**NBA DATA MART PROJECT REPORT**

by Cem Ünsal, Furkan Arslan

202128009, 202028013

Table of Contents

[Chapter 1: Introduction 2](#_Toc183724229)

[General Information About the Chosen Company 2](#_Toc183724230)

[Description of the Problem 2](#_Toc183724231)

[Current Situation 2](#_Toc183724232)

[Chapter 2: Requirements Analysis 3](#_Toc183724233)

[User Expectations 3](#_Toc183724234)

[Questions To Be Answered 3](#_Toc183724235)

[Surveys, Questionnaires, and Interview Reports 3](#_Toc183724236)

[Conceptual Design 4](#_Toc183724237)

[Chapter 3: Dimensional Design of the System 4](#_Toc183724238)

[Dimensional Model Used 4](#_Toc183724239)

[Proposed Model 4](#_Toc183724240)

[Diagrams 5](#_Toc183724241)

[References 7](#_Toc183724242)

# Chapter 1: Introduction

## General Information About the Chosen Company

The chosen company is the National Basketball Association (NBA), a professional basketball league in North America. The league is composed of 30 teams (29 in the United States and 1 in Canada). The NBA headquarters is located at:

Address: 645 Fifth Avenue, New York, NY 10022, USA  
Business Information: The NBA is one of the major professional sports leagues in the United States and Canada. It organizes regular-season games, playoffs, and championships while providing statistical data to enhance the experience for fans, teams, and analysts.

## Description of the Problem

The NBA has extensive data on team performance metrics such as rebounds, assists, blocks, and scores, collected from every game played. However, the data is spread across various systems, making it challenging to perform in-depth analyses of team performances across seasons and locations. Teams, analysts, and management are struggling to gain insights to improve strategies and make data-driven decisions.

## Current Situation

Currently, the NBA relies on transactional databases for storing raw game data. While these databases are optimized for data input and retrieval, they are not ideal for analytical tasks. Users face difficulties in querying complex insights such as performance trends over time, comparative analyses of home vs. away games, and seasonal evaluations. This has led to inefficiencies in decision-making and reduced the ability to generate actionable insights from the data.

# Chapter 2: Requirements Analysis

## User Expectations

The users expect the developed data warehouse system to:

1. Provide a centralized repository for all game performance data.
2. Enable efficient querying and reporting of team performance metrics.
3. Support comparative analysis across different teams, seasons, and game locations.
4. Offer user-friendly access to pre-defined and ad-hoc queries.

## Questions To Be Answered

The developed system should answer questions such as:

* What is the performance comparison between home and away games for a specific team?
* How has a team's performance evolved across different seasons?
* Which teams have the highest average points, rebounds, or assists in the last five years?
* What is the correlation between turnovers and game outcomes for each team?
* Which cities or states host the most successful teams in terms of points scored?

## Conceptual Design

Based on the gathered requirements, the conceptual design focuses on:

1. Central fact table: Capturing team overall performance metrics (e.g., rebounds, assists, blocks, turnovers, and points).
2. Dimension tables: Providing descriptive information such as team details, game dates, and locations.
3. Star and snowflake schemas: Evaluating the most suitable schema for the data mart implementation.

# Chapter 3: Dimensional Design of the System

## Dimensional Model Used

The snowflake schema has been chosen for its normalized structure, ensuring minimal data redundancy and optimized storage. This design is well-suited for handling complex queries across multiple dimensions.

## Proposed Model

The snowflake schema consists of:

1. Fact Table:
   * Game performance on overall metrics such as rebounds, assists, steals, blocks, turnovers, personal fouls, and points for both home and away teams.
2. Dimension Tables:
   * Team Dimension: Contains information about teams, including team name, abbreviation, city, state, and year founded.
   * Game Dimension: Captures game-specific data such as game ID, season ID, game date, and in-game statistics for both home team and away team.
   * Location Dimension: Provides data about team states and cities for detailed analysis.
   * Season Dimension: Tracks seasonal and date information.

## Diagrams

ER-Diagram, star schema diagram, and snowflake schema diagram has been created using ER Assistant. The best choice for this project will be the snowflake schema diagram. It details the relationships between the fact table and dimension tables, ensuring clarity in the data model with no redundancy and it is more flexible.

A diagram of a team

Description automatically generated

*Figure 1: ER-Diagram*

A group of black and white text

Description automatically generated with medium confidence

*Figure 2: Star Schema*

A diagram of a team

Description automatically generated

*Figure 3: Snowflake Schema*

# References

* (Kimball & Ross, 2013)
* (Inmon, 2005)
* (NBA, n.d.)