**Client Name**: AIArtify Inc.  
**Scenario**:  
AIArtify generates anime-style images using prompts from users and shares them across social media platforms. They now want to:

* Analyze image engagement (likes, shares, comments)
* Track GPU usage and generation time per platform
* Identify ethical concerns
* Find high-performing prompts and user behaviors

They've collected a large dataset and want to store it in **HDFS**, manage it via **Hive**, and monitor jobs through **Ambari**.

|  |  |
| --- | --- |
| Column Name | Description |
| image\_id | Unique image identifier |
| user\_id | Unique user ID |
| prompt | Prompt text used |
| likes | Number of likes |
| shares | Number of shares |
| comments | Number of comments |
| platform | Social media platform |
| generation\_time | Time to generate image (in seconds) |
| gpu\_usage | GPU usage during generation (%) |
| file\_size\_kb | Image file size in KB |
| resolution | Image resolution (e.g., 1024x1024) |
| style\_accuracy\_score | How closely it resembles Studio Ghibli (0–100) |
| is\_hand\_edited | Manual edit flag (Yes/No) |
| ethical\_concerns\_flag | Ethical issues flagged (Yes/No) |
| creation\_date | Creation date |
| top\_comment | Most popular comment |

SELECT COUNT(\*) FROM ghibli\_image\_data;

1. Which platform has the highest average engagement (likes + shares + comments)?

SELECT platform,

AVG(likes + shares + comments) AS avg\_engagement

FROM ghibli\_image\_data

GROUP BY platform

ORDER BY avg\_engagement DESC;

2. Which prompts generated the highest style accuracy scores?

SELECT prompt, style\_accuracy\_score

FROM ghibli\_image\_data

ORDER BY style\_accuracy\_score DESC

LIMIT 10;

3. Find the top 5 users with the most images flagged for ethical concerns.

SELECT user\_id, COUNT(\*) AS flagged\_images

FROM ghibli\_image\_data

WHERE ethical\_concerns\_flag = 'Yes'

GROUP BY user\_id

ORDER BY flagged\_images DESC

LIMIT 5;

4. What is the average GPU usage and generation time by platform?

SELECT platform,

AVG(gpu\_usage) AS avg\_gpu,

AVG(generation\_time) AS avg\_time

FROM ghibli\_image\_data

GROUP BY platform;

5. How many images were manually edited and also flagged for ethical concerns?

SELECT COUNT(\*) AS count\_edited\_flagged

FROM ghibli\_image\_data

WHERE is\_hand\_edited = 'Yes'

AND ethical\_concerns\_flag = 'Yes';