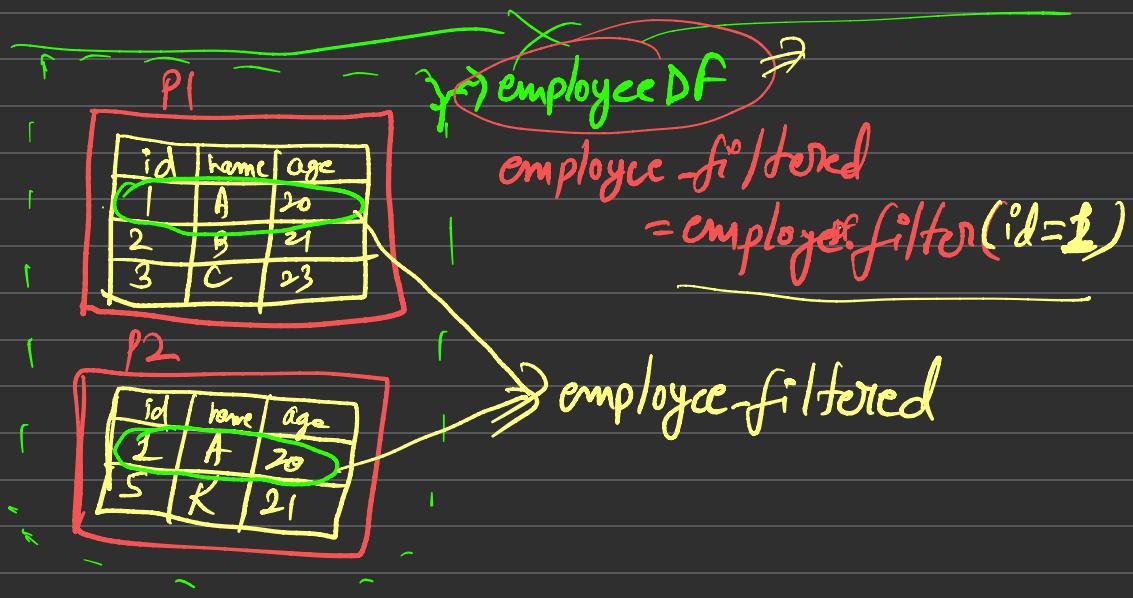
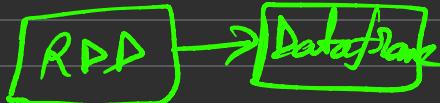


`textdata = spark.read.text("content.txt")`

↙ RDD → X Not Dataframe

`csvdata = spark.read.csv("employee.csv")`

↙ RDD → Dataframe



Lazy Evaluation

Python

```
a=3  
b=2  
c=a+b  
print(c)
```

C++

```
int a; }  
int b; }
```

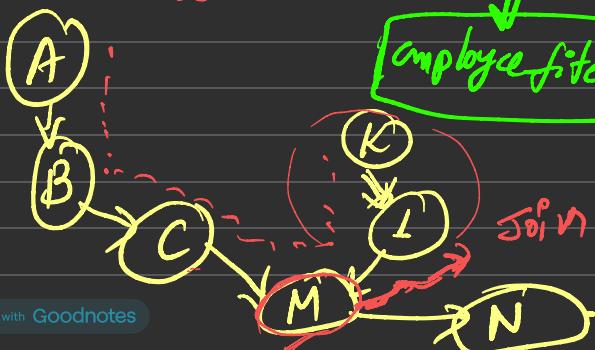
```
a=3; }  
b=2; }
```

```
c=a+b  
cout<<c;
```

Spark Application

spark = SparkSession.builder.getOrCreate()

T {
 employee_df = spark.read.csv("employees.csv")
 employee_filtered = employee_df.filter(id==2)
 employee_filtered.count() } A



Job, Stage, Task

orders = spark.read.csv("orders.csv")

payments = spark.read.csv("payments.csv")

customers = spark.read.csv("customers.csv")

orders.filter = orders.filter(status="Delivered")

join { payments.filter = payments.filter(status="Success")
orders-with-company = orders.filter.withColumn

("Company_name", "Stage")

join2 { customer.filter = customers.filter("Country = USA")
pay-with-order = orders-with-company.join(
pay-with-order.write()) payments.filter)

Pay-with-customer = pay-with-order.join

(customer.filter)

