CompTIA Data+ Certification Exam Objectives

EXAM NUMBER: DA0-001

CompTIA

About the Exam

Candidates are encouraged to use this document to help prepare for the CompTIA Data+ (DA0-001) certification exam. This exam will certify the successful candidate has the knowledge and skills required to transform business requirements in support of data-driven decisions by:

- **Mining data**
- **Manipulating data**
- **Applying basic statistical methods**
- **Analyzing complex datasets while adhering to governance and quality standards throughout the entire data life cycle**

This is equivalent to 18–24 months of hands-on experience working in a business intelligence report/data analyst job role. 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 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"), they 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 testing exam objectives. Please know that all related exam preparation materials will still be valid.

TEST DETAILS

Required exam	DA0-001
Number of questions	Maximum of 90
Types of questions	Multiple choice and performance-based
Length of test	90 minutes
Recommended experience	• 18–24 months of experience in a report/business analyst job role • Exposure to databases and analytical tools • Basic understanding of statistics • Data visualization experience
Passing score	675 (on scale of 100–900)

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	15%
2.0 Data Mining	25%
3.0 Data Analysis	23%
4.0 Visualization	23%
5.0 Data Governance, Quality, and Controls	14%
Total	**100%**

1.0 Data Concepts and Environments

1.1 Identify basic concepts of data schemas and dimensions.

- **Databases**
  - Relational
  - Non-relational
- **Data mart/data warehousing/data lake**
  - Online transactional processing (OLTP)
  - Online analytical processing (OLAP)
- **Schema concepts**
  - Snowflake
  - Star
- **Slowly changing dimensions**
  - Keep current information
  - Keep historical and current information

1.2 Compare and contrast different data types.

- **Date**
- **Numeric**
- **Alphanumeric**
- **Currency**
- **Text**
- **Discrete vs. continuous**
- **Categorical/dimension**
- **Images**
- **Audio**
- **Video**

1.3 Compare and contrast common data structures and file formats.

- **Structures**
  - Structured
    - Defined rows/columns
    - Key value pairs
  - Unstructured
    - Undefined fields
    - Machine data
- **Data file formats**
  - Text/Flat file
    - Tab delimited
    - Comma delimited
  - JavaScript Object Notation (JSON)
  - Extensible Markup Language (XML)
  - Hypertext Markup Language (HTML)

2.0 Data Mining

2.1 Explain data acquisition concepts.

- **Integration**
  - Extract, transform, load (ETL)
  - Extract, load, transform (ELT)
  - Delta load
  - Application programming interfaces (APIs)
- **Data collection methods**
  - Web scraping
  - Public databases
  - Application programming interface (API)/web services
  - Survey
  - Sampling
  - Observation

2.2 Identify common reasons for cleansing and profiling datasets.

- **Duplicate data**
- **Redundant data**
- **Missing values**
- **Invalid data**
- **Non-parametric data**
- **Data outliers**
- **Specification mismatch**
- **Data type validation**

2.3 Given a scenario, execute data manipulation techniques.

- **Recoding data**
  - Numeric
  - Categorical
- **Derived variables**
- **Data merge**
- **Data blending**
- **Concatenation**
- **Data append**
- **Imputation**
- **Reduction/aggregation**
- **Transpose**
- **Normalize data**
- **Parsing/string manipulation**

2.4 Explain common techniques for data manipulation and query optimization.

- **Data manipulation**
  - Filtering
  - Sorting
  - Date functions
  - Logical functions
  - Aggregate functions
  - System functions
- **Query optimization**
  - Parameterization
  - Indexing
  - Temporary table in the query set
  - Subset of records
  - Execution plan

3.0 Data Analysis

3.1 Given a scenario, apply the appropriate descriptive statistical methods.

- **Measures of central tendency**
  - Mean
  - Median
  - Mode
- **Measures of dispersion**
  - Range
    - Max
    - Min
  - Distribution
  - Variance
  - Standard deviation
- **Frequencies/percentages**
- **Percent change**
- **Percent difference**
- **Confidence intervals**

3.2 Explain the purpose of inferential statistical methods.

- ***t*-tests**
- ***Z*-score**
- ***p*-values**
- **Chi-squared**
- **Hypothesis testing**
  - Type I error
  - Type II error
- **Simple linear regression**
- **Correlation**

3.3 Summarize types of analysis and key analysis techniques.

- **Process to determine type of analysis**
  - Review/refine business questions
  - Determine data needs and sources to perform analysis
  - Scoping/gap analysis
- **Type of analysis**
  - Trend analysis
    - Comparison of data over time
  - Performance analysis
    - Tracking measurements against defined goals
    - Basic projections to achieve goals
  - Exploratory data analysis
    - Use of descriptive statistics to determine observations
  - Link analysis
    - Connection of data points or pathway

3.4 Identify common data analytics tools.

(The intent of this objective is NOT to test specific vendor feature sets nor the purposes of the tools.)

- **Structured Query Language (SQL)**
- **Python**
- **Microsoft Excel**
- **R**
- **Rapid mining**
- **IBM Cognos**
- **IBM SPSS Modeler**
- **IBM SPSS**
- **SAS**
- **Tableau**
- **Power BI**
- **Qlik**
- **MicroStrategy**
- **BusinessObjects**
- **Apex**
- **Dataroma**
- **Domo**
- **AWS QuickSight**
- **Stata**
- **Minitab**

4.0 Visualization

4.1 Given a scenario, translate business requirements to form a report.

- **Data content**
- **Filtering**
- **Views**
- **Date range**
- **Frequency**
- **Audience for report**
  - Distribution list

4.2 Given a scenario, use appropriate design components for reports and dashboards.

- **Report cover page**
  - Instructions
  - Summary
    - Observations and insights
- **Design elements**
  - Color schemes
  - Layout
  - Font size and style
  - Key chart elements
    - Titles
    - Labels
    - Legends
  - Corporate reporting standards/style guide
    - Branding
    - Color codes
    - Logos/trademarks
    - Watermark
- **Documentation elements**
  - Version number
  - Reference data sources
  - Reference dates
    - Report run date
    - Data refresh date
  - Frequently asked questions (FAQs)
  - Appendix

4.3 Given a scenario, use appropriate methods for dashboard development.

- **Dashboard considerations**
  - Data sources and attributes
    - Field definitions
    - Dimensions
    - Measures
  - Continuous/live data feed vs. static data
  - Consumer types
    - C-level executives
    - Management
    - External vendors/stakeholders
    - General public
    - Technical experts
- **Development process**
  - Mockup/wireframe
    - Layout/presentation
    - Flow/navigation
    - Data story planning
  - Approval granted
  - Develop dashboard
  - Deploy to production
- **Delivery considerations**
  - Subscription
  - Scheduled delivery
  - Interactive (drill down/roll up)
    - Saved searches
    - Filtering
  - Static
  - Web interface
  - Dashboard optimization
  - Access permissions

4.4 Given a scenario, apply the appropriate type of visualization.

- **Line chart**
- **Pie chart**
- **Bubble chart**
- **Scatter plot**
- **Bar chart**
- **Histogram**
- **Waterfall**
- **Heat map**
- **Geographic map**
- **Tree map**
- **Stacked chart**
- **Infographic**
- **Word cloud**

4.5 Compare and contrast types of reports.

- **Static vs. dynamic reports**
  - Point-in-time
  - Real time
- **Ad-hoc/one-time report**
- **Self-service/on demand**
- **Recurring reports**
  - Compliance reports (e.g., financial, health, and safety)
  - Risk and regulatory reports
  - Operational reports [e.g., performance, key performance indicators (KPIs)]
- **Tactical/research report**

5.0 Data Governance, Quality, and Controls

5.1 Summarize important data governance concepts.

- **Access requirements**
  - Role-based
  - User group-based
  - Data use agreements
  - Release approvals
- **Security requirements**
  - Data encryption
  - Data transmission
  - De-identify data/data masking
- **Storage environment requirements**
  - Shared drive vs. cloud based vs. local storage
- **Use requirements**
  - Acceptable use policy
  - Data processing
  - Data deletion
  - Data retention
- **Entity relationship requirements**
  - Record link restrictions
  - Data constraints
  - Cardinality
- **Data classification**
  - Personally identifiable information (PII)
  - Personal health information (PHI)
  - Payment card industry (PCI)
- **Jurisdiction requirements**
  - Impact of industry and governmental regulations
- **Data breach reporting**
  - Escalate to appropriate authority

5.2 Given a scenario, apply data quality control concepts.

- **Circumstances to check for quality**
  - Data acquisition/data source
  - Data transformation/intrahops
    - Pass through
    - Conversion
  - Data manipulation
  - Final product (report/dashboard, etc.)
- **Automated validation**
  - Data field to data type validation
  - Number of data points
- **Data quality dimensions**
  - Data consistency
  - Data accuracy
  - Data completeness
  - Data integrity
  - Data attribute limitations
- **Data quality rule and metrics**
  - Conformity
  - Non-conformity
  - Rows passed
  - Rows failed
- **Methods to validate quality**
  - Cross-validation
  - Sample/spot check
  - Reasonable expectations
  - Data profiling
  - Data audits

5.3 Explain master data management (MDM) concepts.

- **Processes**
  - Consolidation of multiple data fields
  - Standardization of data field names
  - Data dictionary
- **Circumstances for MDM**
  - Mergers and acquisitions
  - Compliance with policies and regulations
  - Streamline data access

CompTIA Data+ (DA0-001) Acronym List

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

ACRONYM	DEFINITION
AIFF	Audio Interchange File Format
ANOVA	Analysis of Variance
API	Application Programming Interface
AWS	Amazon Web Services
BI	Business Intelligence
BLOB	Binary Large Object
CRM	Customer Relationship Management
CSS	Cascading Style Sheet
CSV	Comma-separated Values
DML	Data Markup Language
DSS	Data Security Standard
ELT	Extract, Load, Transform
ETL	Extract, Transform, Load
FAQs	Frequently Asked Questions
GDPR	General Data Protection Regulation
HTML	Hypertext Markup Language
JDBC	Java Database Connectivity
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
MDM	Master Data Management
MP4	MPEG Layer-4 Audio
NoSQL	Not Only Structured Query Language
ODBC	Open Database Connectivity
OLAP	Online Analytical Processing
OLS	Ordinary Least Squares
OLTP	Online Transaction Processing
P&L	Profit and Loss
PCI	Payment Card Industry
PDF	Portable Document Format
PHI	Personal Health Information
PHP	Hypertext Preprocessor
PII	Personally Identifiable Information
RDBMS	Relational Database Management System
SAS	Statistical Analysis System
SCD	Slowly Changing Dimension
SDLC	Software Development Life Cycle
SSIS	SQL Server Integration Services
SQL	Structured Query Language
TIFF	Tagged Image File Format
TSQL	Transactional SQL
WAV	Windows Wave
XLSM	Excel Macro-enabled Spreadsheet
XLSX	Excel Extensions to Open XML Spreadsheet
XML	Extensible Markup Language

CompTIA Data+ Proposed Hardware and Software List

CompTIA has included this sample list of hardware and software to assist candidates as they prepare for the Data+ 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 samples and are not exhaustive.

HARDWARE

- Desktop/laptop
  - High processing power for large volume analyses
  - Lower processing power for smaller volume analyses
- Internet access
- Cloud environment

SOFTWARE

- SQL environment to run scripts (SQL Lite, Management Studio, etc.)
- Eclipse
- Anaconda
- R Studio
- Database modeling tool
- Microsoft Office Suite
- Visualization tools (Tableau, Power BI, etc.)
- Reporting tools
- Sample datasets (Kaggle)

CompTIA