In partnership with

Learning Note: Career progression in the field of data science

April 2024

Author: Dr Edwin Byusa, PhD

Introduction

Data practitioners are like modern-day explorers, navigating huge volumes of data to uncover insights that drive innovation and progress. In this dynamic field, stagnation is never an option. A data science career is a continuous journey that requires constant learning and practice. Because technologies evolve, methodologies shift, and new tools emerge at such a fast pace, staying abreast of these advancements is not a choice but a necessity.

This learning note will outline some of the paths to progression in the field of data science. It is based on a review of the literature as well as Cenfri's experiences in implementing the Rwanda Economy Digitalisation Programme (REDP). We look at upskilling paths and examine the indispensable role of continuous learning.

Options for those wanting to pursue a career in data science

"Data science" is used throughout this learning note as a generic term for the field of data work, but it is important to note that not all individuals pursuing a career in this field will become data scientists. There is no single career journey, but rather different pathways that individuals can undertake.

Likewise, organisations have different needs. Institutions with smaller data teams may require that a single individual fulfil multiple data roles. Some of the roles across the data value chain are illustrated in the diagram below:

Figure 1. Roles across the data value chain

A progressive pathway toward expertise in data science involves a strategic combination of foundational education, skill development through practical experiences, and continuous learning.

Foundational skills

Mastering Excel serves as a good foundation in the journey toward a career in data science, especially when complemented by academic programmes and specialised courses (Kim et al., 2018). Key competencies for success in data science include basic programming proficiency, comprehension of mathematical concepts, proficiency in advanced Excel and a sound grasp of statistical knowledge.

For individuals aspiring to achieve expertise in this field, obtaining a bachelor's or a college degree in disciplines such as Computer Science, Statistics, Mathematics, Engineering or Economics can lay a solid foundation on which to build specific data skills. Some aspiring data scientists choose to supplement their qualifications by pursuing a master’s degree or specialisation in areas like advanced data analysis, machine learning or big data.

Any role that involves interaction with data is a viable starting position. The more exposure to data, the better. There is so much to learn about data, and a successful career consists of learning as many little things as possible along the way. An internship may be the ideal starting point.

Exposure to data enables individuals to better understand data sources and learn to problem-solve issues that may arise. For example, understanding how separate systems use different date formats might seem simple but these discrepancies can cause significant challenges. Understanding the origin of data enables individuals to develop better models by understanding the underlying data, and its relation to the real world.

Although it is not necessary to be a software engineer, familiarity with the storage, collection and basic reporting of data is valuable. Presenting simple but accurate and relevant figures to decision-makers is a useful skill. Most companies do not use machine learning but would still benefit from having a solid reporting system with

accurate data. Real-world problems are often not the same as classroom problems, making practical experience and skill an important commodity.

Select roles, skills and upskilling possibilities

Role	Description	Foundational knowledge	Competencies	Upskilling
Junior Data Scientist	Works on data projects under the guidance of more experienced colleagues. Responsibilities typically involve cleaning, organising and analysing data using statistical methods, and programming languages such as Python or R.	Proficiency in Python/R programming, expertise in data wrangling, statistical analysis, and basic knowledge of machine learning.	Problem-solving acumen, curiosity, and collaborative teamwork.	Participation in online courses, attendance at workshops, engagement in mentorship programmes, and involvement in hands-on projects.
Data Analyst	Analyses data to help companies make informed decisions.	Proficiency in SQL, Python or R, familiarity with data visualisation tools (e.g., Tableau, Power BI), and basic statistical analysis.	Attention to detail, analytical thinking, and effective communication.	Online courses, boot camps, and hands-on practice with real datasets.
Data Engineer	Data Engineers excel in building and maintaining systems that store and process data. Data engineers work with big data technologies and cloud platforms to ensure data is accessible and ready for analysis.	SQL, Python, big data technologies (e.g., Hadoop, Spark), and cloud computing platforms (e.g., AWS, Azure).	Technical proficiency, problem-solving, and attention to detail.	Online courses, hands-on projects, and certifications in specific technologies.
Systems Architect	Data/System Architects design the blueprints for data management systems, ensuring they meet the needs of both the data scientists and the business (Mohammad et al., 2014).	Data modelling, ETL, database design, data warehousing, and enterprise architecture frameworks.	Strategic thinking, technical proficiency, and effective communication.	Experience in data engineering, certifications in architecture frameworks, and continued learning in data management.

Data Steward	Data Stewards are responsible for ensuring data is accurate, secure, and used responsibly (Tractenberg & Tractenberg, 2019). They develop policies for data usage and ensure compliance with regulations. This is an important role and foundation in the field of data-driven approaches (Teperek et al., 2018).	Data quality management, data governance frameworks, and regulatory compliance (e.g., GDPR, HIPAA).	Attention to detail, organisational skills, and ethical judgment.	Training in data governance, experience in data management roles, and certifications in data quality or governance.
Business Intelligence Analyst	Business Intelligence (BI) Analysts specialise in transforming data into insights for business decisions, this role involves creating dashboards and reports to help leaders understand the data (Chiang et al., 2012).	Data visualisation tools (s.g., Tableau, Power BI), SQL, business acumen, and communication skills.	Analytical thinking, business understanding, and effective communication.	Online courses in BI tools, experience in data analysis, and business courses or an MBA.
Data Scientist	As a more complex problem solver, a Data Scientist is a professional who specialises in analysing, interpreting, and extracting meaningful insights from large and complex datasets (Kim et al., 2018). They use a combination of statistical, mathematical, programming, and domain-specific knowledge to identify patterns, trends, and relationships within data. Data Scientists employ various tools	Advanced machine learning, statistical modelling, data visualisation, and big data technologies (e.g., Hadoop, Spark).	Business acumen, critical thinking, and effective communication.	Work experience, online courses, internships, conferences, and research papers.

	and techniques, such as machine learning algorithms, data mining, and predictive modelling, to solve complex problems and drive decision-making in organisations (Cao, 2017).			
Machine Learning Engineer	ML individuals focus on building and fine-tuning the algorithms that allow computers to learn from data. They require a deep understanding of machine learning, programming, and working with large datasets (Pannala Sai Pranay Reddy & Sai Manvith Markonda, 2023)	Deep learning, natural language processing, computer vision, and software engineering best practices.	Technical proficiency, innovation, and efficiency.	Specialised online courses, open-source contributions, and collaboration with research teams.
Senior Data Scientist	Senior Data Scientists lead projects, provide support and guidance to less experienced team members, and contribute to important business decisions. Senior data scientists possess expert knowledge in data analysis and machine learning (Kim et al., 2018).	Leadership in data science projects, expertise in advanced analytics, and proficiency in cloud computing.	Leadership, mentorship, and strategic thinking.	Leadership training, industry experience, and continuous learning in both data science and executive programmes.
Data Science Manager	Oversees a team of data scientists, ensuring that projects are completed successfully and aligned with the company's goals. They need strong leadership and communication skills.	Project management, team management, advanced data science knowledge, and stakeholder management.	Management skills, decision-making, and effective communication.	Management courses, real-world management experience, and coaching.

As highlighted earlier, progression in the field of data science depends on many factors and will look different for different people. By way of example, an individual might start as a Junior Data Scientist and work on data projects under the guidance of more experienced colleagues. They could then progress to a Data Scientist and, with sufficient experience, take on the role of a Senior Data Scientist, enabling them to lead projects and contribute to important business decisions.

In larger organisations, a Director of Data Science would set the strategic direction for the company's data science efforts, oversee multiple teams, and ensure that all data work aligns with the overall business strategy. In government institutions, this role might be filled by a Chief Data Officer (CDO). As one of the top executives, the CDO assumes responsibility for the company's entire data strategy, making sure data is used effectively across the organisation to achieve business goals. In an alternate example, an individual may begin their career with a broad internship before accepting a position as a Data Analyst. Once they have gained sufficient exposure in the field of data science, they might opt to specialise in, for example, machine learning. This would allow them to go on to secure a position as a Machine Learning Engineer.

An alternative to focusing on data analysis is available to those opting to pursue the data engineering or systems architecture routes.

Further considerations

As data professionals progress in their careers, they might choose to specialise in areas like text analysis (Natural Language Processing), image recognition (Computer Vision), advanced learning algorithms (Deep Learning) or big data engineering. Continuous learning through online courses, workshops, and conferences is crucial to stay updated with the latest technologies and methodologies.

There are other opportunities to stay abreast of new developments such as participating in communities of practice (CoPs), joining online communities with active data competitions that allow participants to test their skills (e.g.; Zindi or Kaggle), attending industry events, and voluntarily contributing to projects. Networking is a valuable tool for career growth in the dynamic field of data science, and all of these activities provide opportunities to meet other professionals and learn from them.

Lesson from the Rwanda Economy Digitalisation Programme

Cenfri's Rwanda Economy Digitalisation Programme (REDP) focuses on leveraging data insights to improve policymaking, catalyse innovation and improve livelihoods. Data-driven decision-making requires institutional buy-in and so, the REDP identified and adopted three levels of capacity development:

1. ***The systems-level*** involves policy and strategic management, norms and procedures, and information management. Some examples here are the development of data frameworks that include data sharing, data standardisation, data classification and data cataloguing guidance. Through this programme, a data-sharing policy draft has been prepared for the Government of Rwanda.
2. ***The institutional level*** consists of people, funds, processes, data, and resource and information management. Some examples here are the development of an organisational data strategy and other data governance

interventions. In addition, different activities are organised targeting internal change management in partner institutions. These activities include executive courses in data literacy among others.

3. ***The personal level*** involves the skills, knowledge, competencies, and practice of implementing, monitoring, reporting and accountability through data-driven approaches. In collaboration with its programme partners and associates, Cenfri has offered several data-related training courses that cater to people with different levels of prior exposure to the field of data science. The 18-month data fellowships provided young graduates with valuable opportunities to apply their academic data knowledge within public sector institutions.

Professionals aspiring for growth within a particular working environment would often need to strategically position themselves by first mastering the system's needs, orientation, priorities, and vision. They can then begin understanding the institutional dynamics and invest in personal development.

Conclusion

While a career in the field of data science can result in numerous growth opportunities across the globe, data professionals should pay attention to advances in artificial intelligence (AI) and carefully consider its future impact on their area of data specialisation. Data professionals can rely on some of the practices mentioned to enhance their skills and position themselves at the forefront of the digital transformation journey.

References

Antamoshkina, O. I., Zinina, O. V., & Olentsova, J. A. (2020). Methodology of building a master's individual educational route for effective development of professional competencies. *Journal of Physics: Conference Series*, *1691*(1). https://doi.org/10.1088/1742-6596/1691/1/012207

Cao, L. (2017). Data science: A comprehensive overview. In *ACM Computing Surveys* (Vol. 50, Issue 3). Association for Computing Machinery. https://doi.org/10.1145/3076253

Chiang, R. H. L., Goes, P., & Stohr, E. A. (2012). Business Intelligence and Analytics education, and program development: A unique opportunity for the Information Systems discipline. In *ACM Transactions on Management Information Systems* (Vol. 3, Issue 3). https://doi.org/10.1145/2361256.2361257

Kim, M., Zimmermann, T., Deline, R., & Begel, A. (2018). Data scientists in software teams: State of the art and challenges. *IEEE Transactions on Software Engineering*, *44*(11), 1024–1038. https://doi.org/10.1109/TSE.2017.2754374

Mohammad, A., Mcheick, H., & Grant, E. (2014). Big data architecture evolution: 2014 and beyond. *DIVANet 2014 - Proceedings of the 4th ACM Symposium on Development and Analysis of Intelligent Vehicular Networks and Applications*, 139–144. https://doi.org/10.1145/2656346.2656358

Pannala Sai Pranay Reddy, & Sai Manvith Markonda. (2023). *From Software Engineer to being a Machine Learning Engineer- A Study on Skills and Responsibilities*.

Papenfuss, T., Sanchez, A., & Rubin, J. (2023). *Organizational Data Journey System Design and Management Program, Executive Director and Senior Lecturer*.

Radovilsky, Z., Hegde, V., & Acharya, A. (2018). Skills Requirements of Business Data Analytics and Data Science Jobs: A. *Comparative Analysis Journal of Supply Chain and Operations Management*, *16*(1).

Ruiz Castro, M., Van der Heijden, B., & Henderson, E. L. (2020). Catalysts in career transitions: Academic researchers transitioning into sustainable careers in data science. *Journal of Vocational Behavior*, *122*. https://doi.org/10.1016/j.jvb.2020.103479

Teperek, M., Cruz, M. J., Verbakel, E., BÃ¶hmer, J., & Dunning, A. (2018). Data Stewardship Addressing Disciplinary Data Management Needs. *International Journal of Digital Curation*, *13*(1), 141–149. https://doi.org/10.2218/ijdc.v13i1.604

Tractenberg, R. E., & Tractenberg, R. (2019). *Becoming a steward of data science*.