

Data engineer

Profile I am an experienced Data Engineer with a past as Business Intelligence developer. Working in IT since 1997. Worked for companies like Booking.com, Rabo, ING and VodafoneZiggo. Mostly in scrum teams. Good analytical and communication skills. Presented knowledge sharing sessions and published IT related articles. As continuous learner I enjoy the big data ecosystem.

Technical Skills

- Programming Languages: Python, pyspark, Linux command line, SQL, scala
- Data Modelling: Dimensional modelling, Data vault
- Cloud:
 - AWS (Emr, Athena, Dms, Ecs, Fargate, Lambda)
 - GCP (BigQuery, Dataflow Dataproc)
 - Databricks

Education

- Ma. (drs), Maatschappijgeschiedenis, Erasmus universiteit Rotterdam, [1986 - 1993] Thesis: ISBN13: 9789065501165. Nominated 'Hiskus-scriptieprijs' 1993/1994
- Oracle systeemontwerp & -bouw, Fast Lane, Maarssen, Bussum [aug. 1996 - feb. 1997]. *SQLPlus*, *PL/SQL*, *SQLForms* 3.0, Developer/2000 (Oracle Forms 4.5, Reports 2.5), NIAM, Oracle 7 DBA, Designer/2000, Unix, ERM

Projects

Data engineer Booking.com (2023-10 - now)

- Tools: Spark, pyspark, Hive, oozie, Airflow, Astronomer, Kubernetes, Snowflake, DBT, linux, docker, Python, Hadoop
- Built etl-pipelines, gave presentations, support migration from on premise hadoop to cloud, snowflake/dbt.
- Optimized (spark) workflows to significantly improve performance and stability. Extended monitoring and logging.
- Poc to optimize ingestion pipelines to significantly improve runtimes and reduce complexity.
- Poc working with DBT on Snowflake.

Data engineer VodafoneZiggo (2021-01 - 2023-09)

- Tools: Hadoop, MapR, Spark, pyspark, scala, Hive, oozie, Airflow, Snowflake, linux, docker, Python, AWS

- Worked on Spark applications on Aws Emr, using mainly PySpark and sometimes Scala. Using sources like: daily or hourly dump files, Kafka streams and api-calls. Set up Snowflake to Data Lake ingestion via a Poc and define standards. Make application (Python, Aws Fargate). Set up Data Lake ingestion via Aws Dms via Poc, discussed architecture. Make application (Aws (Dms, Lambda, Fargate)). Start of project: using Scala and PySpark on onPrem cluster (MapR).

Data engineer Rabo Nederland (2017-07 - 2020-12)

- Tools: Hadoop, Spark, pyspark, Scala, Hive, Kafka, Airflow, linux, Python, Nifi, Hortonworks, Cloudera, sqoop
- Built etl pipelines, both batch and streaming. Worked with parquet, avro, orc, json etc. Used hdfs, HBase and Hive as sink. Data preparation via Google BigQuery. Used Databricks. Gave knowledge sharing sessions about HBase. Set up standards for the new Rabo Data Lake.

Data Warehouse Developer Rabo Nederland (2015-11 - 2017-06)

- Built deployment automation of Oracle and PowerCenter sources using Python. Built etl-flows using Oracle PL/SQL, PowerCenter, Python.

Data Warehouse Developer, Rabo International (2013-07 - 2015-10)

Data Warehouse Developer, Liberty Global (LGI, UPC) (2011-10 - 2013-07)

Data Warehouse Developer, Tele2 (BBNED) (2010-11 - 2011-09)

Data Warehouse Designer/Developer, Abn Amro (Fortis) (2010-01 - 2010-10)

Data Warehouse Developer, Royal Bank of Scotland (RBS) (2008-06 – 2009-12)

Data Warehouse Designer/Developer, Uvv (2006-04 - 2008-06)

Data Warehouse Designer, Ing (2005-05 - 2006-03)

Courses

- Neo4j Fundamentals (Neo4j, Cypher, Graph data modeling), Neo4j GraphAcademy
- Data Engineering on Google Cloud Platform (5 courses), Coursera
- Functional Programming Principles in Scala, Coursera
- Data Vault (modelling), VanderLek Advies
- Dimensional modelling, VanderLek Advies