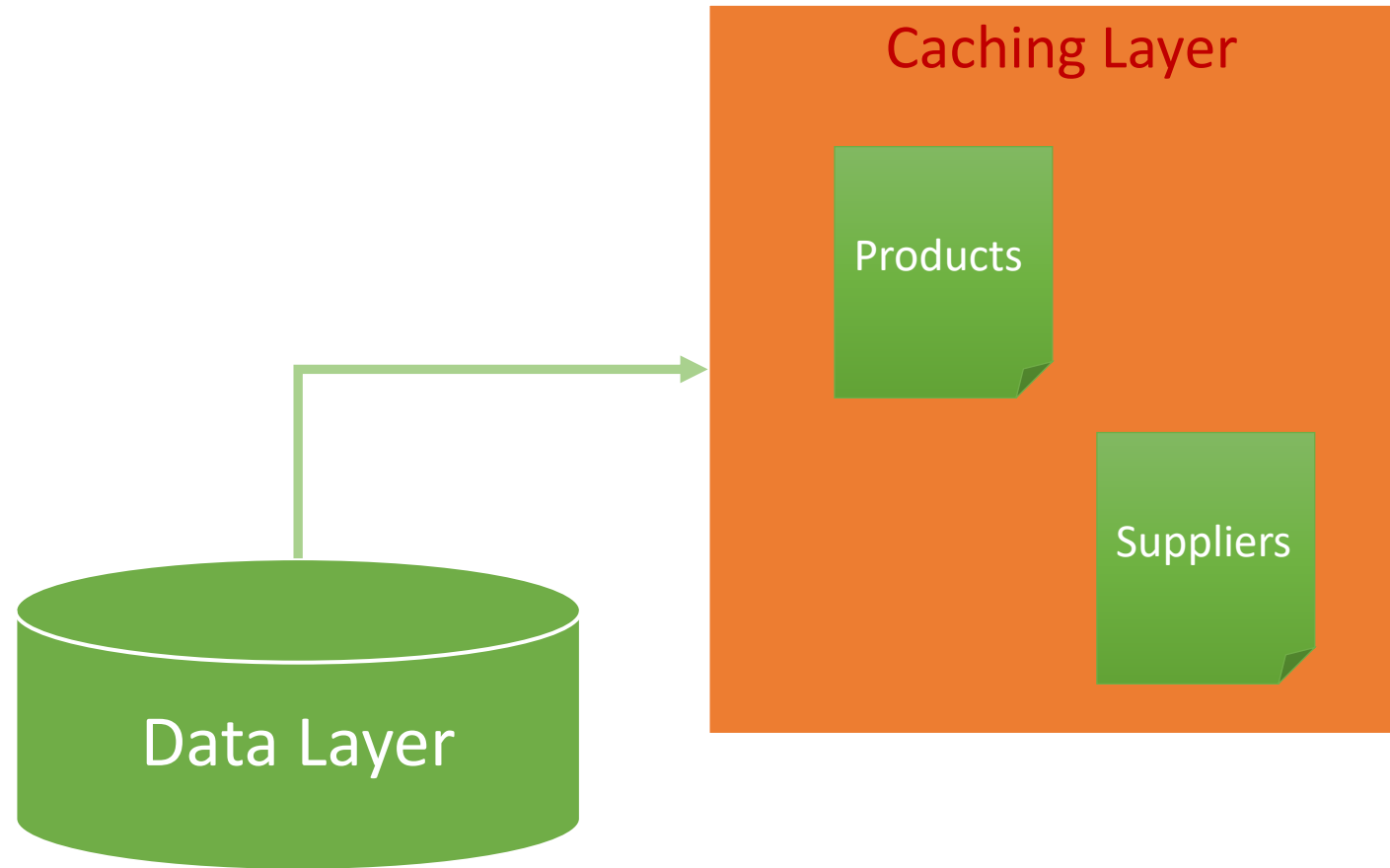




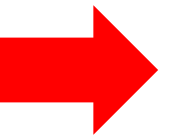
Cache Data in Spark





How do you cache data in Spark?

```
%sql  
cache table test_tbl as  
select * from test_tbl
```





What will it do?

```
%sql  
show views
```

Create a session scope temporary view

	namespace ▲	viewName ▲	isTemporary ▲
1		test_tbl	true





How to know if your view is in cache?

Jobs Stages **Storage** Environment Executors SQL JDBC/ODBC Server Struct

Storage

Parquet IO Cache

Look into Spark UI Storage tab

Data Read from External Filesystem (All Formats)	Data Read from IO Cache (Cache Hits, Compressed)	Data Written to IO Cache (Compressed)	Cache Misses (Compressed)	True Cache Misses
0.0 B	0.0 B	0.0 B	0.0 B	0.0 B

▼ RDDs

You will see your view

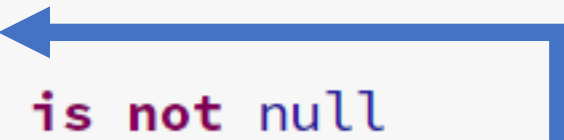
ID	RDD Name	Storage Level
18	In-memory table test_tbl	Disk Memory Deserialized 1x Replicated





How to use cached view?

```
%sql  
select count(distinct callType) as distinct_call_type_count  
from test_tbl  
where callType is not null
```



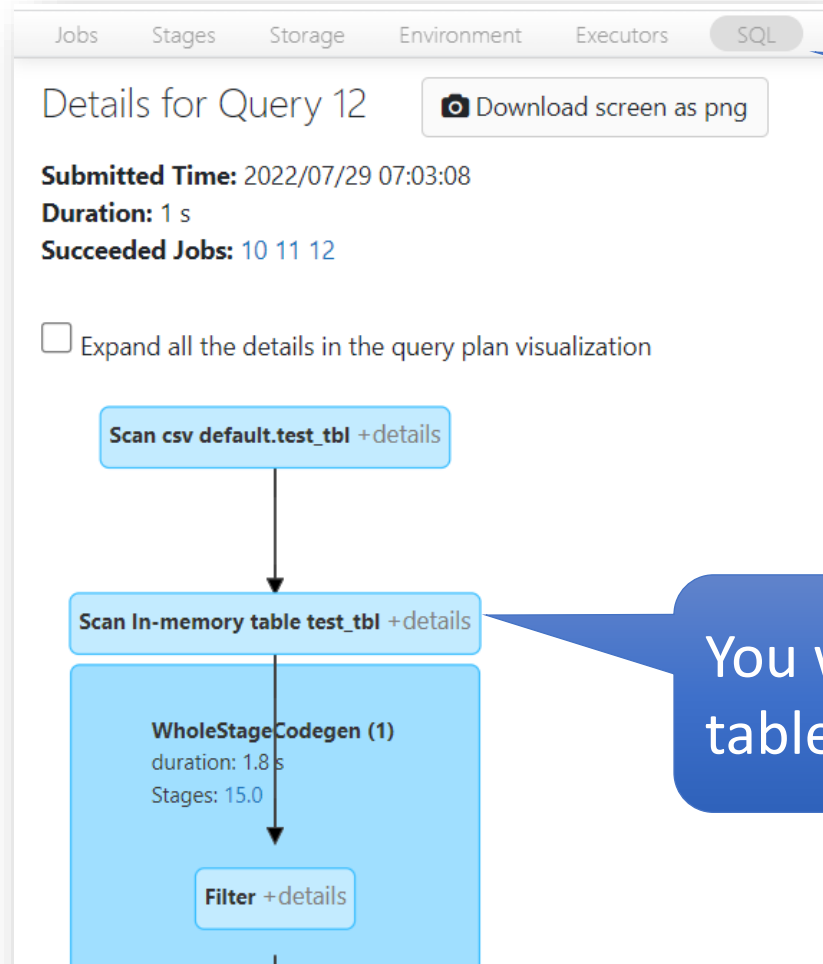
As simple as a table or
view



How to see if your query is using cached view?

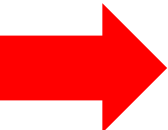


ScholarNest



Look into Spark UI
SQL tab

You will see In-memory
table scan



How to un-cache data from Spark memory?



ScholarNest

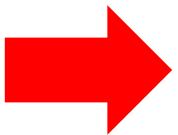
```
%sql  
uncache table test_tbl
```





What if someone forget to un-cache?

```
%sql  
clear cache
```



Summary



ScholarNest

- CACHE TABLE creates a temporary view to cache the data in memory
- You can create cache view with the same as table name
- Refer your cached view to access data from cache
- Cache view is session scoped.
- You can see all the cached views in the Spark UI -> Storage tab
- You will see in-memory table scan when your query refers to cached view
- UNCACHE is a best practice





Spark Data Engineering with Databricks Spark Certification Course

For course inquiries

Email:

sushant@scholarnest.com

WhatsApp:

9353465988