Looking “Normal” and Seeing “Stars”

Table Designs Working Together

Deborah Melkin

SQL Saturday Chattanooga

June 27, 2020

WHO AM I?

- 20 years as a DBA
- Mainly work with SQL Server
- Mainly work with OLTP but have worked with some data marts.
- NESQL Board Member
- SQL Saturday\User Group Speaker
- IDERA ACE Class of 2020
- Speaker Idol Winner 2019
- Pronouns: she\her

Random facts:

- I grew up in TN
- I'm the alto section leader in my choir
- I go to bluegrass jams regularly
- I've been learning guitar and now mandolin
- I am a bit of a musical theater geek
- I became a Red Sox fan after 2003 ALCS, Game 7

Agenda

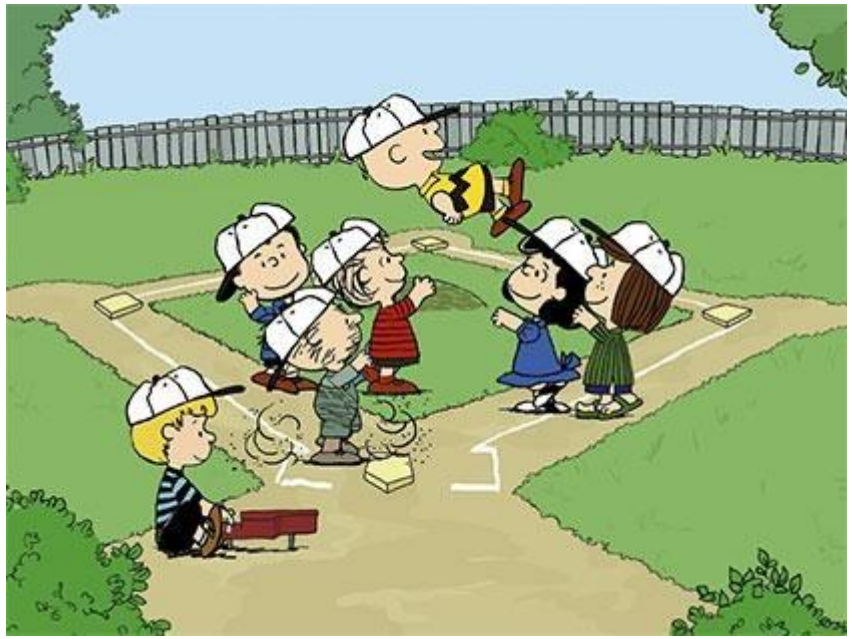
Normal Forms

Star Schemas

Normal Forms and Star Schemas Working Together

Requirements

- There are 162 games scheduled for the regular season.
- Every game has a home team and an away team.
- There are two different leagues:
  - *American League (AL)*
  - *National League (NL)*
- The biggest difference is that pitchers don't bat in the AL.
- Games are broken down into innings.
  - *An Inning is 3 outs per team.*
  - *There are 9 innings in a game.*
  - *Extra innings are played a tie at the end of 9 innings until the tie is broken.*
- A player on a team can play one or more positions.
- For batters, we want to keep track of:
  - *Hits – number of time they hit the ball and made it safely to a base*
  - *RBI (Runs Batted In) – credits the batter for making a play that allows runs to be scored*
- For pitchers, we want to keep track of:
  - *Number of pitches during the inning*
  - *Number of pitches that were strikes*



American League: National League:

https://www.baseball-reference.com/

Normal Forms

The Key,

The Whole Key,

Nothing But the Key,

So Help Me, Codd!

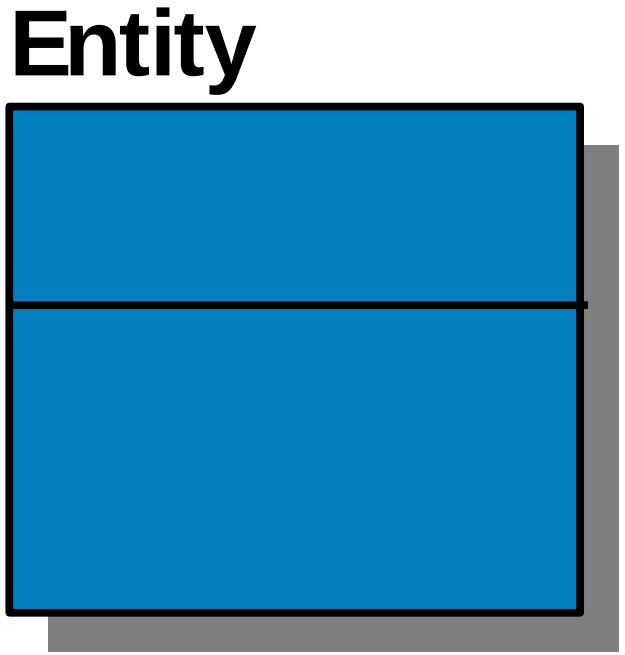
Why Use Normal Forms

- Eliminate data redundancies and incompleteness
- Data manipulation
- Clean database structures
- Extensible schema

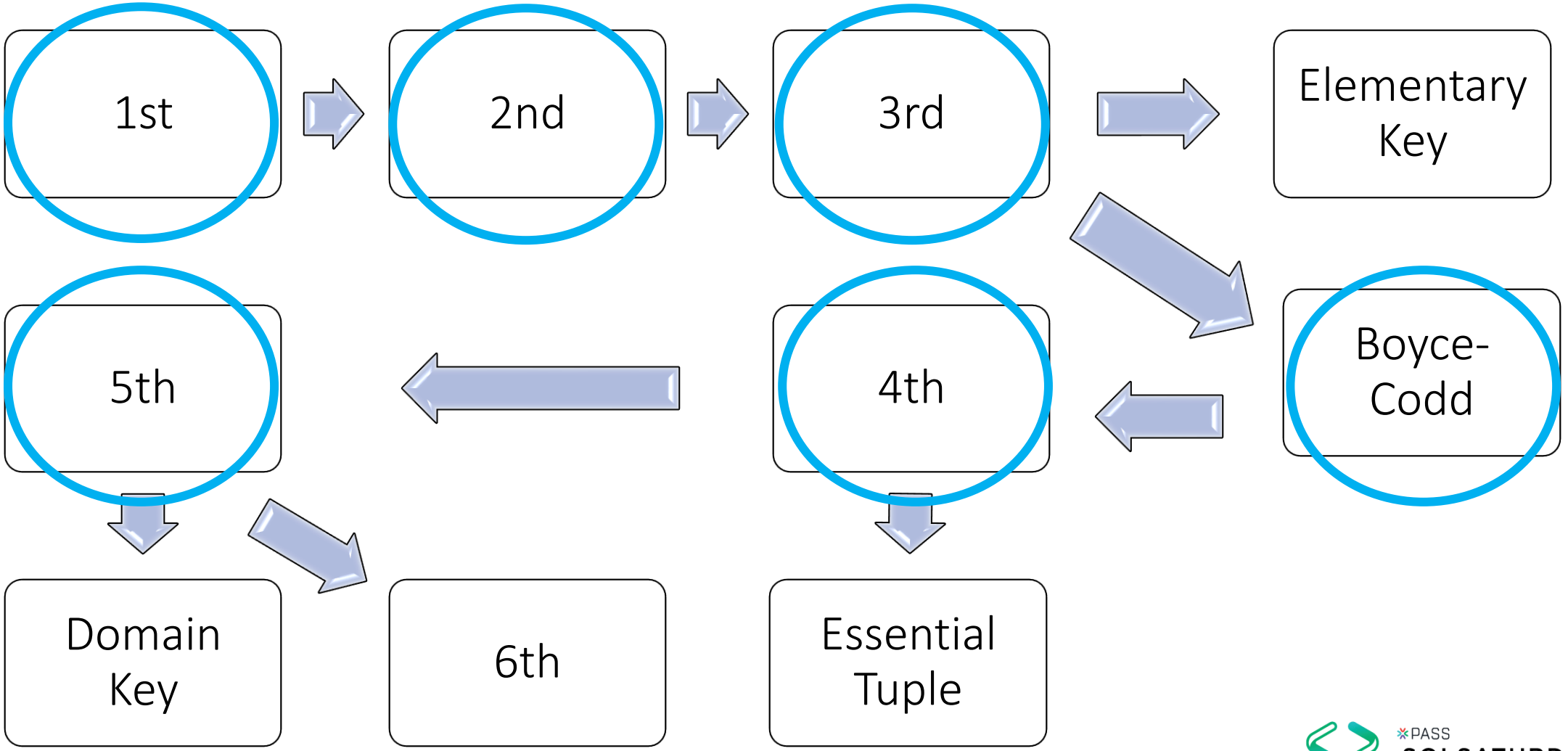
If you don't know what type of database you are working with, it is likely Normal Form.

What is Normalization?

"Tables are normalized when they represent propositions about entities of one type."



Normal Forms



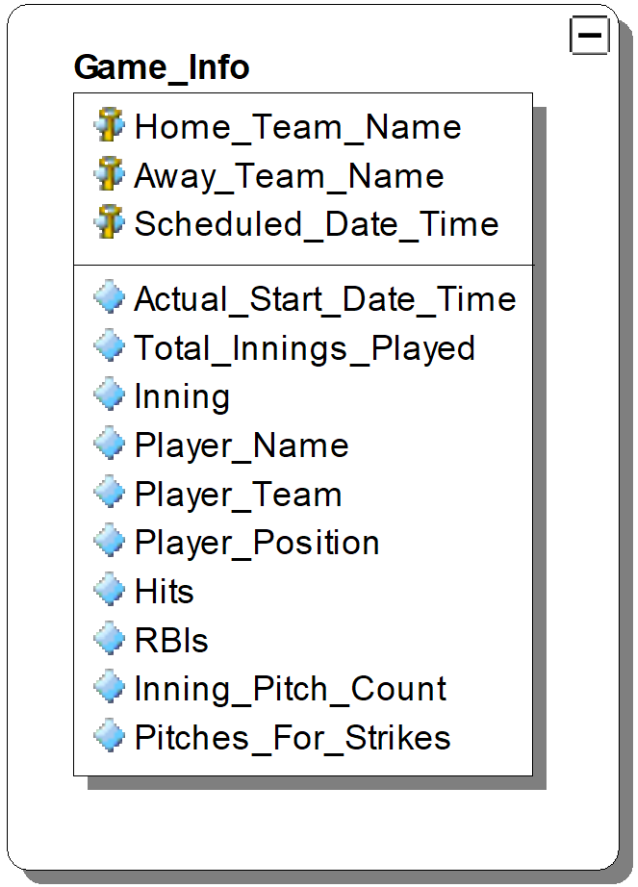
What Level of Normal Do We Want?

FIRST NORMAL FORM	The Key,
SECOND NORMAL FORM	The Whole Key,
THIRD NORMAL FORM	Nothing But the Key,
BOYCE-CODD NORMAL FORM *(optional)*	So Help Me, Codd!

Tables Through The Normal Forms

FIRST NORMAL FORM: ALL COLUMNS ARE ATOMIC.

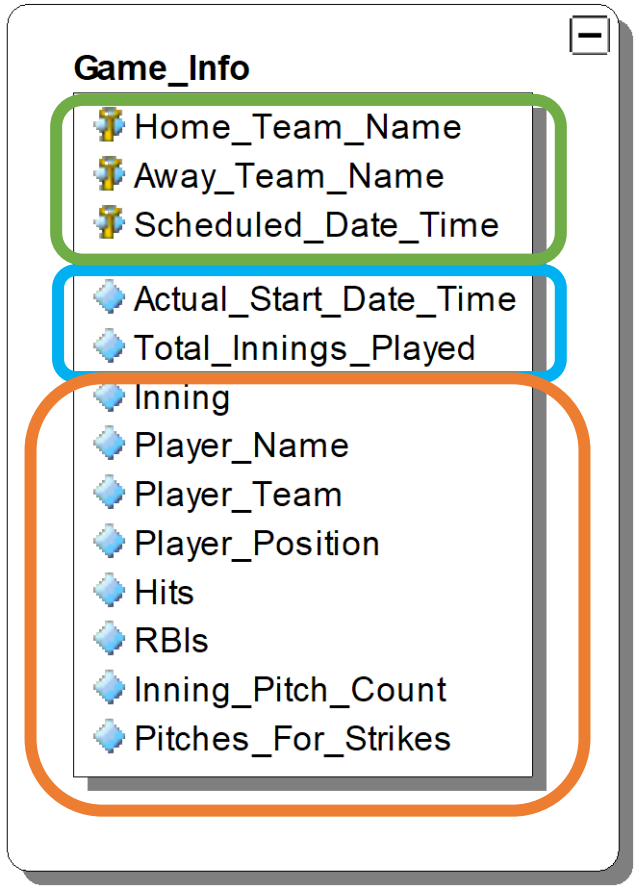
1st Normal Form



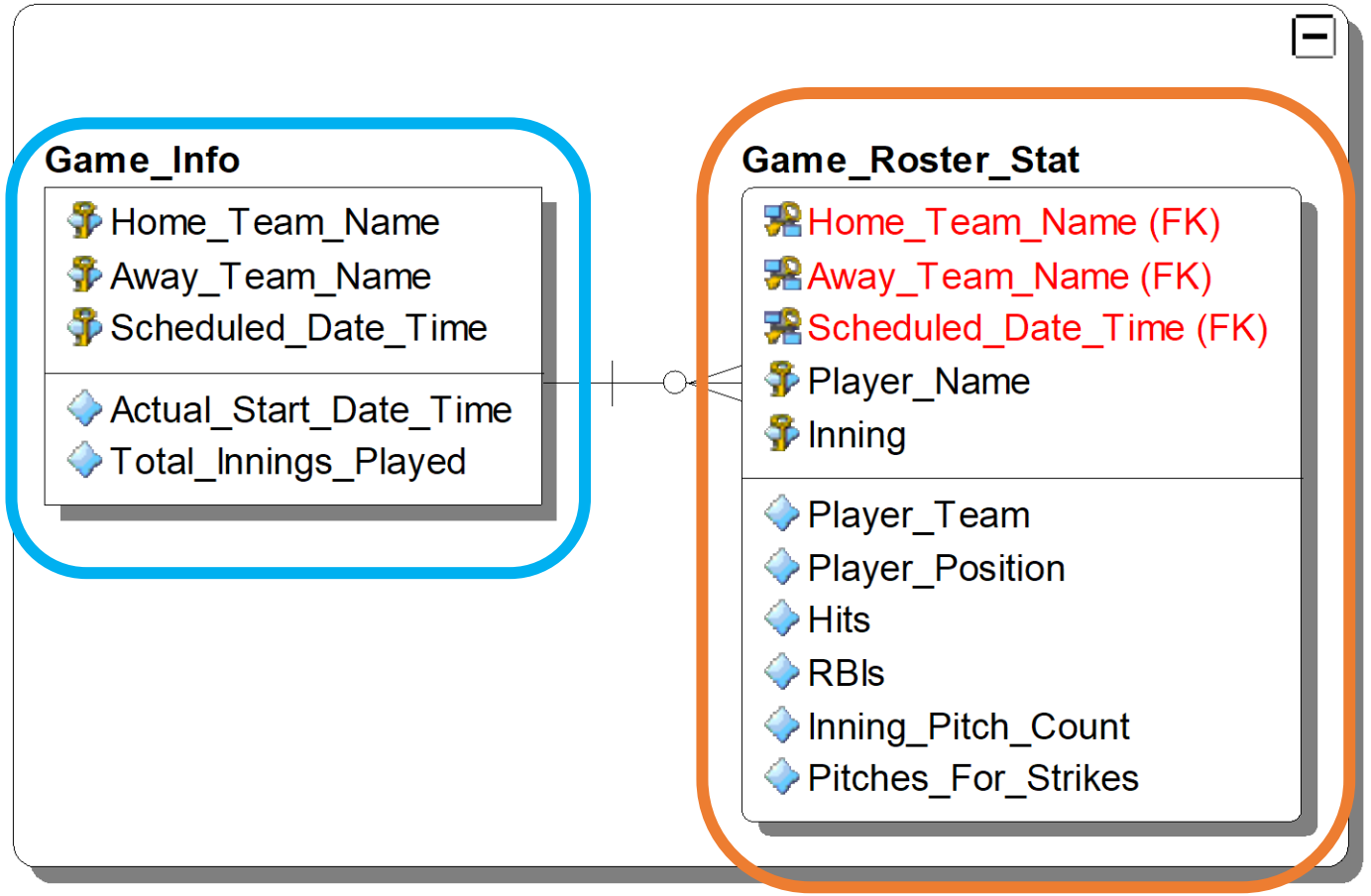
Tables Through The Normal Forms

SECOND NORMAL FORM: AND EVERY NONKEY COLUMN MUST BE FUNCTIONALLY DEPENDENT ON THE ENTIRE KEY.

1st Normal Form



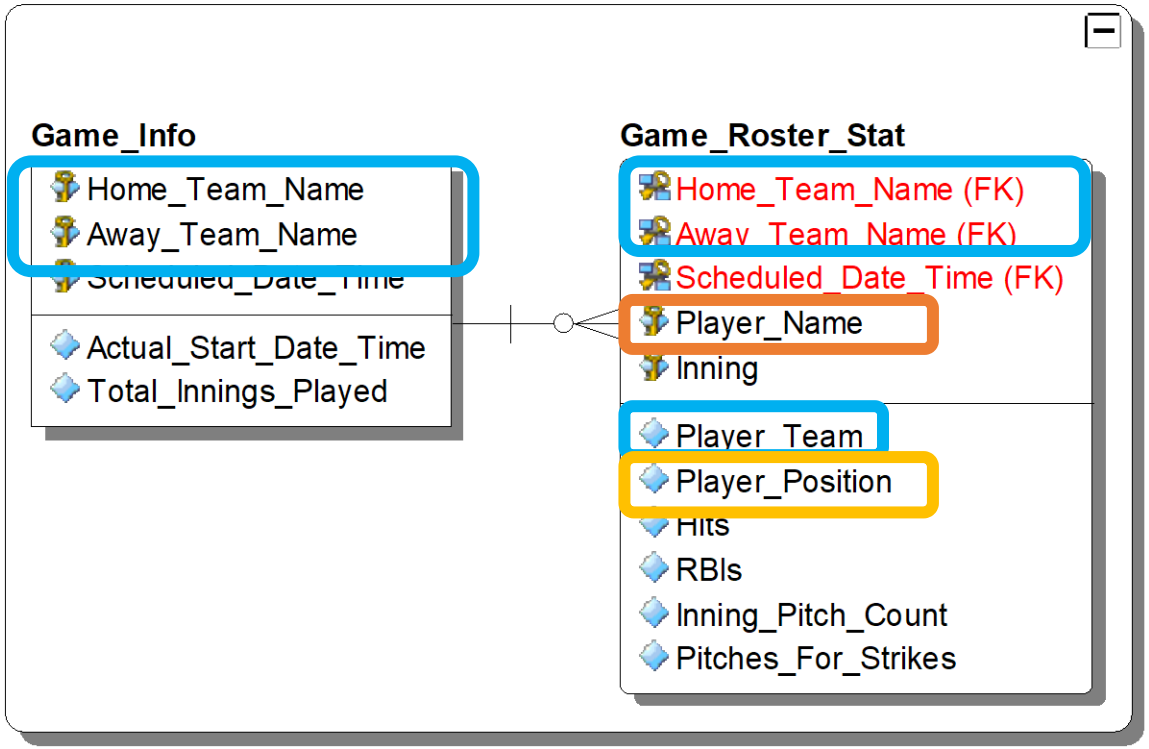
2nd Normal Form



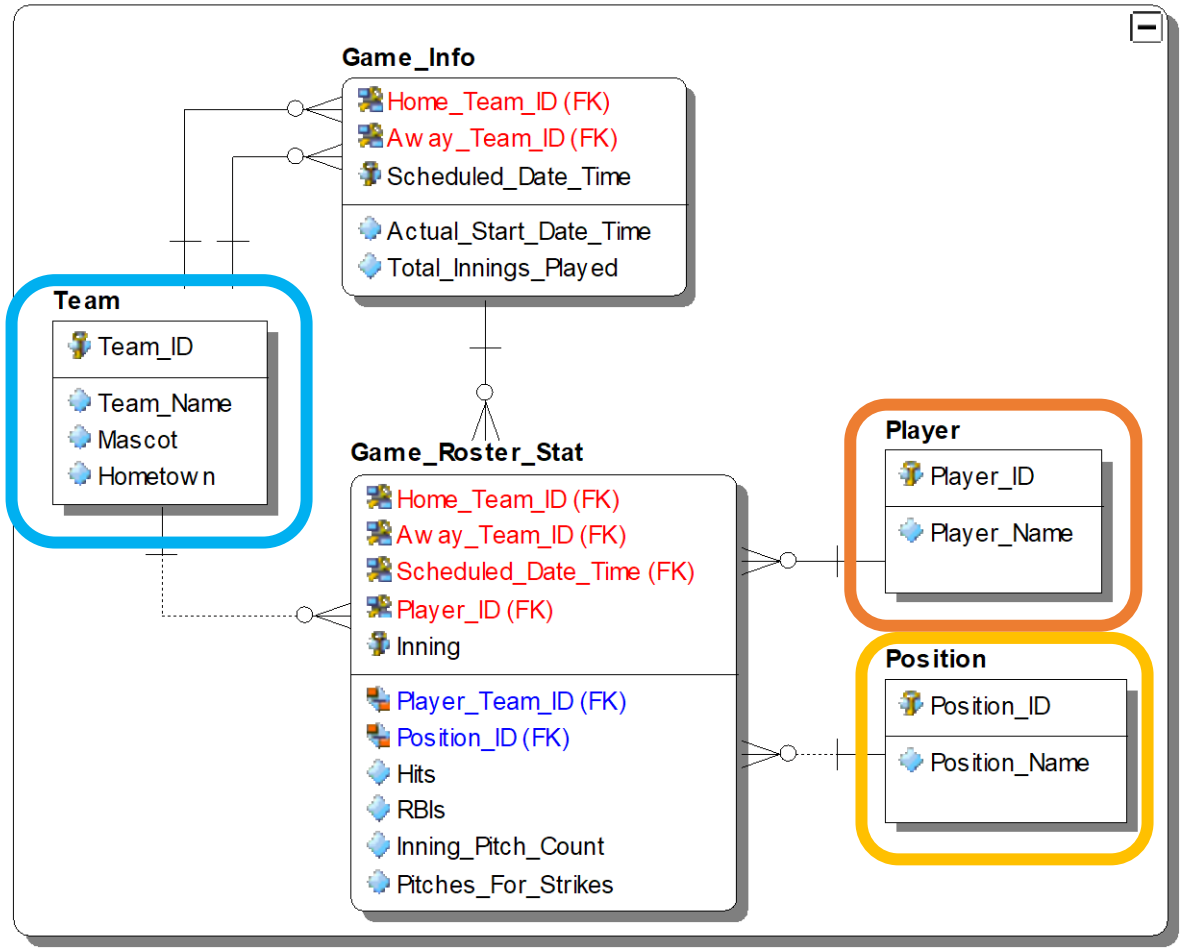
Tables Through The Normal Forms

THIRD NORMAL FORM: AND EVERY NONKEY COLUMN MUST BE NONTRANSITIVELY DEPENDENT ON EVERY KEY.

2nd Normal Form

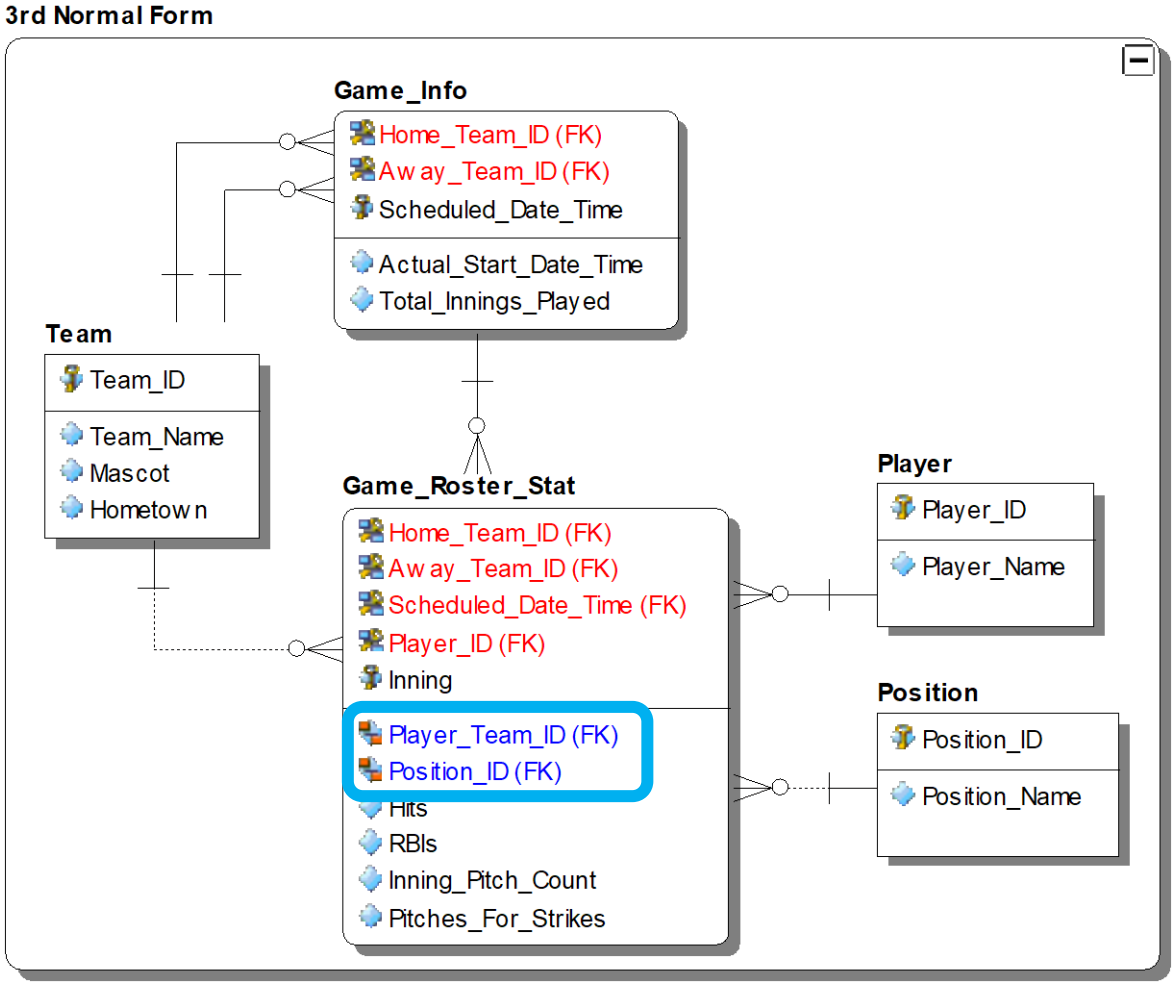


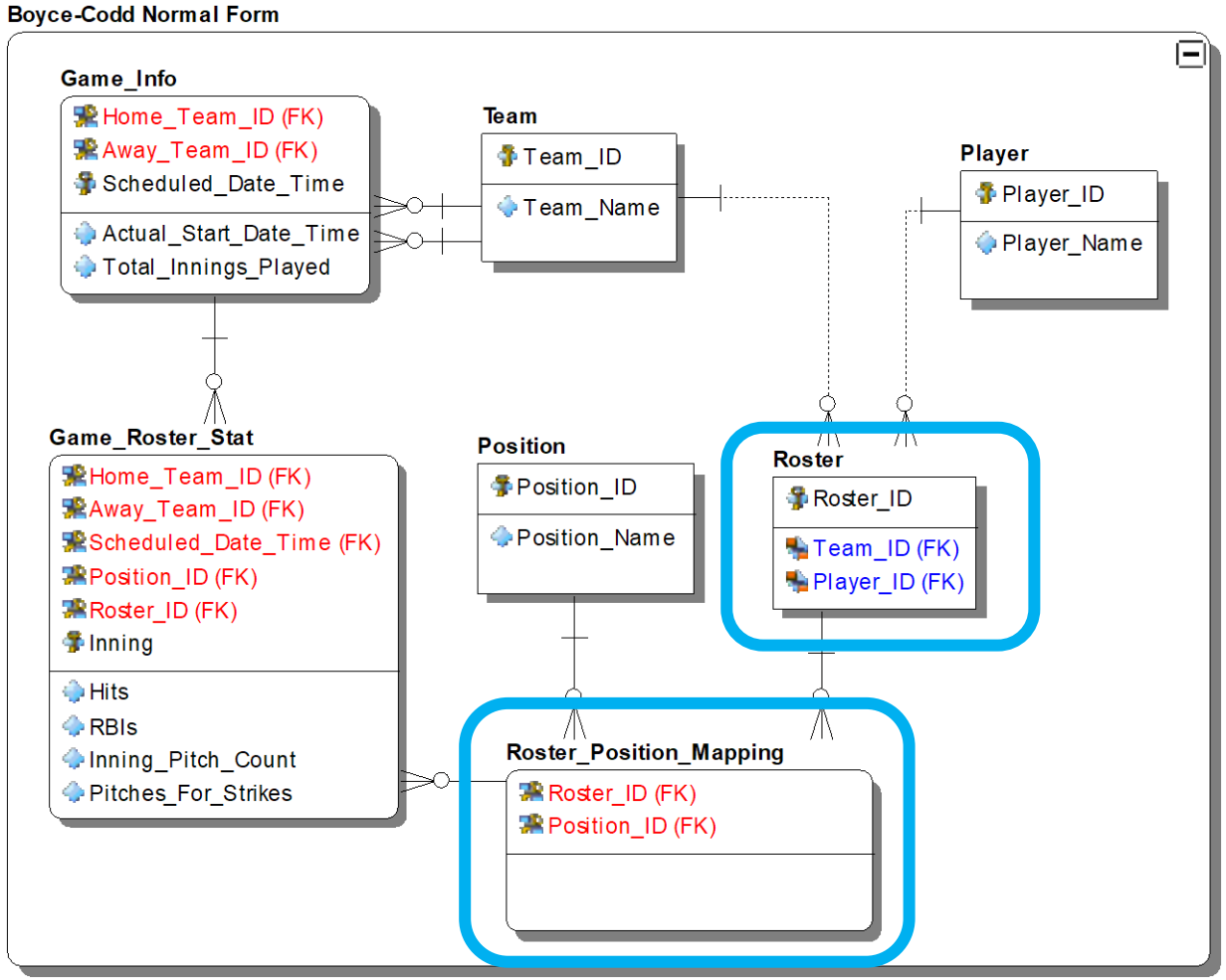
3rd Normal Form



Tables Through The Normal Forms

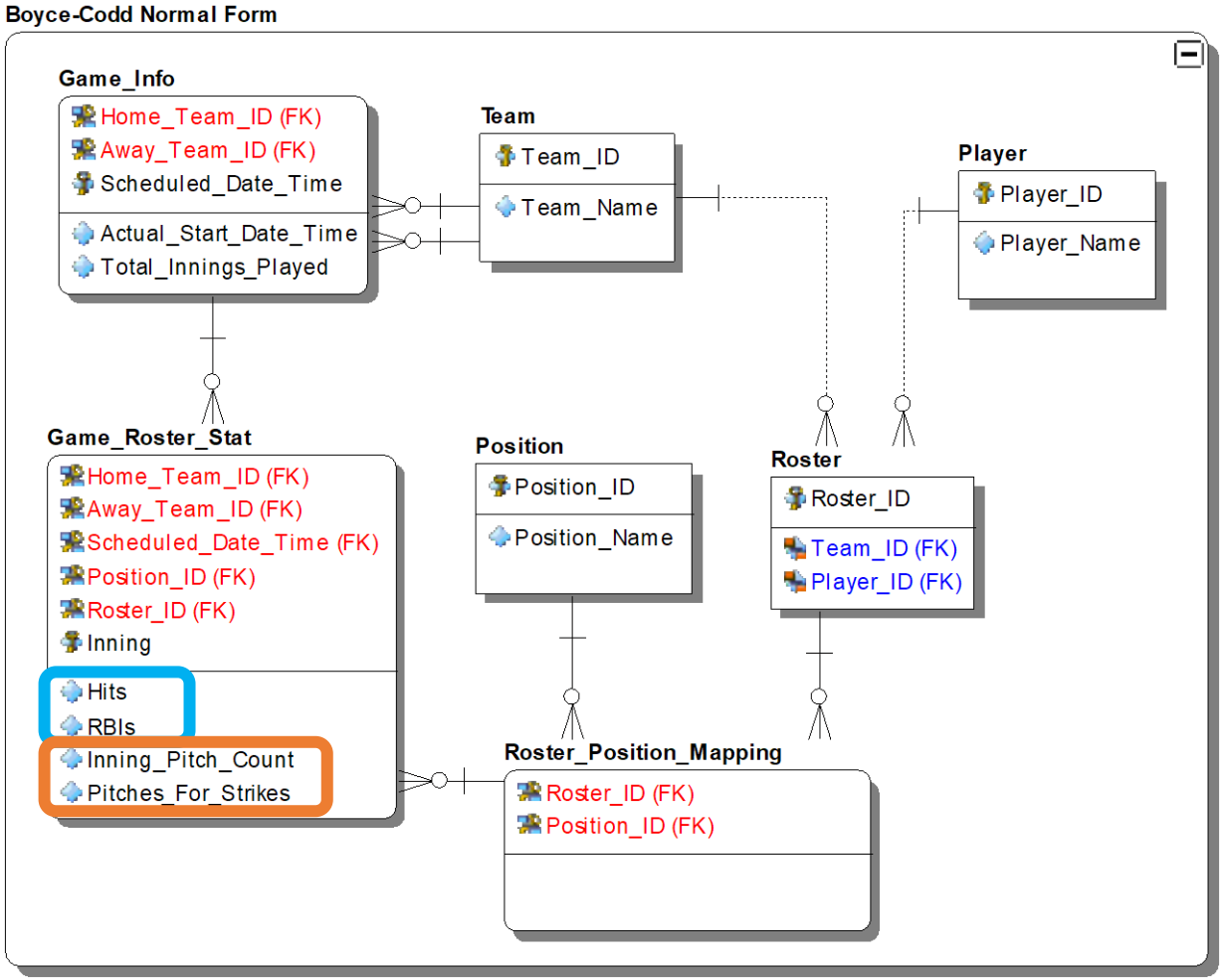
BOYCE-CODD NORMAL FORM: AND EVERY DETERMINANT MUST BE A KEY.

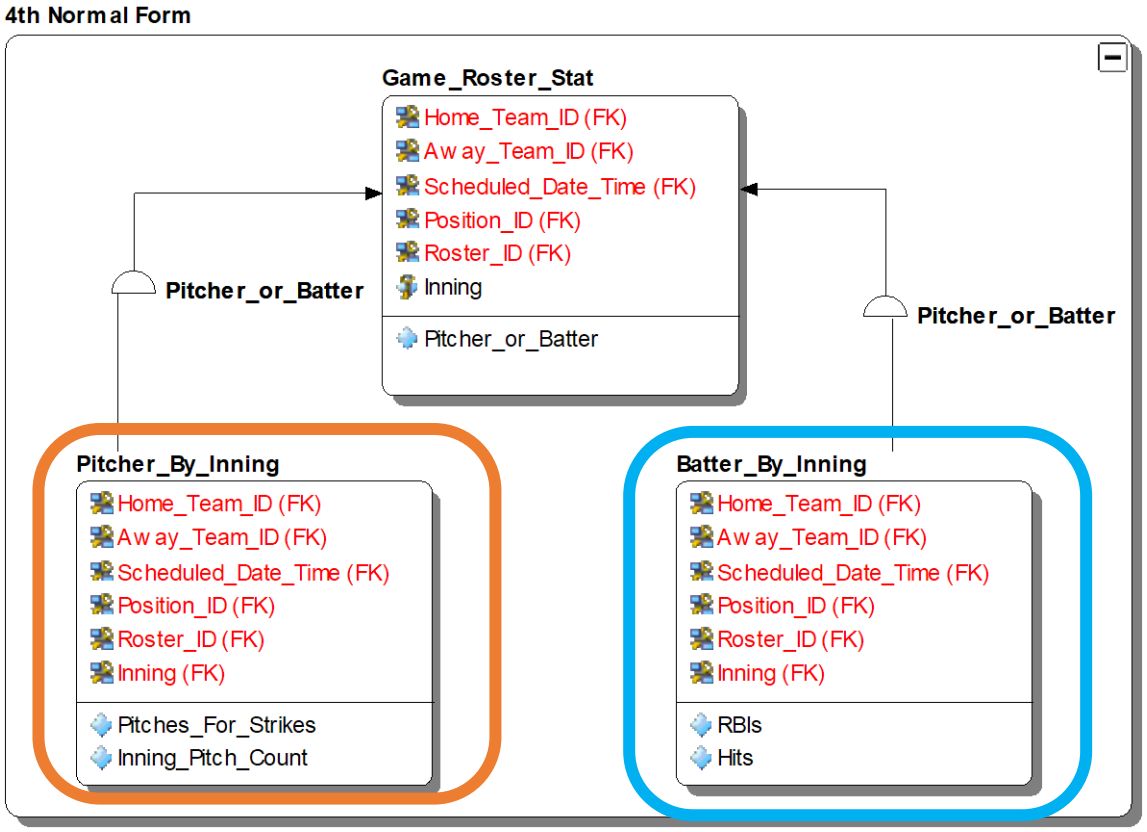




Tables Through The Normal Forms

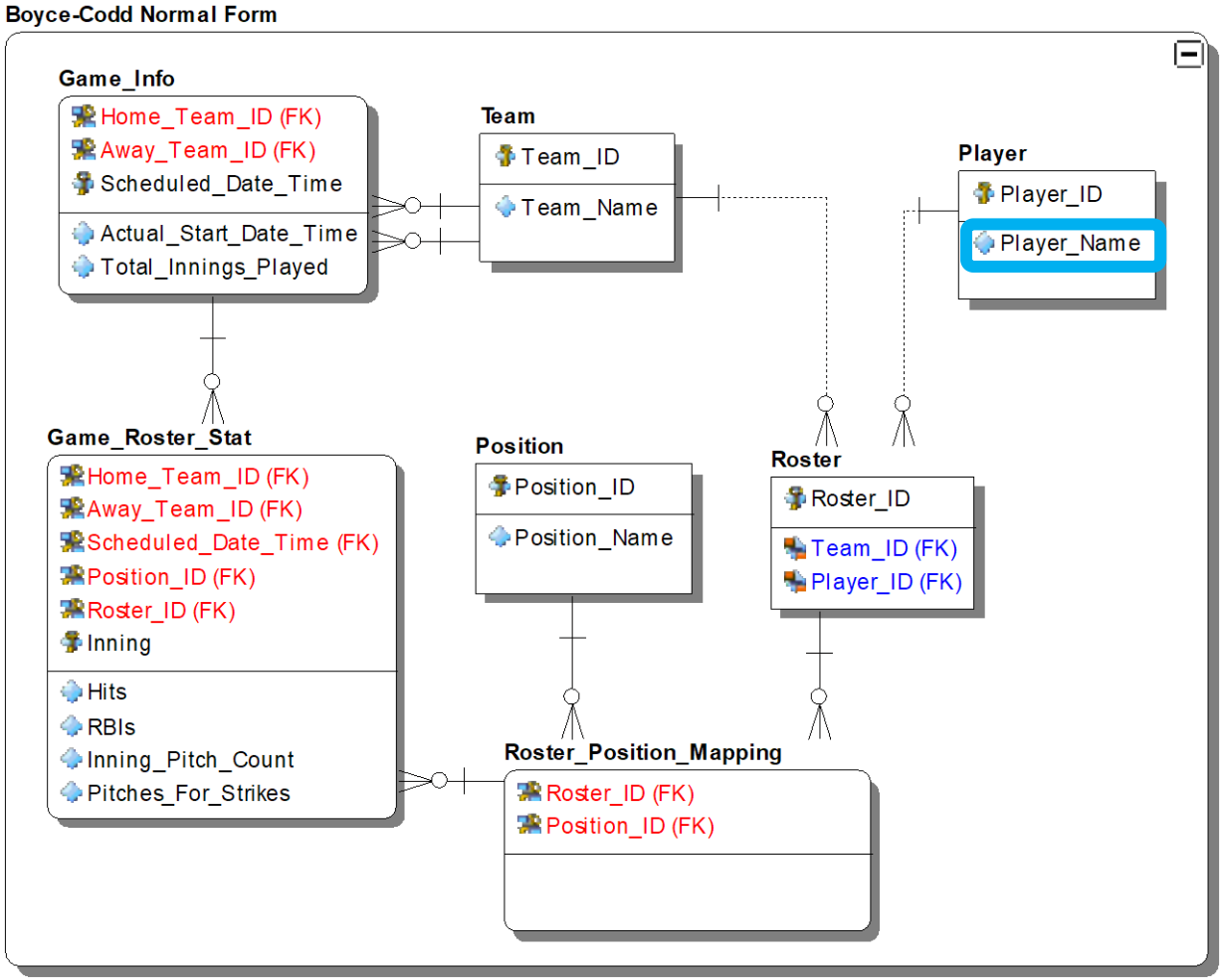
FOURTH NORMAL FORM: AND THERE MUST BE NO NONTRIVIAL MULTIVALUED DEPENDENCIES THAT ARE NOT FUNCTIONAL DEPENDENCIES.

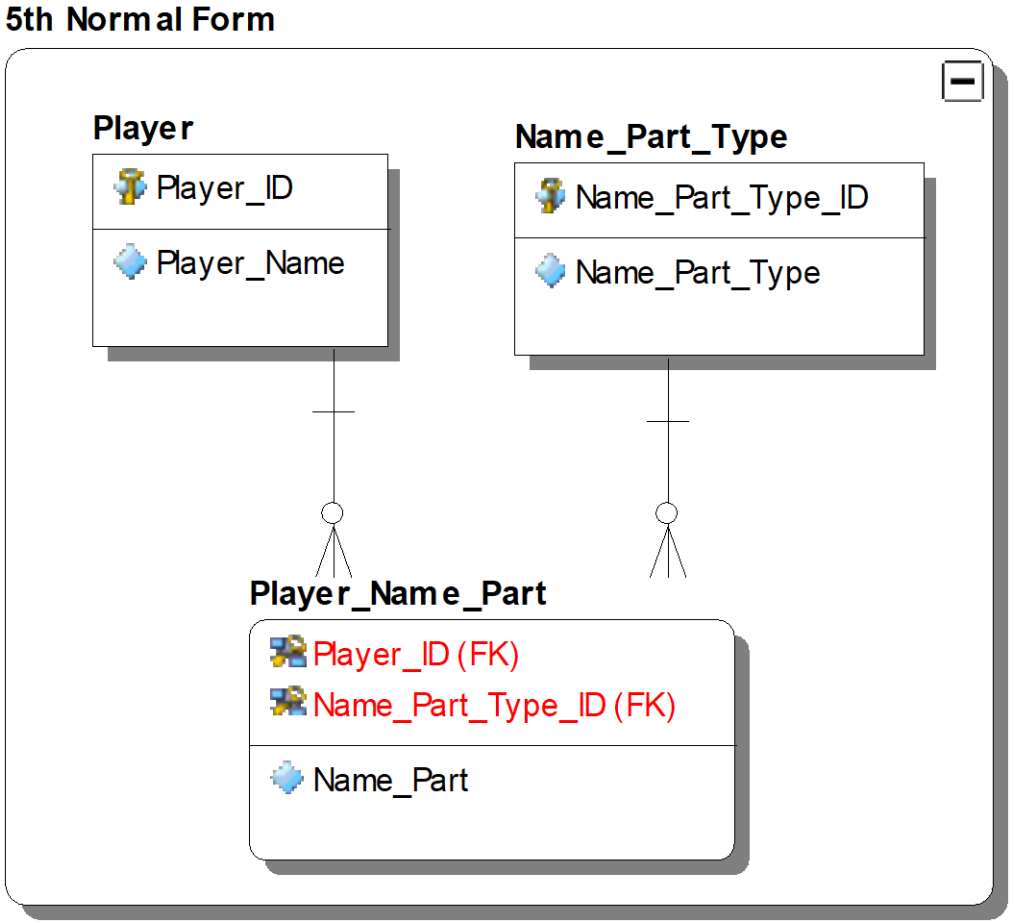




Tables Through The Normal Forms

FIFTH NORMAL FORM: AND EVERY NONTRIVIAL JOIN DEPENDENCY IN THE TABLE IS IMPLIED BY THE KEYS OF THE TABLE.

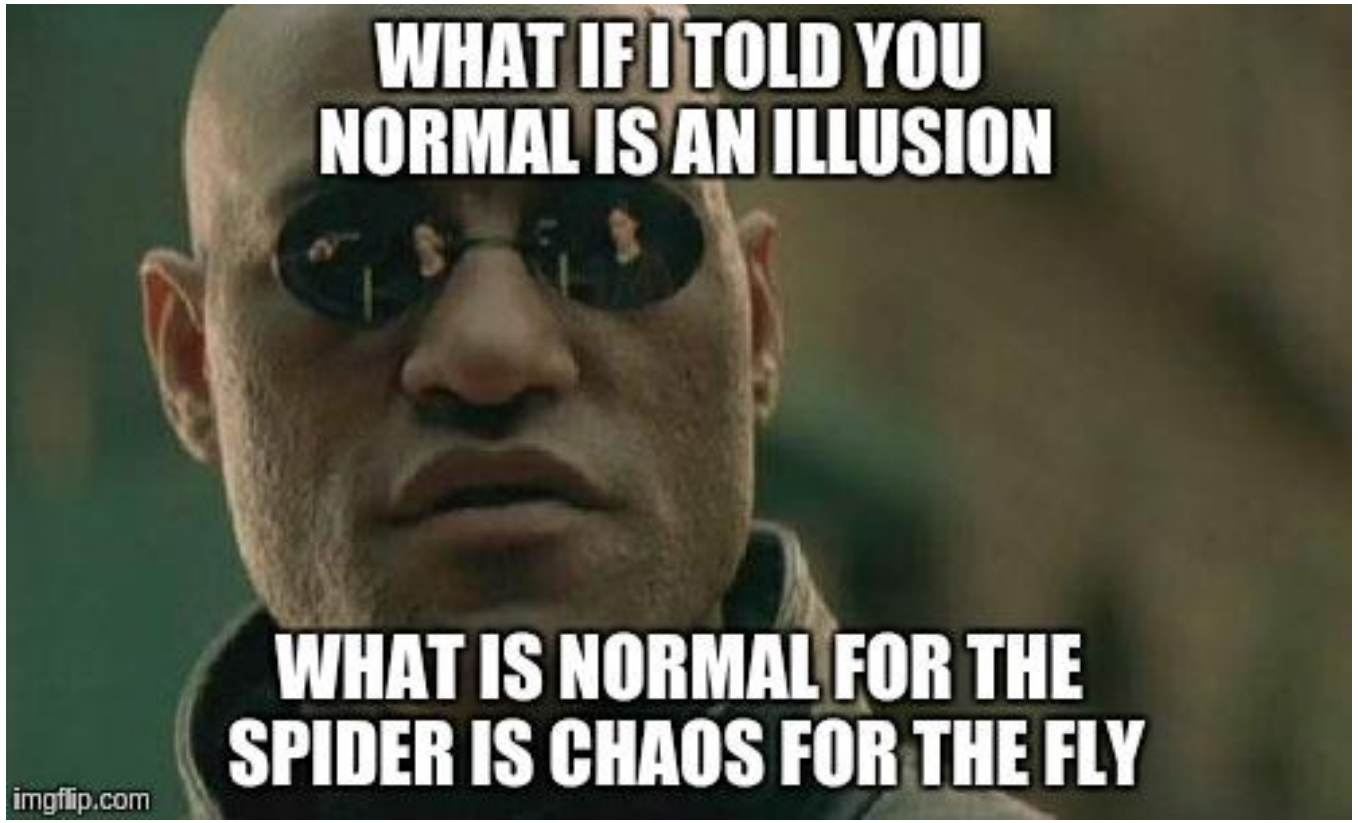




Why Denormal?

Performance

History



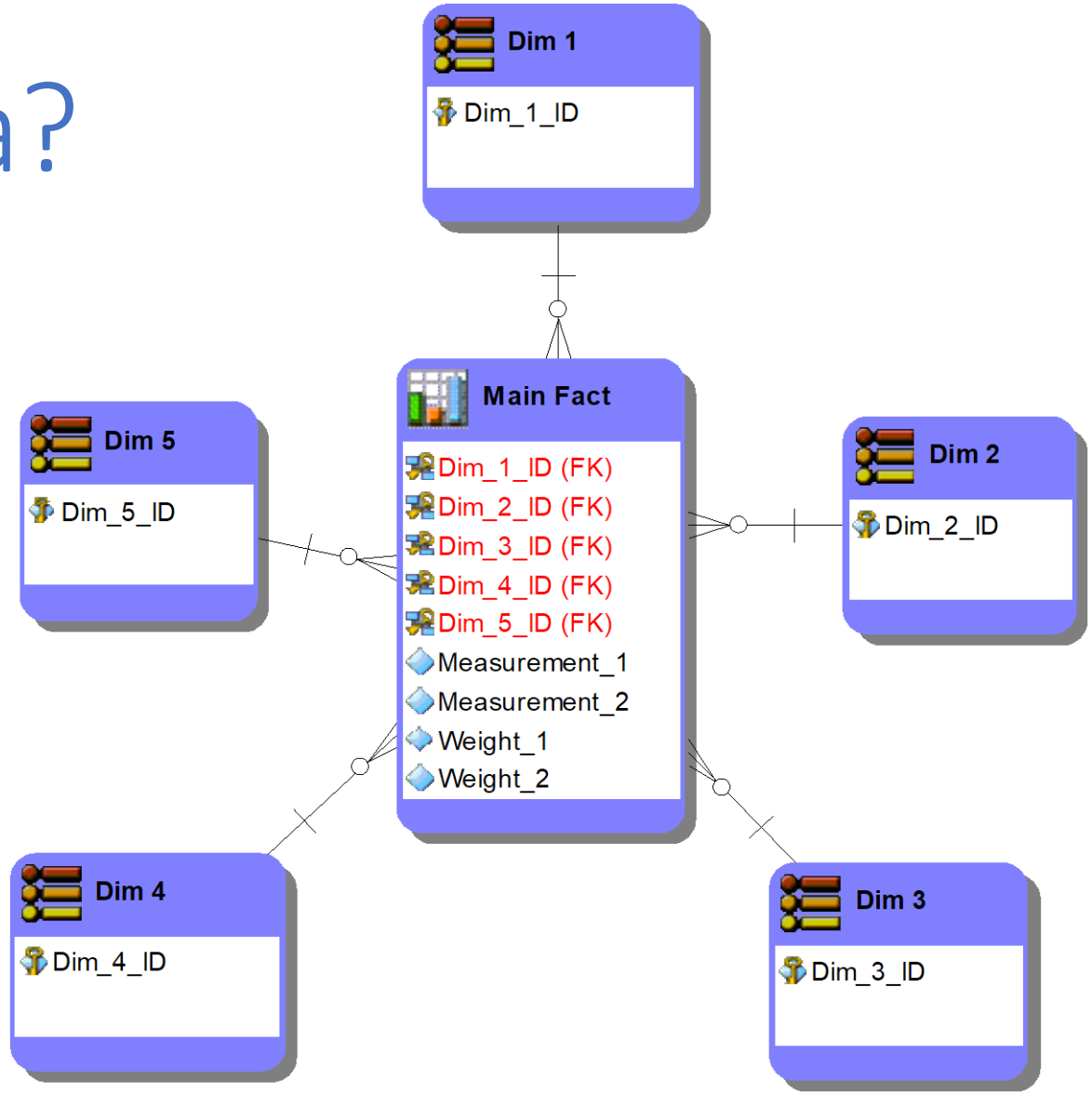
Star Schema

Why Use Star Schema

- Structures the data so it's intuitive to business users
- Fast query performance
- Preferred for data warehouse/business intelligence presentations
- Simplicity and "Understandability"

What is a Star Schema?

"The generic representation of a dimensional model … in which a fact table with a composite key is joined to a number of single level dimension tables, each with a single primary key."

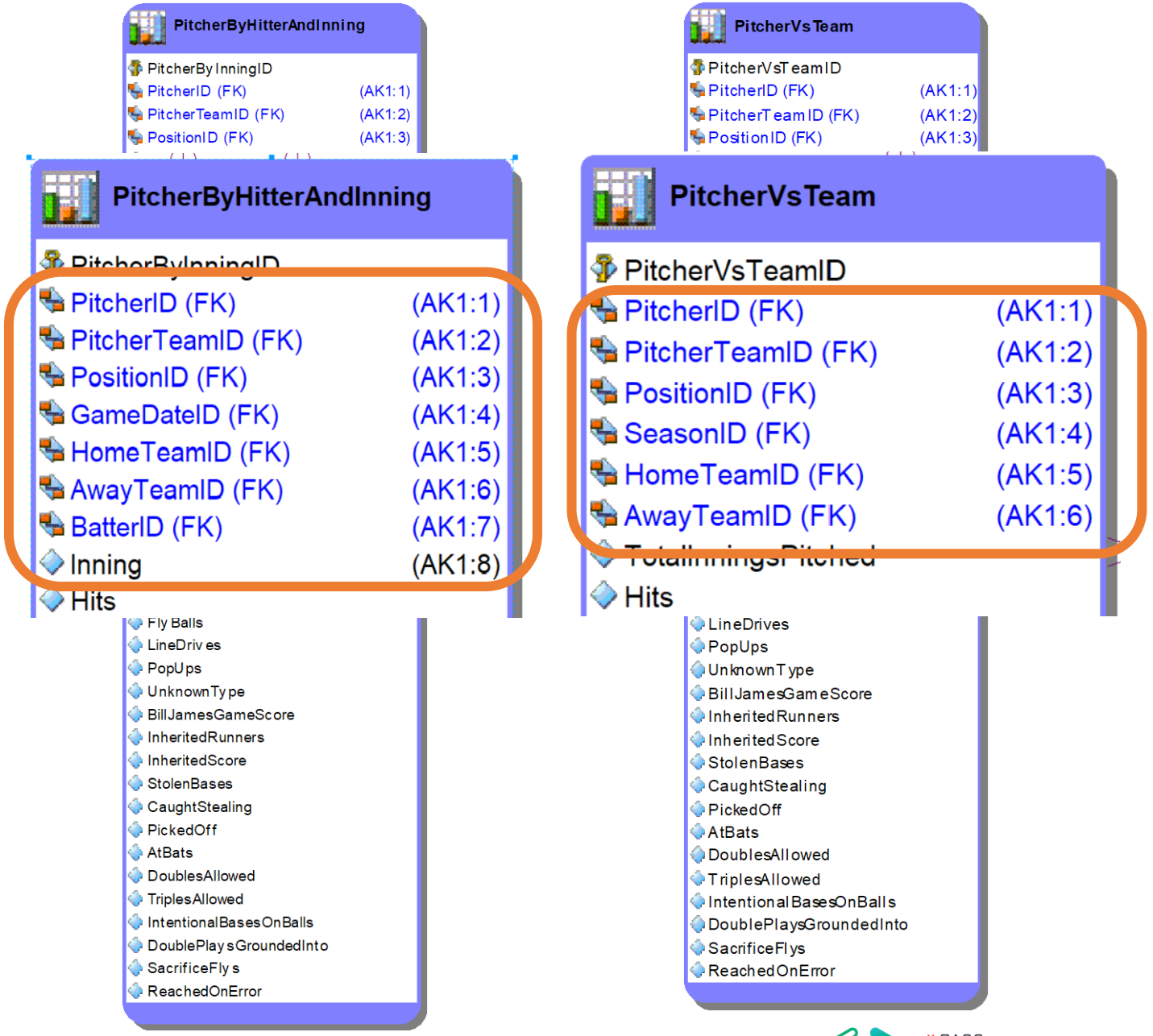


Fact

"In a dimensional model, the central table with numeric performance measurements characterized by a composite key, each of whose elements is a foreign key drawn from a dimension table."

Keyword:

Granularity

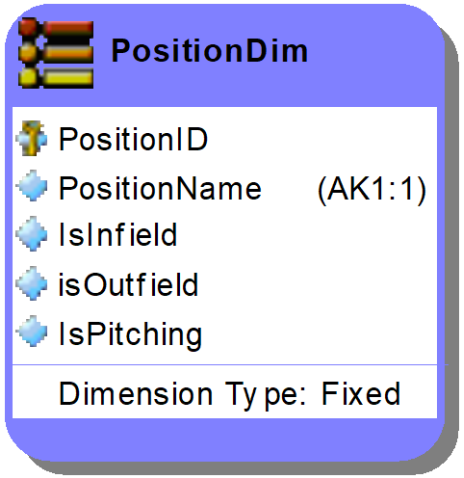


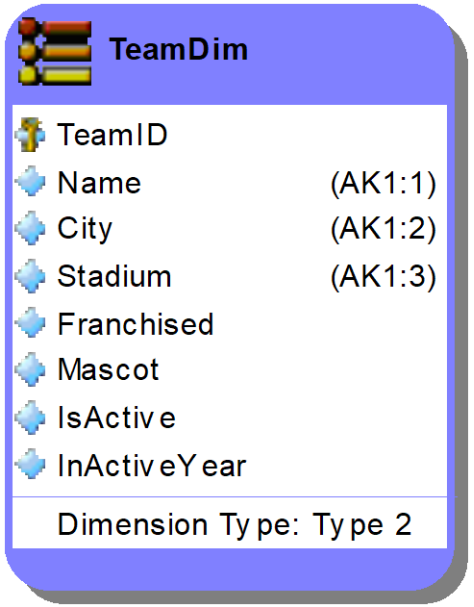
Dimension

"A table in a dimensional model with a single part primary key and descriptive attribute columns."

Multiple types of dimension tables:

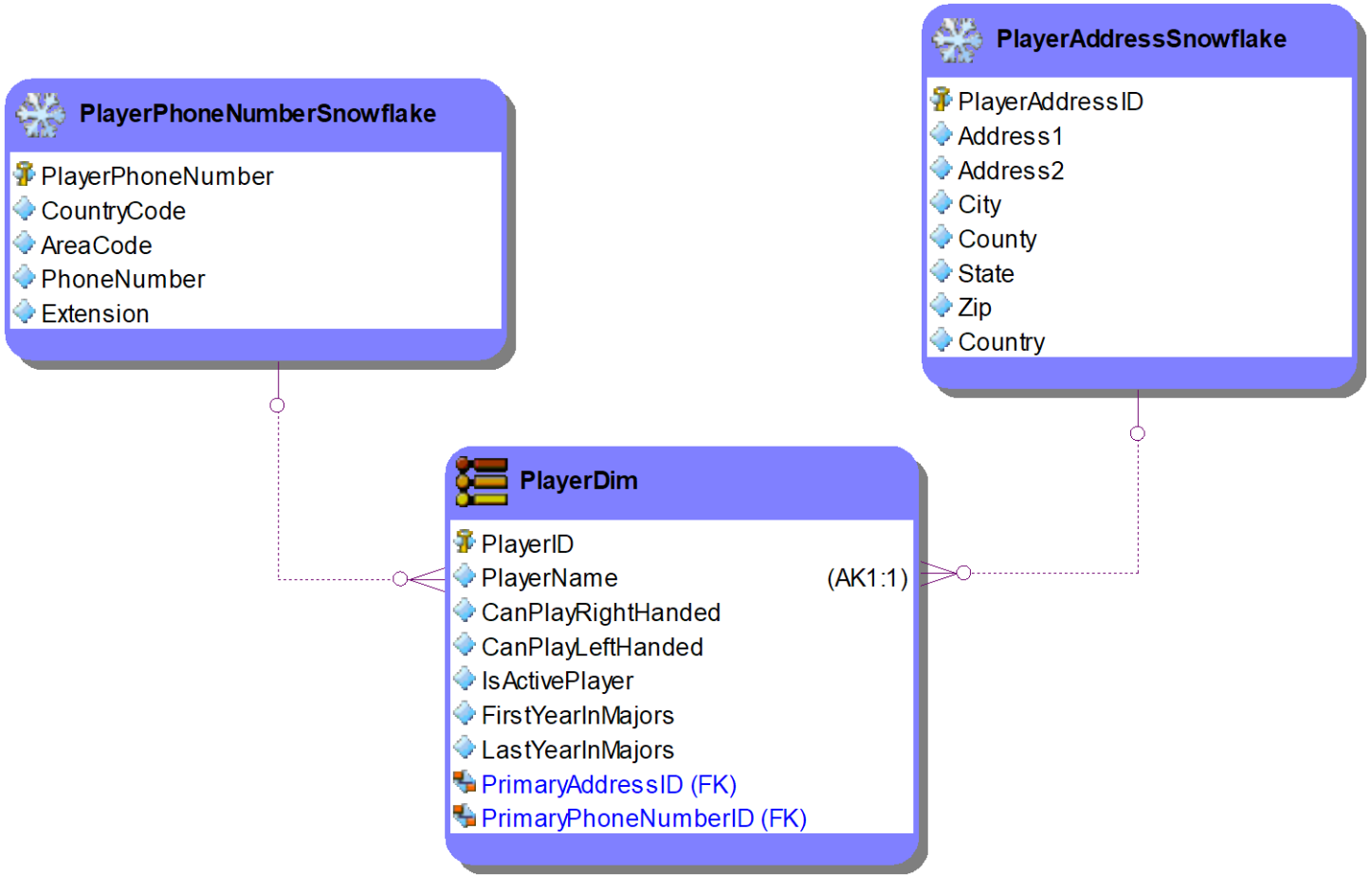
- Conformed
- Junk
- Degenerated
- Role playing
- Slowly Changing Dimensions
  - Types 0 - 7





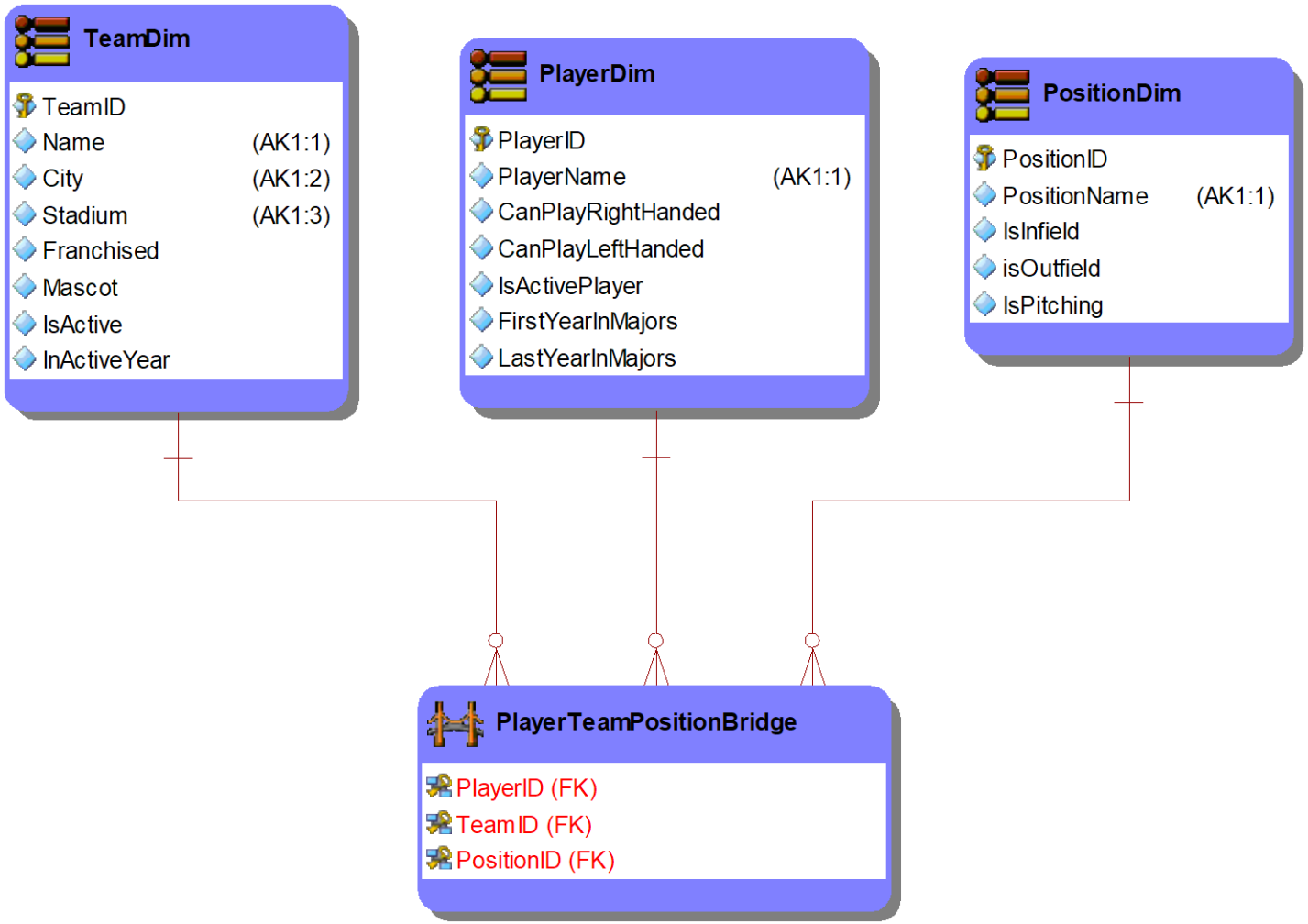
Snowflake

"A normalized dimension where a flat, single table dimension is decomposed into a tree structure with potentially many nesting levels. In dimensional modeling, the fact tables in both a snowflake and star schema would be identical, but the dimensions in a snowflake are presented in third normal form."



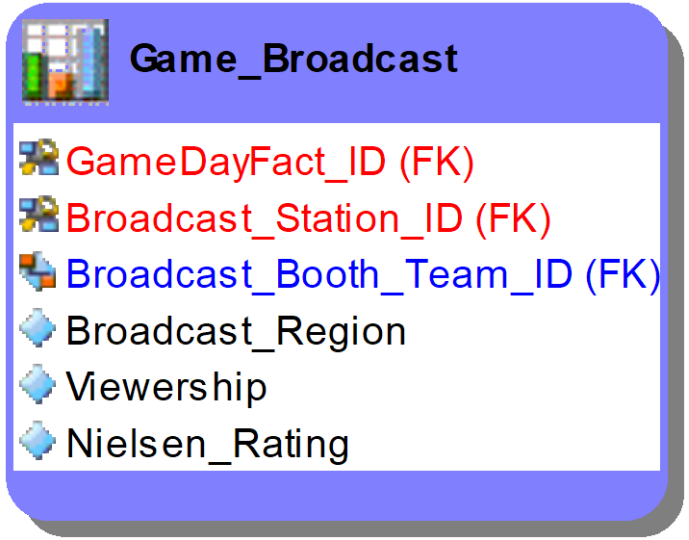
Bridge

"A table with a multipart key capturing a many-to-many relationship that can't be accommodated by the natural granularity of a single fact or dimension table. Serves to bridge between the fact and dimension tables to support many-valued dimension attributes or ragged hierarchies."



Factless Fact

"A fact table that has no facts but captures the many-to-many relationships between the dimension keys. Most often used to represent events or coverage information that does not appear in other fact tables.



Normal and Stars Working Together

Reviewing the differences

NORMAL FORM

- All about the **KEY**
- How the data is **related**
  - Relationship Types have names
- Optimizes **Writes** over Reads
  - The first four normal forms emphasize the ease of updates
  - Too many indexes on a table affect write performance

STAR SCHEMA

- All about the **FACT**
- How the data **functions**
  - Table Types have names
- Optimizes **Reads** over Writes
  - Fact tables store aggregates and measures
  - Index based on data usage

Finding similarities in the differences

NORMAL FORM

- Lookup Tables
- Mapping Tables
- Super\Subtype Tables
- SQL Server Database Engine

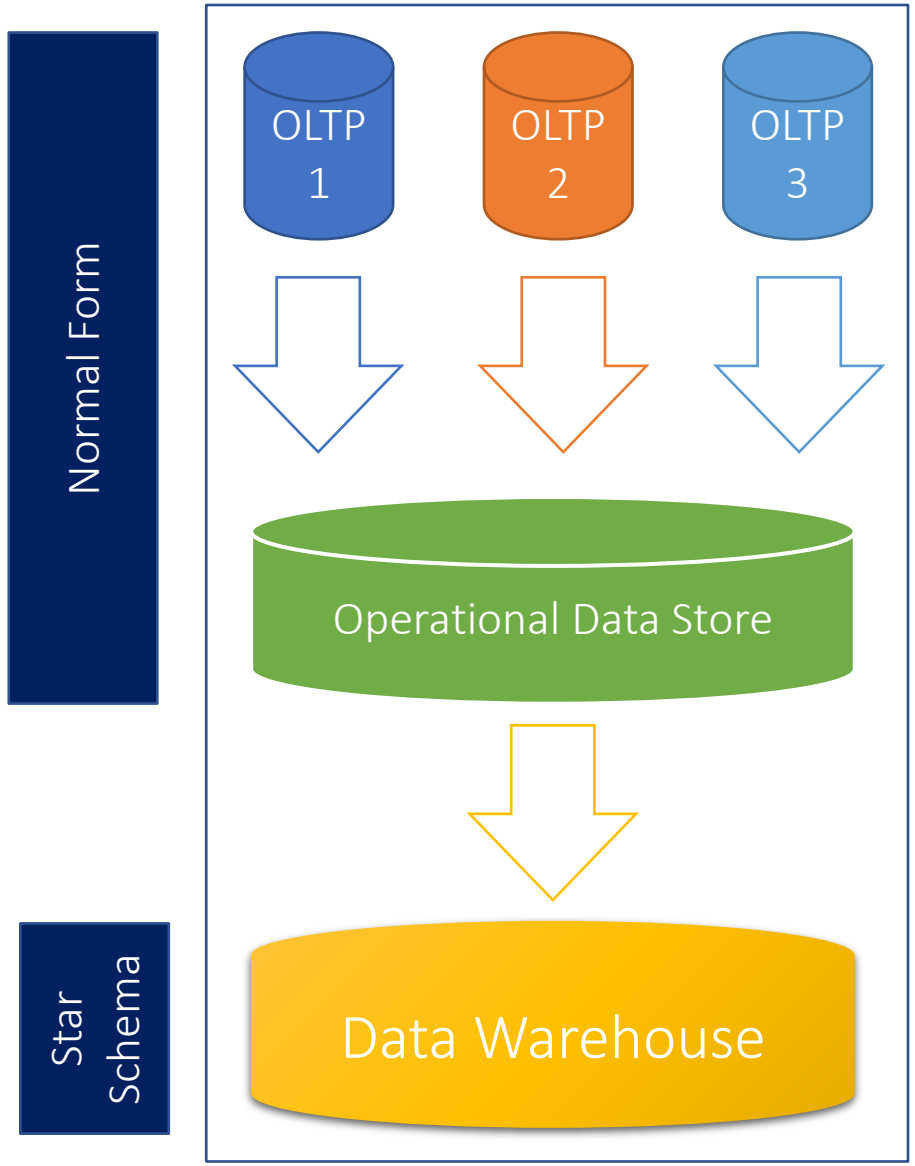
STAR SCHEMA

- Dimensions
- Bridges
- Snowflakes
- SQL Server Database Engine

Operational Data Store

A relational database used to combine data from different sources before ETL\ELT process to move data into the data warehouse or mart.

Some data warehousing methodologies say that you must have an ODS as part of the ETL process.



In Conclusion

Why is this important?

It’s all about the data

Knowing how the data used dictates …

- The type of model you use
- The type of tables you create
- The level of normalization
- The facts and dimensions to create
- How to move the data between the systems

Additional Resources – Normal Forms

- *Inside Microsoft SQL Server 2008: T-SQL Querying*. Itzik Ben-Gan, Lubor Kollar, Dejan Sarka, Steve Kass.
- https://www.bkent.net/Doc/simple5.htm William Kent, "A Simple Guide to Five Normal Forms in Relational Database Theory", Communications of the ACM 26(2), Feb. 1983, 120-125)
- Additional Normal Forms:
- Elementary Key Normal Form: http://what-when-how.com/Tutorial/topic-1114galv/Database-Design-and-Relational-Theory-167.html
- Essential Tuple Normal Form: https://researcher.watson.ibm.com/researcher/files/us-fagin/icdt12.pdf

Additional Resources - Star Schemas

- *The Data Warehouse Lifecycle Toolkit*. Ralph Kimball, Margy Ross, Warren Thornthwaite, Joy Mundy, Bob Becker.
- https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/
- *Building the Data Warehouse*. William H. Inmon.

Any Questions, Comments?

Email: dgmelkin@gmail.com

Twitter: @dgmelkin

Blog: debthedba.wordpress.com

PASS SQLSATURDAY

Thank you to our Global Sponsors and Supporters

Thanks For Coming!