



CompTIA Data+ Certification Exam Objectives

EXAM NUMBER: DAO-002 V2



About the Exam

The CompTIA Data+ certification exam will certify the successful candidate has the knowledge and skills required to:

- Translate business requirements in support of data-driven decisions by acquiring, preparing, and transforming data.
- Use industry-standard tools and emerging technologies to create appropriate reports and visualizations.
- Apply basic statistical methods and analyze complex data sets while adhering to governance and quality standards throughout the entire data life cycle.

This is equivalent to 1.5 to 2 years of hands-on experience that includes exposure to databases and analytical tools, a basic understanding of statistics, and familiarity with data visualization.

These content examples are meant to clarify the test objectives and should not be construed as a comprehensive listing of all the content of this examination.

EXAM ACCREDITATION

The CompTIA Data+ exam is accredited by the ANSI National Accreditation Board (ANAB) to show compliance with the International Organization for Standardization (ISO) 17024 standard and, as such, undergoes regular reviews and updates to the exam objectives.

EXAM DEVELOPMENT

CompTIA exams result from subject matter expert workshops and industry-wide survey results regarding the skills and knowledge required of an IT professional.

COMPTIA AUTHORIZED MATERIALS USE POLICY

CompTIA Certifications, LLC is not affiliated with and does not authorize, endorse, or condone utilizing any content provided by unauthorized third-party training sites (aka “brain dumps”). Individuals who utilize such materials in preparation for any CompTIA examination will have their certifications revoked and be suspended from future testing in accordance with the CompTIA Candidate Agreement. In an effort to more clearly communicate CompTIA's exam policies on use of unauthorized study materials, CompTIA directs all certification candidates to the [CompTIA Certification Exam Policies](#). Please review all CompTIA policies before beginning the study process for any CompTIA exam. Candidates will be required to abide by the [CompTIA Candidate Agreement](#). If a candidate has a question as to whether study materials are considered unauthorized (aka “brain dumps”), he/she should contact CompTIA at examsecurity@CompTIA.org to confirm.

PLEASE NOTE

The lists of examples provided in bulleted format are not exhaustive lists. Other examples of technologies, processes, or tasks pertaining to each objective may also be included on the exam, although not listed or covered in this objectives document. CompTIA is constantly reviewing the content of our exams and updating test questions to be sure our exams are current, and the security of the questions is protected. When necessary, we will publish updated exams based on existing exam objectives. Please know that all related exam preparation materials will still be valid.

TEST DETAILS

Required exam	DA0-002
Number of questions	Maximum of 90
Types of questions	Multiple-choice and performance-based
Length of test	90 minutes
Recommended experience	1.5 to 2 years of experience, including exposure to databases and analytical tools, a basic understanding of statistics, and familiarity with data visualization

EXAM OBJECTIVES (DOMAINS)

The table below lists the domains measured by this examination and the extent to which they are represented.

DOMAIN		PERCENTAGE OF EXAMINATION
1.0	Data Concepts and Environments	20%
2.0	Data Acquisition and Preparation	22%
3.0	Data Analysis	24%
4.0	Visualization and Reporting	20%
5.0	Data Governance	14%
Total		100%



1.0 Data Concepts and Environments

1.1 Explain data concepts.

- Database types
 - Relational
 - Non-relational
- File extensions
 - .csv
 - .xlsx
 - .json
 - .txt
 - .jpg
 - .dat
- Data structures
 - Structured
 - Table
 - Fact table
 - Dimensional table
 - Slowly changing dimension
 - Bridge table
 - Schema
 - Semi-structured
 - JavaScript Object Notation (JSON)
 - Nested structures
 - Unstructured
- Data types
 - String
 - char
 - varchar
 - nvarchar
 - Null values
- Spatial
- Boolean
- Numeric
 - Integer
 - Decimal
 - Float
- Datetime
 - Timestamp
- Large objects
 - Binary large object (BLOB)
 - Character large object (CLOB)
- Globally Unique Identifier (GUID)/Universally Unique Identifier (UUID)

1.2 Identify types of data sources.

- Databases
- Application programming interfaces (APIs)
- Website data
- Files
- Logs
- Data repositories
 - Data lakes
 - Data lakehouses
 - Data marts
 - Data silos
 - Data warehouses

1.3 Identify infrastructure concepts.

- Cloud providers
 - Amazon Web Services (AWS)
 - Azure
 - Google
- Cloud and on-premises infrastructure
 - Private
 - Public
 - Hybrid
- Storage types
 - Object
 - File
 - Local
 - Shared
 - Block
- Containerization



1.4 Identify common data analysis tools.

- Coding environments
 - Integrated development environment (IDE)
 - RStudio
 - Visual Studio (VS) Code
 - Text editor
 - Notebooks
- Business intelligence software
 - Tableau
 - Power BI
 - Looker
- Packages/libraries
 - Anaconda
 - pandas
 - tidyverse
- Programming languages
 - SAS
 - Python
 - R
 - Scala
- Database management system (DBMS)
 - Structured Query Language (SQL) Server Management Studio
 - MySQL Workbench
 - MongoDB Compass
 - DBeaver
 - Toad
 - Azure Data Studio

1.5 Identify artificial intelligence (AI) concepts.

- Generative AI
 - Large language model (LLM)
- Foundational models
- Deep learning
- Natural language processing (NLP)
- Robotic process automation (RPA)
 - Automated reporting



2.0 Data Acquisition and Preparation

2.1 Given a scenario, use data acquisition methods.

- Data integration
- Querying
 - Join
 - Concatenate
 - Filter
 - Union
 - Grouping
 - Aggregate
 - Nested queries
- Basic query optimization
 - Indexing
 - Parameterization
 - Subsets
 - Temporary tables
- Extract, transform, load (ETL)/
extract, load, transform (ELT)
- Data collection
 - Surveying
 - Sampling

2.2 Given a scenario, perform data exploration to identify possible inconsistencies with a data set.

- Missing values
- Duplication
- Redundancy
- Outliers
- Completeness
- Validation

2.3 Given a scenario, perform appropriate data transformation and cleansing techniques.

- String manipulation
 - Regular expressions (RegEx)
- Conversion
- Clustering
 - Binning
- Augmentation
- Exploding
- Scaling
- Standardization
- Imputation
- Parsing
- Merging
- Appending
- Derived variables
 - Calculated fields
- Deletion



3.0 Data Analysis

3.1 Given a set of requirements, determine the appropriate communication approach for data analysis.

- Mock-up
- Accessibility
 - Auditory
 - Visual
- Technical vs. non-technical audience
- Level of detail
- Internal vs. external
- User persona type
 - C-suite vs. individual contributor
- Sensitive vs. non-sensitive
- Key performance indicators (KPIs)

3.2 Given a scenario, select the appropriate statistical method or function.

- Basic statistical methods
 - Prescriptive
 - Descriptive
 - Predictive
 - Inferential
- Functions and measures
 - Mathematical
 - Measures of central tendency
 - Measures of dispersion
 - Logical
 - Date
 - String

3.3 Given a scenario, troubleshoot basic issues using the appropriate tool or method.

- Issues
 - Connectivity-related
 - User-reported
 - Basic SQL code
 - Corrupted data
- Tools and methods
 - Enable logging
 - Validate data source
 - Consult vendor communities/
online resources



4.0 Visualization and Reporting

4.1 Given a scenario, use the appropriate visual elements.

- Types
 - Charts
 - Maps
 - Pivot tables
 - Infographics
- Design elements
 - Labels
 - Legends
 - Branding
 - Color schemes

4.2 Given a scenario, use the appropriate delivery or consumption method.

- Executive summary
- Self-service portal
- Dashboards
 - Static
 - Dynamic
 - Frequency
 - Recurring
 - Ad hoc
- Data versioning techniques
 - Snapshot
 - Real-time

4.3 Given a scenario, troubleshoot issues using report validation techniques.

- Issues
 - Excessive load time
 - Slow refresh rate
 - Large data size
 - Filter not working correctly
 - Stale data
 - Corrupt data
- Techniques
 - Data filtering
 - Review
 - Code
 - Calc
 - Peer
 - Source validation
 - Data structure changes
 - Monitoring alerts



5.0 Data Governance

5.1 Explain data management concepts.

- Integration
- Documentation
 - Data flow diagram
 - Data explainability report
 - Data dictionary
 - Hierarchy structure
 - Data lineage
- Source of truth
- Data versioning
 - Snapshots
 - Refresh intervals
- Metadata

5.2 Summarize concepts related to data compliance.

- Retention
- General Data Protection Regulation (GDPR)
- Jurisdictional requirements
- Replication
- Storage
- Data ethics
- Payment Card Industry Data Security Standard (PCI DSS)
- Audit
- Classification
- Incident reporting
 - Data breach
 - Security

5.3 Compare and contrast data privacy and protection practices.

- Role-based access control
- Encryption
 - In transit
 - At rest
- Data usage
- Data sharing
- National Institute Standards and Technology (NIST)
- Personal identifiable information (PII)
- Personal health information (PHI)
- Anonymization
- Masking

5.4 Compare and contrast data quality assurance practices.

- Requirement testing
- Stress test
- User acceptance test (UAT)
- Source control
- Unit test
- Data health check
 - Data drifts
- Automated data quality monitoring
- Data profiling
 - Quality metrics
- International Organization for Standardization (ISO)

CompTIA Data+ Acronym List

The following acronyms appear on the CompTIA Data+ exam. Candidates are encouraged to review the complete list and attain a working knowledge of all listed acronyms as part of a comprehensive exam preparation program.

ACRONYM	DEFINITION
AI	Artificial Intelligence
API	Application Programming Interface
AWS	Amazon Web Services
BLOB	Binary Large Object
CLOB	Character Large Object
DBMS	Database Management System
ELT	Extract, Load, Transform
ETL	Extract, Transform, Load
FAQ	Frequently Asked Questions
GDPR	General Data Protection Regulation
GUID	Globally Unique Identifier
HTML	Hypertext Markup Language
IDE	Integrated Development Environment
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
LLM	Large Language Model
MDM	Master Data Management
NAS	Network-attached Storage
NIST	National Institute Standards and Technology
NLP	Natural Language Processing
OLAP	Online Analytical Processing
OLTP	Online Transaction Processing
P&L	Profit and Loss
PCI	Payment Card Industry
PHI	Personal Health Information
PII	Personally Identifiable Information
RAM	Random-access Memory
RDBMS	Relational Database Management System
RegEx	Regular Expressions
RPA	Robotic Process Automation
SAN	Storage Area Network
SDLC	Software Development Life Cycle
SQL	Structured Query Language
SSMS	SQL Server Management Studio
SSRS	SQL Server Reporting Services
UAT	User Acceptance Testing
URL	Uniform Resource Locator
UUID	Universally Unique Identifier
VPN	Virtual Private Network
VS	Visual Studio
XML	Extensible Markup Language

CompTIA Data+ Hardware and Software List

CompTIA has included this sample list of hardware and software to assist candidates as they prepare for the Data+ certification exam. This list may also be helpful for training companies that wish to create a lab component for their training offering. The bulleted lists below each topic are sample lists and are not exhaustive.

EQUIPMENT

- Desktop or laptop with at least 16GB of random-access memory (RAM)
- Local storage (e.g., storage area network [SAN] or network-attached storage [NAS])

SOFTWARE

- Text editor (e.g., VS Code and Notepad++)
- Anaconda
- Jupyter Notebook
- Visualization tools (e.g., Tableau and Power BI)
- Microsoft Office Suite
- Reporting tools (e.g., Quarto, Crystal Reports, and SQL Server Reporting Services [SSRS])
- SQL query tools (e.g., Azure Data Studio and SQL Server Management Studio [SSMS])
- Mock-up software (e.g., Lucid, PowerPoint, and Visio)

SERVICES

- Internet access
- Cloud environment (e.g., AWS and Snowflake)
- Cloud storage

OTHER

- Publicly available data sets (e.g., Kaggle and data.world, etc.)