

Exam Code: 000-419

Exam Name: IBM InfoSphere Quality Stage v8

Examination

Vendor: IBM

Version: DEMO

Part: A

1: When running Word Investigation, producing a pattern report will help you do what?

- A.Refine a standardization rule set.
- B.Decide which fields to use for blocking fields.
- C.Discover inconsistencies in data type representation.
- D.Identify which patterns should survive.

Correct Answers: A

2: A client has multiple sources of insurance policies which were consolidated into a single file. Upon initial investigation, you find that each source has its own format for the policy number field. How can you use the Character Concatenate investigation to identify the policy number format across the sources?

- A.Use Character Concatenate with Source Type (C Mask) and Policy Number (C Mask).
- B.Use Character Concatenate with Source Type (T Mask) and Policy Number (C Mask).
- C.Use Character Concatenate with Source Type (C Mask) and Policy Number (T Mask).
- D.Use Character Concatenate with Source Type (T Mask) and Policy Number (X Mask).

Correct Answers: C

3: A client states that phone data is in the North American format: 999-999-9999. Which field mask would determine if the client is correct?

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

Correct Answers: B

4: You are in the process of a business requirements definition for a data warehouse project. Your client tells you that free-form name information is clean and is an excellent field to use for matching and identifying a unique customer. How would you do initial validation of your client's statement? (Choose two.)

- A.Use Word Investigation to look at the name patterns.
- B.Use Character Discrete investigation to look at the individual name components.
- C.Use the Token Report to look at the token anomalies.
- D.Use Character Concatenate investigation of name and address data.

Correct Answers: A C

5: You have run a standardization of your client's name data. Which investigation would help you determine how much of the data was able to be standardized?

- A.Word Investigation of the standardized primary name field
- B.Character Discrete investigation of the Unhandled Pattern field of the name
- C.Character Discrete investigation of the first name and last name fields from the source
- D.Character Concatenation investigation of the first name and last Name fields from the source

Correct Answers: B

6: A customer's source data contains a transaction date field with format YYYYMMDD. The customer wants to see how many records contain a valid year in the transaction date field. Which mask type should they use in the Character Discrete investigation?

- A.YYYYYXXXX
- B.TTTTXXXX
- C.CCCCXXXX
- D.CCCCCCCC

Correct Answers: C

7: You are obtaining customer data from several separate source systems. As part of your design you are using the Unduplicate Match stage to identify duplicate customer records. This stage can deliver up to four outputs as a result of the match processing. What are two of these outputs? (Choose two.)

- A.Composite
- B.Resident
- C.Clerical
- D.Residual

Correct Answers: C D


8: What is a result when you increase the match cut-off setting?

- A.It minimizes the false positives.
- B.It minimizes the false negatives.
- C.It adds agreement weight.
- D.It subtracts agreement weight.

Correct Answers: A

9: Click the Exhibit button.

Referring to the exhibit containing match pass statistics, which three statements are true? (Choose three.)

Match Pass: CustomerNameStreetAddress		
Pass Definition		Pass Statistics
Baseline Run:		
<None>		 Save Current Statistics Delete B
Chart	Type	Current
<input type="checkbox"/>	Data records read	606
<input type="checkbox"/>	Blocks processed	182
<input type="checkbox"/>	OVERFLOW blocks	0
<input type="checkbox"/>	Maximum Data block size (including overflow)	18
<input type="checkbox"/>	Average Data block size (not including overflow)	3
<input checked="" type="checkbox"/>	Pseudo matches	83
<input type="checkbox"/>	Data duplicates	300
<input type="checkbox"/>	EXACT Data duplicates	0
<input checked="" type="checkbox"/>	Clerical pairs	6
<input checked="" type="checkbox"/>	Data residuals (including SKIPS & MISSING)	300

- A.182 match sets containing at least one XA and one or more DA records were found.
- B.300 duplicate records (DA) were identified.
- C.Zero duplicate records (DA) were identified.
- D.217 residual records (RA) resulted from this pass.
- E.83 match sets containing at least one XA and one or more DA records were found.

Correct Answers: B D E

10: In a sample data set of 10,000,000 records, the Social Security number has 90 percent distinct values. To ensure that the Social Security number does not over-discriminate based upon its rare occurrence, which statement is true?

- A.Change the u-probability on the match command screen from .01 to .0000001.
- B.Set special variable handling "CRITICAL".
- C.Set special variable handling "NOFREQ".
- D.Do not use the Social Security number in a match command.

Correct Answers: C

11: For an Unduplicate Match, how does the Match stage determine the master record?

- A.the record within a block with the most unhandled data
- B.the record within a block with the most values populated
- C.the record within a block that matched to itself with the highest weight
- D.the record within a block that occurred last

Correct Answers: C

12: Which technique can be used to reduce the processing time of the Match Frequency stage?

- A.Create a frequency file for all the fields in the input file.
- B.Create a frequency file for only the fields used for blocking and matching.
- C.Create a frequency file for only the fields used for matching.
- D.Create a frequency file for only the fields used for blocking.

Correct Answers: C

13: Click the Exhibit button.

Examine the statistics shown in the exhibit.

What was the largest number of records processed in a block in Match Pass - CUSTOMER_NAME_ADDR?

Match Pass - CUSTOMER_NAME_ADDR		
Chart	Type	Value
<input type="checkbox"/>	Data records read	23
<input type="checkbox"/>	Blocks processed	19
<input type="checkbox"/>	OVERFLOW blocks	0
<input type="checkbox"/>	Maximum Data block size (including overflow)	2
<input type="checkbox"/>	Average Data block size (not including overflow)	1
<input checked="" type="checkbox"/>	Pseudo matches	3
<input type="checkbox"/>	Data duplicates	1
<input type="checkbox"/>	EXACT Data duplicates	0
<input checked="" type="checkbox"/>	Clerical pairs	2
<input checked="" type="checkbox"/>	Data residuals (including SKIPS & MISSING)	20

Match Pass - CUSTOMER_ZIP3		
Chart	Type	Value
<input type="checkbox"/>	Data records read	23
<input type="checkbox"/>	Blocks processed	14
<input type="checkbox"/>	OVERFLOW blocks	0
<input type="checkbox"/>	Maximum Data block size (including overflow)	4
<input type="checkbox"/>	Average Data block size (not including overflow)	2
<input checked="" type="checkbox"/>	Pseudo matches	2
<input type="checkbox"/>	Data duplicates	2
<input type="checkbox"/>	EXACT Data duplicates	0
<input checked="" type="checkbox"/>	Clerical pairs	1
<input checked="" type="checkbox"/>	Data residuals (including SKIPS & MISSING)	20

- A.2
- B.3
- C.6
- D.23

Correct Answers: A

14: You are cleansing customer mailing data to ensure there will be no duplicate customers. As part of your design you have standardized the name and address data and will be using the Unduplicate Match stage to identify the duplicate customer records. Which two fields would typically be considered as good blocking fields in this name and address unduplicate match? (Choose two.)

- A.phonetic postal code
- B.postal code
- C.phonetic street name
- D.customer number

Correct Answers: B C

15: Click the Exhibit button.

In the exhibit, which three actions to the match specification would ensure that the second record is removed from the match set? (Choose three.)

SetID	RecordType	Weight	FirstName	MiddleName	PrimaryName
36	XA	62.33	Rudolph	R	Smith
36	DA	54.97	Rudolph	N	Smith
36	DA	62.33	Rudolph	R	Smith
36	DA	62.33	Rudolph	R	Smith
36	DA	62.33	Rudolph	R	Smith

- A. Add the MatchFirstName as a match column.
- B. Increase the agreement weight on PrimaryName.
- C. Increase the disagreement weight on MiddleName.
- D. Increase the Match cut-off value.
- E. Apply variable special handling of type "Critical Missing OK" on MiddleName.

Correct Answers: C D E

16: What are the two Match Specification probabilities that can be configured? (Choose two.)

- A. m probability
- B. u probability
- C. r probability
- D. pi probability

Correct Answers: A B

17: Which three components must be specified to configure a match specification? (Choose three.)

- A. frequency file
- B. match test database
- C. Standardize stage
- D. sample data set
- E. Match stage

Correct Answers: A B D

18: In your match comparison you have birth dates 1962/03/01 and 1962/08/01 appearing in both files. You suspect there is a data entry error in the month (03 versus 08). Which match comparison type should be used to give a positive score?

- A. LS_CHAR
- B. DISTANCE
- C. CHAR
- D. CNT_DIFF

Correct Answers: D

19: Which two columns are outputs of the Match Frequency stage? (Choose two.)

- A. qsFreqAddressLine1
- B. qsFreqFirstName
- C. qsFreqCount
- D. qsFreqColumnID

Correct Answers: C D

20: Which two processing methods should use candidate selection? (Choose two.)

A.real-time

B.initial load

C.one-time load

D.delta load

Correct Answers: A D