



Privacy & Data Protection Office



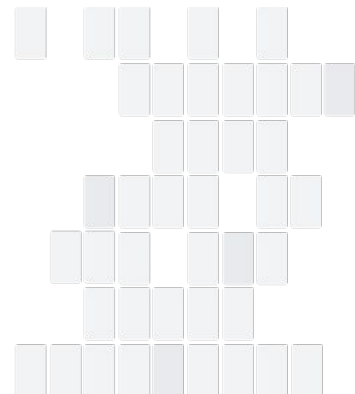
Anon

Improving Usability of Differential Privacy at Scale

PEPR '20, October 2020

Miguel Guevara
Product Manager

Milinda Perera
Software Engineer



1.0

Usability Problem of Differential Privacy



A simple dataset ...

Movie ratings:

Customer ID	Date	Rating	Movie
81478	1999-12-15	3	Mulan
92729	2000-10-24	5	The Piano
245371	2001-01-12	1	Office Space
383404	2005-02-02	4	The Matrix

A simple aggregation ...

Movie counts by date and rating:

```
SELECT
  date, rating,
  COUNT(movie_id) AS movie_count
FROM movie.ratings
GROUP BY 1, 2;
```



Date	Rating	Movie Count
1999-12-15	3	1
2000-10-24	5	60
2001-01-12	1	793
2005-02-02	4	8043

Same aggregation with anonymization ...

Using a [SQL engine](#)* to query with Differential Privacy:

```
SELECT WITH ANONYMIZATION OPTIONS(epsilon = 1.0986, delta = 0.00001)
  date, rating,
  ANON_COUNT(movie_id CLAMPED BETWEEN 0 AND 70) AS movie_count
FROM movie.ratings
GROUP BY 1, 2;
```

* Differentially Private SQL with Bounded User Contribution

Wilson et al., PoPETs, 2020

Same aggregation with anonymization ...

Using a SQL engine to query with Differential Privacy:

```
SELECT WITH ANONYMIZATION OPTIONS(epsilon = 1.0986, delta = 0.00001)
  date, rating,
  ANON_COUNT(movie_id CLAMPED BETWEEN 0 AND 70) AS movie_count
FROM movie.ratings
GROUP BY 1, 2;
```



The usability problem



2.0

Quantifying Privacy and Utility



Quantifying Privacy and Utility

Why? Bridge the usability gap of Differential Privacy

How?

- Define **privacy vs utility** metrics
- Provide infrastructure to **safely and efficiently** compute them **at scale**
- Allow **self-service** model



3.0

Demo!

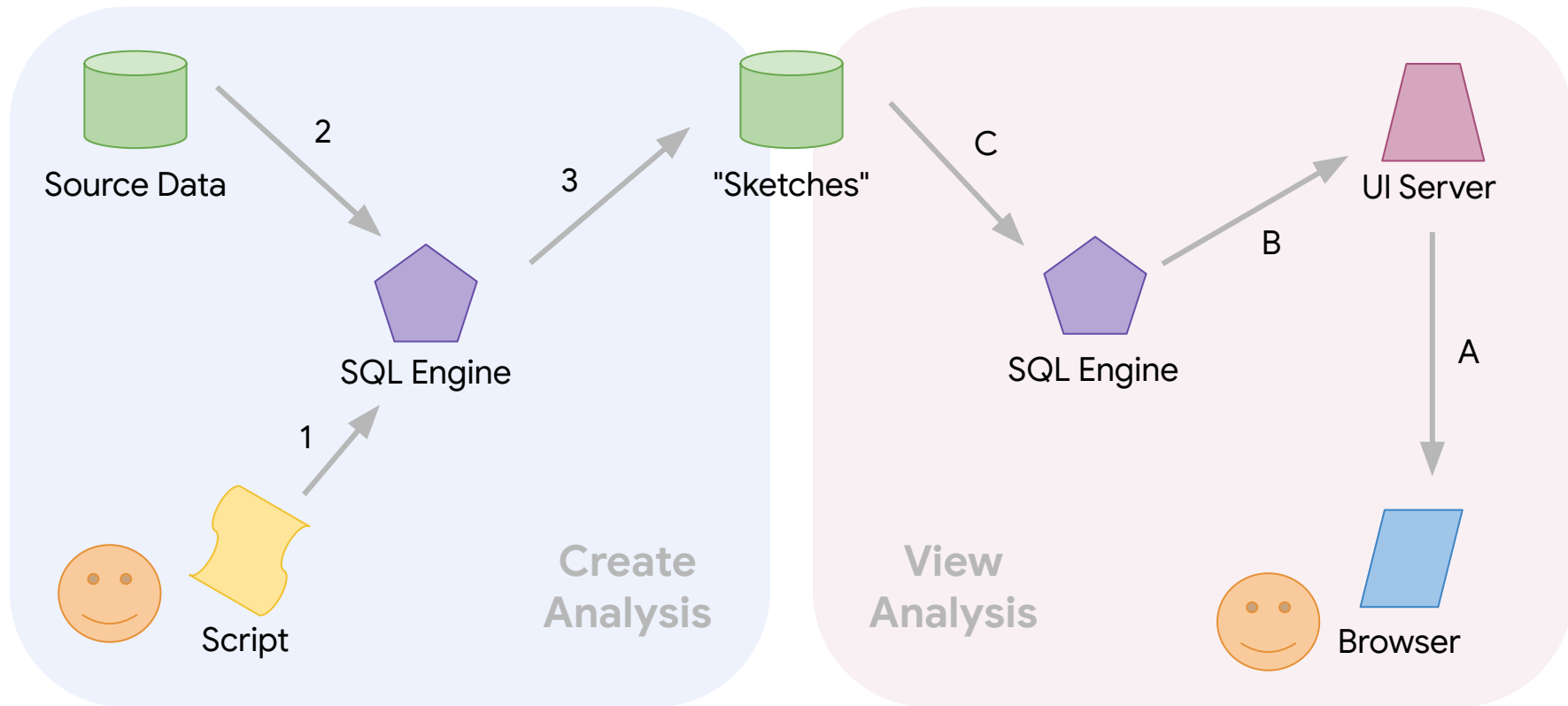


4.0

System Architecture



System Architecture



5.0

Highlights



Highlights

- Median query latency within **seconds!**
- End-to-end analysis for most datasets takes only **minutes**
- **Intuitive** utility metrics for teams

"This is super useful for tuning parameters, we were missing something just like that :)"

- A happy product team

6.0

Future Work



Future Work

- Open source this work
- Local Differential Privacy
- More Functions (AnonMean, AnonMedian, AnonQuantiles)

Thank you!



Anon