SAS Data Quality Using DataFlux Data Management Studio Exam

Navigating the DataFlux Data Management Studio Interface

Navigate within the Data Management Studio Interface

- Register a new Quality Knowledge Base (QKB)
- Create and connect to a repository
- Define a data connection
- Specify Data Management Studio options
- Access the QKB
- Create a name value macro pair
- Access the business rules manager
- Access the appropriate monitoring report
- Attach and detach primary tabs

Exploring and Profiling data

Create and explore a data profile

- Create and explore a data profile
  - Different sources: text file, filtered table, SQL query
- Interpret the results
  - Frequency distribution
  - Pattern frequency distribution
  - Standard metrics
  - Visualizations
  - Alerts

Design data standardization schemes

- Build a scheme from profile results
- Build a scheme manually
- Update existing schemes
- Import and export a scheme

Data Jobs

Create Data Jobs

- Rename output fields
- Add nodes and preview nodes
- Run a data job
- View a log and settings
- Work with data job settings and data job displays
- Best practices (ensure you are following a particular best practice such as inserting notes, establishing naming conventions)
- Work with branching
- Join tables
- Apply the Field layout node to control field order
- Work with the Data Validation node:
    - Add it to the job flow
    - Specify properties/review properties
    - Edit settings for the Data Validation node
- Work with data inputs
- Work with data outputs
- Profile data from within data jobs
- Interact with the Repository from within Data Jobs
- Debug levels for logging
- Determine how data is processed
- Set sorting properties for the Data Sorting Node

Apply a Standardization definition and scheme

- Use a definition
- Use a scheme
- Determine the differences between definition and scheme
- Explain what happens when you use both a definition and scheme
- Review and interpret standardization results
- Explain the different steps involved in the process of standardization

Apply Parsing definitions

- Distinguish between different data types and their tokens
- Review and interpret parsing results
- Explain the different steps involved in the process of parsing
- Use parsing definition
- Interpret parse result codes

Apply Casing definitions

- Describe casing methods: upper/lower/proper
- Explain different techniques for accomplishing casing
- Use casing definition

Compare and contrast the differences between identification analysis and right fielding nodes

- Review results
- Explain the technique used for identification (process of definition)

Apply the Gender Analysis node to determine gender

- Use gender definition
- Interpret results
- Explain different techniques for conducting gender analysis

Create an Entity Resolution Job

- Use a clustering node in a data job and explain its use
- Survivorship (surviving record identification)
    - Record rules
    - Field rules
    - Options for survivorship
- Discuss and apply the Cluster Diff node
- Apply Cross-field matching
- Entity resolution file output node
- Use the Match Codes Node to select match definitions for selected fields.
    - Outline the various uses for match codes (join)
    - Use the definition
    - Interpret the results
    - Match versus match parsed
    - Explain the process for creating a match code
    - Select sensitivity for a selected match definition
    - Apply matching best practices

Use data job references within a data job

- Use of external data provider node
- Use of data job reference node
- Define a target node
- Explain why you would want to use a data job reference (best practice)
- Real-time data service

Understand how to use an Extraction definition

- Interpret the results
- Explain the process of the definition

Explain the process of the definition of pattern analysis

Business Rules Monitoring

Define and create business rules

- Use Business Rules Manager
- Create a new business rule
  - Name/label rule
  - Specify type of rule
  - Define checks
  - Specify fields
- Distinguish between different types of business rules
  - Row
  - Set
  - Group
- Apply business rules
  - Profile
  - Execute business rule node
- Use of Expression Builder
- Apply best practices

Create new tasks

- Understand events
  - Log error to repository
  - Set a data flow/key value
  - Log error to a text file
  - Write the row to a table
- Applying tasks
  - Explain purpose of the data monitoring node
- Review a data monitoring job log
- Review a monitoring report
  - Trigger values
  - Filters

Data Management Server

Interact with the Data Management Server

- Import/export jobs (special case profile)
- Test service
- Run history/job status
- Identify the required configuration components (QKB, data, reference sources, and repository)
- Security, the access control list
- Creation and use of WSDL

Expression Engine Language (EEL)

Explain the basic structure of EEL (components and syntax)

- Identify basic structural components of the code
  - Statements
  - Functions
  - Declarations
- Use EEL
  - Profile
  - Expression node (data job)
    - Tabs (expression, grouping, etc)
    - Order of Operations (pre/post, etc)
  - Expression node (process job)
  - Business rules
  - Custom metrics
    - Use in profile
    - Use in data job (execute custom metric node)
    - Use in business rule
  - Use in data validation node

Process Jobs

Work with and create process jobs

- Add nodes and explain what nodes do
- Interpret the log
- Parameterizing process jobs
- Identify Run options
- Using different functionality in process jobs
- If/then logic
    - Echo
    - Fork
    - Parallel iterator
    - Events and event handling (event listener)
    - Global get/set
    - Expression code features
        - Declaration of events
        - Set output slot
- Embedded data job and data job reference
- Using Work tables, process flow worktable reader
- SAS code execution
- SQL

Macro Variables and Advanced Properties and Settings

Work with and use macro variables in data profiles, data jobs and data monitoring

- Define macro variables:
    - In DM studio
    - In Configuration files
    - With Command line
    - Dynamic
- Use macro variables:
    - In a profile
    - In expression code
    - In a data job
    - In a process job
    - In business rules
- Determine Scoping/precedence (order in which macros are read)
- Compare/Contrast DM Studio versus DM Server

Determine uses for advanced properties

- Multi-locale
  - Use locale guessing
  - Use with a scheme
  - Locale list and locale field
- Apply setting for Max output rows

Quality Knowledge Base (QKB)

Describe the organization, structure and basic navigation of the QKB

- Identify and describe locale levels (global, language, country)
- Navigate the QKB (tab structure, copy definitions, etc)
- Identify data types and tokens

Be able to articulate when to use the various components of the QKB. Components include:

- Regular expressions
- Schemes
- Phonetics library
- Vocabularies
- Grammar
- Chop Tables

Define the processing steps and components used in the different definition types.

- Identify/describe the different definition types
  - Parsing
  - Standardization
  - Match
  - Identification
  - Casing
  - Extraction
  - Locale guess
  - Gender
  - Patterns
- Explain the interaction between different definition types (with one another, parse within match, etc)

Note: All nine sections and 24 main objectives will be tested on every exam. The expanded objectives are provided for additional explanation and define the entire domain that could be tested.