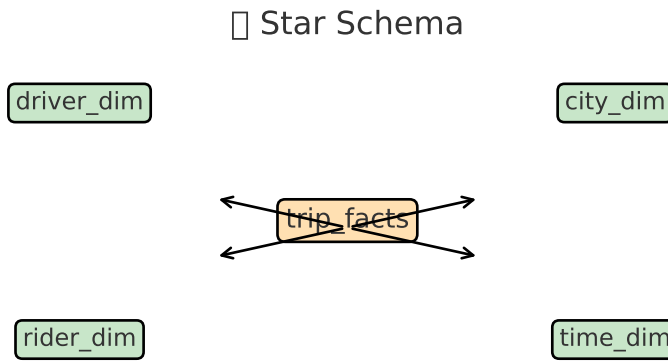
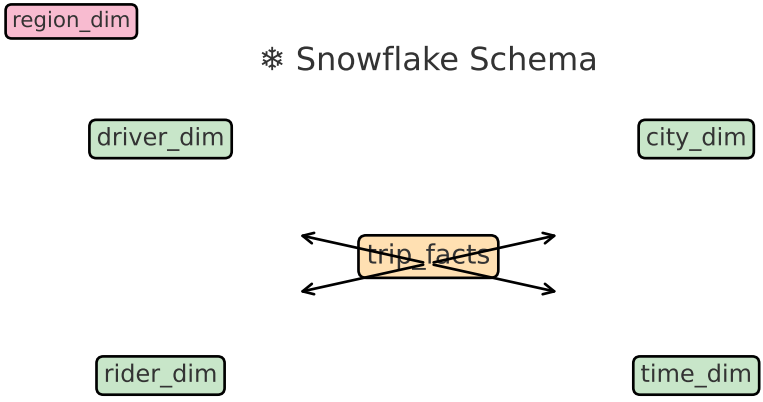
 Data Modeling for Analytics — Interview-Ready Notes Purpose of Data Modeling To design structured data systems that support analytical needs, especially for OLAP use cases like dashboards, KPIs, trend analysis, and ML features. OLAP Concepts: Fact & Dimension Tables Fa Table - Contains measurable, numeric data (KPIs, events) - Primary table in a star/snowflake schem - Examples: sales, trips, payments, logins - Characteristics: - Stores metrics (e.g. amount, distance, duration)   - Has foreign keys linking to dimension tables   - Grows rapidly and gets frequent inserts   - Often partitioned by date for performance Dimension Table - Describes attributes related to fact table - Examples: drivers, riders, cities, time, vehicle_types - Characteristics: - Changes slowly (hence, Slow Changing Dimensions)   - Typically small in size compared to fact tables   - Joins via primary → foreign key to fact tables Star Schema - 1 central fact table - Surrounding denormalized dimension tables - Fast for querying, simpler design ❄ Snowflake Schema - Dimensions are normalized (split into sub-dimensions) - Slightly more comple joins - Better for storage efficiency, data integrity Galaxy Schema - Multiple fact tables shar dimension tables - Used in multi-business models or shared context SCD — Slow Changing Dimensio SCD Type | Description | Keeps History? | Example
-- SCD 0 | Fixed — never changes | Yes | Countr codes, ISO standards SCD 1 | Overwrite old data | No | Update driver rating directly SCD 2 | Add n row + version | Yes | Change in driver's address, rating SCD 3 | Add new column | Partial | Keep o & new pricing tiers SCD 4 | Two tables: current + history | Yes | Active vs Archived offers Hybrid Mix of SCD1 + SCD2 | Yes | Contact update (SCD1), Rating history (SCD2) Tips for Interviews - U real-world analogies: Uber, Amazon, Flipkart - Explain why modeling matters: scalable analytics, change tracking (SCD), performance - Mention terms like: dimensional modeling, grain, primary keys joins, aggregation


⭐ Star Schema
driver_dim
city_dim
trip_facts
rider_dim
time_dim


region_dim
❄ Snowflake Schema
driver_dim
city_dim
trip_facts
rider_dim
time_dim