I have worked for over 20 years as a data specialist in various capacities including Statistician, Data Scientist, Data Engineer, Quantitative Economist, and Project Manager. I have experience in the end-to-end delivery of data and technology; from gathering requirements to designing the data architecture and data model to building the data pipeline and data warehouse to data visualization and analysis. I have a very rich depth and breadth of data experience and passionate about all things data. At the Department of Agriculture, I currently work as Lead Data Scientist and am responsible for the design and implementation of the Azure Data Factory for the Taking Farmers to Market program. This includes data ingestion from a range of data sources (SQL Server, Json, XML, text etc).

At Department of Agriculture, I work daily with DevOps and GitHub. The DevOps instance that runs pipelines on a dedicated VM installation that hosts three Organizations. Each DevOps Organisation has 3 projects (Prod, Test and Dev) mapped to each of these Datawarehouse Environments. I use Azure DevOps to perform CI/CD using GitHub and run a weekly code deployment session to merge all feature branches in the current Agile Sprint to the Live branch. This triggers an Azure DevOps Pipeline script that performs verifications tests and deploys the changes to the Production environment. All commits to feature branches also trigger an Azure DevOps Test script that run Tests for Syntax and checks for Code Standards. I use SourceTree as the visual Git Client connecting to the Azure DevOps repositories. I am responsible for maintenance of the Azure DevOps Instance and VM including Access control and pipeline monitoring.

As Consulting Data Scientist/Data Engineer at IP Australia I investigated the impact of COVID-19 on the agency’s workforce (for IP applications). This required broad and deep data extraction of large data sets using in the AWS S3 environment. Data analysis was undertaken using SQL and Python. National estimates were evaluated against international estimates. To produce these estimates required me to transform files of several formats from external organisations. I also developed a Python web-based extraction tool to rapidly extract data from national and international agencies.

I led the ‘Third Year Review of the Australian Government’s Indigenous Procurement Policy’ (IPP) at Deloitte and was responsible as delivery manager and principal data scientist. I managed a team of economists, accountants, and data scientists using a range of technologies (SAS, R, SQL, Python, Power BI) and assessing agency performance against policy. Engagement with the client was key to obtaining industry and grants data to be integrated for analysis. I mentored the team on analytical processes, develop and analyse a survey of businesses, and developed an alternative measure for compliance reporting. I met with my Deloitte partner and director on advice and issues. I engaged with external stakeholders on data and systems. Prior to the inter-departmental committee (IDC) meeting I advised on additional analysis on three risk areas found in our data analysis; I advised the client that further examination of these risks was required to provide documented assurance and independence. Using the analysis I led a successful argument at the IDC meeting for a shift from volumes-based reporting to volume and value-based reporting (which has been implemented by the client). Separate advice on data governance issues within the client’s Power BI processes was developed for the client and implemented.

I work extensively in data transformation, data analysis, schema interpretation and/or development as well as using SQL and Python scripts. I have also worked on designing and building the Enterprise Data Warehouse in Azure SQL Server. I have extensively worked on optimisation of SQL scripts and pipelines (parallelism, auto scaling, automatic configuration, vulnerability reports etc.) and built a new framework that supports adding new data source items by just configuration changes. I work with other Data Engineers in a fast-paced Agile environment with regular brainstorming sessions and workshops. I also work with colleagues from different business departments (Finance, Customer Service, Marketing etc.) for gathering requirements, proof of concept presentations, ad hoc data analysis requests.

I have extensive experience with health and health-related analytical experience:

* At the National Health Funding Body, I developed audit-style estimates for the Commonwealth Department of Health as well as for each of the states in SAS and SQL. I have also worked extensively in the analysis and visualization of data using Tableau. I used SQL for Data Transformation and designed an optimised ETL solution that tracks changes from the sources using Microsoft CDC and performs a batch processing using a Configuration Table. The data transfer from one layer of the data warehouse to another was performed using Partition switching in/out for negligible down time. I also helped design a data compression strategy that optimised the data transformation process of a legacy data mart and reduce run times by 85%.
* At PwC I delivered a Data Management Stragey, as both Program Lead and Technical Specialist, that targeted the MBS and PBS programs. The analysis provided the Department of Health with an understand of risks relating to its data, storage, technology and handling, and ratings.

I have advanced level experience in SQL over the last 15 years of building Data Warehouses, Data Pipelines and Transformation Logic. I have used most of the advanced features that SQL-based scripting offers and also led program tutorials and debugging. I always use SQL for most of the data transformation and only use ETL tools for Orchestration and Parallelism. I have extensive experience in performance tuning of SQL queries such as by understanding the SQL Engine Execution Plan, Profiler, Indexing, Partitioning, Statistics, and maintenance tasks.