



Why?

## Why data systems?

What about data structures from algorithms class?

Spreadsheets? Text files?

## Why so many design choices?

SQL? Big data? (MapReduce, Hadoop, Spark)

Learning systems? (TensorFlow, deep learning)

## What's exciting for future?

[Shiva@stanford.edu](mailto:Shiva@stanford.edu) -- “Shiva” Shivakumar

## Interests

Peta scale data systems (from cs145 -> Infolab -> now)

## Building new data systems, products (and teams)

- Scaled to billions of consumers, billions of ad \$\$, millions of web publishers, trillions of data rows, million QPS systems
- E.g., AdSense, Search, Dremel/BigQuery, Gmail/Google Apps, Sitemaps, Warp, Google Maps, etc.
- VP @Google, Founder of 2 startups (Urban Engines, Gigabeat), angel investor/advisor in more data/AI startups

# Example: Youtube DB

YouTube

funny cats

About 12,100,000 results

Home

Trending

Subscriptions

History

Watch later

Purchases

Liked videos

CRPTIONS

Popular on YouTu...

Music

Sports

Gaming

FROM YOUTUBE

YouTube Premium

Movies & Shows

**REMOVE CAT PEE STAINS**

**How to Get Rid of Cat Pee Stains**

Ad BISSELL • 2M views

Your cat had an accident on the carpet. BISSELL is here to help!

**CATS make us LAUGH ALL THE TIME! - Ultra FUNNY CATS**

Tiger FunnyWorks • 100K views • 3 days ago

Ultra funny cats and Kitten that will make you cry with laughter! Cats are the best laugh all the time! This is ...

New

**You will LAUGH SO HARD that YOU WILL FAINT - FUNNY C compilation**

Tiger FunnyWorks • 19M views • 8 months ago

Well well well, cats for you again. But this time, even better, even funnier, even more like these furries the ...

**Have you EVER LAUGHED HARDER? - Ultra FUNNY CATS**

Tiger FunnyWorks • 124K views • 1 week ago

Super funny cats and Kitten that will make you scream with laughter! This is the LAUGH challenge ever!

YouTube

cats funny

**Animal Planet Videos**

Baby Cats - Funny and Cute Baby Cat Videos Compilation (2018) Gatos Bebés Video Recopilación

99,337 views

4.7K 768

SHARE

SUBSCRIBE 337K

Baby Cats - Funny and Cute Baby Cat Videos Compilation (2018) Gatos Bebés Video Recopilación

Animal Planet Videos

Published on May 23, 2018

Subscribe Here: <https://goo.gl/qor4XN>

SHOW MORE

Upload video

Go live

Shop now >>

Up next

**Top Cats Vs. Cucumbers - Funny Cat Videos Compilation**

Animal Planet Videos • 23M views

**Funny Elias play with the wheel on the bus and another toys - Elias Adventures**

291 watching

**LIVE NOW**

**LIVE: Rescue kitten nursery! TinyKittens.com**

TinyKittens HQ

1.3K watching

**LIVE NOW**

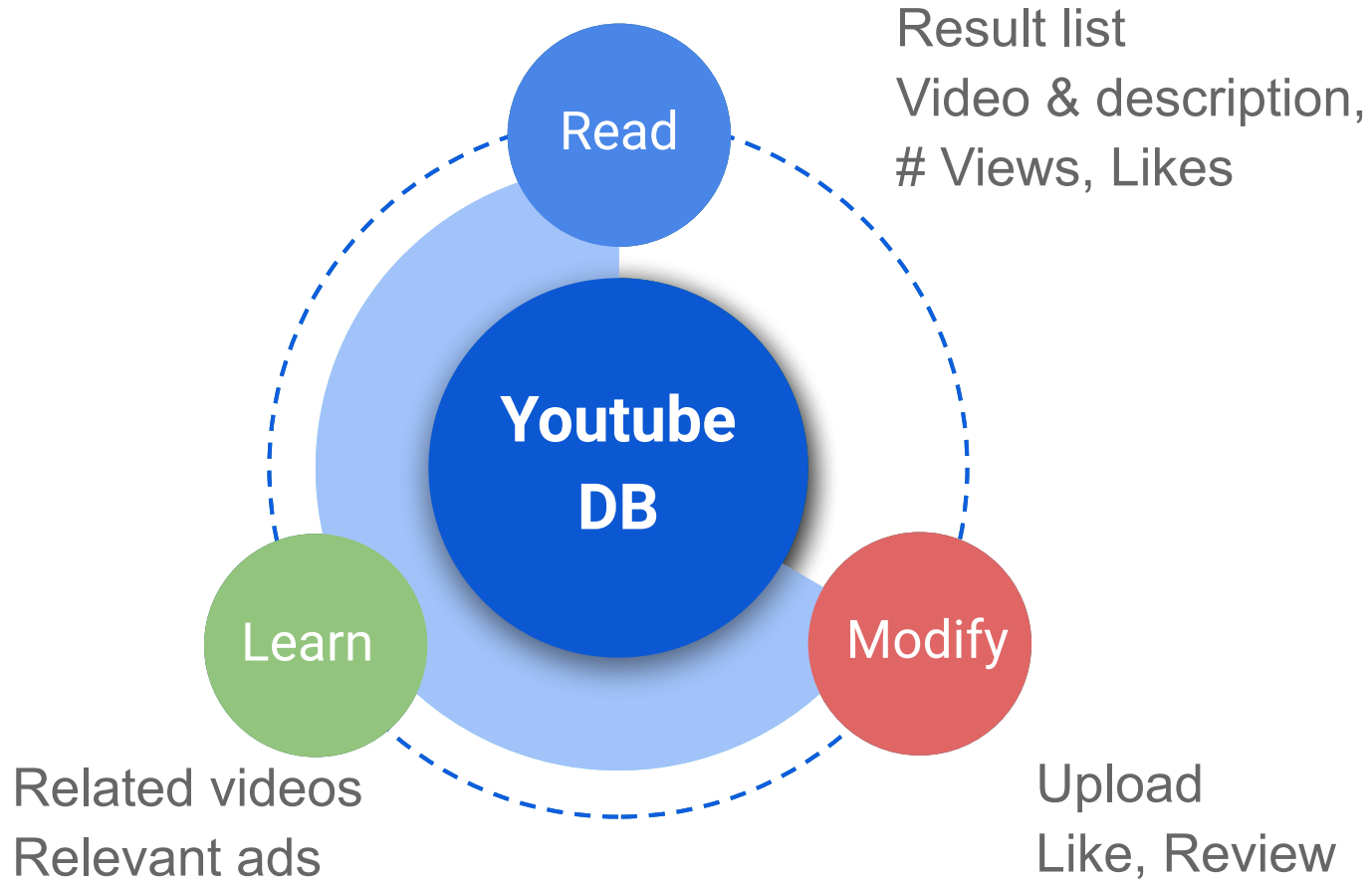
**Puss in boots and the three diabolos [HD]**

Mathew Garcia

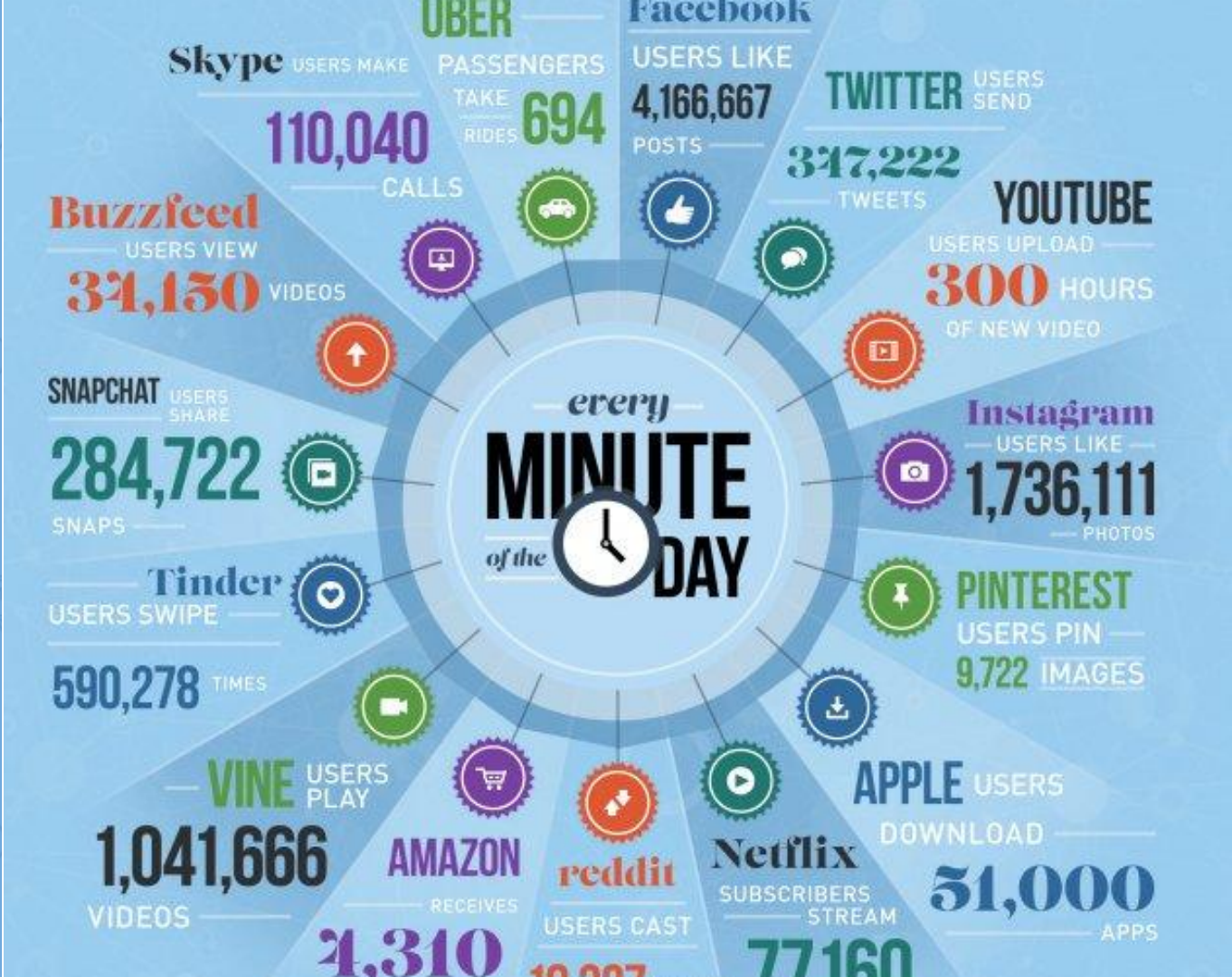
7.6M views

Example

Unpack  
Youtube DB



Every minute  
on the  
Internet







**Example**  
**Self Driving Cars**



# THE COMING FLOOD OF DATA IN AUTONOMOUS VEHICLES

RADAR  
~10-100 KB  
PER SECOND

SONAR  
~10-100 KB  
PER SECOND

GPS  
~50KB  
PER SECOND

CAMERAS  
~20-40 MB  
PER SECOND

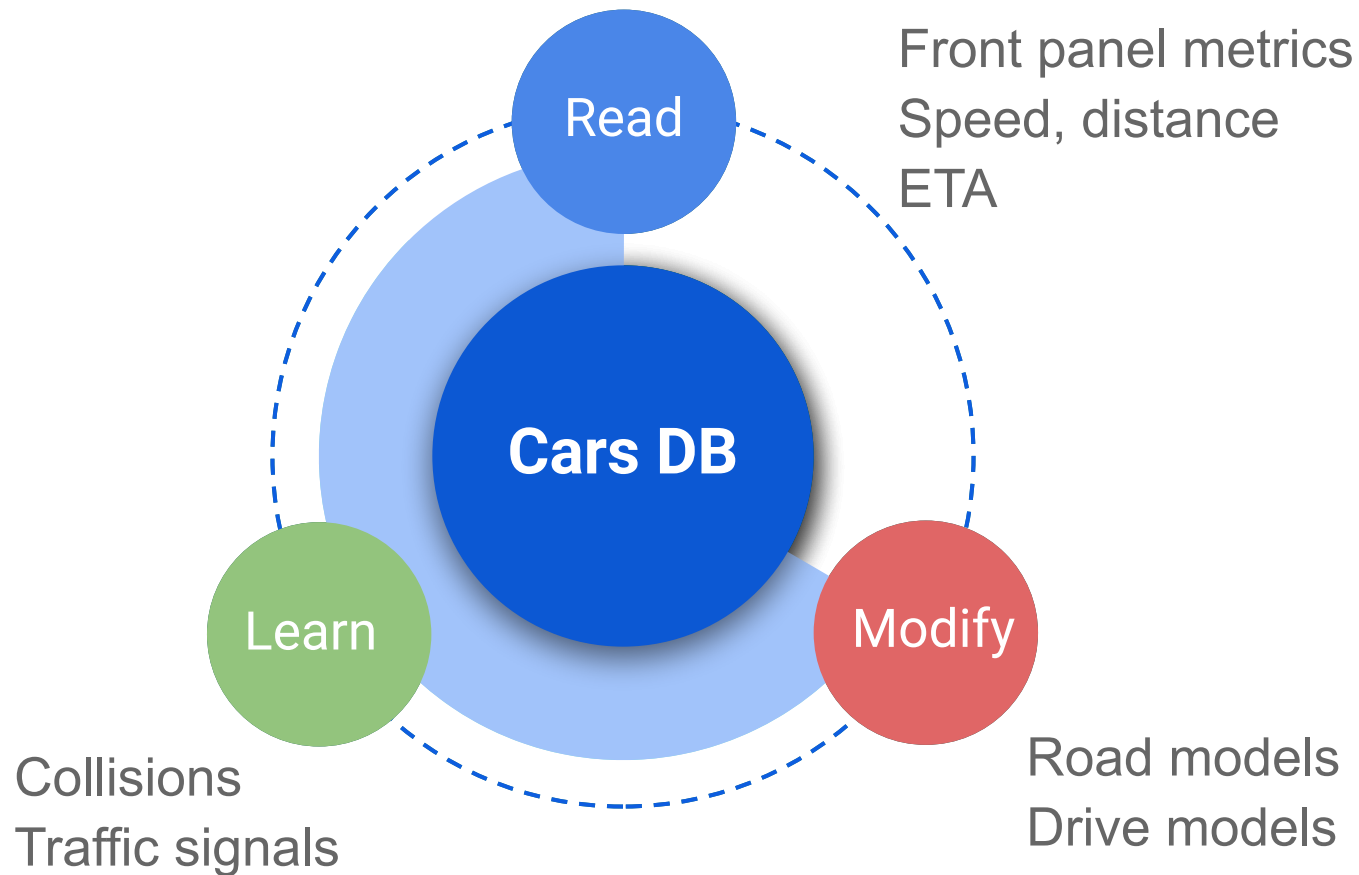
LIDAR  
~10-70 MB  
PER SECOND

AUTONOMOUS VEHICLES  
**4,000 GB**  
PER DAY... EACH DAY



Example

Unpack  
Cars DB

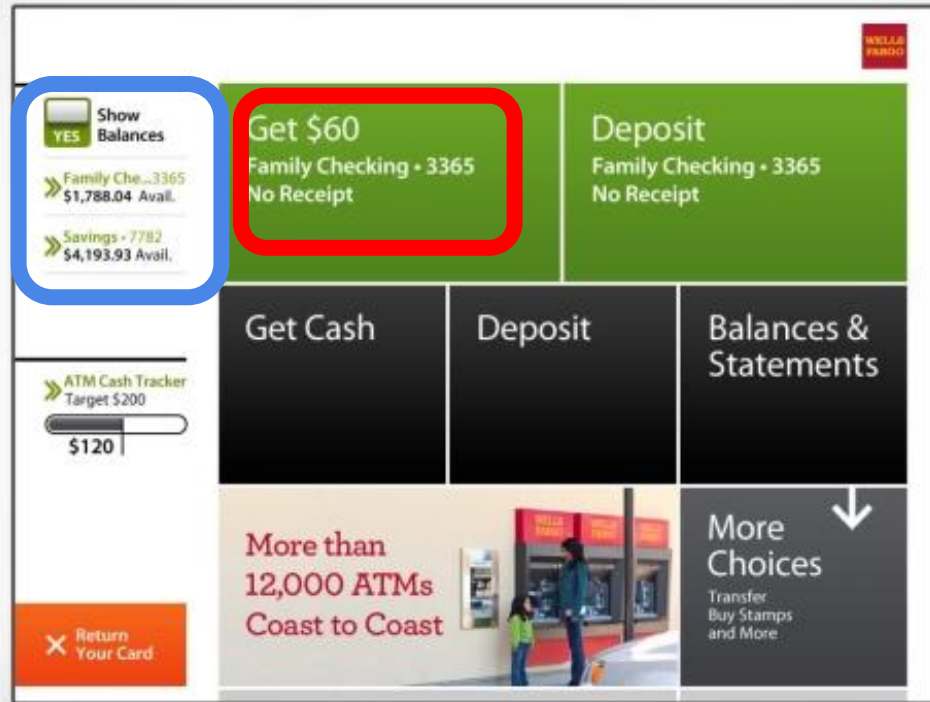




Example

Unpack  
ATM DB:

Transaction



Read Balance  
Give money  
Update Balance

vs

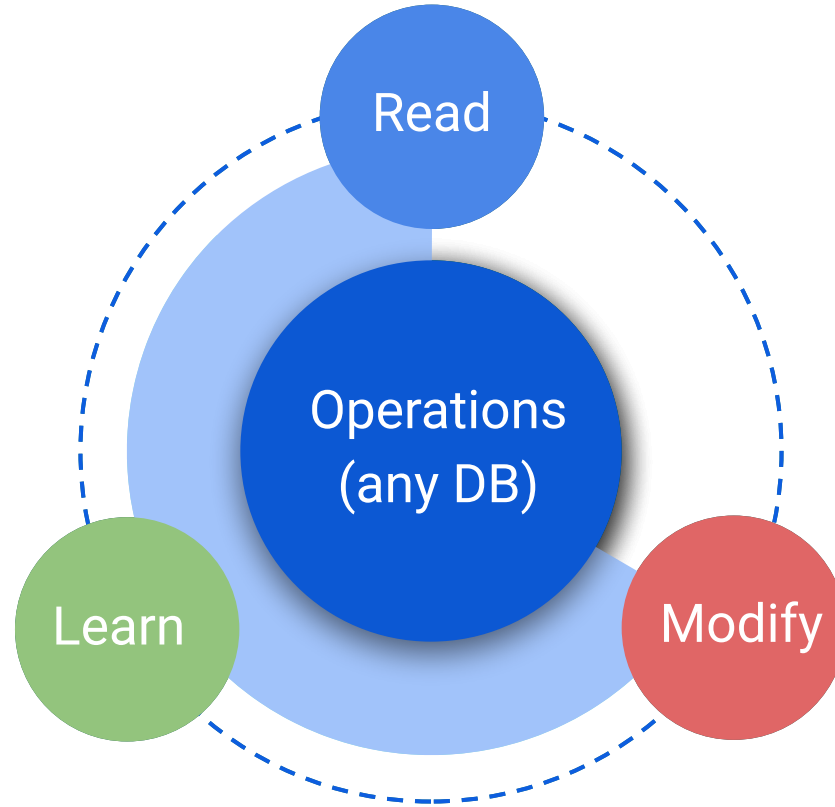
Read Balance  
Update Balance  
Give money

# Goals of Databases

Platform to store, manage data

Supporting

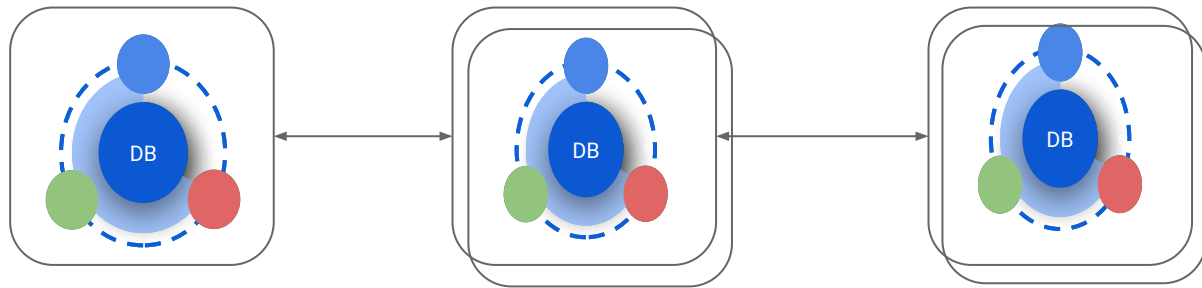
Scale  
Speed  
Stability  
Evolution  
Reliability  
Cost efficiency



# Goals of Data systems



Connect one/many DBs for custom system



Store current data  
(e.g., lot of reads)

Optimize historical  
data (e.g., logs)

Run batch  
Workloads  
(e.g. training)

When to  
build a  
custom data  
system?



The background of the slide is a solid blue color. It is decorated with a repeating pattern of white line-art icons. These icons include a document with lines, a pie chart, an envelope, a speech bubble, a clock, a checkmark inside a circle, a smartphone, a tag, and a presentation board with a line graph.

How?

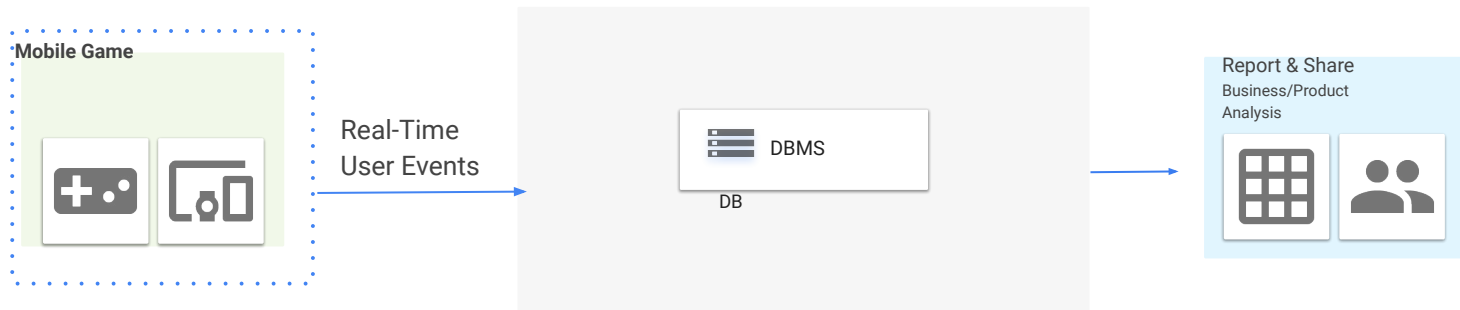
For  
Whom?



# How?

## Example Game App

DB v0



Q1: 1000 users/sec?  
Q2: Offline?  
Q3: Support v1, v1' versions?

App designer

Q7: How to model/evolve game data?  
Q8: How to scale to millions of users?  
Q9: When machines die, restore game state gracefully?

Systems designer

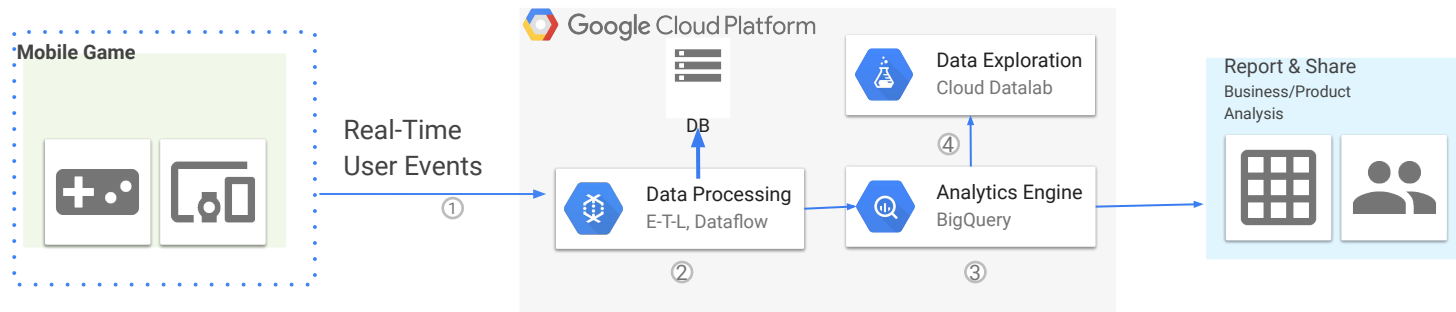
Q4: Which user cohorts?  
Q5: Next features to build?  
Experiments to run?  
Q6: Predict ads demand?

Product/Biz designer

# How?

## Example Game App

### Data system “v1” on Cloud



1 Log user actions

2 Store in DB, after  
Extract-Transform-Load

3 Run queries in a peta  
scale analytics system

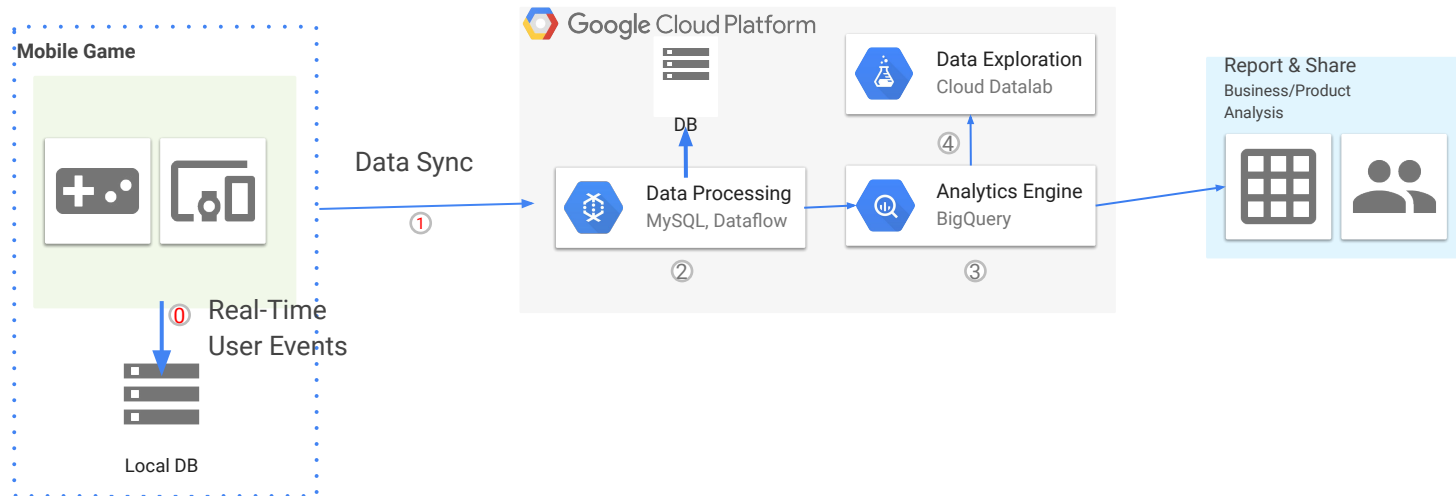
4 Visualize query results

How?

# Example Game App

Data system

“v2” Cloud +  
Local



0 Log user actions  
In local DB

1 Data sync to cloud

2 Store in DB, after ETL

3 Run queries in a petabyte  
scale analytics system

4 Visualize query results



# Course Summary

We'll learn...

- **How to query a database? With concurrent users and crashes / aborts?**

1. Intro

**2-3. SQL**

4. ER Diagrams

5-6. DB Design

**7-8. TXNs**

11-12. IO Cost

14-15. Joins

16. Rel. Algebra



# Course Summary

We'll learn...

- How to query a database? With concurrent users and crashes / aborts?
- **How to design a database?**

1. Intro

2-3. SQL

**4. ER Diagrams**

**5-6. DB Design**

7-8. TXNs

11-12. IO Cost

14-15. Joins

16. Rel. Algebra





# Course Summary

We'll learn...

- How to query a database? With concurrent users and crashes / aborts?
- How to design a database?
- **How to optimize performance?**

1. Intro

2-3. SQL

4. ER Diagrams

5-6. DB Design

7-8. TXNs

**11-12. IO Cost**

**14-15. Joins**

**16. Rel. Algebra**



# Logistics

[cs145.stanford.edu](https://cs145.stanford.edu)

**Assessment** Projects: 35% (5 + 10 + 20), Midterm Exam: 25%, Final Exam: 40%. We will be offering extra credit for (a) insightful in-class and OH participation (b) high quality answers to fellow student questions in piazza.

## Exam dates

- Midterm: 1 hr in class midterm on Nov 1st(Thurs)
- Final: Friday, Dec 14th from 3:30-6:30 p.m. No alternate times, plan ahead.
- **Accommodations** Students who may need an academic accommodation based on the impact of a disability must initiate the request with the Office of Accessible Education (OAE) and notify us at least 7 days (ONE week) prior to the Midterm and/or Final Exam.

## Homeworks/Suggested reading

- Posted on cs145 webpage. We'll use Gradiance. It's an auto-grader to help reinforce concepts AS-You-GO. Strongly encouraged for prepping materials for exams.