

Informe de Arquitectura de Datos: Proyecto NBA (Grupo 7)

Objetivo: Explicar el modelo de datos "Estrella" diseñado para correlacionar el rendimiento físico con el éxito en la cancha.

Diccionario del Modelo de Datos

A. Tabla de Hechos: FACT_GAME (El Centro) Tabla GAME

Es la tabla que registra "qué pasó" en los partidos.

Categoría	Columnas Seleccionadas	Razón Técnica
Categoría	game_id, team_id_home, team_id_away	Para saber en qué partido y qué equipos jugaron.
Categoría	pts_home / pts_away, ast_home / ast_away	Los puntos y asistencias básicos.
Categoría	tov_home / tov_away (Turnovers)	Es vital ver qué tanto pierden el balón según su agilidad.
Categoría	blk_home / blk_away, stl_home / stl_away	Conecta la estatura y envergadura con los tapones.
Categoría	plus_minus_home / plus_minus_away	Indica si el equipo mejoró o empeoró con el rendimiento mostrado.

1. Identificadores

game_id: Es el identificador único de cada partido. Sin él, no podríamos distinguir un juego de otro ni evitar que los datos se dupliquen al sumar.

team_id_home / team_id_away: Son las llaves que conectan con nuestra DIM_TEAM. Nos permiten saber qué equipo estaba en cancha para poder cruzarlo con el físico de sus jugadores.

2. Métricas de Volumen (Rendimiento Básico)

pts_home / pts_away: Es el indicador de éxito más básico. Nos permite ver si los equipos con jugadores más altos anotan más puntos.

ast_home / ast_away (Asistencias): Mide la capacidad de creación de juego. Esto nos servirá para ver si los jugadores más bajos y ágiles (según la tabla de agilidad) son mejores pasadores.

reb_home / reb_away (Rebotes): Vital para tu análisis. Esta columna se cruzará directamente con la estatura (height) de la DIM_PLAYER para confirmar la hipótesis de que a mayor altura, más rebotes capturados.

3. Métricas de Impacto y Defensa (Rendimiento Avanzado)

blk_home / blk_away (Tapones): Esta es la columna "estrella" para tu análisis de envergadura (wingspan). Queremos demostrar que los brazos largos bloquean más tiros.

stl_home / stl_away (Robos): Mide la velocidad defensiva. La cruzaremos con los tiempos de agilidad del Draft Combine.

plus_minus_home / plus_minus_away: Es la calificación final del jugador. Nos dice si el equipo ganó o perdió terreno mientras esos jugadores específicos estaban en la cancha.

Escogimos estas columnas porque son operables matemáticamente y se correlacionan directamente con las pruebas físicas que estamos analizando (salto, altura y envergadura).

B. Dimensión: DIM_PLAYER Tabla Common_player_info

Columna en el Modelo	Origen en la Tabla	Tipo de dato final
ID_Jugador (PK)	person_id	Texto/ID
Nombre	display_first_last	Texto
Estatura	height	Número (tras normalizar)
Peso	weight	Número (tras normalizar)
Posición	position	Texto
País	country	Texto

Justificación de Columnas en DIM_PLAYER

- **person_id:** Es la columna más importante porque es nuestra **Llave Primaria (PK)**. Sin ella, es imposible conectar al ser humano con sus estadísticas de puntos en FACT_GAME o con su rendimiento físico en DIM_PHYSICAL.
- **display_first_last:** Es el nombre legible del jugador. En lugar de trabajar con códigos como "2544"

- **height y weight:** Son el núcleo de nuestra investigación sobre el biotipo. Aunque vienen como texto y requieren normalización, son las variables que usaremos para categorizar a los jugadores (ej. "Jugadores de más de 2 metros") y ver si esa altura realmente les da ventaja en los rebotes.
- **position:** Esta columna es un filtro estratégico. Nos permite segmentar el análisis: no podemos comparar el salto de un pívot (Center) con el de un base (Guard) sin entender que sus roles físicos son distintos.
- **country:** Nos da una dimensión geográfica. Permite analizar si hay tendencias físicas o de rendimiento según la procedencia del jugador, algo muy valioso para el mercado global de la NBA.

C. Dimensión: DIM_PHYSICAL (Extensión de Jugador) Tabla Draft_combine_stats

Categoría	Columna en la Tabla	Por qué seleccionarla
Llave de Conexión	player_id	Indispensable. Es el "gancho" que se unirá al person_id de tu DIM_PLAYER.
Métricas de Salto	max_vertical_leap	Es la medida estándar de explosividad en la NBA.
Métricas de Velocidad	three_quarter_sprint	Ideal para ver qué tan rápido cruzan la cancha los jugadores grandes.
Agilidad	lane_agility_time	Clave para los defensores; mide qué tan rápido mueven los pies lateralmente.
Envergadura	wingspan	Vital. En la NBA se dice que "la envergadura no se puede enseñar". Es mejor que la altura para predecir bloqueos.
Fuerza	bench_press	Para ver si la fuerza física se traduce en más rebotes ganados.

Justificación de Columnas en DIM_PHYSICAL

- **player_id:** Es nuestra **Llave Primaria (PK)** en esta dimensión. Se conecta directamente con el person_id de la tabla de jugadores para que podamos decir: "Este nombre y este apellido tienen este salto vertical".

- **max_vertical_leap (Salto Vertical Máximo)**: Es la métrica reina de la explosividad. La elegimos porque nos permite analizar el rendimiento de rebotes (reb) y tapones (blk) desde la potencia de las piernas, no solo desde la altura.
- **wingspan (Envergadura)**: Esta columna es más importante que la estatura para los defensores. La seleccionamos para probar que un mayor alcance de brazos tiene una correlación positiva directa con los robos (stl) y bloqueos (blk) en la tabla de hechos.
- **lane_agility_time (Agilidad en Pista)**: Mide la velocidad de desplazamiento lateral. Es fundamental para calificar a los jugadores defensivos; un tiempo bajo aquí debería traducirse en un mejor plus_minus defensivo en los partidos.
- **three_quarter_sprint (Velocidad de Pique)**: Nos da la velocidad punta. La elegimos para romper el mito de que los jugadores más pesados o altos son siempre los más lentos.
- **bench_press (Fuerza en Pecho)**: Mide la fuerza del tren superior. Es una métrica clave para los jugadores que "pelean" debajo del aro por los rebotes defensivos y ofensivos.

D. Tabla Puente: MAP_PLAYER_TEAM

Columna	Tipo de Dato	
person_id	TEXT / ID	Es la llave que conecta con la dimensión de jugadores (common_player_info).
team_id	INTEGER / ID	Es la llave que conecta con la tabla de hechos (game) y con la dimensión de equipos.
from_year	REAL / Año	Para saber cuándo empezó el jugador en ese equipo.
to_year	REAL / Año	Para saber cuándo salió.

El conector que rompe la relación de "Muchos a Muchos"

Justificación de las Columnas

- **person_id (PK/FK)**: Es el identificador del jugador. Nos permite saber exactamente a quién nos referimos cuando analizamos su físico en DIM_PLAYER.
- **team_id (FK)**: Es la conexión con la franquicia. Sirve para filtrar la tabla FACT_GAME y ver solo los partidos que le corresponden a ese equipo.

- **from_year y to_year:** Estas columnas son vitales para la **integridad histórica**. Nos permiten asegurar que estamos analizando el rendimiento del jugador en el equipo donde realmente estaba en ese año, y no mezclando datos de cuando era novato con los de cuando era veterano.

1. ¿Por qué creamos la Tabla Puente: MAP_PLAYER_TEAM?

Para evitar la **Relación Muchos a Muchos**

Un jugador (como LeBron James) ha jugado en varios equipos a lo largo de su carrera (Lakers, Heat, Cavaliers). A su vez, un equipo tiene muchos jugadores.

TABLA DIM TEAM

Columna	Tipo de Dato	Significado y Valor para el Análisis
team_id (PK)	INTEGER	Llave Primaria. Es el código único de la franquicia que se conecta con el "Puente" y con la tabla de hechos.
abbreviation	TEXT	El código de 3 letras (ej. <i>LAL, BOS, GS</i>). Ideal para etiquetas en gráficas pequeñas en Power BI.
nickname	TEXT	El nombre de batalla del equipo (ej. <i>Lakers, Celtics</i>). Es lo que el usuario final quiere leer.
city	TEXT	La ciudad sede. Vital para hacer análisis geográficos
year Founded	INTEGER	Permite ver si las franquicias más antiguas tienen una cultura de juego distinta.

Para completar el documento para tus compañeros, la **DIM_TEAM** es la tabla que organiza el caos de los miles de partidos por "Franquicias". En un modelo estrella, las dimensiones sirven para **agrupar y filtrar**, y aquí te explico por qué elegimos estas columnas de la tabla team:

Justificación de Columnas en DIM_TEAM

- **team_id (PK - Llave Primaria):** Es el ADN de la tabla. Sin esta columna, no podríamos conectar la información del equipo con la tabla de hechos (FACT_GAME) ni con la tabla puente (MAP_PLAYER_TEAM). Es lo que garantiza que los puntos de "Lakers" no se mezclen con los de "Celtics".
- **abbreviation (Abreviatura):** Seleccionamos esta columna porque en los tableros de Power BI el espacio es oro. Usar "LAL" o "GSW" permite crear

gráficos de barras y tablas mucho más limpias y fáciles de leer que usar el nombre completo.

- **nickname (Apodo del equipo):** Es la columna que humaniza el dato. En lugar de mostrarle al usuario un código numérico, mostramos "Lakers" o "Bulls". Es fundamental para que cualquier persona que vea el reporte entienda de qué equipo estamos hablando de inmediato.
- **city (Ciudad):** Esta columna es estratégica para el análisis geográfico. Nos permite responder preguntas como: *¿Los equipos de ciudades con mercados grandes atraen a jugadores con mejor perfil físico?* o comparar el rendimiento por zonas del país.

2. ¿Por qué creamos la Dimensión: DIM_TEAM?

Podríamos haber dejado el nombre del equipo dentro del puente o de los hechos, pero eso sería un error de arquitectura.

- **Normalización:** En lugar de escribir "Los Angeles Lakers" 50,000 veces en la tabla de hechos (lo que pesaría mucho), creamos una tabla pequeña donde el nombre aparece **una sola vez**.
- **Atributos Descriptivos:** Aquí guardamos la ciudad (city), la abreviatura (abbreviation) y el apodo (nickname).

La importancia de la Doble Relación:

Esta es la parte más avanzada. Como en un partido hay un equipo **Local** y uno **Visitante**, la DIM_TEAM es la que nos permite conectar con ambos (team_id_home y team_id_away).