

# Projektarbeit

Du hast eine Anstellung als Data Engineer (-in) beim Mischkonzern ‚Müller AG‘ gefunden und bist in der Abteilung externes IT-Consulting angestellt. Als erstes Projekt wird dir ein Auftrag von Extern zugewiesen. Es geht um den Zoo in Pirmasens. Der Zoo besteht seit 36 Jahren. Im Zuge einer umfassenden Modernisierung soll der Betrieb des Zoos zukünftig u.a. it-gestützt ablaufen. Ein integriertes IT-System gibt es jedoch noch nicht. Und hier kommst du bzgl. Themata aus dem Bereich des Data Engineering ins Spiel.

Du sollst nach Rücksprache mit dem Geschäftsführer des Zoos (Ansprechpartner: Herr Teske) die folgenden Schritte durchführen:

1. Erstelle ein vollständiges Datenmodell (ERM) der zukünftigen operativen Datenbank. M:N Beziehungen sind aufgelöst darzustellen. Die Primärschlüssel, Fremdschlüssel und Kardinalitäten (1:1, 1:N, N:1, ...) sind ins Datenmodell einzutragen. Die Relationen sind zu beschriften (Notation wie bei Lösung zur Aufgabe Baustoffgroßhandlung).
2. Setze das Datenmodell aus 1. in eine relationale Datenbank um inklusive Dokumentation (Data Dictionary). Lese die Referenzdaten (siehe Datei ‚Wertebereiche.txt‘) vom Kunden in die Datenbank ein.
3. Die Geschäftsleitung des Pirmasenser Zoos interessiert sich ebenfalls für Business Intelligence. Konkret geht es um ein Data Warehouse. Welches Data Warehouse Schema empfehlst Du hier? Warum? Erstelle ein entsprechendes Datenmodell.
4. Eine hohe Datenqualität ist in diesem Vorhaben sehr wichtig. Erstelle ein Konzept, wie der Datenbestand fortlaufend bearbeitet werden kann und die Datenqualität auf ein >97% Level kommt.
5. Für die Projektergebnisse ist eine Präsentation zu erstellen und beim Kunden zu präsentieren.

**Anmerkung:** zur Umsetzung der oben genannten Punkte sind Gespräche beim Kunden (Herr Teske) zwingend notwendig. Im Rahmen der Anforderungserhebung muss kein Lastenheft erstellt werden. Die Gespräche sind jedoch zu dokumentieren.

**Zusatzpunkte:** Es können Zusatzpunkte erreicht werden, indem das Datenmodell aus 3 (Datawarehouse) ebenfalls in eine relationale Datenbank inklusive Dokumentation (Data Dictionary) umgesetzt werden.

## Organisatorisches

1. Alle der oben genannten 5 Teilaufgaben zählen gleich viel (20 %).  
Der Teil 5 (Präsentation) ist aufgeteilt: hier zählt die abgegebene Präsentation 10% und dein mündlicher Vortrag ebenfalls 10%. Der mündliche Vortrag kann entweder in deutscher oder englischer Sprache gehalten werden.
2. Späteste Abgabe ist Freitag der 02.05.2025 um 10:00 Uhr.
3. Jeder Teilnehmer muss einzeln (also für sich selber) in seinen Projektabgabeordner bis Freitag den 02.05.2025 um 10:00 Uhr hochgeladen haben!
4. Abzugeben bzw. in den Projektabgabeordner hochzuladen sind die folgenden 6 (7) Dateien:
  - 4.1. Datenmodell (ERM) (jpg oder .png Format) der operativen Datenbank,
  - 4.2. relationale Datenbank (.mdb, .accdb oder .db Format),
  - 4.3. Datenbankdokumentation / Data Dictionary (.pdf oder .xls Format),
  - 4.4. Datenmodell Data Warehouse (.pdf oder .png Format),
  - 4.5. Dataquality Konzept (.doc oder .pdf Format),
  - 4.6. Präsentation (.pdf oder .ppt Format).

Im Falle von Data Vault als Datenmodell Data Warehouse ist zusätzlich mindestens eine Star Schema abzugeben.
5. Es muss klar sein, welche Person welchen Abschnitt entwickelt bzw. erarbeitet hat.
6. Am Freitag 02.05.2025 finden dann die Präsentationen statt. Jeder Teilnehmer der jeweiligen Gruppe präsentiert seinen Teil.
7. Die Anwendung bzw. der Einsatz von KI Tools o.ä. zur Erstellung der oben genannten Punkte (siehe 4) sind nicht erlaubt!

## **Projektablauf**

- Legt für Eure Gruppe und euer Projekt einen eigenen Ordner auf dem Netzlaufwerk an.
- Ihr könnt den externen Kunden (Herrn Teske) des Zoo Pirmasens treffen. Die Meetings finden nach Absprache in Gruppe M statt.
- Damit der Kunde sich auf Eure Fragen vorbereiten kann, empfiehlt es sich, die Fragen vorher in euren Ordner in der Dateiablage abzulegen. Darüber hinaus ist es sinnvoll die Antworten und Beschlüsse zu dokumentieren und abzulegen.

## **Zu der Präsentation**

Der Zeitrahmen liegt bei ca. 60 Minuten pro Gruppe bzw. bei ca. 10 – 12 Minuten pro Teilnehmer (je nach Gruppengröße).

Empfohlene Inhalte der Präsentation:

- Aufteilung innerhalb der Gruppe
- Angaben des Kunden, Rahmenbedingungen
- Gespräche mit dem Kunden
- Best Practices aus den Gesprächen mit dem Kunden
- Entity Relationship Modell
- Erstellte Datenbank (Prototyp), Data Dictionary
- Erstelltes Datenmodell Datawarehouse -> warum wurde dieses Schema gewählt?
- Darstellung der Datawarehouse IT-Architektur
- Verwendete Methoden, Techniken, Software, Tools -> Warum diese?
- Vorgehensweise beim Dataquality Konzept
- Inhalt Dataquality Konzept
- Ausblick, was sind die nächsten Schritte im Projekt

## Gruppeneinteilung

Raum	Gruppen	Teilnehmer	Standort
a	1	Elhosh, Ali	Bad Homburg
a	1	Kasem, Suliman	Leipzig
a	1	Rahman, Azrin	Dortmund
a	1	Rosenbauer-Döring, Marijana	Berlin
b	2	Dietzel, Christian	Limburg
b	2	Salazar Rohleder, Axel	Mainz
b	2	Sauer, Carolin	Lübeck
b	2	Sen, Yasin	Mainz
b	2	Sewitz, Sven	Potsdam
c	3	Angrick-Pruner, Gudrun	München
c	3	Djiassi Cheugoue, Antoinette Morgane	Potsdam
c	3	Zander, Kevin	Potsdam
c	3	Zheng, Shuang	Nürnberg
c	3	Cakir, Nurdan	Bochum
d	4	Steindl, Lisa	Leipzig
d	4	Wagner, Alois	München
d	4	Kirschner, Thomas	Hamburg
d	4	Metkar, Shubham (subham)	Pforzheim
d	4	Landsgesell, Florian	Frankfurt