

**Actualtests.com**

The Power of Knowing



Exam : 000-416

Title : WebSphere IIS QualityStage v7.5

Ver : 12-18-2008

---

**QUESTION 1:**

You need to standardize address information using QualityStage Real Time. Which two ways is address information passed to the QSRT Server? (Choose two.)

- A. from a file in QualityStage Data directory
- B. from QualityStage Designer
- C. from QSRT client
- D. from QSRT manager

Answer: A,C

---

**QUESTION 2:**

Which two types of data are analyzed with Character Investigation? (Choose two.)

- A. name
- B. tax ID
- C. address
- D. date

Answer: B,D

---

**QUESTION 3:**

You are using the Date of Birth field as one of the BLOCKING fields. This field is defined with the option: Missing Value = X ( no missing values). In the match, what will happen to the records that contain spaces in the Date of Birth field?

- A. All records with spaces in the Date of Birth field will become residual records.
- B. The value of spaces will have no effect on the match.
- C. All records with spaces in the Date of Birth field will be blocked together.
- D. All records with spaces in the Date of Birth field will be put in one group regardless of other blocking fields.

Answer: C

---

**QUESTION 4:**

While reviewing match results with the customer it is discovered that some of the matched pairs with weights just over the Match Cutoff are false positives. The client would like the match job to be modified such that these records are not identified as matches. How should you resolve this issue?

- A. by adding weight overrides to comparisons

- B. by adding an additional pass to the match job
- C. by increasing the Clerical Review Cutoff
- D. by removing unreliable block variables from passes

Answer: A

---

### QUESTION 5:

QualityStage server is installed on a UNIX machine. The Parallel Extender engine has been newly installed and you wish to enable the QualityStage server to use the Parallel Extender. Which action must you now perform?

- A. Add ORCHHOME to the QualityStage server startup script.
- B. Reinstall all the QualityStage server software.
- C. Add INTBIN to the Parallel Extender PATH.
- D. Run the qs\_license.sh script.

Answer: D

---

### QUESTION 6:

Based on Investigation results it is determined that the phone number field in a file is unreliable. Which statement is true?

- A. The phone number field should not be used as a comparison variable in match.
- B. The phone number field should not be used as a blocking variable in match.
- C. The phone number field should be dropped from incoming records as soon as possible.
- D. The phone number field should not be used in match at all.

Answer: B

---

### QUESTION 7:

#### DRAG DROP

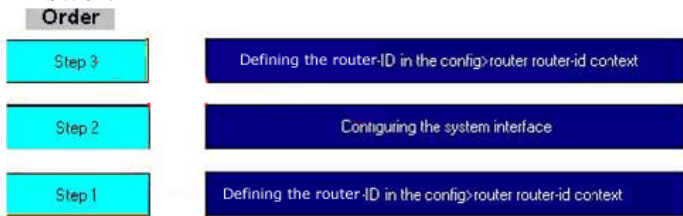
Place the following in order of selection for BGP Router-ID on the 7750 SR.

Order	
place here	Defining the router-ID in the config>router router-id context
place here	Configuring the system interface
place here	Defining the router-ID in the config>router>bgp router-id context

Steps		
Step 1	Step 2	Step 3

Answer:



---

### QUESTION 8:

A QualityStage server is to be installed on Linux, with the client on Windows XP. Which two steps should be completed prior to installing the QualityStage server? (Choose two.)

- A. Create a user and a group both named qsadm.
- B. Create a symbolic link pointing to your QualityStage installation directory.
- C. Install the QualityStage real-time service.
- D. Install the Parallel Extender engine.

Answer: A,B

---

### QUESTION 9:

A customer wishes to household all their customers. The records are all on one file. Which match option should you choose to accomplish this?

- A. Undup
- B. Match Sets
- C. GeoMatch
- D. Match

Answer: A

---

### QUESTION 10:

Which method should be used to create the directory structure under a master project?

- A. Deploy the first job in a project before running it.
- B. Create a run profile.
- C. Add a directory path record to the repository database.
- D. Install the QualityStage Server and the project directories will automatically be created.

Answer: A

---

### QUESTION 11:

Which add-on is used to certify Canadian addresses?

- A. SERP
- B. WAVES
- C. CASS
- D. DPID

Answer: A

---

**QUESTION 12:**

Which three actions can you perform with Pattern Action Language? (Choose three.)

- A. Move individual token into separate output field.
- B. Create phonetic representation for the parsed fields.
- C. Create checksum value for standardized fields.
- D. Concatenate multiple values into a single field.
- E. Perform complex calculation.

Answer: A,B,D

---

**QUESTION 13:**

What is the result of a Survive stage?

- A. a Standardized set of records
- B. a single record per group of records
- C. a set of records with a different layout
- D. a matched set of records

Answer: B

---

**QUESTION 14:**

You are matching on a field which is numeric and contains typographical errors. Which match comparison should you choose?

- A. DATE8
- B. CNT\_DIFF
- C. NUMERIC
- D. ABS\_DIFF

Answer: B

---

**QUESTION 15:**

Which statement describes why a blocking approach to matching is a reasonable trade-off between throughput and accuracy?

- A. Probabilistic matching is superior to deterministic matching.
- B. Each succeeding match pass has diminishing returns.
- C. Usually, any two records will not match.
- D. The accidental agreement is effectively the probability that values agree at random.

Answer: C

---

**QUESTION 16:**

You created a Match stage with the Undup option where three types of records will be generated: Matches, Clerical, Residuals. How many weight cutoffs should you specify?

- A. none
- B. one
- C. three
- D. two

Answer: D

---

**QUESTION 17:**

Which components of the standardization rule set are required?

- A. Classification Table (.CLS)  
Pattern Action File (.PAT)  
LookUp Tables (.TBL)
- B. Classification Table (.CLS)  
Pattern Action File (.PAT)  
Rules File (.RUL)
- C. Classification Table (.CLS)  
Pattern Action File (.PAT)  
Dictionary File (.DCT)  
LookUp Tables (.TBL)
- D. Classification Table (.CLS)  
Pattern Action File (.PAT)  
Dictionary File (.DCT)

Answer: D

---

**QUESTION 18:**

Which two types of data are analyzed with Character Investigation? (Choose two.)

- A. tax ID
- B. name
- C. date
- D. address

Answer: A,C

---

**QUESTION 19:**

Which stage can directly input data using ODBC?

- A. Connect stage
- B. Format Convert stage
- C. Transfer stage
- D. Investigate stage

Answer: B

---

**QUESTION 20:**

What distinguishes probabilistic matching from deterministic matching?

- A. The use of a hierarchy of rule precedence.
- B. The use of decision tables and scoring thresholds.
- C. The use of predefined scoring outcomes.
- D. The use of statistical properties of the data values.

Answer: D

---

**QUESTION 21:**

You have a file of German addresses. How should you standardize it?

- A. Create a standardization stage using US address rules and change it appropriately.
- B. Create a standardization stage using provided German address rules.
- C. Create a CASS stage.
- D. Create your own standardization rules because no German rules are available with QualityStage product.

Answer: B

---

**QUESTION 22:**

Click the Exhibit button.

A developer has run an investigation on the name field of source data and reviewed the results.

Using the exhibit, which assumption can be made on the population of the name field based on the provided report?

00000796	83.022% <<	[X]	Florence811	Bernardy811
00000001	6.800%	[X]	Margaret	Peacock
00000781	3.687% <F	[X]	Aristin	Albert
00000123	0.867% <F	[X]	Julie	Ford814
00000063	1.400% <?	[X]	Amy814	Edwards
00000039	0.867% <F	[X]	Kelsey815	Norman
00000010	0.400% F<<	[X]	Billie Bob741	Smith741
00000011	0.244% ??	[X]	Raj	Batra
00000008	0.178% ?<	[X]	Aily	West735
00000006	0.133% FL<	[X]	Hung Van	Tu732
00000005	0.111% ?F	[X]	Arvin	Bailey
00000008	0.67% <F	[X]	Risk	County
00000001	0.022% FLF	[X]	Kong Woo	Kim
00000001	0.022% FLF	[X]	Billie Bob	Smith
00000001	0.022% FLF	[X]	Hung Van	Tu

- A. A new rule set must be created to handle the name field.
- B. The length of the name field specified in the source data file definition is too long.
- C. The name field can not be used as a blocking variable in match.
- D. Additional information is being appended to some values in the name field.

Answer: D

#### QUESTION 23:

What is used to compensate for errors introduced by using a blocking strategy in match?

- A. weight overrides
- B. survivorship
- C. multiple passes
- D. missing values

Answer: C

#### QUESTION 24:

A user needs to include a word and its misspellings in a standardization process. Which two actions may be used to add this? (Choose two.)

- A. Use the pattern input tab of the standardization override screen.
- B. Add the item to the classification table.
- C. Use the classification tab of the standardization override screen.
- D. Use the Pattern Action Language table.

Answer: B,C

#### QUESTION 25:

The QualityStage Real Time Manager is installed. Which two functions are provided by this service? (Choose two.)

- A. The QualityStage Real Time Manager provides connection support between real time clients and QualityStage Real Time servers.
- B. The QualityStage Real Time Manager starts the QualityStage batch server, if necessary.
- C. The QualityStage Real Time Manager starts QualityStage Real Time servers and tracks them.
- D. The QualityStage Real Time Manager provides load balance statistics.

Answer: A,C

---

**QUESTION 26:**

You are matching on a field which is numeric and contains typographical errors. Which match comparison should you choose?

- A. NUMERIC
- B. ABS\_DIFF
- C. CNT\_DIFF
- D. DATE8

Answer: C

---

**QUESTION 27:**

Why should you run Investigation on the output from standardization?

- A. You can not run an investigation on the output from standardization.
- B. Analyze and quantify standardization results.
- C. Analyze and quantify certification results.
- D. Automatically correct unhandled data.

Answer: B

---

**QUESTION 28:**

You are running a match with one pass where age is one of the blocking fields. However, some of the values in the age field are invalid or blank. Those records are flagged as residuals. The customer would like those records to be matched as well. What should you do?

- A. Run the file through standardization to fix the problem.
- B. Have the customer correct the records in question
- C. Run the file through a Unijoin to fix the age field.
- D. Create a second match pass using different blocking criteria.

Answer: D

---

**QUESTION 29:**

A QualityStage job has patient visit records for input. The job should group all visits for a patient together. How can this grouping be accomplished?

- A. Use the GeoMatch option.
- B. Use the Match Sets option.
- C. Use the Undup option.
- D. Use the GeoMatch Multiple option.

Answer: C

---

**QUESTION 30:**

"Bob" appears frequently in the name field, while the name "Jim" appears infrequently. Which statement is true of an UNCERT match comparison on the name field?

- A. The comparison of Jim to Jim scores the same as the comparison of Bob to Bob.
- B. The comparison of Jim to Bob scores higher than the comparison of Bob to Jim.
- C. The comparison of Jim to Jim scores higher than the comparison of Bob to Bob.
- D. The comparison of Bob to Bob scores higher than the comparison of Jim to Jim.

Answer: C

---

**QUESTION 31:**

Which stage allows QualityStage to read data using ODBC?

- A. Connect
- B. Select
- C. Format Convert
- D. Transfer

Answer: C

---

**QUESTION 32:**

Which field should be used as a critical VARTYPE for a householding match?

- A. Name Type
- B. Unit Value
- C. Number of Name Words
- D. Gender Code

Answer: B

---

**QUESTION 33:**

Which field is appropriate as a first pass blocking field in a name and address match?

- A. Tax ID Number
- B. Title
- C. Employee Status Code
- D. Soundex of Middle Name

Answer: A

---

**QUESTION 34:**

A group of addresses are failing standardization because of an unrecognized street type. Which technique should be used to customize the address rule set?

- A. Use the input pattern tab in the Standardization Overrides feature to correctly handle patterns with this occurrence.
- B. Add the street type item using the classification tab in the Standardization Overrides feature.
- C. Use the input text override screen to correctly handle these occurrences.
- D. Modify the Pattern Action Language to recognize occurrences of this street type.

Answer: B

---

**QUESTION 35:**

What determines the location of the master project?

- A. the location of the QualityStage Designer client
- B. it is an installation option
- C. the run profile
- D. the job's advanced options tab

Answer: C

---

**QUESTION 36:**

Which three actions does the Save Language in a Unijoin stage do? (Choose three.)

- A. Perform arithmetic.
- B. Manipulate fields.
- C. Call subroutines.
- D. Create file backups.
- E. Compare values.

Answer: A,B,E

---

**QUESTION 37:**

Which two statements are true about a job that contains an Investigation stage? (Choose two.)

- A. It can contain multiple Investigation stages, but no other types of stages.
- B. It cannot be run in Real Time.
- C. It can contain other stages, but it cannot contain more than one Investigation stage.
- D. It contains no other stages.

Answer: B,D

---

**QUESTION 38:**

What are three reasons for using the QualityStage CASS stage? (Choose three.)

- A. Isolate non-address items.
- B. Allow the user to receive postal discounts.
- C. Validate commercial names.
- D. Discover undeliverable addresses.
- E. Correct address components.

Answer: B,D,E

---

**QUESTION 39:**

Your customer is producing a mass mailing and is very interested in keeping the cost of mailings down versus getting the most complete coverage. Which two should you consider? (Choose two.)

- A. using household unduplication
- B. using customer unduplication that might under match
- C. using customer unduplication that might over match
- D. skipping address standardization

Answer: A,C

---

**QUESTION 40:**

Which statement is true of RealTran mode for QualityStage Real Time?

- A. RealTran mode must signal end of data.
- B. RealTran mode processes all the input data as a single transaction.

- C. RealTran mode processes are always shared.
- D. RealTran mode processes data on a per transaction basis.

Answer: D

---

**QUESTION 41:**

A QualityStage Real Time application standardizes input records and matches them to standardized records stored in a database table. Which statement is true?

- A. Selection of candidate records should be based on the matching variables specified in the QualityStage Match job.
- B. All records in the reference database should be passed to the QualityStage Match job to guarantee accurate results.
- C. Selection of candidate records should be based on the blocking and matching variables specified in the QualityStage Match job.
- D. Selection of candidate records should be based on the blocking variables specified in the QualityStage Match job.

Answer: D

---

**QUESTION 42:**

A user has an input with a name and three address fields (ADDR1, ADDR2, ADDR3). Investigation has shown that ADDR2 contains no address information but does have additional name data. Which technique should be used to process the name information in ADDR2?

- A. Process this address field along with the ADDR1 and ADDR3 fields in address standardization.
- B. Since this field contains no address data do not include it in any standardization process.
- C. Use a data preparation rule set to get the ADDR2 data into a name domain.
- D. Include the name and ADDR2 fields as input to the name standardization process.

Answer: D

---

**QUESTION 43:**

How would you enable the UNICODE capabilities of QualityStage?

- A. Use the UNICODE enabled stages in the QualityStage Designer.
- B. Install the UNICODE version of the server.
- C. Install the UNICODE version of the QualityStage Designer.
- D. Check the radio button to run the UNICODE server on the job run profile.

Answer: B

---

**QUESTION 44:**

The business users want to determine the percent of valid population in the tax ID number field. Valid tax ID numbers are defined by the business users as being nine digit numeric values that do not contain suspicious values (e.g. 999999999). What type of field mask would provide the business users with the desired view of the tax ID number field?

- A. t mask
- B. c mask
- C. n mask
- D. x mask

Answer: B

---

**QUESTION 45:**

A customer has data in XML format that they want to process with QualityStage. How can QualityStage use the data? (Choose two.)

- A. Use DataStage to pre-process the XML.
- B. Use the XML importer.
- C. Require the input in another format.
- D. Use the Format Convert stage.

Answer: A,C

---

**QUESTION 46:**

Which three fields are generated by CASS processing? (Choose three.)

- A. carrier route code
- B. postal delivery point
- C. census tract
- D. county of residence
- E. valid ZIP + 4

Answer: A,B,E

---

**QUESTION 47:**

A customer wants to split data into two separate streams based on values populated in the SOURCE and TYPE fields. How can this be done?

- A. Create an unduplication job to perform an exact match by specifying the SOURCE and TYPE fields as blocking variables and including no matching variables.
- B. Specify a field in the input data file definitions that include both the SOURCE and TYPE data and use the Select stage to split the data based on the newly created field.
- C. Use the Split stage to filter data and specify the SOURCE and TYPE fields as the select values.
- D. Customize the PREP rule set classification table to tag SOURCE and TYPE values and write records to separate streams based on the newly created tag value.

Answer: B

---

**QUESTION 48:**

Which two statements describe a purpose for standardization? (Choose two.)

- A. Validate ZIP Codes within states.
- B. Provide internally consistent data content and format.
- C. Improve match processing.
- D. Fix invalid addresses.

Answer: B,C

---

**QUESTION 49:**

An address list must be processed through a standardization process. Investigation has discovered the occurrence of a unique address pattern that does not fully standardize. Which technique should be used to handle this?

- A. Drop this pattern with Word Investigation.
- B. Allow the address standardization process to handle what it can and let the rest of the data go to the unhandled data field for subsequent analysis and processing.
- C. Use the input pattern override feature to process this data.
- D. Use the input text override feature to handle this data.

Answer: C

---

**QUESTION 50:**

Your customer is producing a mass mailing and is very interested in keeping the cost of mailings down versus getting the most complete coverage. Which two should you consider? (Choose two.)

- A. skipping address standardization
- B. using customer unduplication that might over match
- C. using customer unduplication that might under match
- D. using household unduplication

Answer: B,D

---

**QUESTION 51:**

Which statement is true about the Standardize stage?

- A. The Dictionary file contains the rules.
- B. The Dictionary file provides a description of the rule set for the Available Processes list.
- C. The Dictionary file identifies and classifies key words.
- D. The Dictionary file defines the fields for the output file.

Answer: D

---

**QUESTION 52:**

Which statement is true?

- A. Standardize stage uses patterns pre-built in Investigation Stage.
- B. Standardize stage uses domain specific database lookup tables.
- C. Standardize stage uses Dictionary files to create patterns.
- D. Standardize stage dynamically builds patterns for the data.

Answer: D

---

**QUESTION 53:**

A freight company has a customer file containing both sold-to and ship-to address fields. These fields need to be standardized separately within the same standardization process. Which method should be used to standardize these address fields?

- A. Use the unhandled pattern override feature to modify the processing of the ship-to address.
- B. Copy the current address rule set using the Rules Management option. Use this rule set along with the original rule set.
- C. Modify the Pattern Action Language in the address standardization rule set to accommodate the second address.
- D. Use the data preparation override feature to combine the address fields.

Answer: B

---

**QUESTION 54:**

Which stage should you use to standardize and validate input addresses from multiple countries in a single input file?

- A. CASS
- B. SERP
- C. WAVES
- D. Standardization using the COUNTRY rule set and then create a separate stream for each country

Answer: C

---

**QUESTION 55:**

In which stage can data be formatted so that the first letter of each word is capitalized?

- A. Parse stage
- B. Collapse stage
- C. Format Convert stage
- D. Transfer stage

Answer: D

---

**QUESTION 56:**

You installed, but have not yet started, the QualityStage server on a UNIX machine. You plan to use the QualityStage server with the IBM Parallel Extender engine. Which two actions must be performed? (Choose two).

- A. You must add the INTBIN environment variable to your PATH.
- B. You must enable the Projects directory for NFS mount.
- C. You must install and configure Parallel Extender before starting the QualityStage server.
- D. You must run the qs\_license.sh script.

Answer: A,C

---

**QUESTION 57:**

A customer has data in XML format that they want to process with QualityStage. How can QualityStage use the data? (Choose two.)

- A. Use the XML importer.
- B. Require the input in another format.
- C. Use DataStage to pre-process the XML.
- D. Use the Format Convert stage.

Answer: B,C

---

**QUESTION 58:**

You are experiencing an increased volume of real time transactions going into the QualityStage Real Time standardization application. What should you do to handle the increased volumes?

- A. Increase thread pool count.
- B. Remove database lookups.
- C. Convert to a batch application.
- D. Remove standardization.

Answer: A

---

**QUESTION 59:**

An IT analyst needs to understand how match scores are being calculated at the field level. Which report or file can be used to find this information?

- A. the Match extract file
- B. the Match Debug report
- C. the Weight Overrides Summary report
- D. the pattern frequency report from Word Investigation

Answer: B

---

**QUESTION 60:**

Which statement is true about QualityStage Standardization?

- A. Input records containing separate first and last name fields do not require standardization to be performed on those fields.
- B. Records should be unduplicated and survived prior to standardization to reduce run time.
- C. It is necessary to standardize all fields in a record before attempting to match records.
- D. Placement of data within the context of a record is used by Standardization to help determine the meaning of the data.

Answer: D

---

**QUESTION 61:**

What is a purpose for standardization?

- A. Correct last name spellings.
- B. Add phonetic values to key words.
- C. Determine the nationality of a personal name.

D. Validate addresses.

Answer: B

---

**QUESTION 62:**

You have a multi-stage job in which you wish to set different starting and ending stages in different runs. Which run mode should be used?

- A. ParallelExtender
- B. deploy
- C. file
- D. data stream

Answer: C

---

**QUESTION 63:**

The QualityStage server is to be installed on UNIX with the client on Windows XP. Which two permissions must be enabled for the QualityStage user ID? (Choose two.)

- A. Create files.
- B. Read database configuration files.
- C. Create directories.
- D. Create users.

Answer: A,C

---

**QUESTION 64:**

Which two statements are true of comparison thresholds within a QualityStage rule set? (Choose two.)

- A. Comparison thresholds are used in blocking records.
- B. Comparison thresholds help overcome spelling errors.
- C. Comparison thresholds determine whether records are considered duplicates.
- D. Comparison thresholds are used in classification tables.

Answer: B,D

---

**QUESTION 65:**

Which match comparison type should be used to give a positive score to two tax ID numbers that differ in one digit (e.g. 555224321 vs. 555224322)?

- A. CNT\_DIFF

- B. CHAR
- C. ABS\_DIFF
- D. NUMERIC

Answer: A

---

**QUESTION 66:**

You assigned a u-prob of 0.1 to the comparison for the Last Name field. Frequency statistics are calculated for all fields. What will be used for the weight calculation during the match run?

- A. U-prob value will be replaced based on the frequency of specific field values.
- B. 0.1 will be multiplied by field frequency.
- C. U-prob must be ignored completely.
- D. 0.1 will be used.

Answer: A

---

**QUESTION 67:**

In which phase is Data Investigation a step in the QualityStage data re-engineering process?

- A. Evaluating Results phase
- B. Understanding the Business Requirements phase
- C. Phase Four
- D. Phase Two

Answer: D

---

**QUESTION 68:**

The p.FRQ report is produced by which stage?

- A. Standardize
- B. Unijoin
- C. Investigate
- D. Match

Answer: C

---

**QUESTION 69:**

The pattern report produced by Word Investigation helps with which task?

- A. discovering inconsistencies in data type representation
- B. deciding which fields to use for blocking fields
- C. refining a Standardization rule set
- D. estimating the system resources needed to run the application

Answer: C

---

**QUESTION 70:**

A clearing house with data from all over the United States has problems recognizing many different spellings for states with long names (e.g. Mississippi). Which technique should be used to modify a rule set to support most misspellings of long words like Mississippi?

- A. Add the word "MISSISSIPPI" to the classification table along with a match comparison threshold.
- B. Use Word Investigation with the USAREA rule set to convert all misspellings of "MISSISSIPPI".
- C. Allow the CASS certification process to correct all state name misspellings.
- D. Modify the pattern action language (PAL) to look for and modify the variant spellings of "MISSISSIPPI".

Answer: A

---

**QUESTION 71:**

The n.DLT report is produced by which stage?

- A. Investigate
- B. Standardize
- C. Unijoin
- D. Match

Answer: A

---

**QUESTION 72:**

The business analyst tells you that they expect to be able to unduplicate over 50% of their customer records using Social Security numbers. How do you determine if this is feasible?

- A. Review the results of Character Investigation on the field containing the Social Security data.
- B. Design the match to accept missing values in the Social Security data field.
- C. Be sure to standardize the Social Security data prior to the match.
- D. Ask others in the IT department if they have the same expectations as the business

analyst.

Answer: A

---

**QUESTION 73:**

You are upgrading the QualityStage server as well as all QualityStage client machines. Before upgrading to the new software you want to back up the rule sets. Where are these rule sets located?

- A. in the RULES directory on the QualityStage server
- B. in the DIC directory on the QualityStage server
- C. in the DATA directory on the QualityStage server
- D. in the RULES directory on the QualityStage client

Answer: D

---

**QUESTION 74:**

Word Investigation produces which three reports? (Choose three.)

- A. language distribution
- B. weight histogram
- C. word classification
- D. token type pattern
- E. word frequency

Answer: C,D,E

---

**QUESTION 75:**

What determines whether a field is populated in the Survive stage?

- A. the target
- B. the group identifier
- C. the output data file definition
- D. the rules

Answer: A

---

**QUESTION 76:**

Which add-on stage will certify Puerto Rican addresses?

- A. WAVES
- B. SERP

- C. CASS
- D. DPID

Answer: C

---

**QUESTION 77:**

The field you want to use for matching has spelling errors. Which match comparison type should be used for this field?

- A. ABS\_DIFF
- B. UNCERT
- C. CHAR
- D. PREFIX

Answer: B

---

**QUESTION 78:**

What is the maximum number of passes allowed in one match stage?

- A. 7
- B. 8
- C. 10
- D. 5

Answer: A

---

**QUESTION 79:**

Which three are valid field masks in an investigation?

- A. O
- B. N
- C. T
- D. C
- E. X

Answer: C,D,E

---

**QUESTION 80:**

What distinguishes probabilistic matching from deterministic matching?

- A. The use of predefined scoring outcomes.
- B. The use of a hierarchy of rule precedence.

- C. The use of decision tables and scoring thresholds.
- D. The use of statistical properties of the data values.

Answer: D

---

**QUESTION 81:**

What is helpful when defining the structure of the target data?

- A. Understanding how to organize data for investigation.
- B. Understanding the data mission that satisfies the business goal.
- C. Understanding the size and complexity of the project.
- D. Understanding the degree of data drift over time.

Answer: B

---

**QUESTION 82:**

Which two are reasons to use Investigation on data? (Choose two.)

- A. Reveal undocumented business rules.
- B. Determine a match pass cutoff.
- C. Verify the reliability of the data.
- D. Reduce the error rate of the data.

Answer: A,C

---

**QUESTION 83:**

You have run a Standardize stage using a NAME rule set. You have set the optional names handling to process all unhandled names as individuals. You notice that the standardization result is different when you test the rule using the Rules Analyzer. Which two are possible causes? (Choose two.)

- A. You did not save the input data first.
- B. You did not enter the proper delimiter.
- C. You did not enter the input data in all upper case.
- D. You did not enter the same input data into the Rules Analyzer.

Answer: B,D

---

**QUESTION 84:**

Which statement is true when creating a Survive stage?

- A. The input records need a group identifier.

- B. The output structure can be different from the input structure.
- C. You cannot define complex rules.
- D. You can only populate one field per rule.

Answer: A

---

**QUESTION 85:**

You have a multi-stage job in which you wish to set different starting and ending stages in different runs. Which run mode should be used?

- A. deploy
- B. data stream
- C. ParallelExtender
- D. file

Answer: D

---

**QUESTION 86:**

The provincial government of Ontario, Canada needs to modify its names rule set to identify corporate numbers located in a business name field. These corporate numbers must have a length of 9. Which statement is true?

- A. This condition can be handled using the input pattern tab in the Standardization Overrides feature.
- B. The Pattern Action Language in the CANAME rule set should be modified for this requirement.
- C. Word investigation with the CANAME rule set can be used to handle this without rule set modification.
- D. The only modification required is adding Ontario corporation numbers to a lookup table in the CANAME rule set.

Answer: B

---

**QUESTION 87:**

A customer's source data contains a birth date field in the format YYYYMMDD. Business requirements define a valid birth year as being less than or equal to the current year and greater than 1900. The customer wants to see how many records contain a valid year in the birth date field and decides to do a Character Discrete Investigation. Which mask type should they use in the Character Discrete Investigation?

- A. YYYYTTTT
- B. CCCCCCCC
- C. CCCCXXXX

D. TTTTTTTT

Answer: C

---

**QUESTION 88:**

What is the maximum number of passes allowed in one match stage?

- A. 7
- B. 10
- C. 8
- D. 5

Answer: A

---

**QUESTION 89:**

The pattern report produced by Word Investigation helps with which task?

- A. discovering inconsistencies in data type representation
- B. refining a Standardization rule set
- C. deciding which fields to use for blocking fields
- D. estimating the system resources needed to run the application

Answer: B

---

**QUESTION 90:**

You want to run a QualityStage job in parallel within DataStage. What is a valid way to do this?

- A. Use the QualityStage Parallel stage.
- B. You must use Windows batch files or UNIX scripts to complete the integration.
- C. Use QualityStage plug-in for DataStage.
- D. Use the Command stage in the DataStage job sequencer.

Answer: D

---

**QUESTION 91:**

A customer has a vendor list containing "doing business as" (DBA) information in the "company name" field. The customer wants to place just the DBA information into another field. Which technique should be used to handle this requirement?

- A. Use the QualityStage name standardization rule set as delivered to process the names.
- B. Add the word "DBA" to the rule set classification table.

- C. Use the custom output definition feature of the QualityStage standardization module.
- D. Set the "DBA" processing flag.

Answer: A

---

**QUESTION 92:**

Which statement is true about QualityStage Standardization?

- A. Placement of data within the context of a record is used by Standardization to help determine the meaning of the data.
- B. Input records containing separate first and last name fields do not require standardization to be performed on those fields.
- C. It is necessary to standardize all fields in a record before attempting to match records.
- D. Records should be unduplicated and survived prior to standardization to reduce run time.

Answer: A

---

**QUESTION 93:**

You are using the Date of Birth field as one of the BLOCKING fields. This field is defined with the option: Missing Value = S ( Spaces). In the match, what will happen to the records that contain spaces in the Date of Birth field?

- A. All records with spaces in the Date of Birth field will be matched together.
- B. All records with spaces in the Date of Birth field will be put in one group regardless of other blocking fields.
- C. The value of spaces will have no effect on the match.
- D. All records with spaces in the Date of Birth field will become residual records.

Answer: D

---

**QUESTION 94:**

The business analyst tells you that they expect to be able to unduplicate over 50% of their customer records using Social Security numbers. How do you determine if this is feasible?

- A. Review the results of Character Investigation on the field containing the Social Security data.
- B. Design the match to accept missing values in the Social Security data field.
- C. Ask others in the IT department if they have the same expectations as the business analyst.
- D. Be sure to standardize the Social Security data prior to the match.

Answer: A

---

**QUESTION 95:**

A standardization rule set may be used in which other stage?

- A. Match
- B. Investigate
- C. Unijoin
- D. Survive

Answer: B

---

**QUESTION 96:**

When multiple rules are specified for the same target in a Survive stage, how is precedence determined?

- A. The first rule is processed and all the rest are ignored.
- B. The rules appearing earlier in the list have precedence.
- C. The rules appearing later in the list have precedence.
- D. The value is specified in the rule priority option.

Answer: C

---

**QUESTION 97:**

Which two are reflected in the match agreement and disagreement weights for a field?  
(Choose two.)

- A. reliability
- B. cutoff
- C. discriminating power
- D. standard deviation

Answer: A,C

---

**QUESTION 98:**

You assigned a u-prob of 0.1 to the comparison for the Last Name field. Frequency statistics are calculated for all fields. What will be used for the weight calculation during the match run?

- A. 0.1 will be multiplied by field frequency.
- B. U-prob must be ignored completely.

- C. 0.1 will be used.
- D. U-prob value will be replaced based on the frequency of specific field values.

Answer: D

---

**QUESTION 99:**

Which run mode is most supportive of incremental job development and debugging?

- A. file
- B. stream
- C. parallel
- D. deploy

Answer: A

---

**QUESTION 100:**

Which scenario would require the use of an Array Field?

- A. When a customer needs a better CASS reference match.
- B. When a record fails area standardization.
- C. When there are many unstructured data fields to compare.
- D. When there is a need to compare a randomly sequenced group of fields.

Answer: D

---

**QUESTION 101:**

Which built-in QualityStage tool would you use to insert a new rule?

- A. Standardization Rules Tester
- B. Standardization Overrides
- C. Standardization Rules Analyzer
- D. Standardization Rules Management

Answer: B

---

**QUESTION 102:**

The results of which process are dependent on frequency analysis?

- A. Match
- B. Standardization
- C. Clerical Review
- D. CASS Certification

Answer: A

---

**QUESTION 103:**

What are two reasons to use a Survive stage? (Choose two.)

- A. to survive a set of matched records
- B. to survive a record in its entirety
- C. to survive the longest record
- D. to survive the best data from several matched records

Answer: B,D

---

**QUESTION 104:**

Which add-on stage will certify Canadian addresses?

- A. SERP
- B. DPID
- C. WAVES
- D. CASS

Answer: A

---

**QUESTION 105:**

Which type of investigation should the developer use to perform match block analysis on the ZIP Code field?

- A. Word Investigation using AREA rule set
- B. Character Discrete Investigation with char mask
- C. Character Discrete Investigation with type mask
- D. Character Concatenate Investigation with char mask

Answer: B

---

**QUESTION 106:**

Which run mode is most supportive of incremental job development and debugging?

- A. parallel
- B. deploy
- C. file
- D. stream

Answer: C

---

**QUESTION 107:**

Which two are valid Survive stage rule techniques? (Choose two.)

- A. all upper case
- B. most frequent non-blank
- C. longest
- D. average value

Answer: B,C

---

**QUESTION 108:**

In which two locations should standard values of individual words be stored for standardization rules? (Choose two.)

- A. QualityStage repository
- B. Dictionary file
- C. Classification table
- D. Look-Up Table

Answer: C,D

---

**QUESTION 109:**

You want to interactively test a Domain-Specific rule set using the QualityStage UI. Which statement is true?

- A. You cannot change the Locale.
- B. The result will be all rule set fields, even the blank ones.
- C. You can enter multiple input strings.
- D. You can select a previously entered input string.

Answer: D

---

**QUESTION 110:**

When do you evaluate the results of a QualityStage data re-engineering process?

- A. only at the end of Phase Two and Phase Three
- B. only after the data has been conditioned and matched
- C. after each phase
- D. only in Phase Four

Answer: C

---

**QUESTION 111:**

A Real Time standardization process needs to recognize a first name that is not in the current QualityStage rule set. The correct gender must also be assigned to this name. Which two tables in the CANAME rule set should be modified to accommodate this requirement? (Choose two.)

- A. the classification table
- B. the dictionary table
- C. the gender table
- D. the firstname lookup table

Answer: A,D

---

**QUESTION 112:**

A QualityStage job has patient visit records for input. The job should group all visits for a patient together. How can this grouping be accomplished?

- A. Use the Match Sets option.
- B. Use the Undup option.
- C. Use the GeoMatch Multiple option.
- D. Use the GeoMatch option.

Answer: B

---

**QUESTION 113:**

Prior to a name standardization run it has been determined that a large percentage of name fields have address information as well. Which technique should be used to address this situation?

- A. Use pattern overrides to isolate the address information.
- B. Let the name standardization rule set isolate the address information.
- C. Correct the Pattern Action Language in the name rule set to eliminate the address information.
- D. Run a preparation rule set against the name field prior to name standardization.

Answer: D

---

**QUESTION 114:**

Which statement is true about the Unijoin and Match stages?

- A. The Unijoin stage allows you to reformat output data while the Match stage only allows variables like WEIGHT and PASS to be appended to output records.
- B. The Match stage can use the frequency of data values when matching records while the Unijoin stage does not.
- C. The Unijoin allows conditional weighting to be applied to specific fields based on data values while the Match stage does not.
- D. The Match stage can perform statistical matching while the Unijoin stage can only perform exact matching.

Answer: B

---

**QUESTION 115:**

What is helpful when defining the structure of the target data?

- A. Understanding the degree of data drift over time.
- B. Understanding how to organize data for investigation.
- C. Understanding the data mission that satisfies the business goal.
- D. Understanding the size and complexity of the project.

Answer: C

---

**QUESTION 116:**

A customer wants to run a QualityStage job from DataStage using the plug-in. Which two statements are true about importing QualityStage meta data into DataStage? (Choose two.)

- A. You must not have any gaps between defined fields.
- B. You must change field names.
- C. You must have MetaStage installed.
- D. You must remove any redefines.

Answer: A,D

---

**QUESTION 117:**

What are two reports generated by a Word Investigation? (Choose two.)

- A. Pattern
- B. Summary
- C. Word Classification
- D. Word Standardization

Answer: A,C

---

**QUESTION 118:**

What is a purpose for standardization?

- A. Add phonetic values to key words.
- B. Determine the nationality of a personal name.
- C. Validate addresses.
- D. Correct last name spellings.

Answer: A

---

**QUESTION 119:**

You are running your QualityStage job using the Parallel Extender mode while varying the degrees of parallel that it is using. What is the key factor that influences performance?

- A. whether the job is CPU-bound or IO-bound
- B. location of the conductor node
- C. number of stages used by the job
- D. location of the buffer pool

Answer: A