Essential Criteria

**1. Experience working with, analysing and maintaining large data sets.**

I have worked for over 20 years as a data specialist in various capacities including Data Scientist, Data Engineer, Statistician, 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 with my education - a bachelor’s degree in Finance and Masters in Applied Statistics – providing a foundation for this engineering experience. 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 Adgriculture I currently work with trade-related data, both agency and external. This requires me

As Consulting Data Scientist/Data Engineer 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 (text, csv, json, AVRO, PARQUET and ORC). I also developed a Python web-based extraction tool (a scraper using Beautiful Soup) to rapidly extract data from national and international health and IP agencies.

**2. Experience with Python**

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 PM&C IT, Finance, and Supply Nation 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. Government procurement data (AusTender) was extracted using an adapted Python-scripted process that joined national procurement data with state-based procurement data (as well as data from ASIC and other public sources). I regularly met with my team through daily stand-ups and individually to discuss issues such as progress, methodology. I regularly met with the project manager and the project sponsor on progress and analysis. 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. Refer: <https://www.niaa.gov.au/sites/default/files/publications/third-year-evaluation-indigenous-procurement-policy.pdf>.

**3. Experience with ingesting, linking and maintain data sets from difference sources**

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.

At … Health I worked exclusively in SQL developing estimates for the Commonwealth Department of Health as well as for each of the states. I have also worked extensively in the analysis and visualization of data using Tableau. At Volkswagen and William Hill, I worked as Datawarehouse Architect and Data Warehouse Lead respectively and designed and built two near real time Enterprise Data Warehouses from scratch. 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 does a row compression on all current data partitions and does a Page Compression on old partitions. I helped optimise the data transformation process of a legacy data mart and reduce run times by 85%. I worked on the delivery of Financial Regulatory reporting needs as per IFRS9 standards of financial risk reporting. I also helped create data dashboards using Power BI and SAP Business Objects.

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 by understanding the SQL Engine Execution Plan, Profiler, Perfmon tools, Indexing, Partitioning, Statistics, maintenance tasks etc.

**4. Experience with cloud technologies**

I have worked with the Microsoft Azure stack for over 5 years. I have used Azure SQL Server for the database, Azure Data Factory and Azure Synapse for data pipelines and data transformation. I developed the

I have built Azure Logic Apps for custom workflows and notifications (the notifications are used by both my team and Platform Services for monitoring use and failures of the ETLs).

At IP Australia (as Consulting Data Scientist) I worked with AWS S3 for the delivery of an analysis of the impact of COVID-19 on the agency’s workforce (for IP applications). Refer also Selection Criteria3.

**Desirable Criteria**

**1. Experience with supporting other software developers in delivering capability.**

**2. Experience with CI/CD solutions.**

At Department of Agriculture, I work 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 the 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.

**Experience as a Data Scientist, Data Engineer, or Data Analyst**

**SQL and Python or JavaScript (Node.js)**

**Batch and real-time (event-based) pipelines e.g., Spark**

I have extensive experience in designing and building batch processing pipelines using either full load or incremental load logic. I always design a configuration-based batching process that includes dynamic addition of new objects. I have designed an optimised batching process using Table Partitioning in conjunction with SQL compression techniques for faster data I/O. I designed an automated process that creates new Table Partitions each night and merges old Partitions. The Automated process also creates new partitions with row compression and converts the old partitions to Page compression for optimised performance. With these techniques the data warehouse was near real time (every 10 minutes). I have extensive exposure to (near) real-time processing delivering projects that required me to develop run a Serverless Spark Pool to run a Python notebook for data transformation.