TrackMan Business Intelligence Challenge

The challenges are an opportunity to demonstrate proficiency with the type of day to day work we would expect you to complete in this role at TrackMan. We want to get a sense of your thought process and the way you do the work. We also want you to get a feel for some of the tasks you'll encounter on the team.

After you complete the challenge, we will use your response as part of the follow up interview. You need to be comfortable with the approach taken and be ready to explain your thinking.

The challenges are intended to be around 2 - 4 hours of work. If there are parts you don't have time to implement, feel free to describe the next steps you would have taken in your submitted work. Also feel free to ask questions, but we will of course only answer to make the task instructions clearer and not how to complete it. In some cases, we might update the challenge with the additional information and reshare so that anyone completing the challenge has the latest information.

Star Schema Data Modelling

In this challenge you will develop a star schema data model. You will transform and lift the schema from the "raw" operational way we receive data from our tracking units and software to the business analysis "semantic" and visualization domain where it is easy to query and report on.

Here are some references if you are unfamiliar with these concepts
- https://radacad.com/power-bi-basics-of-modeling-star-schema-and-how-to-build-it
- https://www.youtube.com/watch?v=vjBprojOCzU
- https://medium.com/@marcosanchezayala/data-modeling-the-star-schema-c37e7652e206

The approach taken is up to you, but your results should optimize for fast query performance and simple flexible aggregations. To choose how to optimize you need to understand how the data will be used. We refer to these as the optimization criteria.

Hence, the challenge can broadly be defined as:
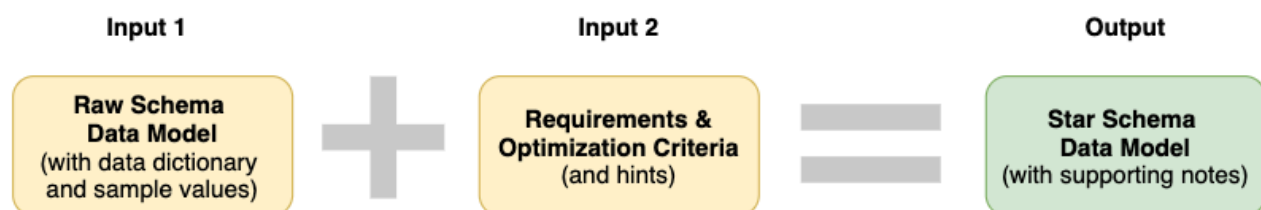
We will supply you the inputs. You should submit back to us the output.

Your Submission

We recommend using draw.io to show your schema solution to this problem. It is a simple cross platform web based drawing tool. Using draw.io will allow you to copy/paste parts of the input model we provide to the output as you develop your ideas. If you're having trouble, you are welcome to use any tool that works for you to provide the schema diagram (you can even hand draw it if feeling brave). The input model is shown below in case you can't open the supplied drawio version.

Using draw.io:Web access is here: https://app.diagrams.net/. If you prefer a local client version, download here: https://github.com/jgraph/drawio-desktop/releases

We strongly recommend supplying notes that support how you made decisions or where you were uncertain, essentially "show your work".
- The reasoning of the overall approach chosen to solve the challenge
- The reasoning for every active choice you made when solving the task
- Evaluation of your own results, where you were still uncertain and why

Please send your output model and supporting notes to the email you have been using to contact TrackMan already.

Your task

When a TrackMan customer uses one of our tracking units, such as the TM4, we refer to each shot they take in front of the unit that we track as a "Stroke". A group of strokes that represent the customer practicing or playing a round over some period is a "Session". Some data changes for each stroke, for example the tracking metrics like ball speed. Some data is constant throughout a session, like the device firmware version.

Input Model
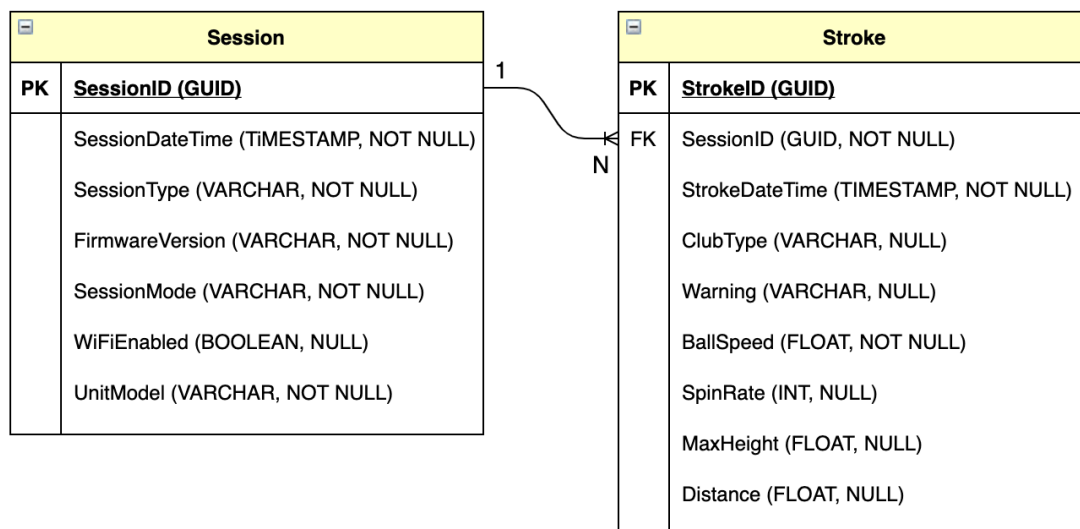
Legend:
- Each field has in brackets after the name the data type of that field and if it is nullable or not
- PK – Primary Key, FK – Foreign Key
- The relationship 1 to N, means one to many

And the companion data dictionary:

Table (size)	Field	Description
Session (30K / day)	SessionID	128 bit GUID that uniquely identifies the session
	SessionDateTime	Timestamp from the start of the session
	SessionType	Label that describes what sort of session the customer had, e.g. practice or game play (approx. 20 labels used)
	FirmwareVersion	Version of the Tracking firmware (approx. 100+ versions)
	SessionMode	Label that describes if the unit was set to be used indoor or outdoor or for putting (approx. 10 labels used)
	WifiEnabled	Flags if Wi-Fi connectivity was used for the session
	UnitModel	Label that describes the model of the unit (approx. 10 labels used)
Stroke (1.5M / day)	StrokeID	128 bit GUID that uniquely identifies the stroke
	StrokeDateTime	Timestamp from contact time of club to the ball
	ClubType	Label that describes the club used in the stroke, if supplied by the customer (approx. 25 labels used)
	Warning	Label describes if there was some technical warning state during the shot (approx. 5 labels used)
	BallSpeed	Decimal number in meters per second that describes the maximum ball speed observed
	SpinRate	Number in revolutions per minute (RPM) that describes the maximum ball spin observed
	MaxHeight	Decimal number in meters that describes the maximum ball height observed
	Distance	Decimal number in meters that describes the observed distance travelled by the ball

Requirements & Optimization Criteria

Please note – each of the optimization criteria have a purpose, they inform a particular element we are looking for in the output data model schema. Please read them carefully. If you cannot see how to utilize that criteria it is likely you have missed an opportunity to optimize the star schema data model.

1. Sessions data will be filtered in the BI tool (or used in a "Group By" in SQL) on each of SessionType, Firmware Version, UnitModel, SessionMode
2. Strokes data will be filtered in the BI tool (or used in a "Group By" in SQL) on each of Firmware Version, SessionMode, ClubType, Warning
3. Strokes data will also be filtered by a new data element BallSpeedThreshold (VARCHAR, NOT NULL), that categorizes the Stroke.BallSpeed in to 3 distinct categories
4. Session data will have a new data element SessionSeconds (INT, NOT NULL), that enables the average or distribution of session lengths to be calculated and visualized (FYI: this can be calculated from difference between first and last Strokes.StokeDateTime objects)
5. The ball tracking metrics in Stroke of BallSpeed, SpinRate, MaxHeight and Distance will only be used to calculate if SpinRate, MaxHeight, Distance are present vs BallSpeed, which is always present (today these are NULL or some positive number with many decimal places).
6. Be able to calculate frequency that Sessions used Wi-Fi connectivity.

Hints

- Star schema theory says tables should either be FACTS or DIMENSIONS. Facts are the data that is summarized (aggregated, filtered etc) and Dimensions are the fields that can be used to filter data.
- VARCHAR, GUID, FLOAT are expensive fields for a fact table, take all opportunity to optimize away
- BOOLEAN, INT are cheap fields for a fact table, use where possible
- When adding new fields, think if these should be a column in a fact table or a new dimension
- See which fields can be removed to optimize the size of the model (only the criteria above need to be supported)

Good Luck!