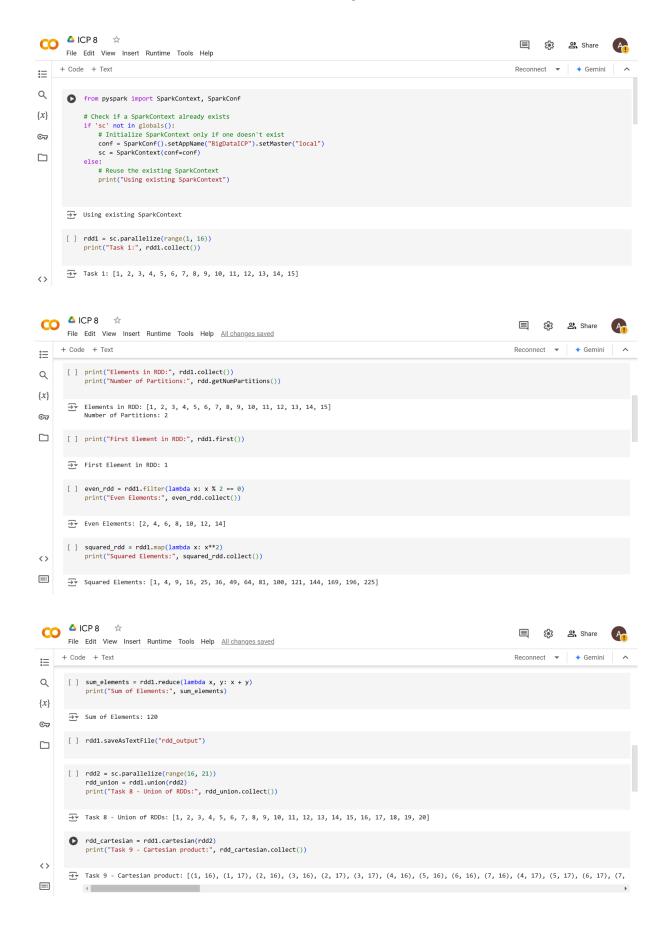
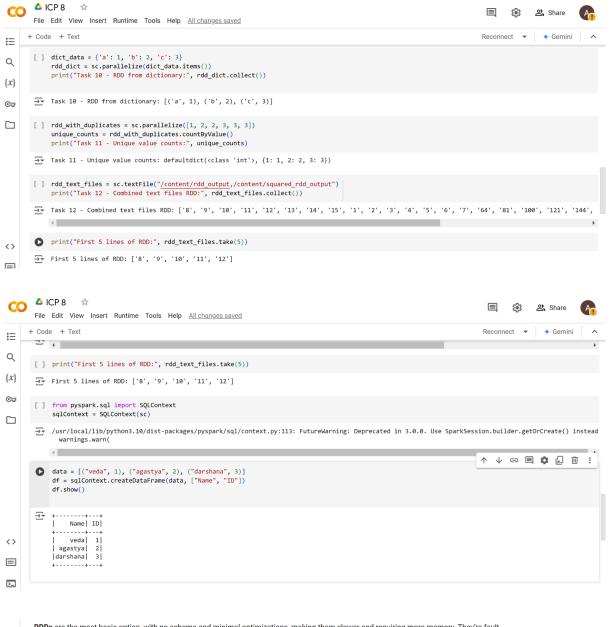
## ICP8





RDDs are the most basic option, with no schema and minimal optimizations, making them slower and requiring more memory. They're fault-tolerant and type-safe, catching errors early, but lack SQL support and are generally harder to use. DataFrames, on the other hand, are structured like tables with schemas, which allows them to be optimized for faster performance. They support SQL queries and are high-level, making them easier to work with, though they detect errors at runtime. Datasets combine the best of both, offering the structure and SQL support of DataFrames, with added type safety and the highest performance due to advanced optimizations. They're also the most memory-efficient, making them ideal for complex application

GitHub Link: https://github.com/w8162583/bda.git