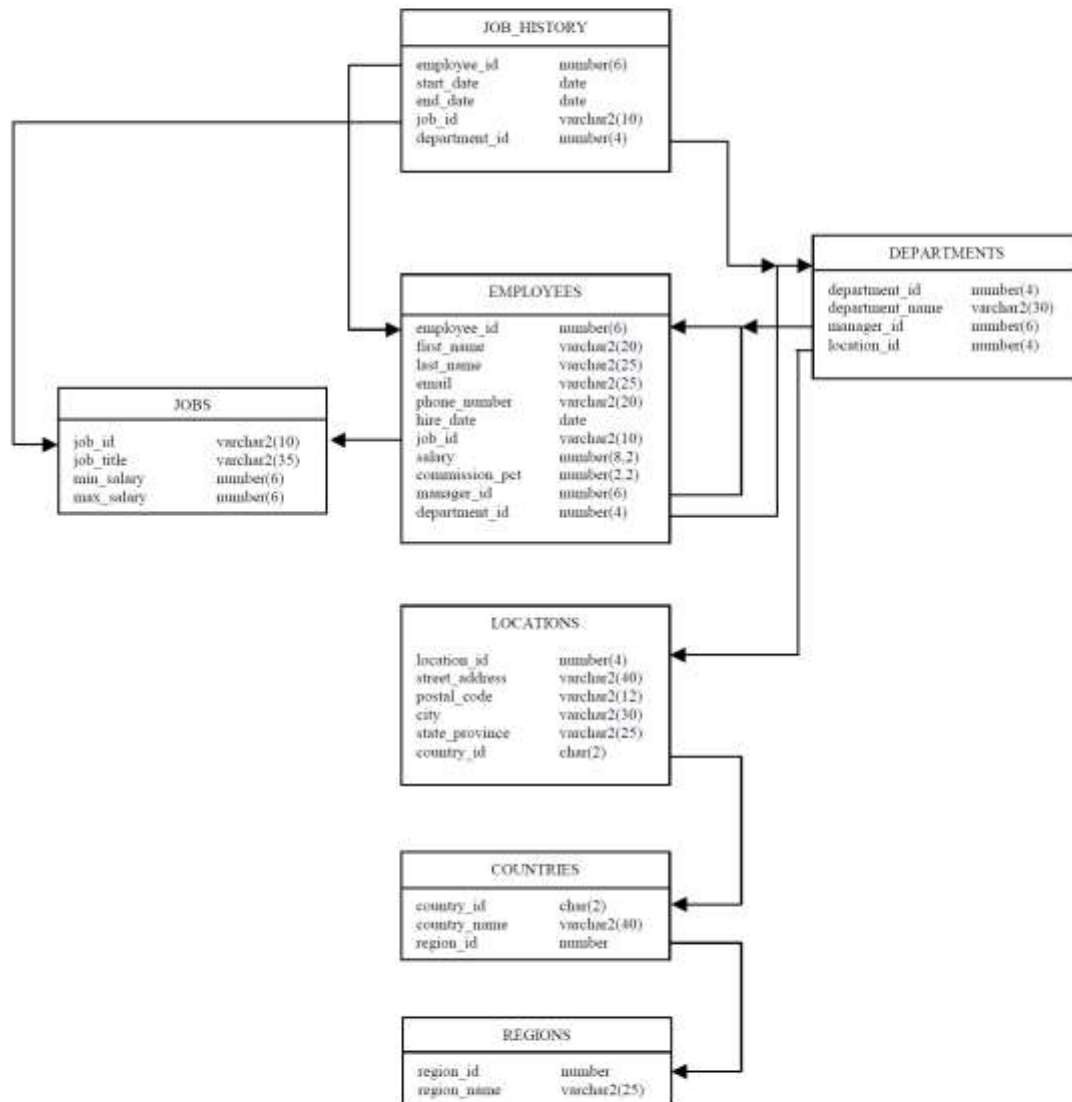


Data Masker for Oracle

Oracle Sample HR Schema - Entity Relationship Diagram

An entity relationship diagram of the tables in the HR sample schema supplied with Oracle version 9i (and above).



HR DATASET

DATA MODELING EXERCISE – STEP BY STEP FORMAT

Scenario Overview (Given to Student)

An organization wants to build an **HR Analytics Dashboard** to analyze:

- Employee strength
- Attrition trends
- Department performance
- Salary distribution
- Experience and tenure analysis

You are provided with raw HR data in Excel/CSV format.

STEP 1: Understand the Dataset (Basic)

Sample Columns Provided:

- EmployeeID
- EmployeeName
- Gender
- Age
- Department
- JobRole
- DateOfJoining
- DateOfExit
- EmploymentStatus
- Salary
- ExperienceYears
- Location
- PerformanceRating

Exercise Questions:

1. Identify **measures** and **attributes** from the dataset.
2. Which columns are **descriptive** and which are **numeric**?
3. Which column uniquely identifies an employee?

◆ STEP 2: Define Business Objectives (Basic)

Business Questions:

- How many employees are currently active?
- Which department has the highest attrition?
- What is the average salary by job role?
- How does experience affect attrition?

Exercise Questions:

1. Convert the above business questions into **analytical requirements**.
2. Identify which questions require **aggregation**.

◆ STEP 3: Decide the Fact Table (Basic → Intermediate)

Exercise Questions:

1. What should be the **Fact Table** in this HR model?
2. What will be the **grain** of the fact table?
 - One row per employee?
 - One row per employee per month?
3. Justify your selected granularity.

◆ STEP 4: Identify Dimension Tables (Intermediate)

Exercise Task:

From the given dataset, identify and design the following dimensions:

1. **Employee Dimension**
2. **Department Dimension**
3. **Job Role Dimension**

4. Date Dimension

5. Location Dimension

Questions:

- What will be the **primary key** of each dimension?
- Which columns belong to each dimension?

◆ STEP 5: Design Star Schema (Intermediate)

Exercise Questions:

1. Draw a **Star Schema** for the HR dataset.
2. Place the **Fact table at the center**.
3. Connect all dimension tables using **foreign keys**.
4. Explain why Star Schema is suitable for HR analytics.

◆ STEP 6: Measures & KPIs Identification (Intermediate)

Required KPIs:

- Total Employees
- Active Employees
- Attrition Count
- Attrition Rate
- Average Salary
- Average Experience

Exercise Questions:

1. Identify which KPIs are **additive**.
2. Which KPIs require **calculated measures**?
3. Where should these measures reside (fact table or BI layer)?

◆ STEP 7: Date Dimension Modeling (Intermediate)

Exercise Task:

Design a **Date Dimension** using:

- DateKey
- Date
- Month
- Quarter
- Year
- Financial Year

Questions:

1. Why should Date be a **separate dimension**?
2. Which date should link to the fact table:
 - DateOfJoining
 - DateOfExit
 - Both?

◆ **STEP 8: Handling Attrition Logic (Advanced)**

Scenario:

Some employees have no DateOfExit (still working).

Exercise Questions:

1. How will you identify **currently active employees**?
2. How will attrition be calculated correctly?
3. Will you store EmploymentStatus in fact or dimension? Why?

◆ **STEP 9: Slowly Changing Dimension (SCD) (Advanced)**

Scenario:

Employees may change:

- Department
- Job Role
- Location

Exercise Questions:

1. Which SCD type is suitable (Type 1 / Type 2)?
2. How will historical data be preserved?
3. What additional columns are required in dimension tables?

◆ STEP 10: Many-to-One & One-to-Many Relationships (Advanced)

Exercise Questions:

1. One Department → Many Employees
Explain the relationship.
2. One Job Role → Many Employees
Where should the foreign key exist?
3. Why should **bi-directional relationships be avoided** in BI tools?

◆ STEP 11: Performance Optimization (Advanced)

Scenario:

Dashboard is slow due to large HR data.

Exercise Questions:

1. What modeling changes will improve performance?
2. Should unnecessary columns be removed?
3. Why is **cardinality** important in HR data models?

◆ STEP 12: Final Validation Questions (Interview Level)

1. Why is **EmployeeID** a better key than EmployeeName?
2. Can salary ever be a dimension? Why or why not?
3. Why should calculated columns be minimized?
4. How does a good data model improve DAX performance?

Practice Outcome

- Explain modeling decisions in **business language**