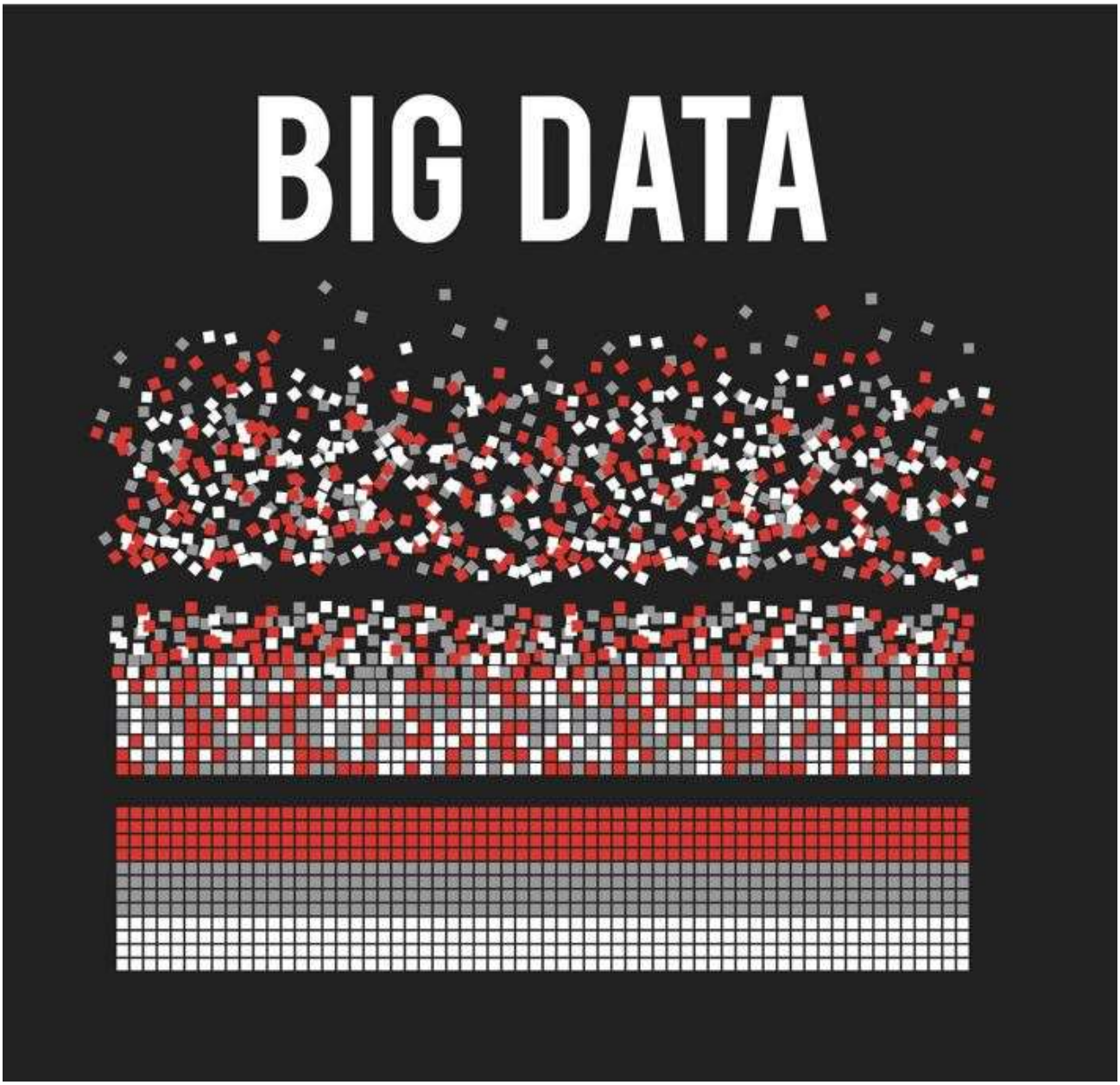

Assessor Guide

BSBXBD402

Test big data samples

Assessment 2 of 4

Short Answer Questions

SWINBURNE OPEN EDUCATION

Version 1.1

Assessment Details

This section is for SUT VE Quality and Compliance review and feedback and must be deleted in the student version of the assessment.

SECTION 1		
UNIT OF COMPETENCY DETAILS		
Code	Title	
BSBXBD402	Test big data samples	
COURSE AND MODULE DETAILS *Assessments may be published in more than one course. Add lines for additional courses as needed.*		
Course Code (UPed)	Module Number (Order)	Module Code (UPed)
SUT4002D	10	M00147
ASSESSMENT TYPE		
Assessment Method: *Select all that apply.*	**Questioning** Choose an item. Choose an item.	

SECTION 2
STUDENT INSTRUCTIONS *The following instructions detail the requirements of the assessment and are captured in the LMS assessment page. This includes a description of the student instructions, associated files and submission instructions.*
Student instructions
This is assessment 2 of 4 assessments for BSBXBD402 Test big data samples. This assessment requires you to answer 7 short answer questions to test your knowledge required of this unit. To be assessed as competent, you must complete all tasks in the spaces/format required. You are required to download your assessment by clicking on the assessment document icon below (see Let's begin) and upload your completed assessment for submission.
Supporting documents
Not Applicable
Files for submission
Submit the assessment document with all tasks completed in the spaces provided.
Submission instructions

PDF File Submissions

Please save all Word documents as PDF files before submitting.

IMPORTANT: Word documents will **not** be accepted.

Most modern web browsers can open and display a PDF file. If you have an older operating system, however, you may need a PDF reader installed on your device such as the Acrobat Reader, available from Adobe.

Windows: Word 2013 and newer

Choose **File** > **Export** > **Create PDF/XPS**.

Windows: Word 2010

1. Click the **File** tab
2. Click **Save As**
   - To see the Save As dialog box in Word 2013 and Word 2016, you have to choose a location and folder
3. In the **File Name** box, enter a name for the file, if you haven't already
4. In the **Save** as type list, click **PDF** (*.pdf).
   - If you want the file to open in the selected format after saving, select the Open file after publishing check box.
   - If the document requires high print quality, click Standard (publishing online and printing).
   - If the file size is more important than print quality, click Minimum size (publishing online).
5. Click **Options** to set the page to be printed, to choose whether markup should be printed, and to select output options. Click **OK** when finished.
6. Click **Save**.

macOS: Office for Mac

To save your file as a PDF in Office for Mac follow these easy steps:

1. Click the **File**
2. Click **Save As**
3. Click **File Format** towards the bottom of the window
4. Select **PDF** from the list of available file formats
5. Give your file a name, if it doesn't already have one, then click **Export**

For more detailed instructions refer to Microsoft Support.

SECTION 3

ASSESSMENT TASK CRITERIA AND OUTCOME

This assessment will be graded as Satisfactory (S) or Unsatisfactory (US).

To achieve Satisfactory; valid, sufficient, authentic, and current evidence of meeting the criteria must be submitted.

Refer to the mapping spreadsheet for details for this unit.

SECTION 4

ASSESSMENT DETAILS

Please refer to SECTION 2 to confirm how the assessment tools will be built and the methods that will be used to collect evidence i.e., Student's will type answers directly into LMS or will upload of files of completed assessment tasks.

The STUDENT INSTRUCTIONS above will be added directly into the LMS.

All associated files will be accessed via the LMS, as will any Assessor Guides, Matrix, Templates etc.

Students and Assessors have restricted permissions in the LMS. Assessor Guides, including model answers, will be available to Assessors ONLY.

The following pages contain the draft assessment which will be built into the LMS once reviewed. This includes:

- ☒ Instructions to students
- ☒ Questions /tasks
- ☒ Templates /tables where applicable
- ☒ Links to supporting files /websites
- ☒ Instructions to assessors
- ☒ Sample answers /examples of benchmark answers

SECTION 5

STAKEHOLDERS AND SIGN OFF

List all that apply for each of the stakeholder roles below.

UPed Learning Designer/Author name	Shevorne De Silva
SOE Quality and Compliance Manager name	Manie Guenot
SUT VE Quality Compliance name	Simon Hitchick
Date approved	12/05/2022

Assessment Instructions

Task overview

This assessment task contains seven (7) short answer questions. Read each question carefully before typing your response in the space provided.

Assessment Information

Submission

You are entitled to three (3) attempts to complete this assessment satisfactorily. Incomplete assessments will not be marked and will count as one of your three attempts.

All questions must be responded to correctly to be assessed as satisfactory for this assessment.

Answers must be typed into the space provided and submitted electronically via the LMS. Hand-written assessments will not be accepted unless previously arranged with your assessor.

Reasonable adjustment

Students may request a reasonable adjustment for assessment tasks.

Reasonable adjustment usually involves varying:

- the processes for conducting the assessment (e.g. allowing additional time)
- the evidence gathering techniques (e.g. oral rather than written questioning, use of a scribe, modifications to equipment)

However, the evidence collected must allow the student to demonstrate all requirements of the unit.

Refer to the Student Handbook or contact your Trainer for further information.

Please consider the environment before printing this assessment.

Question 1

Identify three (3) Australian Privacy Principles (APPs) that relate to testing big data sources and outline the obligations of each.

(Word count: 20 - 40 words for each APP)

Assessor instructions: Students must correctly identify the APPs that relate to testing big data sources and outline the obligations for each within the specified word limit. Assessors may refer to the Guide to Data Analytics and the Australian Privacy Principles (Long URL: https://www.oaic.gov.au/privacy/guidance-and-advice?a=3086) for more information on APPs.

A sample answer is provided below.

#	Australian Privacy Principle	Obligations as it relates to testing big data sources *(20 – 40 words)*
1	APP 3 – Collection of personal information	Organisations should: - collect information only by lawful and fair means - collect sensitive information only with the individual's consent (unless an exception applies).
2	APP5 – Notification	Organisations should not use personal information for a purpose other than the primary purpose it was collected for, unless an exception applies.
3	APP 11 – Security of personal information	Organisations should actively consider whether they are permitted to retain personal information. When personal information is no longer needed it should be destroyed or de-identified.
	Other answers may include: APP10 – Quality of personal information	Organisations should take responsible steps to ensure that personal information: - collected is accurate, up-to-date and complete - used or disclosed is having regard to the purpose of the use or disclosure is accurate, up-to-date, complete and relevant.

Question 2

Outline three (3) features of interactive big data sources.

(Word count: 25 - 35 words)

Assessor instructions: Students must provide three features of interactive, batched and real-time big data sources. The student may use different wording to describe the features. Provided answers need to reflect the characteristics described in the provided exemplar answer.

A sample answer is provided below.

Interactive big data sources
Raw data from this source: 1. is used to find patterns in data 2. is processed within minutes and with low latency 3. can be explored using various platforms that can run ad-hoc queries. Other answers may include: • provides the ability to work with more data • can scale reliably when new data becomes available.

Question 3

Identify and describe the (3) types of data formats found in common big data sources (i.e. batch, real-time or interactive) and provide (3) examples for each. Use the table given below to write your answers.

(Word count: 15 - 35 words for each description)

Assessor instructions: Students must identify the three types of data formats and provide a description and three examples for each format.

A sample answer is provided below.

Data format:	Description *(Word count: 15 - 35 words)*	Examples *(At least 3 for each data format)*
Structured	• Data has a predefined structure. • Can be easily stored in a tabular form in any relational database. • All rows in a table have the same set of columns • Can be easily processed and analysed	1. Relational database management system files (Oracle DB, MySQL, MS SQL DB, Postgres DB, etc) 2. Data from Enterprise systems (e.g. CRM, ERP) 3. Call detail records Other answers may include: • Metadata
Semi-structured	• Data has some structure, but not as rigid • It is best managed in NoSQL (non-relational) databases. Can also be stored in data lakes • Can be searched and processed using the semantic tags and metadata.	1. XML 2. JSON 3. Point of Sale transaction data (e.g. CSVs, spreadsheets) Other answers may include: 4. HTML 5. Social media data
Unstructured	• Data that does not have any structure • There's no pre-defined data format • Cannot be stored in traditional databases and do not contain any identifiable tags. • It is complex to search and process.	1. Email body 2. Audio files (the binary part) 3. Video files (the binary part) Other answers may include: • Blobs of text • Chat logs • Images • Word documents • Archived documents • User comments • Twitter Feeds

Question 4

Outline the following three (3) big data testing methodologies and the data validation protocols followed for each.

Note: You may apply the process to a specific big data testing platform of your choice (e.g. Hadoop, Power BI) when writing your answers.

(Word count: 60 - 80 words per methodology)

Assessor instructions: Students must identify the data validation protocols for each big data testing methodology given. The student may use different wording to describe the data validation protocols. However, the answers provided by the students need to reflect the protocols described in the exemplar answer given.

A sample answer is provided below.

Big data testing methodology	Data validation protocols *(Word count: 60 - 80 words per methodology)*
Data staging validation	• Determine the need to test the complete set of data or a representative sample dataset. • Verify the data obtained from the source system/file to confirm if it is corrupted or not by performing checksums, hashes or format checking. • Check data files to confirm if they were uploaded or imported into the analytical platform correctly. • Verify if the source data synchronises with the data that is uploaded into the analytical platform.
MapReduce validation	• Verify the business logic and align the dataset to relevant parts of the organisation to create a source to target mapping. • Validate the MapReduce process and data model schema to ensure the correct generation of the 'key-value' pair. • Implement data segregation rules on the data according to source to target mapping information. • Check if the aggregation and consolidation of data after the 'reduce' operation. • Take necessary action to clarify and resolve any anomalies identified.
Output validation	• Validating the dashboard report model to ensure correct visualisations are used to display the required data. • Checking the correct application of the transformation rules. • Check data integrity to ensure data is loaded into the target system successfully. • Check output against the specific source system data to ensure there is no data corruption. • Validation of reports with the required data and ensuring that all indicators are displayed correctly.

Question 5

Explain what 'test scripting' is and how it is used to perform big data validations.

(Word count: 45 - 65 words)
Assessor instructions: Students must explain test scripting and how is used to perform data validations.

A sample answer is provided below.

Test scripting is used to automate the testing process. It contains a set of instructions that can be performed on a system/platform to validate the loaded data.
There can be a high degree of scripting required to design test scenarios and test cases.
It can be used to point out any errors faster and easily.

Other answer may include:

- Once written these test scripts can be manipulated and reused to perform tests.
- These scripts for data validation can be written using scripting languages (e.g. Scala) and can be run on the system that contains the data to be validated.
- Test scripts can be used to compare data and validate their accuracy.

Question 6

Outline the **protocols and techniques** used when **performance testing big data throughput** in *PowerBI Performance Analyser*. Use the following table to write your answers for each listed criterion (A to E).

(Word count: 20 - 35 words per criterion)

Note: Refer to Microsoft's guidelines on Power BI Performance Analyser (Long URL: https://docs.microsoft.com/en-us/power-bi/create-reports/desktop-performance-analyzer#using-performance-analyzer)

Assessor instructions: Students must provide answers to all criteria listed in the table related to *PowerBI Performance Analyser*. The student may use different wording to describe the protocols and techniques. However, the answers provided by the students need to reflect the protocols described in the exemplar answer given.

A sample answer is provided below.

Criterion	Protocols and techniques *(35 – 55 words)*
A. Capturing performance information	**Start recording** option is used to begin the data collection process. It collects and displays the performance measurement information in real-time.
B. Computation of durations	This is computed as the difference between the start timestamp and the end timestamp of an operation, sometimes called "wall clock" durations. It is measured in milliseconds (ms).
C. Refreshing all visuals performance data	Use the **refresh visuals** option available in PowerBI Performance Analyser pane to refresh all visuals on the current page of the report.
D. Refreshing an individual visual's performance data	Whilst Performance Analyzer is recording the **Analyse this visual** option can be selected which can be found in the top-right corner of each visual, to refresh that visual, and capture its performance information.
E. Saving performance information	This can be done by selecting the **Export** button. Selecting Export creates a .json file with information from the Performance Analyser pane.

Question 7

When testing and validating big data, one would encounter a variety of issues. These issues then need to be processed and reported.

In this context, outline the **protocols and techniques** that are typically used in an organisation **when processing and reporting issues** related to big data samples.

(Word count: 85 – 115 words)

Assessor instructions: Students must outline the protocols and techniques for processing and reporting issues related big data validation. The student may use different wording to describe the protocols and techniques. However, the answers provided by the students need to reflect the protocols described in the exemplar answer given.

A sample answer is provided below.

- Various ticketing systems can be used (e.g. JIRA) to record, track and process the issues
- Test case templates can be used to document and process and record the results of the issues during various stages of testing.
- These documents which contain evidence of the issues can be used to report the issues to the relevant personnel after testing (e.g. management)
- The protocols for reporting issues may include the use of email, meetings, calls etc. The requirement to report to certain people in the hierarchy of the team/department (e.g. Lead tester, supervisor) may vary.

- Depending on the issue you may need to notify clients or third parties, according to certain communication protocols.

Assessment checklist:

Students must have completed all questions within this assessment before submitting. This includes:

1	7 short answer questions to be completed in the spaces provided.	☐

Congratulations you have reached the end of Assessment [2]!