What L&D Leaders Need to Know about Data Fluency

Adel Nehme, 2020-10-15
About me

- Undergraduate degree in Economics from the American University of Beirut
- MSc in Business Analytics & Data Science from ESSEC Business School & CentraleSupelec
- Data Science Educator and Evangelist @ DataCamp
Our Mission

Our mission is to democratize data science education by building the best platform to learn and teach data skills and make data fluency accessible to millions of people and businesses around the world.
Trusted by over 1,600 data-driven companies
Please share your questions in the chat
Agenda

- The last 2 decades in disruption and the state of digital transformation
- Data transformation underpins digital transformation
- How the data fluency skill gap is a key component of data transformation
- The costs of not addressing the data fluency skill gap
- Examples of data fluency upskilling
- The key components of a scalable data fluency program
- The tools and skills in a data-driven organization
- 8 key data personas to scale in every organization
- What should L&D leaders do next?
- Q&A
The two decades review in disruption

**Finance**
- Money Transfer
  - TransferWise
  - Cash App
- Personal Banking
  - N26
  - Revolut
- Payment Processing
  - stripe
  - Square

**Travel**
- Hospitality
  - Booking.com
  - Airbnb
- Flights
  - TripActions
- Ride Hailing
  - Uber
  - Lyft

**Retail**
- E-Commerce
  - Amazon
  - Shopify
- Delivery
  - Blue Apron

**Transport**
- Bike Sharing
  - Lime
  - HelloBike.
The state of digital transformation

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Description</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>70%</td>
<td>Of digital transformation projects fail at achieving their stated goals</td>
<td><strong>McKinsey</strong></td>
</tr>
<tr>
<td>60-73%</td>
<td>Of organization-data is untouched for analysis</td>
<td><strong>Forrester</strong></td>
</tr>
<tr>
<td>50%</td>
<td>Of organizations do not mention “data analytics” in their corporate strategies</td>
<td><strong>Gartner</strong></td>
</tr>
</tbody>
</table>
The state of digital transformation

- **70%**
  - Of digital transformation projects fail at achieving their stated goals
  - Source: McKinsey

- **60-73%**
  - Of organization-data is untouched for analysis
  - Source: Forrester

- **50%**
  - Of organizations do not mention “data analytics” in their corporate strategies
  - Source: Gartner
“Leaders need to look at data first to succeed in their digital initiatives, rather than treating them as an afterthought to help with ad hoc projects.”—Mike Rollings, Research VP at Gartner
Being data-driven is at the heart of digital transformation

Digital Products, Services & Processes

Successful digital transformation
Being data-driven is at the heart of digital transformation

- What drives our customer-acquisition?
- What are our best marketing channels?
- What segments are we serving best?

- How can we improve our customer experience?
- What features do our customers want the most?
- What features are our customers using the least?

- What should we prioritize in our roadmap?
- How do we become operationally better?
Data-driven is not synonymous with hiring data scientists

**Data reactive**
- No one accesses or uses data in their daily work. Your company rarely reports on or presents data

**Data scaling**
- Very few people have the skills and access they need to analyze, report, and present data confidently

**Data progressive**
- Every team has at least one data fluent employee who can analyze, report, and present their data, regardless of role

**Data fluent**
- Everyone knows how to access the data they need to do their job (this doesn't mean that everyone needs to code)
Data transformation underpins digital transformation

The key differentiators between the disruptors and the incumbents is not technology-based but in their data-driven culture, the insights they draw from data while examining and iterating upon their services, and the data fluency skills they foster.
The steps to data transformation
Data transformation depends on many levers

Tools
- Which data tools to use?

Organization
- Bespoke organizational models

Processes
- Data processes for scale

People
- Upskilling for data fluency

Infrastructure
- Enabling data access

Webinar: Scaling Data Science at your Organization
Data transformation depends on many levers

- **Tools**: Which data tools to use?
- **Organization**: Bespoke organizational models
- **Processes**: Data processes for scale

**People**: Upskilling for data fluency

**Infrastructure**: Enabling data access

[Webinar: Scaling Data Science at your Organization]
The data fluency skill gap

43% of businesses believe the most pressing skill gaps to fill are in “Data analytics”

Source: McKinsey

* survey of 1,000 businesses across industries

34% of CEOs believe skill gaps in “data analytics” is the most crucial threat to their organization

Source: PwC’s 2019 annual CEO survey
The costs of data fluency skill gap

50%  Of employees avoid or find alternative solutions to data-related tasks

74%  Of employees feel overwhelmed by working with data

Source: Accenture & Qlik * survey of 9,000 employees across a range of industries
The costs of data fluency skill gap

50%
Of employees avoid or find alternative solutions to data-related tasks

74%
Of employees feel overwhelmed by working with data

Source: Accenture & Qlik
* survey of 9,000 employees across a range of industries
Addressing the data fluency skill gap

89%

Of Learning and Development leaders believe building data fluency is *moderate to high* priority for their organizations.

2x

Companies with mature data fluency competencies have reported almost twice as much performance improvement as their non-data fluent rivals and twice as much urgency.

Source: *What 300+ L&D Leaders Have Learned About Building Data Fluency*
Organizations are realizing Upskilling is the only way.

- Marks & Spencer launched a data academy for 1,000 employees.
- Amazon launched a machine learning university for developers.
- Airbnb launched an organization-wide data university.
- AT&T launched a $1B—10 year long Upskilling initiative for 140,000 people.
The impact of becoming data fluent

70% of organizations that invested in upskilling are reporting positive ROI

88% of organizations that invested in upskilling on analytics have exceeded business goals

Source: McKinsey

Source: Deloitte
Successful digital transformation requires scaling data fluency upskilling

While different individuals within an organization need to learn different data fluency skills — it enables an organization to speak one common data language and is one of the pillars of a data-driven culture
The challenge in becoming data fluent
Scaling data fluency programs
The components of data fluency

Competency Areas

Tools

Personas
How data flows through an organization

**Raw Data Collection**
- Collect Data from Different Sources

**Data Warehousing**
- Centralized Data Storage

**Data Processing**
- Create Organization Relevant Data

**Data Access**
- Make Organization Relevant Data Access Easy for the Organization

**Data Insights**
- End Consumer Interaction with Organization Relevant Data

- Dashboards
- Machine Learning
- Insights and Analysis
How data flows through an organization

**Raw Data Collection**
- Collect Data from Different Sources

**Data Warehousing**
- Centralized Data Storage

**Data Processing**
- Create Organization Relevant Data

**Data Access**
- Make Organization Relevant Data Access Easy for the Organization

**Data Insights**
- End Consumer Interaction with Organization Relevant Data

**Data**
- Dashboards
- Machine Learning
- Insights and Analysis

**Competencies**
- Data Literacy
- Business Analysis
- Data Analysis
- Data Science
- Statistics
- Machine Learning
- Programming
- Data Engineering
How data flows through an organization

Data Collection:
- Collect Data from Different Sources

Data Warehousing:
- Centralized Data Storage

Data Processing:
- Create Organization Relevant Data

Data Access:
- Make Organization Relevant Data Access Easy for the Organization

Data Insights:
- End Consumer Interaction with Organization Relevant Data

Data:
- Dashboards
- Machine Learning
- Insights and Analysis

Competencies:
- Data Literacy
- Business Analysis
- Data Analysis
- Data Science
- Statistics
- Machine Learning
- Programming
- Data Engineering

Tools:
- Programming Languages
- Databases
- Spreadsheets
- BI Tools
- Command Line Tools
Breaking down data tools

Python is an open-source programming language for data analysis, machine learning, big data and data engineering.

R is an open-source programming language most commonly used for statistical & data analysis, dashboards and visualization.

Scala is an open-source programming language specifically designed for the maintenance and processing of big data.
Breaking down data tools

**Programming Languages**
- **Python** is an open-source programming language for data analysis, machine learning, big data and data engineering.
- **R** is an open-source programming language most commonly used for statistical & data analysis, dashboards and visualization.
- **Scala** is an open-source programming language specifically designed for the maintenance and processing of big data.

**Databases**
- **SQL** is a structured query language that allows to query data from a database. It can come in different dialects and versions.
Breaking down data tools

**Programming Languages**

- **Python**: An open-source programming language for data analysis, machine learning, big data and data engineering.
- **R**: An open-source programming language most commonly used for statistical & data analysis, dashboards and visualization.
- **Scala**: An open-source programming language specifically designed for the maintenance and processing of big data.

**Databases**

- **SQL**: A structured query language that allows to query data from a database. It can come in different dialects and versions.

**Spreadsheets**

- **Spreadsheets**: Allow for easy, intuitive, and drag-and-drop interfaces for manipulating, aggregating, and visualizing data but fall short when processing large amounts of data and often produce bottlenecks when creating reproducible shareable analysis. Most notable amongst them are Google Sheets and Microsoft Excel.
## Breaking down data tools

### Programming Languages

<table>
<thead>
<tr>
<th>Language</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Python</td>
<td>An open-source programming language for data analysis, machine learning, big data and data engineering.</td>
</tr>
<tr>
<td>R</td>
<td>An open-source programming language most commonly used for statistical &amp; data analysis, dashboards and visualization.</td>
</tr>
<tr>
<td>Scala</td>
<td>An open-source programming language specifically designed for the maintenance and processing of big data.</td>
</tr>
</tbody>
</table>

### Databases

<table>
<thead>
<tr>
<th>Language</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL</td>
<td>A structured query language that allows to query data from a database. It can come in different dialects and versions.</td>
</tr>
</tbody>
</table>

### Spreadsheets

<table>
<thead>
<tr>
<th>Language</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spreadsheets allow for easy, intuitive, and drag-and-drop interfaces for manipulating, aggregating, and visualizing data but fall short when processing large amounts of data and often produce bottlenecks when creating reproducible shareable analysis. Most notable amongst them are Google Sheets and Microsoft Excel.</td>
</tr>
</tbody>
</table>

### BI Tools

<table>
<thead>
<tr>
<th>Language</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BI Tools are supercharged dashboarding tools that allow data aggregation, manipulation and visualization at scale for big data. The most notable amongst them are Power BI and Tableau.</td>
</tr>
</tbody>
</table>
Programming Languages

- **Python** is an open-source programming language for data analysis, machine learning, big data and data engineering.
- **R** is an open-source programming language most commonly used for statistical & data analysis, dashboards and visualization.
- **Scala** is an open-source programming language specifically designed for the maintenance and processing of big data.

Databases

- **SQL** is a structured query language that allows to query data from a database. It can come in different dialects and versions.

Spreadsheets

- Spreadsheets allow for easy, intuitive, and drag-and-drop interfaces for manipulating, aggregating, and visualizing data but fall short when processing large amounts of data and often produce bottlenecks when creating reproducible shareable analysis. Most notable amongst them are Google Sheets and Microsoft Excel.

BI Tools

- BI Tools are supercharged dashboarding tools that allow data aggregation, manipulation and visualization at scale for big data. The most notable amongst them are Power BI and Tableau.

Command line tools

- Command line tools are used to systematize file handling, enable version control, work with cloud tools, execute data pipelines developed using other data tools—especially programming languages—easily and scalably. The two most notable are Shell and Git.
Data fluency competency areas

- Data Literacy
- Business Analysis
- Data Analysis
- Data Science
- Statistics
- Machine Learning
- Programming
- Data Engineering
Data fluency competency areas

Data literacy skills allow data-driven decision making and for business stakeholders to communicate with data scientists.

**Basic**
- Data Literacy
- Define what are different data roles and applications

**Intermediate**
- Derive simple descriptive statistics and visualizations

**Advanced**
- Data-driven decision making

Competency areas:
- Data Literacy
- Business Analysis
- Data Analysis
- Data Science
- Statistics
- Machine Learning
- Programming
- Data Engineering
Data fluency competency areas

Business analysis is analyzing data, and prescribing actionable decisions to improve business value

Data Literacy:
- Define what are different data roles and applications

Basic:
- Derive simple descriptive statistics and visualizations

Intermediate:
- Derive simple descriptive statistics and visualizations
- Apply business domain knowledge and report insights

Advanced:
- Data-driven decision making
- Report insights with visualizations and dashboards

Data Analysis:
- Business Analysis
- Data Science
- Statistics
- Machine Learning
- Programming
- Data Engineering
Data fluency competency areas

Data analysis is about analyzing and reporting on business relevant data

- **Data Literacy**
  - Define what are different data roles and applications

- **Business Analysis**
  - Derive simple descriptive statistics and visualizations

- **Data Analysis**
  - Derive simple descriptive statistics and visualizations

- **Data Science**
  - Derive simple descriptive statistics and visualizations

- **Statistics**
  - Report insights with visualizations and dashboards

- **Machine Learning**
  - Data-driven decision making

- **Programming**
  - Report insights with visualizations and dashboards

- **Data Engineering**
  - Mastery of the data analysis workflow

Basic

Intermediate

Advanced
Data fluency competency areas

Data science skills are about digging through and generating data insights from data that provide business value

Data Literacy
- Define what are different data roles and applications

Business Analysis
- Derive simple descriptive statistics and visualizations

Data Analysis
- Derive simple descriptive statistics and visualizations

Data Science
- Mastery of the data analysis workflow

Statistics

Machine Learning

Programming

Data Engineering

Basic
- Derive simple descriptive statistics and visualizations
- Apply business domain knowledge and report insights
- Mastery of the data analysis workflow
- Democratize insights with reports and dashboards

Intermediate

Advanced
- Data-driven decision making
- Report insights with visualizations and dashboards
- Report insights with visualizations and dashboards
- Applying Machine Learning to business problems
Data fluency competency areas

Statistical skills enable testing and proving hypotheses about the data, and making sure analysis is correct.
Data literacy competency areas

Machine learning is about creating predictions from your data

**Basic**
- **Data Literacy**: Define what are different data roles and applications
- **Business Analysis**: Derive simple descriptive statistics and visualizations
- **Data Analysis**: Derive simple descriptive statistics and visualizations
- **Data Science**: Mastery of the data analysis workflow
- **Statistics**: Mastery of the data analysis workflow
- **Machine Learning**: Mastery of Business Applications of Machine Learning

**Intermediate**
- **Data Literacy**: Define what are different data roles and applications
- **Business Analysis**: Derive simple descriptive statistics and visualizations
- **Data Analysis**: Derive simple descriptive statistics and visualizations
- **Data Science**: Mastery of the data analysis workflow
- **Statistics**: Mastery of the data analysis workflow
- **Machine Learning**: Mastery of Business Applications of Machine Learning
- **Programming**: Apply business domain knowledge and report insights
- **Data Engineering**: Democratize insights with reports and dashboards
- **Democratize**: Statistical modeling and experiment design
- **Perform**: Perform the machine learning workflow

**Advanced**
- **Data-driven decision making**: Report insights with visualizations and dashboards
- **Report insights with visualizations and dashboards**: Applying Machine Learning to business problems
- **Report insights with visualizations and dashboards**: Advanced statistical modeling for inference
- **Apply Machine Learning on all different types of data**: Apply Machine Learning on all different types of data
**Data fluency competency areas**

**Programming skills enable automated processes, data workflows and improved performance**

<table>
<thead>
<tr>
<th>Basic</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Literacy</td>
<td>Define what are different data roles and applications</td>
<td>Data-driven decision making</td>
</tr>
<tr>
<td>Business Analysis</td>
<td>Derive simple descriptive statistics and visualizations</td>
<td>Report insights with visualizations and dashboards</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>Derive simple descriptive statistics and visualizations</td>
<td>Report insights with visualizations and dashboards</td>
</tr>
<tr>
<td>Data Science</td>
<td>Mastery of the data analysis workflow</td>
<td>Applying Machine Learning to business problems</td>
</tr>
<tr>
<td>Statistics</td>
<td>Mastery of the data analysis workflow</td>
<td>Advanced statistical modeling for inference</td>
</tr>
<tr>
<td>Machine Learning</td>
<td>Mastery of Business Applications of Machine Learning</td>
<td>Apply Machine Learning on all different types of data</td>
</tr>
<tr>
<td>Programming</td>
<td>Automate processes and benchmark performance</td>
<td>World-class coding best practice</td>
</tr>
<tr>
<td>Data Engineering</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data fluency competency areas

Data engineering is about getting data in the right hands by setting up and maintaining a data infrastructure

Basic
- **Data Literacy**: Define what are different data roles and applications
- **Business Analysis**: Derive simple descriptive statistics and visualizations
- **Data Analysis**: Derive simple descriptive statistics and visualizations
- **Data Science**: Mastery of the data analysis workflow
- **Statistics**: Mastery of the data analysis workflow
- **Machine Learning**: Mastery of Business Applications of Machine Learning
- **Programming**: Automate processes and benchmark performance
- **Data Engineering**: Extract, Transform, and load data

Intermediate
- **Data Literacy**: Derive simple descriptive statistics and visualizations
- **Business Analysis**: Apply business domain knowledge and report insights
- **Data Analysis**: Mastery of the data analysis workflow
- **Data Science**: Democratize insights with reports and dashboards
- **Statistics**: Statistical modeling and experiment design
- **Machine Learning**: Perform the machine learning workflow
- **Programming**: Test code and develop internal tools for data work
- **Data Engineering**: Automate data flows and work with cloud technology

Advanced
- **Data-driven decision making**: Report insights with visualizations and dashboards
- **Business Analysis**: Report insights with visualizations and dashboards
- **Data Analysis**: Applying Machine Learning to business problems
- **Data Science**: Advanced statistical modeling for inference
- **Statistics**: Apply Machine Learning on all different types of data
- **Machine Learning**: World-class coding best practice
- **Programming**: Manage large databases and process big datasets
Data personas for every organization
# Data personas: Data Leaders & Consumers

**Data Leaders & Consumers:** Non-technical role — but needs to consume data insights

## DATA SKILLS

<table>
<thead>
<tr>
<th>Beginner</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>■ Understand what data scientists, machine learning scientists, and data engineers do.</td>
<td>■ Draw common visualizations and extract simple descriptive statistics from data.</td>
<td>■ Have a strong grasp of the fundamentals of business intelligence and BI tools.</td>
</tr>
<tr>
<td>■ Know which questions can (and can't) be answered with data.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>■ Interpret the results of data projects, including calculations and visualizations.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

## COMMONLY USED TOOLS

- Spreadsheets
- BI Tools

## EXAMPLE JOB TITLES

- Chief Marketing Officer
- Human Resources Manager
- Head of Sales
## Data personas: Business Analysts

**Business Analysts:** Responsible for tying data insights to actionable results that increase profitability or efficiency.

### DATA SKILLS

<table>
<thead>
<tr>
<th>Beginner</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>■ Draw common visualizations and extract simple descriptive statistics from data.</td>
<td>■ A deep knowledge of the business domain and the ability to report and communicate insights using data.</td>
<td>■ Democratize access to data insights by creating dashboards and organizing data to answer organizational questions.</td>
</tr>
<tr>
<td>■ Understand the business applications of data.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### COMMONLY USED TOOLS

<table>
<thead>
<tr>
<th>Spreadsheets</th>
<th>BI Tools</th>
<th>Databases</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Excel" /></td>
<td><img src="image" alt="Tableau" /></td>
<td><img src="image" alt="SQL" /></td>
</tr>
</tbody>
</table>

### EXAMPLE JOB TITLES

- Business Analyst
- Marketing Analyst
- Supply Chain Analyst
**Data personas: Data Analysts**

**Data Analysts:** Responsible for generating data insights and reporting them.

### DATA SKILLS

**Beginner**
- Draw common visualizations and extract simple descriptive statistics from data.
- Understand the business applications of data.

**Intermediate**
- A deep understanding of the data analysis workflow, which includes importing, manipulating, cleaning, calculating, and reporting on organization data.
- A strong grasp of business intelligence and BI tools.

**Advanced**
- Democratize access to data insights by creating dashboards and organizing data to answer organizational questions.

### COMMONLY USED TOOLS

- **Spreadsheets**
- **BI Tools**
- **Databases**
- **Programming Languages**

### EXAMPLE JOB TITLES

- Data Analyst
- Operations Analyst
- Marketing Analyst
- Supply Chain Analyst
## Data personas: Data Scientists

### Data Scientists: Investigate, analyze and communicate data insights and drive data agenda

#### DATA SKILLS

<table>
<thead>
<tr>
<th>Beginner</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
</table>
| - Deeply understand the data analysis workflow, which include importing, manipulating, cleaning, calculating, and reporting on organization data. | - Understand fundamental statistics, including distributions, modeling, and inference.  
- Understand supervised and unsupervised ML workflows.  
- Can create dashboards using coding tools such as Python and R. | - Can apply analyses and machine learning workflows to business applications such as finance, marketing, and healthcare.  
- Work with non-standard data types, such as time series, text, geospatial, and images. |

#### COMMONLY USED TOOLS

<table>
<thead>
<tr>
<th>Databases</th>
<th>Programming Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL</td>
<td>Python, R, SQL, Excel</td>
</tr>
</tbody>
</table>

#### EXAMPLE JOB TITLES

- Data Scientist
- Data Analyst
# Data personas: Machine Learning Scientists

## Machine Learning Scientists: Developing and deploying machine learning models

### DATA SKILLS

<table>
<thead>
<tr>
<th>Level</th>
<th>Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Beginner</strong></td>
<td>Deeply understand the data analysis workflow, which include importing,</td>
</tr>
<tr>
<td></td>
<td>manipulating, cleaning, calculating, and reporting on organization data.</td>
</tr>
<tr>
<td><strong>Intermediate</strong></td>
<td>Perform supervised and unsupervised machine learning workflows including</td>
</tr>
<tr>
<td></td>
<td>feature engineering, training models, testing accuracy and making predictions.</td>
</tr>
<tr>
<td></td>
<td>Can apply analyses and machine learning workflows to business applications such as finance, marketing, and healthcare.</td>
</tr>
<tr>
<td><strong>Advanced</strong></td>
<td>Perform deep learning workflows.</td>
</tr>
<tr>
<td></td>
<td>Work with APIs and coding best practices.</td>
</tr>
</tbody>
</table>

### COMMONLY USED TOOLS

- **Programming Languages**: [Python](#), [R](#), [Java](#)
- **Databases**: [SQL](#)
- **Command Line Tools**: [Git](#), [Docker](#)

### EXAMPLE JOB TITLES

- Data Scientist
- Research Scientist
- Machine Learning Engineer
## Data personas: Statisticians

### STATISTICIANS: Scaling experiments, and providing analytical rigour to analysis

### DATA SKILLS

<table>
<thead>
<tr>
<th>Beginner</th>
<th>Intermediate</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deeply understand the data analysis workflow, which include importing, manipulating, cleaning, calculating, and reporting on organization data.</td>
<td>Perform statistical modeling workflows, including feature engineering, training models, testing goodness of fit, and inferring significance. Test hypotheses and design simple experiments such as A/B tests.</td>
<td>Design more complex experiments &amp; understand Bayesian statistics. Understand specialist models, such as survival models, generalized additive models, mixture models, and structural equation models.</td>
</tr>
</tbody>
</table>

### COMMONLY USED TOOLS

<table>
<thead>
<tr>
<th>Databases</th>
<th>Programming Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQL</td>
<td>R, Python</td>
</tr>
</tbody>
</table>

### EXAMPLE JOB TITLES

- Data Scientist
- Inference Data Scientist
- Quantitative Analyst
# Data personas: Programmers

**Programmers:** Deploys solutions to automate repetitive tasks and scale data work

## DATA SKILLS

### Beginner
- Write functions to avoid repetitive code.
- Benchmark and optimize code to improve performance.

### Intermediate
- Deeply understand coding best practices.
- Work with web APIs and develop packages for sharing code.

### Advanced
- Develop data pipelines and work with parallel programming.
- Understand programming paradigms, such as functional programming and object-oriented programming.

## COMMONLY USED TOOLS

**Programming Languages**
- Python
- R
- SQL

**Command Line Tools**
- Git
- Bash

## EXAMPLE JOB TITLES

- Software Engineer
- Data Scientist
- Dev-Ops Engineer
# Data personas: Data Engineers

**Data Engineers:** Gets data into the right hands and maintains data infrastructure

## DATA SKILLS

<table>
<thead>
<tr>
<th>Skill Level</th>
<th>Skills</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginner</td>
<td>- Efficiently extract, transform, and load data from raw data sources into organization databases.</td>
</tr>
</tbody>
</table>
| Intermediate| - Process data and automate data flows using the command line.  
- Process data using cloud platforms. |
| Advanced    | - Manage and optimize databases and process big datasets. |

## COMMONLY USED TOOLS

<table>
<thead>
<tr>
<th>Category</th>
<th>Tools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Programming Languages</td>
<td>Python, R, SQL</td>
</tr>
<tr>
<td>Databases</td>
<td>SQL, NoSQL</td>
</tr>
<tr>
<td>Command Line Tools</td>
<td>Bash, PowerShell</td>
</tr>
</tbody>
</table>

## EXAMPLE JOB TITLES

- Data Engineer
- Software Engineer
- Dev-ops Engineer
## Data personas skill matrix

<table>
<thead>
<tr>
<th>Data Consumers &amp; Leaders</th>
<th>Business Analysts</th>
<th>Data Analysts</th>
<th>Data Scientists</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Spreadsheets</strong>&lt;br&gt;![Spreadsheet Icon]</td>
<td><strong>Spreadsheets</strong>&lt;br&gt;![Spreadsheet Icon]</td>
<td><strong>Spreadsheets</strong>&lt;br&gt;![Spreadsheet Icon]</td>
<td><strong>Programming Languages</strong>&lt;br&gt;![Language Icon]</td>
</tr>
<tr>
<td><strong>BI Tools</strong>&lt;br&gt;![BI Tool Icon]</td>
<td><strong>BI Tools</strong>&lt;br&gt;![BI Tool Icon]</td>
<td><strong>BI Tools</strong>&lt;br&gt;![BI Tool Icon]</td>
<td><strong>Databases</strong>&lt;br&gt;![Database Icon]</td>
</tr>
<tr>
<td><strong>Databases</strong>&lt;br&gt;![Database Icon]</td>
<td><strong>Databases</strong>&lt;br&gt;![Database Icon]</td>
<td><strong>Databases</strong>&lt;br&gt;![Database Icon]</td>
<td><strong>Databases</strong>&lt;br&gt;![Database Icon]</td>
</tr>
<tr>
<td><strong>Data Literacy</strong></td>
<td><strong>Data Literacy</strong></td>
<td><strong>Data Literacy</strong></td>
<td><strong>Data Literacy</strong></td>
</tr>
<tr>
<td><strong>Business Analysis</strong></td>
<td><strong>Business Analysis</strong></td>
<td><strong>Business Analysis</strong></td>
<td><strong>Business Analysis</strong></td>
</tr>
<tr>
<td><strong>Data Analysis</strong></td>
<td><strong>Data Analysis</strong></td>
<td><strong>Data Analysis</strong></td>
<td><strong>Data Analysis</strong></td>
</tr>
<tr>
<td><strong>Data Science</strong></td>
<td><strong>Data Science</strong></td>
<td><strong>Data Science</strong></td>
<td><strong>Data Science</strong></td>
</tr>
<tr>
<td><strong>Statistics</strong></td>
<td><strong>Statistics</strong></td>
<td><strong>Statistics</strong></td>
<td><strong>Statistics</strong></td>
</tr>
<tr>
<td><strong>Machine Learning</strong></td>
<td><strong>Machine Learning</strong></td>
<td><strong>Machine Learning</strong></td>
<td><strong>Machine Learning</strong></td>
</tr>
<tr>
<td><strong>Programming</strong></td>
<td><strong>Programming</strong></td>
<td><strong>Programming</strong></td>
<td><strong>Programming</strong></td>
</tr>
<tr>
<td><strong>Data Engineering</strong></td>
<td><strong>Data Engineering</strong></td>
<td><strong>Data Engineering</strong></td>
<td><strong>Data Engineering</strong></td>
</tr>
</tbody>
</table>
# Data personas skill matrix

<table>
<thead>
<tr>
<th>Machine Learning Scientists</th>
<th>Statisticians</th>
<th>Programmers</th>
<th>Data Engineers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Programming Languages</strong></td>
<td>Programming Languages</td>
<td>Programming Languages</td>
<td>Programming Languages</td>
</tr>
<tr>
<td><strong>Databases</strong></td>
<td>Databases</td>
<td>Databases</td>
<td>Databases</td>
</tr>
<tr>
<td><strong>Command Line Tools</strong></td>
<td>Command Line Tools</td>
<td>Command Line Tools</td>
<td>Command Line Tools</td>
</tr>
<tr>
<td>Data Literacy</td>
<td>Data Literacy</td>
<td>Data Literacy</td>
<td>Data Literacy</td>
</tr>
<tr>
<td>Business Analysis</td>
<td>Business Analysis</td>
<td>Business Analysis</td>
<td>Business Analysis</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>Data Analysis</td>
<td>Data Analysis</td>
<td>Data Analysis</td>
</tr>
<tr>
<td>Data Science</td>
<td>Data Science</td>
<td>Data Science</td>
<td>Data Science</td>
</tr>
<tr>
<td>Statistics</td>
<td>Statistics</td>
<td>Statistics</td>
<td>Statistics</td>
</tr>
<tr>
<td>Machine Learning</td>
<td>Machine Learning</td>
<td>Machine Learning</td>
<td>Machine Learning</td>
</tr>
<tr>
<td>Programming</td>
<td>Programming</td>
<td>Programming</td>
<td>Programming</td>
</tr>
<tr>
<td>Data Engineering</td>
<td>Data Engineering</td>
<td>Data Engineering</td>
<td>Data Engineering</td>
</tr>
</tbody>
</table>

Get the white paper

The L&D Guide to Data Fluency

Download here!
What should L&D leaders do next?
Make L&D a strategic pillar of the organization

Aligning L&D with the business goals

Create a learning culture
Find your data personas

Personalize your learning paths

Measure ROI

Empathy & Patience

Scaling data fluency L&D programs
How DataCamp builds and sustains data fluency

**ASSESS**
Accurately identify strengths and skills gaps across your organization

**LEARN**
Upskill your team with interactive courses, broken into bite-sized chapters

**APPLY**
Use real-world tools and workflows to complete end-to-end analysis

**PRACTICE**
Master new skills with daily exercises on desktop and DataCamp Mobile
No matter your role or experience levels make up your team, skill assessments will provide you with an understanding of your organization’s data skill level.
Gain new skills by doing

From data concepts to coding, every training is interactive. Learners complete hands-on exercises in browser, receive real-time feedback, and are rewarded with XP every time they make progress.
Learn on the go

Grow your data skills on the go with DataCamp Mobile. Make progress in mobile courses and keep your skills sharp with daily 5-minute coding challenges.
Go deeper with projects

Complete real-world analyses and case studies in a risk-free environment. Projects give learners the confidence to transfer their new skills to their work.
Advanced enterprise reporting

Easily understand, measure, and report on the impact training has on your organization and see how you are moving towards your learning goals.
Track your team’s growth and progress

At a glance, you can review your organization’s data skill level, visualize the most popular technology and content areas, and identify your most active learners.
Simple team management

Centrally manage all aspects of your account from your Enterprise dashboard. Invite learners, view adoption recommendations, and create role and skill-based teams.
Add DataCamp to your LMS/LXP

Integrate DataCamp to centralize access to courses and your learner analytics. Out-of-the-box integrations are available for Degreed, Cornerstone OnDemand, SAP SuccessFactors, and EdCast.
Customer Success: Fast track your team’s success

We’ll learn about your business goals

We’ll help you map your learning journey to meet your targets

We’ll partner with you to create long-term value
Thank you

Adel Nehme
Data Science Evangelist @ DataCamp
adel@datacamp
LinkedIn