Database Applications
ISYS1101/1102 | Semester 2 2024
Assignment 1: Database Design and Optimisation

Assessment Type	Database Design, Database Optimisation, Implementation, and Demonstration Individual Assessment
Due Date (M1)	Week 4, during the lab sessions. Further details will be provided on Canvas.
Due Date (M2)	23:59 Sunday 01 September 2024
Demonstrations (M3)	Week 7 - 8, during the lab sessions. Further details will be provided on Canvas.
Silence Period (M1)	None
Silence Period (M2 & M3)	Starts at 5:00PM Friday 30 August 2024
Weight	Milestone 1: 5 marks Milestone 2 and 3 (marked together): 20 marks
Submission	Online, via Canvas. Submission instructions are provided on Canvas.

1 Overview

1.1 Assessment Criteria

This assessment will determine your ability to:

1. analyse the requirements outlined in the problem description;
2. develop a conceptual model to assist you with the design of the database backend required for the system;
3. use an industry-standard ER modeling tool to draw the ER model and convert your ER model into a relational database schema;
4. identify and implementing efficient storage strategies for extremely large tables
5. identify and implementing efficient partition strategies for extremely large tables

6. write efficient queries on extremely large tables and describe the query plans that query optimiser would likely be using and explain how a cost-based query optimiser would execute such query plans;
7. write stored T-SQL procedures and functions to automate common tasks in an SQL Server database.

1.2 Learning Outcomes

This assessment will assess how you attained the following course learning outcomes:
CLO 1: apply advanced data analysis and modeling concepts, physical design, integrity, security and transaction management.
CLO 2: create stored procedures and functions to enhance the usability of a database;
CLO 3: apply techniques for efficient storing, accessing, securing, and recovering of data;
CLO 4: build an efficient database application with an emphasis on storage management, indexing, and query optimization.

---

2 Assessment Details

2.1 Preparation Work

You are required to implement the database backend for the below mentioned application on Microsoft SQL server. In order to successfully complete these tasks you must have completed Week 1 – 4 lab sheets and ensure that your SQL Server account is correctly configured and Azure Data Studio on your laptop is configured correctly.

2.2 Assignment Task Description

Introduction

The Australian Electoral Commission (AEC – https://aec.gov.au) is responsible for providing the Australian people with an independent electoral service which meets their needs and encourages them to understand and participate in the electoral **process**. Australia's manual system of federal elections has one of the most complex and time-consuming counting operations in the world. While it can at times require patience, the federal election counting process delivers (1) integrity to the results, concentrating on (2) accuracy in a (3) highly transparent manner.

While manual process ensures these three key priorities, there are two areas of concern to many stake holders, namely:

1. The time it takes to count votes and the human resources required to complete the process within an acceptable time frame
2. The volume of papers it requires and the environmental impact of running a manual election.

Let's suppose you are employed by a software development company that just received a contract from AEC to build a computerised voting system for federal elections. As in the case with manual elections,

the most important aspect of this system is to ensure the integrity of the voting system, accuracy, and transparency.

System requirements

The system is developed in several phases. The first phase, which you are responsible for, is limited to **federal general elections for House of Representatives**. The following voting processes are not in the scope of this phase:

1. Federal general elections for senate
2. Federal by-elections
3. State and territory elections
4. City council and shire council elections
5. Referendums
6. Any other election services provided by AEC

IGNORE

In this assignment, you are required to analyse the database requirements, design the database backend for the voting system, identify various database optimisations, and implement the system.

Elections

In Australia, federal elections are held every three years. In these elections, the voters get an opportunity to vote for both lower house (the house of representatives, which is in the scope of this assignment) and upper house (the senate, which not part of this assignment). There are currently 151 seats in the lower house, and 76 seats in the upper house.

As the scope of this assignment is limited to **federal general elections for House of Representatives** no further details are provided on the upper house (the senate) and how the elections are conducted for senate seats.

Representatives <- Division member <- Voters

Members of the House of Representatives are elected by the voters registered in each Electoral Division using full preferential voting. Each Electoral Division elects one member.

Electoral Divisions

For the House of Representatives, each state and territory is divided into electoral divisions (or commonly known as electorates or seats). Population determines the number of electoral divisions. To ensure continued equal representation, the boundaries of these divisions have to be redrawn (redistributed) periodically. As of last re-distribution based on 2017 population data, there are 151 electoral divisions in Australia.

State	Population (in 2017)	Electoral Divisions
NSW	7,797,791	47
VIC	6,244,227	38
QLD	4,883,739	30
WA	2,567,788	16

State <- Electoral Division <- Population

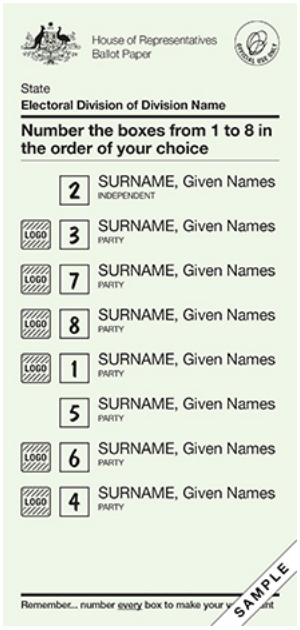
SA	1,716,966	10
TAS	519,050	5
ACT	419,256	3
NT	247,512	2
Total	**24,396,329**	**151**

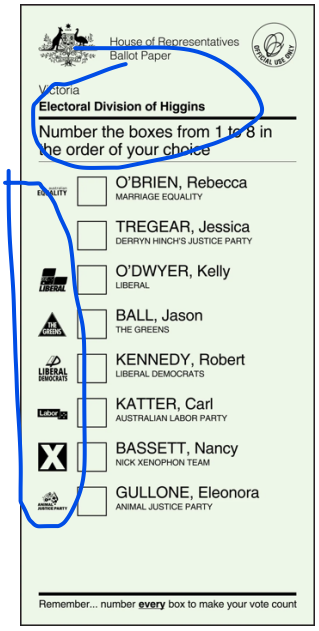
Relationship: One Party has 1 or more divisions.

Election Process

When the sitting government is nearing its term (3 years) or under circumstances it is dissolved, the Australian Electoral Commission (AEC) calls for nominations of candidates. Registered political parties will then nominate their candidates for one or more electoral divisions. Most political parties nominate candidates for many electoral divisions. Independent candidates can nominate themselves for the election.

Once the nomination process is over, AEC will determine the election date and will print ballot papers for each electoral division. A sample ballot paper is shown below. A screenshot of a real ballot paper (Higgins electoral division in 2016 election) is also shown.





<- Electoral division on paper

<-- It also has the party on it.

On the election date, registered voters are required to attend a polling station and cast their vote on a ballot paper similar to above. The actual voting process is much more flexible with pre-poll voting, postal voting, absentee voting, and declaration votes. However, for the scope of this assignment, we only consider regular voting process on election day.

The preferential voting system

Candidates for the house of representatives are elected using the preferential voting system. In this system, the voters are required to cast their order of preferences to ALL candidates contesting in their electoral division. As shown in the above ballot paper, the voter has given first preference to the fifth candidate on the ballot paper. The voter has given their second preference to the first candidate on the ballot paper, and so forth. <- This is an example.

At the end of the election day, after all the polling stations are closed, the counting begins. The counting of preferential votes is a complex process.

Step 1: Count of first preferences (primary vote)

In this step, all of the number "1" votes are counted for each candidate. If a candidate gets more than half the total first preference votes, that candidate will be elected.

Step 2: Distribution of preferences

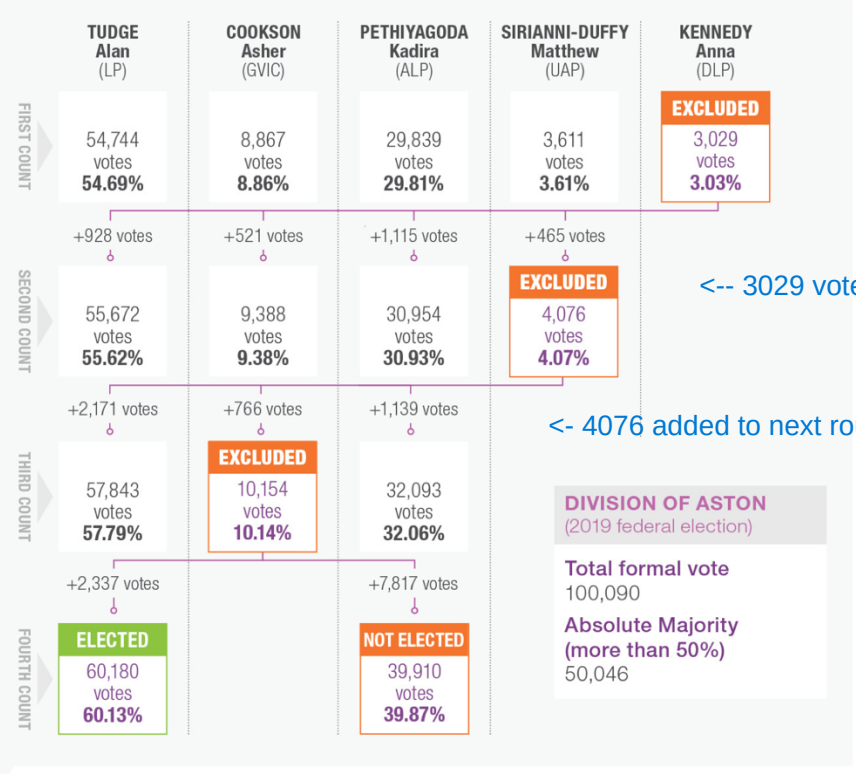
If no candidate has more than half of the votes, the candidate with the fewest votes is excluded. This candidate's votes are transferred to the candidates according to the second preferences of the voters on the ballot papers for the excluded candidate. If still no candidate has more than half the votes, the second-last candidate who now has the fewest votes are excluded and the votes are transferred according to the next preference on the ballot papers. This process is continued until one candidate has more than half the total number of valid votes.

This process is illustrated with a real example (distribution of preferences in Aston electoral division in 2019 election) on the page 32 of the following document.

https://www.aec.gov.au/about_aec/Publications/electoral_pocketbook/2019/2019-electoral-pocketbook.pdf

A screen shot of the above-mentioned page:

Example of a distribution of preferences



At the end of this two-stage counting process, the winning candidate is declared as the new Member of Parliament for the corresponding electoral division.

Components of the proposed system.

The proposed computerised election system must be able to conduct the entire election process for federal house of representatives elections. The major tasks in the election process are listed below:

1. Maintenance of electoral role
2. Maintenance of information required for conduct of an election (such as basic election information, electoral division information, political party information, candidate information, etc. Refer to details below.
3. Election Day – This process should mimic the manual process where a voter visits a polling information, once identification is established a ballot paper issued, marks their preferences and lodges the ballot paper.
4. Counting of ballot papers (counting is a complex process, refer to details below).

This system will maintain the following information.

1. Computerised Electoral Role Table 1: electoralRole or voterRegistry

The system will maintain a computerised electoral role, i.e. a database of registered voters for each electoral division. For each registered voter, following information is stored:

There are many Electoral divisions.
- Each division is made up of many voters

<-- INSERT A "VOTERID"

- Title
- First name*
- Middle names (if any)
- Last name*
- Gender
- Date of Birth*

Address

- Residential Address* (Unit number, street number, street name, suburb, postcode, state) <- De-normalize.
- (no letter box addresses accepted)
- Postal Address (as above, or can be different)
- Contact Details (daytime phone number, mobile phone number, email address)
- Electoral Division (determined by the system based on residential address)

There are around 17,259,000 Australians are currently enrolled to vote (https://www.aec.gov.au/Enrolling_to_vote/Enrolment_stats/index.htm). The increase of the size of the electoral role is approximately proportional to the population growth in Australia. The current population growth in Australia is approximately 1.2%. <-- Use this in storage calculation. (https://population.gov.au/sites/population.gov.au/files/2022-04/2022-23_budget_overview.pdf)

2. Details of Elections

Table 2: ElectionEvent or electionDetails

The following details about elections are required to be stored in the database.

- Election Serial Number (a unique code generated and stored by the system)
- Date of the election
- Type of election (house of representative, senate, by-election, etc)
- Total number of electoral divisions
- Total number of registered voters (the number of registered voters at the closing of the electoral role registrations for the corresponding election)

3. Details of Electoral Divisions

table 3:electoralDivisoin or electoralDetails

The following details about electoral divisions are required to be stored in the database.

https://www.aec.gov.au/About_AEC/Publications/electoral_pocketbook/2019/2019-electoral-pocketbook.pdf

- Electoral Division Name (refer to page 156 – 158 of https://www.aec.gov.au/about_aec/Publications/electoral_pocketbook/2019/2019-electoral-pocketbook.pdf) for full list
- Total number of currently registered voters
- Historical record of registered voters (the historical data are captured at closing date of the electoral role registrations for the past elections. Both the date and no. of voters are stored) <- Table for past elections: ElectionMaster
- Name and party of the current member of parliament

<-Candidate

4. Details of Political Parties

politicalParty

The following details about political parties are required to be stored in the database.

- Party Code (refer to page 160 of https://www.aec.gov.au/about_aec/Publications/electoral_pocketbook/2019/2019-electoral-pocketbook.pdf)
- Name of the party

- Party Logo
- Postal address of the party headquarters
- Secretary of the party
- Contact Person (name and other contact details such as daytime phone number, mobile, and email)

5. Details of Candidates table 5: detailCandidate

The following details about political parties are required to be stored in the database.

- Name
- Political Party Code (or IND, if they are independent)
- Contact Details name and other contact details such as daytime phone number, mobile, and email)
- Election Code
- Electoral Division Contesting

6. Computerised Ballot papers cast

The computerised ballot paper captures and stores voters' preferences (similar to what's written on a paper-based ballot paper.

Referential INtergity rtansactions

[Very important] To ensure integrity and confidentiality of the voting process, once a voter is issued a computerised ballot paper, there should not have any identification records to positively identify who cast that vote. As such, only the following data are stored with each computerised ballot paper.

ElectionMasterCode ->

- Election Code table 6: compPaper This is BallotPreferences in the answers.
- Electoral Division ElectorateName
- Preferences cast (i.e which candidate got the first preference, who got the second preference, etc) <-- Should there be NEW TABLE?

However, there must be a mechanism in place to record the issuance of a ballot paper to a voter. The issuance record must capture the following information:

ElectionMasterCode -> Table 7: issuanceRecord

- Election code
- Electoral Division ElectorateName
- Polling Station Name
- Identifying details of the voter – these data should be sufficient to uniquely refer to a voter in the electoral role VoterID
- Timestamp

7. Election results Table

At the end of counting process, for each electoral division, the following result data are stored.

ElectionMasterCode ->

- Election Code
- Electoral Division electionDivisionID
- Primary vote for each candidate (i.e. first preferences) <- First round of preferential voting

- Preferential vote count for each candidate, at the each iteration of elimination process (refer to page 32 of https://www.aec.gov.au/about_aec/Publications/electoral_pocketbook/2019/2019-electoral-pocketbook.pdf)

Assignment Tasks

Milestone 1:

You are required to build the data model for this application using an Entity-Relationship Diagram. You must use UML notation for your ER diagram, however, you are free to use any diagramming tool. This diagram should be of professional quality with sufficient details that any database personnel should be able to comprehend.

Then, convert your data model into the physical database design (database schema) and finally generate the DDL script to build the back-end database schema for the application. Ensure that primary keys and foreign keys are correctly identified.

In order to complete this milestone, you are required to demonstrate your workings and the final DDL script to your tutor during Week 4 lab sessions. It is very important to get it marked off by the tutor before you proceed to Milestone 2. If your design does not meet the business requirements, your tutor will allow you to make any amendments (only once) and present again for evaluation.

There are no Canvas submissions for this milestone.

Milestone 2:

TASK 1: Identify **three** tables, in your final schema, that are expected to be extremely large and are expected to grow over time. Tables to use: electionResults, preferentialVoting, ballot

For each of these tables: Week 3 lab + Lecture calculation

- Describe the expected record size (in megabytes), the estimated initial table size, and estimated table size after 10 years of use. With a 3-year election cycle, it is expected that 4 elections are to be held during this period (e.g. 2025, 2028, 2031, 2034)

TASK 2: Write SQL queries for the following tasks:

1. Assume that the total number of voters column (in the Electoral Division table) is empty. This data is to be computed by aggregating data from Voter Registry. Write an SQL query to display the total number of voters registered in each of the electoral division. Your query should produce a report consisting of the electoral division name and total number of voters only. The result should be displayed in the descending order of the total number of voters. A sample result is shown below: Aggregtion: GROUP BY, HAVING

Division	Electors on 20	Ce
MACARTHUR	133,501	
PATERSON	132,123	
MAYO	130,767	
ADELAIDE	130,151	
COWPER	129,887	
SPENCE	129,243	
STURT	129,151	
LONGMAN	129,110	
HINDMARSH	128,814	
BOOTHBY	128,766	

2. The names of candidates appear on the ballot paper in a totally randomized order, i.e. no political party or a group gets any advantage of having their candidates on the top of the list, or candidates with names starting A always appear at the top, etc. Write an SQL query to produce candidate lists for all electoral divisions for the 2022 federal election (election event id: 20220521). The result set must be sorted by electoral division name, and then candidates within each electoral division must be randomized. A sample result is shown below:

Electorate	Candidate Name	Political Party
Adelaide	GRANTHAM, Amy	Liberal
Adelaide	GEORGANAS, Steve	Australian Labor Party
Adelaide	ALLWOOD, Sean	United Australia Party
Adelaide	McMILLAN, Matthew	FUSION: Science, Pirate, Secular, Climate Emergency
Adelaide	GALDIES, Rebecca	The Greens
Adelaide	GERHARD, Faith	Australian Federation Party
Adelaide	ALLWOOD, Gayle	Pauline Hanson's One Nation
Aston	SPELMAN, Rebekah Jane	United Australia Party
Aston	TUDGE, Alan	Liberal
Aston	COOKSON, Asher	The Greens
Aston	BRUCE, Ryan	TNL
Aston	IBBOTSON, Craig	Pauline Hanson's One Nation
Aston	ROCHE, Liam	Liberal Democrats
Aston	DOYLE, Mary	Australian Labor Party
Ballarat	BARNES, John	The Greens
Ballarat	GREEN, Ben	Liberal
Ballarat	GRAHAM, Alex	Independent
Ballarat	SEDGMAN, Kerryn	Australian Federation Party
Ballarat	PRYSE-SMITH, Terri Elizabet	United Australia Party
Ballarat	TAXIS, Rosalie	Pauline Hanson's One Nation
Ballarat	McGRATH, Julia	Liberal Democrats
Ballarat	KING, Catherine	Australian Labor Party
...	...	...

3. Registered voters who do not vote at an election receive a penalty (typically, a fine). Write an SQL query, using IN or NOT IN clause, to generate a report that lists the names and addresses of registered voters who did not vote in 2022 general election (election event id: 20220521) and also not voted in 2019 general election (election event id: 20190518).

Refer to Week 1 lab: HARD QUESTIONS (exists questions)

For each of the queries:

- Produce the SQL query, but do not include the results set.
- Identify what indexes would help. Identify the type of index and columns that are used to build these indexes (justify your design).
- Show the SQL commands for building these indexes in SQ Server.
- Show the query execution plans both before the index is added and after adding the index.

. Write index plans. Show how SQL query works.

- Explain how the index was utilised (or not) and why. What join algorithms were used? What changes would you need to make for the index to be properly utilised, or for a different join algorithm to be used instead? (Provide concrete details of the changes).

<- Look at QUERY PLAN + refer to Week 3 lecture v2

TASK 3: Describe a suitable partition strategy for the three extremely large tables you identified in step 1. Include details of the partitioning type and which columns/partition key to be used. You must include the SQL DDL statements used to implement your partition strategy.

Week 4

You must justify your design decisions. Include details about which of the above queries it will improve the performance of and how it helps with concrete examples. (You must explain in clear terms – such as partition pruning, partition joins, and parallel SQL, applicable to each of these queries.)

Stored procedure to do computation.

TASK 4: Before a voter is allowed to vote, to ensure the integrity of the election system, the system should check if he/she had voted earlier on this election. Write a stored function – `previouslyVoted()`, to check if the voter had voted before.

This function reads the election code, electoral division, voter identification as inputs and returns a Boolean value (true, if voted before and false, if not voted before).

https://www.w3schools.com/sql/sql_stored_procedures.asp

Manually counting process. Table: VoteCoutnRecord

TASK 5: Write a T-SQL stored procedure – `primaryVoteCount()`, to implement the first preferences counting stage of the counting process. This stored procedure requires election code and electoral division name as inputs. It will extract relevant Computerised Ballot Papers for the chosen election and electoral division and will count the first preferences received by each candidate. At the end of the counting process, it will update Election Results table with primary votes (first preferences) received by each candidate.

BONUS TASK : Write a T-SQL stored procedure – `distributePreferences()`, to implement the preference redistribution stage of the counting process. This stored procedure requires election code and electoral division name as inputs.

It will extract relevant Computerised Ballot Papers for the chosen election and electoral division and will go through the redistribution passes until a winner is chosen. At the end of each redistribution pass, it will update Election Results table with preference votes received by each candidate at the completion of the pass. The eliminated candidates' tally can be denoted by a 0.

Milestone 3: This is what changed.

You are required to do a demo of your complete application hosted on the school's Microsoft SQL server. These demos will be conducted during the Weeks 8 – 9). You will be provided with same sample data and the tutor will test out the functionality by using these sample data.

Tutors will use a standard testing plan for all submissions and **you will only receive your second milestone marks if you can demonstrate the functionality of your application in this milestone.**

---

3 Submission

Follow the instructions on Canvas to complete your submission for the project for each Milestone.

3.1 Milestone 1 Submission

There are no submission requirements for Milestone 1. You are required to demonstrate your workings to the marker during Week 4.

The marker will check the following elements in your design:

- The conceptual model for the database backend, in the form of an entity – relationship diagram using UML notation. You may use any tool to generate the diagram, Miro Board is one such potential tool;
- The physical database design, in the form of a database schema;
- A SQL DDL script to create tables and other related database objects in SQL Server.

3.2 Milestone 2 Submission

Submit your report containing answers to 5 tasks above, including sql code, and any other associated files in a zip file using the Canvas Assignment page. It is your responsibility to make sure the submission is (1) complete; (2) correctly zipped; (3) clearly labelled files. Please verify that your submission is correctly submitted by downloading what you have submitted to see if the files include the correct contents.

3.3 Milestone 3 Submission

No additional submissions required for this milestone.

3.4 Assessment Declaration

When you submit work electronically, you agree to the RMIT assessment declaration.

3.5 Silence Period

For **Milestone 1**, there is no silence period.

For **Milestones 2 & 3**, a silence period will take effect from 5:00PM Friday 30 August 2024

This means questions about this assignment will be not answered, whether they are asked on Canvas Discussion Forum, by email, or in person. The silence period is in place because staff members are generally unavailable over the weekend. Additionally, to be fair to all students giving presentations, we will not respond to questions about the Milestone 3 presentations during Week 8.

Make sure you ask your questions with plenty of time for them to be answered.

3.6 Late Submissions & Extensions

A penalty of 10% per day is applied to late submissions up to 5 business days, after which you will receive zero marks.

Short extensions may be granted by the course coordinator up to 1 business day *before* the due date in accordance with RMIT Assessment Adjustment process. However, extensions are not guaranteed and require suitable documentation. The course coordinator may refer requests to Special Considerations.

Special Consideration *may result in an equivalent assessment*, which may take the form of a timed assessment assessing the same knowledge and skills of the assignment and are generally granted on an individual basis. For more information refer to the RMIT Special Consideration process.

3.7 Supported software for assessment and grading

Your assignment solution must be implemented in SQL Server and hosted on school's SQL Server database. Markers won't mark any other variations of implementations.

4 Marking Guidelines

4.1 Milestone 1

- Conceptual model using entity-relationship model 3/5
- The database design, in the form of a database schema 2/5

4.2 Milestone 2 and 3

The marks are divided into the following categories:

- Implementation of the database on SQL Server: 2/20
- Storage strategy for large tables and implementation: 3/20
- Sample queries: 3/20
- Partition strategy for large tables and implementation: 3/20
- previouslyVoted() stored function: 2/20
- primaryVoteCount() stored procedure: 4/20
- BONUS MARKS: distributePreferences() stored procedure: 5 extra marks
- Demonstrate the functionality, presentation skills and answering questions: 3/20

The detailed breakdown is provided on the marking Rubric available on Canvas.

5 Academic Integrity and Plagiarism (Standard Warning)

Academic integrity is about the honest presentation of your academic work. It means acknowledging the work of others while developing your own insights, knowledge and ideas. You should take extreme care that you have:

- Acknowledged words, data, diagrams, models, frameworks and/or ideas of others you have quoted (i.e., directly copied), summarised, paraphrased, discussed or mentioned in your assessment through the appropriate referencing methods

- Provided a reference list of the publication details so your reader can locate the source if necessary. This includes material taken from Internet sites. If you do not acknowledge the sources of your material, you may be accused of plagiarism because you have passed off the work and ideas of another person without appropriate referencing, as if they were your own.

RMIT University treats plagiarism as a very serious offence constituting misconduct. Plagiarism covers a variety of inappropriate behaviours, including:

- Failure to properly document a source
- Copyright material from the internet or databases
- Collusion between students

For further information on our policies and procedures, please refer to the RMIT Academic Integrity Website.

The penalty for plagiarised assignments includes zero marks for that assignment, or failure for this course. Please keep in mind that RMIT University uses plagiarism detection software.