

Request for Change

Date: 19 February 2021

Requestor: NATO Communications and Information Agency

POC: Dr. Gernot FRIEDRICH

Email: gernot.friedrich@ncia.nato.int

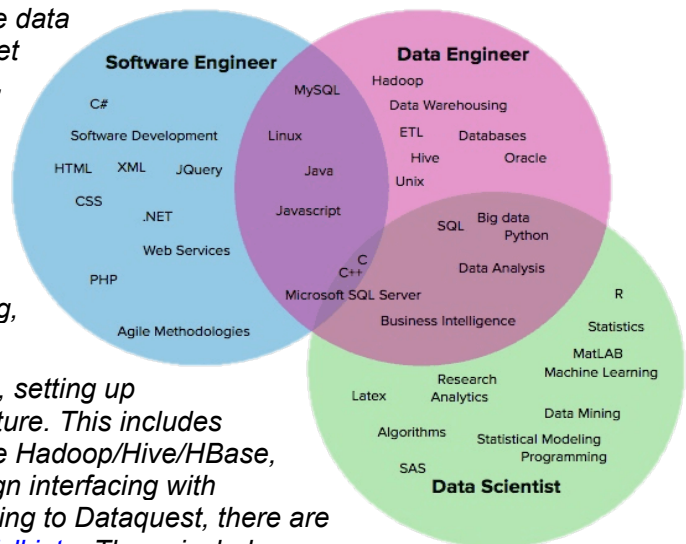
Target Version: SFIA v8

Rational:

Over the last decade, data infrastructure teams were stood up in larger companies and the new competency or profession of a “Data Engineer” became to emerge while Data Science also became a new discipline and was maturing. Inspired by software engineering, data engineers build tools, infrastructure, frameworks, and services. In fact, it is arguable that data engineering is much closer to software engineering than it is to a data science.

In relation to previously existing professions, the data engineering field can be thought of as a superset of business intelligence, data mining, warehousing, data analytics and that includes elements from software engineering and data science. This discipline also integrates specialization around the operation of so called “big data” distributed systems, along with concepts around machine learning and the extended Hadoop ecosystem, stream processing, and in computation at scale.

The data engineering jobs also cover designing, setting up and operating the organization’s data infrastructure. This includes tasks like setting up and operating platforms like Hadoop/Hive/HBase, Spark, or using hosted cloud services and design interfacing with associated BI and Data Analytics tools. According to Dataquest, there are [three main categories that data engineers can fall into](#). These include:



- **Generalist:** Generalists are typically found on small teams or in small companies. In this setting, data engineers wear many hats as one of the few “data-focused” people in the company. Generalists are often responsible for every step of the data process, from managing data to analysing it. Dataquest says this is a good role for anyone looking to transition from data science to data engineering, since smaller businesses will not need to worry as much about engineering “for scale.”
- **Pipeline-centric:** Often found in midsize companies, pipeline-centric data engineers work alongside data scientists to help make use of the data they collect. Pipeline-centric data engineers need “in-depth knowledge of distributed systems and computer science,” according to Dataquest.
- **Database-centric:** In larger organizations, where managing the flow of data is a full-time job, data engineers focus on analytics databases. Database-centric data engineers work with data warehouses across multiple databases and are responsible for developing table schemas.

There are a several certifications that are specific to data engineering:

- **Data Science Council of America (DASCA) Certified:**
 - [Associate Big Data Engineer](#)
 - [Senior Big Data Engineer](#)
- [Certificate in Engineering Excellence Big Data Analytics Optimization \(CPEE\)](#)
- **Amazon Web Services (AWS) Certified:** [Big Data Specialist](#)

- *Microsoft Certified: [Azure Data Engineer Associate](#)*
- *[Cloudera Certified Professional \(CCP\): Data Engineer](#)*
- *[Google Certified Professional Data Engineer](#)*
- *[IBM Certified Data Engineer – Big Data](#)*

It is suggested to add a new skill “Data Engineering” at levels 4, 5 and 6 to the SFIA Framework.

Data Engineering (DATE)

Overall description:

Designing, building, operationalizing, securing and monitoring distributed data infrastructures. Data engineers are managing, optimizing, overseeing and monitoring data retrieval, storage and distribution as well as preparing data for analytical or operational use to enable data-driven decision-making. The identification of data sources, processing concepts and methods and their translation into a coherent design, with a particular emphasis on security and compliance; scalability and efficiency; reliability and fidelity; and flexibility and portability. Data infrastructures (pipelines and stores) must be compatible with the overarching IT architectures, and adhere to legislative and corporate policies. Data engineers harvest both structured and unstructured data sets from different sources systems; integrating, consolidating and cleansing data; and structuring and storing it for use for predictive analytics, machine learning and data mining purposes. Data engineering requires a solid conceptual understanding of Data architecture, data models, relational and non-relational database design, query execution and optimization, artificial intelligence, machine learning models and logical operations.

Level 4 description:

Designs data pipelines and stores using appropriate modelling techniques following agreed architectures, design standards, data engineering patterns and methodology. Identifies and evaluates alternative design options and trade-offs. Creates multiple design views to address the concerns of the different stakeholders and to handle both functional and non-functional requirements. Builds reference data sets and models, simulates or prototypes the behaviour and quality of proposed data engineering solutions to enable approval by stakeholders. Delivers and maintains data engineering solutions. Reviews, verifies and improves own designs against original requirements.

Level 5 description:

Adopts and adapts appropriate data engineering design methods, tools and techniques selecting appropriately from predictive (plan-driven) approaches or adaptive (iterative/agile) approaches, and ensures they are applied effectively. Designs large or complex data engineering solutions including interfacing with Business Intelligence, Data Analytics toolsets, AI/ML models etc. Undertakes impact analysis on major design options and trade-off. Makes recommendations, assesses, and manages associated risks. Reviews others' data engineering designs to ensure selection of appropriate technology (e.g. on premise, cloud based or hybrid data engineering solutions), efficient use of resources, and integration of solution components. Ensures that the data engineering solution design balances functional and non-functional requirements. Contributes to development of data engineering design policies and standards and selection of architecture components and building blocks. Plans and directs the automation of traditional data collection and management and business intelligence functions.

Level 6 description:

Develops organisational policies, standards, guidelines, and advanced data engineering patterns for data engineering solution designs. Champions the importance and value of data analytics principles and the selection of appropriate data engineering solution designs. Drives adoption of and adherence to relevant policies, standards, strategies and architectures. Leads data engineering solution design activities for strategic, large and complex data analytics pipelines and warehouses. Develops effective implementation and procurement strategies, consistent with specified requirements, architectures and constraints of performance and feasibility. Develops data engineering solutions requiring introduction of new technologies or new uses for existing technologies. Evaluates the suitability of on premise, cloud based and hybrid data engineering solutions considering policies, standards, cost-effectiveness, performance and corporate strategy.