

Data Science Engineer

Data science engineers work alongside data scientists to feed, deploy, monitor and maintain models and other data products. They have understanding in data science as well as computer science expertise pertaining to production systems operations (DataOps/MLOps).

The roles and responsibilities of a data science engineer may include:

- prototyping and demonstrating solutions for clients in customer environments to enable further development;
- developing end-to-end (Data/MLOps) pipelines based on an in-depth understanding of cloud platforms, AI lifecycle, and business problems to ensure analytics solutions are delivered efficiently, predictably, and sustainably;
- building automation software to operate systems needed for data storage, data management, data science notebooks, distributed training, model repository, feature repository, continuous delivery, model serving, and monitoring;
- operating production AI systems and making sure they are available, scalable, and performant;
- building and sharing the technical expertise necessary to analyze and recommend enterprise-grade solutions for operationalizing AI or advanced analytical models;
- communicating best practices and tools among data science teams in order to improve productivity and avoid common mistakes;
- assembling different pieces in order to build an end-to-end, reliable, enterprise-grade production system;
- setting the required architecture and deployment processes for AI, from data ingestion to production and maintenance;
- providing technical advice to management and other scientists as it pertains to the operationalization of models.

Behavioural Competencies

Communication

Listening to others and communicating in an effective manner that fosters open communication.

Why this competency matters

Data Science Engineers must use active listening skills to ensure they clearly understand both explicit and implicit messages from team members. They must be able to adapt their communication style to the widely varying needs of many groups - management, project teams, technical staff, and clients. They also need to produce clear documentation related to methods and equations used. Data Science Engineers must be prepared to serve as liaison between the IT/IM organization and the machine learning project.

Foundation	Intermediate	Advanced
1	2	3

<ul style="list-style-type: none"> • Listens actively to ensure messages are understood. • Presents ideas and information in a clear and concise manner. • Produces documentation appropriate for the medium through which information is being presented. • Documents processes clearly, using plain language. 	<ul style="list-style-type: none"> • Probes to discover underlying needs, interests, issues, and motivations. • Adapts style, mode, and tone based on client reactions and issues being addressed. • Articulates linkages between evidence and recommended course of action. • Documents processes, tailoring the level of detail to multiple audiences. 	<ul style="list-style-type: none"> • Interprets complex and possibly contradictory information. • Uses varied communication vehicles and opportunities to promote dialog and develop shared understanding and consensus. • Writes documentation that conveys nuances to facilitate in-depth understanding of the topic. • Makes compelling case for all stages of a proposed initiative to senior decision makers. • Conveys and justifies complex recommendations to senior management in clear and non-technical terms.
---	--	--

Teamwork

Working collaboratively with others to achieve common goals and positive results.

Why this competency matters

Data Science Engineers need to continuously work with others while providing advice on the operationalization of machine learning models and other data products. They interact with internal and external clients and stakeholders as part of effective machine learning solution development. They must be prepared to use creativity and flexibility in addressing client needs. They must be able to work as part of a diverse team as different resources are often brought together to interact and develop options. When necessary, they must be able to take a leadership role, emphasizing team goals, helping to define the problem, and treating others with respect.

Foundation	Intermediate	Advanced
2	3	4
<ul style="list-style-type: none"> • Share relevant information to facilitate collaboration. • Accomplishes own share of work. • Works with others to clearly define problems and identify suitable solutions. 	<ul style="list-style-type: none"> • Helps stakeholder mitigate resistance to change in their environment. • Builds on successful initiatives to gain support for ideas. 	<ul style="list-style-type: none"> • Coaches, challenges, and helps others develop their skills. • Guides others in making complex decisions.

Analytical Thinking

Understanding when data can be used to inform or to support, as well as the process of interpreting data into identifiable problems and research questions.

Why this competency matters

Data Science Engineers must be able to both understand and respond to complex issues. They see the connections between problems and issues and manipulate that information in order to develop short- and long-term plans. They are capable of adapting their thinking style, using cause and effect relationships to analyze problems in a step-by-step way, interpreting information and developing recommendations for management, clients and other stakeholders. They systematically organize and compare various aspects of a problem or situation in order to resolve problems in a sound, logical and decisive manner.

Foundation	Intermediate	Advanced
2	3	4
<ul style="list-style-type: none">● Ask the right questions to identify relevant problems/issues.● Recognize connections and develop short-term plans and recommendations.● Use cause and effect relationships to analyze problems systematically.● Resolve problems in a sound, logical and decisive manner.	<ul style="list-style-type: none">● Understand and respond to complex analytical issues.● Develop short- and long-term plans and recommendations.● Use cause and effect relationships to analyze problems, interpret information and develop recommendations.	<ul style="list-style-type: none">● See connections between problems and issues to develop short- and long-term recommendations for management, clients, and stakeholders.● Systematically organize and compare various aspects to resolve problems.● Resolve problems in a sound, logical and decisive manner.

Ethics and privacy

Ensuring access, quality, and security while cleaning, processing, and transforming data for analytics to ensure access to accurate, reliable and high value information in support of data science and machine learning operations.

Why this competency matters

Data Science Engineers must understand the ethical basis of managing large data sets that contain private information in them, and be able to describe the advantages and disadvantages of the use of record level data to achieve business outcomes. They discuss ethical concerns with stakeholders and, when necessary, seek out and use appropriate disclosure procedures. They take a balanced approach to managing risk by implementing appropriate privacy and security measures, and share evidence, research and decision making openly. They have knowledge of the responsible use of artificial intelligence. They comply with ethical guidelines in the design and use of systems which automate decision

making. Data Science Engineers must be able to develop a method of collecting, storing and sharing data in accordance with laws, regulations and the ethical standards of the organization. They need to understand data relevancy to be able to assess biases in algorithms and ensure their outcomes are fair to everyone.

Foundation	Intermediate	Advanced
2	3	4
<ul style="list-style-type: none"> Understands ethical concerns arising from the use of historical data in making models or inferences. Applies ethical guidelines/regulations of the organization consistently. Able to identify and assess bias and ensure fair outcomes from analyses. Raises ethical concerns when necessary. 	<ul style="list-style-type: none"> Able to assess the advantages and disadvantages of record level data, and their impact on analyses. Familiar with ethical framework of the organization and uses appropriate disclosure procedures when necessary. 	<ul style="list-style-type: none"> Guides others in making complex ethical decisions. Fosters an environment of transparency, trust and respect. Advises on ways of accommodating and benefiting from differences and between groups. Ensures that standards and safeguards are in place to protect organizational integrity.

Promote innovation and guide change

Actively encourages exploration of data to solve business problems through bold thinking, experimentation and intelligent risk taking. Willing to entertain the possibility of failure, and learns from it to improve future undertakings.

Why this competency matters

Data Science Engineers have the courage and resilience to challenge convention. They seek opportunities for innovation and propose creative practices, concepts or products. They adjust practices to address lessons learned following setbacks and mistakes, and implement plans that respond to changes in direction and priorities. They demonstrate resilience, composure and a positive outlook in an environment of uncertainty and ambiguity.

Foundation	Intermediate	Advanced
0	2	3
N/A	<ul style="list-style-type: none"> Seek opportunities for automation, bringing forth creative practices, concepts or products. Demonstrates a positive attitude to change and a 	<ul style="list-style-type: none"> Stays informed of emerging trends, identifying new requirements. Continuously acquires and applies new knowledge to improve job performance.

	<p>willingness to try new approaches.</p> <ul style="list-style-type: none"> • Responds constructively to ambiguity and uncertainty; contributes ideas for improvements. 	<ul style="list-style-type: none"> • Explains rationale for change and promotes the benefits of change.
--	---	--

Achieving results

Mobilize and manage resources to deliver on the priorities of the Government, improve outcomes and add value. Consider the context, risks and business intelligence to support high-quality and timely decisions.

Why this competency matters

Data Science Engineers coordinate and prioritize work activities to contribute to organizational objectives and results. They increasingly play a key role in defining measurable KPIs for projects and programs. They follow through on operational plans and revise them when priorities or conditions change. They inform decision-making with sound understanding of context, data and evidence. They take ownership of MLOps and DataOps processes and acknowledge the impact and outcome of those processes. They take ownership and acknowledge the impact and outcome of their decisions.

Foundation	Intermediate	Advanced
2	3	4
<ul style="list-style-type: none"> • Accomplishes own share of work. • Responds to changes in organizational priority, revising own work when conditions change. • Plans for contingencies to deal with unforeseen events or setbacks. • Breaks activities into smaller components to facilitate completion. 	<ul style="list-style-type: none"> • Evaluates project plans to ensure that goals are reached. • Contributes to annual plans for the work unit considering a range of factors in the planning process. 	<ul style="list-style-type: none"> • Tracks progress of projects and adjusts as needed to meet strategic and/or operational objectives. • Negotiates commitments and deadlines. • Ensures development and use of objective criteria to measure and improve organisational work. • Develops strategic plans considering short-term requirements as well as long-term direction.

Technical Competencies

Data Management

Ensuring access, quality, and security while cleaning, processing, and transforming data for analytics to ensure access to accurate, reliable and high value information in support of data science and machine learning operations.

Why this competency matters

Data Science Engineers must be able to demonstrate working level knowledge of DBMS applications and data lakes. They are able to query and process data from structured and unstructured sources, move data between cloud and on-premise environments, and implement data integrity safeguards.

Foundation	Intermediate	Advanced
2	3	4
<ul style="list-style-type: none">• Make effective use of relational databases and data lakes to address operational needs.• Query and process data from both structured and unstructured sources.• Move data between cloud and on-premise environments while maintaining appropriate safeguards.• Leverage diagnostic and monitoring tools to resolve problems.	<ul style="list-style-type: none">• Proficiently use of relational (SQL) databases and non-relational (NoSQL) databases.• Is able to store, process, monitor, and analyze streaming data, for example data from IoT sensors.• Make recommendations on logical models.	<ul style="list-style-type: none">• Research, pilot, evaluate new technologies and standards.• Mentor people and provide input/guidance to cross-functional teams.• Solve unusual problems or problems with significant impact on the business.

Programming

Knowledge and ability to design, define, construct, enhance, support, and maintain software associated with the operationalization of data or machine learning operations.

Why this competency matters

Data Science Engineers must be able to use version control platforms to assist with collaboration. They consider privacy, accessibility, usability and interoperability. They must have knowledge of both commercial and open source software packages and solutions related to data science. They have knowledge of software construction, testing, configuration, deployment infrastructure and the range of system development methodologies and operating standards.

Data Science Engineers will develop, code, test, and review complex data management workflows. They must be able to understand how changes would affect ML operations. They make recommendations/decisions in application and program design, standards, and enhancements. They must be able to analyze and model business functions, processes and information flow within or between systems. They build software supporting systems related to data storage, data science notebooks, model repositories, feature repositories, model serving and monitoring. They also put models in production and ensure they are available, scalable and performant.

Foundation	Intermediate	Advanced
2	3	4
<ul style="list-style-type: none"> • Demonstrate ability to write efficient and maintainable code. • Write a program to parse data and develop machine learning models. • Ability to write program to retrieve data from cloud or through APIs. 	<ul style="list-style-type: none"> • Ability to develop reusable code artifacts that can be called. • Ability to develop efficient programs which can be adapted for distributed and parallel processing. • Demonstrate ability to create and deliver portable analytical and ML solutions. • Analyze and model business functions, processes and information flow within or between systems. 	<ul style="list-style-type: none"> • Demonstrate knowledge in multiple applications, data management systems and technologies or in a single area of expertise. • Make recommendations/decisions in application and program design, standards and enhancements. • Debug very complex or urgent problems. • Provide guidance/mentor on programming practices and techniques to individuals and cross functional teams.

Mathematics and Statistics

Knowledge in a range of mathematical and statistical techniques, to understand and be able to apply them, and to know their underlying assumptions and limitations.

Why this competency matters

Data Science Engineers must have an understanding of algebra and probability theories and techniques that will be applied at multiple stages of data science work. They must be able to carry out operations on matrices, study the basic properties of functions and relations, and to indicate classes of equivalence relations. They understand the theoretical basis of analysis of variance, can describe the assumptions underlying statistical techniques, and understand the consequences of the assumptions not holding. They are able to depict the expected output of factor analysis, and effectively and accurately interpret statistical output. They can compare selected statistical methods and specify differences between them, selecting the most relevant statistical method for a specific analytical problem.

Foundation	Intermediate	Advanced
1	2	3
<ul style="list-style-type: none"> • Applies existing techniques and approaches. • Assesses trends and developments in the subject area, and identifies the potential impact on concepts, methods, techniques and 	<ul style="list-style-type: none"> • Demonstrate ability to perform exploratory data analysis and identify important relationships between variables. • Apply statistical techniques to extract valuable dataset from noise. 	<ul style="list-style-type: none"> • Provides subject matter advice to users within and outside the organization. • Writes and presents detailed subject matter and technical concepts, definitions and terms using common terminology.

<p>procedures.</p> <ul style="list-style-type: none"> • Conduct descriptive analysis of data. • Analyze and evaluate complex data problems. 	<ul style="list-style-type: none"> • Understands the pros and cons of various statistical tests and when they should be applied. • Possess knowledge of several statistical concepts, including statistical significance, regression, and hypothesis testing. 	
---	---	--

Project management

Knowledge and ability to apply agile project management principles and practices during the planning, implementation, monitoring, and completion of projects, ensuring effective management of scope, resources, time, cost, quality, risk, and communications.

Why this competency matters

Data Science Engineers need to understand the different project management approaches applicable to data and digital projects, including agile methodologies and project reporting. They apply formal project management principles and practices during the planning, implementation, monitoring and completion of projects. They identify issues and escalate appropriately to minimize project impacts. They participate in the development of project plans (e.g. project charters, work breakdown structure, estimates, change management plans, communication plans). They are also comfortable working on projects of different size, from proofs of concept to large and constantly evolving projects, through iterative development. They are able to independently manage small projects or components of larger projects, working closely with other team members to deliver work in small increments.

Foundation	Intermediate	Advanced
1	2	3
<ul style="list-style-type: none"> • Work as part of an agile team. • Identify and escalate issues and potential delays. • Sets priorities for tasks in order of importance. 	<ul style="list-style-type: none"> • Understand project reporting. • Develop simple project plans, including work breakdowns and estimates. • Manage small, straightforward projects or specific components of larger projects. 	<ul style="list-style-type: none"> • Manage a multi-stage data science project. • Develop a project plan, including timelines, deliverables, milestones and costs. • Identify potential roadblocks and risks.

Machine Learning

Possess a combination of knowledge and skills in developing self-learning algorithms, including the application of open source machine learning algorithms and libraries.

Why this competency matters

Data Science Engineers must be able to understand different types of machine learning techniques (supervised, unsupervised, semi-supervised), related algorithms, and their advantages and disadvantages in terms of performance, speed, interpretability, and other factors in order to effectively operationalize the work of analysts. They must understand the differences between machine learning and other analytical techniques.

Foundation	Intermediate	Advanced
1	2	3
<ul style="list-style-type: none"> • Uses advanced tools and techniques to perform data exploration such as data mining, web scraping. • Appropriately uses predictive modelling, time series/forecasting, clustering, principal component analysis and other techniques. 	<ul style="list-style-type: none"> • Understanding “how” and “which” machine learning techniques is appropriate to solve business problems. • Ability to apply machine learning algorithms as defined in libraries to build and train AI solutions. • Explains results obtained to stakeholders. 	<ul style="list-style-type: none"> • Represent subject matter on development projects, task forces or committees in various roles. • Develop and implement machine learning algorithms for use in building and training AI solutions. • Possess knowledge about recent advances in application of DevOps and MLOps techniques.

Data Visualization

The process of translating data into an accessible format utilising various tools and processes such as charts, graphs, maps, dashboards and other formats to aid others in seeing and understanding trends, outliers and patterns in data.

Why this competency matters

Data Science Engineers must be able to prepare data sets for visualization in a format best suited to the communication medium. They are able to communicate complex concepts by applying the adequate visualization technique to the data or analytical output at hand. They are able to simplify complex theories and data through visualization by focusing on key areas. Data Science Engineers are able to integrate visualization into the data science workflow.

Foundation	Intermediate	Advanced
1	2	2
<ul style="list-style-type: none"> • Basic understanding of the most appropriate visualization technique for various data types. • Ability to share results through dashboards or applications. • Ability to apply best 	<ul style="list-style-type: none"> • Evaluates graphical representations of data for accuracy or misrepresentation. • Includes correct and relevant references, labels and citations. • Produces visual representations of workflows to aid understanding. • Leverages tools to produce visuals from workflow definitions. • Demonstrate ability to create appropriate visualization to present patterns in a dataset. 	

practices and techniques when creating data visualization.	<ul style="list-style-type: none"> Ability to present and interpret data visualization concisely to management or business stakeholders.
--	---

Storytelling

Conveying results of work coherently and understandably through data visualization to present phenomena from a new perspective, using different approaches to build narratives in order for stakeholders to identify the best course of action.

Why this competency matters

Data Science Engineers must be able to translate data science outputs into an appropriate visual design, defining the context of the story. They leverage best practices in visual design to streamline and ensure story clarity while demonstrating the scientific basis for the analysis.

Foundation	Intermediate	Advanced
0	1	2
N/A	<ul style="list-style-type: none"> Presents information clearly and coherently Displays holistic information, telling complete stories rather than presenting selective evidence. 	<ul style="list-style-type: none"> Assesses audience needs, familiarity with data and understanding of subject matter. Ensures data presentations link directly to original questions and/or line of thinking.

Business acumen

Understanding and dealing with the risks and opportunities that will likely lead to a positive outcome. Effectively communicating ideas to management, clients, and the public.

Why this competency matters

Data Science Engineers are able to deal with large amounts of knowledge and translate it effectively for a non-technical audience. They maintain a working knowledge of current and upcoming trends, and are able to acquire the foundations of relevant disciplines, concepts, and tools. Their knowledge and analytical skills of business objectives provide answers to current problems, and are able to propose actionable insights that can improve product or service quality. They work with the client to fully understand their needs, and regularly report on progress for feedback. They are able to understand the need to adapt the production process to the expected product and functionality.

Foundation	Intermediate	Advanced
0	2	3

N/A	<ul style="list-style-type: none">● Uses the organization's formal and informal channels to accomplish work.● Applies analytical knowledge and skills to address current problems.● Leverages client's priorities and objectives to enable required actions and make recommendations.	<ul style="list-style-type: none">● Uses the organization's formal and informal channels to accomplish work.● Seeks perspectives from clients to accomplish work.● Applies analytical knowledge and skills to address current problems.● Leverages client's priorities and objectives to enable required actions and make recommendations.
-----	---	---