

PLANIFICACIÓN TEMPORAL DEL EQUIPO DE TRABAJO

EQUIPO DE TRABAJO:

- David Camuñas Sánchez
- Mario Pérez Sánchez - Montañez

ÍNDICE

INTRODUCCIÓN	3
TECNOLOGÍAS UTILIZADAS:	3
JUSTIFICACIÓN DE LA UTILIZACIÓN DE ESTAS TECNOLOGÍAS	3
PLANIFICACIÓN TEMPORAL DEL EQUIPO DE TRABAJO	4
INICIO DEL PROYECTO: Lunes 10 de febrero.	4
FIN DEL PROYECTO: Martes 25 de febrero.	4
PLANIFICACIÓN EXTENDIDA	4
Lunes 10 de febrero	4
Miércoles 12 de febrero	4
Viernes 14 de febrero	4
Martes 26 de febrero	4
Miércoles 19 - Viernes 21 de febrero	4
Sábado 22 de febrero	5
Viernes 24 de febrero	5

INTRODUCCIÓN

Nuestro equipo de trabajo está formado por dos personas y el proyecto a realizar es un almacén de datos (Data Warehouse).

Los **datos** almacenados en este almacén tratan sobre estadísticas de la liga estadounidense de baloncesto, conocida como la NBA (National Basketball Associations). En concreto, estos datos pertenecen a la temporada 2018 - 2019.

TECNOLOGÍAS UTILIZADAS:

- Para el análisis y la extracción de datos, se llevará a cabo el proyecto con python. En detalle se realizará un estudio estadístico de los diferentes datos. Estos datos han sido obtenidos de la web oficial de la NBA (<https://es.global.nba.com/statistics/>), junto con la web Kaggle.
- Como motor de base de datos se utilizara SQLite. Nuestra base de datos estará formada por una serie de cuatro tablas.

JUSTIFICACIÓN DE LA UTILIZACIÓN DE ESTAS TECNOLOGÍAS

- Se ha elegido **python** como tecnología de extracción y tratamiento de datos, debido a que un lenguaje muy intuitivo, poco tipado y lo más importante que dispone de muchas librerías para el análisis de datos, las cuales, facilitan bastante la manipulación de los mismos.
- En cuanto al motor de base de datos, finalmente se ha optado por **SQLite**, debido a que es un motor con un nivel de compatibilidad alto con los distintos tipos de tecnologías, pero particularmente con python.

PLANIFICACIÓN TEMPORAL DEL EQUIPO DE TRABAJO

INICIO DEL PROYECTO: Lunes 10 de febrero.

FIN DEL PROYECTO: Martes 25 de febrero.

PLANIFICACIÓN EXTENDIDA

- **Lunes 10 de febrero**
 - Identificación de las herramientas de análisis de datos y de los motores de base de datos.
 - Finalmente se ha elegido python y SQLite.
- **Miércoles 12 de febrero**
 - Generación del archivo JSON para almacenar los datos obtenidos del sitio web.
 - Y así poder realizar más tarde los distintos estudios estadísticos de forma local en ese fichero.
- **Viernes 14 de febrero**

Problemas detectados en la obtención de datos en archivo con formato JSON.

 - Corrección de los problemas, finalmente se opta por la creación de varios archivos csv, donde cada archivo csv representa los datos que posteriormente se volcaran en las tres tablas de la Base de Datos.
- **Martes 26 de febrero**
 - Generación de los scripts en python para la realización del volcado de los datos desde la web a los archivos csv.
 - Creación del notebook, donde se irán realizando los distintos data frames y consultas (creación de gráficas) para favorecer el estudio de los datos.
- **Miércoles 19 - Viernes 21 de febrero**
 - Generación de data frames a partir de los archivos csv previamente creados. Para ello se utiliza varias librerías de python, como la librería “pandas”.
 - Creación de la base de datos, junto con las distintas tablas que la compondrán. Como son:
“PLAYERS”, “TEAMS”, ROOKIES” y “PLAYERS_TEAMS”.

Debido a la mayor complejidad de estas tareas, se ha otorgado un mayor tiempo de dedicación. Esto se suplirá trabajando el equipo el día: Sábado 22 de febrero.
- **Sábado 22 de febrero**
 - Creación de varias gráficas, para facilitar la visualización de los datos y su estudio.
 - Realización de varias consultas SQL, en las que se pedirá la obtención de una serie de datos almacenados en nuestra base de datos o Data Warehouse.

- **Viernes 24 de febrero**

- Realización de las transparencias pertenecientes a la presentación de nuestro Data Warehouse (almacén de datos).