# TASKS

**Task 1:** **Create an Entity Relationship Diagram (ERD)** for a relational database (showing keys, attributes, datatype, named relationships and optionality) to meet the case study requirements for the *new* system. (Appendix 1). You can normalise the document in Appendix 2 to inform your design.

**Task 2: Design and develop a prototype database** based on your ERD from task 1. You can use any tools (i.e., Microsoft SQL Management Studio / My SQL). You should show the SQL for your queries and their results. Ensure that there is enough test data to adequately demonstrate the functionality.

**Task 3: Design a Data Warehouse** to support Business Intelligence (BI) for this company. Based on data from your ERD, design a star schema for a particular functional area of the business (your choice). Describe how the data will be transformed as it is moved into the data warehouse. Explain how the facts / measures could be used to improve the business.

*You should state any assumptions that you make in your design and ensure that they do not contradict the (limited) information in the Case Study.*

APPENDIX 1: CASE STUDY AND REQUIREMENTS

# CASE STUDY

**Southdowns Electric Mountain Bike Rental (SEMBR)**

You are working for SEMBR, an electric bike hire company, similar to this company here: [Cannonball Bikes.](https://www.cannonballbikes.co.uk/)

*Current System:*

There is one rental centre in Eastbourne. They offer daily / weekly / monthly hire for electric bikes which allow people to ride on the South Downs (a

national park). They have personal and business customers. SEMBR stores customer information, and many customers make repeat rentals. They currently stock two types of bike for hire:

* Full suspension (High Season £50 per day / Low Season £40 per day)
* Hardtail (High Season £40 per day / Low Season £30 per day)

Each bike hire is recorded separately (i.e., if someone want 3 bikes for one day, that is three hires). The cost of rental is more in the high season (May-August) than low season (September-April). This is based on the start date. All hirers must be over 18s (for insurance and safety reasons). Customers must produce photo ID to make a booking. Customers return bikes to the Eastbourne centre. Hires are recorded in an online booking diary system, using a spreadsheet template. There have been issues with double-booking / bikes being unavailable due to servicing and repairs etc.

*New System:*

SEMBR have been invited to partner with Walk The Trails (WTT), a company similar to this: [Natural Adventure Company.](https://www.thenaturaladventure.com/tours/england/walking-the-south-downs-way/?utm_term=&utm_campaign=&utm_source=adwords&utm_medium=ppc&hsa_acc=9391358381&hsa_cam=15853739946&hsa_grp=&hsa_ad=&hsa_src=x&hsa_tgt=&hsa_kw=&hsa_mt=&hsa_net=adwords&hsa_ver=3&gclid=CjwKCAjwyY6pBhA9EiwAMzmfweVtgcXPHB6SikNZ5taP6SAJ7vBOHUXgnaNe9rJjTXlzHVUWlkJsGxoC17MQAvD_BwE) WTT currently runs self-guided walking tours of the [South](https://www.nationaltrail.co.uk/en_GB/trails/south-downs-way/)

[Downs Way](https://www.nationaltrail.co.uk/en_GB/trails/south-downs-way/) - a national trail for walking, biking, riding etc. that runs between Winchester and Eastbourne (160km). They wish to add self-guided bike tours to their offer. They will deal with supplying the accommodation bookings, route maps and daily luggage transfers as they do for their existing walking customers. There may be a small discount for bikes hired via WTT.

WTT would like SEMBR to supply electric mountain bikes. The bikes will need to be picked up and dropped off from both ends of the South Downs Way, as customers may travel the route in either direction. SEMBR will open a new rental centre in Winchester. They hope to minimise driving the bikes between locations. Staff will work across both locations as required. SEMBR will also offer support and emergency repairs over the route, fixing and replacing bikes along the trail, so that the tourist(s) can continue with their booked holiday. WTT will provide a holiday ID number for each bike booking (which may be for more than one bike). Other bookings will still be available to individual customers, as per the current system. The range of bikes may be extended e.g. different sized frames.

To achieve this, SEMBR will need to update their information system. A feasibility report has concluded that the company should use a relational database with a web front-end for their new system.

It is your job to write a design report for the database element.

APPENDIX 1: CASE STUDY AND REQUIREMENTS

The new system should support the following requirements:

1. Register customers for first use, including name, address, dob. A photo ID must be human-verified human to check age – this can be completed by WTT or SEMBR staff.
2. Record information for each bike, including type (currently two types, may be more in future), unique frame number (UFN), service history and overall condition.
3. Show location of the bike – currently and on any given future date (based on bookings).
4. Record information about any emergency repairs (WTT bikes only), including whether SEMBR issued a replacement bike. Include the holiday number from WTT.
5. Calculate the total number of bikes on hire for a given period.
6. Calculate the total income from hired bikes for a given period – split this between WTT and other hires.

APPENDIX 2: EXAMPLE DOCUMENT FOR NORMALISATION

CUSTOMER and HIRE record for normalisation

This is the prototype view of a customer record showing their hire history.

It will not contain all required data – you will need to add to what is here in your design (but should normalise just this data to inform your data model). Note that the earlier hires for this customer are via WTT, but the later ones are individual bookings.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CUSTOMER INFORMATION**  CustomerID: *AAA1111*:  Forename: *Jennie* Surname: *Harding* DOB: *13/07/1969*  Tel mobile: *07782632485*  Tel home: *012330522649*  Address: *45 Madeup Street, Brighton*  Postcode: *BN4 5GH* | | |  | **ID VALIDATION**  ID Type: *passport / driving license / other* Date Verified: *04/08/2023*  Verified By (staffID): *GR2r*  Staff Name: *Gill Roller* | | | | **ID SCAN**  A passport with a coin and a passport  Description automatically generated with medium confidence | | |  |
| **UFN** | **BIKE TYPE** | **RENTAL START \*** | **PICKUP LOCATION** | | **DAYS** | **COST PER**  **DAY (£)** | **TOTAL PAID** | **RETURN LOCATION** | **RETURN SIGNOFF** | **SIGNOFF BY** | **WTT REF (OPTIONAL)** |
| F14 | Full suspension | 01/08/2023 | Eastbourne EB12 8RG | | 10 | 50 | 500 | Winchester  WC8 2LR | yes | GR2  Gill Roller | WTT123 |
| F22 | Full suspension | 01/08/2023 | Eastbourne EB12 8RG | | 10 | 50 | 500 | Winchester  WC8 2LR | yes | GR2  Gill Roller | WTT123 |
| H86 | Hardtail | 18/09/2023 | Winchester  WC8 2LR | | 5 | 30 | 150 | Winchester  WC8 2LR | no\*\* | LW8  Lee Wilson |  |
| F22 | Full suspension | 12/10/2023 | Eastbourne EB12 8RG | | 2 | 40 | 80 | Eastbourne EB12 8RG | \*\*\**NULL* | NULL |  |

\*note: varying price for same type of bike between August and October – this will change each year

\*\*note: bike damaged this hire, so it would have had to be repaired (details not shown here)

\*\*\*note: bike not yet returned (hire still in progress)

SEMBR-WTT ELECTRIC BIKE RENTALS DATABASE DESIGN

Task 1: Entity Relationship Diagram (ERD)-Normalisation

A table of information with text

Description automatically generated with medium confidence

A table with a list of items

Description automatically generated with medium confidence

Figure : Normalisation

Some of the Assumptions from normalised attributes (figure1):

1. WTT\_REF is referring to a table of WTT

2. Each booking has a unique identifier, e.g., BookingID.

3. ID Type and Date Verified referring to Validation column with CustomerID as FK

4. With the normalised data and information from the case study, some attributes and tables are added to make Entity-Relationship Diagram.

A black background with white rectangles

Description automatically generated

Figure : Entity Relationship Diagram

Cardinality and Optionality:

1. **A Customer can make zero, one, or more Bookings (One-to-Many):**
   * Optionality: Optional on the "many" side (a customer might not have any bookings).
2. **A Booking is associated with one Bike, and a Bike can be associated with multiple Bookings (Many-to-One):**
   * Optionality: Mandatory on the "many" side (each booking must be associated with a bike).
3. **Each Bike is associated with a specific Location (Many-to-One):**
   * Optionality: Mandatory on the "many" side (each bike must be associated with a location).
4. **Staff work at a specific Location (Many-to-One):**
   * Optionality: Mandatory on the "many" side (each staff member must work at a location).
5. **EmergencyRepairs are associated with a Booking (Many-to-One):**
   * Optionality: Mandatory on the "many" side (each emergency repair must be associated with a booking).
6. **A Booking can be associated with multiple EmergencyRepairs (One-to-Many):**
   * Optionality: Optional on the "many" side (a booking might not have any emergency repairs).
7. **One customer can have many ID validations. (One-to-Many):**
   * Optionality: Mandatory on the "many" side (a customer must have ID validations).
8. **An emergency repair can have zero, one, or more replaced bikes (One to Many):**
   * Optionality: Optional on the "many" side (an emergency repair might not have any replaced bikes).
9. **One Bike can have zero, one, or more service histories (One to Many):**
   * Optionality: Optional on the "many" side (a bike might not have any service histories).
10. **One location (pickup/return) can deal with many bookings (One to Many):**
    * Optionality: Mandatory on the "many" side (each location must deal with bookings).
11. **One Staff can handle many validations (One to Many):**
    * Optionality: Mandatory on the "many" side (each validation must be staff verified).

See Appendix E, Data Types (Figure 1-8), for a detailed breakdown of the attributes mentioned earlier. It dives into specifics like data types, keys, NULL, default settings, and other key traits. (Nationalarchives.gov.uk, 2021)

Task 2: Database Prototype (Based on ERD)

The entities outlined in the database schema (refer to Appendix E: Data Types Figure 1-8) have been designed to align with the specified requirements of the database. Subsequently, these entities have been enriched with illustrative dummy data, as evidenced in Appendix D: Dummy Data Figure 1-8. A comprehensive representation of the database prototype, including primary keys, attributes, and inter-entity relationships, is demonstrated in Appendix C: Prototype.

|  |  |
| --- | --- |
| Query / Requirement (database focus) | MET |
| 1. Register customers for first use, including name, address, dob. A photo ID must be human-verified human to check age – this can be completed by WTT or SEMBR staff. *(See Appendix D. Dummy Figure 3 & 8)* | YES |
| 2. Record information for each bike, including type (currently two types, may be more in future), unique frame number (UFN), service history and overall condition. *(See Appendix D. Dummy Figure 1 & 5)* | YES |
| 3. Show location of the bike – currently and on any given future date (based on bookings). *(See Appendix C. Queries Figure 1)* | PARTIAL |
| 4. Record information about any emergency repairs (WTT bikes only), including whether SEMBR issued a replacement bike. Include the holiday number from WTT. *(See Appendix D. Dummy Figure 4)* | YES |
| 5. Calculate the total number of bikes on hire for a given period. *(See Appendix C. Queries Figure 2)* | YES |
| 6. Calculate the total income from hired bikes for a given period – split this between WTT and other hires. *(See Appendix C. Queries Figure 3)* | YES |

Task 3: Data Warehouse for Business Intelligence

To design a star schema for a Data Warehouse supporting Business Intelligence (BI) (Harding, n.d.) for SEMBR, I will focus on the functional area related to bike rentals. The central fact table will be the "Booking" table, and the dimension tables will include "CustomerInformation",”ID\_Validation”, "Location", "Bike", "Bike\_Service\_History", ”Emergency\_Repairs”, ”Staff”.

Star Schema Design:

* **Fact Table: Booking**
  + **BookingID (Primary Key):** Unique identifier for each rental transaction.
  + **CustomerID (Foreign Key):** References the CustomerInformation dimension.
  + **BikeID (Foreign Key):** References the Bike dimension.
  + **Rental\_Start:** Date when the rental was initiated.
  + **Total\_Paid:** Total earning per booking.
  + **Duration:** Duration of the rental in days.
  + **Pickup\_LocationID (Foreign Key):** References the Location dimension for pickup.
  + **Return\_LocationID (Foreign Key):** References the Location dimension for return.
  + **Return\_SignOff:** Sign-off status for bike return (YES/NO), NULL if still in the process.
  + **SignOffBy (Foreign Key):** References the Staff dimension for sign-off.
  + **WTT\_REF:** Reference number for Walk The Trails bookings.
  + **Discount:** Discount applied to the rental via WTT.
* **Dimension Tables: CustomerInformation Dimension:**
  + **CustomerID (Primary Key):** Unique identifier for each customer.
  + **Forename:** First name of the customer.
  + **Surname:** Last name of the customer.
  + **DOB:** Date of Birth of the customer.
  + **TeleMobile:** Mobile phone number of the customer.
  + **TeleHome:** Home phone number of the customer.
  + **Address1:** First line of the customer's address(street).
  + **Address2:** Second line of the customer's address(city).
  + **PostCode:** Postal code of the customer's address.
* **ID\_Validation Dimension:**
  + **PhotoID (Primary Key):** Unique identifier for each photo ID validation.
  + **VerificationStatus:** Verification status (Verified/Not Verified).
  + **ID\_Scan:** Image of customer ID’s.
  + **Date\_Verified:** Verification date.
  + **CustomerID:** References to Customer dimension table.
  + **StaffID:** References to Staff dimension table.
* **Bike Dimension:**
  + **BikeID (Primary Key):** Unique identifier for each bike.
  + **BikeType:** Type of the bike (Full suspension/Hardtail).
  + **Frame\_Size:** Size of the bike frame (Small/Medium/Large).
  + **Overall\_Condition:** Overall condition of the bike (Good/Excellent/Bad).
  + **Bike\_Status:** Current status of the bike (Booked/Available/In-Service).
  + **LocationID:** References to Location dimension table.
* **Bike\_Service\_History Dimension:**
  + **ServiceID (Primary Key):** Unique identifier for each bike service entry.
  + **ServiceDate:** Date of the bike service.
  + **Description:** Description of the service.
  + **Cost:** Cost of the service.
  + **BikeID:** References to Bike dimension table.
* **EmergencyRepairs Dimension:**
  + **Emergency\_RepairID (Primary Key):** Unique identifier for each emergency repair entry.
  + **Description:** Description of the emergency repair.
  + **ReplacementBike (Foreign Key**): References the Bike dimension for the replacement bike.
  + **Cost:** Cost of the emergency repair.
  + **BookingID:** References Booking fact table.
  + **WTT\_REF:** Reference number for Walk The Trails bookings.
* **Location Dimension:**
  + **LocationID (Primary Key):** Unique identifier for each location.
  + **Location:** Name of the location (Winchester/Eastbourne).
* **Staff Dimension:**
  + **StaffID (Primary Key):** Unique identifier for each staff member.
  + **Forename:** First name of the staff.
  + **Surname:** Last name of the staff.
  + **Company:** Company the staff belongs to (WTT/SEMBR).
  + **LocationID (Foreign Key):** References the Location dimension.

**Data Transformation:**

1. **ETL Process (Extract, Transform, Load)** (Connolly and Begg, 2015):
   * **Extract:** Data will be extracted from the operational database tables (CustomerInformation, Location, Bike, Booking, Bike\_Service\_History, ID\_Validation, Staff, EmergencyRepairs).
   * **Transform:** Data will be transformed to fit the structure of the star schema, including data cleansing, formatting, and the creation of derived attributes (e.g., Total Cost, Total Paid).
   * **Load:** Transformed data will be loaded into the corresponding dimension and fact tables in the data warehouse.
2. **Data Quality and Consistency** (Connolly and Begg, 2015)
   * Implement data validation checks during the ETL process to ensure data quality.
   * Handle missing or inconsistent data by applying default values or cleaning mechanisms.
3. **Indexing:**
   * Create appropriate indexes on foreign keys and other columns to optimize query performance.

**Example:** Let's discuss how data transformation could occur for the *CustomerInformation*dimension based on the attributes provided:

1. **Cleaning and Validation:**
   * Ensure that all required fields, especially the **CustomerID** and **DOB**, are present and not null.
   * Validate that the **DOB** falls within the specified range, in this case, not later than '2005-11-27'.
2. **Standardization:**
   * Standardize phone numbers to a consistent format (e.g., remove special characters, ensure a specific country code format).
   * Standardize the **PostCode** to a consistent format if needed.
3. **Derivation of Additional Attributes:**
   * Create a new attribute for age based on the **DOB** to provide additional insights into customer demographics.
4. **Handling Missing or Default Values:**
   * Set default values for optional fields like **TeleMobile**, **TeleHome**, **Address1**, **Address2** if they are missing.
5. **Data Type Conversion:**
   * Ensure that data types are consistent with the target schema. For example, if the original data uses a different date format for **DOB**, convert it to match the required format.
6. **Data Quality Checks:**
   * Implement checks to identify and handle any outliers or anomalies in the data

**BI Use Cases (facts/measures):**

1. **Rental Analysis:**
   * Analyse rental patterns by location, bike type, and time to identify peak seasons and popular bike choices.
   * Evaluate the effectiveness of marketing strategies and discounts.
2. **Customer Behaviour Analysis:**
   * Understand customer preferences, repeat rental behaviour, and demographics.
   * Identify and target specific customer segments for promotions.
3. **Location Performance Analysis:**
   * Evaluate the performance of rental centres in Winchester and Eastbourne.
   * Optimize bike allocation and servicing based on location demand.
4. **Booking Sign-off Analysis:**
   * Monitor and analyse the efficiency of bike return sign-off processes.
   * Identify any patterns related to late returns or damaged bikes.
5. **WTT Partnership Impact:**
   * Analyse the impact of the partnership with Walk The Trails on rental volumes and revenue.
   * Evaluate the effectiveness of any discounts offered through WTT.

Implementing this star schema will provide a solid foundation for BI, enabling SEMBR to gain valuable insights into their bike rental operations and make informed business decisions.

References

1. Connolly, T.M. and Begg, C. (2015). *Database systems: a practical approach to design, implementation and management*. Harlow: Pearson Education Limited.

2. Harding, J. (n.d.). *Data Management.* [Lecture] Data Modelling, SQL Review - Database Design Testing, Business Intelligence, Data Security / Legal and Ethical Considerations.

3. Nationalarchives.gov.uk. (2021). Available at: [https://webarchive.nationalarchives.gov.uk/ukgwa/+/http://www.cabinetoffice.gov.uk/media/254290/GDS%20Catalogue%20Vol%](https://webarchive.nationalarchives.gov.uk/ukgwa/+/http://www.cabinetoffice.gov.uk/media/254290/GDS%20Catalogue%20Vol%25)

4. GDPR (2018). *General Data Protection Regulation (GDPR)*. [online] General Data Protection Regulation (GDPR). Available at: <https://gdpr-info.eu/>. (GDPR, 2018)

5. Banerjee, D. (2022). *What is SQL ENUM?* [online] Scaler Topics. Available at: <https://www.scaler.com/topics/enum-sql/>. (Banerjee, 2022)

6. Atlassian (n.d.). *Auto Increment Primary Key in SQL Server*. [online] Atlassian. Available at: https://www.atlassian.com/data/admin/how-to-define-an-auto-increment-primary-key-in-sql-server#:~:text=CREATE%20TABLE%20books%20(%20id%20INT%20NOT%20NULL%20IDENTITY%20PRIMARY%20KEY [Accessed 14 Dec. 2023]. (Atlassian, n.d.)

APPENDICES

Appendix B. Data Types

*Figure 1: Bike*

A screenshot of a computer code

Description automatically generated

*Figure 2: Booking*

*A computer code with text

Description automatically generated*

*Figure 3: Emergency Repairs*

*A screenshot of a computer code

Description automatically generated*

*Figure 4: Customer*

*A screenshot of a computer code

Description automatically generated*

*Figure 5: ID Validation*

A screenshot of a computer

Description automatically generated

*Figure 6: Location*

*A close up of text

Description automatically generated*

*Figure 7: Bike Service History*

*A screenshot of a computer

Description automatically generated*

*Figure 8: Staff*

*A screenshot of a computer

Description automatically generated*

Appendix C. Queries

Note that, because all data are artificial, some query results might seem illogical.

*Figure 1: Current and future Location*

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

***Figure 2: Total number of bikes on hire***

A screenshot of a computer program

Description automatically generated

***Figure 3: SEMBR and WTT income***

A screenshot of a computer

Description automatically generated

Appendix C. prototype



Appendix D. Dummy

Note that, all entities were loaded with 10 rows.

*Figure 1: Bike*

A screenshot of a computer

Description automatically generated

*Figure 2: Booking*

*A screenshot of a computer

Description automatically generated*

*Figure 3: Customer*

*A screenshot of a computer

Description automatically generated*

*Figure 4: Emergency Repairs*

*A screenshot of a computer program

Description automatically generated*

*Figure 5: Bike Service History*

A screenshot of a computer

Description automatically generated

*Figure 6: Location*

*A screenshot of a computer

Description automatically generated*

*Figure 7: Staff*

*A screenshot of a computer

Description automatically generated*

*Figure 8: ID Validation*

*A screenshot of a computer

Description automatically generated*