# Justification for Data Modeling: OLTP to OLAP Star Schema

This document justifies the transformation of the original OLTP data schema into a star schema for OLAP purposes. The business goal is to enable efficient analysis of ride and payment data from Divvy bike trips based on various dimensions.
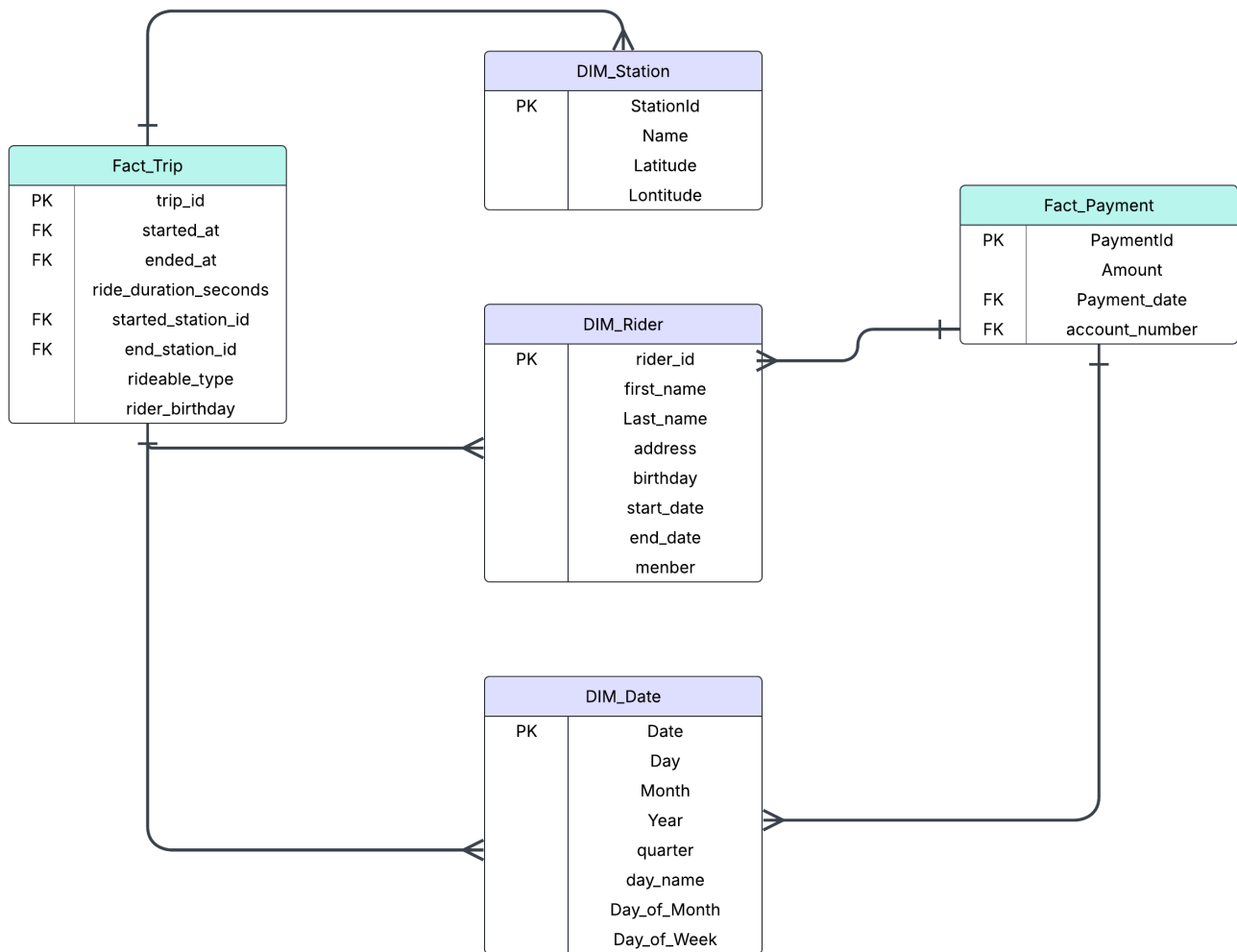
Business Requirements

- Analyze ride duration based on date/time, stations, rider age, and membership type

- Analyze money spent per month/quarter/year and per member by age

- Extra: Analyze spending per member based on ride frequency and time spent on bike

Original OLTP Schema
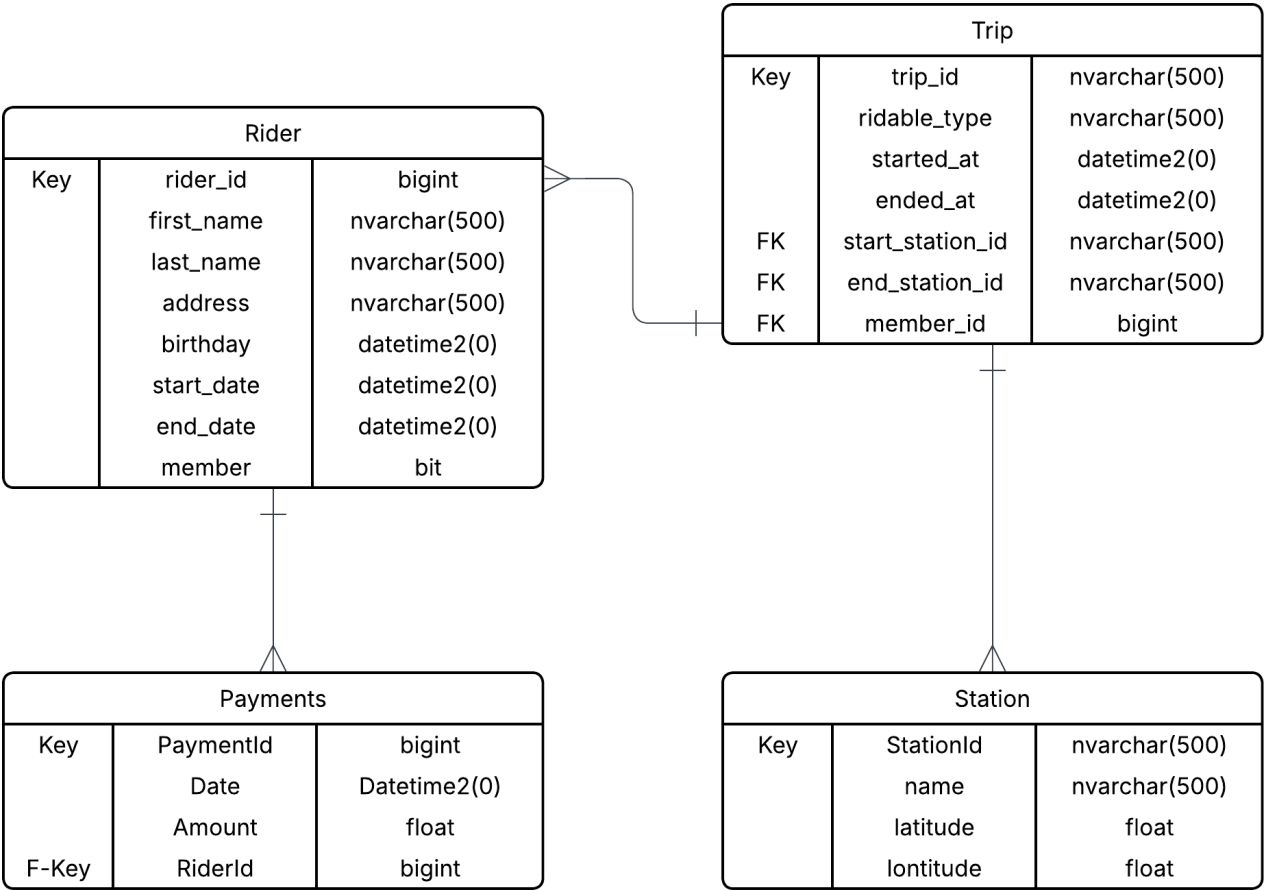
The original schema is normalized, designed for transaction processing. It has high granularity and multiple tables: Rider, Trip, Payments, and Station. Relationships between these entities enable accurate transactional operations, but are less efficient for analytical queries due to complexity and joins.

Star Schema for OLAP

DIM_Station

PK	StationId
	Name
	Latitude
	Lontitude

Fact_Trip

PK	trip_id
FK	started_at
FK	ended_at
	ride_duration_seconds
FK	started_station_id
FK	end_station_id
	rideable_type
	rider_birthday

Fact_Payment

PK	PaymentId
	Amount
FK	Payment_date
FK	account_number

DIM_Rider

PK	rider_id
	first_name
	Last_name
	address
	birthday
	start_date
	end_date
	menber

DIM_Date

PK	Date
	Day
	Month
	Year
	quarter
	day_name
	Day_of_Month
	Day_of_Week

The new schema uses a fact-dimension model for analytical efficiency. Central to the model are the fact tables:

- Fact_Trip: contains ride duration and links to date, station, and rider dimensions

- Fact_Payment: contains payment amounts and links to dates and riders

Supporting dimensions:

- DIM_Date: supports temporal analysis (day, month, quarter, etc.)

- DIM_Station: enables spatial analysis of ride patterns

- DIM_Rider: supports demographic and membership analysis

Alignment with Business Goals

This schema enables the following OLAP capabilities:

- Analyze ride duration across days, hours (via DIM_Date)

- Filter rides by start/end stations (via DIM_Station)

- Segment riders by age at ride time and membership status (via DIM_Rider)

- Aggregate payments by time and rider attributes (via Fact_Payment and DIM_Date)

- Extra credit: calculate per-member spending normalized by average rides/minutes per month