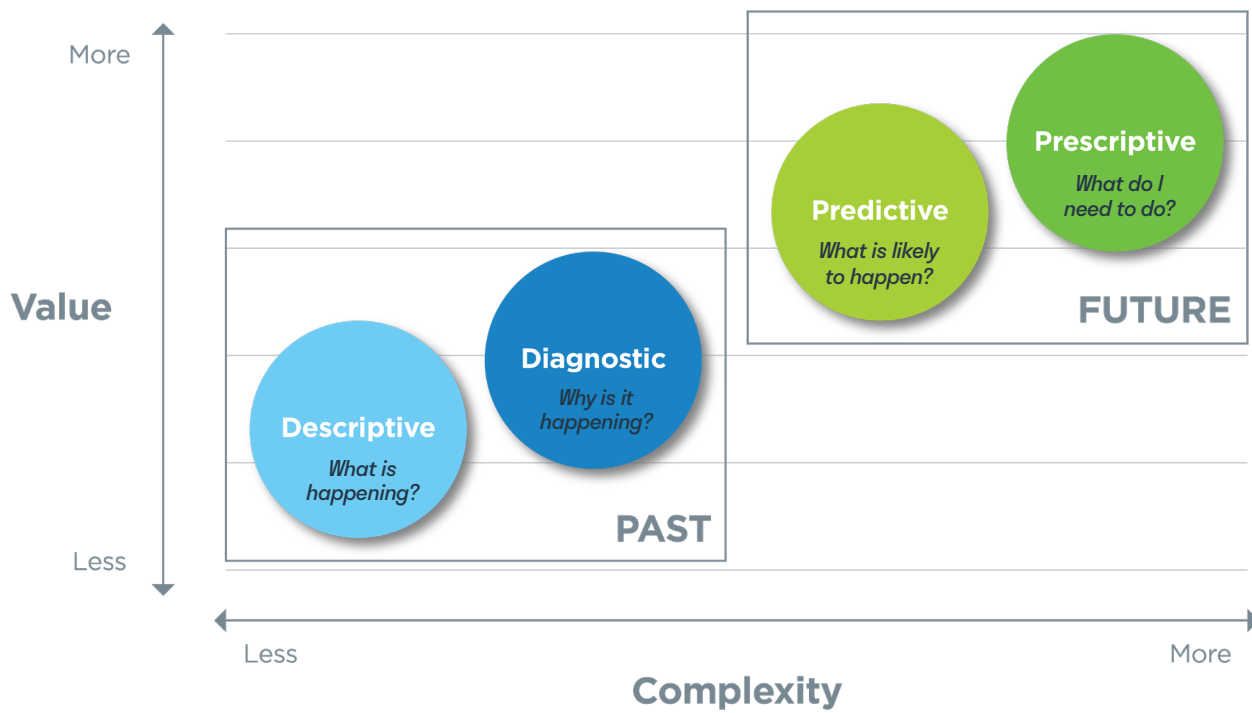


The Ultimate Guide to Data Analytics

What Is Data Analytics?

Data analytics involves using data, techniques and tools that identify patterns and trends, which in turn generate actionable insights that support informed decision-making.

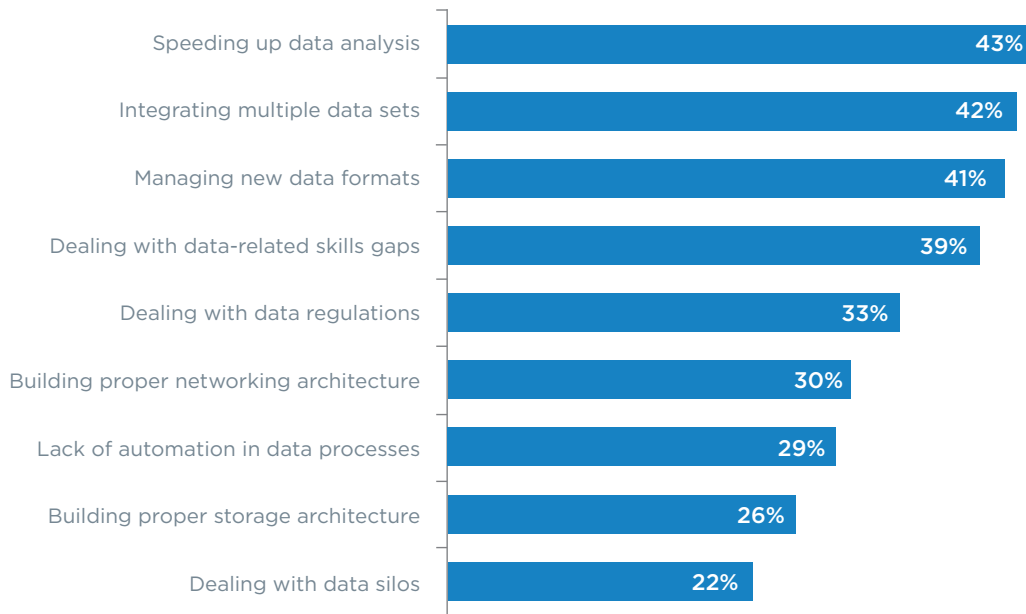
4 Types of Data Analytics



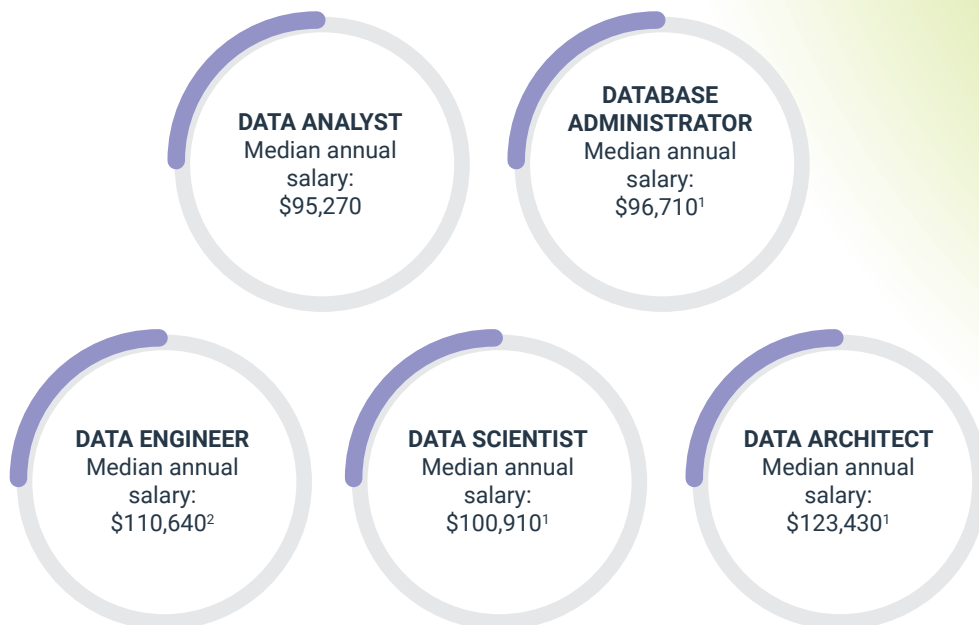
Main Stages of the Data Lifecycle



Common Challenges in Data Management

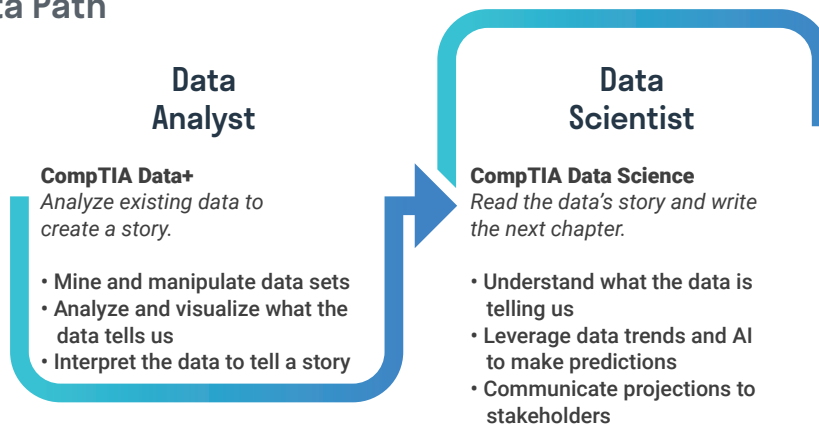


Job Roles in the Field of Data

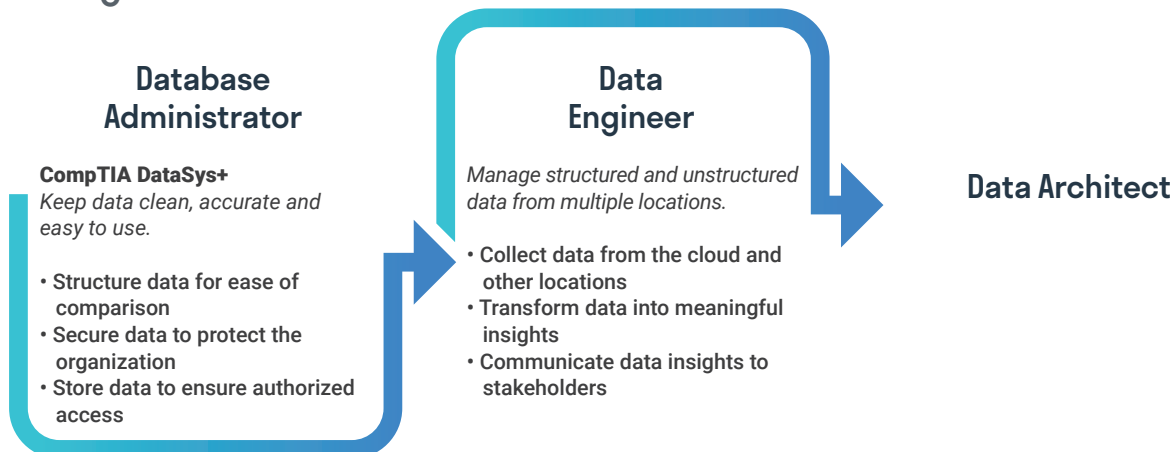


The CompTIA Data Pathway

Interpret Data Path



Manage Data Path



Why Choose a Data Analytics Career?

- Recognize patterns and trends
- Understand how data is compiled
- Help organizations reach their goals
- Future-proof your career
- Help stakeholders make better decisions
- Improve productivity
- Provide imperative information for organizations
- Identify and leverage sources for a competitive advantage
- Mitigate risk

266%
projected tech
growth rate above
the national rate for
data scientists and
data analysts.

How To Get a Job in Data Analytics

Step 1: Determine your career goals

Step 2: Learn the fundamentals

Step 3: Get qualified

Step 4: Practice your skills

Step 5: Network and form relationships with industry experts

Step 6: Start applying for introductory data analytics jobs

Step 7: Land your dream job

Tips for Succeeding in a Data Analytics Career

- Define your area of interest
- Practice data storytelling
- Upskill and reskill
- Create a portfolio
- Stay up to date with the industry

Data Analytics Skills

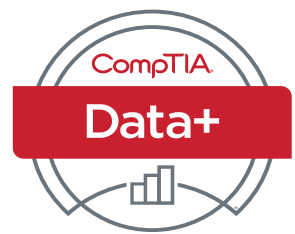
TECHNICAL/HARD SKILLS	PROFESSIONAL/SOFT SKILLS
Understanding of data processes and principles, and ability to apply basic statistical methods	Communication (written and verbal)
Data analytics and visualization software	Storytelling and presenting
Machine learning algorithms and models	Decision-making
Ability to mine, manipulate, analyze, interpret and visualize data	Problem-solving
Python, R and other programming languages	Analytical mindset
Data optimization	Creative thinking
Predictive modeling	Attention to detail
Database management	Time management
Adherence to governance and quality standards throughout the entire data life cycle	Organizational

Tools and Technologies

TOOL	DESCRIPTION
Structured Query Language (SQL)	A special-purpose programming language used for managing and manipulating relational databases.
NoSQL Database	A database management system that does not rely on a traditional relational database structure, often used for handling unstructured or semi-structured data.
Python	A general-purpose programming language with extensive libraries and frameworks that enable data analysis, machine learning and visualization.
Microsoft Excel	A simple, but powerful spreadsheet program commonly used for data entry, basic data analysis and visualization.
R Programming	A programming language commonly used for statistical computing and graphics.
Tableau	A data visualization software that allows for visualization, analysis and understanding of data.
SAS	A software suite for advanced analytics, multivariate analysis, business intelligence, data management and predictive analytics.
Power BI	A business analytics service provided by Microsoft that offers interactive visualizations and business intelligence capabilities.
AWS/Amazon Web Services	A cloud-based platform offering a variety of services and tools for storage, computing, analytics and machine learning.
Hadoop Ecosystem	An open-source software framework used for collecting and processing large amounts of data.
In-Memory Database	A database management system that stores data in memory to achieve faster processing speeds.
Apache Spark	An open-source data processing engine that provides faster and more efficient processing of large datasets.
Data Lake/Data Warehouse	A centralized repository that allows for the storage of unstructured and structured data at any scale.
Google Analytics	A web analytics service offered by Google that tracks website traffic and provides insights into user behavior.

Glossary

KEY TERM	DESCRIPTION
Ad-hoc report	Reports that are generated on request.
Aggregation	Process of collecting data to present it in summary form.
Algorithm	A set of heuristics and calculations that creates a model from data.
Analysis	The process of analyzing the past or future and making an informed decision.
Artificial intelligence (AI)	Computer systems that are able to perform tasks that normally require human intelligence.
Business intelligence (BI)	Process of using strategies and technologies to analyze business information and transforming it into actionable recommendations for key stakeholders.
Cloud	A virtual space that allows users to access to storage, files, software and servers through internet-connected devices.
Cluster analysis	Grouping objects in a way that objects in the same cluster are more similar to each other than they are to objects in a different cluster.
Data	Facts and statistics collected together for reference or analysis.
Data analytics	The science of analyzing raw data to make conclusions about that information.
Data mining	Process used to extract usable data from a large dataset. Data mining involves data collection, warehousing and computer processing. In order to segment and evaluate the data, data mining uses advanced algorithms.
Data modeling	The process of diagramming data flows.
Data visualization	Data visualization presents a clear picture of what the data actually means. Using bar graphs, pie charts, tables and other visuals, data visualization makes the data easier for those making business decisions to comprehend.
Mean	Average of all of the numbers in a dataset.
Median	The middle number in a dataset (when in order).
Mode	The most common number in a dataset.
Querying	Requesting data from a database.
Recurring reports	Reports that are compiled and presented on a regular, ongoing basis. <i>Examples:</i> <ul style="list-style-type: none"> • Compliance reports (financial, health and safety) • Risk and regulatory reports • Operational reports (performance, key performance indicators)
Regression analysis	A statistical method used to measure the relationship between variables.
Standard deviation	A measure of how dispersed the data is in relation to the mean.
Structured data	Standardized, clearly defined and searchable data.
Text analytics	The process of drawing meaning out of written communication. Usually, text analytics software relies on text mining and natural language processing (NLP) algorithms to find patterns and meaning.
Unstructured data	Datasets (typical large collections of files) that aren't stored in a structured database format.
Values	Information that has been translated into letters, numbers, and/or symbols so that it can be read, moved and manipulated by a machine.



¹ Source: U.S. Bureau of Labor Statistics (BLS)
² Source: Lightcast

CompTIA.

© 2023 CompTIA, Inc., used under license by CompTIA, Inc. All rights reserved. All certification programs and education related to such programs are operated exclusively by CompTIA, Inc. CompTIA is a registered trademark of CompTIA, Inc. in the U.S. and internationally. Other brands and company names mentioned herein may be trademarks or service marks of CompTIA, Inc. or of their respective owners. Reproduction or dissemination prohibited without the written consent of CompTIA, Inc. Printed in the U.S. 10407- May 2023