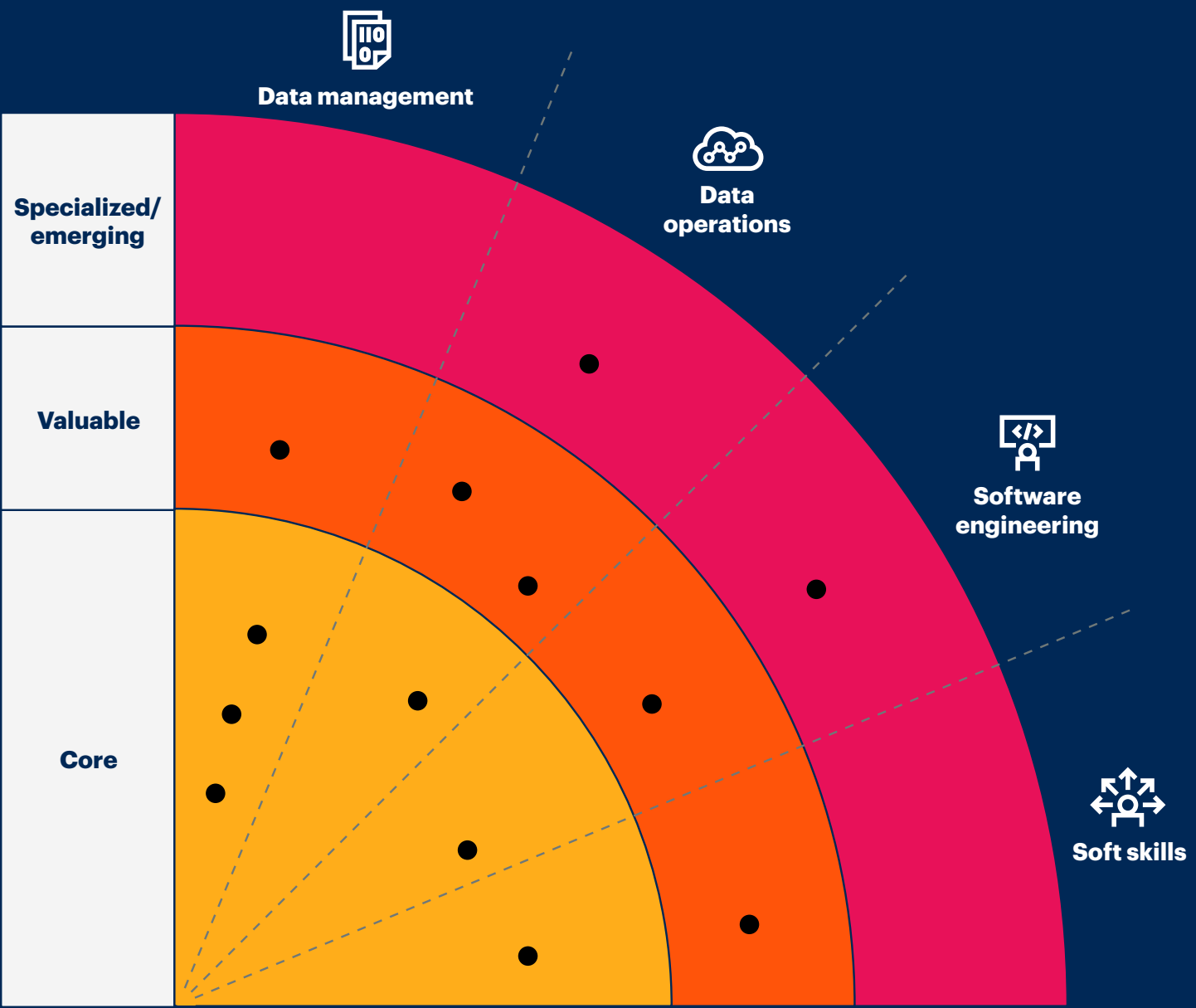
Gartner

Data and Analytics

Gartner Research

Essential Skills for Data Engineers



Essential Skills for Data Engineers to Succeed

6 May 2024 - ID G00810403 - 40 min read
By: Mayank Talwar, Zain Khan, Shubhankar Nandi
Initiatives: Data Management Solutions for Technical Professionals; Lead a World-Class D&A Organization

Data engineers play a key role in unlocking the value of data by designing and building systems to collect, store, transform, operationalize and deliver data at scale. Data and analytics technical professionals can use this document to identify and develop essential skills for data engineering.

Key Findings

- Data engineers focus primarily on building, managing and optimizing data pipelines that facilitate data movement. They also embrace generative AI (GenAI) and cloud adoption and develop processes to effectively build, store, manage and deliver data for a variety of use cases and consumers.
- As part of their role, data engineers are responsible for reducing manual data work and improving productivity. They employ innovative tools, techniques and architectures to automate common, repetitive, and tedious data preparation and integration tasks.
- Data engineers bring a level of discipline and repeatability to the deployment of data pipelines that can lead to lower costs, faster deployments as well as improvements in the quality, integrity and availability of data.

Recommendations

- Evolve your data engineering capabilities by developing skills aligned to four categories: data management, DataOps, software engineering and soft skills. Start with skills that most appeal to you or that require the most attention in your organization.
- Prioritize the core data engineering skills, such as SQL and database fundamentals, programming, data processing, data storage and data orchestration. These core skills are widely adopted, heavily used and have proven to provide significant benefits.
- Hone your skills and understanding of GenAI for both personal development and to achieve your organization's AI goals. As a data engineer, your skills will be needed to build and scale data engineering pipelines that extract, chunk, embed, and retrieve the right data and metadata that will support large language model (LLM) design patterns such as retrieval augmented generation (RAG).
- Develop soft skills, such as collaboration and communication, to build the relationships and credibility that you will need to successfully enable your organization's data engineering function.

Skills Overview

Data engineering is the discipline of translating data into "usable forms" in a controlled manner. It is used in building and operating D&A applications, data pipelines and data platforms. Through collaboration between the business and IT, data engineering makes the appropriate data accessible and available to various data users (e.g., data scientists or data analysts) at the right time.

Data engineers are responsible for building, managing and operationalizing data pipelines to support D&A use cases, while complying with data governance and data security requirements. Without data engineers, D&A initiatives will be more prone to additional costs, deployment delays and data integration, quality and availability issues.

Data engineers need to work alongside business users, domain experts, data analysts, data scientists and data architects to frame business problems, integrate the needed data and determine the best way to provision that data on demand. In the absence of dedicated machine learning (ML) engineers, they also need to support data scientists in training, and developing and operationalizing data science models, which requires proficiency in programming languages like Python.

Data and analytics technical professionals interested in, or looking to upskill to, data engineer roles must learn a few essential skills (if they haven't already). This research helps technical professionals answer the following question:

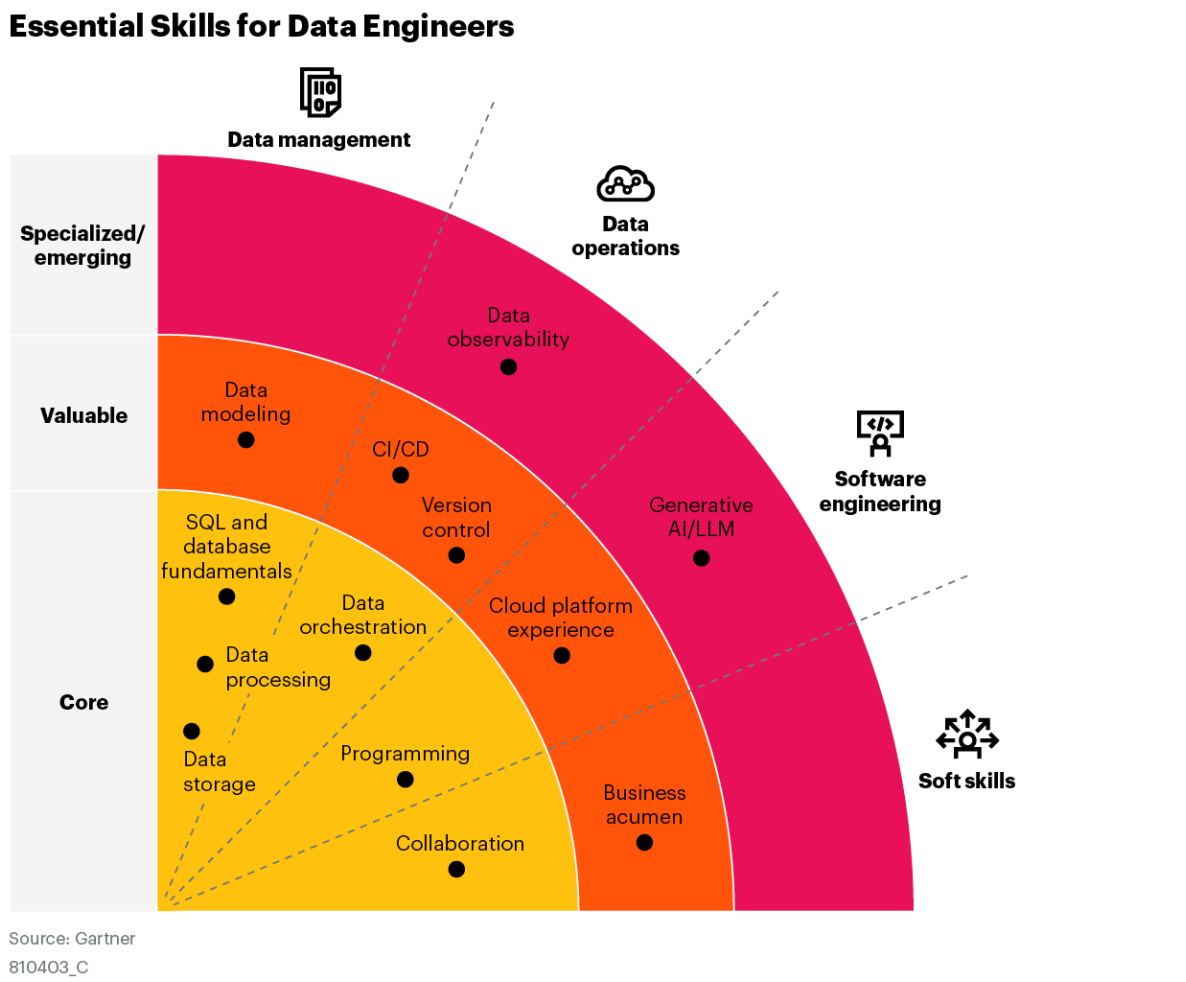
What skills must I learn to build a modern data engineering system?

Data Engineer Skills Scope

Figure 1 illustrates the skills that are most relevant for data engineering. These skill sets represent foundational capabilities and skills that have proven effective in a wide range of organizations. The following three sections of this research describe the:

- Core skills that you should learn first
- Valuable skills that you will need but require the foundation provided by the core skills
- Specialized skills required to address specific use cases

Figure 1: Essential Skills for Data Engineers



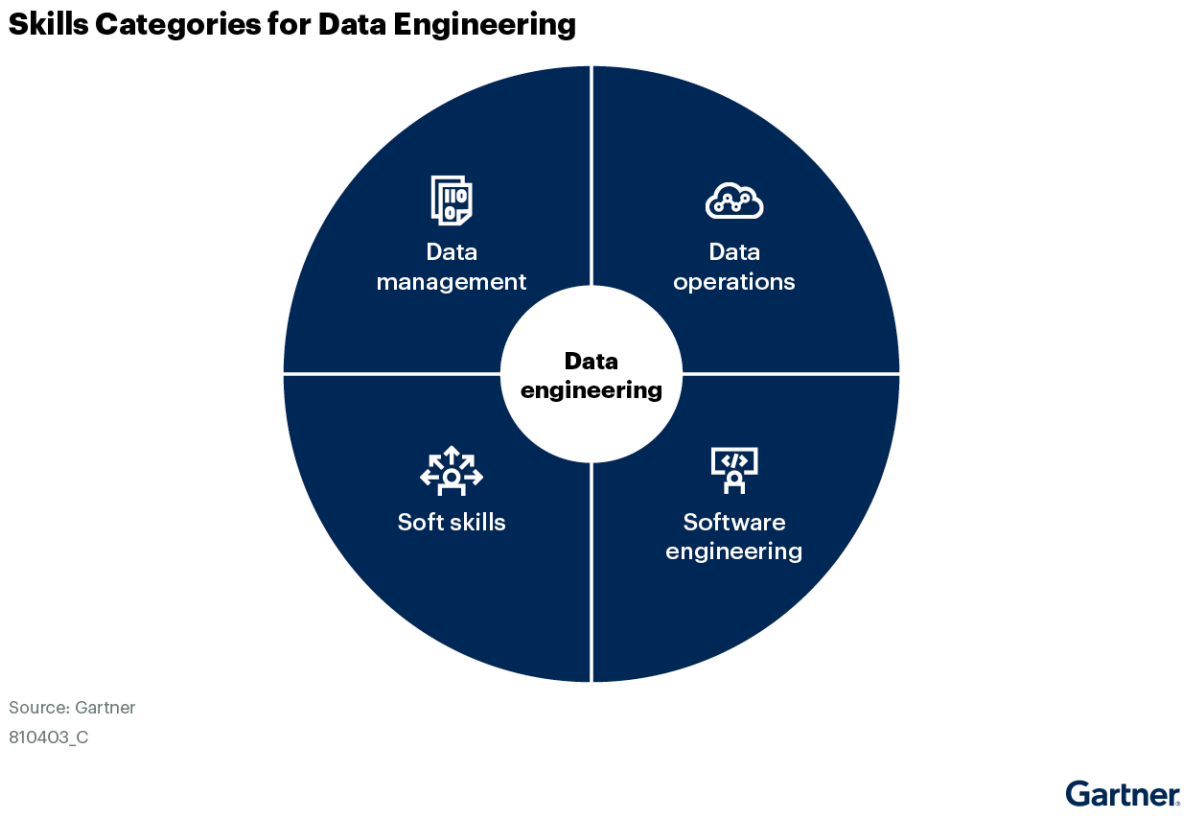
Skills Categorization

Skills for data engineers could be broken down into four main categories (see Figure 2):

- **Data management:** This category represents technology options and skill sets that data engineers must have to efficiently handle data integration, data storage and data/metadata management.
- **DataOps:** This category represents the skill sets that must be possessed by data engineers to manage the overall life cycle of data from its collection, to storage, to processing/transformation to analysis.

- **Software engineering:** This category of skills encompasses practices for building, managing and operating composable D&A applications. It covers coding and scripting languages (such as Python), as well as practices like data validation and code versioning.
- **Soft skills:** This category includes skills such as collaboration and business acumen to help data engineers engage more productively with other teams in their organization and achieve larger organizational goals.

Figure 2: Skills Categories for Data Engineering



Core Skills

Every data engineer should be familiar with the following fundamental core skills: SQL and database fundamentals, data processing, data storage, data orchestration, programming and collaboration.

SQL & Database Fundamentals

Structured Query Language (SQL) is arguably the most important core skill needed to query the database. It is a language designed to manage data in a relational database management system (RDBMS). However, SQL remains a part of every system, as even nonrelational database systems are adding SQL support.

SQL Basics

Data engineers use SQL to curate and structure data; transform data for extraction, transformation and loading/extract, load, transform (ETL/ELT) pipelines; model business logic; build databases and deliver essential KPIs. It is pervasive across a wide range of tools and platforms, making it both a critical and an extensible skill. For example, dbt, a data transformation tool, enables data engineers to transform data in their warehouses by simply writing select SQL statements. Database management systems (DBMSs), such as PostgreSQL, MySQL and others, use SQL. SQL is also the backbone of big data and analytics frameworks like Apache Hadoop, analytics and distributed processing engines like Apache Spark and Presto, and business intelligence (BI) tools like Microsoft Power BI and Salesforce (Tableau).

Fundamental Database Concepts

After gaining a basic understanding of SQL, data engineers must learn some key concepts to manage data present in databases, these are:

- **Different data types:** Defines what value the column can hold — integer, character, money, date and time, binary and so on.

Actionable, objective insight

Position your IT organization for success. Explore these additional, complementary resources and tools for data leaders and their teams:

Article

Four Key Pillars of a Data-Centric Approach to AI

Adopt a data-centric mindset to position your organization for success in the era of unstructured AI data.

Read More

Technical Execution

Data and Analytics Architects

Access action plans for the key trends for data and analytics architects.

Learn More

Webinar

How Tech Architects Can Get the Most Out of GenAI Coding Assistants

Mitigate risks and apply AI coding tools effectively to enhance your development processes.

Watch Now

Tool

Gartner Data and Analytics Maturity Score for CDAOs

Assess and improve your function's readiness.

Learn More

Already a client?
Get access to even more resources in your client portal. Log In

Connect With Us

Get actionable, objective insight that drives smarter decisions and stronger performance on your mission-critical priorities. Contact us to become a client:

U.S.: 1 855 811 7593

International: +44 (0) 3330 607 044

Become a Client

Learn more about Gartner for Technical Professionals

gartner.com/en/chief-information-officer

Stay connected to the latest insight

Attend a Gartner conference

View Conference

Gartner®