

Discord's Data Platform



Discard's Data Platform

Why	companies	need	a	data	plat	Jarm	?
					•		

- business decisions and Strakgies

- power data science - train ML models

Data doesn't lie

Making sense from both the

- deep product insights

Can they not work with dutabases directly?

- each service has a separate PB
- databases can be of different kinds

eg: profile service -> MongoDB

eg: profile service - Mysal data requires them at the payment service - Mysal same place (DB)

- cannot make cross DB "joins" and "queries"

Joins and computation

- analytics queries will choke transactional PB throughput

Hence all the data is duplicated and stoned in a centralized location - Data Warehouse / lake

4 unified view of data / information

Grunching stat / insights is easy

Discord's Data Playform - Derived

different end use cases

Paymenk Paymenk Paymenk Porders Por					
Profile Databases Payment Databases Big Query, Redshift Unified view of the data 1. Non-relational Data Sources Rig Query, Redshift Unified view of the data	Taxasackianal			Data	cuare house
Orders Orders III unified view of the data 1. Non-relational Data Sources 2 2 1. Non-relational Data Sources	O whale as at			Big	avery, Redshift
So, what is the challenge here? 1. Non-relational Data Sources 1. Non-relational Data Sources	Paymenk — — —	\rightarrow			
So, what is the challenge here? 1. Non-relational Data Sources Of the data	Onders				united view
So, what is the challenge here?					•
1 Q . 1 MI	So, what is the challenge here?				
1 Q . 1 MI	1. Non- Relational Data Spirates	2	7	2	
2. Different data views for	1. Tota Metalionar Data socials	_ ^_	\sim	\sim	
	2. Different data views for	Insight	Broduct	ML	

eg: Table 1 + Table 2 + Table 3 -> Recommendation

Table 1 + Table 5 - Retention analysis

Table 2 + Table 4 — Payment insights

Note: different tables need to be joined/merged

and kept to efficiently power usecases

* End usecase should need to do minimal joins

Auth Profile Order

User detail Core Tables

User detail Derived Tables

Insights Data Science ML

Creating a new derived table

New table is configured in YAML along with details Such as

The sac habite 12 ton Jiganita in

1. columns of new table

2 strakgy: merge, append, replace

- 4 while writing the new values what should we do
- 4. Window 5. partition-by
- 6. cluster-by Sort within pointition
- 7. dataset 8. sql → select query that 1s executed and

output is put in the derived table

