

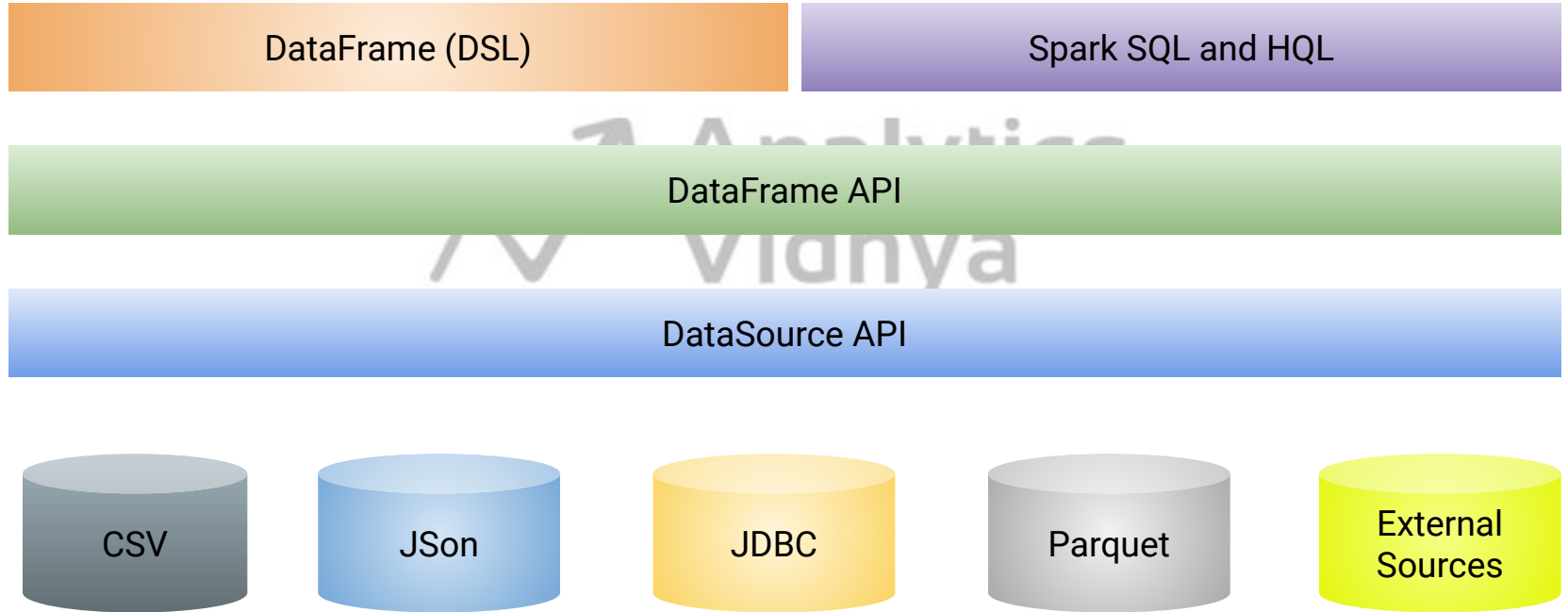


Spark SQL

Spark SQL

- Spark API for structured data processing.
- Contains information about structure of data and computation being performed.
- Spark SQL uses internal optimizations.

Spark SQL Architecture



Spark SQL Use cases

- Banking: Credit card fraud detection
- Real time analysis of stock market data
- Genomic sequencing: Determining the sequence of DNA of an organism
- And more

Integrated



```
In [4]: #import findspark
#findspark.init()
from pyspark import SparkContext, SparkConf
from pyspark.sql import SparkSession
conf=SparkConf()

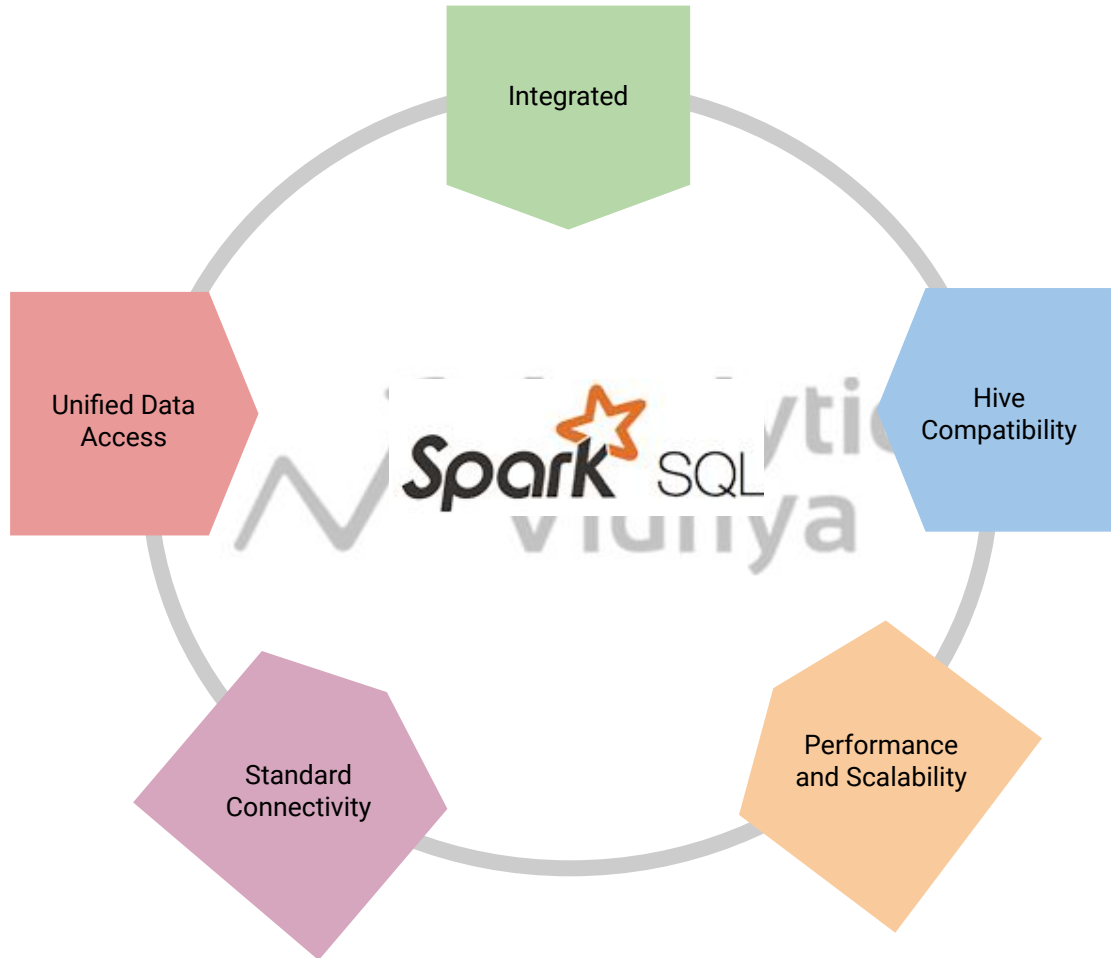
spark = SparkSession.builder.config(conf=conf).appName('DataValidation').getOrCreate()
```

```
In [5]: df=spark.read.format("csv").option("header",True).load("/home/siddharth/Documents/CDP/Datasets/African_Mobile_Data.csv")
```

```
In [6]: df.createOrReplaceTempView("mobile")
```

```
In [8]: spark.sql("select * from mobile limit 10").show()
```

Date	Country	City	Region	Segment	Sales	Profit
1/1/13	Cote d'Ivoire	Abidjan	Western	Commercial	\$656.96	\$6.57
1/1/13	Madagascar	Antananarivo	Eastern	Public	\$875.94	\$-70.08
1/1/13	Rwanda	Kigali	Eastern	Public	\$258.35	\$18.08
1/1/13	Zimbabwe	Harare	Eastern	Residential	\$875.62	\$-35.02
1/2/13	Ethiopia	Addis Ababa	Eastern	Residential	\$509.93	\$10.20
1/3/13	Chad	N'Djamena	Middle	Commercial	\$443.13	\$4.43
1/3/13	Malawi	Blantyre	Eastern	Residential	\$651.78	\$19.55
1/3/13	Mauritius	Port Louis	Eastern	Public	\$896.64	\$-62.76
1/4/13	Sao Tome and Prin...	Sao Tome	Middle	Commercial	\$226.92	\$13.62
1/5/13	Burundi	Bujumbura	Eastern	Commercial	\$928.07	\$74.25



Thank You!!

