Qualifications and experience we consider to be essential for the role:

• 5+ years of experience in Data Engineering: SQL, DWH (Redshift or Snowflake), Python (PySpark), Spark and associated data engineering jobs.

• Experience with AWS ETL pipeline services: Lambda, S3, EMR/Glue, Redshift(or Snowflake), step-functions (Preferred)

• Experience with building and supporting cloud based ETL (Extract Transform Load) Data Pipelines

• Excellent communication & presentation skills

• Good to have working experience on RESTful API frameworks (Flask/FastAPI), Messaging Queue service (Kafka)

• Should be able to work on a problem independently and prepare client ready deliverable with minimal or no supervision

• Experience with working in an agile environment, development life cycle and with diverse stakeholders (like IT, Business, Project Managers etc)

• General Insurance domain experience is preferred

• Data Management Skillsets:

• Ability to develop, enhance data models and identify ETL optimization opportunities. Exposure to ETL tools is going to help in the work

• Should have strong grasp of advanced SQL functionalities (joins, nested query, procedures, PL/SQL)

• Strong grasp of python libraries and concepts around PySpark, Numpy, Pandas, functions, constructors etc

• Strong ability to translate functional specifications / requirements to technical requirements

We are looking for somebody with a working knowledge of building data pipelines and the underlying infrastructure. Experience in data warehouse design undertakings, following best practices during implementation is a big plus. You have worked with (or are keen to do so) Data Analysts, Data Scientists and Software Engineers.

Practical knowledge of (or strong desire to learn) the following or similar technologies:

• Python

• Airflow

• Databases (PostgreSQL)

• Data Warehousing (Redshift / Snowflake)

• SQL (We use DBT for modelling data in the warehouse)

• Data Architecture including Dimensional Modelling

Otherwise an interest in learning these, with the support of the team, is essential. We're looking for people with a commitment to building, nurturing, and iterating on an ever-evolving data ecosystem.

Other beneficial skills include:

• Familiarity with Docker and/or Kubernetes (EKS)

• Implementation / Contribution to building a Data Lake or Data Mesh

• Having worked with a wide variety of AWS services

• Experience in using infrastructure as code tools (e.g. Terraform)

Responsibilities

• Understand distributed technologies and the best practices around them

• Build and maintain services/features/libraries that serve as a definitive example for new engineers

• Design and write effective complex Spark jobs (data processes, aggregations, pipeline)

• Design and write complex asynchronous, highly parallel low latency APIs and processes

• Work as part of an Agile team to maintain, improve, and monitor their data collection processes using Scala and Java

• Apply industry practices such as TDD and SOLID

• Understand and be able to apply data structures and algorithms

• Understand data architecture and use appropriate design patterns

• Design database(s) and database tables across a range of data storage technologies

• Support our Data Science team to help deliver their machine learning models into production environments

Skills

• Bachelor's degree in Computer Science, similar technical field of study or equivalent practical experience.

• Commercial experience developing Spark Jobs using Scala

• Commercial experience using Java and Scala (Python is nice to have)

• Experience in data processing using traditional and distributed systems (Hadoop, Spark, AWS - S3)

• Experience designing data models and data warehouses.

• Experience in SQL, NoSQL database management systems (PostgreSQL and Cassandra)

• Commercial experience using messaging technologies (RabbitMQ, Kafka)

• Experience using orchestration software (Chef, Puppet, Ansible, Salt)

• Confident with building complex ETL workflows (Luigi, Airflow)

• Good knowledge of working cloud technologies (AWS)

• Good knowledge using monitoring software (ELK stack)

• Motivated problem-solving skills, ability to bring ideas forward and adapt solutions to complex challenges

What is denormalised and normalised

Levels of nf and explain

What is facts in data warehousing

Data set and data frame

**Sql-views**

* Avoid full access
* Security purpose

Rollback on view-------study

To be successful in this role you will have: • experience in deploying applications and/ or data services to the cloud using CI/ CD approach, preferably with experience of containerised micro-service delivery • experience in agile software development within Azure • experience in designing and developing API's for data - e.g. search API's, feed API's etc • experience in designing and developing presentation layers for data rich applications • experience with developing with Azure storage services, especially Azure SQL server • the ability to pass security clearance, backed by the right to work in the UK

Key Responsibilities

• Create and maintain optimal data pipeline architecture.

• Assemble large, complex data sets that meet functional / non-functional business requirements.

• Identify, design, and implement internal process improvements: automating manual processes, optimizing data delivery, re-designing infrastructure for greater scalability, etc.

• Build the infrastructure required for optimal extraction, transformation, and loading of data from a wide variety of data sources using SQL and Azure ‘big data’ technologies.

• Build analytics tools that utilize the data pipeline to provide actionable insights into customer acquisition, operational efficiency and other key business performance metrics.

• Work with stakeholders including the Executive, Product, Data and Design teams to assist with data-related technical issues and support their data infrastructure needs.

• Keep our data separated and secure across national boundaries through multiple data centers and regions.

• Create data tools for analytics and data scientist team members that assist them in building and optimizing our product into an innovative industry leader.

• Work with data and analytics experts to strive for greater functionality in our data systems.

• Showcasing proof points and case studies.

Qualifications

Minimum Requirements:

• We are looking for a candidate who has attained a Graduate degree in Computer Science, Statistics, Informatics, Information Systems or another quantitative field.

• Advanced working SQL knowledge and experience working with relational databases, query authoring (SQL) as well as working familiarity with a variety of databases.

• Advanced working SQL/nosql, ADLS, Databricks, ADF, Azure DevOps

• Experience building and optimizing ‘big data’ data pipelines, architectures and data sets.

• Experience performing root cause analysis on internal and external data and processes to answer specific business questions and identify opportunities for improvement.

• Strong analytic skills related to working with unstructured datasets.

• Build processes supporting data transformation, data structures, metadata, dependency and workload management.

• A successful history of manipulating, processing and extracting value from large, disconnected datasets.

• Working knowledge of message queuing, stream processing, and highly scalable ‘big data’ data stores.

• Strong project management and organizational skills.

• Experience supporting and working with cross-functional teams in a dynamic environment.

12 types of Databases and when to use them:  
  
1) Relational Databases (SQL):  
- Use when your data is structured and consistent.  
- Supports ACID transactions and complex queries.  
- Examples: MySQL, PostgreSQL, Oracle.  
  
2) Key-Value Store:  
- Use when data model is based on key-value pairs and requires high scalability and availability.  
- Provides lightning-fast data retrieval and high throughput.  
- Examples: Aerospike, DynamoDB  
  
3) Document Databases:  
- Handles semi-structured data with varying fields.  
- Provides schema flexibility and horizontal scaling.  
- Examples: MongoDB, Couchbase  
  
4) Graph Databases:  
- Perfect for data with complex relationships.  
- Used in applications like social networks and recommendation engines.  
- Examples: Neo4j, Amazon Neptune.  
  
5) Columnar Databases:  
- Data is stored by columns instead of rows to optimize reading from a column.  
- Great for applications that involve storing massive data sets and running analytical queries.  
- Examples: HBase, Redshift.  
  
6) Time-Series Databases:  
- Opt for this when dealing with time-series data like IoT sensor readings or server logs.  
- Great for efficient storage and retrieval of time-stamped data.  
- Examples: InfluxDB, Prometheus.  
  
7) In-Memory Databases:  
- When speed is of the essence, and you can afford to sacrifice persistence.  
- Ideal for caching, real-time analytics, and high-frequency trading.  
- Redis and Memcached are popular choices.  
  
8) Wide-Column Stores:  
- Use in applications with large volumes of data and high write throughput.  
- Best suited for analytical workloads and applications that require high availability.  
- Apache Cassandra is a prominent example.  
  
9) Search Engines:  
- When your primary use case revolves around full-text search.  
- Essential for applications that require searching of data content.  
- Elasticsearch and Solr are popular choices.  
  
10) Spatial Databases:  
- Used to store geographical and location-based data.  
- Choose for applications that require Spatial indexing, geospatial analytics.  
- Examples include PostGIS, CartoDB  
  
11) Blob Datastore:  
- Use in applications that requires storing large documents, images, audio and video files.  
- Provides high availability, durability and cost effective storage.  
- Examples include HDFS, Amazon S3  
  
12) Ledger Databases:  
- Used for maintaining a transparent, immutable, and cryptographically verifiable transaction log.  
- Useful for applications dealing with financial transactions and supply chain systems  
- Examples: Amazon QLDB, Azure SQL Ledger  
  
Tech world offers a rich ecosystem of databases and there's no one-size-fits-all solution.  
The choice of a database depends on your specific use case, data model, scalability needs, and budget.