Comparison of Roles in Data Science

> **Introduction**
>
> This document presents a detailed comparison of the most common roles in data science, covering their characteristics, scope, data volume handled, technical and soft skills, common tools, application sectors, and other relevant aspects.

1. Data Scientist

- **Characteristics / Functions:** Predictive modeling, advanced analysis, extracting insights from heterogeneous data.

- **Scope / Impact:** Supports strategic and tactical decision-making through deep analysis.

- **Data Volume:** Large volumes, both structured and unstructured.

- **Hard Skills:**

    - Advanced statistics and machine learning.
    - Programming in Python, R, SQL.
    - Data wrangling and cleaning.

- **Soft Skills:** Critical thinking, clear communication, scientific curiosity, ability to synthesize.

- **Common Tools:** Python (Pandas, scikit-learn, TensorFlow), R, Jupyter Notebooks, SQL.

- **Common Sectors / Areas:** Tech, healthcare, retail, telecommunications, research.

2. Data Engineer

- **Characteristics / Functions:** Designs and maintains pipelines and architectures for data processing and storage.

- **Scope / Impact:** Essential for ensuring data availability and quality across the organization.

- **Data Volume:** Very large volumes, batch and real-time processing.

- **Hard Skills:**

    - Programming in Python, Java, Scala.
    - Knowledge of relational and NoSQL databases.
    - Experience with distributed systems (Hadoop, Spark).
    - Cloud computing (AWS, GCP, Azure).

- **Soft Skills:** Problem solving, teamwork, attention to technical detail.

- **Common Tools:** Apache Hadoop, Spark, Kafka, Airflow, cloud services.

- **Common Sectors / Areas:** Large enterprises, fintech, e-commerce, cloud services.

- **Note on Scala:** Scala is a language commonly used by data engineers for working with distributed systems like Apache Spark due to its high efficiency and compatibility with the Big Data ecosystem.

3. **Data Analyst**

- **Characteristics / Functions:** Performs exploratory analysis, reporting, and visualization to support operational decision-making.

- **Scope / Impact:** Tactical support in the organization through clear and useful information.

- **Data Volume:** Medium-sized, generally structured data.

- **Hard Skills:**

    - Advanced SQL, Excel.
    - Visualization using Tableau, Power BI.
    - Basic descriptive statistics.

- **Soft Skills:** Effective communication, attention to detail, curiosity for pattern detection.

- **Common Tools:** Excel, SQL, Tableau, Power BI, basic Python.

- **Common Sectors / Areas:** Business, marketing, retail, healthcare, logistics.

4. **Machine Learning Engineer (ML Engineer)**

- **Characteristics / Functions:** Implements, deploys, and maintains machine learning models in production, optimizing performance and scalability.

- **Scope / Impact:** Crucial for transforming prototypes into scalable automated solutions.

- **Data Volume:** Variable, from moderate to large.

- **Hard Skills:**

    - Advanced programming (Python, C++).
    - Knowledge of software engineering and DevOps.
    - ML frameworks (TensorFlow, PyTorch).

- **Soft Skills:** Interdisciplinary collaboration, adaptability, problem-solving focus.

- **Common Tools:** TensorFlow, PyTorch, Docker, Kubernetes, REST APIs.

- **Common Sectors / Areas:** Tech, AI, startups, banking, healthcare.

5. **Business Intelligence Specialist (BI Specialist)**

- **Characteristics / Functions:** Designs dashboards and reports for KPI monitoring and visualization, facilitating managerial decisions.

- **Scope / Impact:** Direct support to executive and managerial levels with structured, up-to-date information.

- **Data Volume:** Aggregated and structured data.

- **Hard Skills:**

  - SQL and data modeling.
  - BI tools (Power BI, Tableau, QlikView).

- **Soft Skills:** Business-oriented mindset, clear communication, effective presentation.

- **Common Tools:** Power BI, Tableau, QlikView.

- **Common Sectors / Areas:** Companies of all sectors, finance, sales, operations.

6. **Applied Data Scientist**

- **Characteristics / Functions:** Applies data science techniques adapted to specific industry or domain problems.

- **Scope / Impact:** Solving concrete problems and improving sector-specific processes.

- **Data Volume:** Varies depending on the project.

- **Hard Skills:** Data science, programming, deep domain knowledge.

- **Soft Skills:** Interdisciplinary communication, analytical thinking, results orientation.

- **Common Tools:** Depends on the sector (Python, R, domain-specific tools).

- **Common Sectors / Areas:** Healthcare, finance, marketing, manufacturing.

7. **Data Architect**

- **Characteristics / Functions:** Defines the organization's data architecture, ensuring integration, governance, and standards.

- **Scope / Impact:** Strategic long-term impact on data management within the organization.

- **Data Volume:** Very large and complex, multi-source.

- **Hard Skills:**

    - Data modeling, SQL and NoSQL databases.
    - Data security and governance.
    - Cloud computing and distributed architectures.

- **Soft Skills:** Leadership, strategic vision, project management.

- **Common Tools:** ETL tools, databases, AWS/GCP/Azure.

- **Common Sectors / Areas:** Large corporations, banking, telecommunications, consulting.

8. **Financial Data Analyst**

- **Characteristics / Functions:** Analysis of financial metrics, forecasting models, risk management, and fraud detection.

- **Scope / Impact:** Supports financial management and decision-making in corporate finance.

- **Data Volume:** Medium to large, structured financial and accounting data.

- **Hard Skills:**

    - Advanced Excel, SQL.
    - Accounting and financial knowledge.
    - Sector-specific analytical tools.

- **Soft Skills:** Analytical thinking, professional ethics, clear communication.

- **Common Tools:** Excel, SQL, Python, financial software.

- **Common Sectors / Areas:** Finance, banking, insurance, financial consulting.

9. Risk Data Analyst

- **Characteristics / Functions:** Analyzes data to identify, evaluate, and mitigate financial, operational, or market risks.

- **Scope / Impact:** Contributes to proactive risk management, protecting the organization's financial health.

- **Data Volume:** Medium to large, including historical and real-time data.

- **Hard Skills:**

    - Statistical and predictive modeling.

    - Programming in Python, R, or SQL.

    - Knowledge of risk regulations and standards.

- **Soft Skills:** Critical analysis, attention to detail, clear communication with legal and financial areas.

- **Common Tools:** Python, R, SQL, financial analysis tools.

- **Common Sectors / Areas:** Banking, insurance, consulting, auditing.

Conclusions

Scala is widely used mainly by Data Engineers for developing Big Data systems due to its high efficiency and compatibility with frameworks like Apache Spark. Each data science role has a specific set of responsibilities and skills, and they all collaborate to enable organizations to make informed decisions and build data-driven solutions.