

What are **two** benefits of storing this data in MongoDB with JSON over a relational database management system such as Postgres? Please reference specific examples from the **business** collection to back up your claims.

Format your answer as follows:

1. Benefit #1, Example #1.
2. Benefit #2, Example #2.

Limit each benefit to one sentence and each example to one sentence for a total of at most four sentences.

1. Benefit #1: MongoDB's flexible schema allows businesses to include different sets of fields without requiring a fixed structure, Example #1: Flying Elephants at PDX has a RestaurantTakeOut attribute, where Oskar Blues taproom does not.
2. Benefit #2: MongoDB supports nested fields, enabling more natural representations of complex attributes, Example #2: Oskar Blues Taproom stores hours and attributes as nested documents, capturing daily open hours and amenities like WiFi and BikeParking in a single, self-contained structure.

0.0.1 Question 2d

In the last question, you performed equivalent left joins in both Postgres and Mongo. Now, examine their query plans, paying special attention to `executionTimeMillis`. Which database system was faster? What gives the database system you chose an advantage over the other? Keep your response to at most three sentences.

PostgreSQL was significantly faster, was an `executionTimeMillis` of approximately 4ms compared to MongoDB's 416ms. This performance gap is due to PostgreSQL's use of optimized hash join and efficient indexing strategy. In contrast, MongoDB's `$lookup` stage performs a collection scan and is generally slower for large joins unless explicitly optimized with embedded documents or denormalization.

1. What do you notice about how the columns of `business_df` are constructed, e.g. how are fields and subfields represented in the dataframe?
2. How are values that are not found in every document handled in the pandas dataframe?

Hint: You will need to horizontally scroll to view all the column names and values.

Format your answer as follows and use one sentence per question:

1. Sentence 1
2. Sentence 2

1. Nested fields and subfields in `business_df` are flattened using dot notation, where each subfield becomes a separate column with a name like `attributes.Smoking`.
2. Values that are not found in every document are represented as NaN (Not a Number) in the pandas dataframe.

