imputation utilising relevant techniques and establish whether data enrichment is necessary for achieving the project objectives.	Extensive and/or in depth knowledge of complex data exploration techniques and methods and experience in applying them creatively and effectively. Ability to design, lead and coordinate exploratory analysis activities including descriptive analysis techniques, identification of relevant trends and relationships within the data, data visualisation and conducting missing data analysis and imputation for complex (e.g. noisy or sparse) data. Establishes whether data enrichment is necessary for achieving large-scale project objectives as well as the nature and sources of the required data.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

PREPARE AND PREPROCESS DATA FOR MODELLING – ANALYTICS SOLUTION DESIGN
Ability to design data science/analytics solutions based on the business and data understanding.

AWARENESS	PRACTITIONER	SENIOR	LEAD
High-level awareness of a wide range of core data science/analytics techniques, their advantages, disadvantages and areas of application.	Uses familiarity with an increasing number of data science/analytics techniques, their advantages, disadvantages, assumptions and application to identify which techniques may work best.	Ability to design analytical modelling approach for moderate-scale projects or for components of large-scale projects utilising sound knowledge of data science techniques.	Design and supervise the design of Data Science/Analytics solution for large or complex projects.
	Knowledge of statistical and machine learning techniques such as classification, linear regression modelling, clustering and decision trees.	Chooses the technique optimal for the task (for example, decision trees, advanced regression techniques such as LASSO methods, random forests etc.) based on knowledge of the data and the technique's constraints, assumptions, interpretability, robustness and application.	Chooses the technique optimal for the task (for example, decision trees, advanced regression techniques such as LASSO methods, random forests etc.) based on detailed knowledge of the data and the technique's constraints, assumptions, interpretability, robustness and areas of application from a wide range of best-practice Data Science/Analytics techniques.
	Understands the data, and performance requirements for the problem to select the best technique.	Understanding of business requirements and constraints including potential trade-offs between speed and accuracy,	
	Develops ability to identify the cause of errors, such as the impact of outliers , and works logically to identify potential solutions.	Maintains knowledge of data science trends.	Utilises understanding of organisational context to negotiate performance requirements and implementation of best-practice solutions.
	Develops knowledge of current data science/analytics trends.		Expands knowledge of data science/analytics trends.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

PREPARE AND PREPROCESS DATA FOR MODELLING – DATA PRE-PROCESSING

Preprocess and transform the data to ensure that it is in the optimal format, layout or shape for the project purposes.

AWARENESS	PRACTITIONER	SENIOR	LEAD
Awareness of the necessity of data preprocessing stage, its objectives and time and resource requirements.	Ability to create required data set utilising understanding of routine problems, data formats, applicability of the data to the problem and standard modelling techniques. Ability to fuse data sources using knowledge of data pre-processing techniques such as transformation, integration, normalisation, feature extraction, to identify and apply appropriate methods.	Ability to create required data set utilising understanding of the organisational problem, applicability of the data to the problem, data format and a range of modelling techniques. Ability to fuse data sources using knowledge of a variety of data pre-processing techniques such as transformation, integration, normalisation, feature extraction, to identify and apply appropriate methods.	In-depth knowledge of and experience in using a wide variety of more complex data manipulation and transformation techniques such as transformation, integtration, normalisation and feature extraction to fuse and reshape complex, multi-source data. Utilises in-depth knowledge and experience of organisational problems, data formats and data modelling to lead the team in application of data manipulation techniques for large or complex projects.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

☑ **CREATE, TEST AND VALIDATE MODELS**– CREATE MODELS
Ability to develop a data science/analytical model.

AWARENESS	PRACTITIONER	SENIOR	LEAD
Awareness of the necessity of this stage, its objectives and time and resource requirements.	Knowledge of statistical methods and best-practice advanced modelling techniques (predictive modelling, advanced clustering, association rules etc).

Experience in using modelling techniques to model structured, uncomplicated data.

Ability to source additional information, ideas and solutions through a variety of sources such as research and relevant libraries. | Knowledge of a wide range of statistical methods and best-practice advanced modelling techniques (for example, predictive modelling, advanced clustering, text mining, social network analysis, association rules etc).

Experience in experimenting with, selecting and developing the modelling techniques most suitable for the organisational objective, organisational context and increasingly complex, unstructured and multiple data sets (for example including different types of data such as streaming data, raw text data). | Extensive knowledge of a range of best-practice advanced modelling techniques (for example, text mining, social network analysis, predictive modelling, advanced clustering, association rules etc).

Significant experience in selecting and combining modelling techniques that are most likely to deliver maximal accuracy and clarity of insight for a complex organisational objective and context.

Experience in building accurate, valid and efficient modelling solutions using a variety of complex data types (for example, streaming data, raw text data etc). |

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

☑ **CREATE, TEST AND VALIDATE MODELS**– TEST AND VALIDATE MODELS
Ability to apply relevant testing and model validation techniques.

AWARENESS	PRACTITIONER	SENIOR	LEAD
Awareness of the necessity of this stage, its objectives and time and resource requirements	Ability to assist in applying best-practice model fit testing, tuning and validation techniques (such as Chi square, ROC curve, root mean square error etc) to assess model performance utilising general understanding of data attributes that impact testing process, such as size of data set and how it is partitioned. Uses understanding of how variables and features impact model performance to test and validate best performance.	Ability to apply best-practice techniques for model testing and tuning, to assess accuracy, fit, validity and robustness for moderate-scale models and model ensembles (such as Chi square, ROC curve, cross validation, root mean square error) utilsing experience in model testing and detailed knowledge of factors that impact the testing process such as the size of the data set and how it is partioned. Uses in-depth understanding of variables and features that impact model performance to test and validate best performance.	Ability to identify, apply and lead the team in applying suitable best-practice techniques for model testing and tuning, to assess accuracy, fit, validity and robustness for large-scale or multi-stage models and model ensembles utilising extensive experience and knowledge of factors that impact the testing process. Identifies model evaluation metrics based on detailed undertanding of organisational and analytical requirements. Utilises knowledge of relevant policy and standards to ensure testing information is documented and maintained by the team.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

DEPLOY MODELS – DEPLOY MODELS
Ability to integrate the model into the business system.

AWARENESS

Awareness of the fact that models typically need to be deployed / integrated into existing organisational systems.

PRACTITIONER

Utilises knowledge of existing infrastructure, architecture and a range of big data technologies to ensure data science/analytics models, techniques and technologies can be deployed into production and are sustainable and maintained over time.

Utilises knowledge of software development pipeline and a range of technologies to support development of user interface and to troubleshoot issues.

SENIOR

Utilises detailed knowledge of existing infrastructure, architecture and big data technologies and the organisational context to deploy data science/analytics models, techniques and technologies and ensure they are sustainable and maintained overtime.

Ability to build a user interface and support use of model through collaboration with key stakeholders and understanding of the problem and organiational context.

LEAD

Ability to lead team in deploying large-scale or multi-stage models or model ensembles utilising extensive knowledge of existing infrastructure, architecture and big data technologies.

Utilises collaboration skills, in-depth knowledge of the problem and organiational context to develop innovative user interfaces and support use of the model.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

BUSINESS INSIGHTS – BUSINESS INSIGHTS DEVELOPMENT
Ability to develop clear and actionable business insights based on the analysis results.

AWARENESS	PRACTITIONER	SENIOR	LEAD
Awareness of the need to develop clear and actionable business insights based on the modelling results.	Translates data science/analytics outputs into clear and actionable insights and supports for decision making utilisng general understanding of the organisational context, meaning of the data and knowledge of the practical application of data science/analytics outputs. Ability to identify trends and patterns in data sets and the implications for the organisation.	Translates data science/analytics outputs into clear and actionable insights and supports for decision making utilising detailed knowledge of the organisational context, data analytics capability, meaning of the data, and experience in practical application of data science outputs. Identifies trends and patterns in complex data sets and the implications for the organisation.	Ability to lead the team in developing actionable insights and decision making supports based on complex or multiple data science/ analytics outputs utilising extensive knowledge of the organsation context and strategies, the data analytics capability and extensive experience in the practical application of outputs/models. Utilises organisational and data science/analytics knoweldge to identify long-term trends with the potential to impact operational outcomes.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

BUSINESS INSIGHTS – PRESENT TO STAKEHOLDERS
Ability to present insights to the stakeholders.

AWARENESS

Awareness of the need to present insights to the stakeholders.

PRACTITIONER

Ability to assist to present insights to stakeholders in easy-to-understand language and provide visual data summaries.

Utilises knowledge of data visualisation software tools and the audience to identify best fit for the information to be conveyed.

SENIOR

Ability to present insights to stakeholders utilizing experience in translating complex information into easy-to-understand language and ability to provide visual data summaries illustrating and supporting the insights.

Uses knowledge of a variety of data visualization software tools to select the best option for the data and stakeholders.

Understands audience and tailors communication to the audience.

LEAD

Ability to present insights to the stakeholders for large or complex projects and gain business validation utilising extensive communications and influencing skills and experience in conveying complex information in easy-to-understand language.

Experience in identifying the best visualisation tool based on detailed knowledge of the data, organisational context and data analytics capability.

Significant undertanding of a wide variety of audiences and strong ability to tailor communication to their needs.

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

MONITOR AND ASSESS MODELS
Ability to monitor and assess model validity.

AWARENESS	PRACTITIONER	SENIOR	LEAD
Awareness of the fact that models typically need to be assessed and recalibrated as needed.	Ability to assist in monitoring predictive performance of the model to ensure it is up to date and delivering valid results.	Ability to monitor predictive performance of the model or set of models to ensure it is up to date and delivering valid results.	Ability to coordinate performance monitoring of large-scale or multi-stage models or sets of models to ensure they are up to date and delivering valid results.

TECHNICAL COMPETENCIES

DATA SCIENCE TOOLS
Ability to identify and utilise appropriate data science/analytics tools and programming languages.

AWARENESS	PRACTITIONER	SENIOR	LEAD
Appreciation of the fact that custom coding may be required to customise every step of the data science/analytics life-cycle. Awareness of the mainstream programming languages available (e.g. R, Python etc.)	Familiarity with the wide range of mainstream commercial and open source data science/analytics software tools, their constraints, advantages, disadvantages and areas of application. Intermediate skills in using at least one such tool. Familiarity with programming languages (such as R or Python). Basic programming skills. Ability to interpret an existing script of moderate complexity.	Familiarity with the wide range of data science/analytics commercial and open source software tools, their constraints, advantages, disadvantages, areas of application and mainstream packages relevant to technical stages of data science/analytics projects. Expertise with at least one such tool. Intermediate to advanced skills in programming languages used for data science/analytics (such as R or Python) and ability to apply these for data acquisition, preprocessing, modelling and model deployment. Ability to interpret and modify existing scripts and conduct quality checks.	Familiarity with the wide range of data science/analytics commercial and open source software tools, their constraints, advantages, disadvantages, areas of application and best-practice packages. In-depth expertise with at least one or two such tools. Advanced skills in programming languages used for data science/analytics (such as R or Python). Ability to apply these for data acquisition, preprocessing, modelling and model deployment. Ability to coordinate quality checks of scripts for one or more projects as well as to maintain and monitor a library of team scripts and coordinate its review and update.

TECHNICAL COMPETENCIES

BIG DATA ANALYTICS
Understanding of, and ability to work with Big Data concepts, architectures and methodologies.

AWARENESS	PRACTITIONER	SENIOR	LEAD
Appreciation of 'big data' and its relationship to data science/analytics and data management. Awareness of the context when 'big data' solution is required. Awareness of the concept of the MapReduce programming model, how it addresses big data problems. Awareness of big data technologies.	Familiarity with the latest Big Data technologies and the complexities of integrating with existing information systems and data science/analytics solutions to use some technologies in a 'Big Data' environment. Ability to assist in creating a Big Data environment as well as capturing and evaluating results in a big data setting. Assist in developing custom programs to address unique data science/analytics problems and deploying Big Data solutions on large scale data sets in the enterprise.	Knowledge of the latest big data technologies, methods of analysis and relevant software tools. Ability to identify a big data problem and select the appropriate big data technologies for a problem, using a variety of technologies in a 'big data' environment. Ability to create a big data environment by setting up and deploying the latest big data tools, design a solution that encapsulates systems dealing with both structured data and unstructured data and capture and evaluate results in a big data setting. Deploy big data solutions on large scale data sets in the enterprise and develop custom programs to address unique data science/analytics problems.	In-depth knowledge of big data technologies, the specifics of integrating them with existing information systems and using them for data science/analytics solutions. Ability to design and lead data science/analytics projects including creation of a big data environment by setting up and deploying tools, capturing and evaluating results and deploying big data solutions on large scale data sets in the enterprise. Ability to lead a team in identifying a big data problem, selecting the adequate techniques and performing data acquisition, data audit, cleaning, preprocessing, model development & testing and deployment. Ability to design and implement a multi-stage solution that encapsulates systems dealing with both structured and unstructured data. Shares knowledge, experience and skills with team members through coaching and mentoring.

CORE COMPETENCIES

PROJECT MANAGEMENT
Ability to scope, plan, manage, monitor and evaluate data science/analytics projects.

AWARENESS	PRACTITIONER	SENIOR	LEAD
Awareness and understanding of all stages of data science/analytics life-cycle and specifics of the data science/analytics project management.	General understanding of all stages of Data Science/Analytics life-cycle and project management. Ability to assist in the scoping, planning and delivery of projects under the direction of Senior Analyst or Lead Analyst, including documenting business requirements.	Detailed understanding all stages of Data Science/Analytics life-cycle and Data Science/Analytics project management, the relevant time and resources requirements. Ability to manage moderate-scale projects and assist in management of large-scale or multi-stage projects. Experience in working with stakeholders and the collection of business requirements for Data Science projects including establishing the business need, key stakeholders, scope, resourcing and success criteria for a specific issue. Ability to prepare a project plan, communicate the plan to the team and allocate the tasks..	In-depth understanding of all stages of Data Science/Analytics life-cycle and specifics of Data Science/Analytics project management, the relevant resources, time requirements etc. Ability to manage large-scale Data Science/Analytics projects and assist in managing Data Science/Analytics programs. Ability to coordinate the team's activities in collecting business requirements, establishing the need, scope, goal, key stakeholders and success criteria. Negotiates with key stakeholders to identify best practice outcomes aligned to organisational objectives.

CORE COMPETENCIES

THINKING AND PROBLEM SOLVING – CRITICAL AND ANALYTICAL THINKING
Use research skills, critical thinking and data science/analytics techniques to interpret data and analyse business environments.

PRACTITIONER

Ability to consider an issue from different perspectives and question information validity and assumptions.

Ability to work systematically through a problem, use reserch skills to identify potential solutions from a variety of sources and knowledge of Data Science/Analytics techniques to assist the team to interpret data and analyse business environments.

SENIOR

Ability to critically evaluate complex business information from multiple perspectives, including questioning assumptions and validity.

Ability to identify relevant information from a wide variety of sources and use research skills and data science/ analytics techniques to interpret data, analyse business environments, and develop advanced solutions for real world problems.

LEAD

Ability to critically evaluate and analyse complex business information from differing and multiple perspectives to make strategic and impactful decisions.

Ability to support the team to utilise critical and analytical thinking through role modelling and utilising knowledge of problem solving techniques.

CORE COMPETENCIES

THINKING AND PROBLEM SOLVING – CREATIVITY AND OPENNESS TO INNOVATION

Ability to identify, develop and engage with new and different approaches and perspectives.

PRACTITIONER

Ability to learn and ask questions about new approaches, perspectives, tools and techniques.

SENIOR

Ability to maintain an openness to new approaches, seek different perspectives and experiment with new techniques.

Ability to adapt or develop new methods, techniques and learn new software packages as available in the literature and recent publications.

LEAD

Ability to support innovation in the team through asking questions, maintaining openness to new ideas and encouraging experimentation.

Ability to identify and support the team in adapting or developing innovative methods and learn new software packages as available in the literature and recent publications.

DIRECTOR LEVEL COMPETENCIES

DATA SCIENCE/ANALYTICS SOLUTION LIFE-CYCLE COMPETENCIES

Extensive knowledge of data science and analytics concepts, principles and practices across areas including (but not limited to) data wrangling, machine learning, statistical model building, validation, assessment, monitoring and recalibration and model deployment.

Experience in applying these to new areas and industries and leading the research and implementation of innovative approaches.

Significant expertise in managing the production and maintaining the library of relevant scripts and utilisation of other data analytical software tools.

Ability to provide strategic advice to Senior Executives, contributing to the development of longer term organisational strategy.

Ability to develop and deliver data awareness raising programs across the organisation. Ability to develop and deliver on data program aims and objectives to maximise the organisation's successful achievement of its key objectives.

Oversee and manage data quality assessment, data cleaning and data pre-processing practices and processes maintaining up-to-date best practice approaches.

Uses in-depth knowledge of organisational data capability to provide leadership and direction across strategic and highly technical projects ensuring maximal leverage of internal data assets and external data.

Uses in-depth knowledge of the organisation and ability to build relationships to oversee the business analysis and requirement collection processes by the data science/analytics teams; data exploration practices and processes.

Maintains awareness of trends and best practice approaches.

DIRECTOR LEVEL COMPETENCIES

TECHNICAL COMPETENCIES

SOFTWARE TOOLS AND PROGRAMMING LANGUAGES

High level of knowledge of a wide range of data-related commercial and open source software tools, their constraints, advantages, disadvantages, areas of application and best-practice packages.

Ability to ensure that the organisation maintains optimal up-to-date toolset and the relevant skill set to achieve organisational strategies.

BIG DATA SKILLS

High level of awareness of wide range of the best-practice big data storage and distribution systems and relevant software tools, their constraints, advantages and disadvantages.

Ability to design organisational strategy in data science utilising big data ensuring that the organisation maintains an optimal, up-to-date big data - modelling toolset and the relevant skill set.

DIRECTOR LEVEL COMPETENCIES

CORE COMPETENCIES
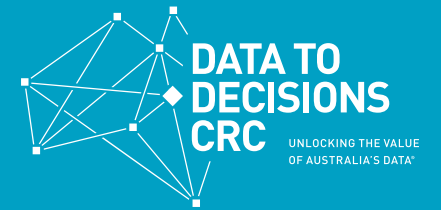
MANAGE DATA SCIENCE PROGRAMS

Experience in coordination and management of data science/analytics programs, including applying and improving established approaches and techniques as well as introducing innovative methods.

Ability to lead strategic planning for longer term or large data science/analytics programs that may have an organisation-wide impact.

Ability to manage effective preparation and quality of executive reports that provide data-driven insights into strategic issues.

Identifies and leads innovative solutions utilising in-depth knowledge of the organisation's business and stakeholders to communicate the implications at a strategic level.

Develops and implements effective knowledge sharing processes, ensuring that team members maintain a high level of technical expertise.