

CompTIA Data+
Certification Exam
Pre-draft Exam Objectives
Exam Number: DA0-002

- Pre-draft Exam Objectives summarize the tasks and skills identified in the Job Task Analysis (JTA) workshop that provide directional information about the upcoming exam version.
- The Draft Exam Objectives will replace the Pre-draft Exam Objectives after approximately two months when the skills have been peer-evaluated and validated through a JTA survey of job role practitioners.
- Pre-draft Exam Objectives may contain typos and errata that will be corrected during the development process.
- CompTIA will not accept feedback on the Pre-draft Exam Objectives document. If errors are found, please wait until the Draft Exam Objectives are posted, and then provide feedback using the Draft Exam Objectives Feedback form.

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
 - (i) Fact table
 - (ii) Dimensional table
 - 1. Slowly changing dimension
 - (iii) Bridge table
 - Schema
 - Semi-structured
 - JSON
 - (i) 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
- APIs
- Web scraping
- 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)
 - SQL Server Management Studio
 - SQL Workbench
 - Compass
 - DBeaver
 - Toad
 - Azure Data Studio

1.5 Identify artificial intelligence 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
- ETL/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
- 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 (PCI)
- Audit
- Classification
- Incident reporting
 - Data breach
 - Security

5.3 Compare and contrast data privacy and protection practices.

- Role-based access control (RBAC)
- Encryption
 - In transit
 - At rest
- Data usage
- Data sharing
- National Institute Standards and Technology (NIST)
- Personal Identifiable Information (PII)
- Protected 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)