**Spark在书写的时候，可以用三种方式：**

**approach 1: use RDD (not recommend)  
approach 2: use Dataframe, register the RDD to a dataframe (recommend for DE)  
approach 3: use SQL (recomend for data analysis or DS)**

标准 SQL 的书写顺序 vs 执行逻辑顺序（非常经典）：

A screenshot of a phone

AI-generated content may be incorrect.A screenshot of a phone

AI-generated content may be incorrect.

DataFrame 没有严格的“语句顺序”，而是通过 **链式调用**（method chaining）来表达操作。但大致逻辑和 SQL 对应。

A screenshot of a computer code

AI-generated content may be incorrect.

A screenshot of a computer screen

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.

A screenshot of a computer

AI-generated content may be incorrect.