1. Demonstrated experience as a Data Engineer in large and complex enterprise environments.

In my tenure as a Senior Data Engineer at CBA, my methodology is structured around first engaging with stakeholders to thoroughly understand their needs. This involves a meticulous analysis of their requirements, bridging any gaps, and seeking further clarifications from business stakeholders when necessary. In partnership with Data Designers and Architects, my focus is to ensure that our designs integrate smoothly with the existing infrastructure, thus fostering a unified and efficient data ecosystem.

My experience encompasses managing extensive datasets, as evidenced by my role in developing data pipelines for files containing millions of records, facilitating their ingestion into the foundational Raw layer of our Data Warehouse. Notably, my projects have included the migration of vast quantities of data from on-premises storage to cloud-based data warehouse solutions such as Snowflake and Redshift. I pioneered a cloud framework to streamline the complex ingestion of fixed-width files from mainframe sources, enhancing the efficiency and scalability of our data engineering efforts. This innovation not only accelerated the development of data pipelines but also expedited the availability of data for business stakeholders, enabling swift insight generation.

Leading the charge in the creation and upkeep of scalable data pipelines for handling significant data volumes, my role extensively utilized AWS services and PySpark. This involved the adept use of Amazon Elastic Container Service (ECS) for the effective orchestration of Docker containers, thereby optimizing both resource use and deployment strategies. Furthermore, I engaged with Amazon Elastic Container Registry (ECR) for the secure management of container images, which streamlined our continuous integration and delivery processes, ensuring quicker and more dependable deployments. My responsibilities also included the adept configuration and scaling of AWS Glue job instances and Airflow cluster instances, aligning with the project's demands and underscoring my ability to navigate complex enterprise settings with a priority on securing data storage and transmission.

2. Demonstrated experience in the development, implementation, compliance, and review of policies and procedures.

With a solid background of over six years in Data Engineering, my expertise encompasses the development of comprehensive data pipelines, facilitating every phase of data processing including ingestion, transformation, and reporting. My technical proficiency spans a variety of cutting-edge tools and technologies, including PySpark for processing, Airflow for workflow automation, and Docker for application containerization, alongside extensive experience with diverse database systems like Redshift, Snowflake, BigQuery, Teradata, and Oracle. Further enhancing my technical repertoire, I've adeptly integrated a plethora of AWS services such as S3, Glue, ECS, ECR, EMR, EC2, Lambda, StepFunctions, Athena, and DynamoDB into scalable data solutions.

A highlight of my career at Commonwealth Bank of Australia (CBA) was leading a critical project within the Comprehensive Credit Reporting (CCR) initiative, focusing on the intricate task of extracting data from XML response files. This role required me to meticulously analyze data dictionaries to identify Personally Identifiable Information (PII) elements, collaborate with data stewards for verification, and manage the update process in Alation/Collibra for data governance. I was instrumental in implementing security measures, such as hashing PII data and establishing stringent access controls to databases containing sensitive information, thus preventing unauthorized access by consumer and reporting teams. By developing a versatile Python and Spark-based framework for data ingestion from XML files, I delivered a solution that not only improved efficiency but also fostered reusability across teams, significantly saving development time and efforts. My contribution in crafting PySpark data transformation jobs, incorporating complex logic while adhering to Enterprise Data Modeling standards for data warehouse integration, was recognized and praised.

My skill set is comprehensive, including in-depth knowledge and hands-on experience with PySpark, Airflow, Docker, and various database technologies. Additionally, my ability to leverage AWS services effectively ensures the delivery of robust, scalable data infrastructure solutions.

3. Demonstrated experience in complex problem-solving and issues management, and may coordinate and perform detailed or sensitive projects that impact strategic, political, or operational outcomes.

In my role at CBA within the Business Banking sector, I faced the intricate task of architecting a data pipeline from scratch utilizing AWS services and PySpark. The project was complex, requiring the consolidation of requirements from approximately 60 distinct reports previously generated using Excel, MySQL SSIS packages, and SAS. Transitioning this data—while maintaining the integrity of the logic from the SSIS packages—into a newly established data warehouse, first through Hive and subsequently into Snowflake, presented a substantial challenge. As the Lead Data Engineer, my responsibilities included fostering collaboration among scrum masters, designers, architects, and business stakeholders. This was achieved through regular discussions and the strategic use of Jira for backlog management, allowing us to break down the requirements into manageable segments for phased and efficient delivery.

The project's complexity was further heightened by the absence of a centralized base data repository, necessitating a detailed process to pinpoint the necessary data points for each report. This process involved securing files from various source system administrators, followed by their integration into the data lake or raw layer. Upon this foundation, I constructed a transformation layer to refine the data, making it suitable for analysis and reporting through Tableau. My comprehensive approach covered the entire spectrum of data pipeline development—from data sourcing and ingestion, through transformation and enrichment, to making it accessible for business intelligence. Another critical aspect of this project was managing Personally Identifiable Information (PII). I ensured the meticulous handling of PII data in accordance with data policy regulations, utilizing the AWS Comprehend service to automatically identify PII data within the pipeline. Additionally, I supported Data Stewards and Designers in recognizing PII columns, guided by the data dictionary provided by Source System Subject Matter Experts (SMEs). This rigorous approach to data privacy and security underscored my commitment to adhering to stringent data protection standards, further demonstrating my capability in addressing complex problems and coordinating sensitive projects that have significant strategic, political, and operational implications.

4. Demonstrated experience actively managing key stakeholder relationships and identifying relevant stakeholders’ expectations and concerns to develop and communicate methodologies and practices to achieve outcomes.

During my role at Commonwealth Bank of Australia (CBA), I took a proactive approach in managing and nurturing relationships with key stakeholders, a testament to my adeptness in understanding and aligning with their needs and expectations. This involved not just routine engagements but also the delivery of tailored data solutions that addressed specific business challenges. By conducting regular solution overviews and designing experimental data models, I showcased my ability to translate complex data concepts into actionable insights for stakeholders. My interaction with them went beyond formal meetings; I was often in direct communication, sharing data extracts on an ad-hoc basis and swiftly incorporating feedback to refine and enhance these extracts. This iterative process was indicative of a vibrant and responsive stakeholder relationship.

My responsibilities extended to providing after-hours support, reflecting my commitment to ensuring uninterrupted business operations and stakeholder satisfaction. As a senior data engineer, handling BAU (Business as Usual) issues was part of my remit. This included transparent communication about any delays or issues, thereby setting realistic expectations and mitigating any potential frustration. One of my notable contributions was automating ad-hoc queries based on stakeholder feedback, significantly reducing their dependency on the data team and enabling them to access insights independently. Moreover, I facilitated the creation and walkthrough of design and mapping documents, ensuring that both the data team and business stakeholders shared a common understanding of project deliverables. This practice not only fostered confidence among stakeholders in the data team's capabilities but also minimized the risk of design or requirement discrepancies, thereby enhancing project outcomes and stakeholder satisfaction.

5. Demonstrated experience in Object-Oriented Language (Java, Python, go, JavaScript, SQL,Cypher).

With a solid foundation in multiple object-oriented programming languages, my experience encompasses over seven years of working with Python, during which I developed tools for mocking substantial volumes of data for volume testing in non-production environments. My proficiency in SQL extends even further, with fourteen years of experience and an advanced level of expertise in this language. While I possess a good understanding of JavaScript and Go, my hands-on experience with these languages is limited. Nonetheless, my eagerness to learn and adapt to new technologies has always driven me to quickly acquire new programming skills as required. This adaptability and passion for technology were pivotal in my roles at Latitude Financial Services, where I leveraged Python and PySpark for intricate data processing tasks. My tenure at NAB and Macquarie Bank further underscored my technical capabilities, as I successfully applied Java and Scala to develop sophisticated data engineering solutions, demonstrating my comprehensive skill set and versatility across a range of object-oriented languages.

6. Demonstrated experience in processing and analytics services. (Spark, Java map reduce, Knative, Ray, NIFI).

Throughout my career, I have amassed considerable experience with a range of processing and analytics services, cementing my expertise in the domain of big data and its applications. My journey with Apache Spark, a cornerstone of my skill set, spans multiple roles and projects at leading financial institutions such as CBA, NAB, Macquarie, and Latitude Finance. In these capacities, I leveraged Spark's powerful processing capabilities to architect and implement sophisticated big data pipelines, facilitating efficient data analysis and insights generation. Further, my proficiency includes Java MapReduce, which I have utilized for developing distributed data processing solutions, enhancing the scalability and performance of data analytics tasks.

In addition to these core technologies, I have ventured into the realms of Knative and Ray through proof-of-concept projects. With Knative, I explored the serverless framework's potential for building and deploying modern, container-based applications, aiming to optimize cloud environments for greater operational efficiency. My engagement with Ray allowed me to delve into parallel and distributed computing, experimenting with its capabilities to devise more effective data processing methodologies. Although my experience with Apache NiFi is more theoretical at this stage, my keen interest in upskilling in this area reflects my continuous pursuit of knowledge and proficiency in cutting-edge technologies that promise to advance the efficiency and effectiveness of data processing and analytics services.

7. Knowledge or experience in working with Open-search, Neo4j, Redis, Oracle, Kafka, Kubernetes.

Throughout my tenure at Macquarie Bank, my project engagements have immersed me in a diverse technological landscape, enriching my professional toolkit with a wide array of sophisticated tools and platforms. My journey includes significant hands-on experience with Kafka, where, over a span of three years, I architected and implemented real-time data streaming pipelines. My work with Kafka wasn't just about moving data; it involved navigating complex streaming scenarios to ensure that data flows were not only efficient but also reliable, demonstrating an advanced proficiency in managing real-time data infrastructure. Parallel to my Kafka endeavors, I have amassed seven years of experience with Oracle, developing robust database solutions that underpin critical business operations. My role involved designing and managing tables and views that form the backbone of our data storage strategy, showcasing a deep understanding of database management and optimization.

In addition to these technologies, my professional narrative includes a solid grounding in Kubernetes. I have a comprehensive grasp of its core concepts, architecture, and components, including pods, services, deployments, and replicasets. This knowledge has empowered me to deploy applications effectively, manage scaling operations, orchestrate zero-downtime updates, and ensure high availability across the board. While my engagement with Open-search and Neo4j has been more theoretical, I possess a foundational understanding of these technologies as well. This blend of practical experience and theoretical knowledge positions me uniquely, enabling me to leverage these technologies to drive innovation, streamline operations, and deliver tangible business outcomes.