



Projektarbeit Zoo

Data Engineering Modul - Gruppe 1

Gruppenaufteilung

- ▶ Entity Relationship Model - alle zusammen
- ▶ Relationale Datenbank in MS Access - Margarita
- ▶ Data Dictionary - Valerie
- ▶ Datawarehouse Data Vault und Star Schema - Kanimozhi
- ▶ IT-Architektur - Valerie
- ▶ Data Quality Konzept - Thomas



Angaben des Kunden, Rahmenbedingungen

- ▶ Der Zoo besteht seit 36 Jahren
- ▶ Bisher gab es kein IT-System, die Mitarbeiter haben alles handschriftlich in Notizbüchern notiert
- ▶ Es soll ein integriertes IT-System erstellt werden, wofür die Vorarbeit im Bereich Data Engineering nötig ist
- ▶ Es gibt im Zoo:
 - ▶ 70 Mitarbeiter
 - ▶ 50 externe Tierärzte
 - ▶ 120 Lieferanten
 - ▶ 6000 Tiere

Angaben des Kunden, Rahmenbedingungen

- ▶ Die Altdaten sollen per Scans digitalisiert werden, jedoch sind für das IT-System nur die Ist-Daten relevant
- ▶ Die Ist-Daten werden von Datentypisten dem neuen IT-System hinzugefügt
- ▶ Zwei IT-Fachkräfte sollen zukünftig angestellt werden, welche das neue IT-System betreuen werden
- ▶ In der Zukunft sollen Online-Rundgänge möglich sein, sowie Tierpatenschaften angeboten werden

Gespräche mit dem Kunden

- ▶ Angaben zu Lohnzahlungen und Tierarztkostenabrechnung werden dem IT-System in einem späteren Arbeitsschritt hinzugefügt
- ▶ Rundgänge bestehen aus mehreren Strecken, eine Strecke führt von einem Gehege zu einem anderen, Strecken können auch in mehreren Rundgängen vorkommen
- ▶ Der Mehrwertsteuersatz kann zunächst außer acht gelassen werden



Entity Relationship Model

- ▶ Wir haben gemeinsam ein ERM für den Zoo in Pirmasens erstellt
- ▶ Enthaltene (Haupt-)Tabellen:
Futter, Futterarten, Gattungen, Gebäude, Gehege, Gehegeart, Krankengeschichte, Lager, Lieferanten, Mahlzeiten, Mitarbeiter, Pflegearten, Rundwege, Strecken, Tierarten, Tierärzte, Tiere, Unverträglichkeit, Vertretung(Mitarbeiter), Vertretung(Tierarzt)
- ▶ Alle M:N-Beziehungen wurden durch die folgenden Tabellen aufgelöst:
Futter-Lieferanten, Gebäude-Gehege, Mahlzeit-Futter, Mahlzeit-Tiere, Rundwege-Strecken, Tierarten-Mitarbeiter, Tierarzt-Tier, Tiere-Gehege, Tiere-Pflege
- ▶ Referenztabelle, im ERM noch nicht in Beziehung gesetzt:
Anrede, Einheit Futter-/Lagermenge, Einheit Gewicht, Einheit Größe(Gehege), Einheit Größe(Tier), Geschlecht(MA), Geschlecht(Tier), ja/nein, Land, MwSt. Satz in %, PLZ Ort, Titel, Vertretungsgrund, Währung
- ▶ Die 1:N-Beziehungen wurden farblich markiert:
 - ▶ Rot: Linie zwischen Relationship und 1
 - ▶ Blau: Linie zwischen Relationship und N
- ▶ [Link](#)

Datenbank (Prototyp)

- ▶ Die Zoo Datenbank [02_Zoo_Database_Version10.accdb](#) wurde in Microsoft Access entworfen, das dem Kunden folgende Vorteile bietet:
 - mehrbenutzer-fähig
 - sehr gut für Dateneingaben
 - Lösungen sind oft ohne Code möglich
 - Integration mit anderen MS Office Produkten
 - Zugriff auf viele Datenformate, einfacher Import und Export
 - Berichtsgenerator
 - umfangreicher Abfragegenerator

Data Dictionary

- ▶ Data Dictionary soll zum besseren Verständnis der Datenbank beitragen
- ▶ Aufbau:
 - ▶ Oberer Bereich: Die in Beziehung gesetzten Tabellen
 - ▶ Unterer Bereich: Die Referenztabellen



Data Dictionary

- ▶ Spalten:
 - ▶ Tabelle
 - ▶ Spalte
 - ▶ Felddatentyp
 - ▶ Schlüssel: Angabe, ob das Feld ein Primär- oder Fremdschlüssel ist
 - ▶ FK bezieht sich auf: Angaben zur Tabelle und dem Feld auf welches sich der Fremdschlüssel bezieht
 - ▶ Indizierung: Angabe, ob das Feld indiziert wurde
 - ▶ Anforderungen an den Wert: Angaben, ob das Feld leer sein darf, ob es eine Nummerierung gibt oder ob es sich um ein Boolean Feld handelt
 - ▶ Beschreibung: Erläuterung, des Feldinhalts
- ▶ Link

Datenmodell Datawarehouse (Datavault)

► Vorteile von Data Vault

- Einfachheit: Data Vault ist ein einfacher, aber leistungsstarker Ansatz für Data Warehousing, der von Unternehmen jeder Größe leicht implementiert werden kann
- Flexibel: Der modulare Aufbau von Data Vault macht es einfach, Daten nach Bedarf hinzuzufügen oder zu löschen, und ermöglicht es Ihnen, Ihre BI-Lösungen schnell zu iterieren
- Beweglichkeit: Data Vault bietet die nötige Flexibilität, um schnell auf veränderte Geschäftsanforderungen zu reagieren und große Datenmengen in das Data Warehouse zu laden.

Datenmodell Datawarehouse (Datavault)

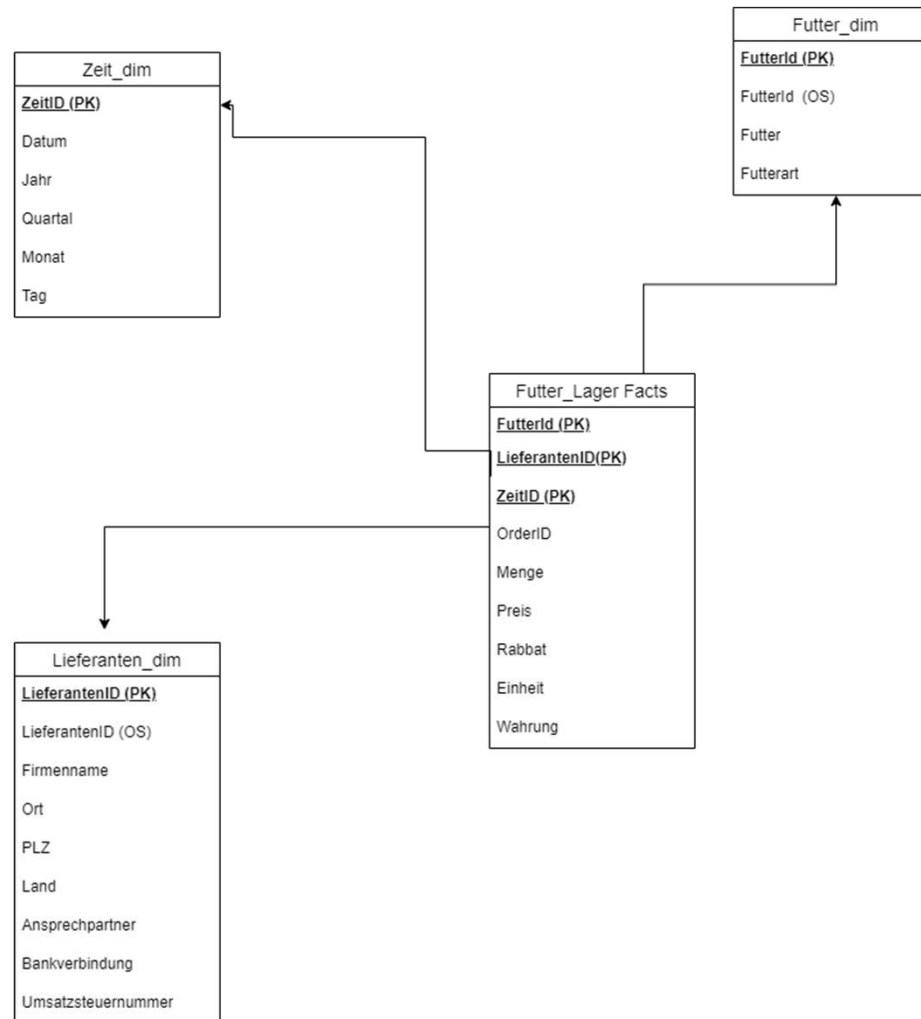


Data Vault_Zoo

- HUB
- Links
- Satellites

Datenmodell Datawarehouse (Star Schema)

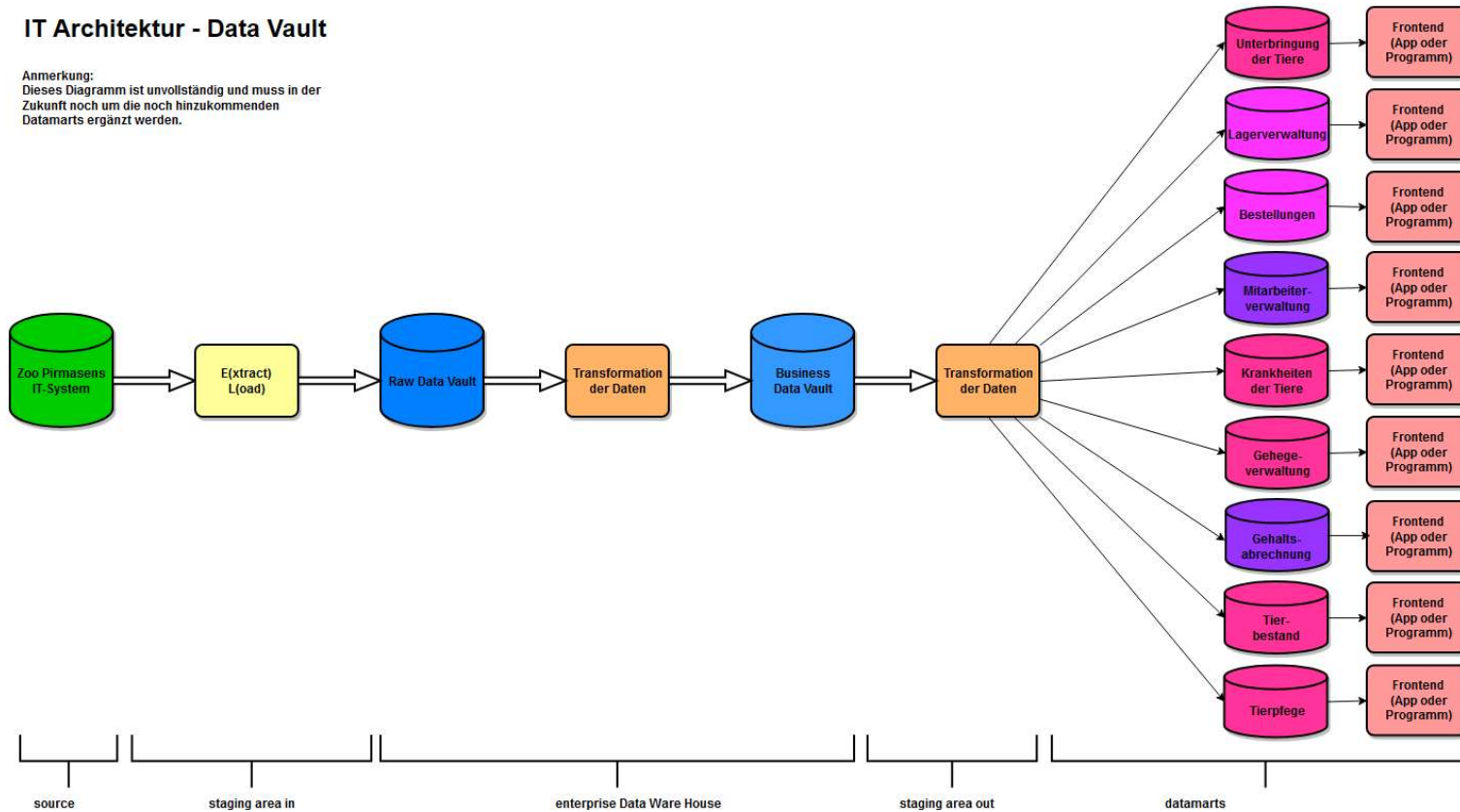
- Faktentabelle- Futter_Lager
- Dimensionstabellen: Lieferanten, Futter and Time.
- Measures : Preis, Rabatt, Währung, Menge, Abgang and Zugang
- Star Schema für die Futterbestellung.
- Datenqualität und Konsistenz
- Historische Analyse
- Skalierbarkeit
- Leistung



Datawarehouse IT Architektur

IT Architektur - Data Vault

Anmerkung:
Dieses Diagramm ist unvollständig und muss in der
Zukunft noch um die noch hinzukommenden
Datamarts ergänzt werden.



Verwendete Methoden, Software, Tools

- ▶ <https://app.diagrams.net/> - Diagrammen (ERM System, IT-Architektur)
- ▶ Microsoft Excel - Dokumentation der Datenbankstrukturen, Data Warehouse Konzept (Data Vault)
- ▶ Microsoft Word - Dataquality Konzept
- ▶ Microsoft Access - Die Erstellung der Datenbank
- ▶ PowerPoint - Präsentation des Projekts
- ▶ Google Drive - Datenaustausch

Data Quality Konzept



Data Quality Konzept

► Ausgangssituation



Data Quality Konzept

► Zielvorstellung



Data Quality Konzept

► Altdatenmigration



Data Quality Konzept

► Das Betriebssystem



Data Quality Konzept

► Der ELT Prozess (Datawarehouse)



Data Quality Konzept

► Fazit



Ausblick

Was sind die nächsten Schritte im Projekt?

