**INFO 6210**

**Data Management and Database Design**

**Physical Data Model and Normalization**

**Assignment 2**

Group Name: Shaktiman

Group Member 1: Suman Rawat (NUID 001058600)

Group Member 2: Shakti Chetan (NUID 001375816)

In this assignment, we are assumed to be working for a company called Nerd Analytics and that you are completely in charge of the database. We will create a physical database, normalize, actively stream social media data to the database we created in Assignment 1.

We have taken new dataset related to NBA Game. This dataset contains following entities:

1. games: all games from 2004 season to last update with the date, teams and some details like number of points, etc.

2. games\_details: details of games dataset, all statistics of players for a given game

3. players: players details (name)

4. teams: all teams of NBA

We have used raw csv data file as source. Extracting the raw csv data file for the video games sales data. This file includes all sales related data for different regions of US and Europe. Sales of the games are present in a yearly based fashion. To validate the scraped data, we will provide a list of best-selling video list to the csv file extracted data frame, this will give us the global and north America sale of the bestselling video games.

In this assignment, we have created a physical database, normalize, actively stream social media data to the database. Another group of statisticians and machine learning experts will be using the data that we have modeled, gather, clean and database analyze Social Media for an NBA Games domain. For NBA game domain we have entities like games, player, teams. We have performed database normalization on the game\_details table. Normalization involves refactoring a table into smaller (and less redundant) tables but without losing information; defining foreign keys in the old table referencing the primary keys of the new ones. The objective is to isolate data so that additions, deletions, and modifications of an attribute can be made in just one table and then propagated through the rest of the database using the defined foreign keys. We have checked whether our tables are in First normal form (1NF), Second normal form (2NF) and Third normal form (3NF). If not, we have restructured the database so that all the tables are in Third normal form; that is, we have normalized the database.

**Conceptual Schema of the Database**

A picture containing screenshot

Description automatically generated

**ER Diagram**

A screenshot of a social media post

Description automatically generated

**Physical Model**

A screenshot of a social media post

Description automatically generated

**Citations:**

1. NBA Dataset: https://www.kaggle.com/nathanlauga/nba-games#ranking.csv

2. ER Diagram: https://www.draw.io/

3. Normalization: https://www.studytonight.com/dbms/database-normalization.php

4. https://www.elstel.org/database/dbschemacmd.html.en

5. <https://www.nba.com/>

**Contribution Statement:**

1. Conception or design of the work – Shakti and Suman

2. Data collection for Web scraping: Suman

3. Data collection for web API and raw csv file: Shakti Chetan

4. Drafting the article: Suman

5. Creating Conceptual Schema: Shakti Chetan

6. Creation of ER Diagram: Suman Rawat

7. Creation of Physical Model: Shakti Chetan

6. Critical revision of the article: Suman and Shakti