

Data Engineering: Gruppenarbeit zu einer Immobilienvermittlung

Arbeitstitel "ImmoMatch"

Aufteilung innerhalb der Gruppe

Alle Aufgaben wurden von uns gemeinsam immer wieder untereinander vor- und nachbesprochen. Daher gab es eher so etwas wie Hauptzuständige für jeden Bereich, aber im Zeitverlauf des Projektes hat jeder an jeder Stelle Input geliefert.

Aufgabe	Hauptzuständige	Präsentation	Bemerkung
Anforderungsdokumentation	Jürgen, Torben	Torben	
ERM	Jürgen, Natalia	Natalia	
Prototyp / DB Erstellung	Jürgen, Natalia	Jürgen	
DWH Modell	Daniel, Muhammed, Torben	Daniel	
Data Quality	Muhammed, Daniel, Torben	Muhammed	

Verwendete Methoden, Software und Tools

- Draw.io
 - Haben wir verwendet, da es kostenlos zugänglich war und dennoch ansprechende grafische Darstellungen lieferte, die im Laufe der Arbeit von jedem von uns korrigiert werden konnten.
 - Außerdem sehr vielseitig, sodass wir dort sowohl eine Prozessmodellierung, ein ER-Modell und ein Schema für das Datawarehouse machen konnten.
- MySQL Workbench
 - Wurde als erster Test für den Prototyp und zur Fehlervermeidung bei der DB-Erzeugung verwendet, da wir den Umgang mit nicht nur manuellen Tools wie Access üben wollten, aber dennoch eine SQL Basis (von uns) gewünscht war.
- SQLiteBrowser

- Die tatsächliche Umsetzung des Prototyps erfolgte in SQLite, weil das Programm für alle zugänglich war.
- Google Drive: Docs, Präsentation
 - Da wir sehr interaktiv gearbeitet haben und jeder bei jedem Teil mitgeholfen hat, kam für uns nur eine gleichzeitige Arbeit und damit eine Cloudlösung infrage.

Angaben des Kunden, Rahmenbedingungen

- Aufgrund der Covid-19 Situation ist das Unternehmen wirtschaftlich belastet.
- Es wird ein zusätzlicher Geschäftszweig aufgebaut im Immobilienbereich.
- Dabei geht es vor allem darum, dass angebotene Objekte mit Käuferwünschen zusammengebracht werden
- Ein integriertes IT-System gibt es noch nicht.
- Gespeichert werden Information über
 - Verkäufer, Objekte, Lage, Käufer (Kunden), Kundentypen, Kontakte mit den Kunden, Wünsche von Käufern, Besichtigungen, Exposé, Kauf, Mitarbeiter, weitere Beteiligte beim Kauf.
 - Einige typische DB-Interaktionen, die auf der Basis möglich sein sollen, sind unten in einem eigenen Abschnitt aufgezeigt.
- Das Unternehmen wächst aktuell auf der Basis seiner 16 Mitarbeiter sehr schnell
 - Es ist daher eine Entwicklung eines agileren und mächtigeren IT-Systems nötig
 - Aktuell finden laut Kunden zurzeit täglich 1000 Transaktionen statt (10% Telefon, Rest Internetformular)
 - Das Geschäft beschränkt sich auf den nationalen Rahmen
- Es sind keine nicht-funktionalen Anforderungen gestellt worden außer:
 - Spätere reibungslose Zusammenwirkung der operationalen Datenbank mit einer noch zu erstellenden grafischen Oberfläche für die Mitarbeiter
- Es gibt bereits eine Internetformularlösung, die eine automatisierte und streng-formatierte Ausgabe erzeugt, die in Excel eingelesen werden kann. Eine Übertragung in eine erzeugte Datenbank sollte kein Problem sein.
- Da der Kunde noch im Aufbau seines Geschäftes ist, ist die Nachverfolgbarkeit der Datenherkunft für ihn sehr wichtig, um die weitere Geschäftsausrichtung aktiv zu steuern.
 - Es muss also als funktionale Anforderung nachverfolgt werden können aus welcher Quelle: Telefon, Internet, eingekaufte Leads (+genaue Quelle) die Daten kommen.
 - Weitere Analysen, die ein DWH möglich machen muss, sind auch unten beispielhaft skizziert
- **Das Hauptziel unserer Arbeit ist es dem Kunden zu ermöglichen seine Ziele zu messen und zu evaluieren (Leads, Käufer, Verkäufer, Transaktionen,...), am Markt zu bleiben und den Umsatz zu steigern**

Typische Aktionen, die mit dem neuen operativen IT-System möglich sein sollen:

- Abfragen nach einzelnen Attributen:

- Kundennummer, Nachname, Objektid, Haustyp, Lage, Preismaximum, ...
- Abfragen nach Kombinationen:
 - Beispiele: Lage&Typ, Lage&Typ&Preis
 - Objekte eines Verkäufers
- Abfragen nach Auswertungen:
 - Besichtigungen pro Objekt
 - Abfrage Anzahl Exposés pro Kunde
 - Abfrage Anzahl versandter Exposés nach Objekt
 - Exposés(-kunden) eines Objekts, Exposés nach Kunden
- Abfragen mit vielen Attributen:
 - Mögliche Objekte für einen Kunden (Matching)
- Funktionen:
 - Mailing an alle Kunden, ...

Typische Analysen, die mit dem neuen DWH-System möglich sein sollen:

- Tagesaktualität ausreichend, aber Historisierung sehr wichtig, da das Nachfassen bei alten Kunden einen geschätzten Umsatz von 30% ausmacht.
- Auswertung der Leadqualität
- Analysen DWH nach den folgenden Dimensionen (Beispiele)
 - Art des Objektes (Welche Art von Objekt wurde vom Kunden am meisten nachgefragt?)
 - Lage (In welcher Lage besteht mehr oder weniger Nachfrage nach Objekten?)
 - Ort (Welche Stadt hat mehr oder weniger Nachfrage von Käufern oder Verkäufern?)
 - Preis (Wie der Preis des Objekts den Käufer beeinflusst?)
 - Zeit (Wie viele Besichtigungen gab es einem Quartal?)
 - Welcher Mitarbeiter verkauft am besten?, etc.
- Welche der erfassten Merkmale sind kaufentscheidend?

Gespräche mit dem Kunden

Anbei eine Übersicht der Kundengespräche und deren Protokollierung.

Gesprächstermine	Datum/Zeit	Bemerkung
"Gespräch 1"	2021-05-31 10:00	
"Gespräch 2"	2021-05-31 11:00	"Es gibt keine weiteren Kundendaten"

"Gespräch 3"	2021-05-31 14:50	
"Gespräch 4"	2021-05-31 15:25	
"Gespräch 5"	2021-06-01 9:30	
"Gespräch 6"	2021-06-01 12:26	
"Gespräch 7"	2021-06-01 14:45	

Gesprächsprotokolle

- Gespräch 1 zur Anforderungskklärung:
 - Vorhandensein von Daten angesprochen
 - *Schon vorhanden*
 - *Über Telefon (10%) von Mitarbeiter eingetragen / Internetformular sowohl Käufer/Verkäufer (90%)*
 - *Strukturierte E-Mail*
 - *Internetformular in die direkte Datenbank umsetzbar*
 - *Mehrere Excelfiles*
 - *Nationales Geschäft*
 - *Matching*
 - *Eingekaufte Leads*
 - *16 Mitarbeiter (stark wachsend)*
 - *1000 Anfragen pro Tag (Anfragen, Termine, Exposés)*
 - *Zentralisierung, Integrierung*
 - *Nicht-funktional*
 - *Deutsch*
 - *Ausblick: GUI zur optimalen Nutzung*
 - *Felder im bestehenden Internetformular Kaufinteressent*
 - *Allgemeine Informationen (Kunde):*
 - *Anrede (Herr, Frau, Firma); Titel; Vorname; Nachname; Strasse; Plz; Ort; Telefon geschäftlich; Telefon privat; Telefon mobil; eMail; sonstige Bemerkungen; Kontaktwunsch (wann); gewünschte Kontaktart (telefonisch, per eMail, Whatsapp, etc.)*
 - *Kundenwünsche:*
 - *Art des Objektes (1FH, Mehrfamilienhaus, etc.); Preisvorstellung von bis (Preisminimum / Preismaximum); Größe in qm; Ort (wo wird das Objekt gesucht); Lage; Anzahl Zimmer; sonstige Bemerkungen*
 - *Felder im bestehenden Internetformular Verkäufer*

- *Allgemeine Informationen (Kunde):*
 - *Anrede (Herr, Frau, Firma); Titel; Firma; Ansprechpartner (bei Firma); Vorname; Nachname; Strasse; Plz; Ort; Telefon geschäftlich; Telefon privat; Telefon mobil; eMail; sonstige Bemerkungen; Kontaktwunsch (wann); gewünschte Kontaktart (telefonisch, per eMail, Whatsapp, etc.)*
- *Verkaufsinfo:*
 - *Art des Objektes (1FH, Mehrfamilienhaus, etc.); Preisvorstellung (Preisminimum); Größe in qm; Ort (wo wird das Objekt gesucht); Lage; Anzahl Zimmer; sonstige Bemerkungen*
 - *Exposée*
- *Felder bei Leads von Interessenten*
 - *Allgemeine Informationen (Kunde):*
 - *Anrede; Firma; Vorname; Nachname; Strasse; Hausnummer; PLZ; Ort; Email; Telefonnummer*
 - *Interessensinfo:*
 - *Art des Objektes (1FH, Mehrfamilienhaus, etc.), Preisvorstellung von bis, Größe in qm, Ort (wo wird das Objekt gesucht), Anzahl Zimmer*
- *Felder bei Leads Verkäufer*
 - *Allgemeine Informationen (Kunde):*
 - *Anrede; Firma; Ansprechpartner (wenn Firma); Vorname; Nachname; Strasse; Hausnummer; Plz; Ort; Email; Telefonnummer;*
 - *Verkaufsinfo:*
 - *Art des Objektes (1FH, Mehrfamilienhaus, etc.); Preisvorstellung von bis; Größe in qm; Ort; Lage; Anzahl Zimmer; Exposée (meistens dabei, manchmal nicht)*
- *Anforderungen an DW*
 - *Frequenz*
 - *Täglich ausreichend*
 - *Größe? Anzahl an Transaktionen*
 - *1000 Aktionen täglich*
 - *Historisierung*
 - *Nachfassen bei alten Kunden → erneute Kontaktaufnahme (30% des Umsatzes)*
 - *Für Sicherung der Daten und als Leads*
 - *Tagesaktuell ausreichend*
 - *Analysen (zur Überprüfung des DWH)*

- Analysen DWH nach den folgenden Dimensionen
 - Art des Objektes
 - Lage
 - Ort
 - Preis
 - Zeit
 - Merkmale des Objektes
 - Mitarbeiter
 - Kontakte
 - Käuferwünsche
 - Käufer
- → welche der o.g. Merkmale sind kaufentscheidend
- Ziele
 - am Markt bleiben
 - Umsatz steigern
 - Leads auswerten
 - Käufer
 - Verkäufer
 - Umsatz
 - Anzahl Transaktionen (Anzahl Verkäufe)
- Wachstum? Abteilungen?
 - Aktuell 16 Mitarbeiter. Schnell wachsend!
- Aktionen (zur Überprüfung der DB)
 - Neuen Verkäufer anlegen
 - Neues Objekt anlegen (zum Verkauf), Aufnahme Objekt
 - Neuen kaufinteressenten anlegen
 - Neuer Wunsch anlegen (von Kaufinteressent), Aufnahme Kundenanforderungen
 - Telko mit Käufer, Fragen beantworten
 - Telko mit Verkäufer, Fragen beantworten
 - Expose zusenden
 - Besichtigungstermin vereinbaren
 - Besichtigungstermin durchführen
 - Notartermin
 - Kauf vom Objekt
- Gespräch 3:
 - Ist ein Mitarbeiter fester Ansprechpartner für einen Kunden oder können hier Wechsel auftreten (m:n oder 1:n)?
 - Kann auch anders sein. Kunde kann mehrere Ansprechpartner haben
 - Es wird aber durch Historisierung versucht Kunden mit ihnen bekannten Mitarbeitern zu verknüpfen

- Muss Verkäufer bei Besichtigung dabei sein oder reicht der Makler?
 - Verkäufer mit dabei
- Ist Kauftermin gleich Notartermin? Oder unabhängig vom Objektkauf?
 - Ablauf Kauf:
 - Einigung über Preis
 - Kommunikation an Notar
 - Finanzierungsnachweis
 - Dann Termin Notar
 - Nach Geldeingang -> Bestätigung und Kauftermin und Finalisierung
- Gespräch 4:
 - Leads
 - Mehrere Anbieter
 - Durchgabe der Attribute
 - Übermittlung auf verschiedensten Wegen in csv Format
 - Nachfrage der Kundendaten nach Leadumwandlung
 - Ladeflag mit Quelle
 - Nachverfolgbarkeit der Leadqualität
- Gespräch 5:
 - Diskussion über die Leadverarbeitung
 - Klarstellung des "Vorqualifizierungs"-prozesses
 - Ein erfolgreicher Lead ist nur ein Käufer oder Verkäufer, aber nicht ein lediglich gewonnener Kunde
- Gespräch 6
 - Klarstellungen zum Thema Kontakte
 - Käufer/Verkäufer soll spezifizierbar sein
 - Wenn vorhanden Objektid
 - Vorstellungen zum DWHM
 - Klärung zur Kardinalität von Kundenwünschen im DWH
 - Nachfrage zur Trennung von "Inbound Channel" und "Outbound Channel"
- Gespräch 7
 - Nachfrage zu Measures im DWH
 - bzgl. Provision
 - Provision muss eingeführt werden, da diese variieren kann

Best Practices aus den Gesprächen mit dem Kunden

- Gespräche mit und Wünsche vom Kunden mitschreiben
 - Am besten Protokolle gegenzeichnen lassen
- Wiederholung des eigenen Problem-/Lösungsverständnisses bestätigen lassen
- Originaldaten/Musterdaten zukommen lassen

- Unklarheiten hinterfragen
- Geschäftsprozesse immer wieder beschreiben lassen

Entity Relationship Modell nach CHEN

Benötigte Entitäten, Attribute und Feldtypen des ERM

Hier sind vorab die obigen Größen für die Erstellung des ERM vorstrukturiert, die später für den DB Prototypen erweitert werden müssen.

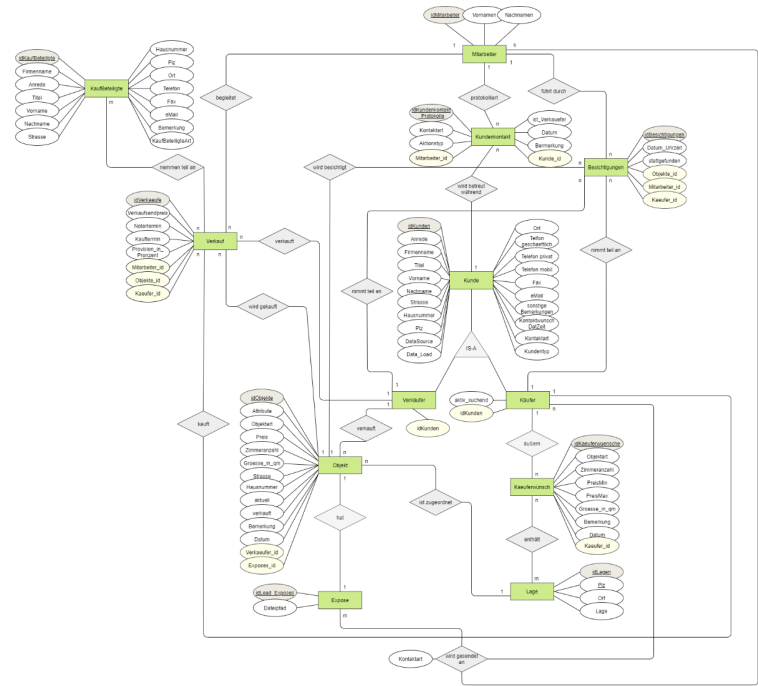
Entität	Attribute	Feldtyp
Kunde	<u>idKunden (PK)</u> DataSource Data_Load Anrede Firmenname Titel Vorname Nachname Strasse Hausnummer PLZ Ort Telefon geschaeftlich Telefon privat Telefon mobil Fax eMail Sonstige Bemerkungen KontaktwunschDatZeit Kontaktart Kundentyp	INT VARCHAR DATETIME VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR
Käufer	<i>Is-a Kunde</i> idKunde aktiv_suchend	INT BOOL
Verkäufer	<i>Is-a Kunde</i> idKunde	INT
Mitarbeiter	<u>idMitarbeiter (PK)</u> Vorname Nachname	INT VARCHAR VARCHAR
Objekte	<u>idObjekte (PK)</u> Objektarten Lagen Mindestpreis (€) Zimmeranzahl Groesse_in_qm Strasse Hausnummer Bemerkung aktuell Verkauft Datum Verkaeufser_id (FK) Exposes_id (FK)	INT VARCHAR VARCHAR DECIMAL INT INT VARCHAR VARCHAR VARCHAR BOOL BOOL DATETIME INT INT
Expose	<u>idExposes (PK)</u> Dateipfad	INT VARCHAR

Lagen	<u>idLagen (PK)</u> PLZ Ort Lage	INT VARCHAR VARCHAR VARCHAR
Käuferwunsch	<u>idKaeuferwuensche (PK)</u> Objektarten Zimmeranzahl PreisMin PreisMax Groesse_in_qm Bemerkung Datum <i>Kaeufer_id (FK)</i>	INT VARCHAR INT DECIMAL DECIMAL INT VARCHAR DATETIME INT
Besichtigung	<u>idBesichtigungen (PK)</u> Datum_Uhrzeit Stattgefunden <i>Objekte_id (FK)</i> <i>Mitarbeiter_id (FK)</i> <i>Kaeufer_id (FK)</i>	INT DATETIME BOOL INT INT INT
Kundenkontakt	<u>idKundenkontakt Protkolle (PK)</u> Kontaktart Aktionstyp ist_Verkaeufer Datum Bemerkungen <i>Kunde_id (FK)</i> <i>Mitarbeiter_id (FK)</i>	INT VARCHAR VARCHAR BOOL DATETIME VARCHAR INT INT
Verkauf	<u>idVerkaeufe (PK)</u> Verkaufsendpreis Provision_in_prozent Notartermin Kauftermin <i>Objekte_id (FK)</i> <i>Mitarbeiter_id (FK)</i> <i>Kaeufer_id (FK)</i>	INT DECIMAL DECIMAL DATETIME DATETIME INT INT INT
Kaufbeteiligte	<u>idKaufbeteiligte (PK)</u> KaufBeteiligteArt Firmenname Anrede Titel Vorname Nachname Strasse Hausnummer PLZ Ort Telefon geschaeftlich Telefon privat Telefon mobil Fax eMail Bemerkung	INT VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR
KaeuferLeadData	<u>idKunden VLead Data (PK)</u> Data_Source DS_LoadDate Kontaktdatum Kontaktversuche_Anzahl	INT VARCHAR DATETIME DATETIME INT

	Kunden_id Daten_unbrauchbar Anrede Firmenname Vorname Nachname Strasse Hausnummer PLZ Ort Telefonnummer eMail Art des Objektes Preisvorstellung_Min Preisvorstellung_Max Groesse_in_qm Ort_Objekt Zimmeranzahl Sonstige Bemerkungen	INT BOOL VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR DECIMAL DECIMAL INT VARCHAR VARCHAR VARCHAR
LeadVerkaufsinteressenten	idKunden_KLead_Data (PK) Data_Source DS_LoadDate Kontaktdatum Kontaktversuche_Anzahl Kunden_id Daten_unbrauchbar Firmenname Anrede Vorname Nachname Strasse Hausnummer PLZ Ort Telefonnummer eMail Art des Objektes Preisvorstellung_Min Groesse_in_qm Ort_Objekt Lage_des_Objektes Zimmeranzahl Sonstige Bemerkungen Lead_Exposes_id (FK)	INT VARCHAR DATETIME DATETIME INT INT BOOL VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR DECIMAL DECIMAL INT VARCHAR VARCHAR INT VARCHAR INT INT
Lead_Exposes	idLead_Exposes (PK) Dateipfad	INT VARCHAR

Vorschau des ERM

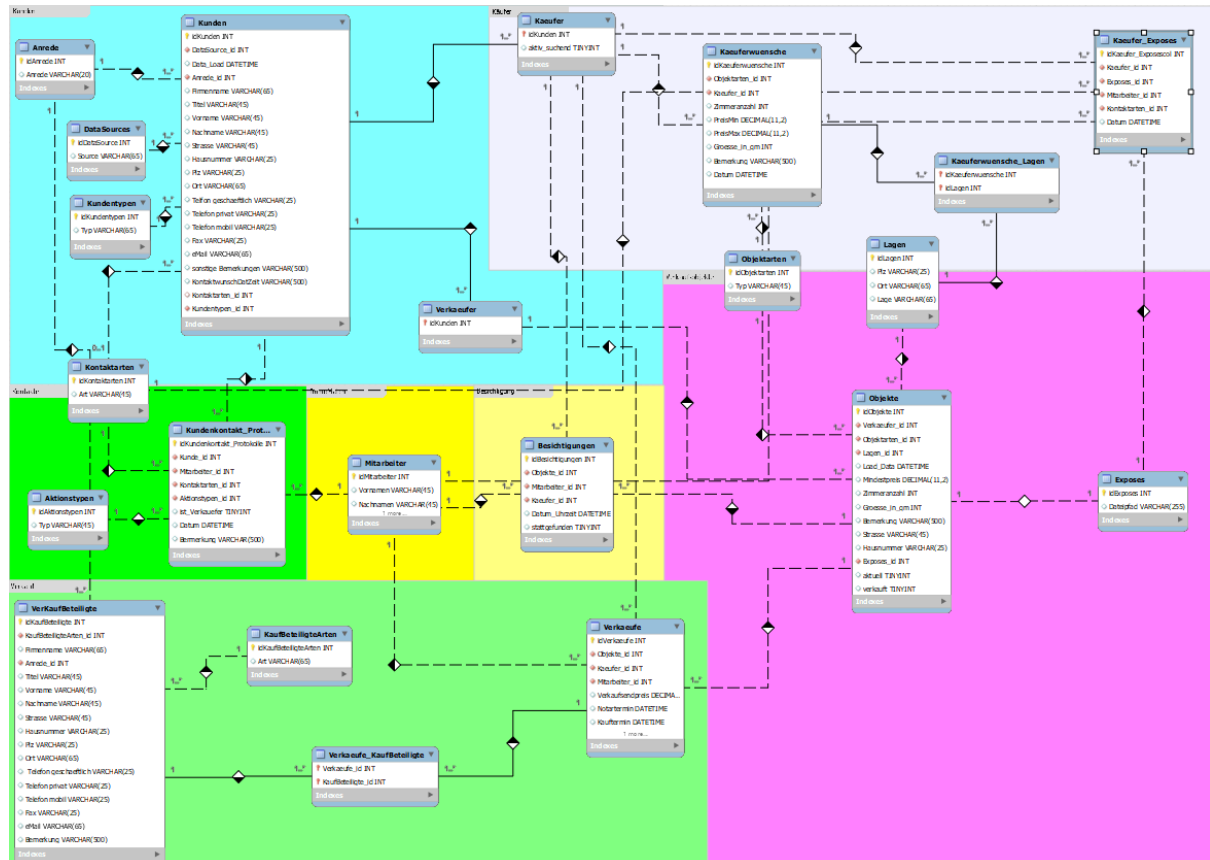
Unten ist eine Vorschau des Entity-Relationship-Modells zu sehen. Eine höhere Auflösung der Abbildung ist in der separaten Datei "ERM-ImmoMatch.png" zu finden.



Erstellte Datenbank (Prototyp)

Vorschau des DB-Prototypen

Unten ist eine grafische Vorschau des DB Prototypen zu sehen. Eine höhere Auflösung der Abbildung ist in der separaten Datei "ImmoMatch_MySQL_WB.png" zu finden.



Der SQL-Code zur Erstellung der Datenbank inklusiver aller Tabellen und Felddaten ist in "ImmoMatch_mysqlite.sql" zu finden. Die Datentypen wurde nach eigenen Verständnis definiert und werden beim Testlauf mit dem Kunden angepasst. Exemplarisch anbei ein SQL-Statement der Kundentabelle:

```
DROP TABLE IF EXISTS `Kunden`;
```

```
CREATE TABLE IF NOT EXISTS `Kunden` (  
  `idKunden` INT NOT NULL,  
  `DataSources_id` INT NOT NULL,  
  `Data_Load` DATETIME NULL,  
  `Anrede_id` INT NOT NULL,  
  `Firmenname` VARCHAR(65) NULL,  
  `Titel` VARCHAR(45) NULL,
```

```

`Vorname` VARCHAR(45) NULL,
`Nachname` VARCHAR(45) NULL,
`Strasse` VARCHAR(45) NULL,
`Hausnummer` VARCHAR(25) NULL,
`Plz` VARCHAR(25) NULL,
`Ort` VARCHAR(65) NULL,
`Telfon geschaeftlich` VARCHAR(25) NULL,
`Telefon privat` VARCHAR(25) NULL,
`Telefon mobil` VARCHAR(25) NULL,
`Fax` VARCHAR(25) NULL,
`eMail` VARCHAR(65) NULL,
`sonstige Bemerkungen` VARCHAR(500) NULL,
`KontaktwunschDatZeit` VARCHAR(500) NULL,
`Kontaktarten_id` INT NULL,
`Kudentypen_id` INT NOT NULL,
PRIMARY KEY (`idKunden`),
CONSTRAINT `fk_Kunden_Anrede1`
FOREIGN KEY (`Anrede_id`)
REFERENCES `Anrede` (`idAnrede`),
CONSTRAINT `fk_Kunden_Kontaktarten1`
FOREIGN KEY (`Kontaktarten_id`)
REFERENCES `Kontaktarten` (`idKontaktarten`),
CONSTRAINT `fk_Kunden_Kudentypen1`
FOREIGN KEY (`Kudentypen_id`)
REFERENCES `Kudentypen` (`idKudentypen`),
CONSTRAINT `fk_Kunden_DataSources1`
FOREIGN KEY (`DataSources_id`)
REFERENCES `DataSources` (`idDataSources`));

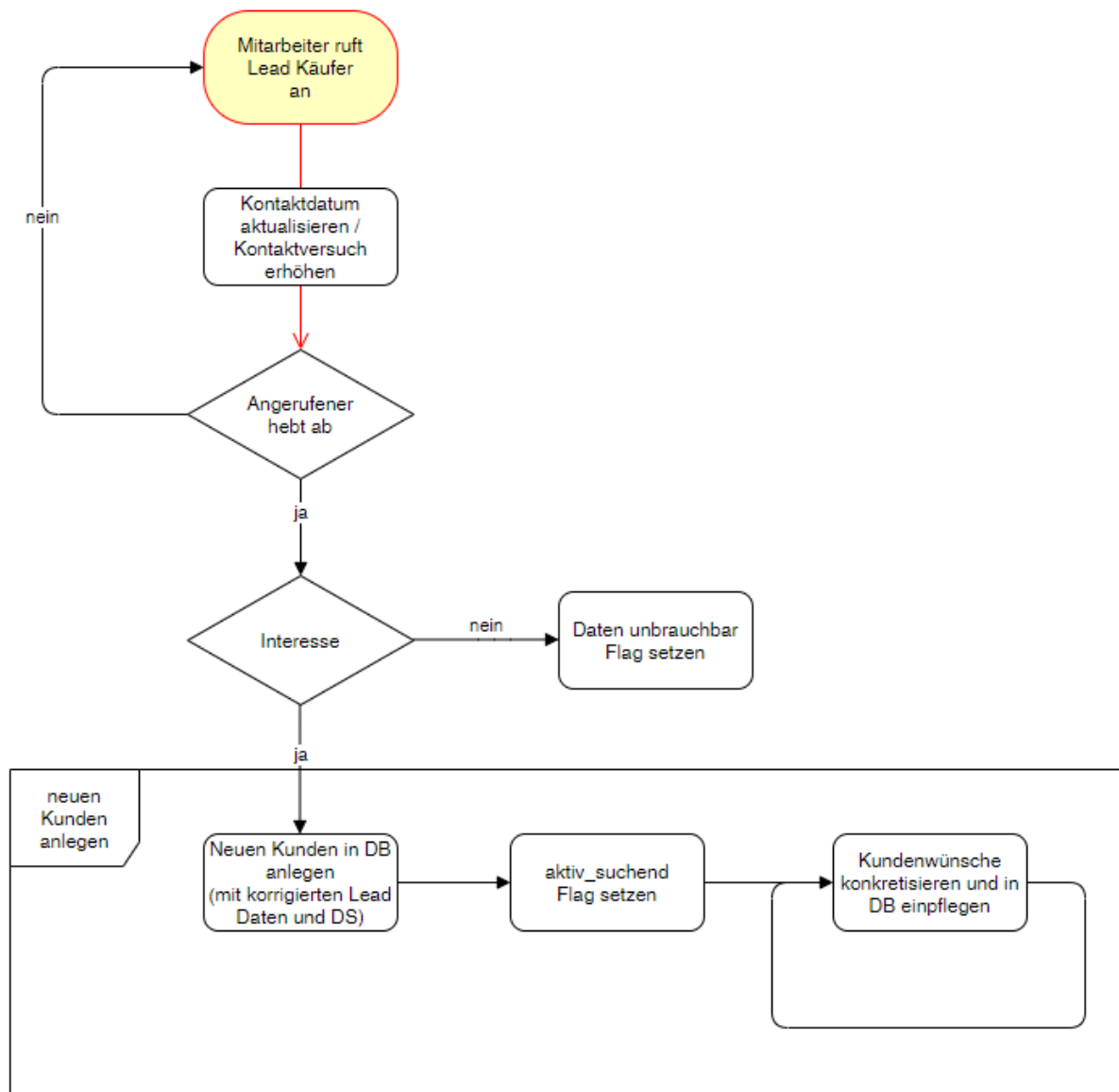
```

Die Datenbank wurde aus dem in MySQL Workbench erstellten Modell erzeugt. Die Syntax wurde auf SQLite angepasst und es wurde die SQLite-Datenbank "ImmoMatch.db" erzeugt.

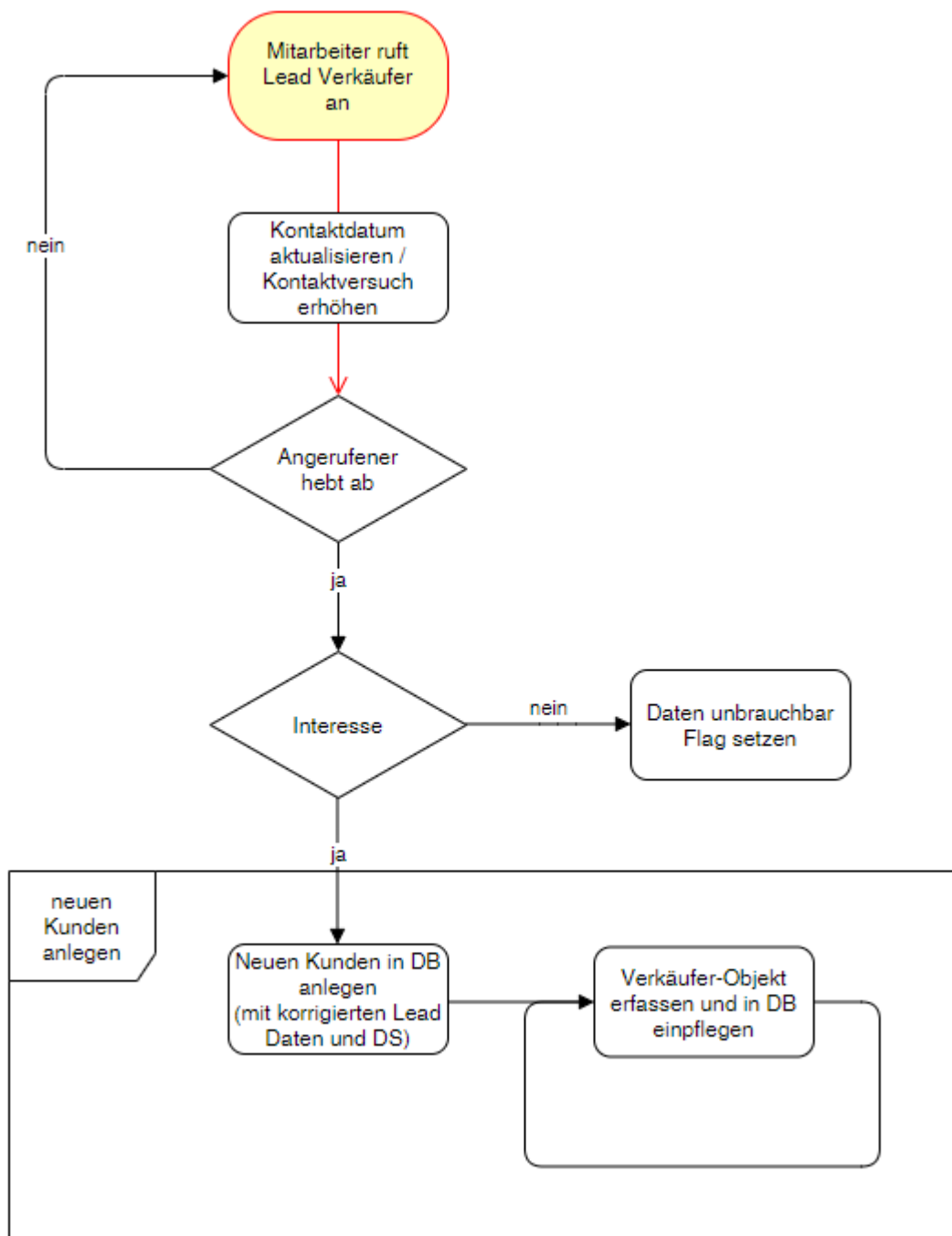
Verarbeitung und Überführung der Leaddaten

Die Erfahrung zeigt, dass die Qualität von Lead-Daten immer sehr durchwachsen sind. Dies zeigt sich oft in ungenauen, falschen und nicht aktuellen Datensätzen. Aus diesem Grund wurden die Lead-Datensätze in separaten Tabellen gespeichert. Die Kundenakquise erfolgt durch einen persönlichen Anruf. Besteht Interesse beim Lead, werden die Lead-Daten mit den Kunden aktualisiert und als neuen Kunden mit Quellenangabe in die "ImmoMatchDB" hinzugefügt. Somit erfolgt eine Übernahme der Lead-Daten erst nach einer erfolgreichen "Vorqualifizierung". Die zugehörigen Prozesse sind in einem Aktivitätsdiagramm (UML) modelliert (siehe nachfolgende Darstellungen).

Käufer Akquise



Verkäufer Akquise



Dokumentation

Im Folgenden findet sich die tabellarische Dokumentation aller Entitäten und Attribute der Prototypendatenbank. Eine Bemerkung und Kommentare sind angegeben, wenn der Feldname nicht selbsterklärend ist.

Bemerkung (basierend auf der Nutzung von MySQL Workbench): Schematisch ist in unserer Dokumentation ein Primärschlüssel an *“idENTITY”* zu erkennen. Der zugehörige Fremdschlüssel heißt *“ENTITY_id”*.

Entität	Attribute	Feldtyp	Bemerkung
Kunden	<u>idKunden (PK)</u> DataSources_id (FK) Data_Load Anrede_id (FK) Firmenname Titel Vorname Nachname Strasse Hausnummer PLZ Ort Telefon geschaeftlich Telefon privat Telefon mobil Fax eMail Sonstige Bemerkungen KontaktwunschDatZeit Kontaktarten_id (FK) Kundentypen_id (FK)	INT INT DATETIME INT VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR	Auswahl aus separater Entität Auswahl aus separater Entität Bei Firma: Ansprechpartner = "Vorname Nachname" Bsp: "Mi/Do/Fr 1600-1830" Auswahl aus separater Entität Auswahl aus separater Entität
Anrede	<u>idAnrede (PK)</u> Anrede	INT VARCHAR	"Firma, Familie, Herr, Frau"
Kontaktarten	<u>idKontaktarten (PK)</u> Art	INT VARCHAR	Telefonisch, per E-Mail, WhatsApp, etc.
Kundentypen	<u>idKundentypen (PK)</u> Typ	INT VARCHAR	"Privat, Makler, Bauträger"
DataSources	<u>idDataSources (PK)</u> <u>Source</u>	INT VARCHAR	"Telefon, Internetformular, \$LEADSOURCENAMES\$"
Kaeufer	<u>idKunden (FK=PK)</u> aktiv_suchend	INT BOOL	Durch <u>idKunden</u> eindeutig identifiziert <i>Kann hiermit inaktiv gesetzt werden</i>
Verkaeufer	<u>idKunden (FK=PK)</u>	INT	Durch <u>idKunden</u> eindeutig identifiziert
Mitarbeiter ¹	<u>idMitarbeiter (PK)</u> Vorname Nachname	INT VARCHAR VARCHAR	

¹ Es wurde bewusst auf weitere Datenfelder verzichtet. Eine eindeutige Zuordnung sollte über die ID möglich sein. Personaldaten sollten aus Datenschutzgründen separat geführt werden.

Objekte	<u>idObjekte (PK)</u> <i>Verkaefer_id (FK)</i> <i>Objektarten_id (FK)</i> <i>Lagen_id (FK)</i> Mindestpreis (€) Zimmeranzahl Groesse_in_qm Bemerkung Strasse Hausnummer <i>Exposes_id (FK)</i> aktuell verkauft ² Datum	INT INT INT INT DECIMAL INT INT VARCHAR VARCHAR VARCHAR INT BOOL BOOL DATETIME	Auswahl aus separater Entität Auswahl aus separater Entität Auswahl aus separater Entität Abgelegt in separater Entität Siehe Fußnote <i>Datum der Eintragung</i>
Exposes	<u>idExposes (PK)</u> Dateipfad	INT VARCHAR	<i>*.doc, *.ppt, *.pdf, etc. hier hinterlegt</i>
Objektarten	<u>idObjektarten (PK)</u> Typ	INT VARCHAR	"1 FH, 2-3 FH, Mehrfamilienhaus, Gewerbeobjekt, Wohnungen, Lofts, etc."
Lagen	<u>idLagen (PK)</u> PLZ Ort Lage	INT VARCHAR VARCHAR VARCHAR	<i>Meint hier Stadtteil³</i>
Kaeuferwuensche	<u>idKaeuferwuensche (PK)</u> <i>Objektarten_id (FK)</i> <i>Kaeufer_id (FK)</i> Zimmeranzahl PreisMin PreisMax Groesse_in_qm Bemerkung Datum	INT INT INT INT DECIMAL DECIMAL INT VARCHAR DATETIME	<i>Datum der Eintragung</i>
Kaeufer_Exposes (Zwischentabelle)	<u>idKaeufer_Expose (PK)</u> <i>Kaeufer_id (FK)</i> <i>Exposes_id (FK)</i> <i>Mitarbeiter_id (FK)</i> <i>Kontaktart_id (FK)</i> Datum	INT INT INT INT INT DATETIME	<i>Datum der Eintragung</i>
Kaeuferwuensche_Lage (Zwischentabelle)	<u>PK = FK1 + FK2</u> <i>Kaeuferwuensche_id (FK)</i> <i>Lagen_id (FK)</i>	INT INT	<i>Wir gehen hier davon aus, dass ein zusammengesetzter PK aus den Fremdschlüsseln ausreichend ist</i>
Besichtigung	<u>idBesichtigungen (PK)</u> <i>Objekte_id (FK)</i> <i>Kaeufer_id (FK)</i> <i>Mitarbeiter_id (FK)</i> Datum_Uhrzeit stattgefunden	INT INT INT INT DATETIME BOOL	<i>Verkäufer ist über Objekt in Besichtigung vorhanden</i> <i>Möglichkeit einen Termin als ausgefallen zu markieren</i>
Kundenkontakte_Protkoll	<u>idKundenkontakt_Protkoll (PK)</u>	INT INT	

² Kann auch ein Vermerk über Verkauf durch die Konkurrenz sein. Verkauf durch uns über *Kauftermin*-Eintrag in *Verkaeufe* abgedeckt

³ Stadtteil-Modellierung war vom Kunden gewünscht.

	<i>Kunden_id (FK)</i> <i>Mitarbeiter_id (FK)</i> <i>Kontaktart_id (FK)</i> <i>Aktionstyp_id (FK)</i> ist_Verkaeuer Datum Bemerkungen	INT INT INT INT BOOL DATETIME VARCHAR	Auswahl aus separater Entität Auswahl aus separater Entität <i>Zur Festlegung, ob Kunde als Käufer/Verkäufer Kontakt aufnimmt</i>
Aktionstypen	<u>idAktionstypen (PK)</u> Typ	INT VARCHAR	Liste: "Anfrage, Aufnahme als Kunde, Expose Anfrage, Terminanfrage, ..."
Verkaeufe	<u>idVerkaeufe (PK)</u> <i>Kaeufer_id (FK)</i> <i>Objekte_id (FK)</i> <i>Mitarbeiter_id (FK)</i> Verkaufsendpreis Provision_in_prozent Notartermin Kauftermin	INT INT INT INT DECIMAL DECIMAL DATETIME DATETIME	<i>Festgelegter Provisionssatz</i> <i>Ein festgelegter Kauftermin spezifiziert gleichzeitig einen erfolgreichen Verkauf durch uns.</i>
Kaufbeteiligte	<u>idKaufbeteiligte (PK)</u> <i>KaufBeteiligteArten_id (FK)</i> Firmenname Anrede_ID (FK) Titel Vorname Nachname Strasse Hausnummer PLZ Ort Telefon geschaeftlich Telefon privat Telefon mobil Fax eMail Bemerkung	INT INT VARCHAR INT VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR	Auswahl aus separater Entität Bei Firma: Ansprechpartner = "Vorname Nachname" Auswahl aus separater Entität
KaufBeteiligteArten ⁴	<u>idKaufBeteiligtenArten (PK)</u> Art	INT VARCHAR	"Notar, Rechtsanwalt, Objektteilbesitzer"
Verkaeufe_Kaufbeteiligte (Zwischentabelle)	<u>PK = FK1 + FK2</u> <i>Verkaeufe_id (FK)</i> <i>KaufBeteiligte_id (FK)</i>	INT INT	<i>Wir gehen hier davon aus, dass ein zusammengesetzter PK aus den Fremdschlüsseln ausreichend ist</i>
KaeuferLeadData ⁵	<u>idKunden_VLead_Data (PK)</u> Data_Source DS_LoadDate Kontaktdatum Kontaktversuche_Anzahl Kunden_id Daten_unbrauchbar Anrede Firmenname Vorname	INT VARCHAR DATETIME DATETIME INT INT BOOL VARCHAR VARCHAR VARCHAR	<i>Zum Zählen der erfolgreichen Kontaktversuche</i> <i>Sollte der Lead zu einem Kunden wird hier bei der Umwandlung in einen Kunden die Kunden_id hinterlegt</i> <i>Flag zur Markierung schlechter Leaddaten</i>

⁴ Hierdurch ist auch der Fall abgedeckt, dass der Verkäufer nicht Besitzer des Objekts ist.

⁵ Wie auch an anderer Stelle noch spezifiziert werden die Leads durch einen separaten Loadingprozess an getrennter Stelle eingeladen und erst nach Vorqualifizierung (bei erfolgreicher Kontaktaufnahme und mit Datenüberprüfung) in die Kundenentität übertragen

	Nachname Strasse Hausnummer PLZ Ort Telefonnummer eMail Art des Objektes Preisvorstellung_Min Preisvorstellung_Max Groesse_in_qm Ort_Objekt Zimmeranzahl Sonstige Bemerkungen	VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR DECIMAL DECIMAL INT VARCHAR VARCHAR VARCHAR	
LeadVerkaufsintere ressenten ⁶	idKunden_KLead_Data (PK) Data_Source DS_LoadDate Kontaktdatum Kontaktversuche_Anzahl Kunden_id Daten_unbrauchbar Firmenname Anrede Vorname Nachname Strasse Hausnummer PLZ Ort Telefonnummer eMail Art des Objektes Preisvorstellung_Min Groesse_in_qm Ort_Objekt Lage_des_Objektes Zimmeranzahl Sonstige Bemerkungen Lead_Exposes_id	INT VARCHAR DATETIME DATETIME INT INT BOOL VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR VARCHAR DECIMAL INT VARCHAR VARCHAR INT VARCHAR INT	Zum Zählen der erfolgreichen Kontaktversuche Sollte der Lead zu einem Kunden wird hier bei der Umwandlung in einen Kunden die Kunden_id hinterlegt Flag zur Markierung schlechter Leaddaten
Lead_Exposes	<u>idLead_Exposes (PK)</u> Dateipfad	INT VARCHAR	

⁶ siehe ⁵

Das DWH Modell

Verwendetes Architekturmodell DWH

Aufgrund der Anforderungen wurde das DWH mit einem Galaxy Schema erstellt. Wir sind auf ein Galaxyschema angewiesen, weil wir zwei voneinander relativ unabhängige feingranulare Fakttabellen haben. Die Kundenkontakte führen zwar zu Verkäufen, aber sind auch davon unabhängige Interaktionen. Diese wäre mit einem konventionellen Star- oder Snowflakeschema nur schwer umzusetzen.

Dieses komplexe, logische Architekturkonzept besitzt ein gutes Abfrageverhalten.

Als Galaxien bestehen die beiden Fakt-Tabellen und deren Dimensionen.

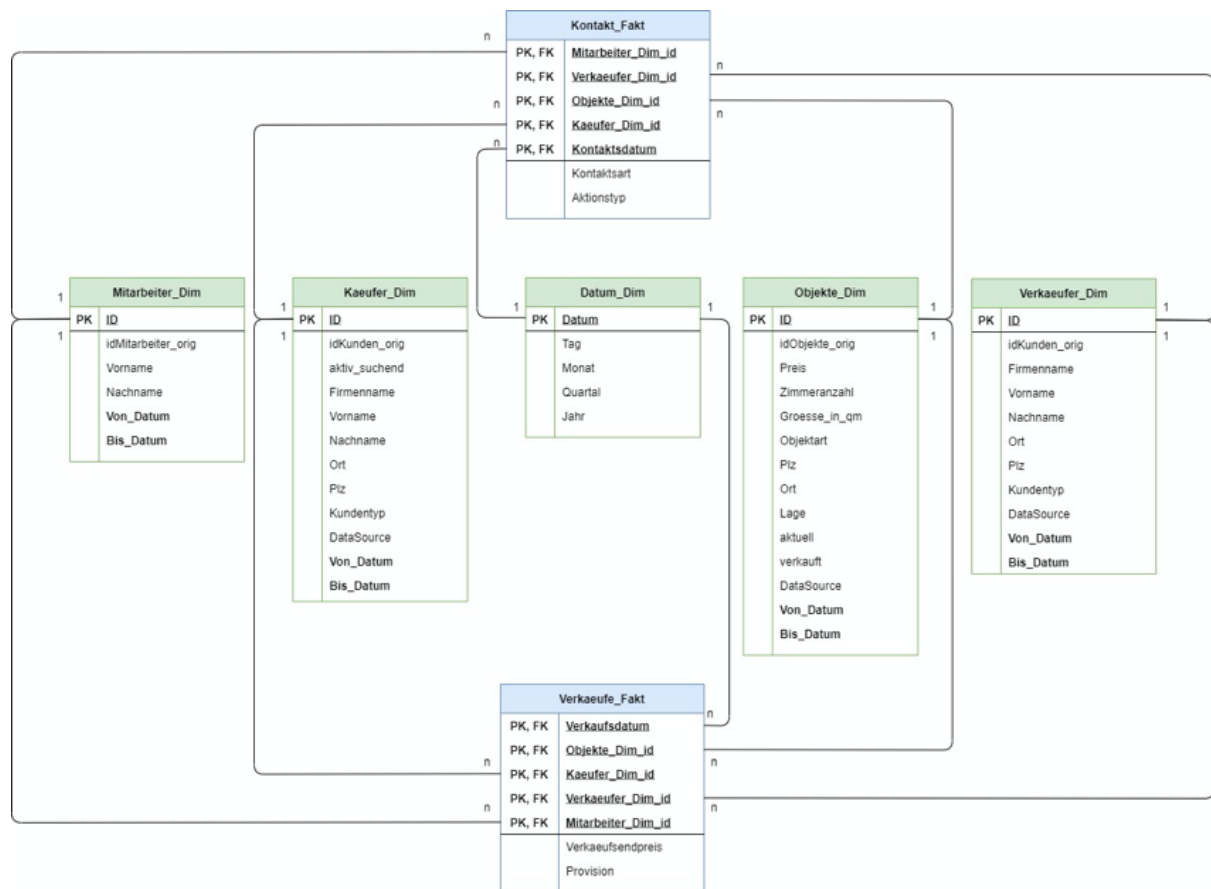
Fakt-Tabellen:

- Kontakttabelle
- Kauftabelle

Dimensionale Tabelle:

- Mitarbeiter_Dim
- Kunden_Dim
- Datum_Dim
- Objekte_Dim
- Verkaeuer_Dim

Erstelltes Datenmodell DWH



Schema zur Übertragung der operativen Daten in das Datawarehouse anhand von Beispielen

Unten sind die Tabellen aus der operativen Datenbank zu sehen und wie die Daten dieser im Datawarehouse in die gemeinsame Faktentabelle Kontakte_Fakt übertragen werden können. Joins oder Abfragen, die dabei gemacht werden müssen, sind über einen Verweis auf den entsprechenden Schlüssel, über den diese möglich sind, dargestellt.

Mitarbeiter_Dim_ID	Kaeufer_Dim_id	Verkaeuffer_Dim_ID	Objekte_Dim_ID	Kontaktdatum+Uhrzeit	Kontaktart	Aktionstyp
99999999	1113 1	99999999 Ist_Verkaeuffer	99999999	20.05.21 2	Internetformular 3	Kundenerstellung 4
99999999	1113	99999999	2905	20.05.21	Internetformular	Objekterzeugung
11	217	99999999	99999999	21.05.21	Email	Exposees angefragt
11 1	217 2	19 3	23	21.05.21 4	Email 5	Exposees gesendet 6
11	217	305 Über Exposed	527	21.05.21	Email	Exposees gesendet
11	217	19 Über Objekt	23	23.05.21	Telefon	Besichtigungstermin vereinbart
11 1	217 2	19 3	23	27.05.21 4	Direkter Kontakt	Besichtigungstermin durchgeführt 5

Kundenkontakt_Protokolle	
IdKundenKontakt_Protokolle (PK)	
Kunde_id (FK)	
Mitarbeiter_id (FK) 1	
Kontaktarten_id (FK) 3	
Aktionstyp_id (FK) 4	
ist_Verkaeuffer	
Datum 2	
Bemerkung	

Besichtigungen 5	
idBesichtigungen (PK)	
Objekte_id (FK) 3	
Kaeufer_id (FK) 2	
Mitarbeiter_id (FK) 1	
Datum_Uhrzeit 4	
stattgefunden	

Kaeufer_Exposes 6	
idKaeufer_Expose (PK)	
Kaeufer_id (FK) 2	
Exposes_id (FK) 3	
Mitarbeiter_id (FK) 1	
Kontaktart_id (FK) 5	
Datum 4	

Demonstration einiger vom Kunden gewünschter Abfragen

Im Folgenden wurden auf Basis der Kundengespräche einige der konkreten Fragestellungen an die Business Intelligence ausformuliert und kurz angedeutet mit welchen Datenquellen diese Analysen umgesetzt werden können. Dies demonstriert die Funktionalität des von uns gewählten DW-Modells:

Analysen DWH nach den folgenden Dimensionen

- *Art des Objektes:* Welche Art von Objekt wurde vom Kunden am meisten nachgefragt.
Kontakt_Fakt(Objekte_Dim_id, Kaeufer_Dim_id)
Objekte_Dim(Objekte_Dim_id, Objektart)
Kaeufer_Dim(Kaeufer_Dim_id, Vorname,Nachname)
- *Lage:* In welcher Lager besteht mehr oder weniger Nachfrage nach Verkaeuffer.
Verkaeuffer_Fakt(Objekte_Dim_id, Verkaeuffer_Dim)
Objekte_Dim(Objekte_Dim_id, Lage)
Verkaeuffer_Dim(Verkaeuffer_Dim_id, Vorname,Nachname)
- *Ort:* Welche Stadt hat mehr oder weniger Nachfrage von Verkäufern.
Verkaeuffer_Fakt(Verkaeuffer_Dim, Provision)
Verkaeuffer_Dim(Verkaeuffer_Dim_id, Vorname,Nachname)

- *Preis:* Wie der Preis des Objekts den Käufer beeinflusst.
Verkaeuer_Fakt(Objekte_Dim_id, Kaeufer_Dim_id, Verkaeufsendpreis)
Kaeufer_Dim(Kaeufer_Dim_id, Vorname, Nachname)

- *Zeit:* In welchem Monat des Jahres gab es mehr Verkäufe.
Verkaeuer_Fakt(Verkaufsdatum, Provision)
Datum_Dim(Datum, Monat)

- *Merkmale des Objektes:* Welche Zimmeranzahl hat das Objekt und welche wurden vom Kundentyp am meisten nachgefragt.
Kontakt_Fakt(Objekte_Dim_id, Kaeufer_Dim_id)
Objekte_Dim(Objekte_Dim_id, Zimmeranzahl, Objektart)
Kaeufer_Dim(Kaeufer_Dim_id, Vorname, Nachname, Kundentyp)

- *Mitarbeiter:* Welcher Mitarbeiter hat am meisten Umsatz gemacht.
Verkaeuer_Fakt(Mitarbeiter_Dim_id)
Mitarbeiter_Dim(Mitarbeiter_Dim_id, Vorname, Nachname)

- *Kontakte:* Wie viele Käufer sind mit welchem Arbeiter zugeordnet.
Kontakt_Fakt(Mitarbeiter_Dim_id, Kaeufer_Dim)
Mitarbeiter_Dim(Mitarbeiter_Dim_id, Vorname, Nachname)
Kaeufer_Dim(Kaeufer_Dim_id, Vorname, Nachname)

- *Käuferwünsche:* Was ist der Durchschnitt von Käuferwünsche in jeder Stadt?
Verkaeuer_Fakt(Kaeufer_Dim_id)
Kaeufer_Dim(Kaeufer_Dim_id, Vorname, Nachname)

- *Käufer:* Wie oft haben die Käufer einen bestimmten Objektart gekauft.
Verkaeuer_Fakt(Kaeufer_Dim_id, Kontaktsdatum, Objekte_Dim_id)
Kaeufer_Dim(Kaeufer_Dim_id, Vorname, Nachname)
Datum_Dim(Datum, Jahr, Monat, Tag)

- → *welche der o.g. Merkmale sind kaufentscheidend*

Data Quality⁷

Potenzielles Problem

Scope	Dirty Data	Reasons
Unzulässiger Wert	datum = 30.02.2050	Werte außerhalb des Admin Bereichs
Attributabhängigkeit verletzt	Notartermin > Kauftermin	Kommt je nachdem
Eindeutigkeit verletzt	(Ort = Berlin, PLZ = 123) (Ort =Mannheim, PLZ = 123)	Uniqueness für PLZ
Referentielle Integrität verletzt	name = "Peter Müller", idObjekte = "123"	referenzierte Objekte (123) nicht definiert
Fehlende Werte	Telefonnummer = +(49)15....	falsches Format oder zusätzliche Informationen
Schreibfehler	Preis = 23.49 oder 23,49	falsche Symbole
Kryptische Werte, Abkürzung	title = Doktor title = Dr.	verschiedene Einträge
Eingebettete Werte	Name = "Georges Spaninen"	mehrere Werte in einem Attribut
Falsche Zuordnung	Stadt = "Spanien" oder "Spain"	mehrere Sprachen
Widersprüchliche Werte	Stadt = "Berlin", PLZ=33333	Stadt und PLZ müssen abhängig sein
Transpositionen	Name = "Torben E." Firma = "Bosch T"	normalerweise in einem Freiformfeld
Duplikate	kunden1(name = "Peter Müller", ...) kunden2(name = "P. Müller", ...)	gleiche Kunden aufgrund einiger Dateneingabe Fehler zweimal vertreten oder Frau und Mann trägt unabhängig voneinander ein
Datenkonflikte	kunden1(name = "Peter Müller", Kauftermin = 12.06.2021) kunden2(name = "Peter	dieselbe reale Entität wird durch unterschiedliche Werte beschrieben

⁷ zum theoretischen Hintergrund siehe Artikel; E. Rahm, H. H. Do: Data Cleaning: Problems and Current Approaches. IEEE Techn. Bull. Data Eng., Dec. 2000

	Müller", Kauftermin = 17.06.2021)	
Falsche Referenz	kunden1(idObjekte =17, ...)	referenzierte primär Schlüssel Kunde (17) ist aber falsch definiert

Data Bereinigung

Datenanalyse

Probleme	Data	Ansatz
Unzulässiger Wert	Kardinalität	Falls Kardinalität > Erwartung >>>> Problem
Unzulässiger Wert	max, min	max, min sollte nicht außerhalb des zulässigen Bereichs liegen
Unzulässiger Wert	var, sd	Varianz, Abweichung der statistischen Werte sollte nicht höher als der Schwellenwert sein
Schreibfehler	Attribute	Beim Sortieren nach Werten werden oft falsch geschriebene Werte neben korrekten Werten angezeigt
Fehlende Werte	Nullwerte	Prozentsatz/Anzahl der Nullwerte
Fehlende Werte	Standardwerte	Die Standardwert kann darauf hinweisen, dass der tatsächliche Wert fehlt
Falsche Zuordnung	Attribute	Vergleichen des Attributwertesatzes einer Spalte einer Tabelle mit referenz Tabelle
Duplikate	Kardinalität	Einzigartigkeit von Kardinalität
Duplikate	Attribute	Sortieren von Werten nach der Häufigkeit des Auftretens; mehr als 1 Vorkommen weist auf Duplikate hin

Definieren von Datentransformationen

Der Datentransformationsprozess besteht normalerweise aus mehreren Schritten, wobei jeder Schritt schema- und instanzbezogene Transformationen (Mappings) durchführen kann. Damit ein Datentransformations- und Bereinigungssystem einen Transformationscode

generieren kann und somit den Umfang der Selbstprogrammierung reduziert, ist es notwendig, die erforderlichen Transformationen in einer geeigneten Sprache anzugeben. Verschiedene ETL-Tools bieten diese Funktionalität, indem sie proprietäre Regelsprachen unterstützen. Ein allgemeinerer und flexiblerer Ansatz ist die Verwendung der Standard Abfragesprache SQL, um die Datentransformationen durchzuführen und die Möglichkeit anwendungsspezifischer Spracherweiterungen zu nutzen

Konfliktlösungen

- **Attribute Splitting:** Freiformattribute erfassen oft mehrere Einzelwerte, die extrahiert werden sollten, um eine genauere Darstellung zu erreichen und weitere Bereinigungs-schritte wie Instanzabgleich und Duplikateliminierung zu unterstützen.
- **Validierung und Korrektur:** Vergleich mit Source data, und auf Eingabefehler und versucht diese soweit wie möglich automatisch zu korrigieren. Die Rechtschreibprüfung basierend auf der Wörterbuchsuche ist nützlich, um Rechtschreibfehler zu erkennen und zu korrigieren.
- **Standardisierung:** Attributwerte sollten in ein konsistentes und einheitliches Format umgewandelt werden. Beispielsweise sollten Datums- und Uhrzeiteinträge in ein bestimmtes Format gebracht werden; Namen und andere Zeichenfolgendaten sollten entweder in Groß- oder Kleinschreibung usw. umgewandelt werden. Textdaten können verdichtet und vereinheitlicht werden, indem Wortstamm gebildet, Präfixe, Suffixe und Stoppwörter entfernt werden. Darüber hinaus sollten Abkürzungen und Kodierungsschemata durchgängig durch Konsultieren spezieller Synonymwörterbücher oder Anwendung vordefinierter Konvertierungsregeln aufgelöst werden.

Möglichkeiten zur Vermeidung von Fehler

Fehlervermeidungsstrategie	Vermeidung von....
DropDown Menüs	Unzulässiger Wert, Schreibfehler, kryptische Werte, Abkürzungen, falsche Zuordnung, Ort, Anrede
Pflichtfelder	Fehlende Werte, Kontaktdaten
einige Einschränkung definieren	Schreibfehler, Telefonnummer
Beispiel eines gewünschten Datensatz	Unzulässiger Wert, Falsche Zuordnung

z.B. in grau hinterlegt	Email Adresse, Telefonnummer
Korrekte Datensätze verwenden, einkaufen	falschen Daten

Tabellen Datenanalyse

Der Kunde stellte einen Beispieldatensatz "Auszug Tabelle Verkäufer.xlsx" zur Verfügung.

Anrede	Vorname	Nachname	Ansprechpartner	PLZ	Straße	Hausnummer	Ort	Tel FN	Mobil 1	e-mail	Art des Objektes	Preis	Größe in qm	Ort	Exposé
Familie	Eisenecker			68723	Alhornweg	4	Mannheim	284657372			1 FH	256000	88	Mannheim	ja
Familie	Madani			68199	Maxstr.		1Erfurt	937462526			2-3 FH	275000	122	Heidelberg	ja
Familie	Emmerich			69118	Rainweg	4	Ladenburg	947362637			MFH	294000	156	Hamburg	ja
Firma	Engel			68161	Pf	7	Brühl	334837626			2 FH	313000	190	Dresden	ja
Firma	Erbes			70372	Neuhelmer Str.	10	Ludwigshafen	336902259			2-3 FH	332000	224	Berlin	ja
Firma	Erbus			68167	Friedrich-Franckhorn-Str.	13	Heidesheim	337542801			MFH	351000	258	Karlsruhe	ja

Anhand der oben beschriebenen Strategien haben wir die vom Kunden erhalten Daten analysiert. Dabei fielen folgende Problem auf:

- Fehlende Primärschlüssel (bisher nicht vorhanden)
- Vorname, PLZ: fehlende Daten (z.B. Frau Juliane Bilger hat keine PLZ)
- Vorname: Schreibfehler, Abkürzung (z. B Frau Aydemir hat Ihre Name als P. geschrieben)
- Ansprechpartner: fehlende Daten, Datenintegrität (die meisten sind leer)
- Straße: Formatfehler (z.B. Frau Juliane Bilger Straße: "Albert")
- Falls Firma, gibt es Problem mit Vor und Nachname
- Ansprechpartner Partner und Email sind leer
- Es gibt zwei Spalte -> Ort (der Ort für Kunden und Objekt als Ort benannt)
- Für Tel FN, ein entry ist weniger als anderen, deswegen kann es so sein, dass User nur 9 zahlen eingeben soll
- Preis und Größe in qm sollen Ausreißer Daten überprüft werden
- Duplikate, Schreibfehler können logisch überprüft werden (z.B Sortieren von Werten wie Nachname, Ort)

Ausblick, was sind die nächsten Schritte im Projekt

- Einpflegen aller bestehenden Daten
- Andere Währung wie CHF vorsehen
- Einführung einer Prozessnummer für bessere Nachverfolgbarkeit/Analyse
- Verfeinerung der Datentypen (nach Rücksprache mit Kunden)
- Aufbau einer GUI
- UML Modellierung
- Testphase
- Rollout
 - Schulung
 - Dokumentationen
- Rückmeldung / Verbesserung
- Mitarbeiterdaten ausbauen, hinzufügen, vereinheitlichen