**Data Engineer Job Description Template**

**Job Overview**

We are looking for a savvy Data Engineer to join our growing team of analytics experts. The hire will be responsible for expanding and optimizing our data and data pipeline architecture, as well as optimizing data flow and collection for cross functional teams. The ideal candidate is an experienced data pipeline builder and data wrangler who enjoys optimizing data systems and building them from the ground up. The Data Engineer will support our software developers, database architects, data analysts and data scientists on data initiatives and will ensure optimal data delivery architecture is consistent throughout ongoing projects. They must be self-directed and comfortable supporting the data needs of multiple teams, systems and products. The right candidate will be excited by the prospect of optimizing or even re-designing our company’s data architecture to support our next generation of products and data initiatives.

**Responsibilities for Data Engineer**

* 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 AWS ‘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 AWS 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.

**Qualifications for Data Engineer**

* Advanced working SQL knowledge and experience working with relational databases, query authoring (SQL) as well as working familiarity with a variety of databases.
* 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.
* We are looking for a candidate with 5+ years of experience in a Data Engineer role, who has attained a Graduate degree in Computer Science, Statistics, Informatics, Information Systems or another quantitative field. They should also have experience using the following software/tools:
  + Experience with big data tools: Hadoop, Spark, Kafka, etc.
  + Experience with relational SQL and NoSQL databases, including Postgres and Cassandra.
  + Experience with data pipeline and workflow management tools: Azkaban, Luigi, Airflow, etc.
  + Experience with AWS cloud services: EC2, EMR, RDS, Redshift
  + Experience with stream-processing systems: Storm, Spark-Streaming, etc.
  + Experience with object-oriented/object function scripting languages: Python, Java, C++, Scala, etc.

**Job Information**   
  
**Technical Responsibilities**

* Hands-on Engineering of Data Processing applications
* Experience with Big Data technologies, assisting customers in building software solutions on-prem or cloud.
* Very strong understanding of SQL alongside traditional/conventional data warehousing design patterns.
* Hands-on experience in developing Big Data applications using Hadoop technologies such as Spark (Scala or PySpark), Map Reduce (Java), YARN, HDFS, Hive, Impala, Sqoop, Oozie, HBase, Kudu.
* Good understanding of UNIX shell scripting.
* Develop analytical functionality and complex transformation that will be finally deployed in production data platforms.
* Good understanding of Big data design patterns
* Hands-on experience troubleshooting, optimizing, and enhancing the big data pipeline and bring improvements.
* Experience & good proficiency with at least one programming language (Java, Scala, Python)
* Extensive experience with at least one major Hadoop platform (Cloudera, Hortonworks, MapR)
* Worked and deployed at least one data engineering application using cloud-native data platform and services (example EMR or Redshift in AWS or Data factory in Azure or BigQuery in GCP)
* Strong troubleshooting and performance tuning skills.
* Very well versed with continuous integration and deployment procedure considering the bigdata stack on-prem or cloud environment.
* Ability to analyze business requirement user stories and translate them into system requirement specifications.
* Experience working under the agile delivery methodology.

**Behavioral Requirement**

* Demonstrated ability to work independently
* Good communication skills and documentation skills.
* Good & Collaborative Team Player.
* Good organizational and time management skills
* An ability to work to deadlines
* A good eye for detail
* Must be ready to learn and adapt to new technologies.

**Qualifications Requirements**

* BE/BTech in computer science Or MCA with sound industry experience (2-4 years)
* A minimum of 2 years' experience in hands-on development
* Experience in Java, Scala or Python (**Production level coding**)
* Experience in a cloud-based environment with PaaS & IaaS
* Work iteratively in a team with continuous collaboration

**About the Role**   
  
True Fit is looking for a Software Developer to work on our Data Normalization Squad in Mumbai, India. The Data Normalization Squad designs and builds the software responsible for building Fashion Genome (Data Lake) which is the core behind True Fits personalized fashion recommendations.

**Responsibilities**

* **Develop Software:** Design, implement, test, and support high-quality software, used by millions of consumers every day, in an agile, iterative development culture.
* **Build Services & apply Machine Learning:** Implement, test, and support Big Data-driven, ML-based algorithms along with services and web applications to serve them to end-users.
* **Focus on Performance:** Make that software hit our Service-Level Agreements, improving data lake coverage to get more and more products available for personalised recommendations.
* **Learn:** Contribute to team success by learning new technologies and algorithms, often while designing and building the software.
* **Collaborate:** Work closely with other team members to invent, prioritize, build, and support our applications.

**Qualification & Skills**

* **1+ years as a professional software engineer** with hands-on experience with functional and/or object-oriented programming languages like: Scala, Kotlin, Java, C++, Python.
* **Experience developing and releasing commercial software**, especially software with a large user base or strict performance requirements.
* **Collaboration skills:** Work together with members of various internal teams, including listening and communicating.
* **Execution skills:** Gets things done both independently and collaboratively, and understands when each is appropriate.
* **Knowledge of Signal Processing:** Algorithms, Linear Algebra, Machine Learning, Computer Systems/Architecture.
* **Experience with Machine learning**, especially in commercial environments.
* **Passion:** Has an interest in our business domain (fashion), solution space (Big Data Machine Learning), or technology stack.
* **Undergraduate degree in Computer Science or a quantitative field** (e.g., Math, Physics, Engineering).

Introduction   
• The role will involve running workshops to define the problem statement, identifying the data that is required, cleaning up the data, running feature engineering techniques and building / testing ML/DL models. A combination of one of more of IBM AI technologies from the Watson portfolio, Open Source and 3rd Party technologies will be used to be deliver the required outcomes.   
  
Your Role and Responsibilities

* The responsibility of a Data Scientists will include working with Sales, Tech Sales & IBM Clients to define an approach to Machine/Deep Learning problems, building the models and delivering them.

Required Technical and Professional Expertise

* Python - 3-5 Years
* Machine learning Model (Regression, Classification, Unsupervised techniques, ensemble modeling) - 2-3 Years (Mtech) / 0-1 Years (PhD)
* Deep Learning models (Image, NLP, Structured data) - 1-2 Years (Mtech) / 0-1 Years (PhD)
* Tech Sales, Sellers, Architects
* Cloud native technologies
* Agile development

Preferred Technical and Professional Expertise

* Specialisation in any of the following fields - NLP, NLU, NLG, Computer Vision, Facial Recognition - 0-2 Years
* Good communication
* Active listener

AWS

**Position Title**   
F1 AWS Technical Architect   
  
07-Jul-2020   
  
**Job ID**   
**297822BR**   
  
**Job Description**   
Handle the development of architectural standards, and the direction and design of architecture for a business or technical domain. Manage efforts to elaborate and communicate the architecture of the given domain and/or the development of architectural solutions for projects and programs.   
Develop architectural solutions for a complex business area or deliver architectural services or governance activities. Elaborate the solution architecture for a specific solution based on architecture-significant requirements and company-wide architecture standards. Utilize architecture patterns to suggest the most adequate utilization of technical platforms in support of the holistic solution architecture design. Maintain and evolve Architecture tools and platforms, principles, policies and standards   
•Deliver technical spikes / evaluations on the AWS platform   
•Coordinate and document tech spike scope with relevant stakeholders   
•Implement tech spike against defined scope and strongly document design / outcomes.   
•Present summary outcomes and recommendations session on completion of spike to stakeholder audience.   
•Support product teams in trouble-shooting deep technical issues and providing technical guidance. Drives and contributes to high quality platform architecture documentation that is transparent, consistent, well-structured and accurate.   
•Support use case on boarding onto the platform and architectural analysis / documentation of the use case architecture.   
  
**Minimum requirements**   
•University degree or equivalent. –   
•At least 8-12 years of experience in IT, predominantly in Big Data and Advanced Analytics   
•Strong AWS hands-on expertise and programming background   
•Good knowledge of Big Data frameworks and related technologies (Spark, Kafka, Hadoop, NoSQL...)   
•Experienced in applying IaC and DevOps best practices   
•Hands-on architect having strong experience in programming languages primarily Java, Python, Unix shell scripts   
•Strong in API based architecture and concept. Able to do quick PoC using API integration and development   
•Good knowledge of DevOps tools (Artifactory, CI/CD Frameworks)   
•Basic knowledge of machine learning and data science concepts and related technologies (RStudio, Jupyter)   
•Good communication, stakeholder engagement and technical writing skills   
•Agile and scrum methodologies knowledge.   
  
Why consider Novartis?   
  
799 million. That’s how many lives our products touched in 2019. And while we’re proud of that fact, in this world of digital and technological transformation, we must also ask ourselves this: how can we continue to improve and extend even more people’s lives?   
  
We believe the answers are found when curious, courageous and collaborative people like you are brought together in an inspiring environment. Where you’re given opportunities to explore the power of digital and data. Where you’re empowered to risk failure by taking smart risks, and where you’re surrounded by people who share your determination to tackle the world’s toughest medical challenges.

DevOps Engineer  
  
**Job Description:**   
  
Designs, develops, troubleshoots and debugs software programs for software enhancements and new products. Develops software including operating systems, compilers, routers,   
  
The Vertica DevOps team seeks engineers to further enhance the automated systems that facilitate the development of our Big Data Analytics Platform.   
  
In this role you will design and implement improvements to internal systems using cutting-edge techniques and tools. Projects may include adding automated reporting to various script frameworks, building monitoring tools that observe testing progress and product health, developing centralized job scheduler systems to efficiently run tests, building log analysis tools to reduce manual inspection, and others. There are opportunities at all points along the DevOps toolchain to match your interests. It's an exciting time to join the team! Vertica is growing quickly to meet the high demand for Big Data solutions, and the DevOps team spearheads our culture of automation.   
  
**Required Skills**

* Bash & Python scripting
* Excellent communication and collaboration skills
* Integration and maintenance of automated tools
* Strong analytical and problem solving skills

**Desired Skills**

* Configuration software: Ansible
* Programmatic experience with AWS or GCP storage & machine images
* Experience on various Linux distributions
* Docker management and troubleshooting
* Familiarity with simple RESTful systems