Lecture 32:
• Exam 2 overview
• Hive wrap up

Announcements:
• HW-5 out
• R-5 out
• Exam 2: Friday (4/19)
• Last Assignment: Out next Wednesday-ish

---

Exam 2 Overview

Basics:
• Closed notes, etc.
• Worth 50 points (like Exam 1)
• 4 multi-part questions

Rough Question Mapping: see study guide for details
• Traditional query processing
• Open Data Formats
• BigQuery ML, Spark/Dataproc, Data Warehousing/Looker Concepts
• Architectures (shared-*, etc.) and Hive basics
Hive Example

Example (†): Generate tables using HiveSQL’s multi-table insert

- Facebook status updates and profile information
- To identify trends in status changes for a day by gender and school

```
FROM (  
    SELECT s.status, p.school, p.gender  
    FROM status_updates s JOIN profiles p ON (s.user_id = p.user_id)  
    WHERE s.ds = '2009-03-20'  
) s1

-- load school summary table
INSERT OVERWRITE TABLE school_summary PARTITION (ds='2009-03-20')  
SELECT s1.school, COUNT(*)  
GROUP BY s1.school

-- load gender summary table
INSERT OVERWRITE TABLE gender_summary PARTITION (ds='2009-03-20')  
SELECT s1.gender, COUNT(*)  
GROUP BY s1.gender

(†) From: Thusoo et al. “Hive – A Warehousing Solution Over a Map-Reduce Framework”, VLDB’09
```