**Introduction**

When working with multiple data files in Apache Spark, understanding the structure and data types of the datasets is crucial for efficient processing and transformation. The inferSchema option in Spark allows automatic detection of column data types, eliminating the need for manual specification.

**Benefits of Using inferSchema**

1. **Automatic Data Type Detection**: Spark automatically determines the data types of columns, reducing the need for manual type assignments.
2. **Improved Data Accuracy**: Ensures that numerical columns are correctly interpreted as integers, floats, or doubles instead of defaulting to strings.
3. **Time Efficiency**: Saves time by avoiding manual schema definitions, especially when dealing with multiple files with similar structures.
4. **Consistency Across Files**: When applied to multiple files, inferSchema ensures uniformity in column data types, reducing potential inconsistencies.
5. **Better Performance in Queries**: Inferring correct data types optimizes query execution as numerical operations and aggregations perform better on appropriately typed columns.
6. **Ease of Data Exploration**: Helps in understanding the dataset quickly without prior knowledge of the schema, which is useful in exploratory data analysis.

**Why inferSchema Was Used**

In the given scenario, inferSchema was applied to study and determine the data types of the provided 10 files. The key reasons for using it include:

* **Understanding Data Structure**: Before performing transformations, it was essential to identify column types.
* **Avoiding Manual Schema Definition**: Since multiple files were involved, manually defining schemas for each file would be inefficient.
* **Ensuring Consistency**: The automatic schema inference helped maintain consistency across the files, making subsequent processing smoother.
* **Facilitating Data Processing**: Correct data types improved Spark’s ability to handle computations and queries efficiently.

**Conclusion**

The use of inferSchema in Spark proved to be a practical approach to automatically determine data types, ensuring accuracy, efficiency, and consistency. This feature is particularly useful when dealing with large datasets with unknown schemas, ultimately enhancing data processing workflows.