

Projektarbeit

Schütze, Hans-Joachim

Chang, Yinshui

Lvov, Phillip

Nzwessa, Constantin



Übersicht

01_ Aufteilung innerhalb der Gruppe	07 Umsetzung Data Warehouse
02_Angaben des Kunden, Rahmenbedingungen	08 Gesamt IT-Architektur
03_Gespräche mit dem Kunden	09 Software, Tools
04_Best Practices aus den Gesprächen mit dem Kunden	10_ Vorgehensweise beim Dataquality Konzept
05_Entity Relationship Modell	11_Inhalt Dataquality Konzept
06_Erstellte Datenbank (Prototyp), Data Dictionary	12_Ausblick, was sind die nächsten Schritte im Projekt

Aufteilung innerhalb der Gruppe

1. Erstelle ein vollständiges Datenmodell (ERM) der operativen Datenbank. M:N Beziehungen sind aufgelöst darzustellen. Die Primär-, Fremdschlüssel und Kardinalitäten sind ins Datenmodell einzutragen. Die Relationen sind zu beschriften.

2. Setze das Datenmodell aus 1. in eine relationale Datenbank um inklusive Dokumentation (Data Dictionary). Lese die Referenzdaten vom Kunden in die Datenbank ein.

3. Die Geschäftsleitung des Pirmasenser Zoos interessiert sich ebenfalls für Business Intelligence. Konkret geht es um ein Datawarehouse. Welches Datawarehouse Schema empfiehlst Du hier? Warum? Erstelle ein entsprechendes Datenmodell inkl. passender Datawarehouse IT-Architektur.

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. Präsentation erstellen und vorstellen

Angaben des Kunden, Rahmenbedingungen

Grundlagen des Zoo Pirmasens

Im Gespräch mit dem Auftraggeber erhalten wir relevante Daten:

- Derzeit betreut der Zoo Pirmasens rund 6.000 Tiere.
- Derzeit arbeiten **70 Mitarbeiter** in fester Position, z. B. Pfleger, die sich um die Tiere kümmern.
- **50 externe Ärzte** stehen **bei Krankheitsfällen von Tieren** zur Hilfe. Wenn ein Tier meldepflichtig erkrankt ist, muss der Arzt die Meldung anweisen.
- In der jetzigen Situationen Vertretungen ein Problem sein - auch hier kann würde eine digitale Lösung optimieren.
- Im Zoo gibt es 2 hausinterne IT-Kräfte in Festanstellung



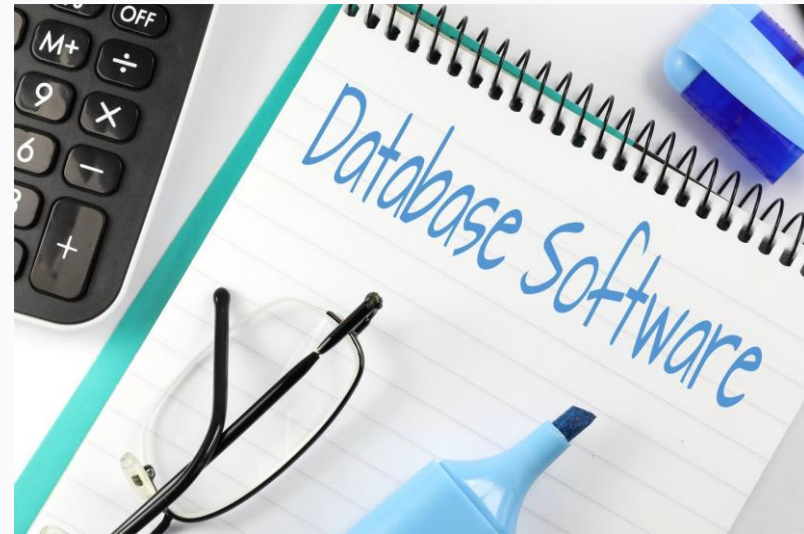
Kirche Hl. Pirminius, Pirmasens (via pixabay)

03_Gespräche mit dem Kunden

1. Termin mit den Auftraggebern

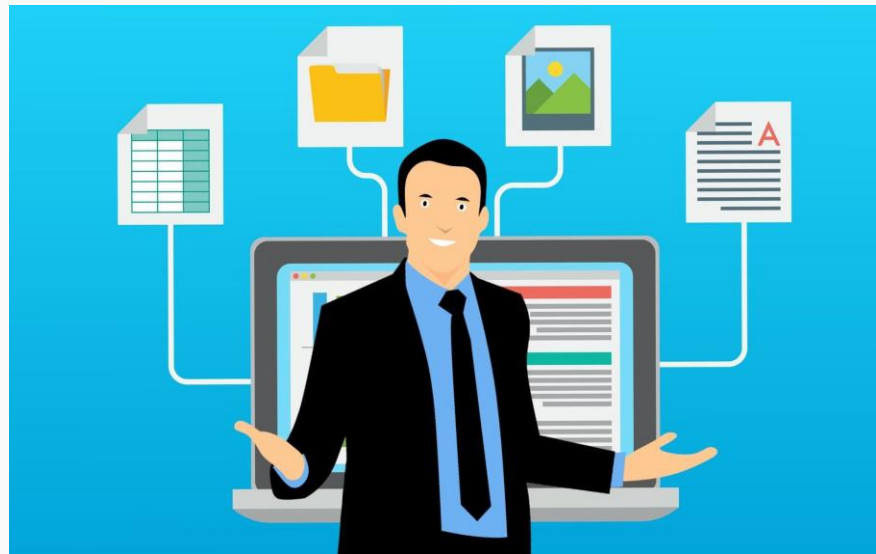
- Es gibt kein Legacy-System - die Datenbank muss von Grund auf erstellt werden.
- Datenqualität sollte bei 100% liegen (>97% ist zu wenig)
 - **Meldepflichtige Krankheiten**
 - **Sicherheit Mitarbeiter steht an erster Stelle**
 - **Betriebswirtschaftlich sinniges Wirtschaften mit Ressourcen**
- Es werde **zwei IT-Mitarbeiter angestellt**, vor allem für die Wartung
- Operative Daten muss das System akzeptieren & annehmen
- **Realität muss mit den Daten jederzeit übereinstimmen.**
Wie kann das realisiert werden? → Data Quality Concept

...

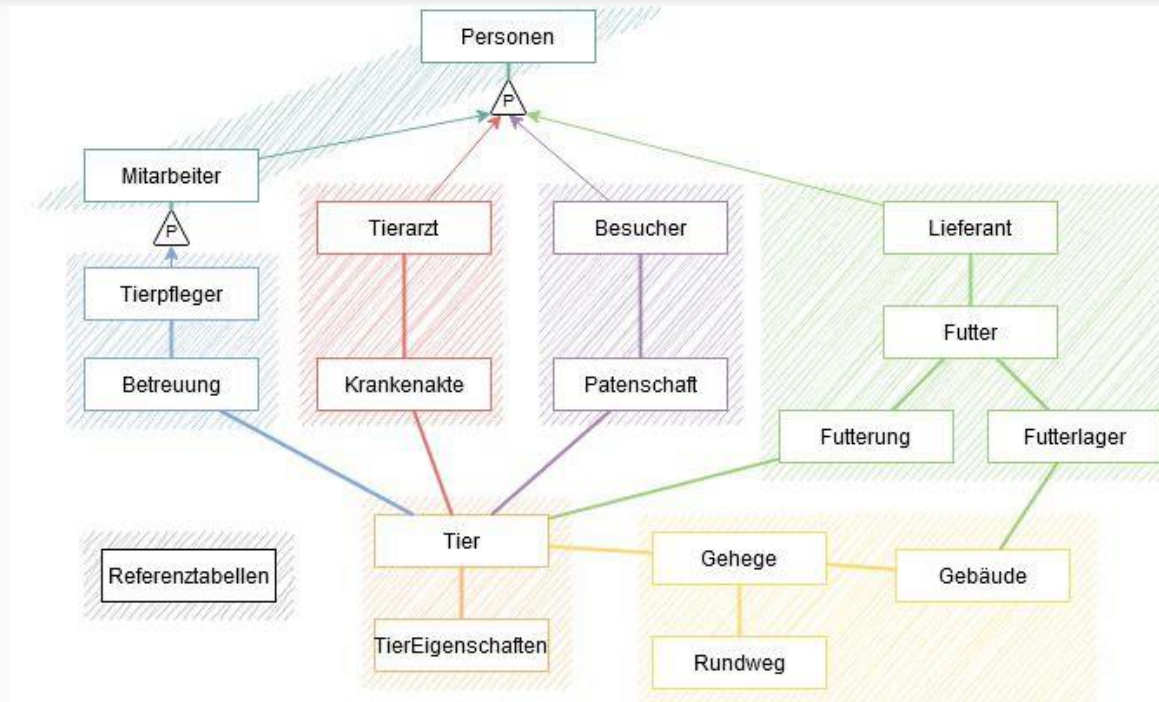


Best Practices aus den Gesprächen mit dem Kunden

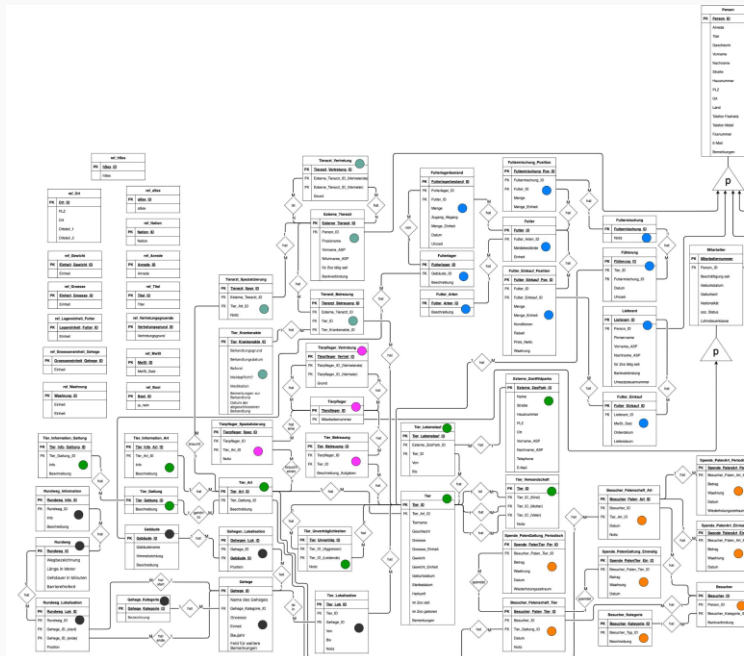
1. Klare Kommunikation sicherstellen
2. Aktives Zuhören praktizieren
3. Transparente Berichterstattung
4. Flexibilität zeigen
5. Erwartungen verwalten
6. Kundenzufriedenheit priorisieren



05_Entity Relationship Modell



05_Entity Relationship Modell



HJS

YC

PL

CN

06_Erstellte Datenbank (Prototyp), Data Dictionary

06_Erstellte Datenbank (Prototyp), Data Dictionary

Tabelle

Feldname

Felddatentyp

Laenge_Zeichen

PK (Primärschlüssel)

FK (Fremdschlüssel)

Referenz

AI (Autoinkrement)

NN (Nicht Null)

Beschreibung

06_Erstellte Datenbank (Prototyp), Data Dictionary

Tabelle	Feldname	Felddatatype	Laenge_Zeichen	PK	FK	Referenz	AI	NN	Beschreibung	String-Länge
Besucher	Besucher_ID	INTEGER		True	False		True	True	Eindeutige Besucher_ID, Besucher des Zoos	41
Besucher	Person_ID	INTEGER		False	True	Person	False	True	Person_ID für jeden Besucher	28
Besucher	Besucher_Kategorie_ID	INTEGER		False	True	Besucher_Kategorie	False	True	Person_ID für jeden Besucher	28
Besucher	Bankverbindung	TEXT	50	False	False		False	False	Bankverbindung für zahlende Besucher	36
Besucher_Kategorie	Besucher_Kategorie_ID	INTEGER		True	False		True	True	Eindeutige Besucher_Kategorie_ID für unterschiedliche Besucher	62
Besucher_Kategorie	Besucher_Typ_ID	INTEGER		False	True	Besucher_Typ	False	True	Besucher_Typ_ID für Besucher mit unterschiedlichen Präferenzen	62
Besucher_Kategorie	Beschreibung	TEXT	50	False	False		False	False	Die verschiedenen Kategorien oder Gruppen von Besuchern im Zoo oder Wildpark.	77
Besucher_Patenschaft_Art	Besucher_Paten_Art_ID	INTEGER		True	False		True	True	Eindeutige Besucher_Paten_Art_ID, für Besucher, die für Paten verfügbar sein könnten	84
Besucher_Patenschaft_Art	Besucher_ID	INTEGER		False	True	Besucher	False	True	Besucher_ID für Besucher des Zoos	33
Besucher_Patenschaft_Art	Tier_Art_ID	INTEGER		False	True	Tier_Art	False	True	Tier_Art_ID für Tiere, für die eine Patenschaft gebucht wurde	61
Besucher_Patenschaft_Art	Datum	Numeric		False	False		False	False	Datum oder den Zeitpunkt, der mit einer bestimmten Art von Patenschaften in einem Zoo oder Wildpark verbunden ist	113
Besucher_Patenschaft_Art	Notiz	TEXT	50	False	False		False	False	enthält Informationen über die verschiedenen Arten von Patenschaften, die von Besuchern für Tiere im Zoo übernommen werden können.	130
Besucher_Patenschaft_Tier	Besucher_Paten_Tier_ID	INTEGER		True	False		True	True	Eindeutige Besucher_Paten_Tier_ID für die Patenschaft eines Besuchers für ein Tier	82
Besucher_Patenschaft_Tier	Besucher_ID	INTEGER		False	True	Besucher	False	True	Besucher_ID für Besucher	24
Besucher_Patenschaft_Tier	Tier_Gattung_ID	INTEGER		False	True	Tier_Gattung	False	True	Tier_Gattung_ID für die Patenschaft eines Tieres	48
Besucher_Patenschaft_Tier	Datum	Numeric		False	False		False	False	Datum, an dem eine bestimmte Patenschaft für ein Tier in einem Zoo oder einer Tieranlage begonnen hat oder erneuert wurde.	122
Besucher_Patenschaft_Tier	Notiz	TEXT	255	False	False		False	False	Notiz oder Anmerkung zu einer Patenschaft für ein bestimmtes Tier in einem Zoo oder Wildpark	93
Besucher_Typ	Besucher_Typ_ID	INTEGER		True	False		True	True	Eindeutige Besucher_Typ_ID des Besucher Type	44
Besucher_Typ	Name	TEXT	50	False	False		False	False	Name der kategorien von Zoo Besuchern	36
Externe_Tierarzt	Externe_Tierarzt_ID	INTEGER		True	False		True	True	Eindeutige Externe_Tierarzt_ID für Externe Tierärzte	52
Externe_Tierarzt	Person_ID	INTEGER		False	True	Person	False	True	Personen_ID, die als externe Tierärzte fungieren	48
Externe_Tierarzt	Praxisname	TEXT	50	False	False		False	False	Praxisname für einen Externen Tierarzt	38
Externe_Tierarzt	Vorname_ASP	TEXT	50	False	False		False	False	Vorname des Ansprechpartners des Externe_Tierarzt	49
Externe_Tierarzt	Nachname_ASP	TEXT	50	False	False		False	False	Nachname des Ansprechpartners des Externe_Tierarzt	50
Externe_Tierarzt	für Zoo tätig seit	TEXT	50	False	False		False	False	Externer Tierarzt ist für den Zoo tätig seit ...	48
Externe_Tierarzt	Bankverbindung	TEXT	50	False	False		False	False	Bankverbindung des Externen Tierarztes	38
Externe_ZooWildparks	Externe_ZooPark_ID	INTEGER		True	False		True	True	Eindeutige Externe_ZooPark_ID des externe Zoo Wildparks	55
Externe_ZooWildparks	Name	TEXT	50	False	False		False	False	Name der externe Zoo Wildparks	30
Externe_ZooWildparks	Straße	TEXT	50	False	False		False	False	Strassenname der Anschrift des externe Zoo Wildparks	52
Externe_ZooWildparks	Hausnummer	TEXT	10	False	False		False	False	Hausnummer der Anschrift des externe Zoo Wildparks	50
Externe_ZooWildparks	PLZ	TEXT	10	False	False		False	False	Postleitzahl (D) der Anschrift des externe Zoo Wildparks	55
Externe_ZooWildparks	Ort	TEXT	50	False	False		False	False	Ortschaft der Anschrift des externe Zoo Wildparks	50
Externe_ZooWildparks	Vorname_ASP	TEXT	50	False	False		False	False	Vorname des Ansprechpartners des externe Zoo Wildparks	54

07 Umsetzung Data Warehouse

Warum Data Warehouse?

Der Schritt zum ersten Data Warehouse ist für Zoo Pirmasens von entscheidender Bedeutung, da er eine Reihe von Vorteilen und Möglichkeiten bietet:

- **Historische Datenverfügbarkeit:**
Speichert historische Daten für Trendanalysen und Prognosen
- **Erhöhte Datenkonsistenz und -qualität:**
Bereinigung und Standardisierung von Daten zur Verbesserung der Datenqualität.
- **Effizienzsteigerung:**
Optimiert für schnellere Berichterstellung und Abfragen. usw...
- **Zentralisierung der Daten:**
Zusammenführung von Daten aus verschiedene Quellen an einem zentrale Ort.

07 Umsetzung Data Warehouse

Warum Data Vault?

Der Data Warehouse schema Data Vault wurde gewählt wegen folgende Vorteile gegenüber alternative schemen wie Star, Snowflake, oder Galaxy.

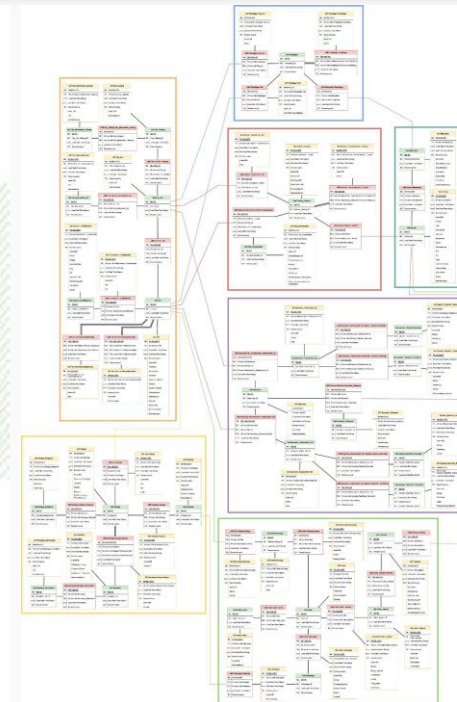
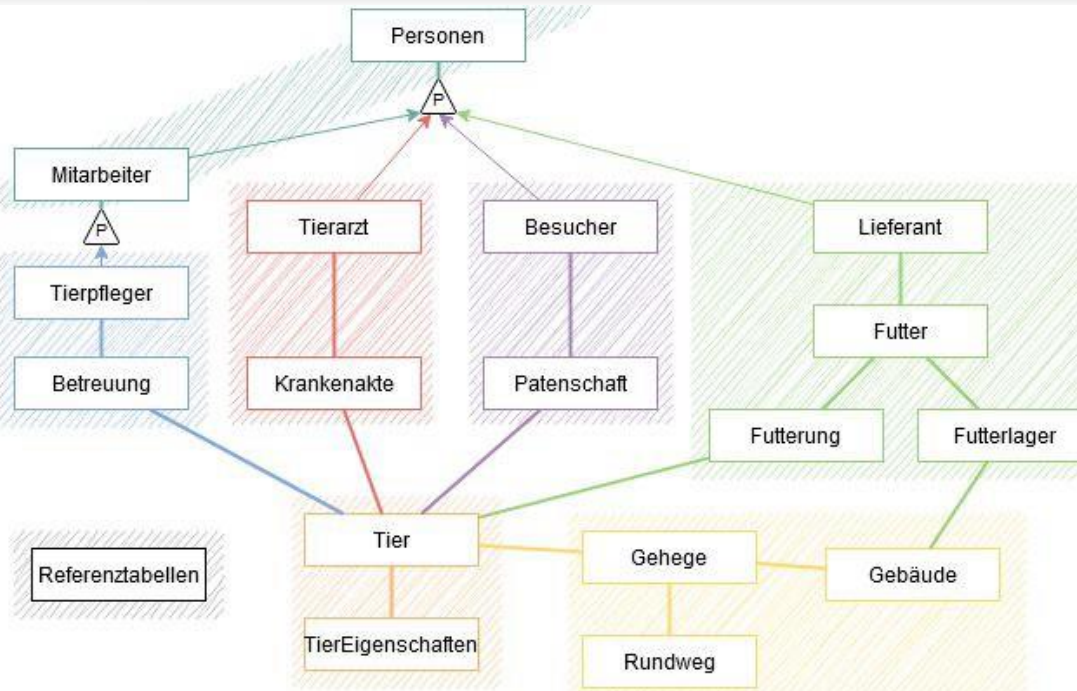
- **Erweiterbarkeit** (In Einklang mit Erweiterungsbestreben des Zoos)
 - Datenbank problemlos erweiterbar durch Zufügen von Sateliten
- **Schnelligkeit** (Verminderte Belastung der operativen Workflows)
 - Extraktion und Loading process ist von Transformationsschritten getrennt
 - Rapide Ladegeschwindigkeit und Automatisierung möglich
- **Einfachheit** (Weniger Personal notwendig)
 - Hubs - Links - Satellites
 - Quick Time to Market - Verminderte Kosten
- **Auditierbarkeit** (Weniger Schwierigkeiten mit Behörden, PETA, etc)
 - Rohdaten in unanbeädert form werde archiviert und historisiert
- **Cutting-Edge** (Awesome!!)

Hub Tabellenname	
HK	BK+RS
BK	PK von Tabelle
LDTs	Load Date Time Stamp
RS	Record source

LINK Tabelle1 Tabelle2	
HK	FK1+FK2+RS
FK1	HK von Hub Tabelle1
FK2	HK von Hub Tabelle2
LDTs	Load Date Time Stamp
RS	Record source

SAT Tabellenname	
HK	FK+RS+LDTs
FK	HK von Hub Tabellenname
LDTs	Load Date Time Stamp
EDTs	End Date Time Stamp
RS	Record source
	Einträge

07 Umsetzung Data Warehouse

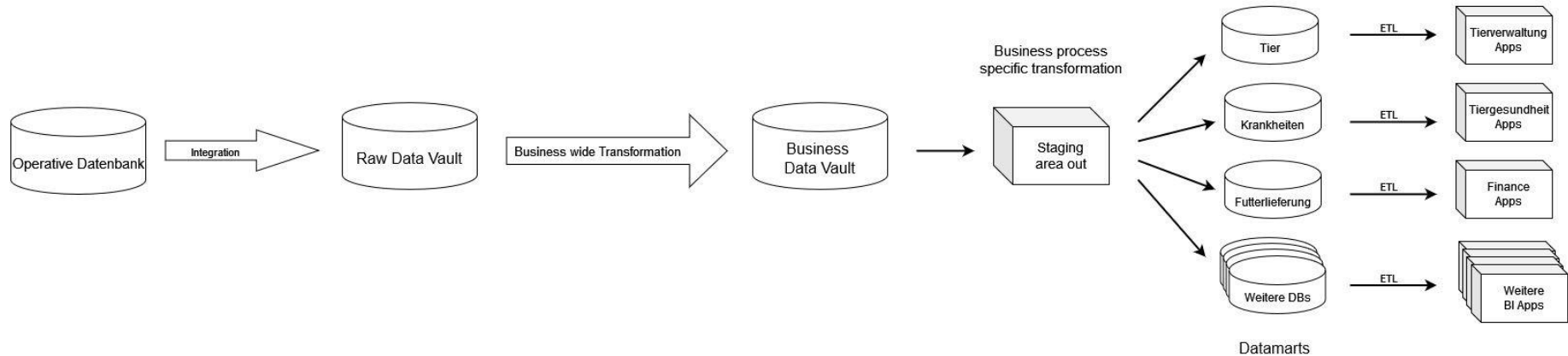


Hub Tabellenname	
HK	<u>BK•RS</u>
BK	PK von Tabelle
LDTS	Load Date Time Stamp
RS	Record source

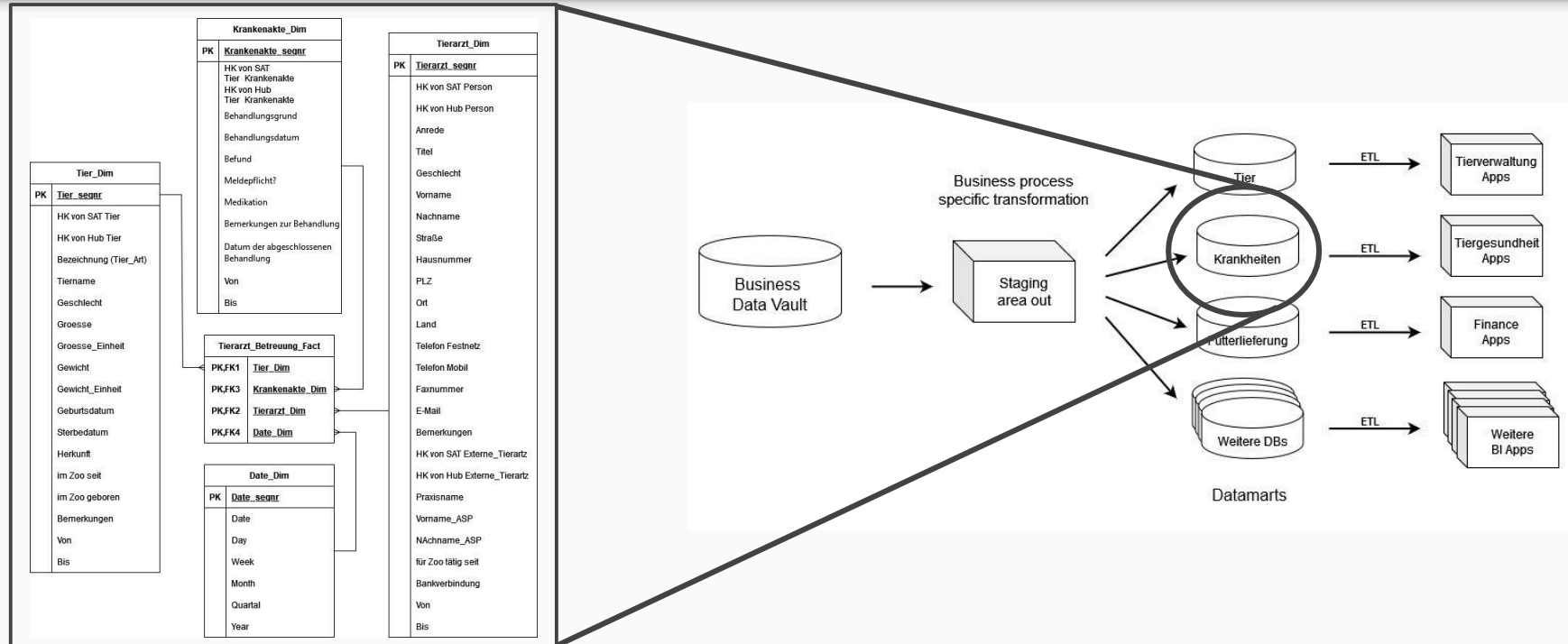
LINK Tabelle1 Tabelle2	
HK	<u>FK1•FK2•RS</u>
FK1	HK von Hub Tabelle1
FK2	HK von Hub Tabelle2
LDTS	Load Date Time Stamp
RS	Record source

SAT Tabellenname	
HK	<u>FK•RS•LDTS</u>
FK	HK von Hub Tabellenname
LDTS	Load Date Time Stamp
EDTS	End Date Time Stamp
RS	Record source
	Einträge

08 Gesamt IT-Architektur



08_Darstellung der Datawarehouse IT-Architektur



09 Software, Tools

Collaborative tools

- Diagrams.net
- Google Docs, Sheets, Slides

Datenbankverwaltungstool

- SQLite

Automatisierung

- Jupyter Labs/Python



Image by FreePik

[illegible]

10_ Vorgehensweise beim Datenqualität Konzept

- Definition von Qualitätsstandards: Festlegung von Kriterien der Datenqualität, basierend auf den Geschäftsanforderungen und -zielen.
- Datenbereinigung: Durchführung von Bereinigungsmaßnahmen wie Duplikaterkennung, Fehlerkorrektur und Standardisierung, um die Datenqualität zu verbessern.
- Implementierung von Daten Qualitätsprozessen: Einführung von Prozessen und Verfahren zur kontinuierlichen Überwachung, Messung und Verbesserung der Datenqualität.

10_ Vorgehensweise beim Dataquality Konzept

- Schulung und Sensibilisierung: Schulung der Mitarbeiter über die Bedeutung von Datenqualität und deren Auswirkungen auf Geschäftsprozesse, um das Bewusstsein und Akzeptanz zu fördern.
- Kontinuierliche Überwachung und Optimierung den System.

11_Inhalt Dataquality Konzept

Data-Qualität ist entscheidend im Bereich Business Intelligence, besonders im Data Engineering.

- Übersicht verschaffen: Umfassende Analyse der vorhandenen Datenstruktur und -inhalte
- Spaltenbeschreibung überprüfen: Prüfung der Spaltenüberschriften für korrekte und aussagekräftige Information
- Häufigkeit von Werten überprüfen: Analyse der Worthäufigkeit für Mustererkennung und Daten Qualitätsverbesserung

11_Inhalt Dataquality Konzept

- Datenvalidierung: Implementierung von Regeln zur Datenvalidierung für Datenintegrität.
- Referenzielle Integrität: Sicherstellung korrekter Beziehungen zwischen Tabellen durch Fremdschlüssel.
- Duplikaterkennung: Mechanismen zur Identifizierung von Duplikaten für Datenredundanz Vermeidung
- Vollständigkeit der Daten: Verwendung von Not Null-Constraints für vollständige Datensätze.

11_Inhalt Dataquality Konzept

- Konsistenz der Daten:Regelmäßige Überprüfung auf Datenkonsistenz und Beziehungen.
- Datenbereinigung:Prozesse zur regelmäßigen Datenbereinigung und aktualisierung.
- Überwachung und Berichterstattung:Mechanismen zur Überwachung und Berichterstellung für Problemerkennung und -behebung.

11_Inhalt Dataquality Konzept

Data-Qualität im Zoo: Wichtige Aspekte und Entitäten

Mitarbeiter: Sicherstellung korrekter Formatierung von Kontaktdaten und Überprüfung von Einstellungsdaten.

Tier: Kontrolle der Konsistenz bei Eltern-Tierbeziehungen und Überwachung von Geburts- und Sterbedaten.

Tierarzt: Validierung von Zuordnungen zu Tieren und detaillierte Dokumentation von Krankheitsfällen.

11_Inhalt Dataquality Konzept

- Tierpfleger: Prüfung der Arbeitsbelastung und Validierung von Zuordnungen zu Tieren.
- Gehege: Überwachung der Belegung und Validierung von Zuordnungen zu Tieren.
- Lieferant: Regelmäßige Überprüfung von Lieferantenkonditionen und Validierung von Kontaktdaten.
- Futterart / Futter: Kontrolle des Futtermittelsverbrauchs und der Lagerbestände sowie Validierung von Lieferanteninformationen.
- Tierpatenschaften: Erfassung und Aktualisierung von Patenschaften sowie Überprüfung der Zuordnung zu Tieren.

12_Ausblick, was sind die nächsten Schritte im Projekt

- Schulung der Mitarbeiter
- Pläne für einen Online-Shop
- Virtuelle Rundgang
- Business wide Transformation and Business data vault
- Business process specific transformation an Datamarts
- Business intelligence concepts
-

Second point

Lorem ipsum dolor sit amet,
consectetur adipiscing elit, sed do
eiusmod tempor incididunt ut labore et
dolore magna aliqua

Incididunt ut labore et dolore

Consectetur adipiscing elit, sed do
eiusmod tempor incididunt ut labore et
dolore magna aliqua

XX⁰%

Use this slide to show a major stat. It can help enforce the presentation's main message or argument.

Final point

A one-line description of it



“This is a super-important quote”



- From an expert

This is the most
important takeaway
that everyone has to
remember.

Thanks!

Contact us:

Your Company

123 Your Street

Your City, ST 12345

no_reply@example.com

www.example.com

