Linked in

Audience Engagement API: A Privacy Preserving Data Analytics System at Scale

Presenter: Ryan Rogers

Collaborators: Subbu Subramaniam, Sean Peng, David Durfee, Seunghyun Lee, Santosh Kumar Kancha, Shraddha Sahay, Parvez Ahammad

PEPR '20

Agenda

1 Overview of Differential Privacy

2 Application

3 Overall Privacy System

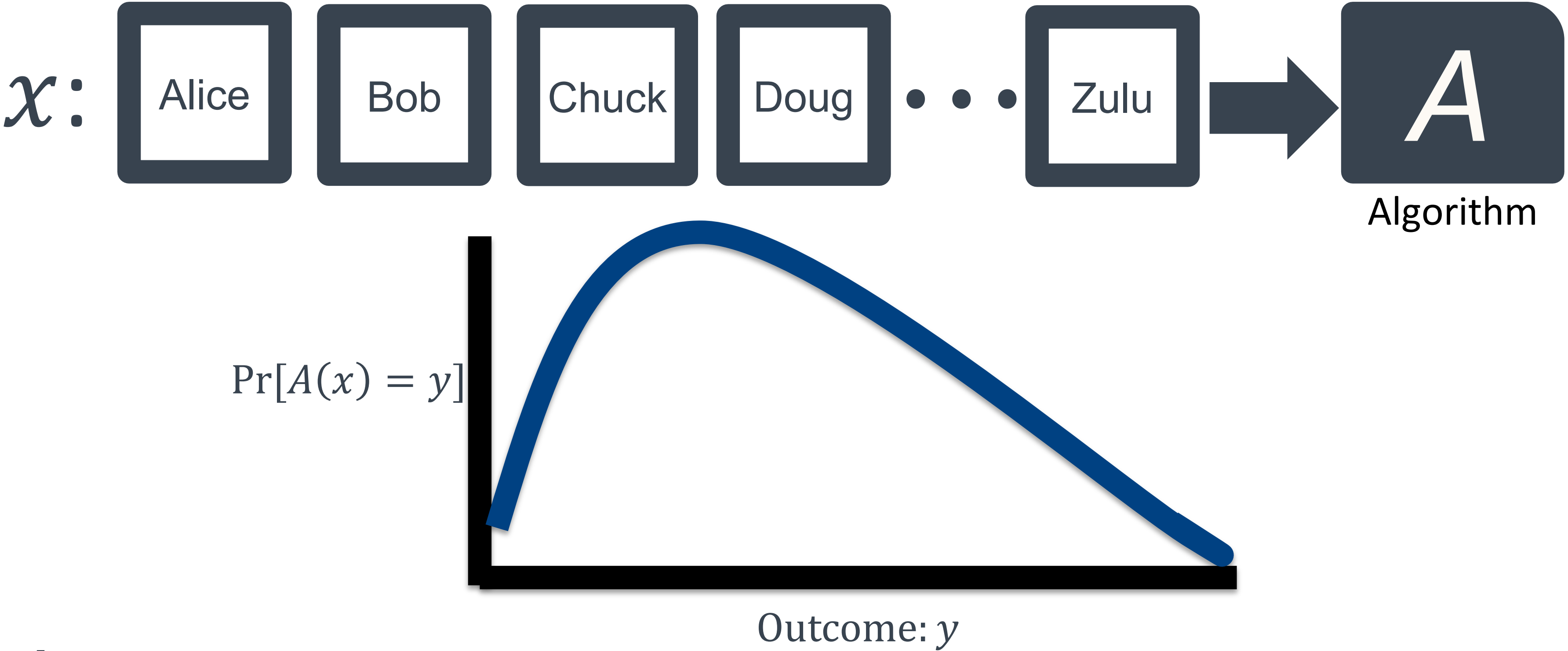
Mission

Utilize data while protecting the privacy of users.

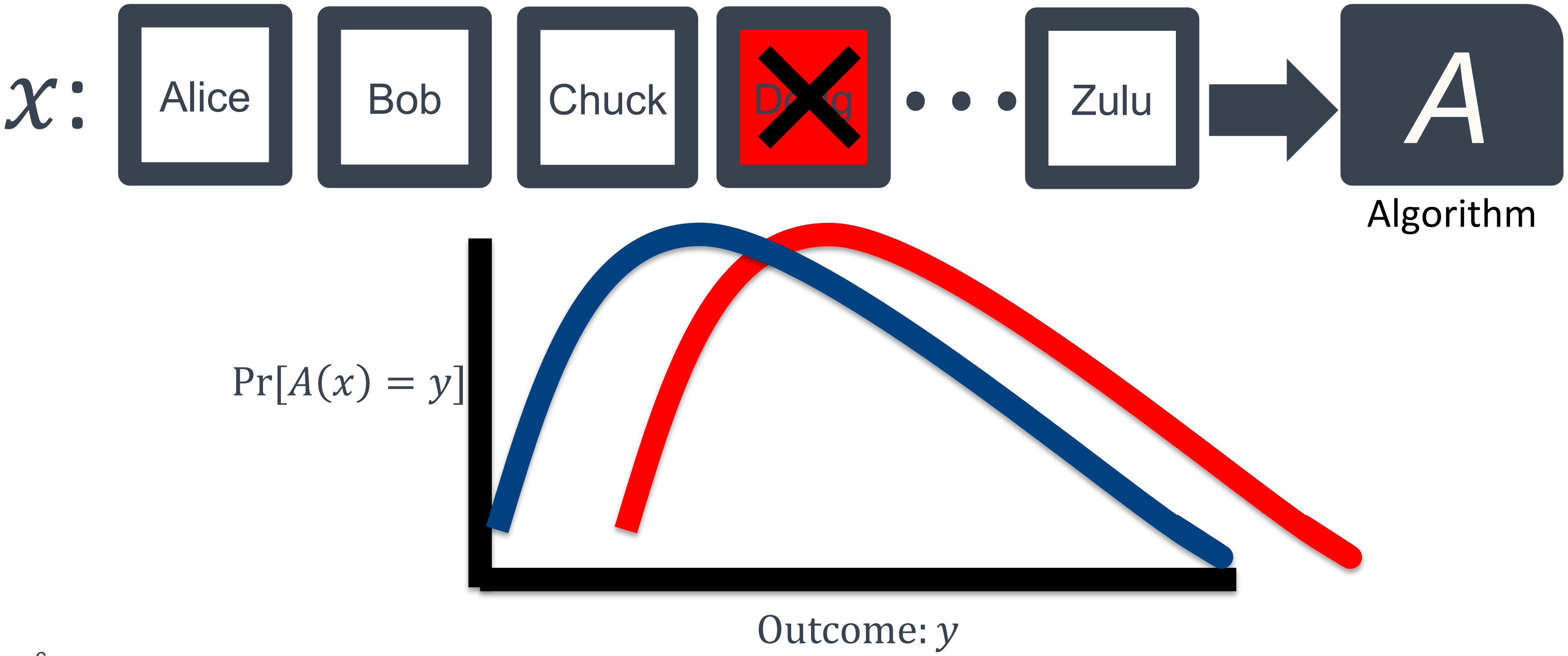
Reasons for Data Privacy

- We want to be ”Members first”
- “Anonymized data isn’t” – Cynthia Dwork
  - 87% of U.S. is uniquely identified by (DOB, Gender, Zip)
- Potential attacks:
  - Reconstruction attacks
  - Differencing attacks
  - Membership inference attacks

Differential Privacy [Dwork, McSherry, Nissim, Smith ‘06]



Differential Privacy [Dwork, McSherry, Nissim, Smith ‘06]



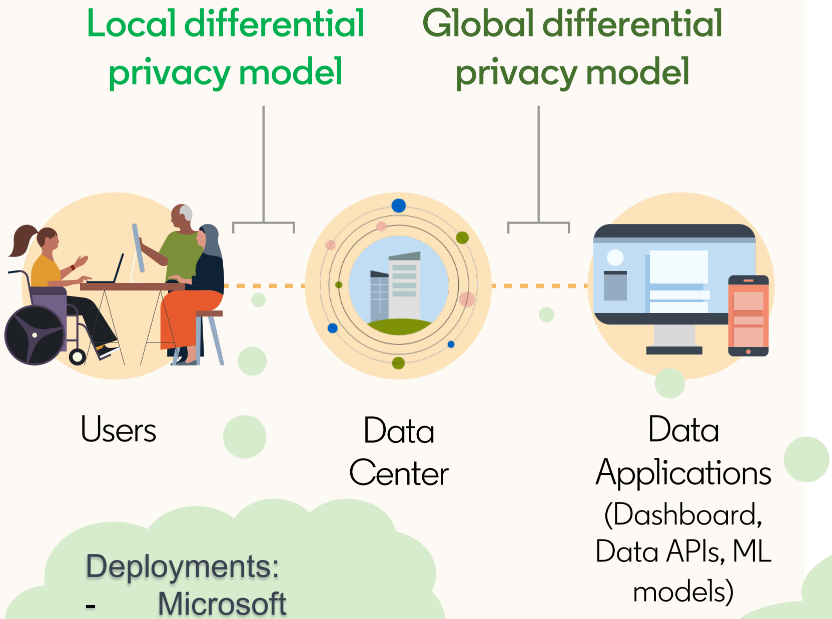
Differential Privacy [Dwork, McSherry, Nissim, Smith '06]

A randomized algorithm $A{:}\,\mathcal{D} \rightarrow \mathcal{Y}$ is (ε, δ) −DP if for any neighboring data sets $x, x' \in \mathcal{D}$ and any outcome $S \subseteq \mathcal{Y}$ we have:

$$P(A(x) \in S) \leq e^{\varepsilon} P(A(x') \in S) + \delta$$

Privacy loss

Models and Deployments of Differential Privacy



Local differential privacy model

Global differential privacy model

Users

Data Center

Data Applications (Dashboard, Data APIs, ML models)

Deployments:
- Microsoft
- Google
- Apple

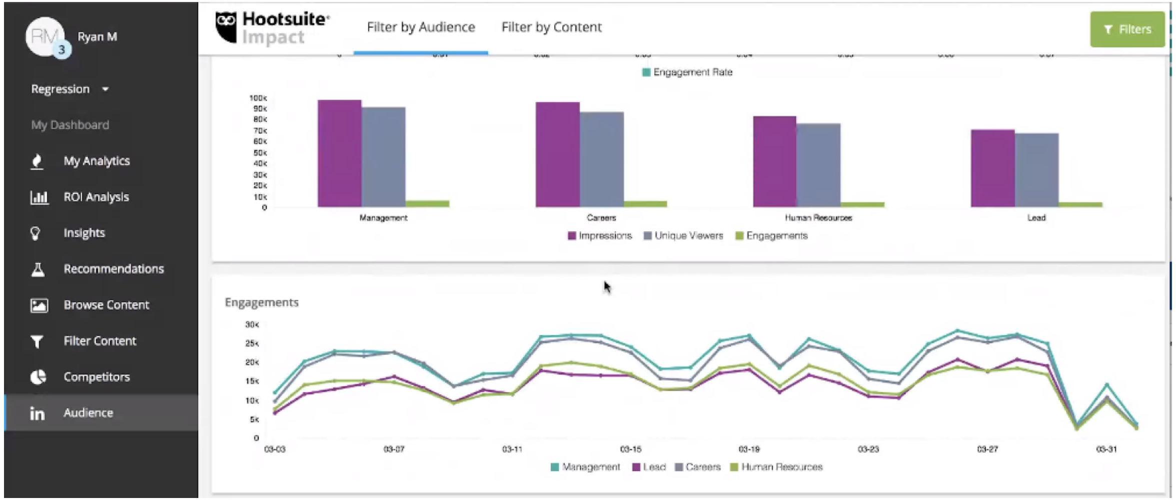
- Traditional data protection techniques are not sufficient to defend data privacy
- Differential Privacy ensures data learnings are the same with/without a single member’s data

Deployments:
- 2020 Census
- Microsoft Open Data DP Project with Harvard
- Google’s Open Source Library

Audience Engagement API

- API Product to provide insights on LinkedIn engagement content and audience data
- Provides information about member data to external marketing partners
- Built on top of ***Pinot*** for fast, real-time data analytics

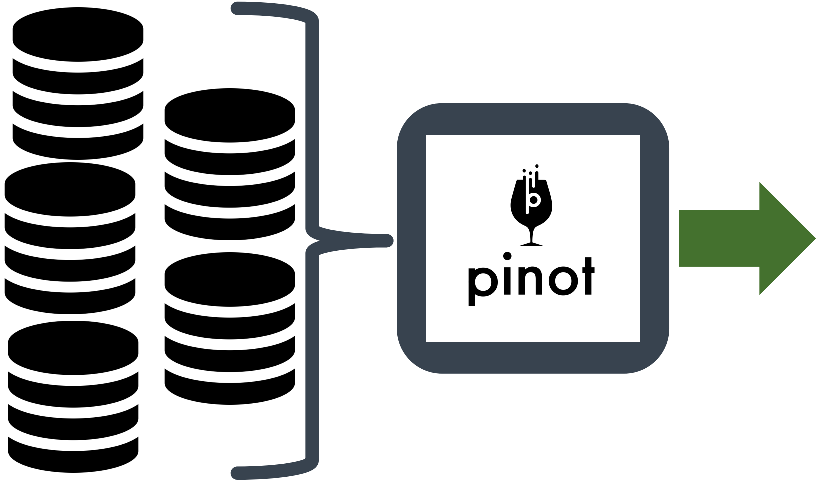
pinot

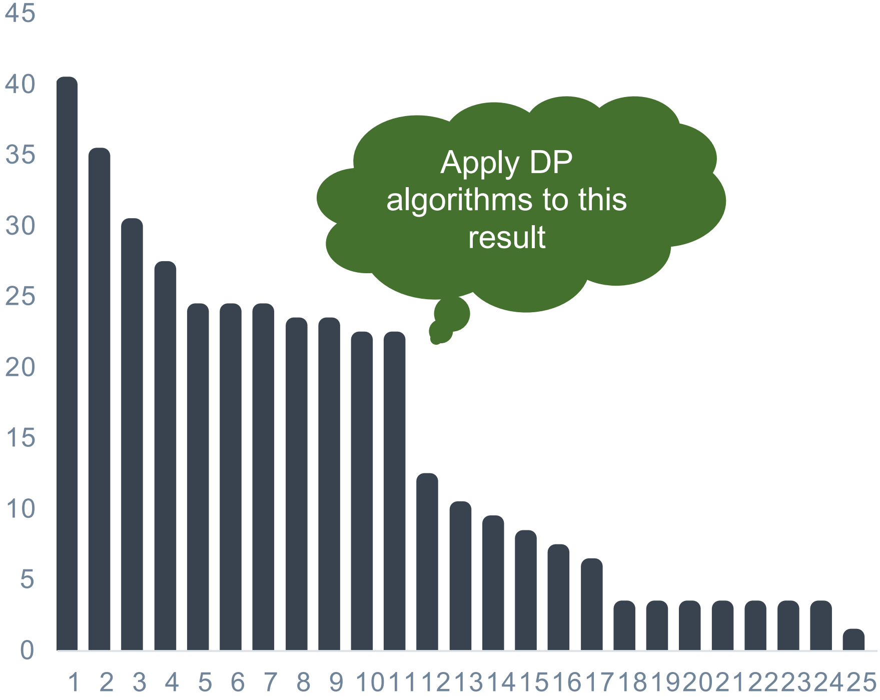


Understanding the Task

- Advertiser can interact adaptively with the API
- Differencing attacks are a concern
- Want to provide both real-time analytics and privacy
- Queries are general top-*k* queries
- Questions that need to be addressed:
  - How much can a single user affect the outcome of these queries?
  - How many queries can the advertiser ask?

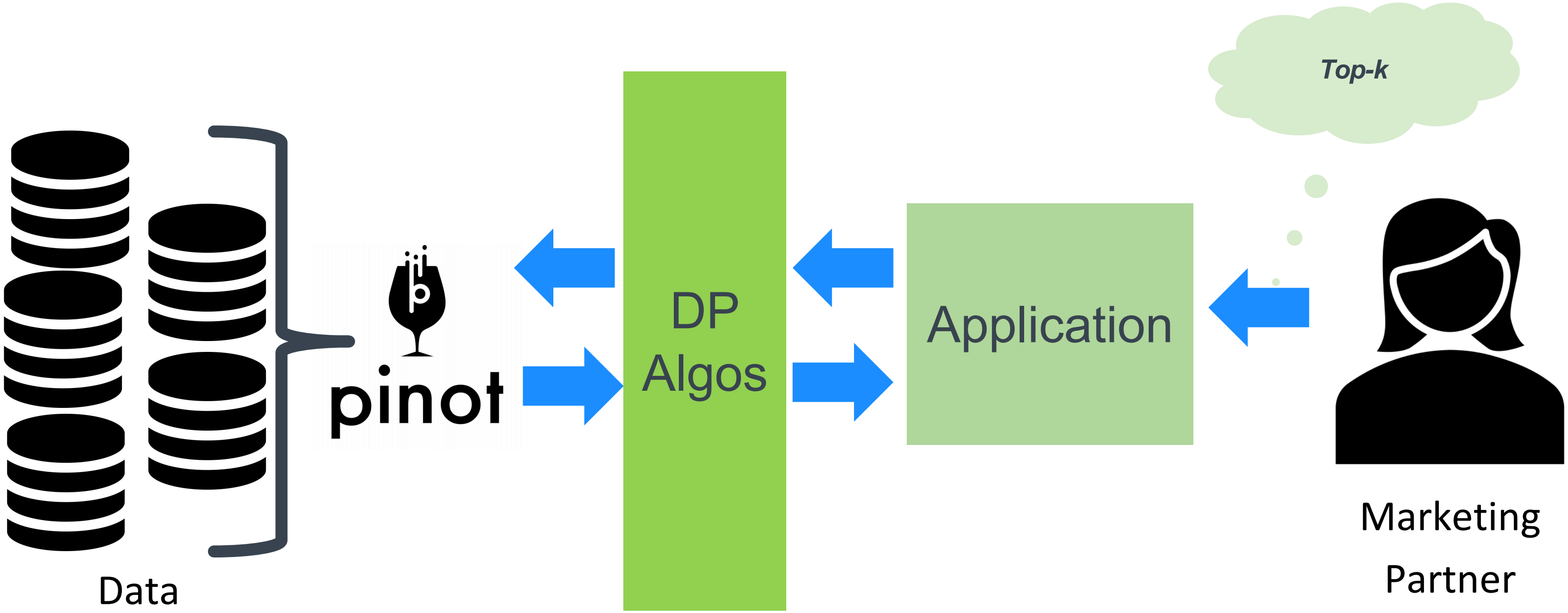
Existing Systems for Data Analytics





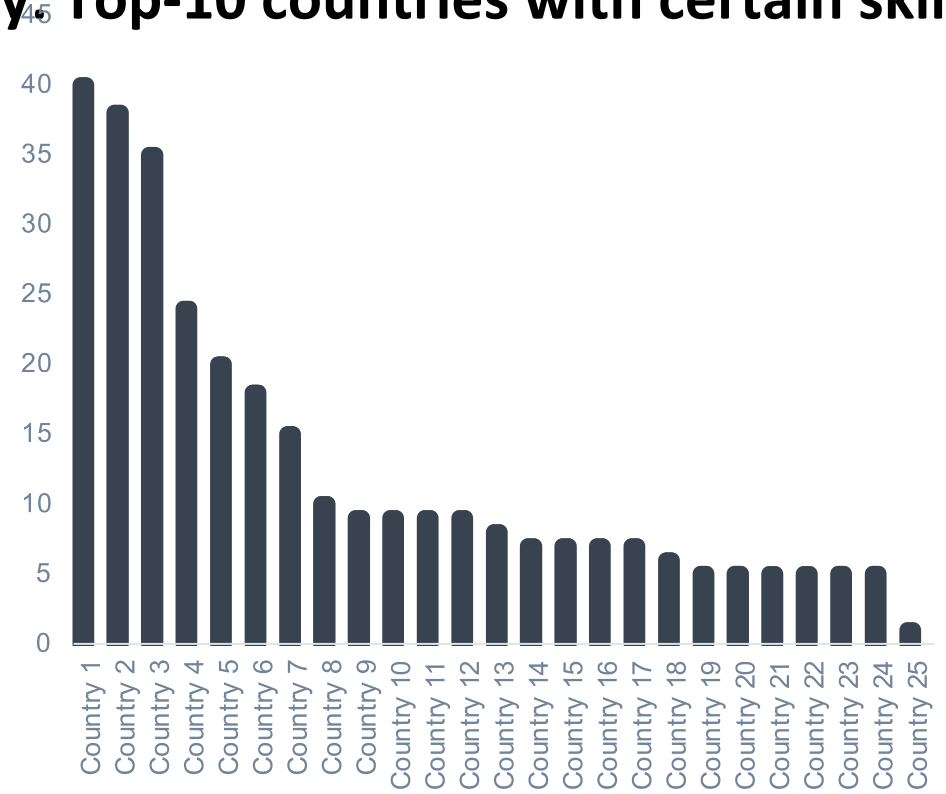
Apply DP algorithms to this result

Overall Privacy System



Sensitivity of the Query

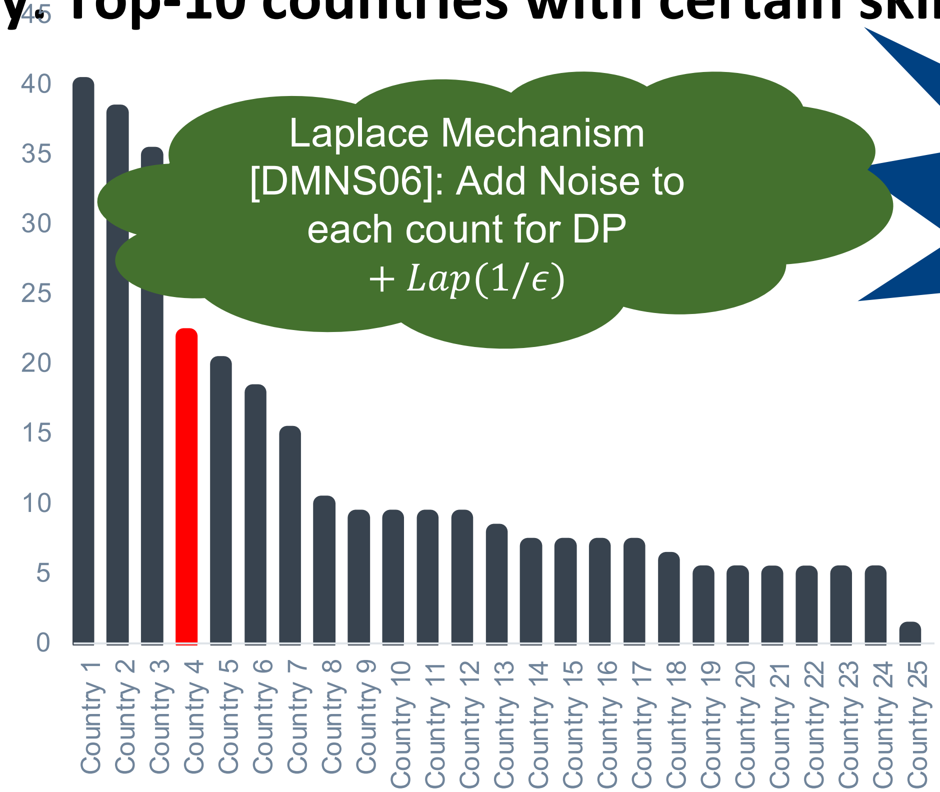
Query: Top-10 countries with certain skill set?



Aggregate Data

Sensitivity of the Query

Query: Top-10 countries with certain skill set?



Laplace Mechanism [DMNS06]: Add Noise to each count for DP

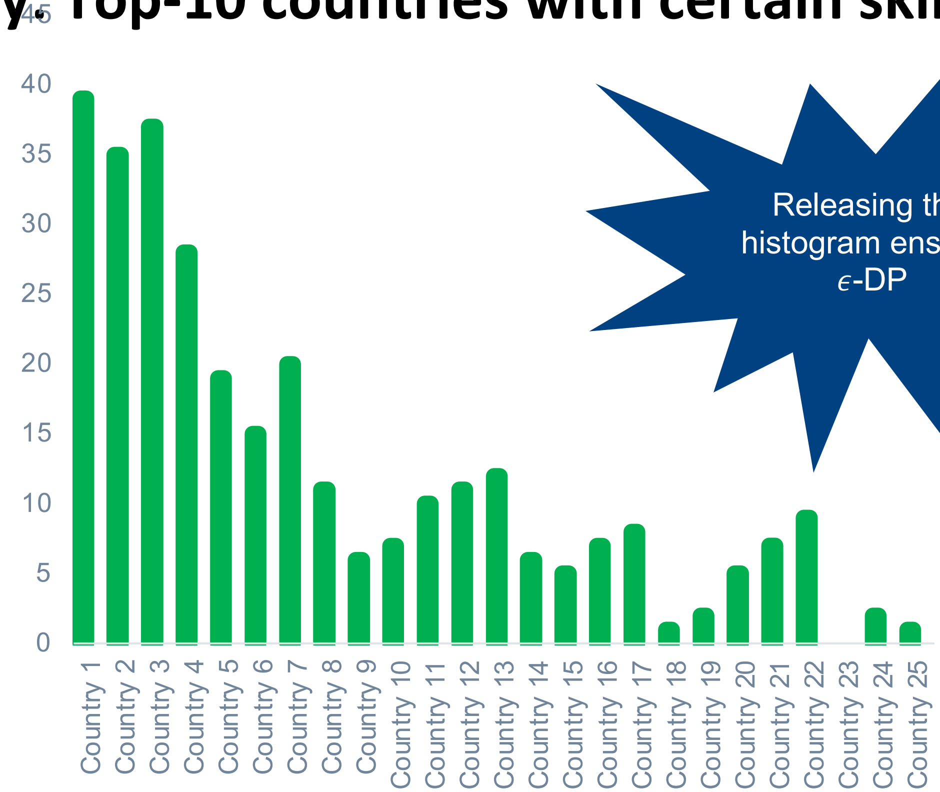
$+\,Lap(1/\epsilon)$

User can impact only one count

Aggregate Data

Sensitivity of the Query

Query: Top-10 countries with certain skill set?

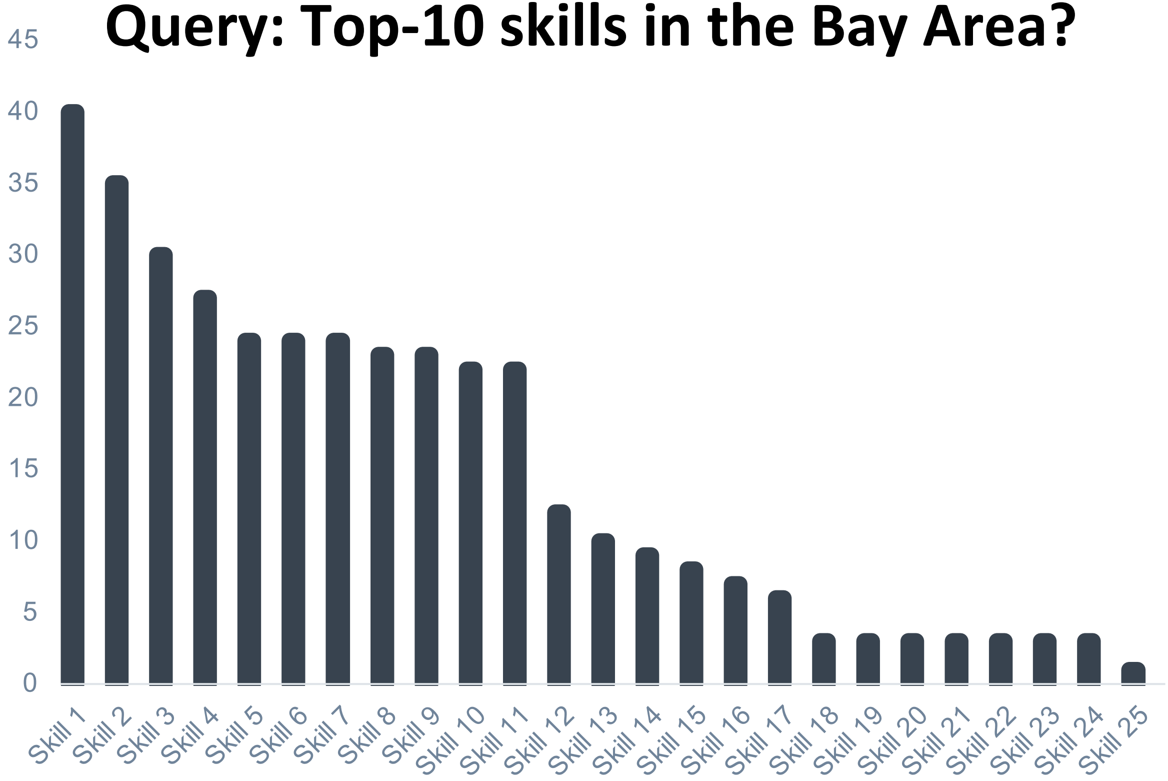


Releasing this histogram ensures ϵ-DP

Aggregate Data

Sensitivity of the Query

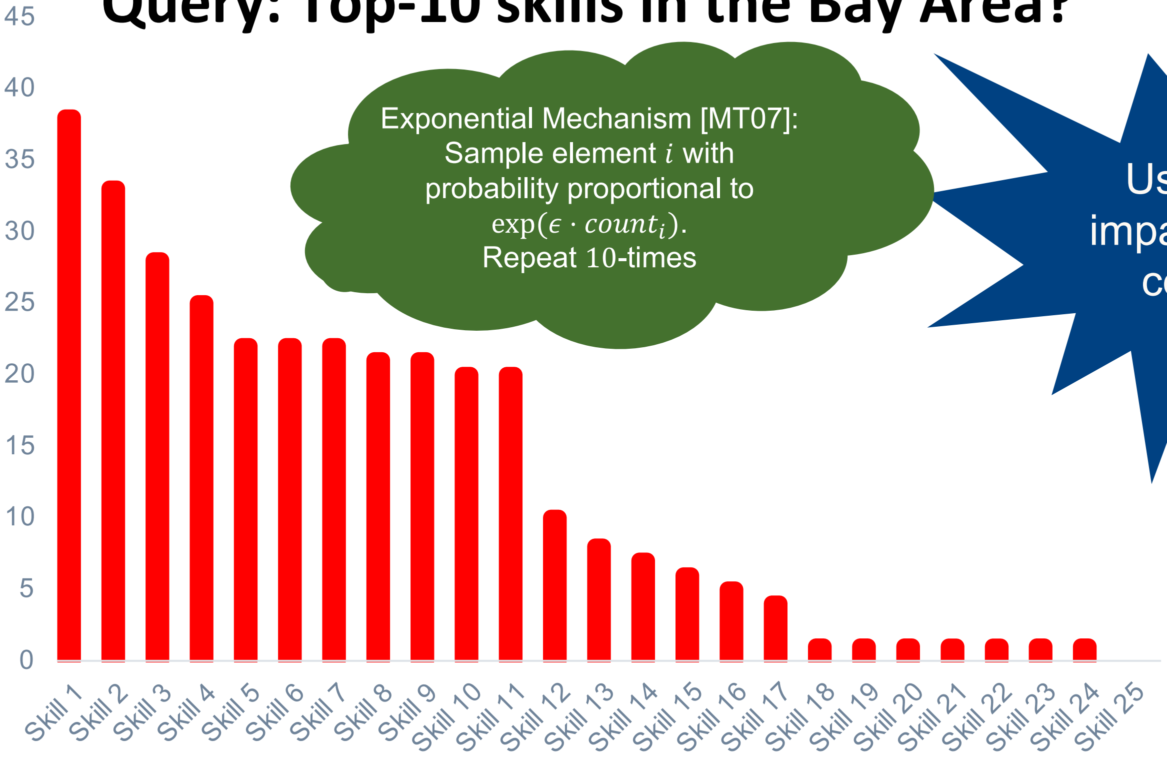
Query: Top-10 skills in the Bay Area?



Aggregate Data

Sensitivity of the Query

Query: Top-10 skills in the Bay Area?



Exponential Mechanism [MT07]:
Sample element i with
probability proportional to
$\exp(\epsilon \cdot count_i)$.
Repeat 10-times

User can impact many counts!

Aggregate Data

Sensitivity of the Query

Query: Top-10 skills in the Bay Area?

Exponential Mechanism [MT07]:
Sample element i with
probability proportional to
$\exp(\epsilon \cdot count_i)$.
Repeat 10-times

Releasing only elements in top-k
(not their counts) ensures
$k\epsilon$-DP

Known Algorithms for User Level DP

Δ-Restricted Sensitivity	Unrestricted Sensitivity
Algorithm: Laplace Mechanism [DMNS‘06]	**Algorithm: Exponential Mechanism** [McSherry, Talwar ‘07]

Unknown Domain Setting

- Previous algorithms require knowing the full data domain
- They require adding noise to counts even when the true count is zero
- Typically, the domain is unknown or very large (e.g. all possible articles)

Algorithms for User Level Privacy

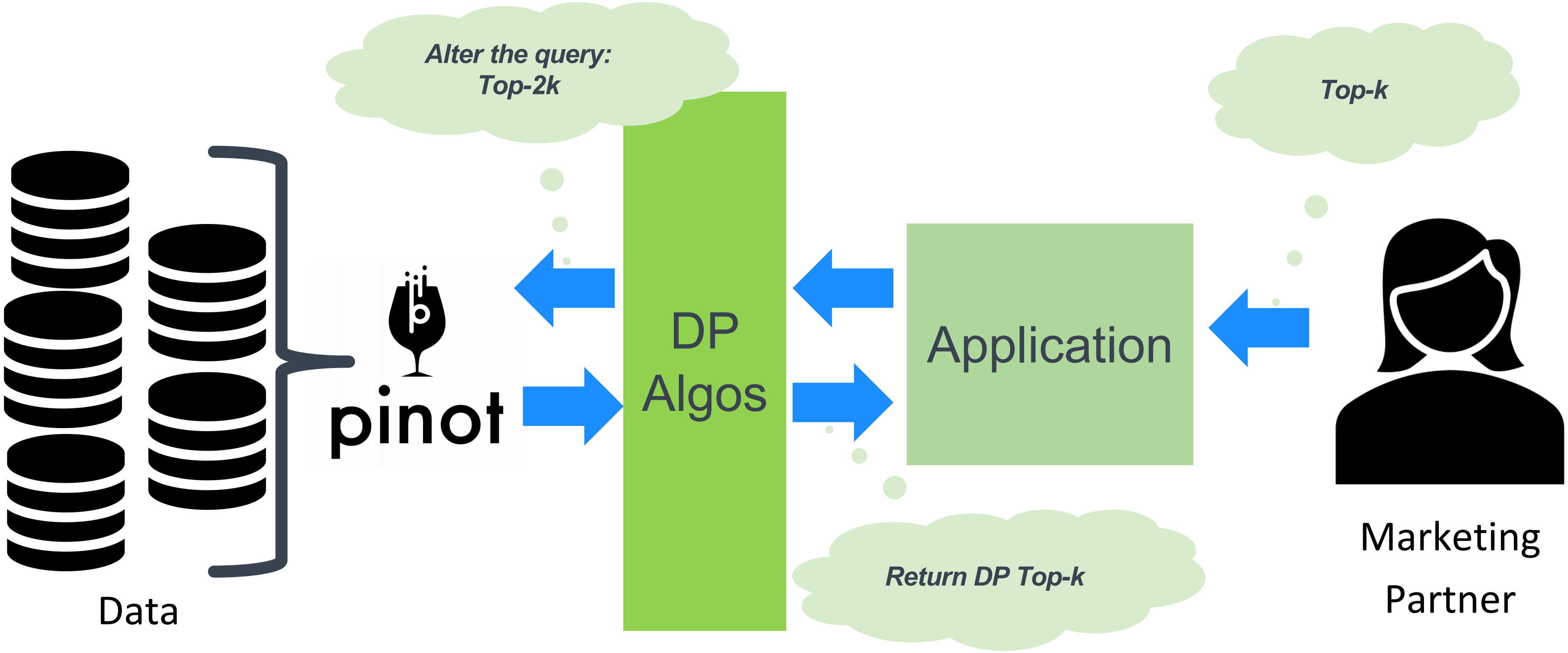
User Level DP Algorithms	**Restricted Sensitivity**	**Unrestricted Sensitivity**
Known Domain	Laplace Mechanism [DMNS’06]	Exponential Mechanism [MT’07]

Algorithms for User Level Privacy

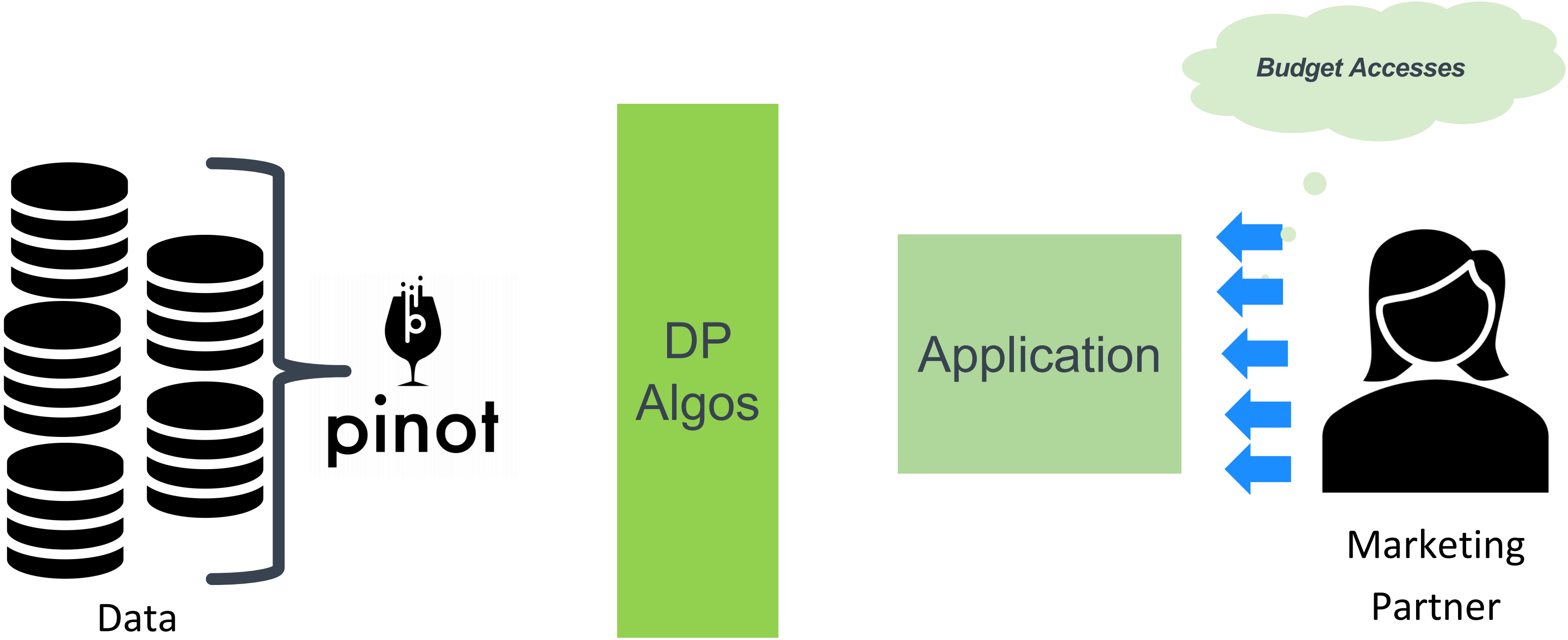
User Level DP Algorithms	**Restricted Sensitivity**	**Unrestricted Sensitivity**
Known Domain	Laplace Mechanism [DMNS’06]	Exponential Mechanism [MT’07]
Unknown Domain	UnkLap Mechanism [Durfee, **R**’19]	UnkExp Mechanism [Durfee, **R**’19]

NeurIPS’19 Spotlight:
https://arxiv.org/abs/1905.04273

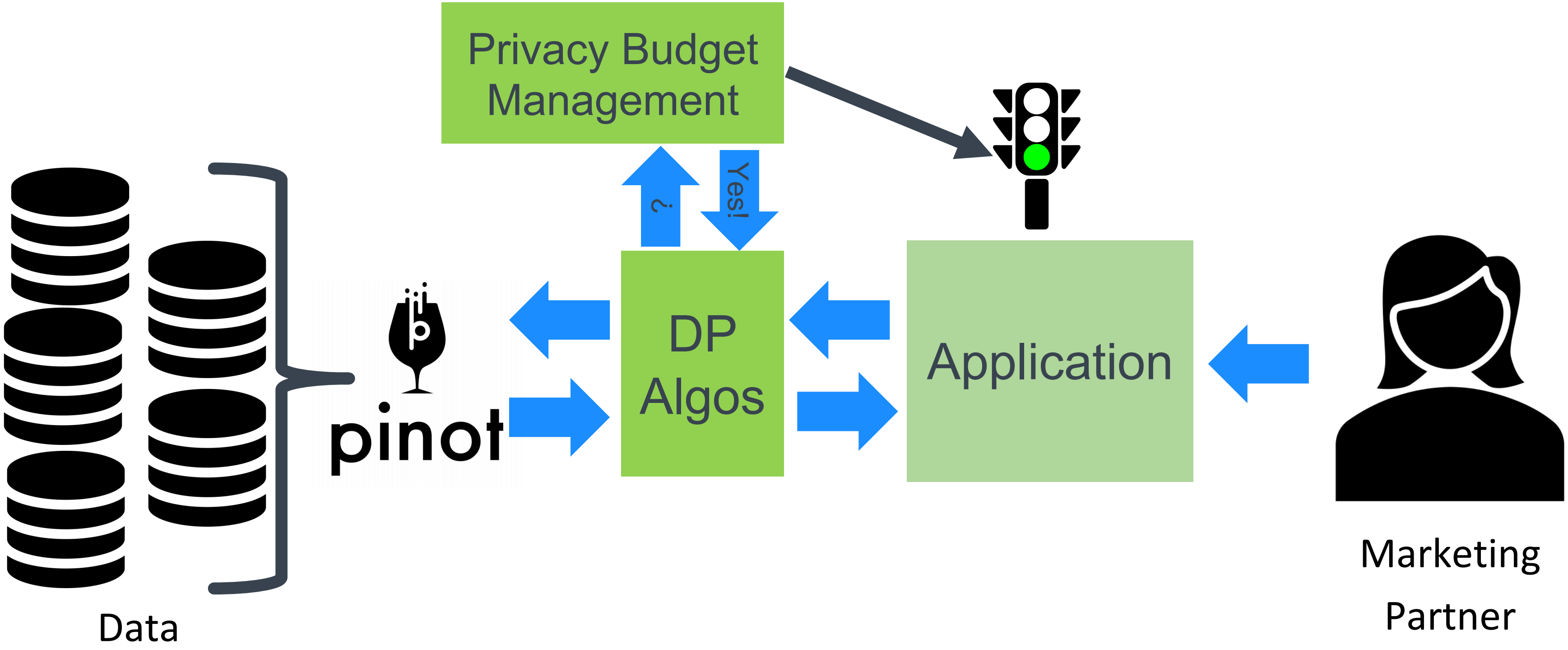
Overall Privacy System



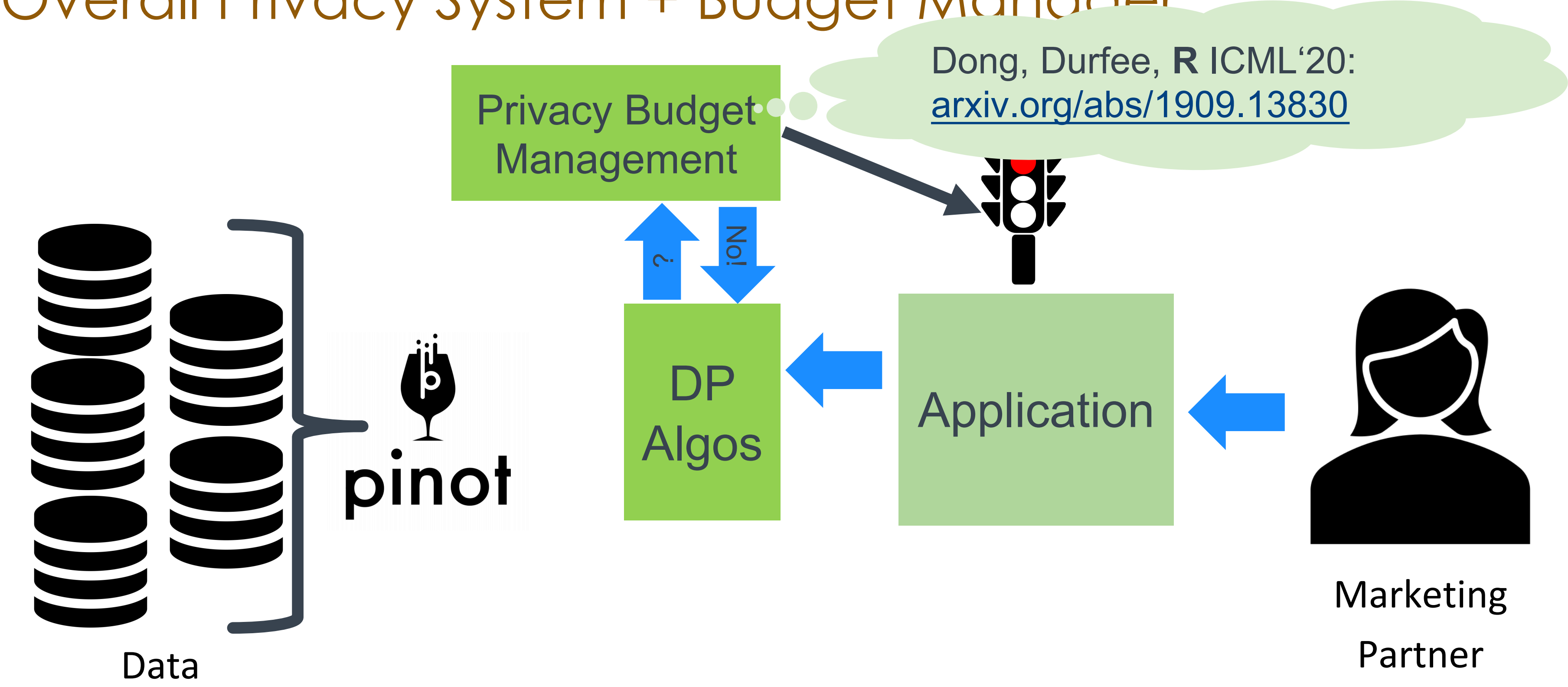
Overall Privacy System



Overall Privacy System + Budget Manager



Overall Privacy System + Budget Manager



Dong, Durfee, **R** ICML‘20: arxiv.org/abs/1909.13830

Thank you!

Linked in