Beginning Dimensional Modeling

Joe Cline
SENIOR DATA ENGINEER & MODELER
@d8ajoe www.linkedin.com/in/josephcline

In This Module

Why have a separate model for business intelligence and analytics?

Basic terminology and concepts

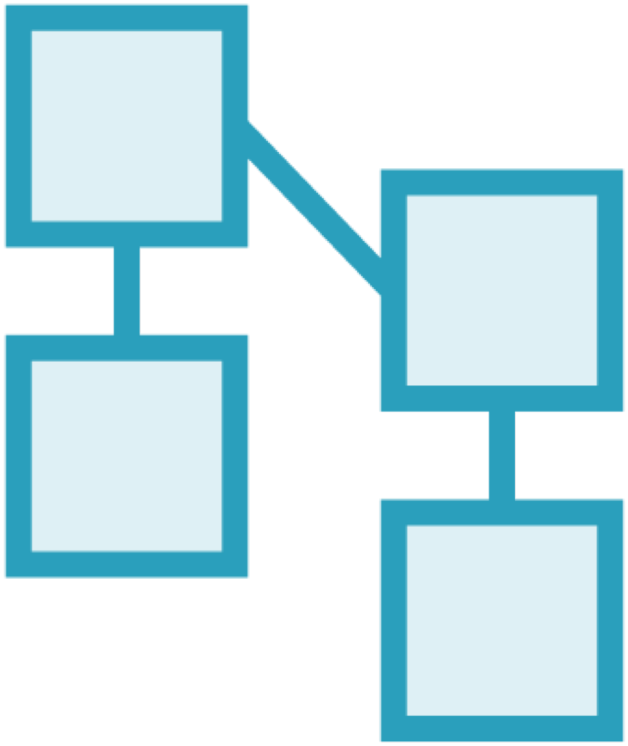
Key performance indicators (KPIs)

Demonstration

- Create a dimensional model

Video 2 starts here (placeholder)

The Current Relational Data Model

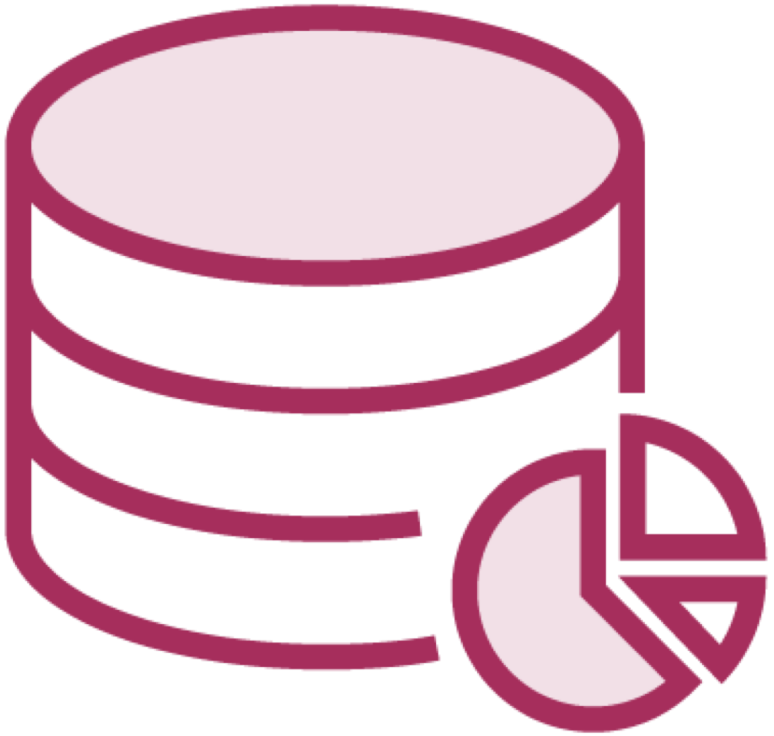


Hotel data

Guest data

The guest's stay data

Future Requirements



Perform EDA (Exploratory Data Analysis)

Hotel performance reports

Rates and inventory optimization

Write Intensive Operations

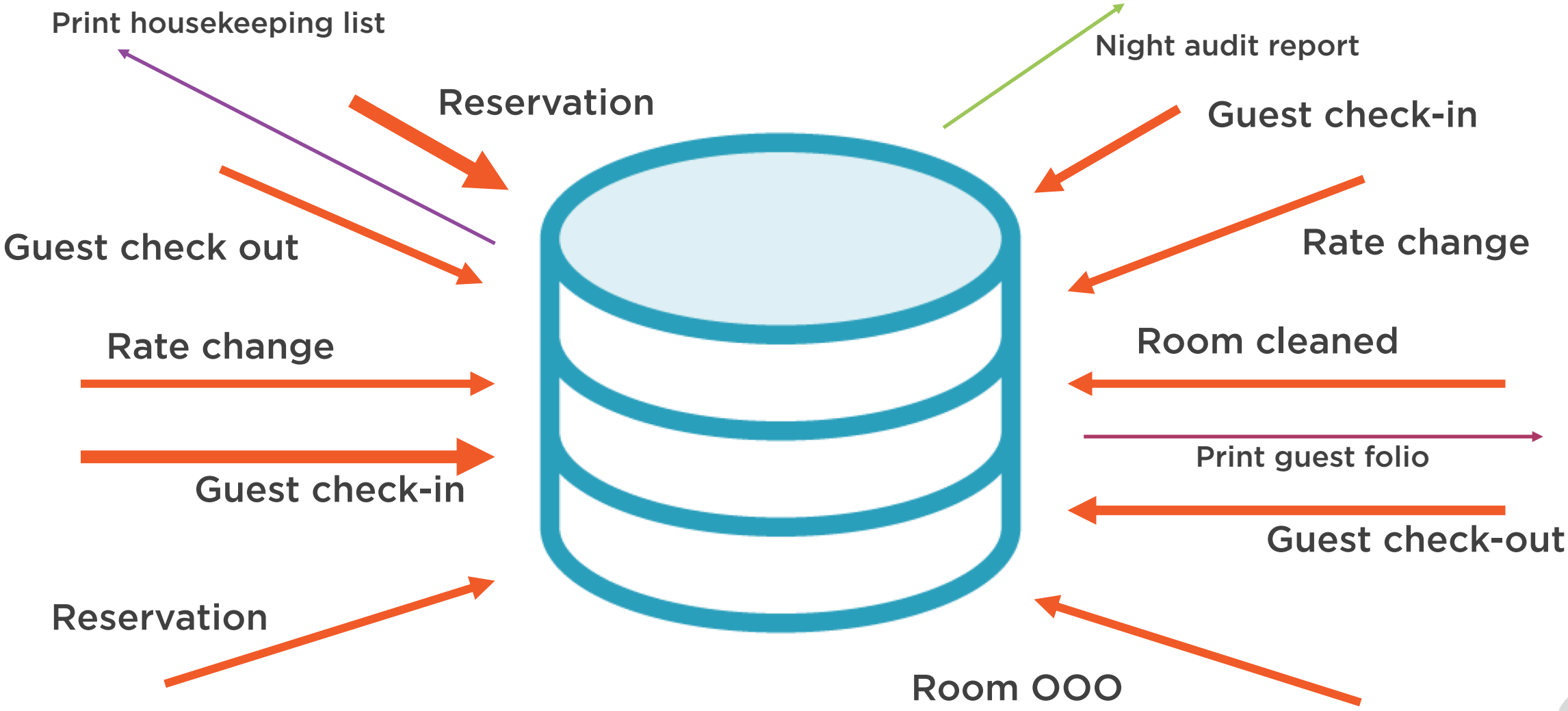
Inserting new guest rows

Updating existing guest rows

Taking rooms in and out of inventory

Making rate changes across dates

Highly Write Intensive



Online Transactional Processing

Also known as OLTP, is a type of database design for applications with a higher write transaction to read transaction ratio

Read Intensive

Analytics and reporting are read intensive

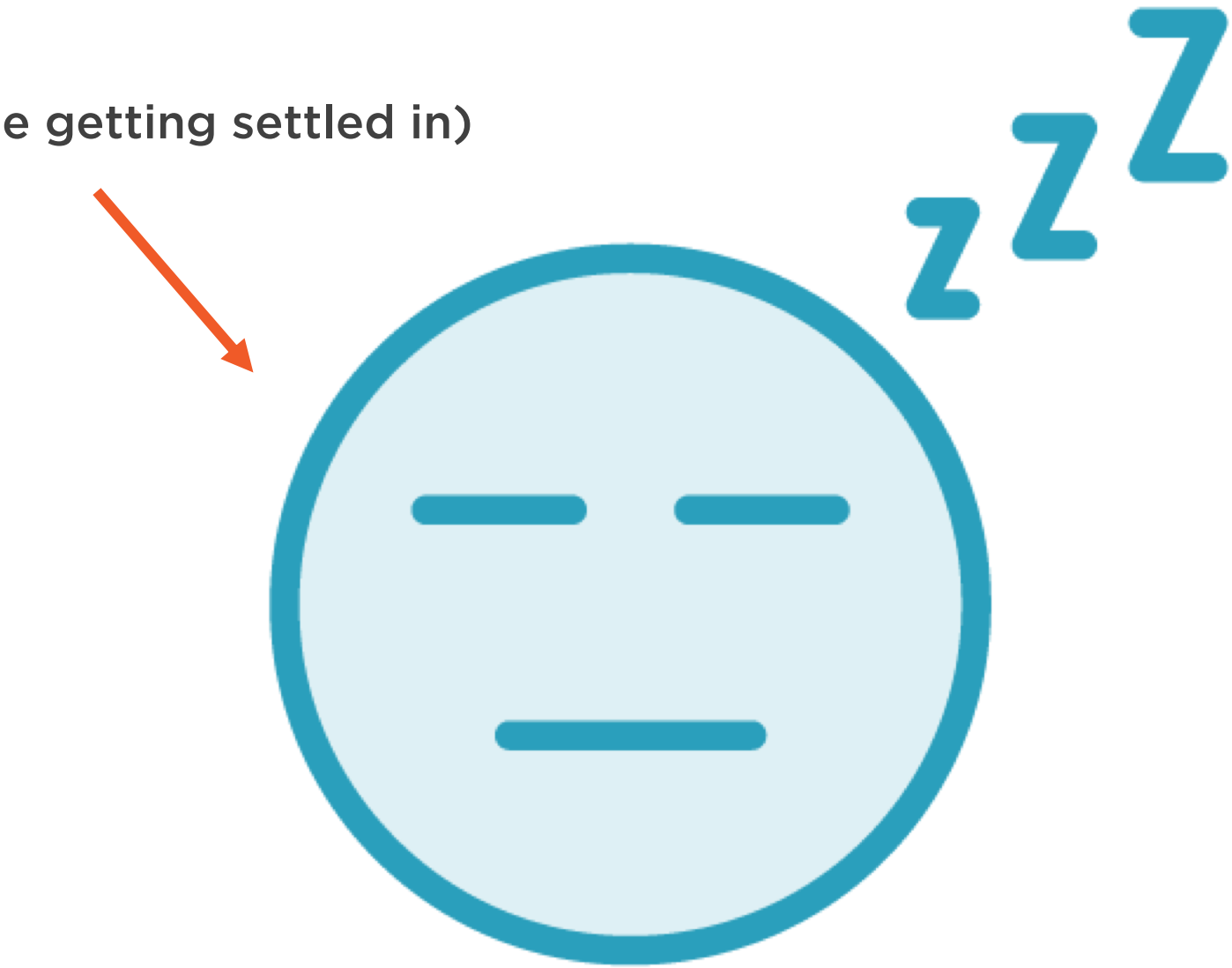
Data gets pulled into memory

Aggregate operations performed

Slows down OLTP databases

Waiting sucks!

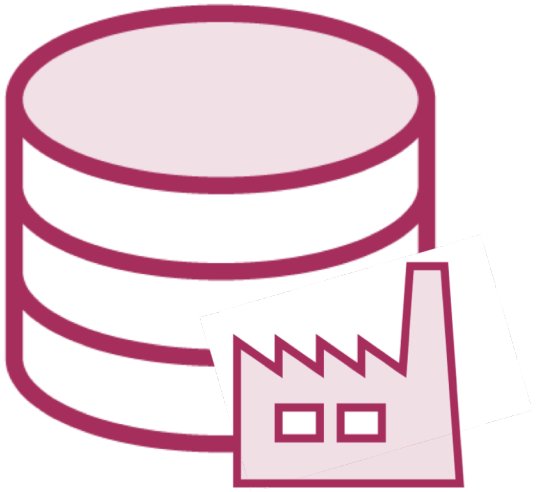
(This is me getting settled in)



Decision Support System

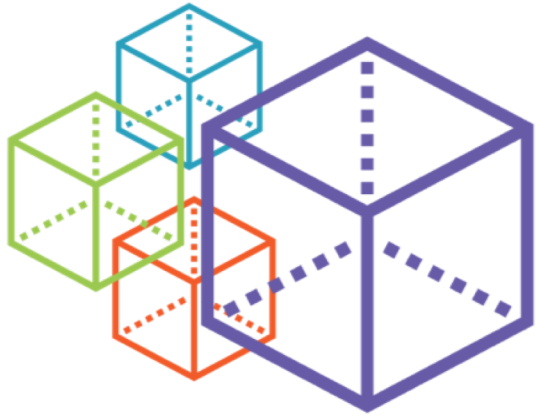
Also known as DSS, is a computational system configured and designed for analytical processes to support the business decision make process

Examples of Decision Support Systems



Data Warehouse

Top-down; traditional; been around since 80s



OLAP Cubes

Bottom-up; purpose specific; a logical data warehouse

Data Lake

A result of big data and Hadoop,;raw data; ad-hoc analysis

It’s the data, stupid.

(Just kidding, you are *so* smart)

Dimensional Data Model

A data model designed with fact and dimension tables used for analytical processing in decision support systems

Video 3 starts here (Placeholder only)

Terminology and Concepts

Facts, measures, and fact tables

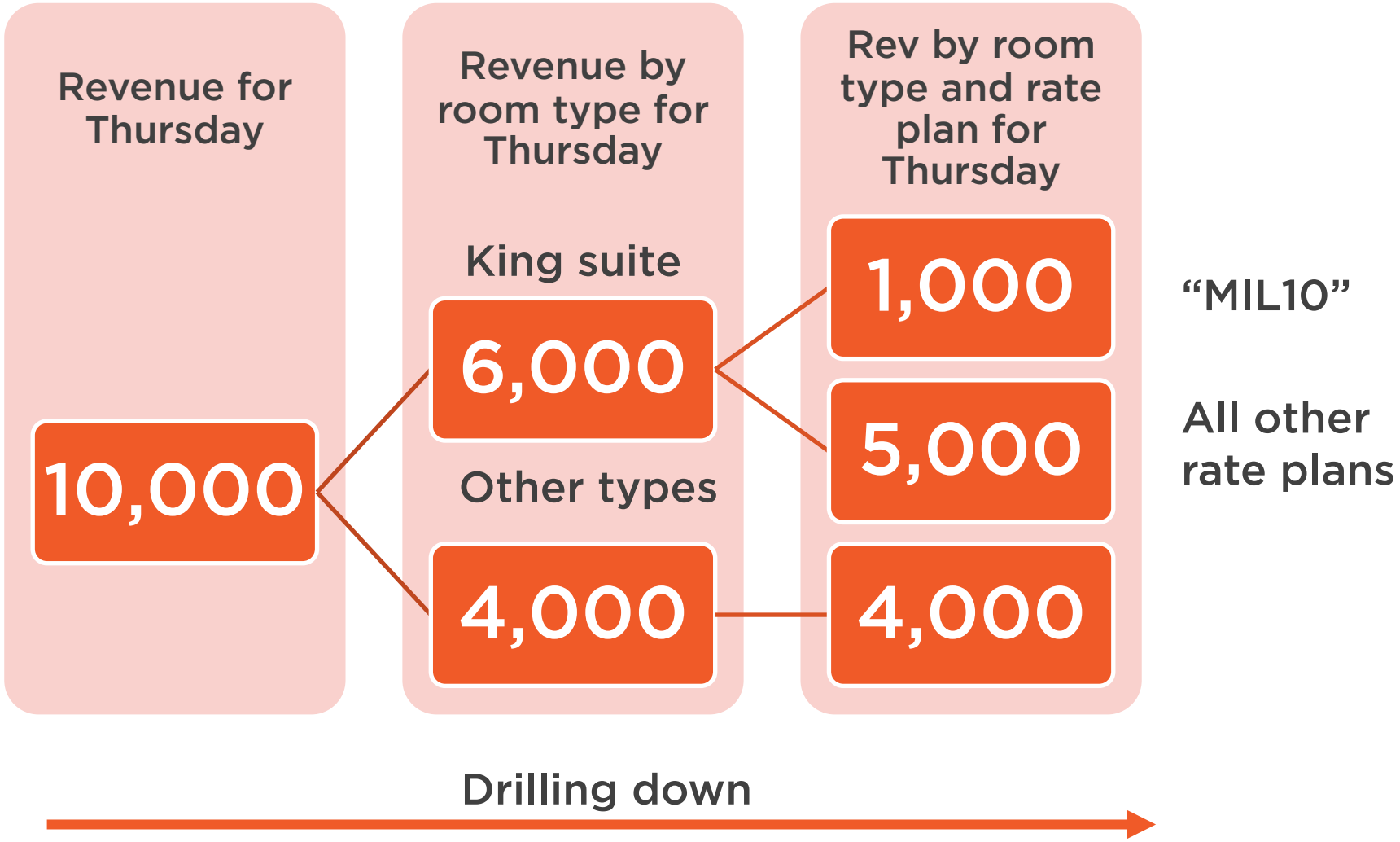
Dimensions, dimension tables, and the special Time dimension

Star and snowflake schema

Metrics and KPIs

Slicing, granularity and drilling down

An Example of Granularity



Video 4 starts here (Placeholder only)

Terminology and Concepts

OLAP cubes

Slowly changing dimensions

Dimensional hierarchy

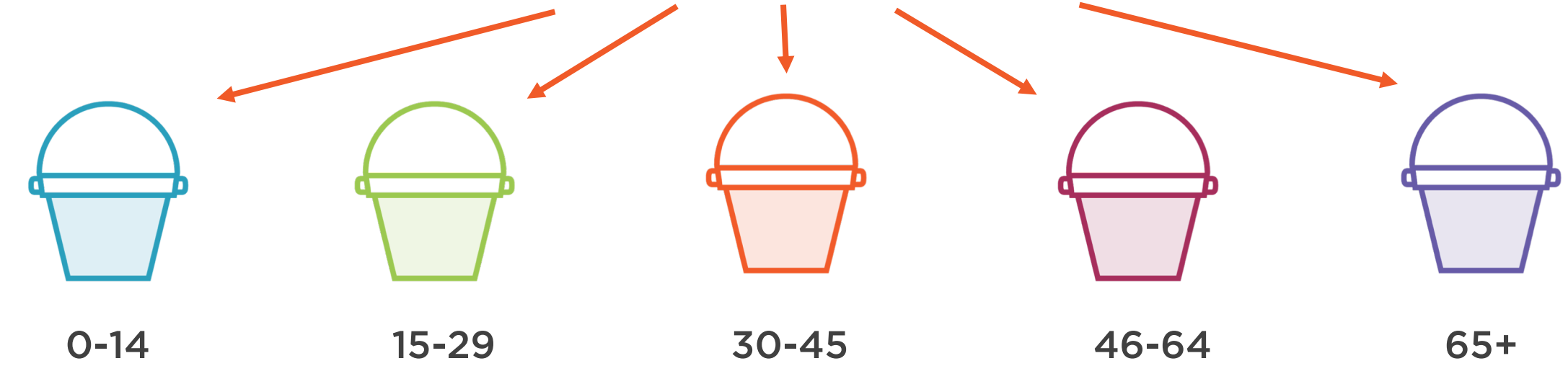
More on drilling down

Online Analytical Processing

Also known as OLAP, is a type of database design for applications with a higher read transaction to write transaction ratio; used in a DSS

Slowly Changing Dimension: Type 0

First name	Last name	Birthdate	Address	City...
Jim	Doe	1/1/1970	123 E. Street	Anytown



Slowly Changing Dimension: Type I

Timestamp	First name	Last name	Birthdate	Address	City
2-17-2010 01:01:00	Jim	Doe	1-1-1970	123 E. Street	Anytown

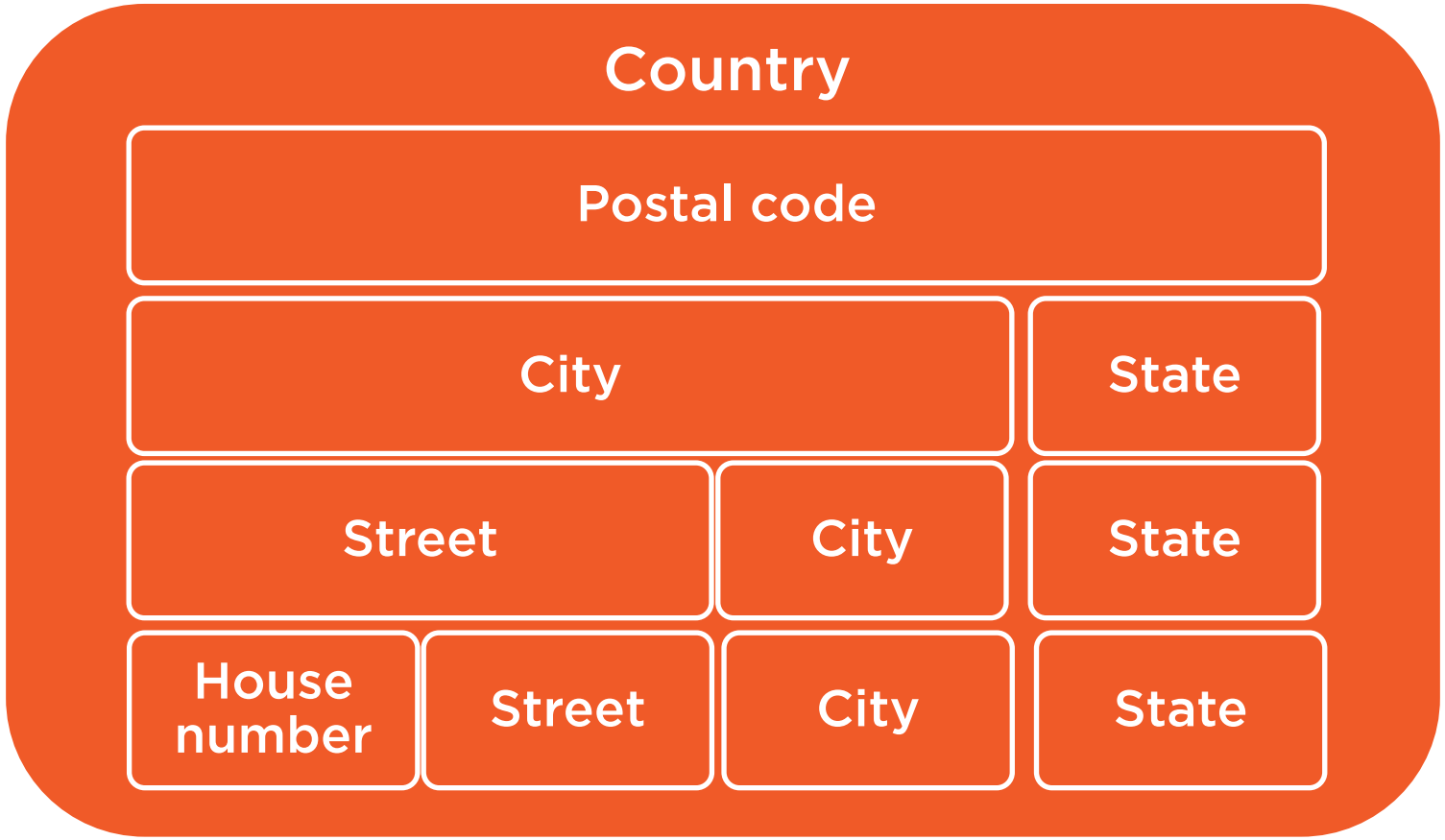
Slowly Changing Dimension: Type I

Timestamp	First name	Last name	Birthdate	Address	City
6-23-2014 01:04:00	Jim	Doe	1-1-1970	456 W. Blvd	Otherville

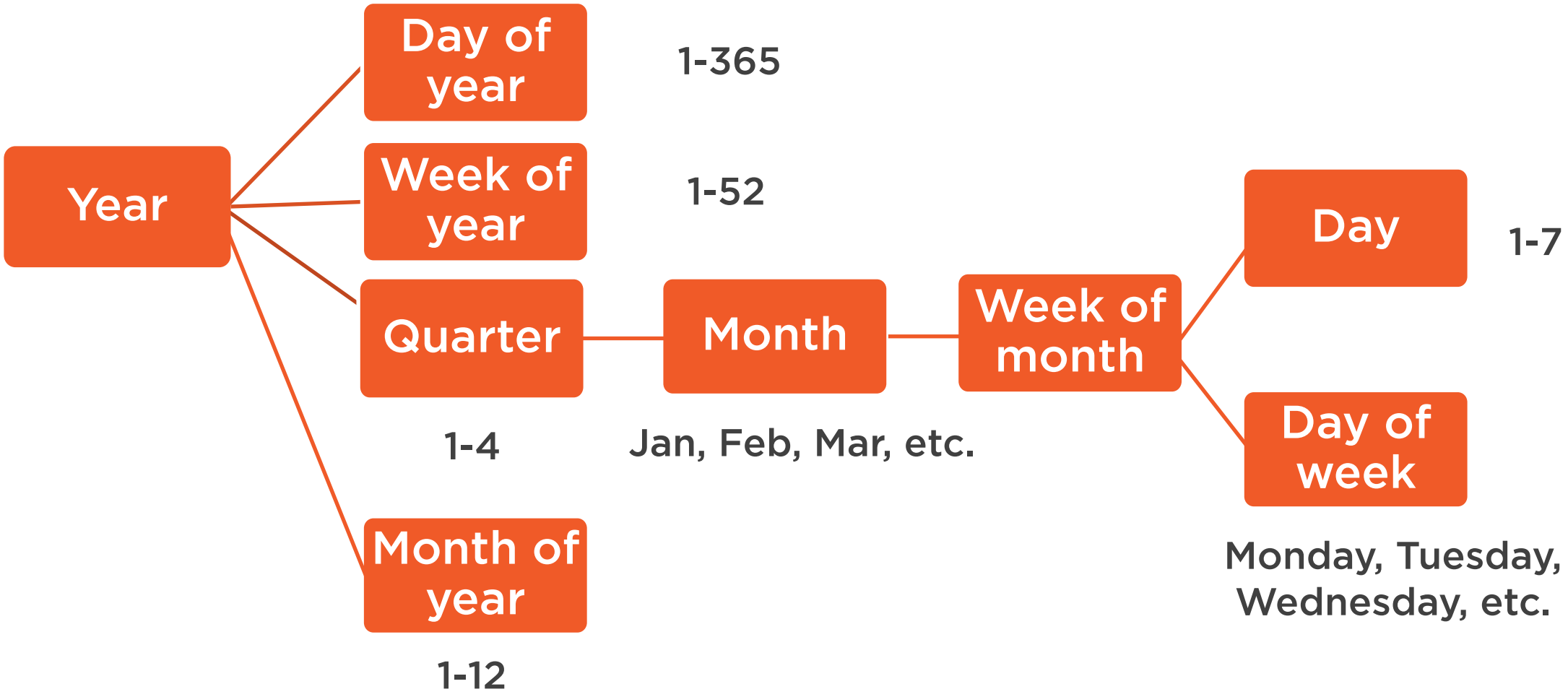
Slowly Changing Dimension: Type II

Timestamp	First name	Last name	Birthdate	Address	City
2-17-2010 01:01:00	Jim	Doe	1-1-1970	123 E. Street	Anytown
6-23-2014 01:04:00	Jim	Doe	1-1-1970	456 W. Blvd	Otherville

Dimensional Hierarchy

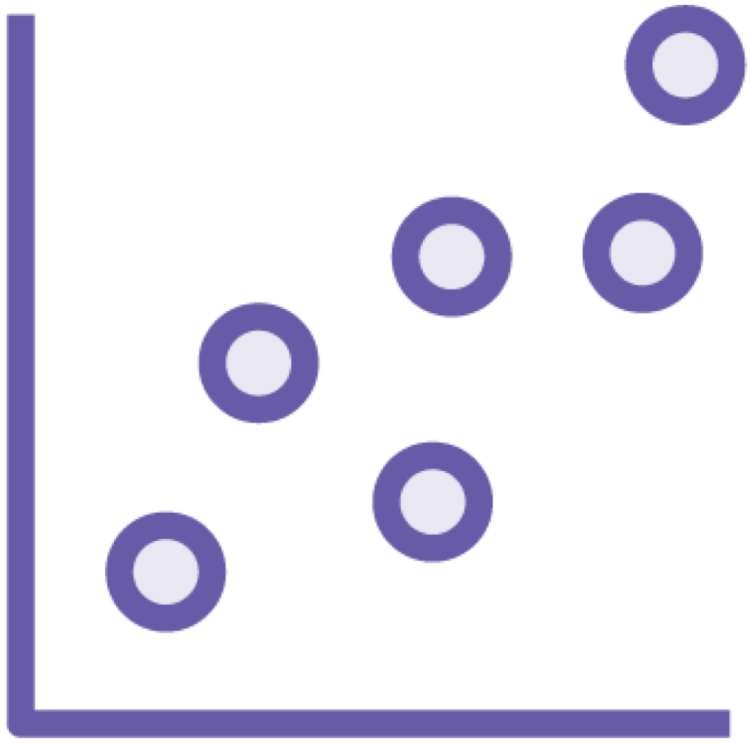


Time Dimension



Video 5 starts here (placeholder only)

Hotel Industry Key Performance Indicators

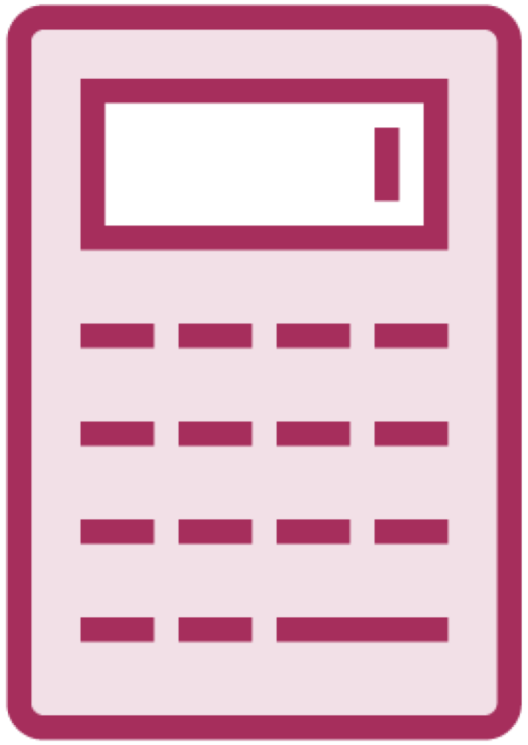


Occupancy percentage

ADR - average daily rate

RevPAR – revenue per available room

Hotel Industry KPI Formulas



Occupancy percentage

- *Occupancy % = rooms occupied / rooms available*

ADR

- *ADR = total room revenue / total rooms occupied*

RevPar

- *REVPAR = total room revenue / total rooms available*

Demo

Time to create a dimensional data model

Summary

- Why we have a separate data model for analytic processing
- Terms and concepts of dimensional modeling
- KPIs considered in a dimensional model
- Demo to create a dimensional model

Next: Building a Database from DDL