

End-Abgabe: Smart City IoT Projekt

Business Database – Struktur und Design

11. November 2025

1.1 Projektbeschreibung

Ziel des Projekts ist der Aufbau eines integrierten Datenmodells für ein Smart City IoT System, das Energieverbrauch, Wetterdaten, Sensorinformationen und Marktpreise zentral zusammenführt. Das System soll eine effiziente Analyse von Energieverbrauch, Netzlast und Einflussfaktoren ermöglichen. Die Daten stammen aus CSV-Dateien und werden in einer relationalen MySQL-Datenbank gespeichert. Später werden ETL-Prozesse (Extract, Transform, Load) und Data-Warehouse-Analysen auf dieser Struktur aufbauen.

Erste Projektschritte:

- Erstellung der Projektbeschreibung und Definition der Ziele.
- Entwicklung des ER-Modells zur Abbildung logischer Beziehungen.
- Überführung des Modells in ein Relationenmodell (3NF) für eine saubere Normalisierung.
- Vorbereitung auf spätere DWH-Fragen (Analysen auf Fakten- und Dimensionstabellen).

1.2 Business DB – Struktur und Design

ER-Diagramm – Logische Struktur

Das Entity-Relationship-Diagramm (Abbildung 1) stellt die logische Grundstruktur der Smart City IoT Datenbank dar. Es zeigt die wichtigsten Entitäten: *EnergieAblesung*, *Sensor*, *Zeit*, *Wetter*, *MarketInfo* und *Lokation*. Das Ziel des Modells ist es, die Datenflüsse zwischen Sensoren, Messwerten, Wetterdaten und Marktinformationen transparent darzustellen und die logischen Abhängigkeiten zwischen diesen Entitäten abzubilden.

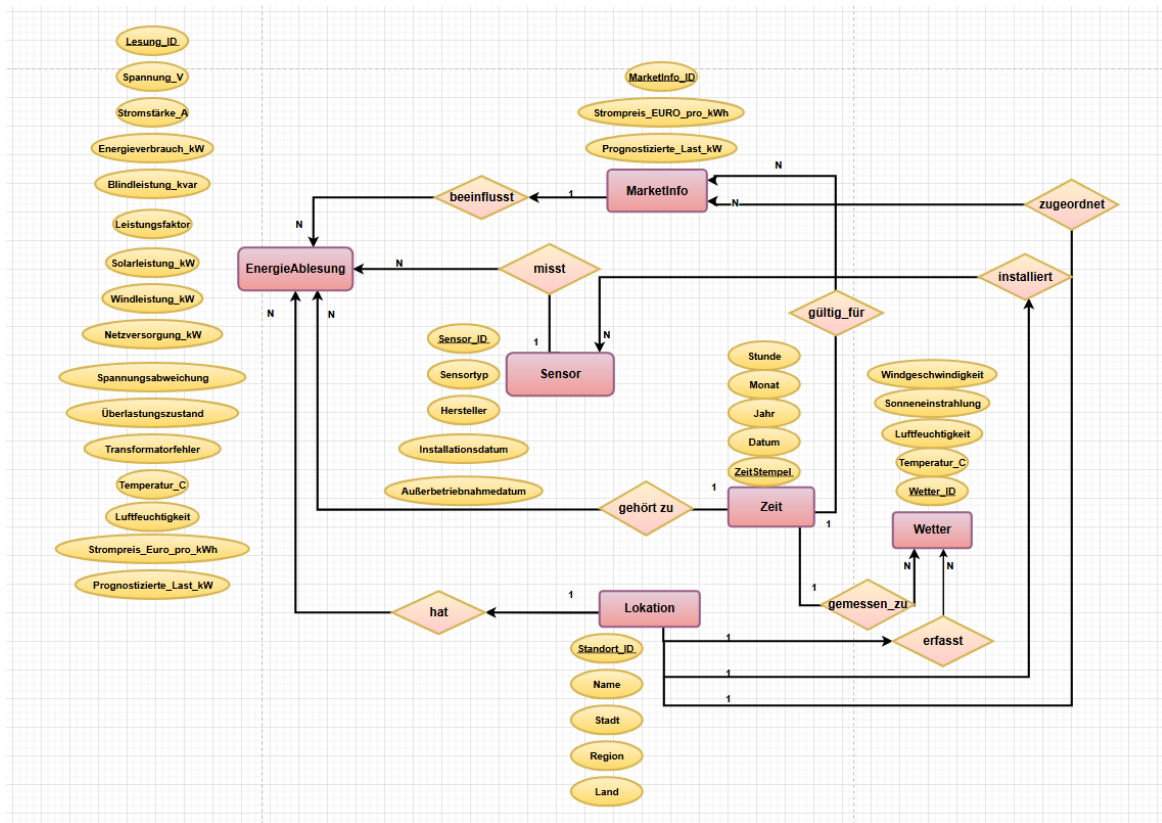


Abbildung 1: ER-Diagramm – Grundstruktur der Datenbank

ER-Diagramm mit Beziehungen und Kardinalitäten

In der erweiterten Version (Abbildung 2) werden zusätzlich die Beziehungen und Kardinalitäten dargestellt. Die Kardinalitäten zeigen, wie viele Datensätze der jeweiligen Entitäten miteinander in Beziehung stehen (z. B. Standort → Sensor: 1:N).

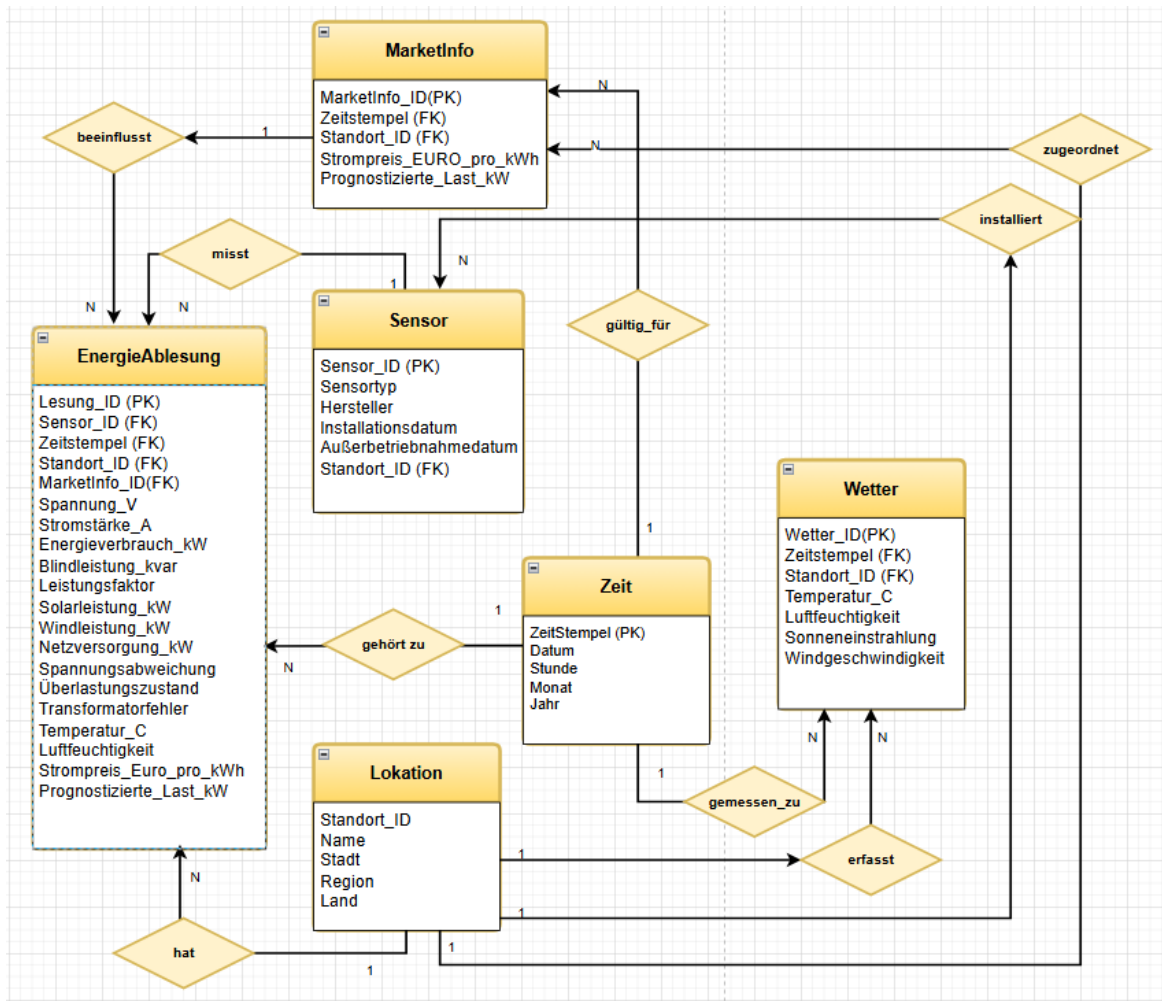


Abbildung 2: ER-Diagramm – Version mit Beziehungen und Kardinalitäten

Objektbeschreibung der Tabellen

Die Objektbeschreibung dokumentiert die logische Struktur, Primär- und Fremdschlüssel sowie die wichtigsten Attribute aller Tabellen des Datenmodells. Jede Tabelle wurde im Rahmen des Object-Push-Cycle so konzipiert, dass sie einen klar abgegrenzten fachlichen Zweck erfüllt und sich nahtlos in die Integrations- und Analysestrecke einfügt.

Tabelle: Lokation Primärschlüssel: Standort_ID. – Dimension mit geografischen Informationen (Name, Stadt, Region, Land).

Tabelle: Zeit Primärschlüssel: Zeitstempel. – Zeitdimension für stundengenaue Analysen (Datum, Stunde, Monat, Jahr).

Tabelle: ZeitTeil Primärschlüssel: ZeitTeil_ID, FK: Zeitstempel. – Detaillierte Zeitmerkmale (Jahr, Monat, Stunde, Quartal, Wochentag).

Tabelle: Sensor Primärschlüssel: `Sensor_ID`, FK: `Standort_ID`. – Technische Metadaten zu installierten Sensoren.

Tabelle: Wetter Primärschlüssel: `Wetter_ID`, FK: `Zeitstempel`, `Standort_ID`. – Meteorologische Messwerte (Temp., Luftfeuchte, Sonneneinstrahlung, Wind).

Tabelle: MarketInformation Primärschlüssel: `MarketInfo_ID`, FK: `Zeitstempel`, `Standort_ID`. – Strompreis und prognostizierte Last.

Tabelle: EnergieAblesung Primärschlüssel: `Lesung_ID`, FK: `Sensor_ID`, `Zeitstempel`, `Standort_ID`, `MarketInfo_ID`. Zentrale Faktentabelle mit elektrischen Messgrößen (Spannung, Stromstärke, Verbrauch, Blindleistung, Leistungsfaktor, ...) sowie Marktattributen.

Normalisierung (Normalisierungsgrad)

1. Normalisierung – Erste Normalform (1NF)

Die erste Normalform (1NF) verlangt, dass alle Attribute in einer Tabelle **atomar** sind, also nur einen einzelnen Wert enthalten. Mehrwertige oder wiederholende Gruppen sind nicht erlaubt. Jede Tabelle muss einen eindeutigen Primärschlüssel besitzen, der jede Zeile eindeutig identifiziert.

In unserem Datenmodell (*Smart City IoT System*) erfüllen alle Tabellen diese Bedingungen automatisch:

- Alle Attribute sind eindeutig und enthalten nur einen Wert (z. B. eine Temperatur, eine Spannung, einen Stromwert usw.).
- Es gibt keine wiederholenden Spalten (wie `Temperatur_1`, `Temperatur_2` ...).
- Jede Tabelle besitzt einen klar definierten Primärschlüssel (z. B. `Standort_ID`, `Sensor_ID`, `Lesung_ID`).

Da keine mehrwertigen Attribute oder Gruppen vorkommen, liegt das gesamte Modell bereits in der **ersten Normalform (1NF)** vor. Es besteht somit kein Verstoß gegen 1NF, und eine zusätzliche Anpassung ist nicht erforderlich.

2. Normalisierung – Zweite Normalform (2NF)

Die zweite Normalform (2NF) verlangt, dass eine Relation bereits in der ersten Normalform (1NF) ist und dass jedes Nicht-Schlüsselattribut vollständig funktional vom gesamten Primärschlüssel abhängt. Das bedeutet, dass es keine partiellen Abhängigkeiten geben darf – also kein Attribut darf nur von einem Teil eines zusammengesetzten Primärschlüssels abhängen.

In unserem Datenmodell (*Smart City IoT System*) besitzt jedoch keine Tabelle einen zusammengesetzten Primärschlüssel. Alle Tabellen, wie z. B. `Lokation`, `Zeit`, `Sensor`, `Wetter`, `MarktInformation`

und **EnergieAblesung**, haben jeweils einen einfachen Primärschlüssel (z. B. **Standort_ID**, **Zeitstempel**, **Sensor_ID**, **Lesung_ID** usw.).

Daher hängen alle Nicht-Schlüsselattribute vollständig von ihrem jeweiligen Primärschlüssel ab, und es existieren keine partiellen funktionalen Abhängigkeiten. Das bedeutet, dass sich das gesamte Datenmodell bereits in der **zweiten Normalform (2NF)** befindet und keine weitere Aufteilung oder Anpassung erforderlich ist.

3. Normalisierung – Dritte Normalform (3NF)

Eine Relation befindet sich in der dritten Normalform (3NF), wenn sie bereits in der 2NF ist und keine **transitive Abhängigkeit** zwischen Nicht-Schlüsselattributen besteht. Das bedeutet, dass jedes Nicht-Schlüsselattribut direkt vom Primärschlüssel (*Primary Key*) abhängen muss und nicht über ein anderes Nicht-Schlüsselattribut.

In der ursprünglichen Tabelle **Zeit** gab es jedoch eine transitive Abhängigkeit:

$$Datum \rightarrow (Jahr, Monat, Stunde)$$

Dadurch war ein Verstoß gegen 3NF vorhanden, weil **Jahr**, **Monat** und **Stunde** indirekt von **Datum** und nicht direkt vom Primärschlüssel **Zeitstempel** (PK) abhingen.

Um diese Abhängigkeit zu entfernen, wurde **Zeit** in zwei Tabellen aufgeteilt:

1. **Zeit** – enthält den Primärschlüssel (**Zeitstempel**) und optional das Datum.
2. **ZeitTeil** – steht in einer 1:1-Beziehung zu **Zeit** und enthält die Attribute **Jahr**, **Monat**, **Stunde**, **Quartal** und **Wochentag**.

Damit hängen alle Attribute wieder direkt vom Primary Key ab, und das Datenmodell ist vollständig **3NF-konform**. Somit wurde die Datenintegrität gewährleistet und Redundanzen wurden erfolgreich eliminiert.

Fazit: Das Smart City IoT Datenmodell erfüllt sämtliche Anforderungen der ersten, zweiten und dritten Normalform. Durch diese Normalisierung wird sichergestellt, dass alle Tabellen konsistent, effizient und redundantfrei strukturiert sind.

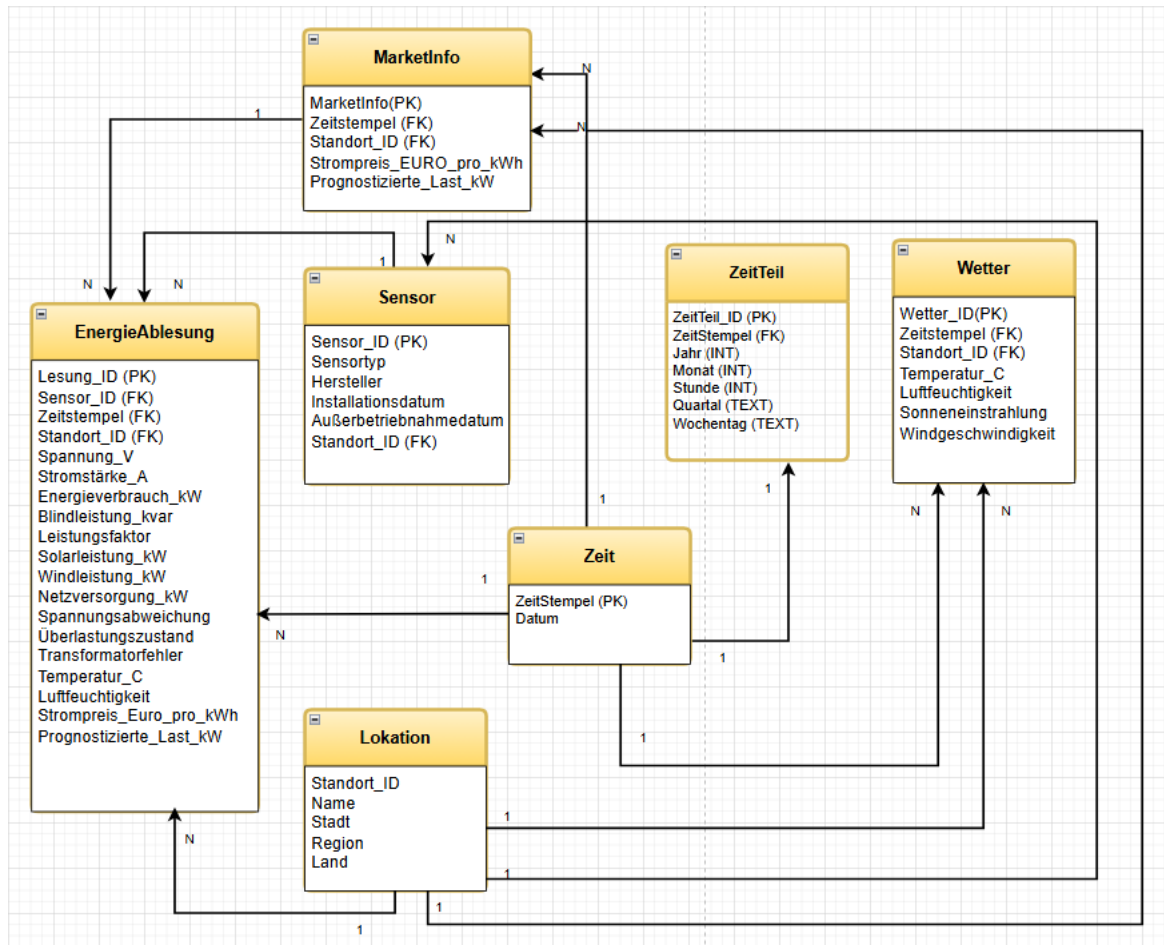


Abbildung 3: Relationales Modell (3NF) mit Aufteilung Zeit/ZeitTeil

Verbindung zum Data Warehouse

Das 3NF-konforme Modell bildet die stabile Grundlage für das **Data Warehouse**. Die im Abschnitt **1.3 DWH Fragen** definierten Analysen greifen auf die Faktentabelle *EnergieAblesung* und die zugehörigen Dimensionen zu und ermöglichen effiziente Zeitreihen-, Standort- und Preisanalysen.

Datum: 11. November 2025

1.3 DWH Fragen – Beispielfragen für Analysen

DWH Fragen

1. **Energieverbrauch pro Standort und Tag** *Beschreibung:* Ermittlung des täglichen Gesamtenergieverbrauchs je Standort zur Analyse des regionalen Verbrauchsverhaltens. *Verwendete Tabellen:* EnergieAblesung, Zeit, Lokation *KPI:* SUM(Energieverbrauch_kW)
2. **Spitzenlast pro Standort (Peak Hour)** *Beschreibung:* Bestimmung der Stunde mit der höchsten Last je Standort, um Engpässe und Lastspitzen zu identifizieren. *Verwendete Tabellen:* EnergieAblesung, Zeit *KPI:* MAX(Energieverbrauch_kW)
3. **Einfluss der Temperatur auf den Energieverbrauch** *Beschreibung:* Analyse des Zusammenhangs zwischen Temperatur und Verbrauch zur Bewertung wetterbedingter Einflüsse. *Verwendete Tabellen:* EnergieAblesung, Wetter *KPI:* Korrelation(Temperatur_C, Energieverbrauch_kW)
4. **Verhältnis erneuerbarer Energien zur Gesamtlast** *Beschreibung:* Bestimmung des Anteils von Solar- und Windenergie am gesamten Energieverbrauch je Standort und Tag. *Verwendete Tabellen:* EnergieAblesung *KPI:* (Solarleistung + Windleistung) / Gesamtverbrauch
5. **Genauigkeit der Marktprognose** *Beschreibung:* Vergleich der prognostizierten und tatsächlichen Last zur Ermittlung der Prognosegüte. *Verwendete Tabellen:* MarktInformation, EnergieAblesung *KPI:* Prognoseabweichung (%)
6. **Überlastungsanalyse pro Standort** *Beschreibung:* Zählung der Überlastungszustände zur Bewertung der Netzstabilität und Identifikation möglicher Risiken. *Verwendete Tabellen:* EnergieAblesung *KPI:* COUNT(Überlastungszustand)
7. **Durchschnittlicher Strompreis pro Tag** *Beschreibung:* Analyse des durchschnittlichen Strompreises in Verbindung mit der täglichen Energieabnahme. *Verwendete Tabellen:* MarktInformation, Zeit, Lokation *KPI:* AVG(Strompreis_Euro_pro_kWh)
8. **Wetterabhängige Verbrauchsanalyse** *Beschreibung:* Vergleich des Verbrauchs unter unterschiedlichen Wetterbedingungen zur Modellierung von Verbrauchsmustern. *Verwendete Tabellen:* EnergieAblesung, Wetter *KPI:* Energieverbrauch unter Temperatur-/Windbedingungen
9. **Standortranking nach Gesamtlast** *Beschreibung:* Ermittlung der Standorte mit dem höchsten Gesamtverbrauch zur Priorisierung von Effizienzmaßnahmen. *Verwendete Tabellen:* EnergieAblesung, Lokation *KPI:* SUM(Energieverbrauch_kW)
10. **Jahres- und Monatsvergleich des Verbrauchs** *Beschreibung:* Vergleich des Gesamtverbrauchs zwischen Monaten und Jahren zur Erkennung von Trends und saisonalen Mustern. *Verwendete Tabellen:* EnergieAblesung, Zeit *KPI:* SUM(Energieverbrauch_kW) pro Monat/Jahr

Zusammenfassung

Die definierten DWH-Fragen bilden die Grundlage für Auswertungen, Dashboards und Berichte im Smart City IoT System. Sie ermöglichen:

- die Überwachung des Energieverbrauchs und der Effizienz pro Standort,

- die Bewertung von Wetter- und Marktbedingungen,
- die Erkennung von Spitzenlasten und Überlastungsrisiken,
- die Verbesserung von Prognosen und Energieoptimierung.