RDDs vs DFs - When to Use?

Difference	RDD	Dataframe
Structured data	Difficult to handle structured and semi-structured data.	Much easier and simpler to handle structured data.

Difference	RDD	Dataframe
Structured data	Difficult to handle structured and semi-structured data.	Much easier and simpler to handle structured data.
Infer schema	Can not infer schema.	Can infer schema. Schema can be defined as well.

Difference	RDD	Dataframe
Structured data	Difficult to handle structured and semi-structured data.	Much easier and simpler to handle structured data.
Infer schema	Can not infer schema.	Can infer schema. Schema can be defined as well.
Optimization	No in-built optimization.	Optimised using Catalyst optimizer.

Difference	RDD	Dataframe
Structured data	Difficult to handle structured and semi-structured data.	Much easier and simpler to handle structured data.
Infer schema	Can not infer schema.	Can infer schema. Schema can be defined as well.
Optimization	No in-built optimization.	Optimised using Catalyst optimizer.
Data manipulation	Using functional programming.	Using high-level operations that are much more expressive.

Difference	RDD	Dataframe
Structured data	Difficult to handle structured and semi-structured data.	Much easier and simpler to handle structured data.
Infer schema	Can not infer schema.	Can infer schema. Schema can be defined as well.
Optimization	No in-built optimization.	Optimised using Catalyst optimizer.
Data manipulation	Using functional programming.	Using high-level operations that are much more expressive.
Expertise	Data engineers or software engineers aware of functional programming.	Anyone comfortable with building SQL like queries.

Analytics Vidhya

When to use RDDs?

- Need to deal with unstructured data.
- Comfortable with writing functional programming code.
- Forgo code optimization.

When to use DFs?

- Need to deal with structured or semi-structured data.
- Not comfortable with writing functional programming code.
- Code optimization.

Thank You!!