Charger sessions and telemetry

At Wallbox we have a growing network of chargers sending events to our cloud servers. Here we present you with 2 event streams:

(you can download the 2 event streams as JSON files from here: Charger logs.zip)

- charger_log_session: events coming from each charging session (one per finished session). A charging session is defined as the time between the EV gun is first plugged by the user, until it is finally unplugged. Therefore, this time is necessarily equal or longer than the actual charging time.

Field name	Description	Units / Type
`id`	Charging session identifier.	
`charger_id`	Identifier of the charger being used.	
`user_id`	Identifier of the person using the charger.	
`energy`	Total session energy supplied to the car.	Watts-Hour (Wh)
`total_cost`	Total cost calculated as the energy consumed multiplied by the unit energy cost (whether default (0.2 €/kWh) or set by the admin of the charger).	Euros (€)

`start_time`	Session starting time.	UNIX timestamp
`end_time`	Session ending time.	UNIX timestamp

- Charger_log_status: events coming from chargers every 30 seconds. They contain useful information such as the current, voltage and temperature for each of the 3 phases corresponding to the three-wire circuit of the charger.

Field name	**Description**	**Units / Type**
`id`	Event identifier.	
`charger_id`	Identifier of the charger being used.	
`ac_current_rms_1`	AC current RMS in phase 1. Value 438 indicates a maximum calculated current of 4.38A.	Centiamperes (cA)
`ac_current_rms_2`	AC current RMS in phase 2.	Centiamperes (cA)

`ac_current_rms_3`	AC current RMS in phase 3.	Centiamperes (cA)
`ac_voltage_rms_1`		Volts (V)
`ac_voltage_rms_2`		Volts (V)
`ac_voltage_rms_3`		Volts (V)
`temperature_1`	Temperature measured on phase 1.	Celsius (ºC)
`temperature_2`	Temperature measured on phase 2.	Celsius (ºC)
`temperature_3`	Temperature measured on phase 3.	Celsius (ºC)
`charger_timestamp`	Timestamp of measurement.	
`server_timestamp`	Timestamp of database insert. It is the same for as many rows as they were sent in the batch process of updating the database.	

Generate a briefing of each charging session with:

- Session time
- Charging profile with variables such as:
    - Temperature

    - Voltage
    - Current
    - Energy
    - Accumulated energy
- Take the charging profile and generate a report with only big changes along the charging session (¿more than 10% increment compared to last value?)
- Which is the average session time per charger?

Generate a notebook with some queries and visualizations to answer all these questions.

Possible languages:

- SQL
- Scala/Java/Kotlin (Spark?)
- Python (Spark/Pandas?)

Architecture questions

- How would you store the data so that it could accommodate different use cases (visualization, ad-hoc analysis, EDA, predictive analytics)
- How would you design a data pipeline that ingests and processes all these events in near-real-time to generate charging profile dashboards for each charging session? Imagine we could show these metrics in a small dashboard in our myWallbox portal/app.
- How would you manage to avoid schema changes breaking down data pipelines?
- How would you do to generate some anomaly detection models in near-real-time based on the received telemetry data?