**Database Applications – 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.*

The 3 tables I expect to be the largest in the database and grow over time are: voterRegistry, ballotPreferences & issuanceRecord.

This is because all of them will contain over a million records (based on the Australian population who are eligible to vote) which as more people will be eligible to vote.

*Why I didn’t select the other tables*

**electionMaster**: Each row will record information for each election will be every few years which will not be a lot of information.

**electoralDivision**: There are 152 divisions which can increase over time. However in comparison to millions of records, that is not large.

**electoralDivisionHistory**: This is a associative table between electoralDivision and electionMaster, mainly containing the primary keys of each and the number of voters that year.

**electionEvent**: This contains detailed information of the election that would be currently be run. Not a lot of data.

**Ballot:** This will have a millions of records reflecting the amount of voters. However the selected tables have more attributes in each records so greater total size needed.

**candidate**: Parties can have various candidates per electoral division. Not large enough in comparison to millions of voters, voters and

**contests**:

**politicalParty**: There is a limited number of parties in Australia. While there are various independents, that still won’t make a large amount of data.

**prefCountRecord**: Not large enough compared to millions of records.

**preferenceTallyPerRoundPerCandidate**: Not large enough compared to millions of records.

Calculations

**voterRegistry**

1. Estimated Record Size

We'll first estimate the size of each column in the voterRegistry table.

**Formula:** Estimated Record Size = Size of column 1 + Size of column 2…

We’ll estimate the size of each column in bytes:

Bytes in a integer: <https://docs.oracle.com/cd/E19253-01/817-6223/chp-typeopexpr-2/index.html>

Bytes in varchar: <https://dev.mysql.com/doc/refman/8.4/en/char.html#:~:text=Values%20in%20VARCHAR%20columns%20are,and%20the%20character%20set%20used>.

General bytes: <https://www.w3schools.com/sql/sql_datatypes.asp>

Detailed documentation: <https://learn.microsoft.com/en-us/sql/t-sql/data-types/data-types-transact-sql?view=sql-server-ver16>

**VARCHAR(N)**: N bytes (where N is the maximum number of characters)

**INTEGER**: 4 bytes

**DATE**: 3 bytes (in a packed representation)

**title (VARCHAR)**: 60 characters maximum  
Each character in a VARCHAR field typically takes 1 byte. So, title = 60 bytes.

**voterID (INTEGER)**: 4 bytes  
An INTEGER is usually 4 bytes.

**firstName (VARCHAR)**: 60 characters maximum  
firstName = 60 bytes.

**middleName (VARCHAR)**: 60 characters maximum  
middleName = 60 bytes.

**lastName (VARCHAR)**: 60 characters maximum  
lastName = 60 bytes.

**gender (VARCHAR)**: 60 characters maximum  
gender = 60 bytes.

**dateOfBirth (DATE)**: 3 bytes  
A DATE field typically takes 3 bytes.

**residentUnitNumber (VARCHAR)**: 20 characters maximum  
residentUnitNumber = 20 bytes.

**residentStreetNumber (INTEGER)**: 4 bytes  
residentStreetNumber = 4 bytes.

**residentStreetName (VARCHAR)**: 60 characters maximum  
residentStreetName = 60 bytes.

**residentSuburb (VARCHAR)**: 30 characters maximum  
residentSuburb = 30 bytes.

**residentPostcode (INTEGER)**: 4 bytes  
residentPostcode = 4 bytes.

**residentState (VARCHAR)**: 30 characters maximum  
residentState = 30 bytes.

**postalUnitNumber (INTEGER)**: 4 bytes  
postalUnitNumber = 4 bytes.

**postalStreetNumber (INTEGER)**: 4 bytes  
postalStreetNumber = 4 bytes.

**postalStreetName (VARCHAR)**: 60 characters maximum  
postalStreetName = 60 bytes.

**postalSuburb (VARCHAR)**: 60 characters maximum  
postalSuburb = 60 bytes.

**postalPostcode (INTEGER)**: 4 bytes  
postalPostcode = 4 bytes.

**postalState (VARCHAR)**: 60 characters maximum  
postalState = 60 bytes.

**daytimePhone (INTEGER)**: 4 bytes  
daytimePhone = 4 bytes.

**mobile (VARCHAR)**: 20 characters maximum  
mobile = 20 bytes.

**emailAddress (VARCHAR)**: 40 characters maximum  
emailAddress = 40 bytes.

**divisionName (VARCHAR)**: 50 characters maximum  
divisionName = 50 bytes.

Sum up the sizes of all columns to get the total size of one record:

Total Record Size=60+4+60+60+60+60+3+20+4+60+30+4+30+4+4+60+60+4+60+4+20+40+50 = 780 bytes

Convert this to megabytes (MB).

**Conversion Steps**

1. **Bytes to Kilobytes (KB)**:
   * 1 KB = 1024 bytes
   * To convert bytes to KB:
   * Record Size in KB= (780 bytes /1024) ≈ 0.7617 KB
2. **Kilobytes to Megabytes (MB)**:
   * 1 MB = 1024 KB
   * To convert KB to MB:
   * Record Size in MB= (0.7617 KB / 1024) ≈ 0.000744 MB

**2. Initial Table Size**

Given 17,259,000 voters (number of expected records):

Initial Table Size=17,259,000×0.000744 MB≈12,836.6 MB

So, the initial table size is approximately **12,836.6 MB**.

**3. Future Table Size After 10 Years**

With a population growth of 1.2% per year, over 10 years growth.

*Growth Factor= (1+ Annual Growth Rate) ^Number of Years*

* **Annual Growth Rate**: 1.2% or 0.012
* **Number of Years**: 10 years

Now, plug these into the formula:

= (1+0.012)^10

= 1.012^10 ≈1.126

Apply the growth factor of 1.126 to the initial table size:

Future Table Size= 12,836.6 MB× 1.126≈**14,455.2 MB**

**issuanceRecord**

**Step 1: Determine the Size of Each Column**

* **voterID (INTEGER)**: 4 bytes  
  An INTEGER is usually 4 bytes.
* **electionEventID (VARCHAR)**: Assuming a maximum of 60 characters  
  Each character in a VARCHAR field typically takes 1 byte, so electionEventID = 60 bytes.
* **issueDate (DATE)**: 3 bytes  
  A DATE field typically takes 3 bytes.
* **ballotIssue (Timestamp)**: 8 bytes  
  A Timestamp field generally takes 8 bytes.
* **pollingStation (VARCHAR)**: 50 characters maximum  
  pollingStation = 50 bytes.

**Step 2: Calculate the Total Record Size**

Sum up the sizes of all columns to get the total size of one record:

Total Record Size=4+60+3+8+50=125 bytes

**Step 3: Convert the Record Size to Megabytes (MB)**

**Conversion Steps:**

1. **Bytes to Kilobytes (KB)**:
   * 1 KB = 1024 bytes
   * To convert bytes to KB: Record Size in KB =(125 bytes / 1024) ≈ 0.1221 KB
2. **Kilobytes to Megabytes (MB)**:
   * 1 MB = 1024 KB
   * To convert KB to MB:
   * Record Size in MB= (0.1221 KB /1024) ≈ 0.000119 MB

So, each record in the issuanceRecord table is approximately **0.000119 MB**.

**Step 4: Estimate the Initial Table Size**

Using the same number of voters (17,259,000):

Initial Table Size=17,259,000×0.000119 MB≈2,053.5 MB

So, the initial table size is approximately **2,053.5 MB**.

**Step 5: Estimate the Table Size After 10 Years**

Apply the growth factor of 1.126 to the initial table size:

Future Table Size= 2,053.5 MB \* 1.126 ≈ **2,312.9 MB**

**ballotPreferences.**

Step 1: Determine the Size of Each Column

* ballotID (INTEGER): 4 bytes  
  An INTEGER is usually 4 bytes.
* candidateID (VARCHAR(20)): 20 characters  
  Each character in a VARCHAR field typically takes 1 byte, so candidateID = 20 bytes.
* preference (INTEGER): 4 bytes  
  An INTEGER field takes 4 bytes.

Step 2: Calculate the Total Record Size

Sum up the sizes of all columns to get the total size of one record:

Total Record Size=4+20+4=28 bytes

Step 3: Convert the Record Size to Megabytes (MB)

Conversion Steps:

1. Bytes to Kilobytes (KB):
   * 1 KB = 1024 bytes
   * To convert bytes to KB: Record Size in KB=28 bytes1024≈0.0273 KB
2. Kilobytes to Megabytes (MB):
   * 1 MB = 1024 KB
   * To convert KB to MB: Record Size in MB=0.0273 KB1024≈0.0000267 MB

So, each record in the ballotPreferences table is approximately 0.0000267 MB.

Step 4: Estimate the Initial Table Size

Given the number of voters and assuming that **each voter has 10 preferences** (since in preferential voting, voters rank multiple candidates):

Initial Number of Records=17,259,000×10=172,590,000 records

Now calculate the initial table size:

Initial Table Size=172,590,000×0.0000267 MB≈4,607.4 MB

So, the initial table size is approximately 4,607.4 MB.

Step 5: Estimate the Table Size After 10 Years

Given the initial table size of 4,607.4 MB, we'll apply the growth factor directly to this result.

Future Table Size=Initial Table Size \* Growth Factor

With the growth factor of 1.126:

Future Table Size = 4,607.4 MB×1.126 ≈ **5,188.7 MB**

Task 2

Task 3

Task 4

Task 5

Bonus Task