

Cohere: Managing Differential Privacy in Large - Scale Systems



Nicolas Küchler
ETH zürich



Emanuel Opel
ETH zürich



Hidde Lycklama
ETH zürich



Alexander Viand
intel



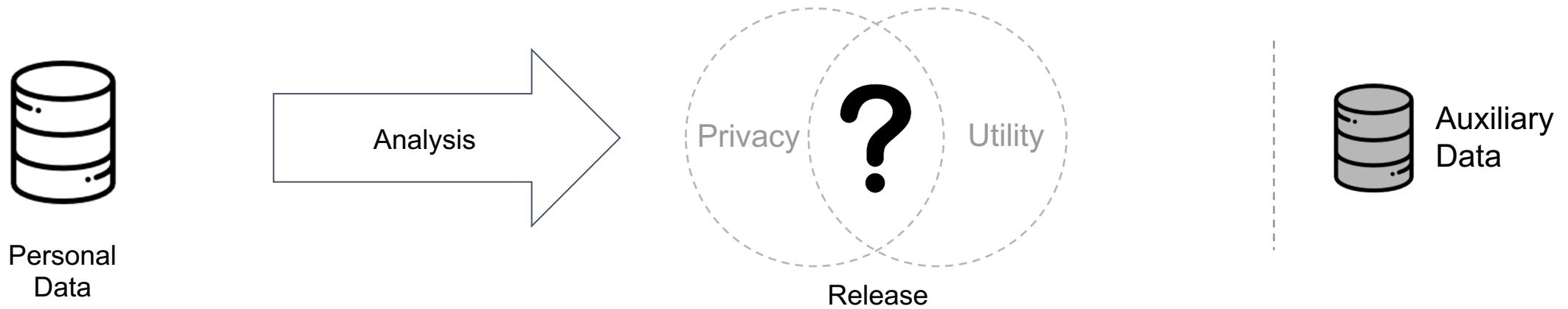
Anwar Hithnawi
ETH zürich

Presented At
IEEE Security & Privacy 2024

Contact Us
www.pps-lab.com

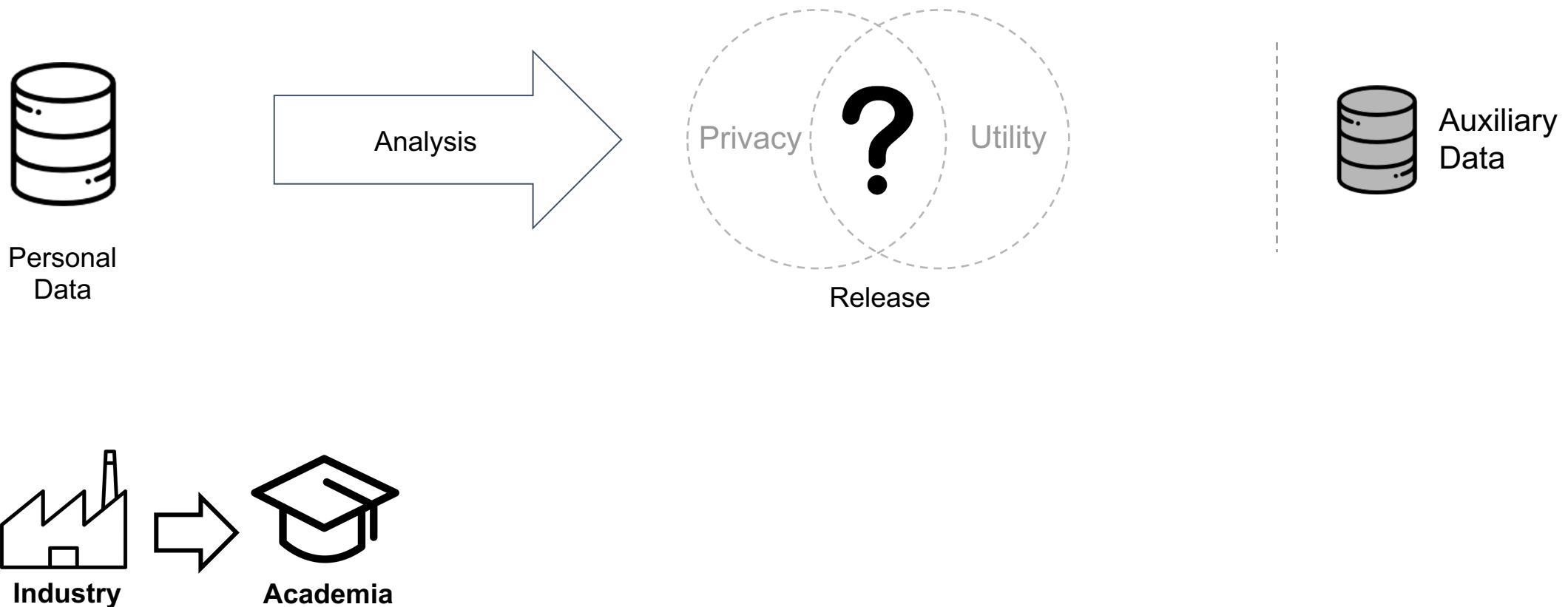
Statistical Release

How can we release useful information without compromising privacy?



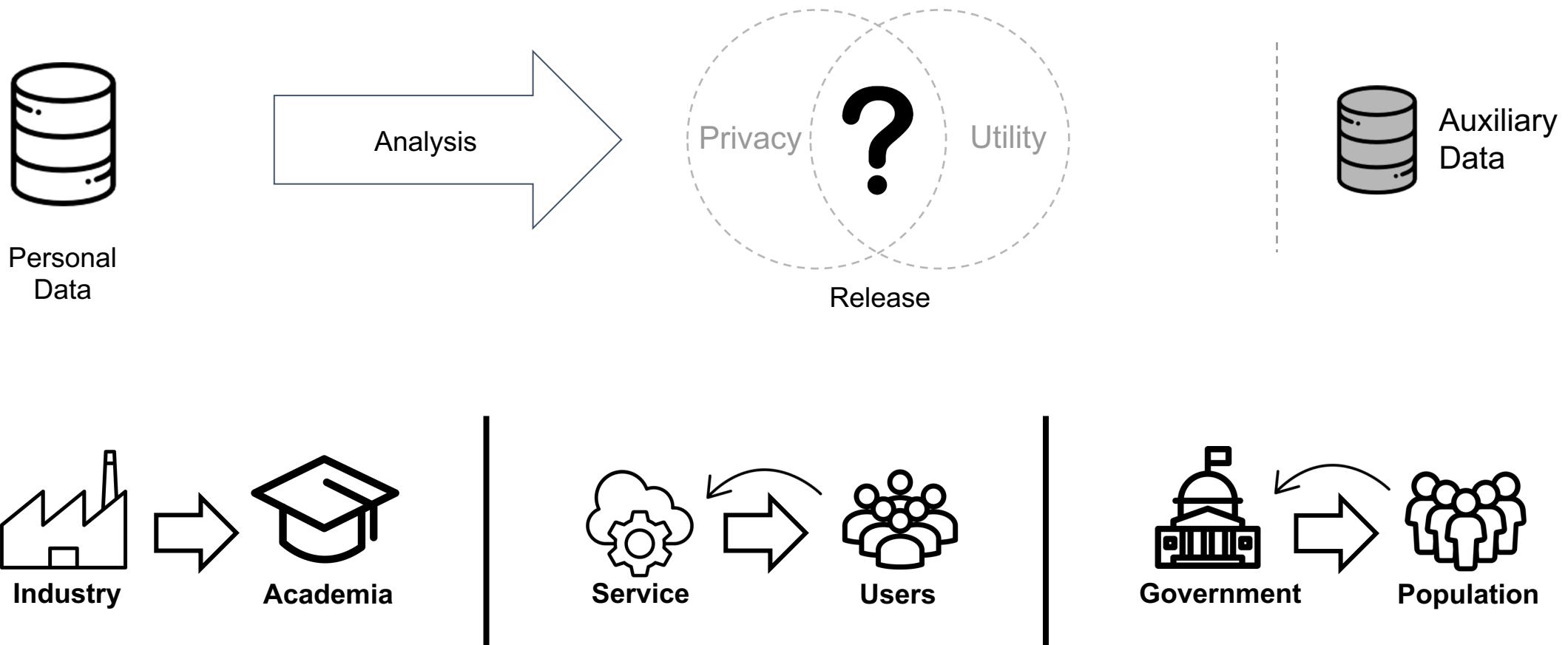
Statistical Release

How can we release useful information without compromising privacy?



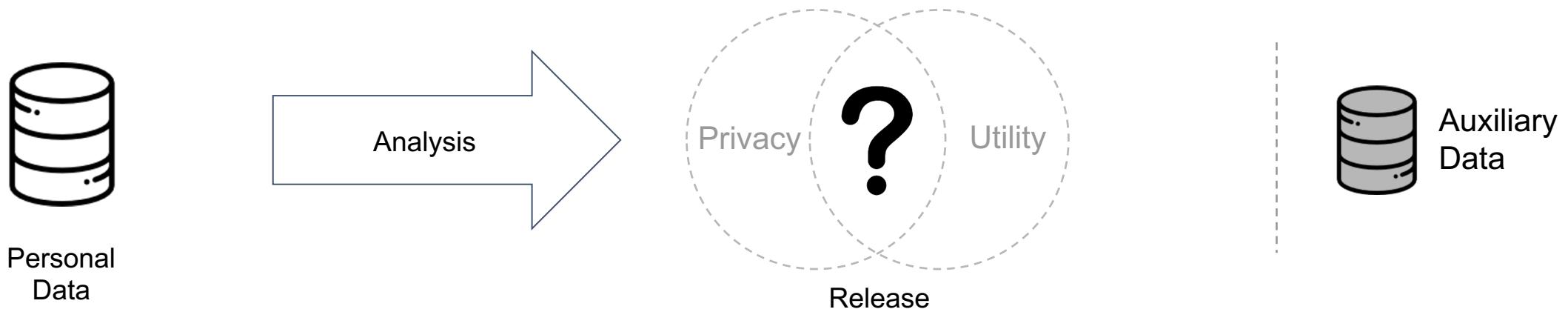
Statistical Release

How can we release useful information without compromising privacy?



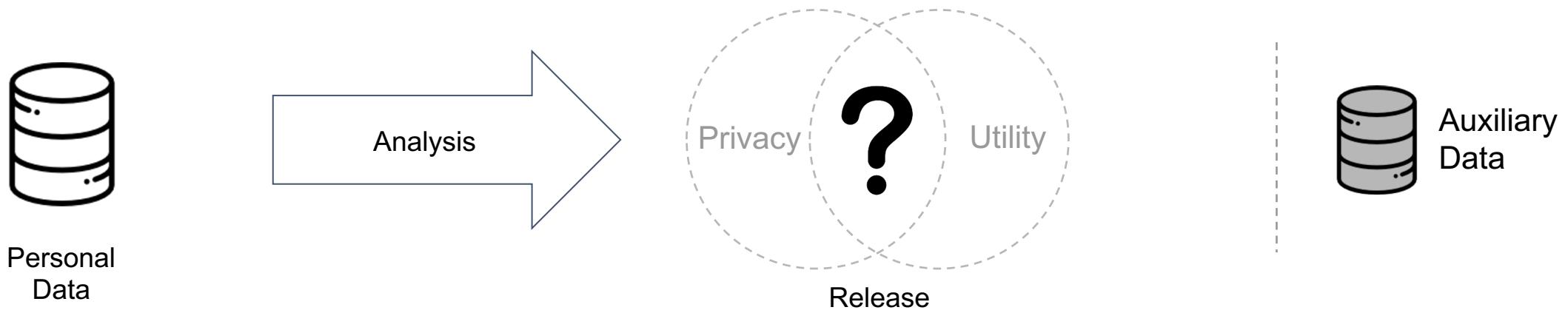
Statistical Release

How can we release useful information without compromising privacy?



Statistical Release

How can we release useful information without compromising privacy?

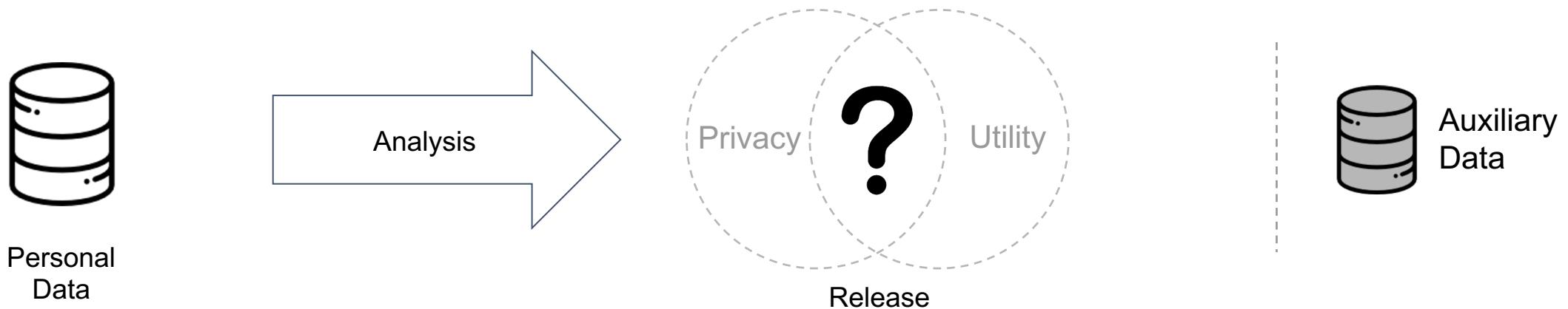


- **Anonymization**
Redact Personally Identifiable Information

Name	Region	...	Value
[REDACTED]	CH		100
[REDACTED]	DE		237

Statistical Release

How can we release useful information without compromising privacy?



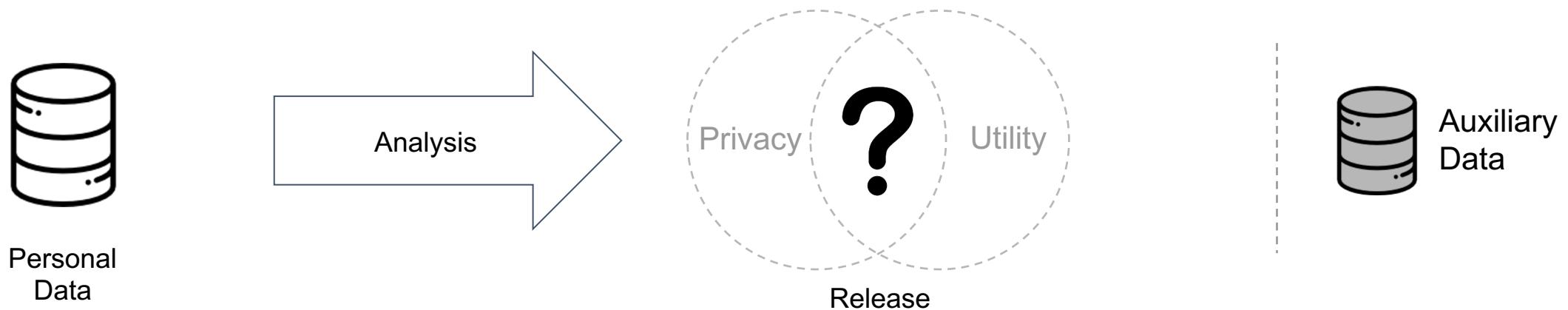
- **Anonymization**
Redact Personally Identifiable Information
- **Release Aggregates**

Name	Region	...	Value
	CH		100
	DE		237



Statistical Release

How can we release useful information without compromising privacy?



- **Anonymization**
Redact Personally Identifiable Information
- **Release Aggregates**

Name	Region	...	Value
CH			100
DE			237



Privacy Attacks

← Re-Identification (NYC TAXI)

{ Database Reconstruction (United States Census 2010)
Membership Inference (LLM)

Differential Privacy

Mathematical definition of privacy in the context of statistical releases

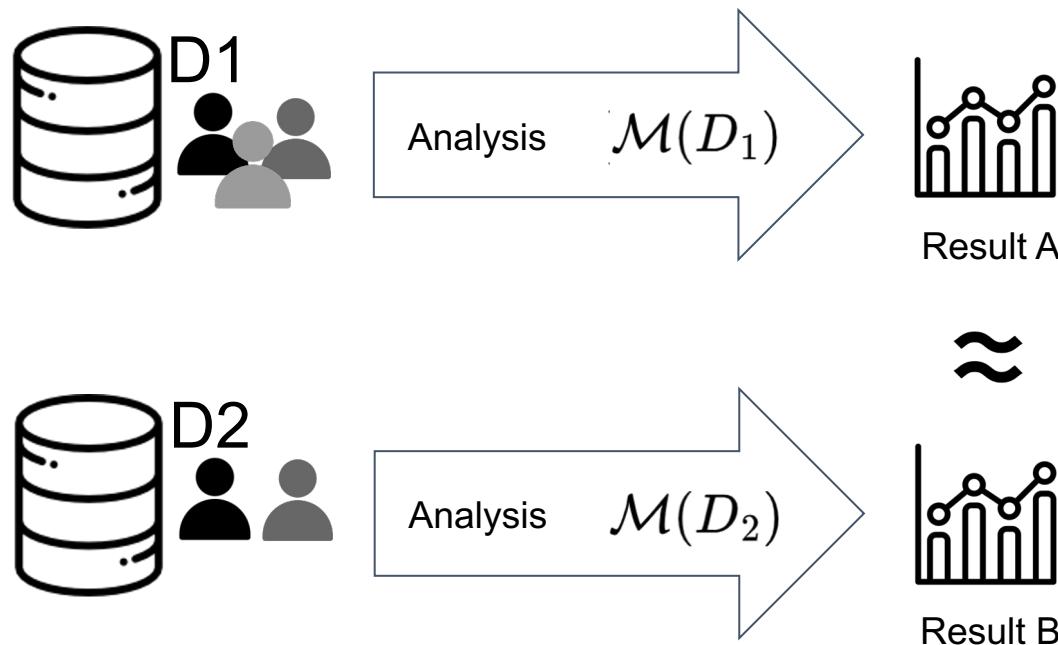
Differential Privacy

Mathematical definition of privacy in the context of statistical releases



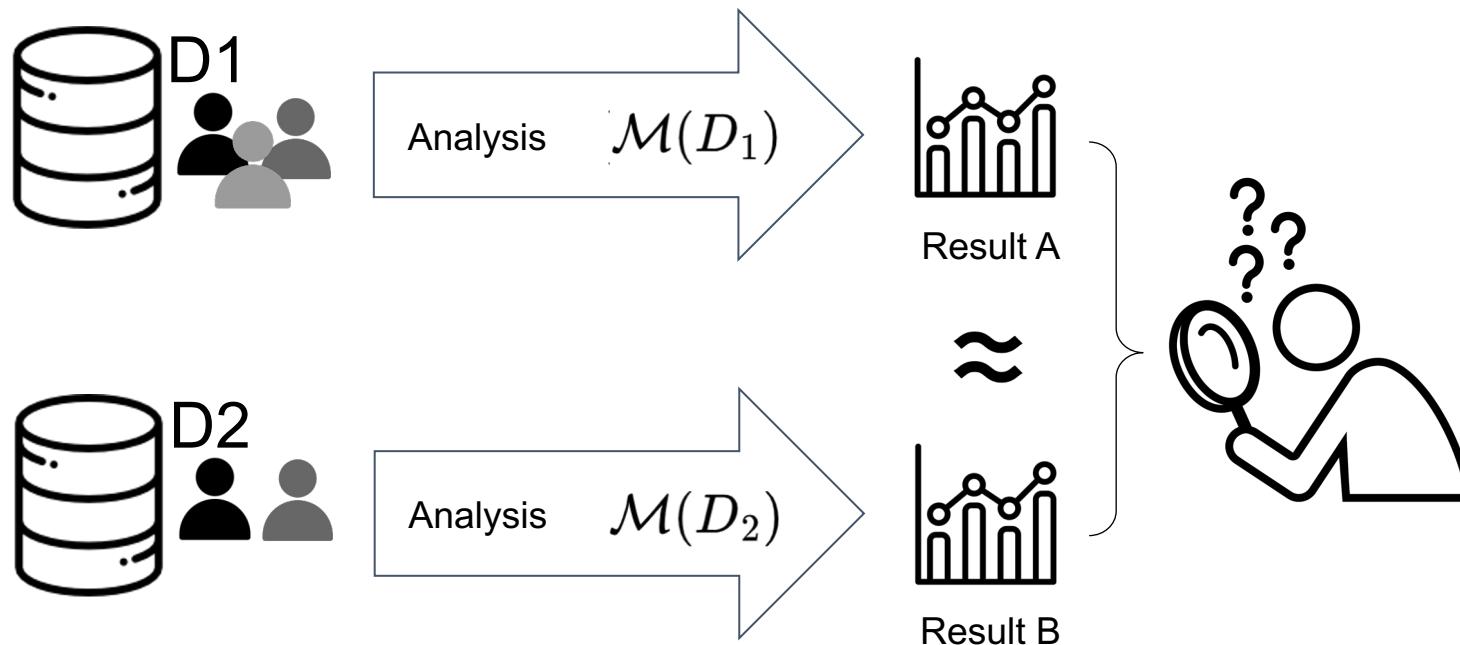
Differential Privacy

Mathematical definition of privacy in the context of statistical releases



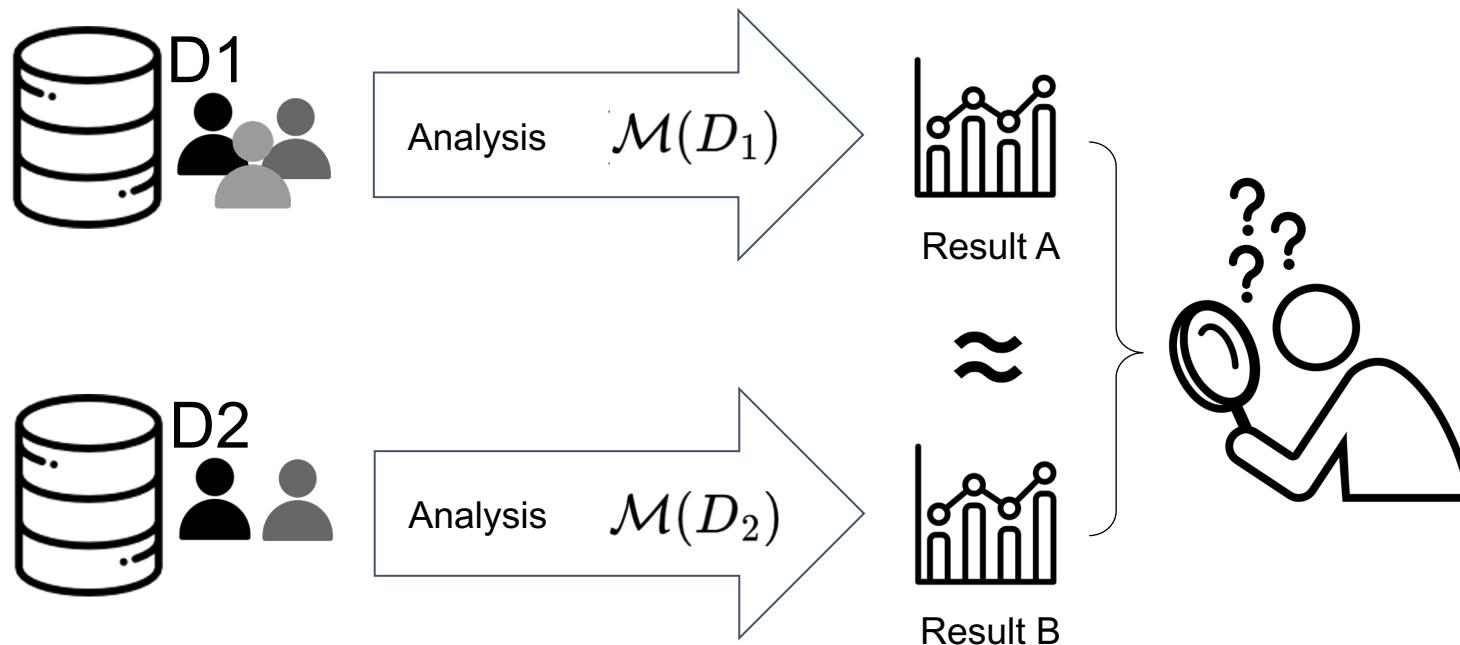
Differential Privacy

Mathematical definition of privacy in the context of statistical releases



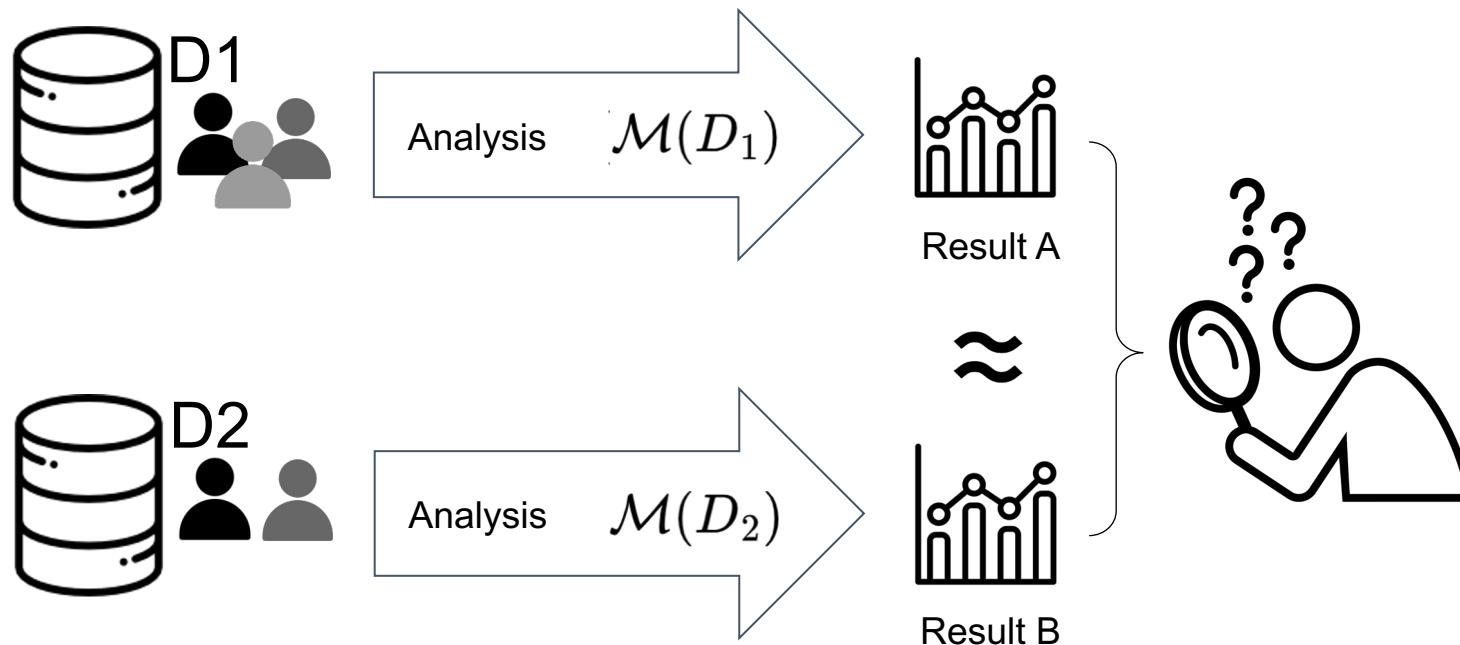
Differential Privacy

Mathematical definition of privacy in the context of statistical releases



Differential Privacy

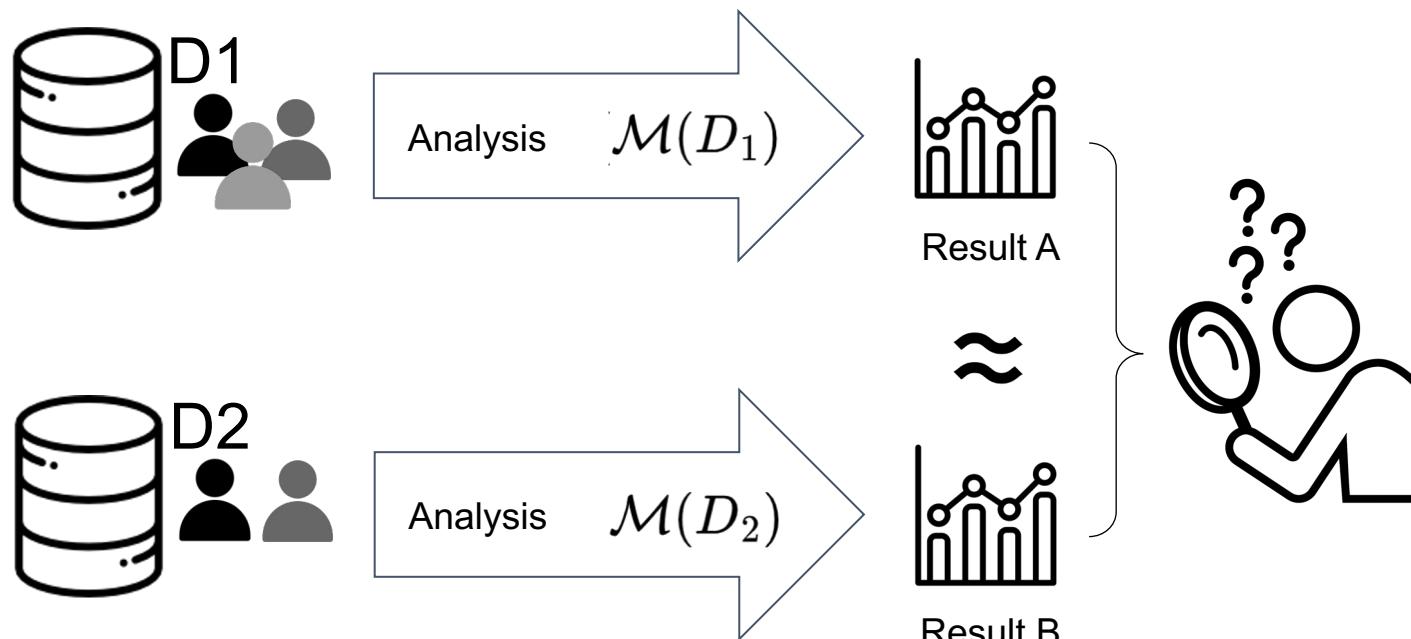
Mathematical definition of privacy in the context of statistical releases



$$\Pr[\mathcal{M}(D_1) \in \mathcal{S}] \leq e^\epsilon \cdot \Pr[\mathcal{M}(D_2) \in \mathcal{S}] + \delta$$

Differential Privacy

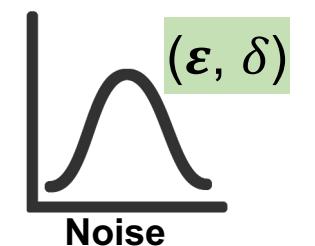
Mathematical definition of privacy in the context of statistical releases



$$\Pr[\mathcal{M}(D_1) \in \mathcal{S}] \leq e^\epsilon \cdot \Pr[\mathcal{M}(D_2) \in \mathcal{S}] + \delta$$

Intuition

Data +



Differential Privacy

Mathematical definition of privacy in the context of statistical releases

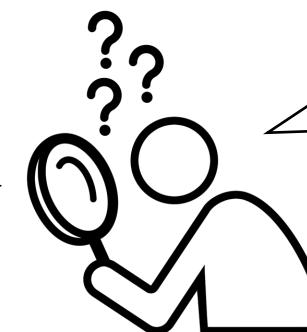


Result A



Result B

$$\Pr [\mathcal{M}(D_1) \in \mathcal{S}] \leq e^\epsilon \cdot \Pr [\mathcal{M}(D_2) \in \mathcal{S}] + \delta$$



Has 's data been included?

Privacy Cost

low
 (ϵ, δ)

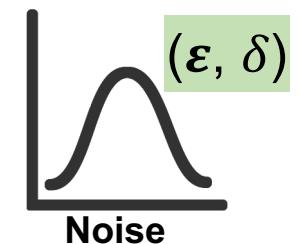
Privacy Leakage



Users

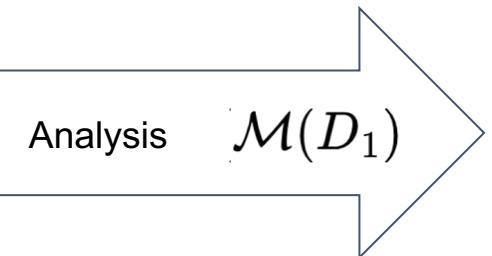
Intuition

Data +



Differential Privacy

Mathematical definition of privacy in the context of statistical releases



$$\Pr[\mathcal{M}(D_1) \in \mathcal{S}] \leq e^\epsilon \cdot \Pr[\mathcal{M}(D_2) \in \mathcal{S}] + \delta$$



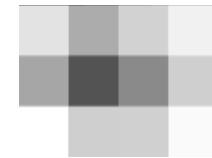
Has 's data been included?

Privacy Cost



(ϵ, δ)

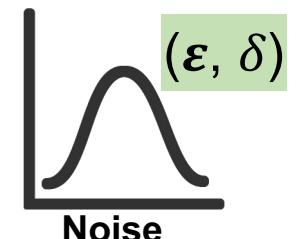
Privacy Leakage



Users

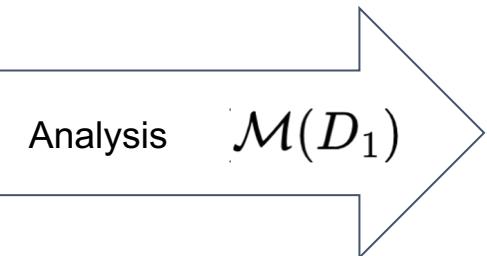
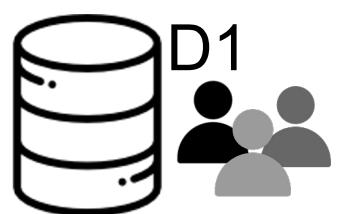
Intuition

Data +

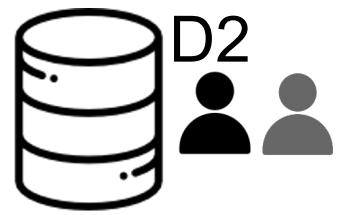


Differential Privacy

Mathematical definition of privacy in the context of statistical releases



Result A



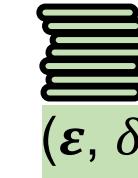
Result B

$$\Pr [\mathcal{M}(D_1) \in \mathcal{S}] \leq e^\epsilon \cdot \Pr [\mathcal{M}(D_2) \in \mathcal{S}] + \delta$$

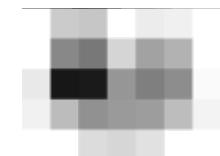


Has 's data been included?

Privacy Cost



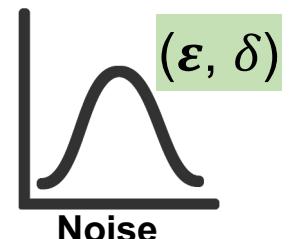
Privacy Leakage



Users

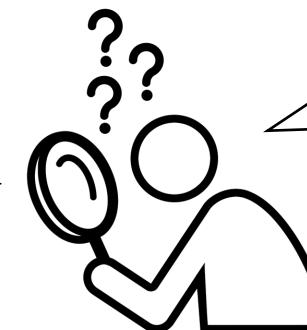
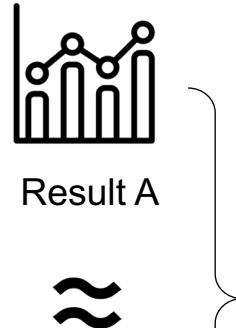
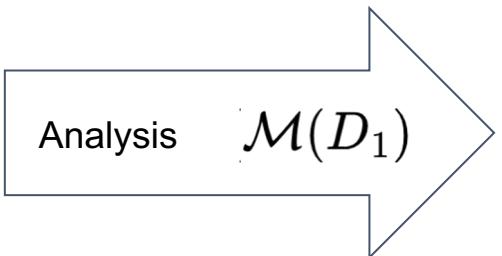
Intuition

Data +



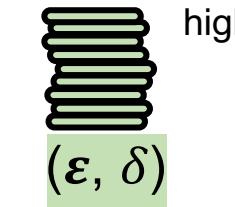
Differential Privacy

Mathematical definition of privacy in the context of statistical releases

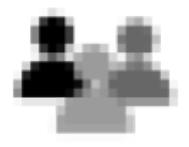


Has 's data been included?

Privacy Cost



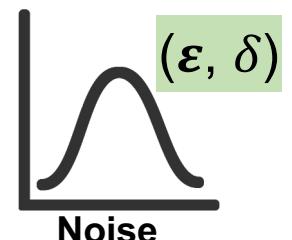
Privacy Leakage



Users

Intuition

Data +



$$\Pr [\mathcal{M}(D_1) \in \mathcal{S}] \leq e^\epsilon \cdot \Pr [\mathcal{M}(D_2) \in \mathcal{S}] + \delta$$

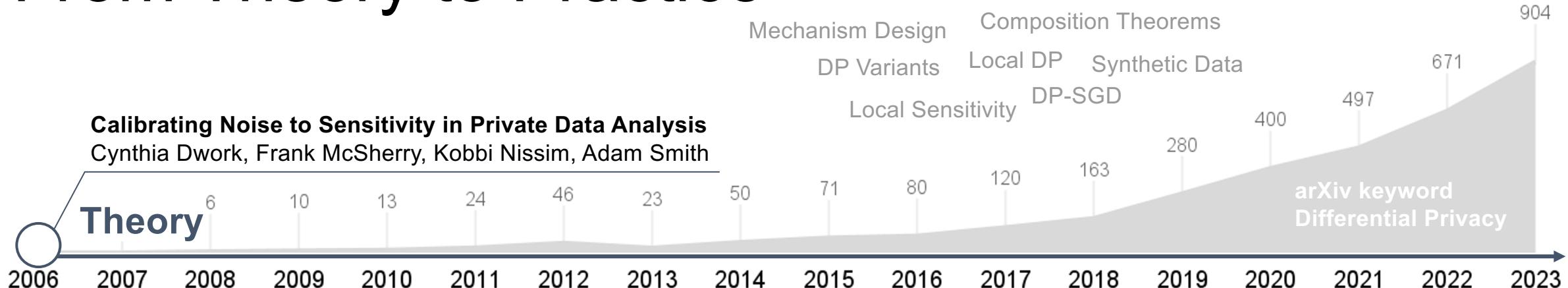
From Theory to Practice

Calibrating Noise to Sensitivity in Private Data Analysis

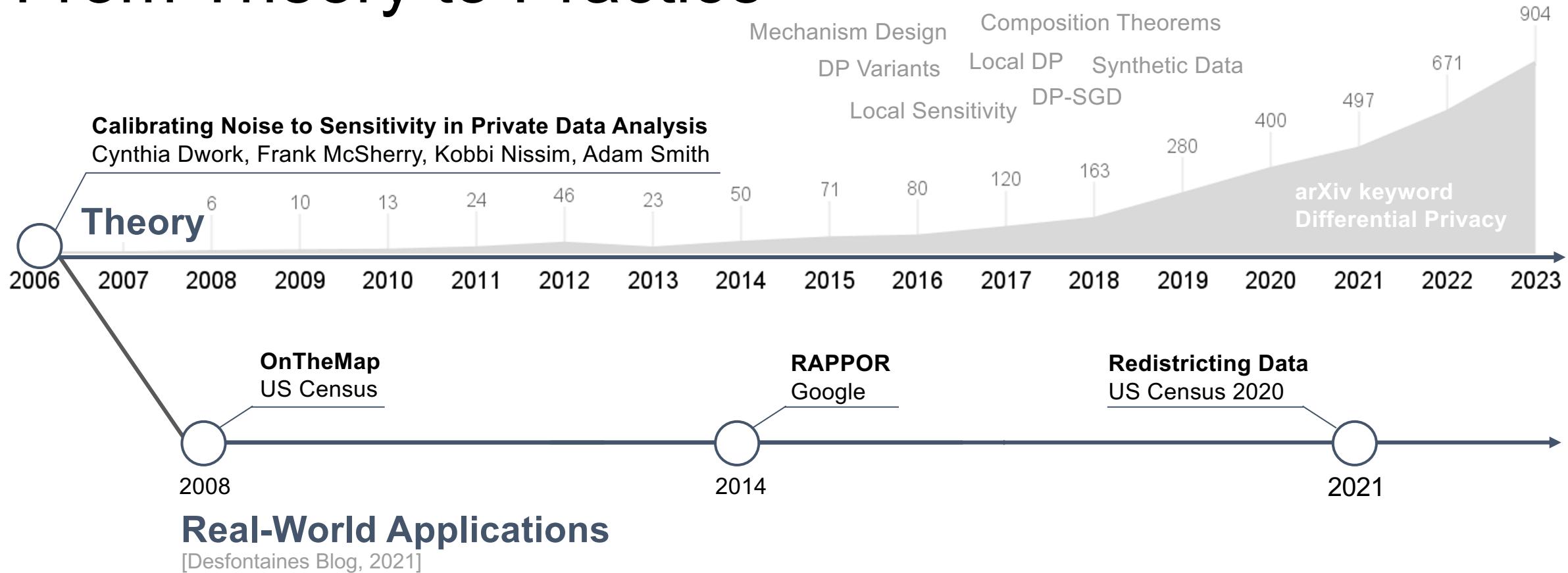
Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith



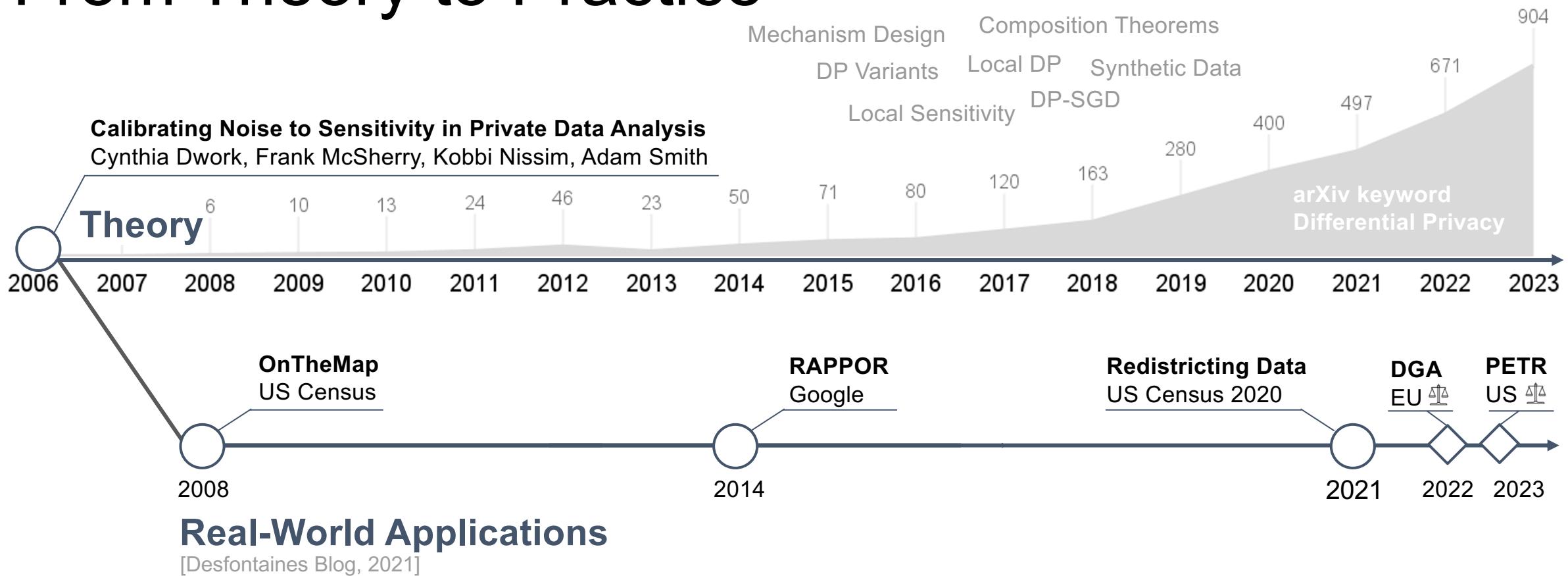
From Theory to Practice



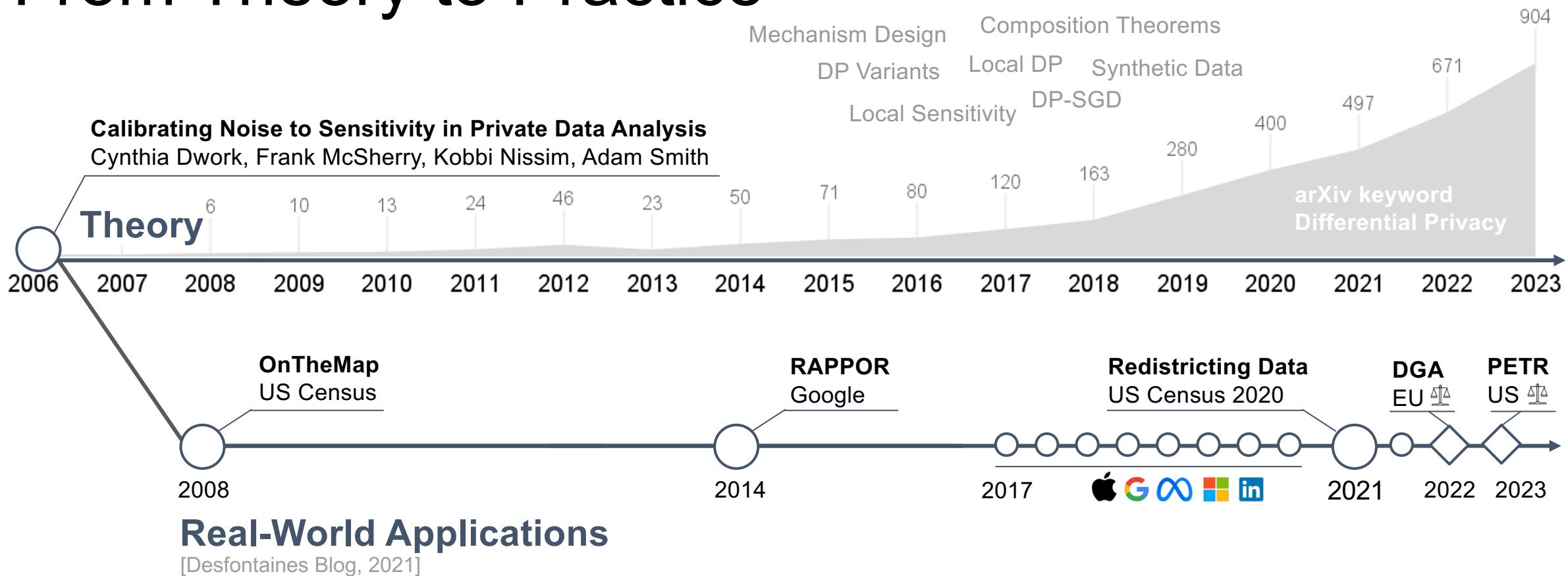
From Theory to Practice



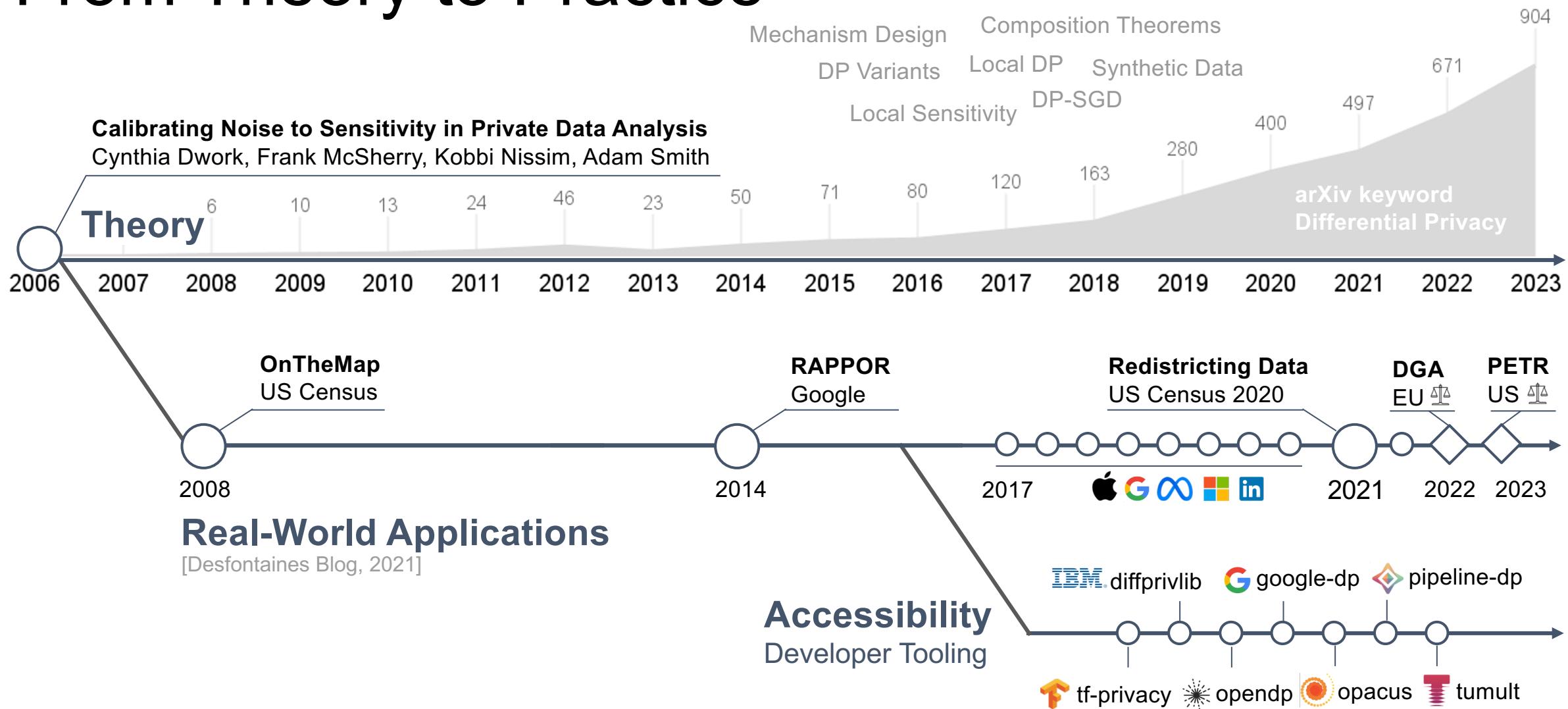
From Theory to Practice



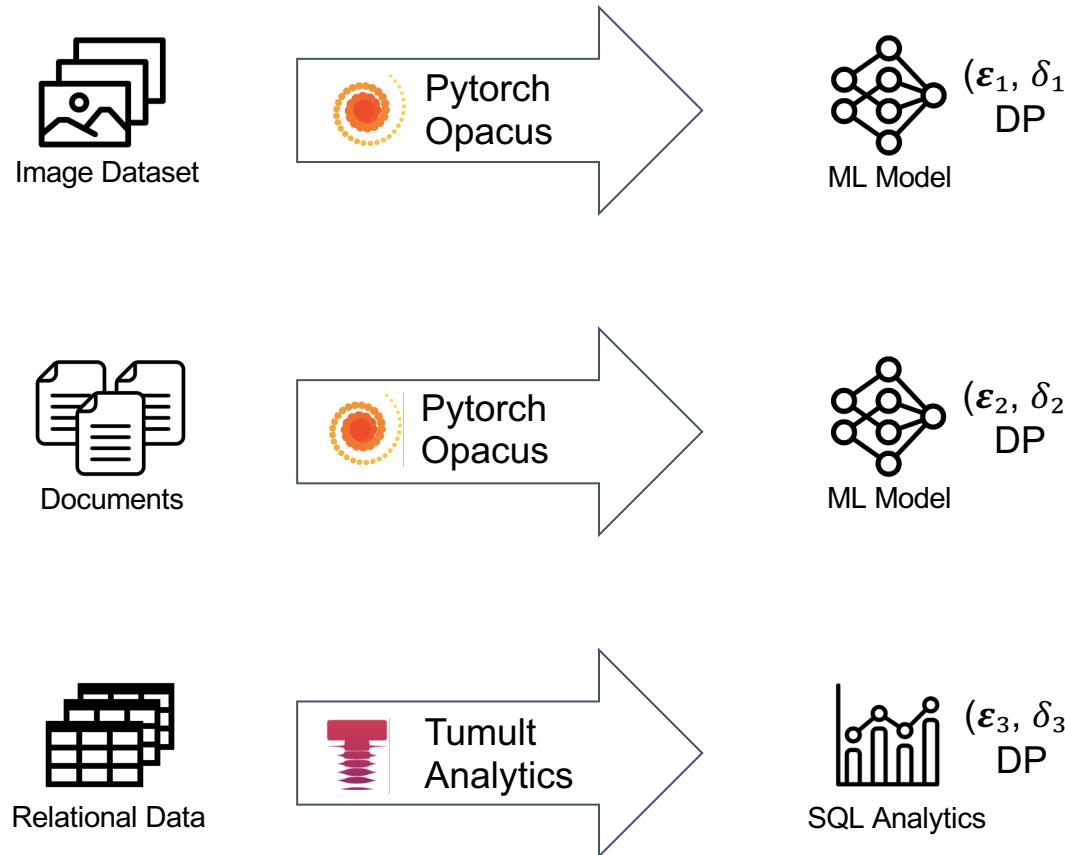
From Theory to Practice



From Theory to Practice



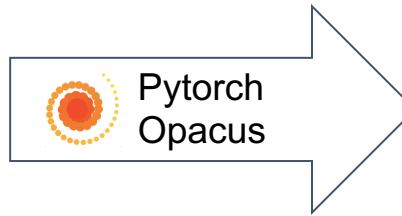
Deploying DP Applications



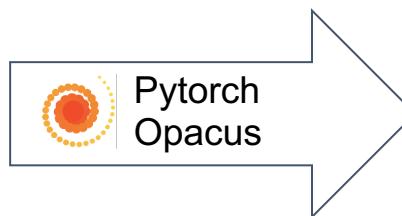
Deploying DP Applications



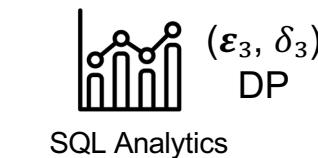
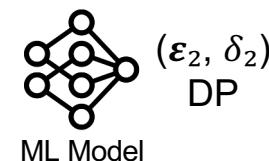
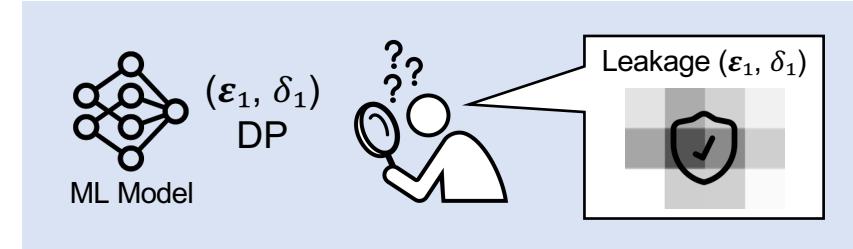
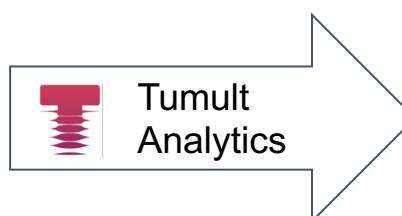
Image Dataset



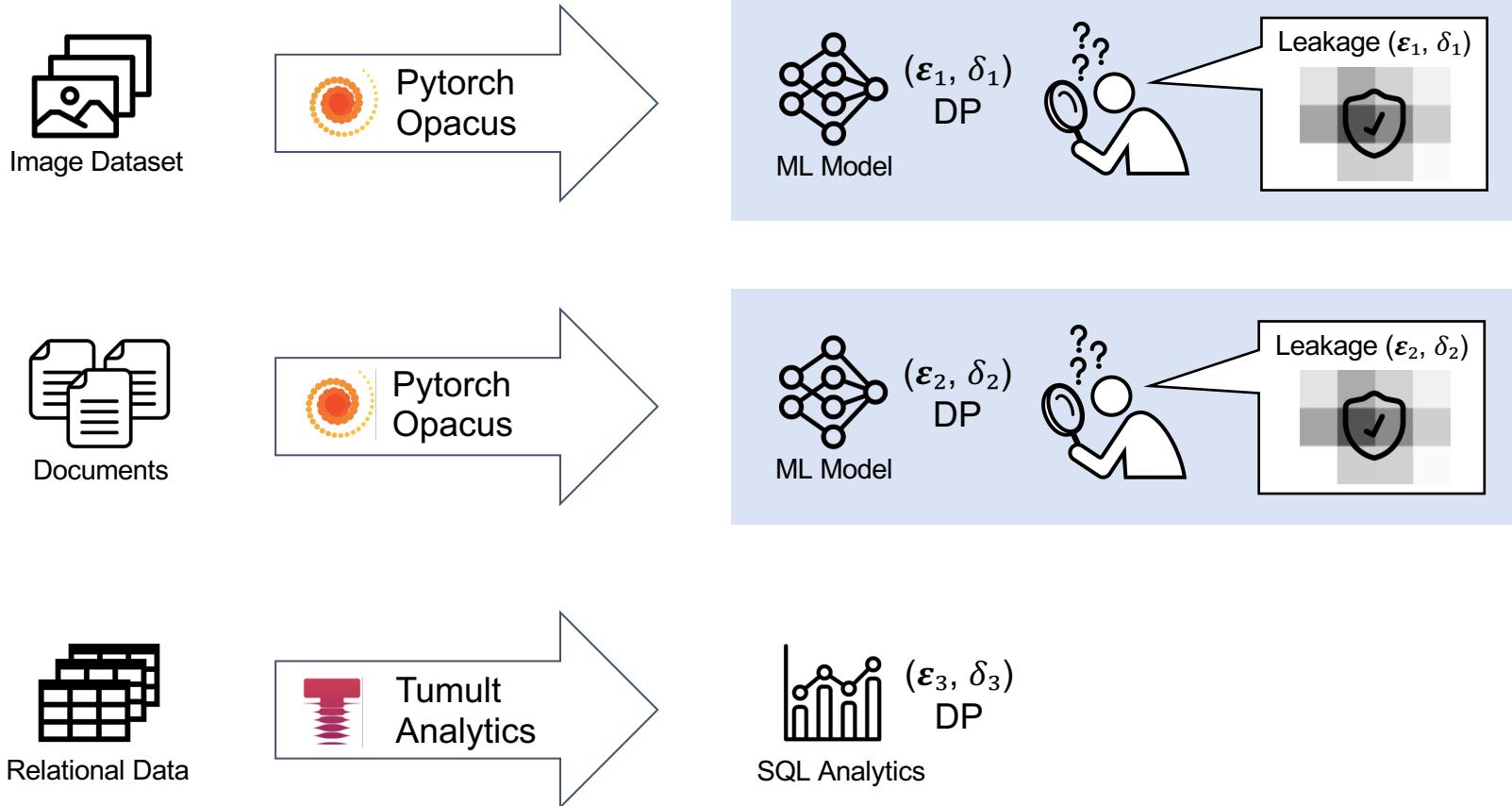
Documents



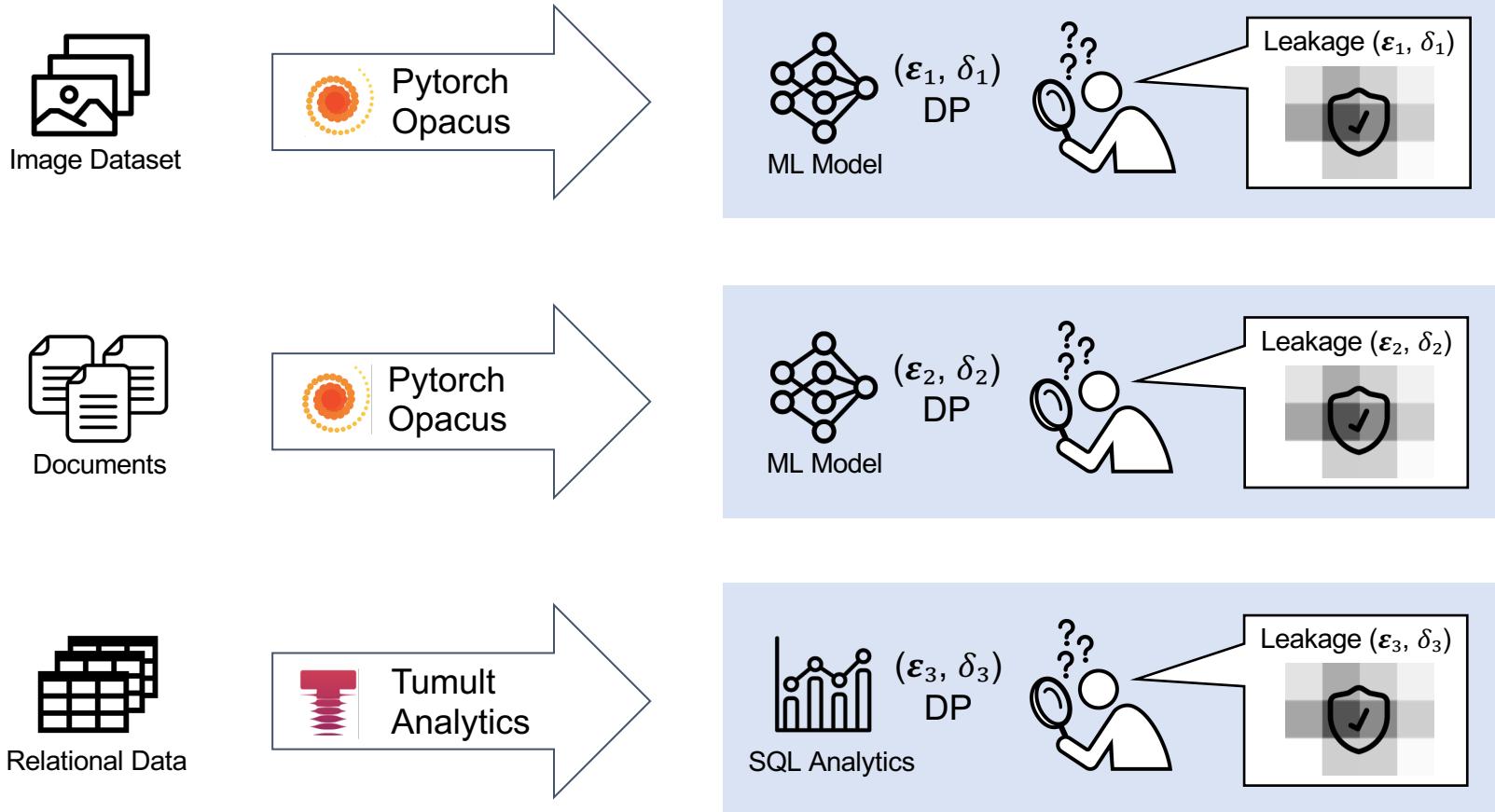
Relational Data



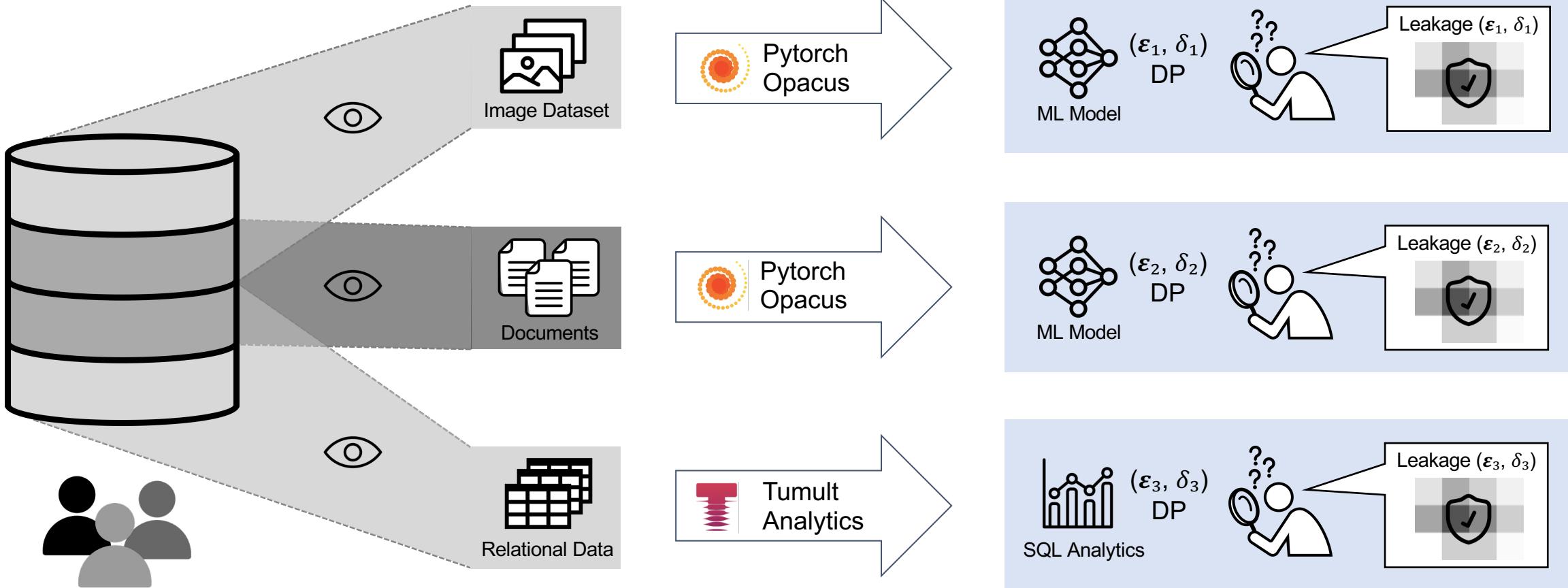
Deploying DP Applications



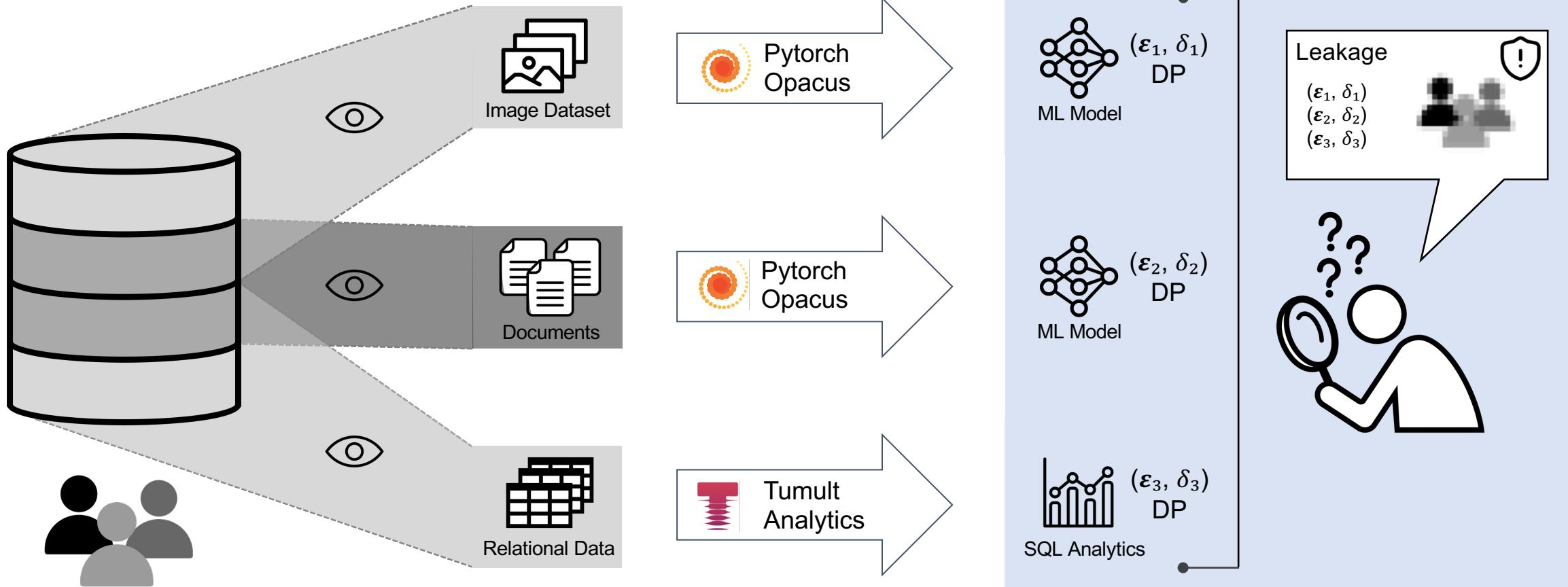
Deploying DP Applications



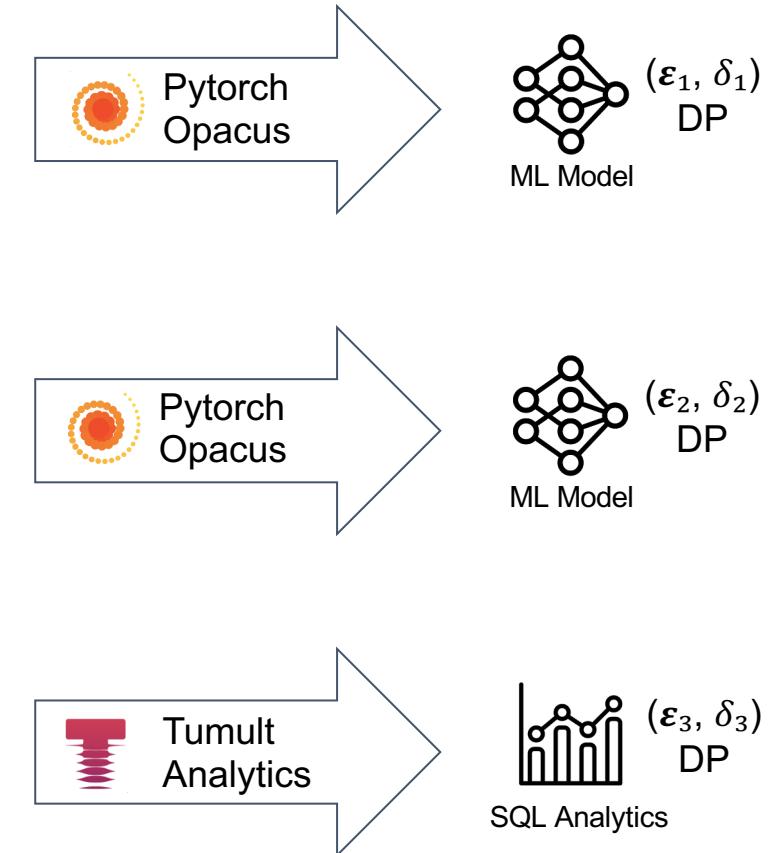
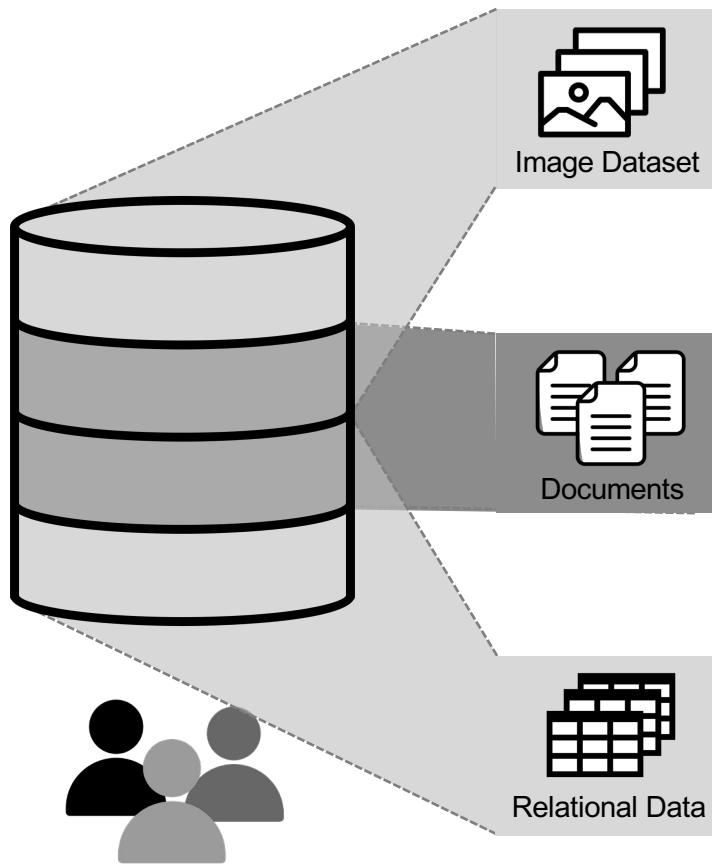
Deploying DP Applications



Deploying DP Applications



Cohere: Managing DP

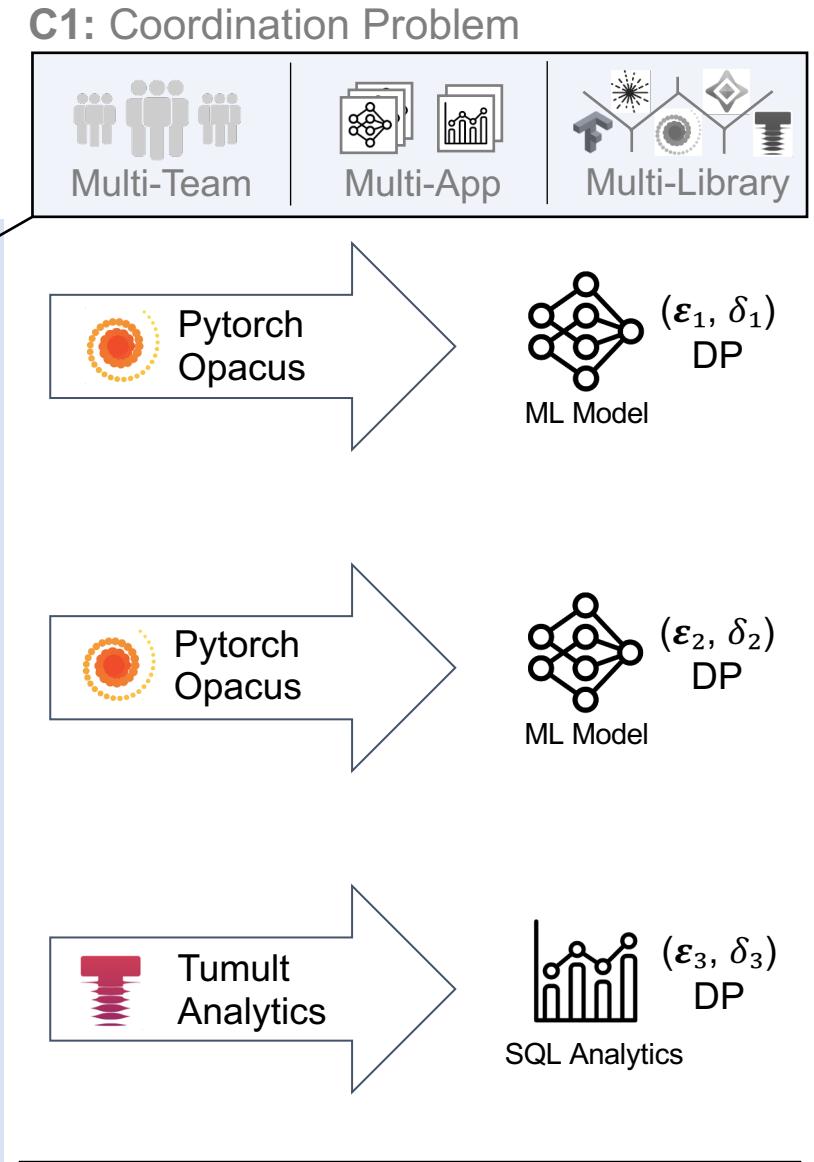
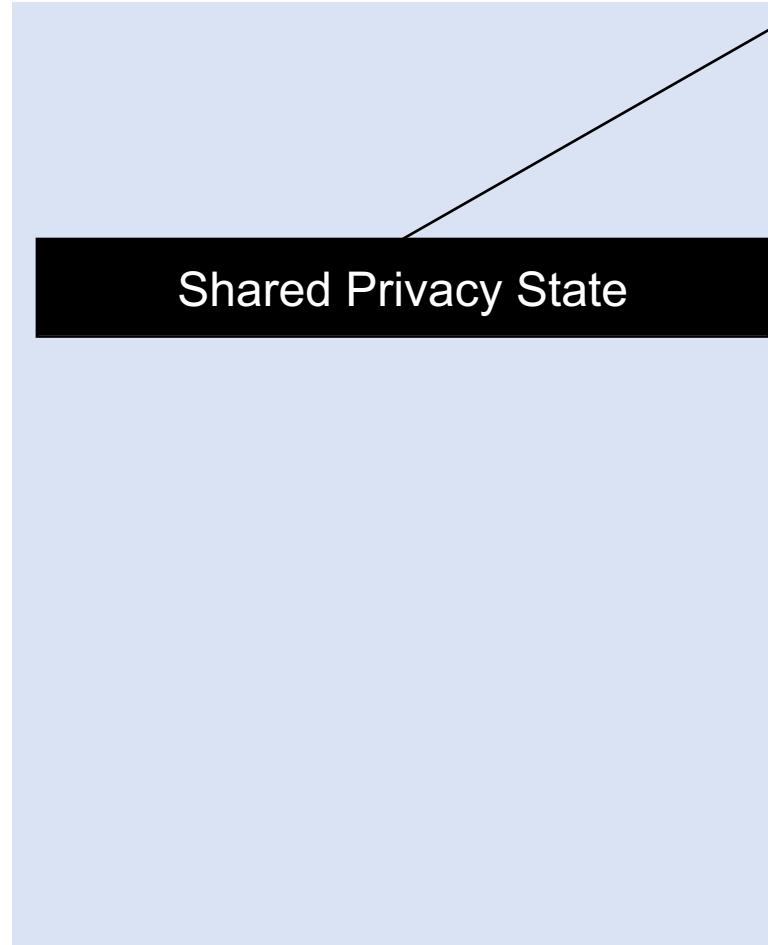
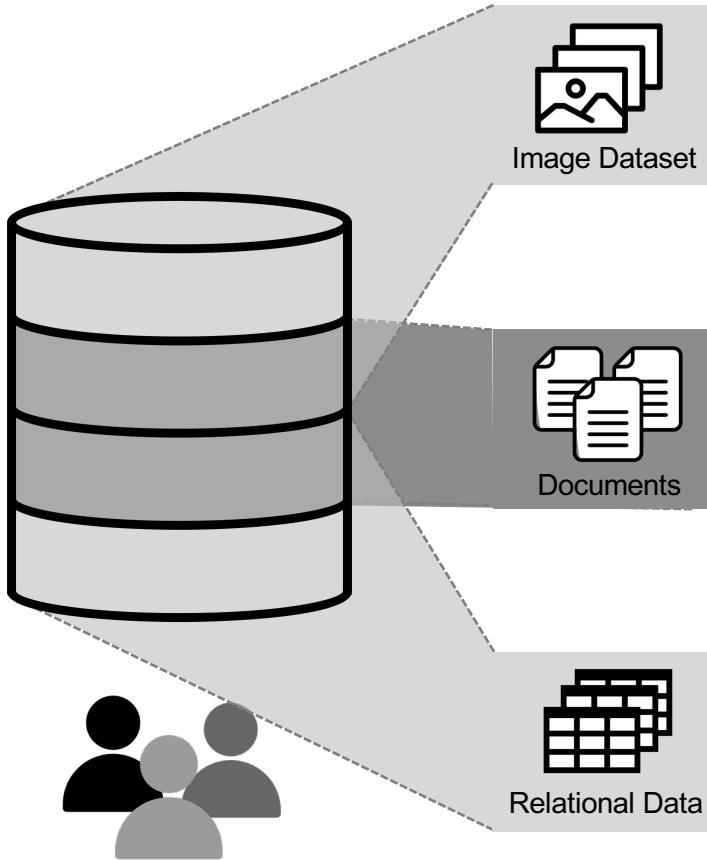


Data Layer

DP Management Layer

Application Layer

Cohere: Managing DP

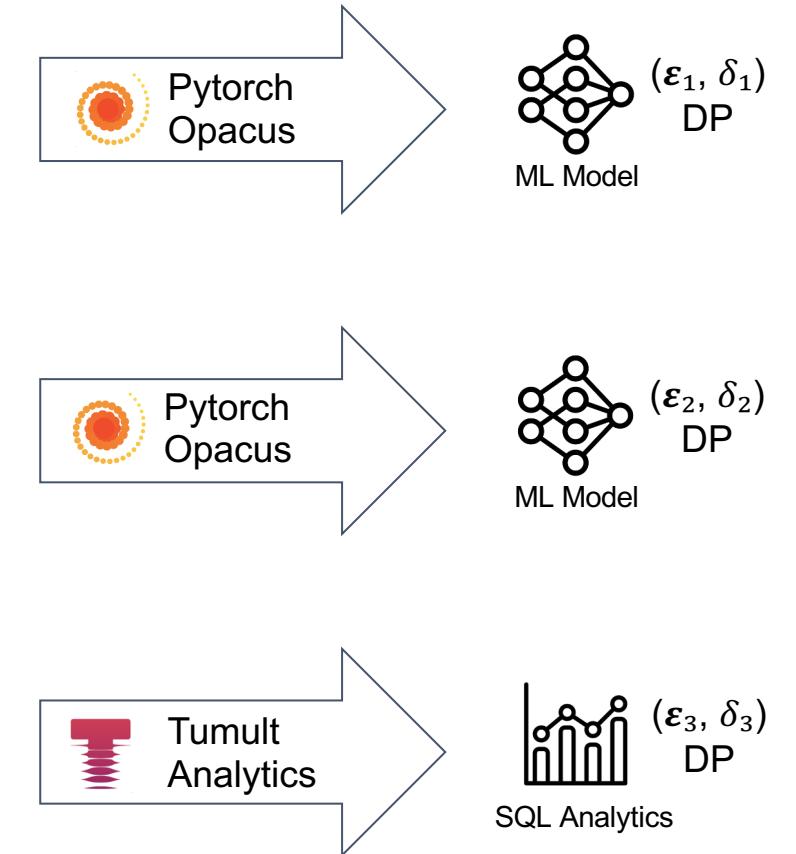
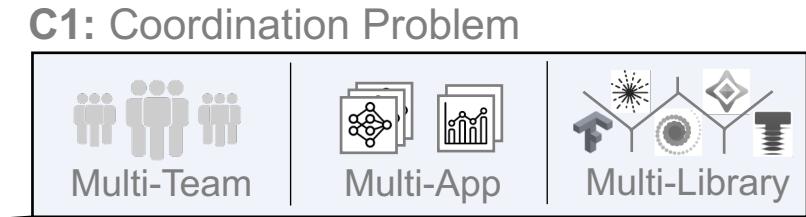
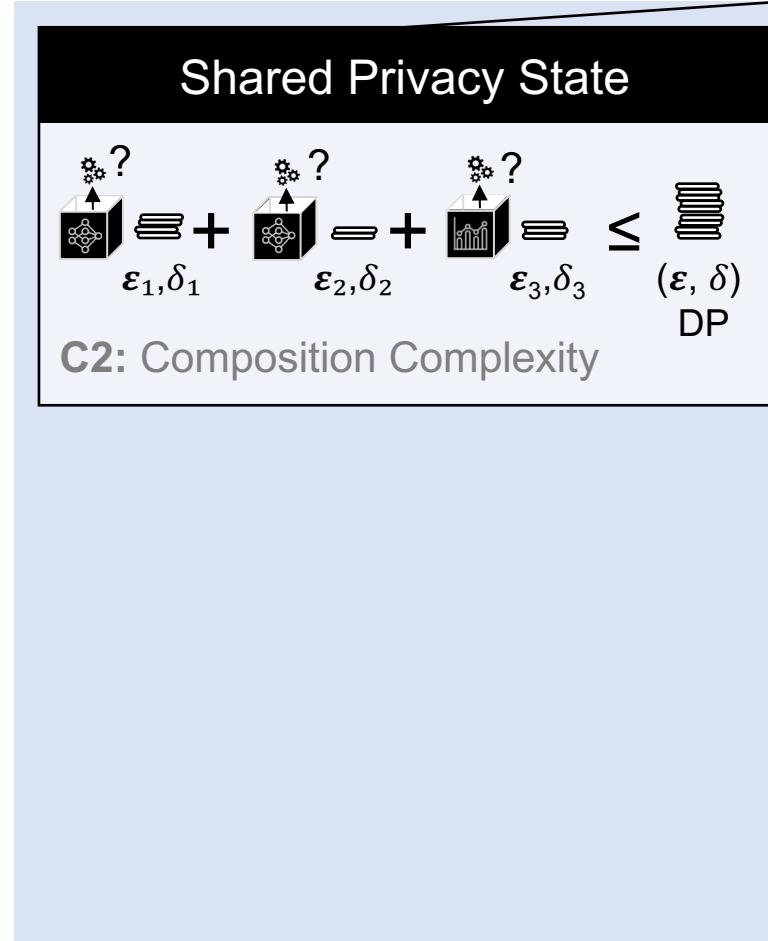
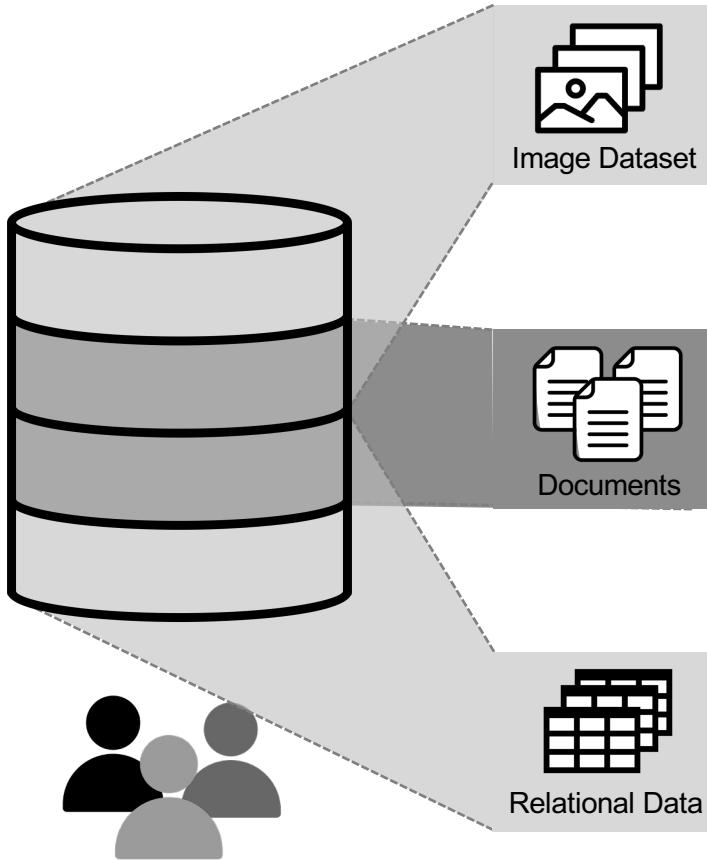


Data Layer

DP Management Layer

Application Layer

Cohere: Managing DP

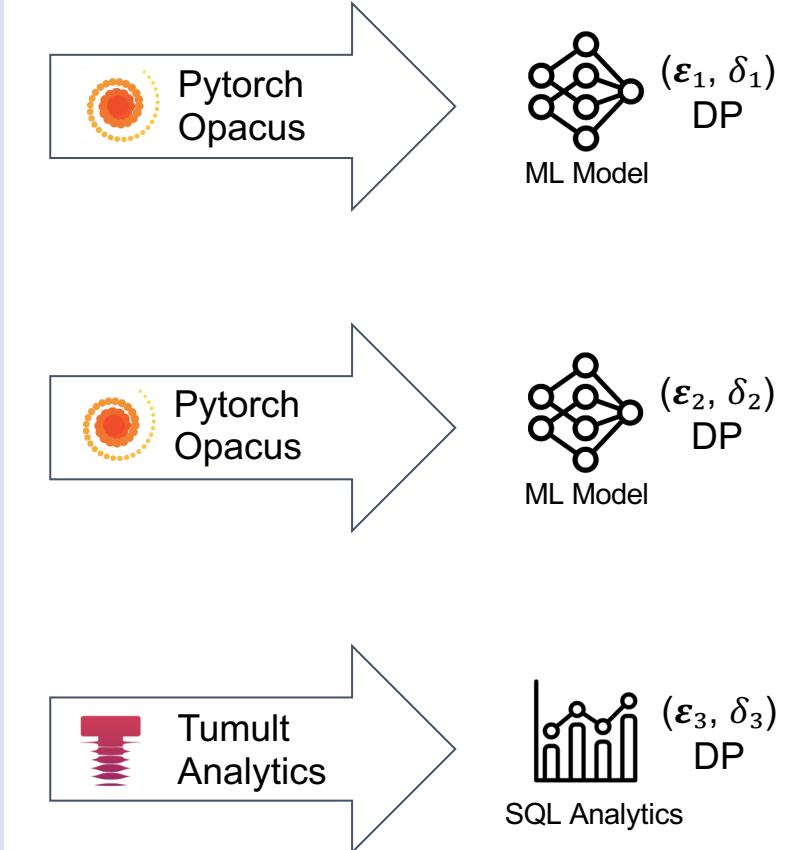
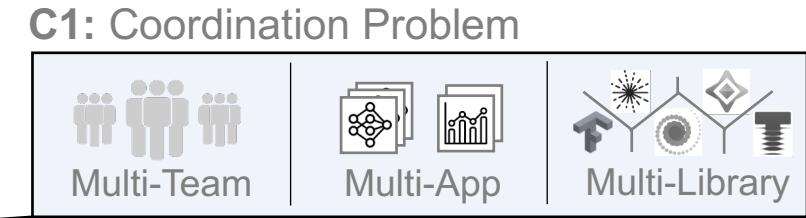
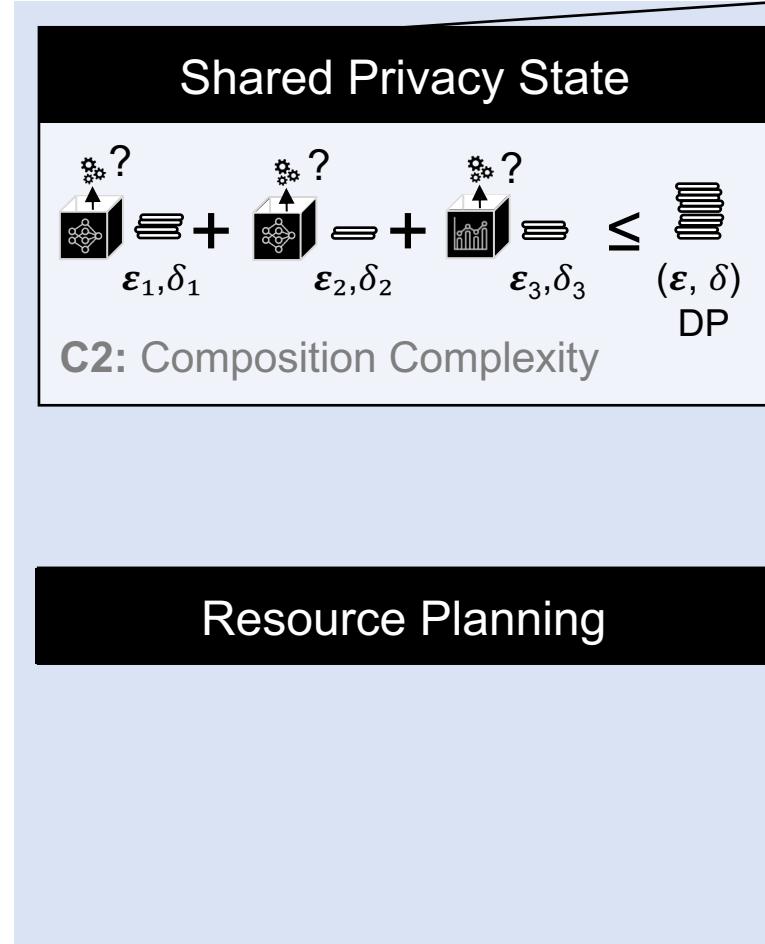
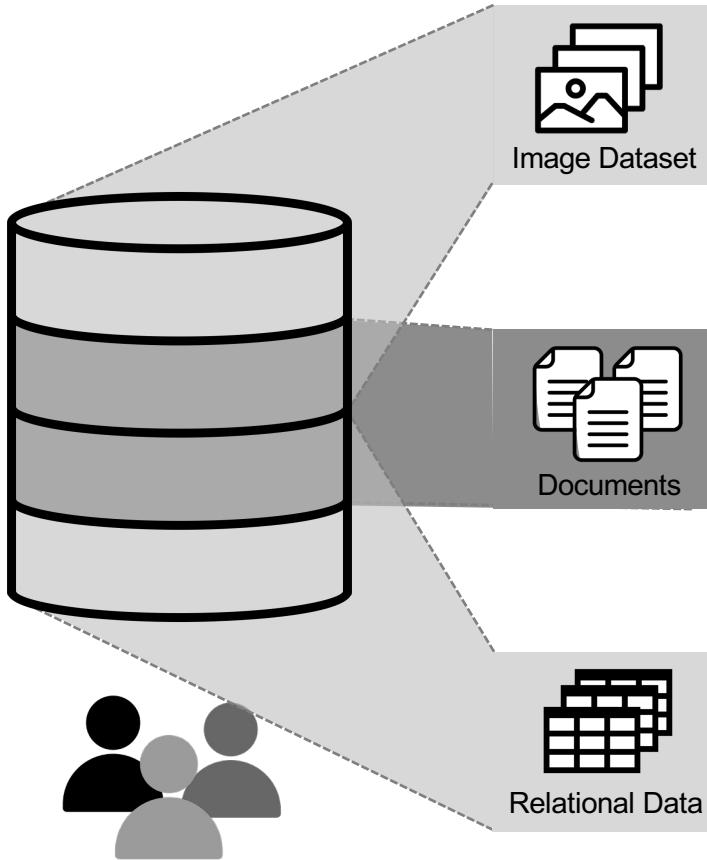


Data Layer

DP Management Layer

Application Layer

Cohere: Managing DP

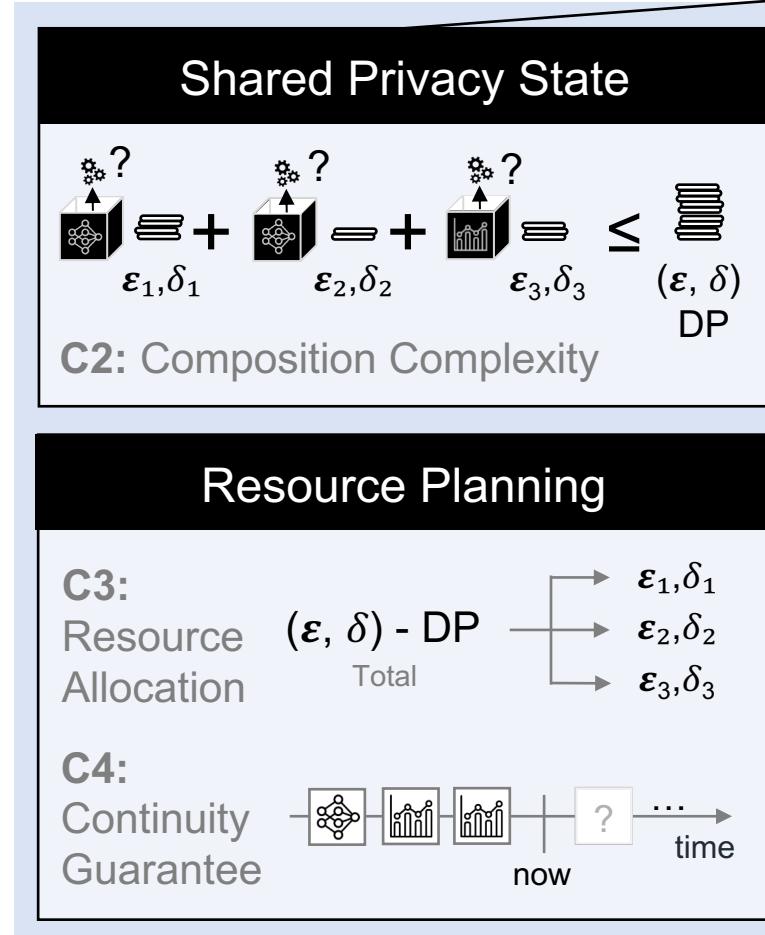
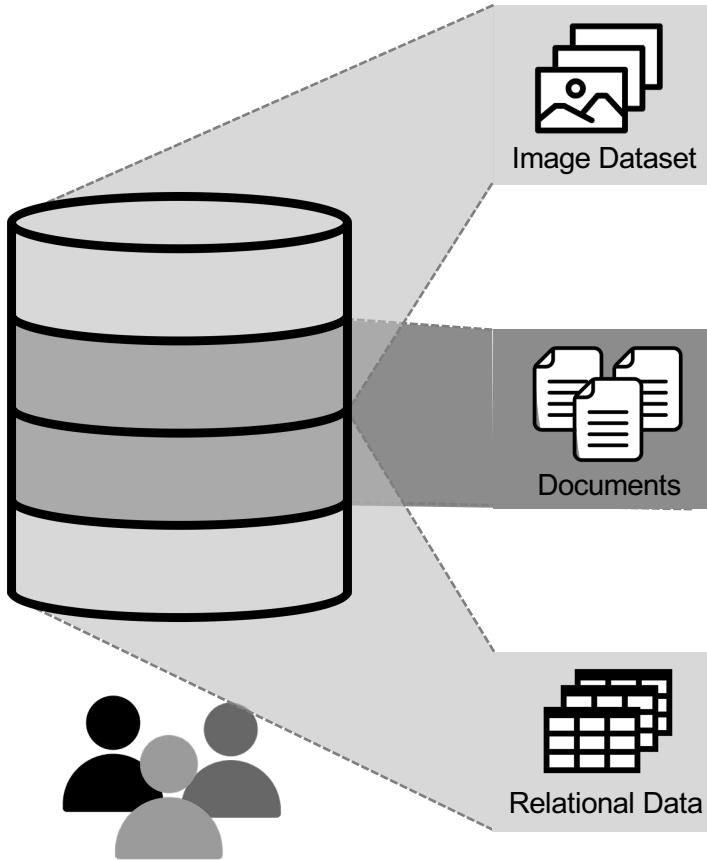


Data Layer

DP Management Layer

Application Layer

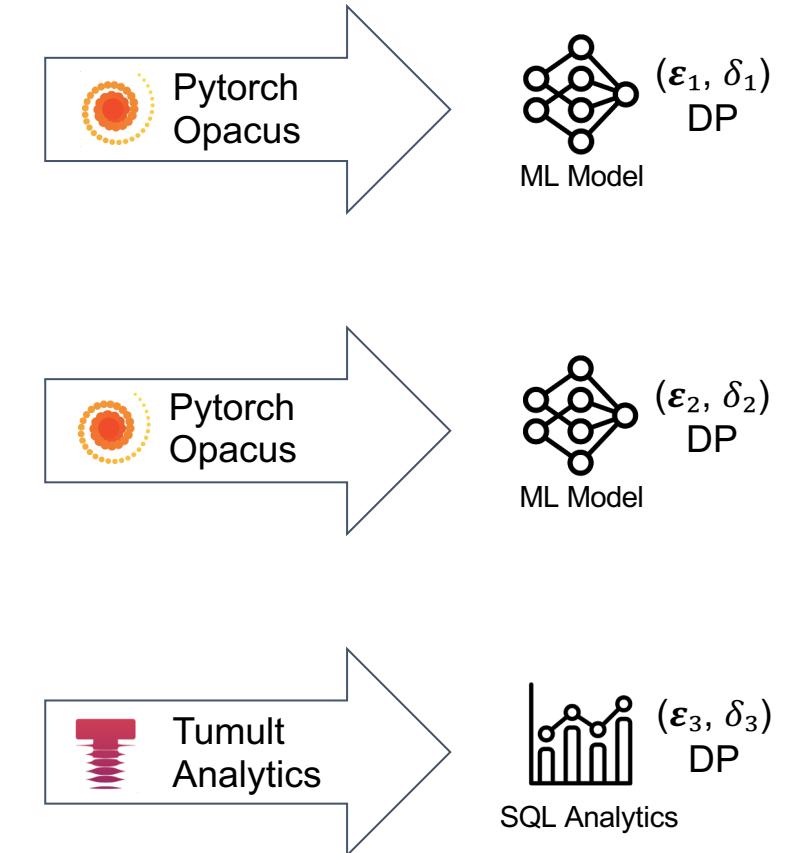
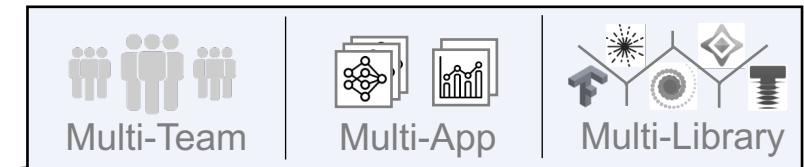
Cohere: Managing DP



Data Layer

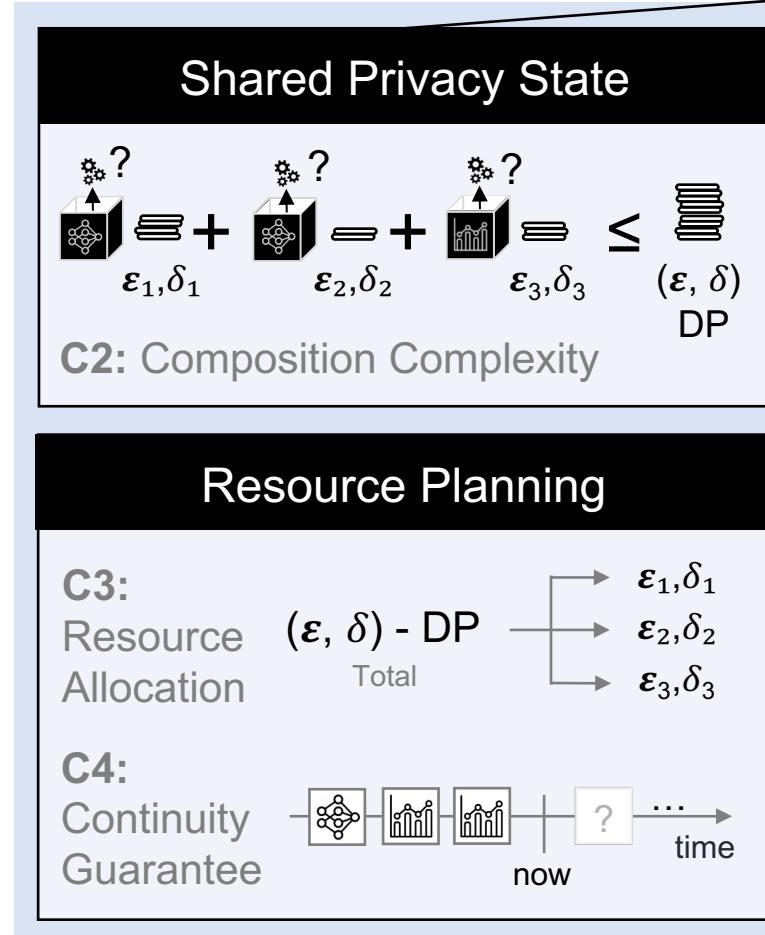
DP Management Layer

C1: Coordination Problem



Application Layer

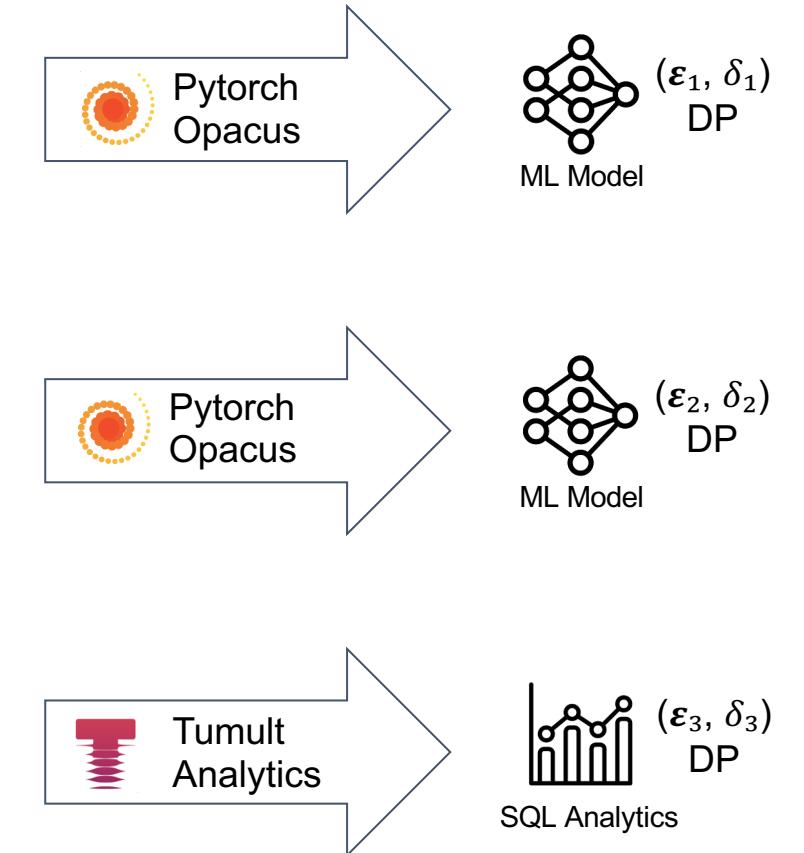
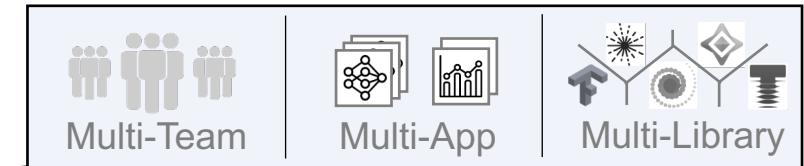
Cohere: Managing DP



Data Layer

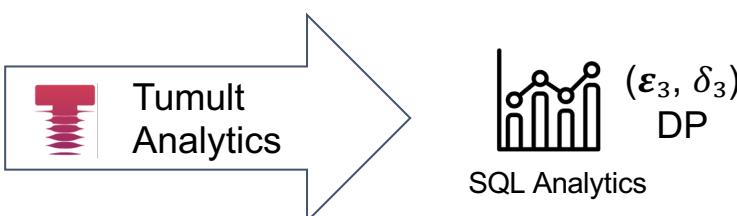
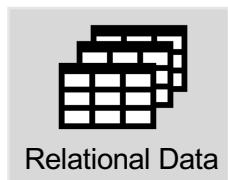
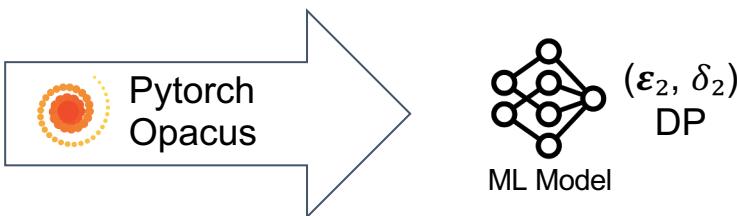
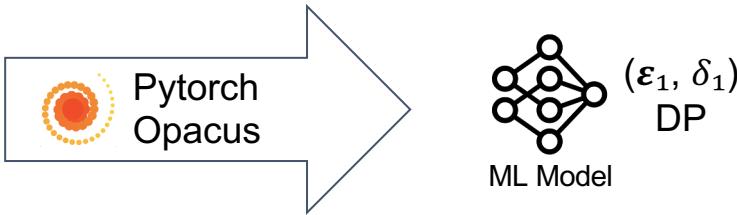
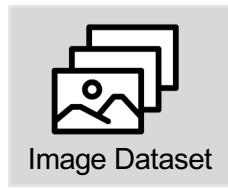
DP Management Layer

C1: Coordination Problem



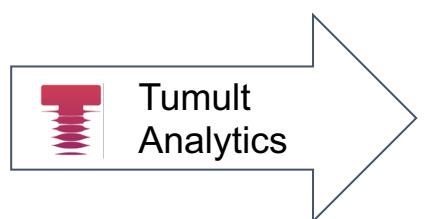
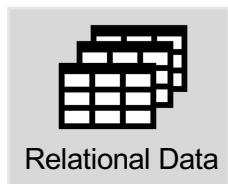
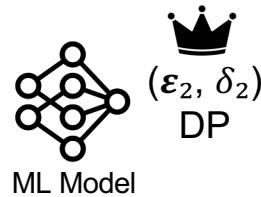
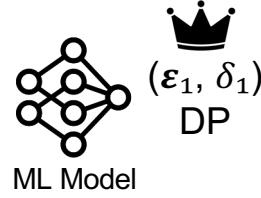
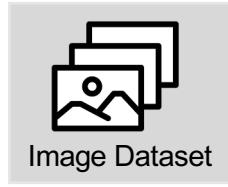
Application Layer

Unifying the Application Layer



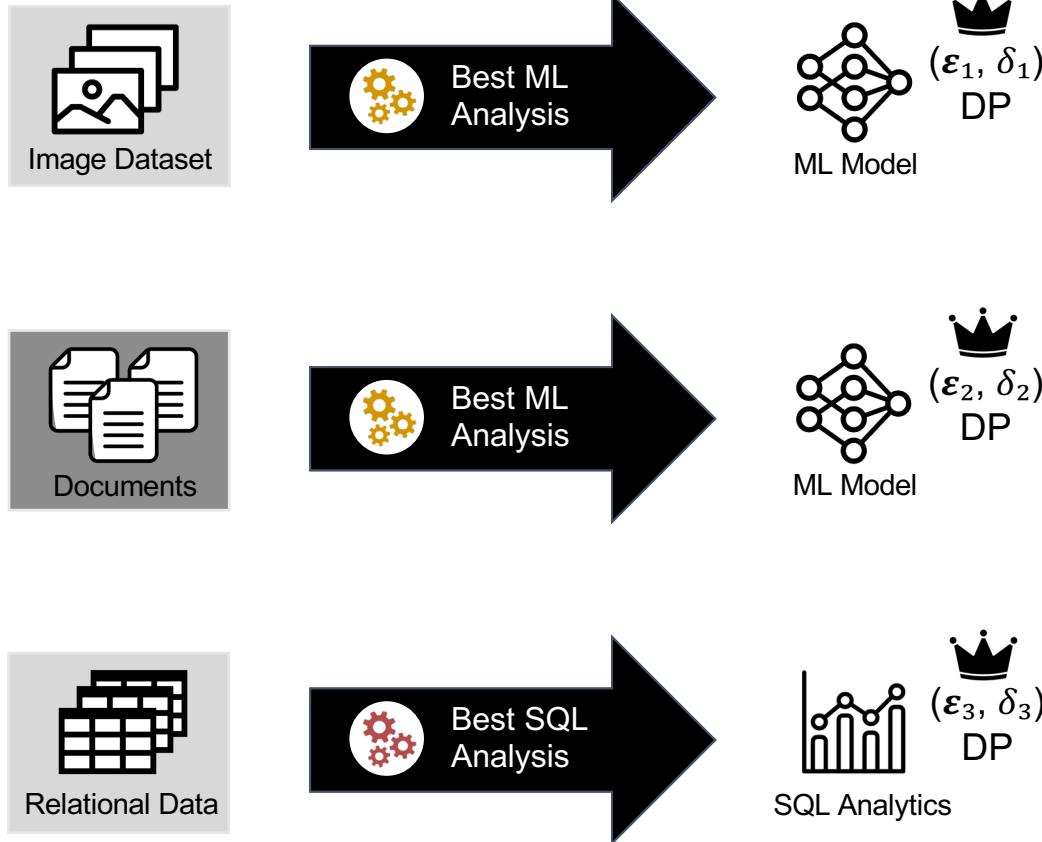
Application Layer

Unifying the Application Layer



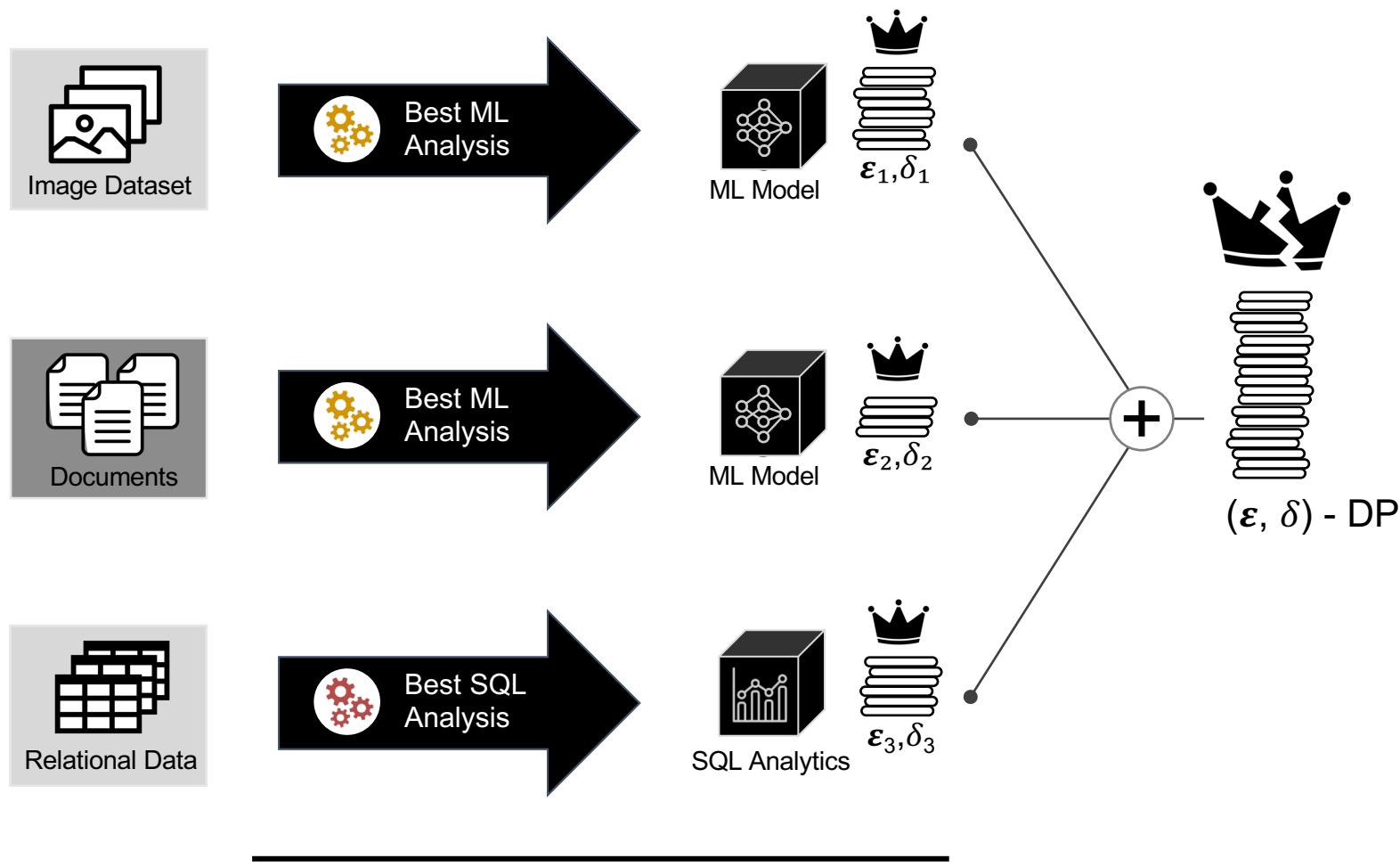
Application Layer

Unifying the Application Layer



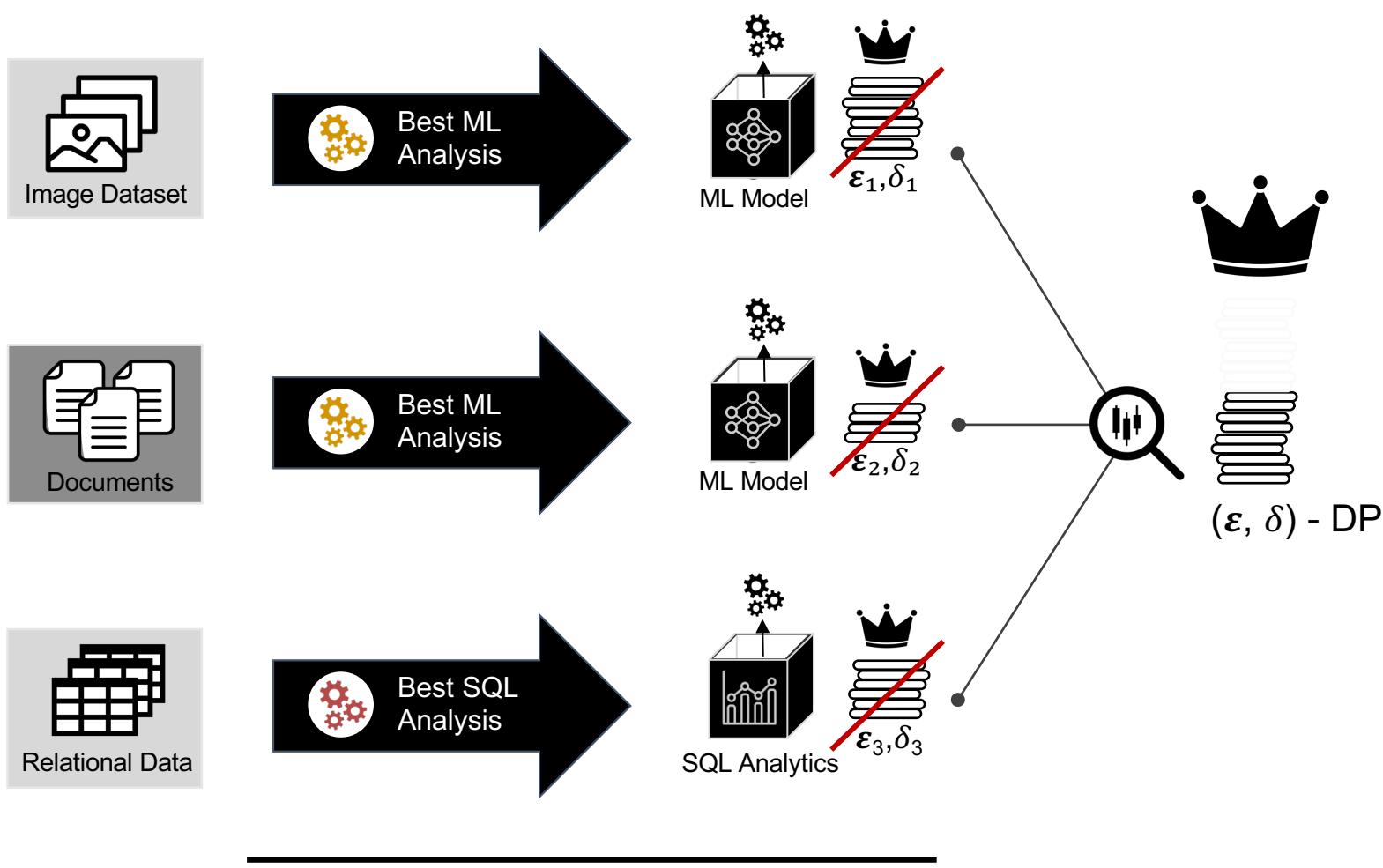
Application Layer

Unifying the Application Layer



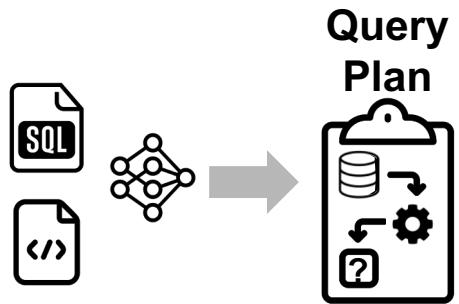
Application Layer

Unifying the Application Layer

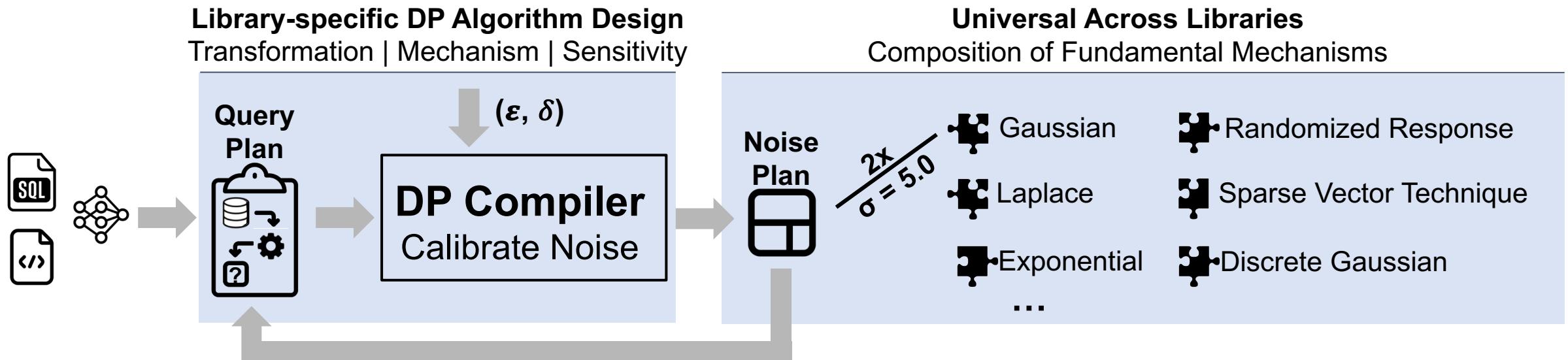


Moving Beyond
Local Optima

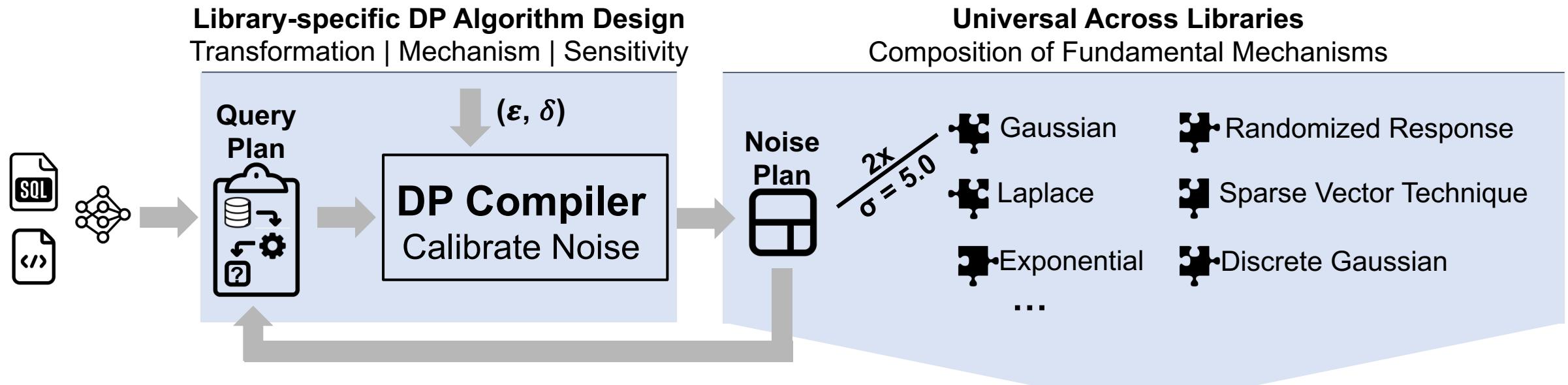
DP Libraries: In a Nutshell



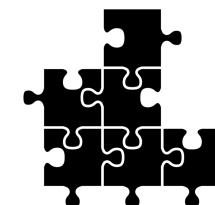
DP Libraries: In a Nutshell



DP Libraries: In a Nutshell

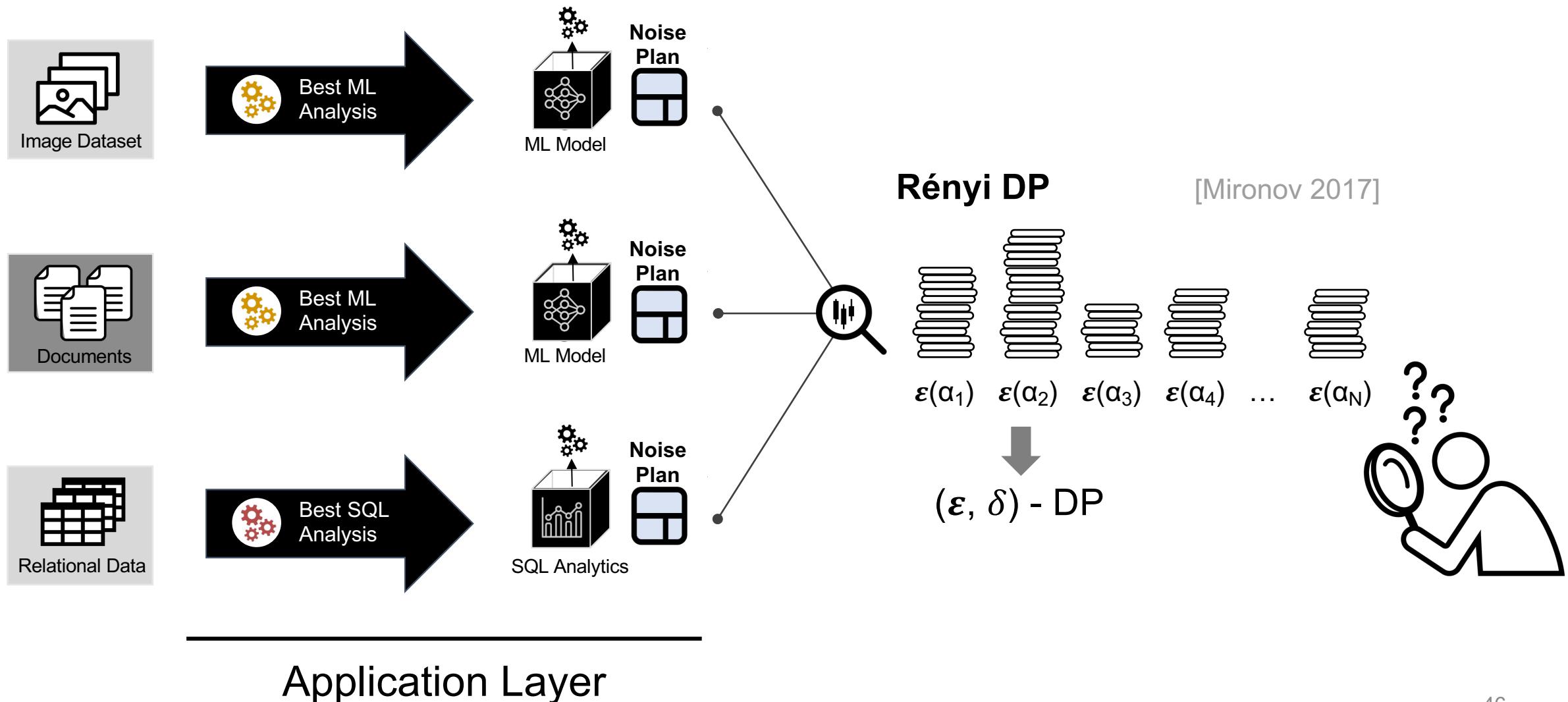


If we can compose all fundamental mechanisms, we can support a variety of heterogeneous libraries through a unified noise plan.

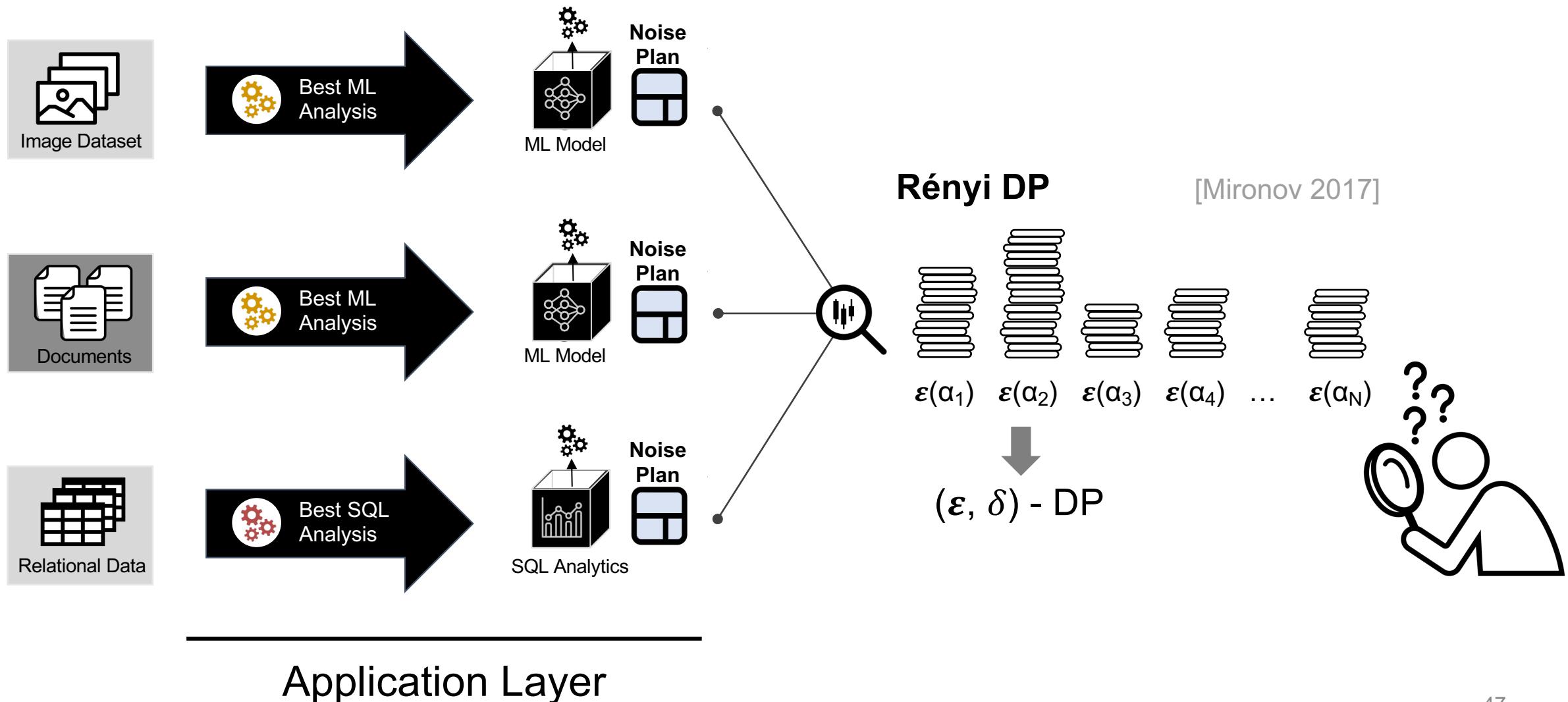


Composition of Fundamental Mechanisms

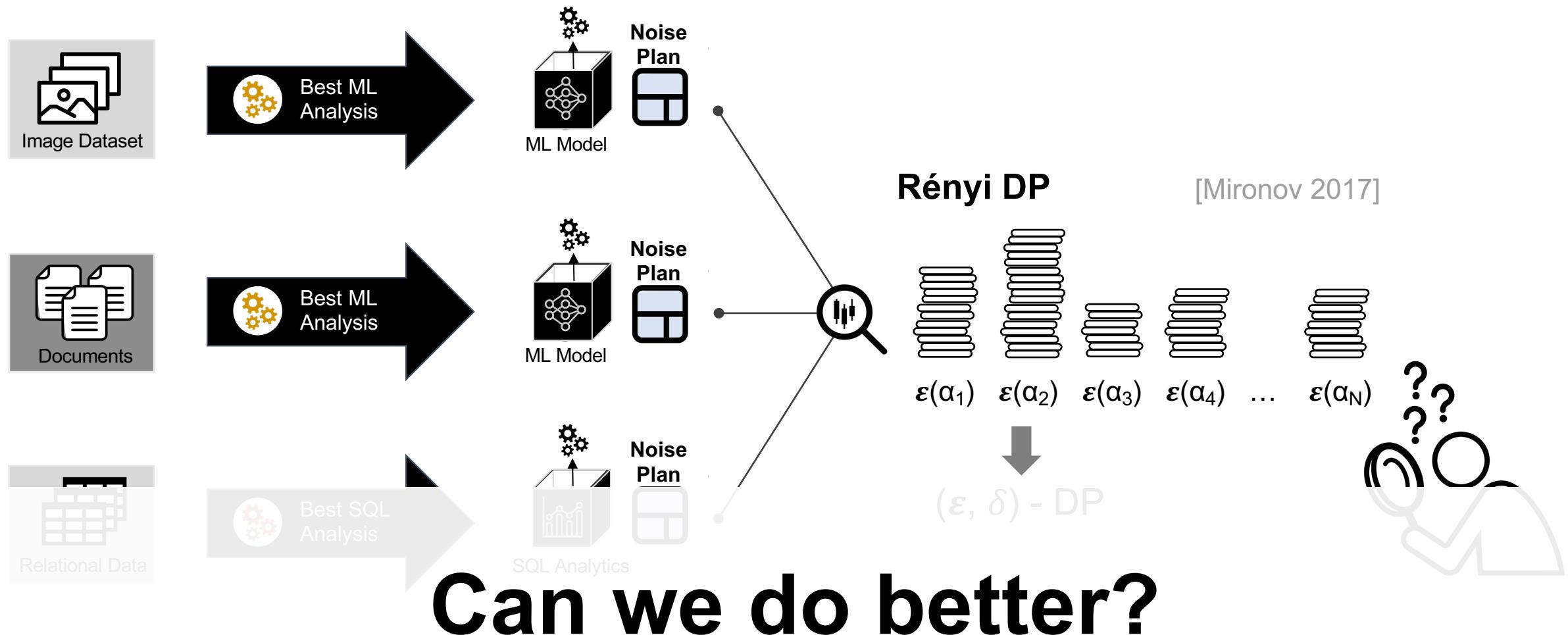
Unifying the Application Layer



Unifying the Application Layer

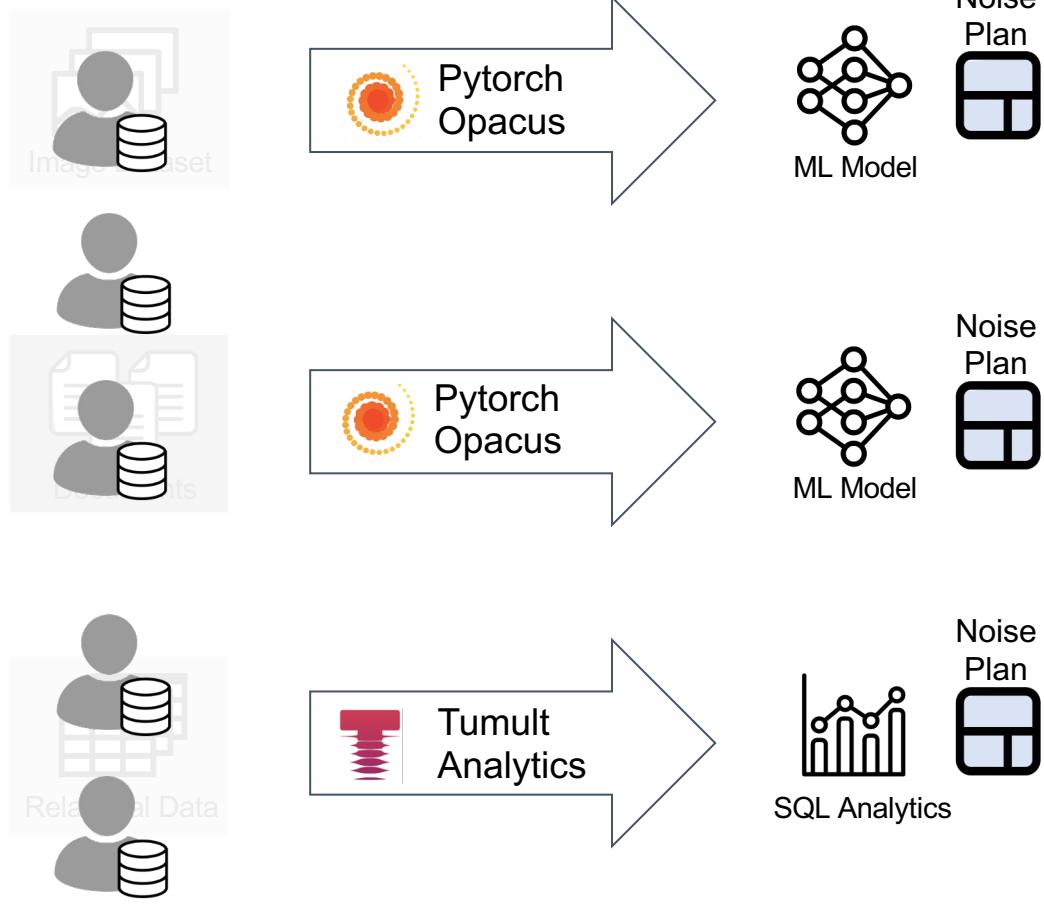


Unifying the Application Layer



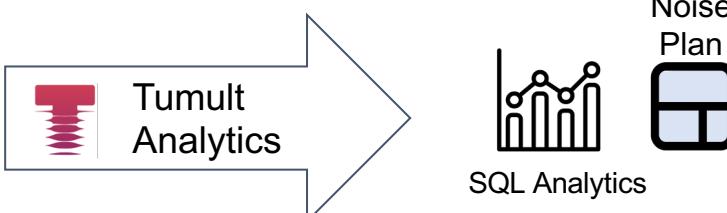
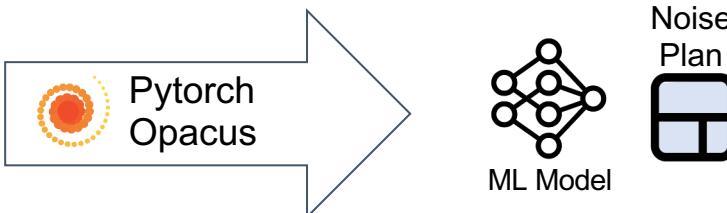
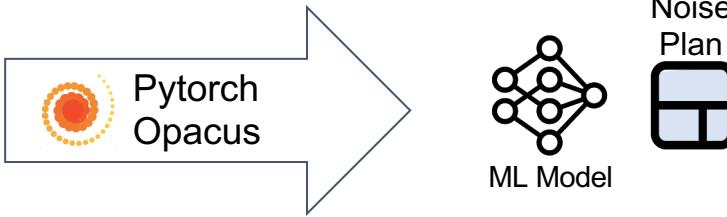
Application Layer

Benefiting from Data Access Patterns



Application Layer

Benefiting from Data Access Patterns

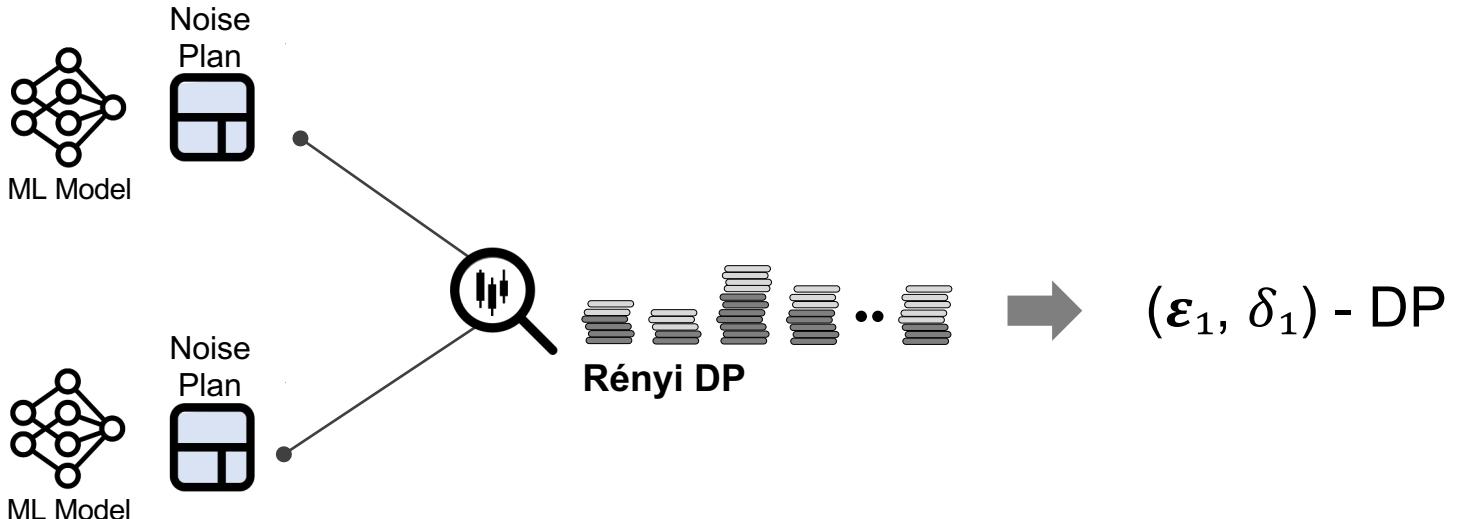
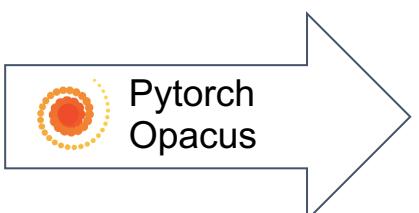
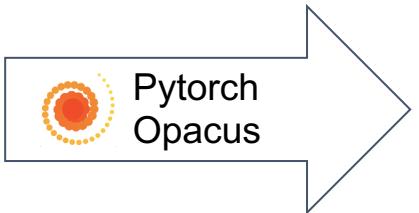


Parallel Composition

[McSherry 2009]

Application Layer

Benefiting from Data Access Patterns

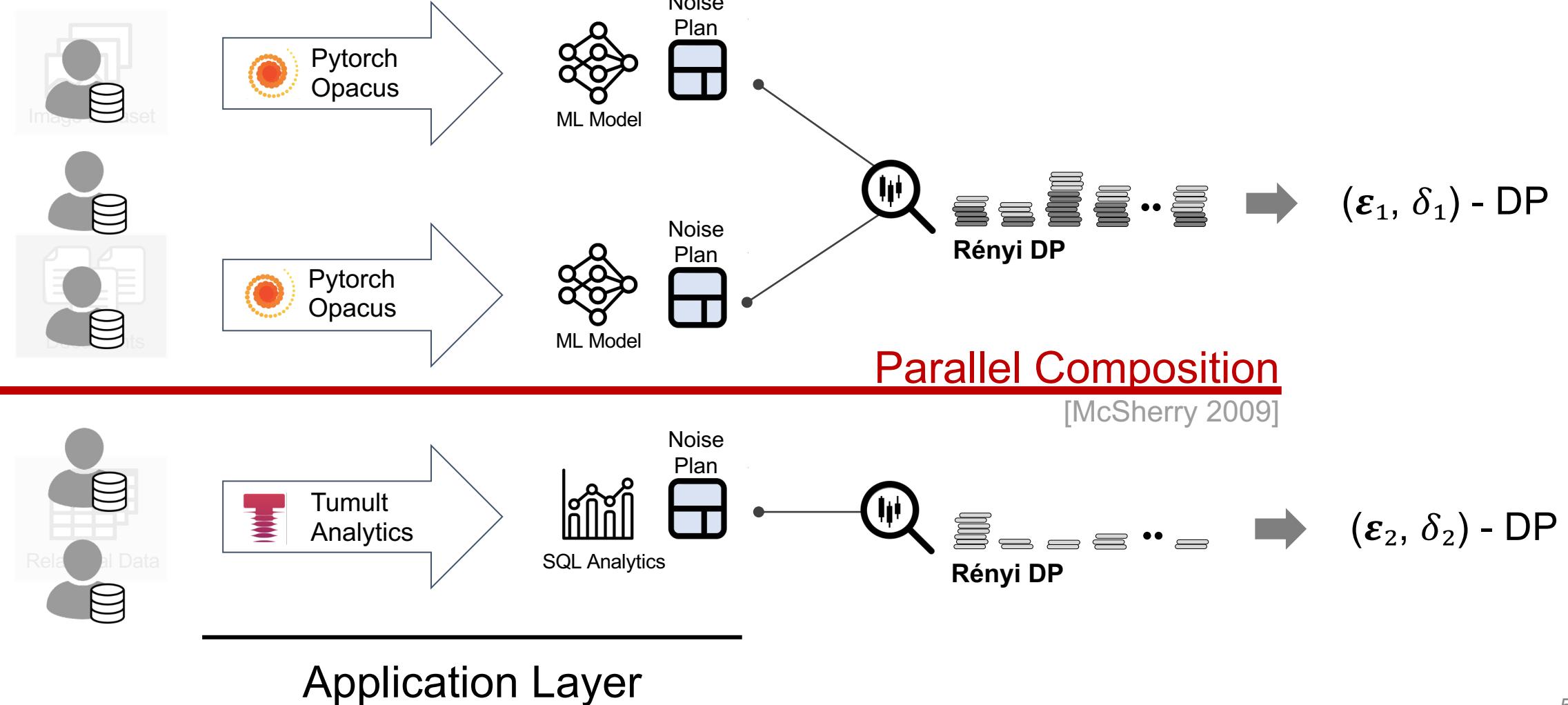


Parallel Composition

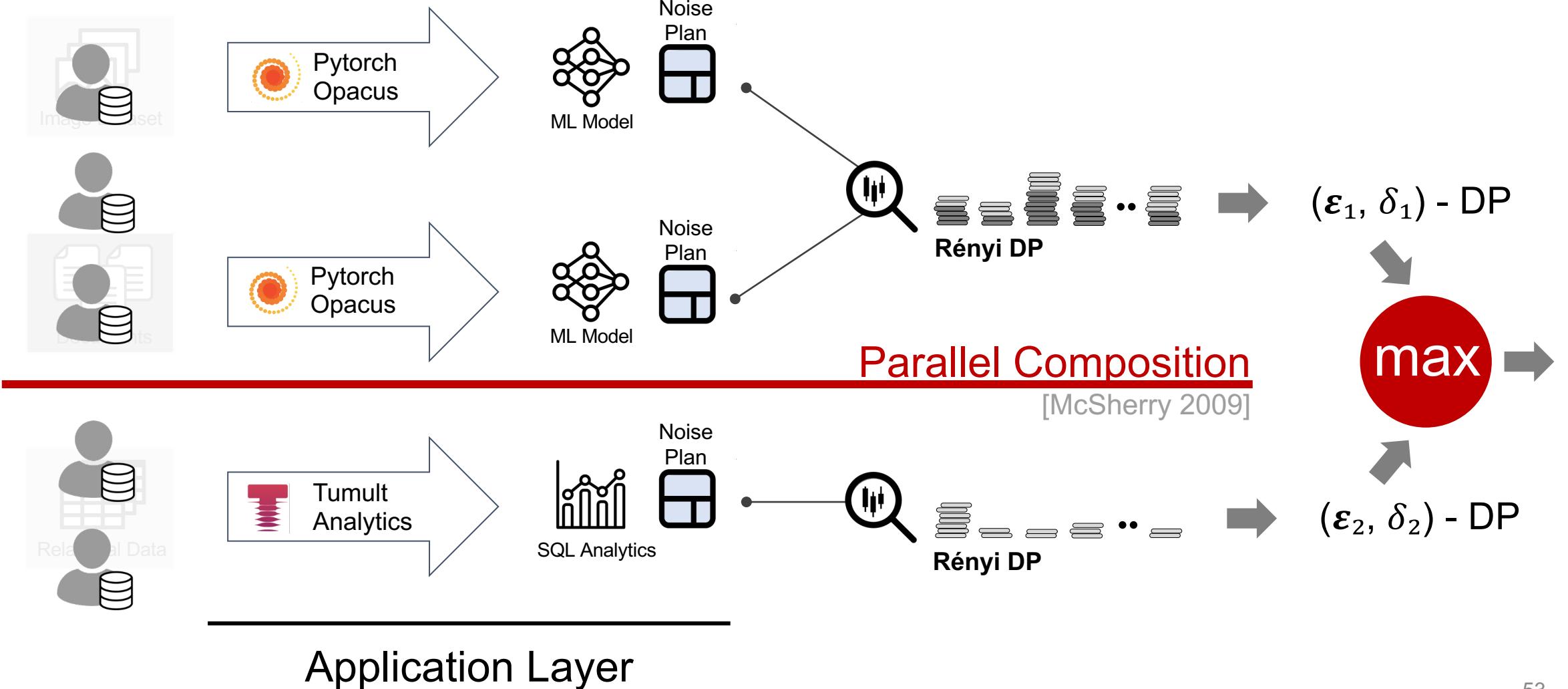
[McSherry 2009]

Application Layer

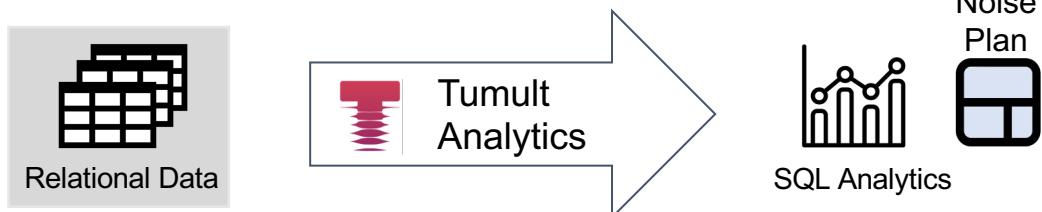
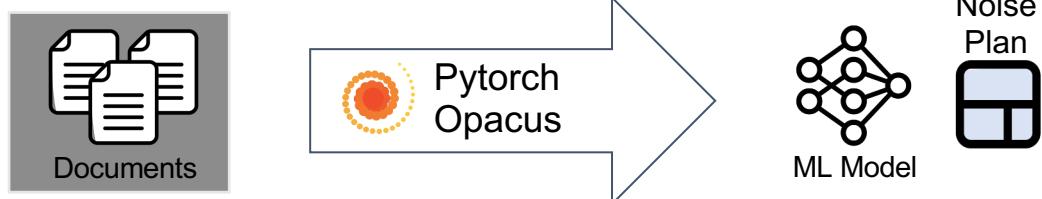
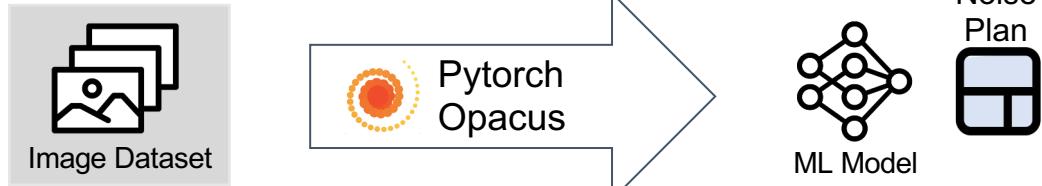
Benefiting from Data Access Patterns



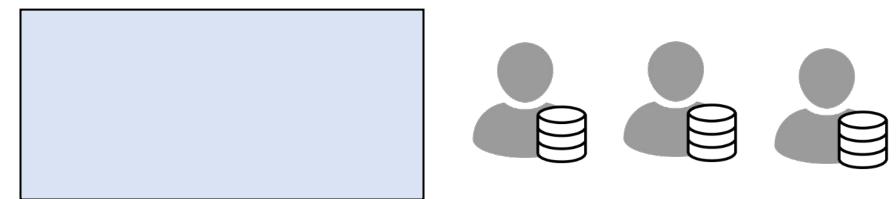
Benefiting from Data Access Patterns



Benefiting from Data Access Patterns



Application Layer

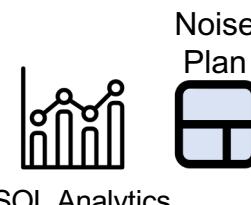
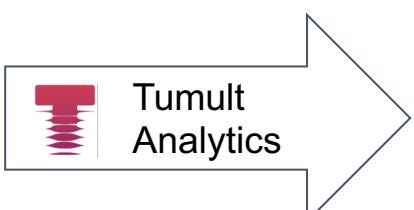
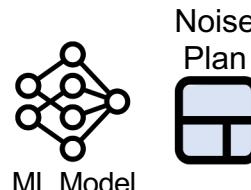
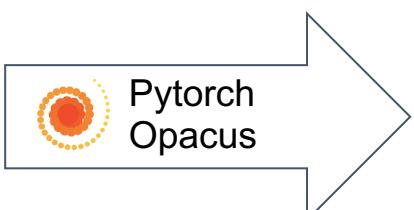
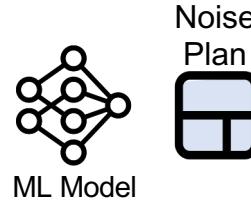
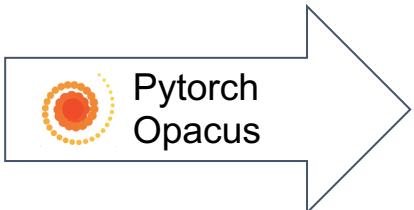
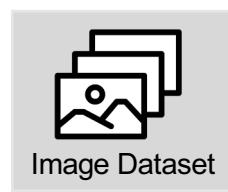


Block Composition
[Lécuyer SOSP'19]



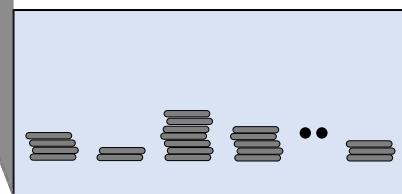
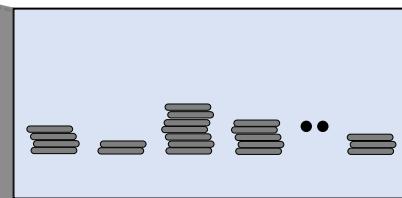
Total: $\max(\square, \square)$

Benefiting from Data Access Patterns



Application Layer

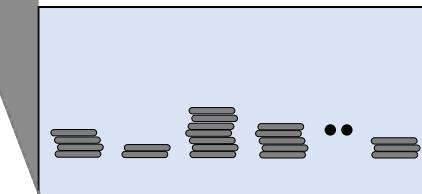
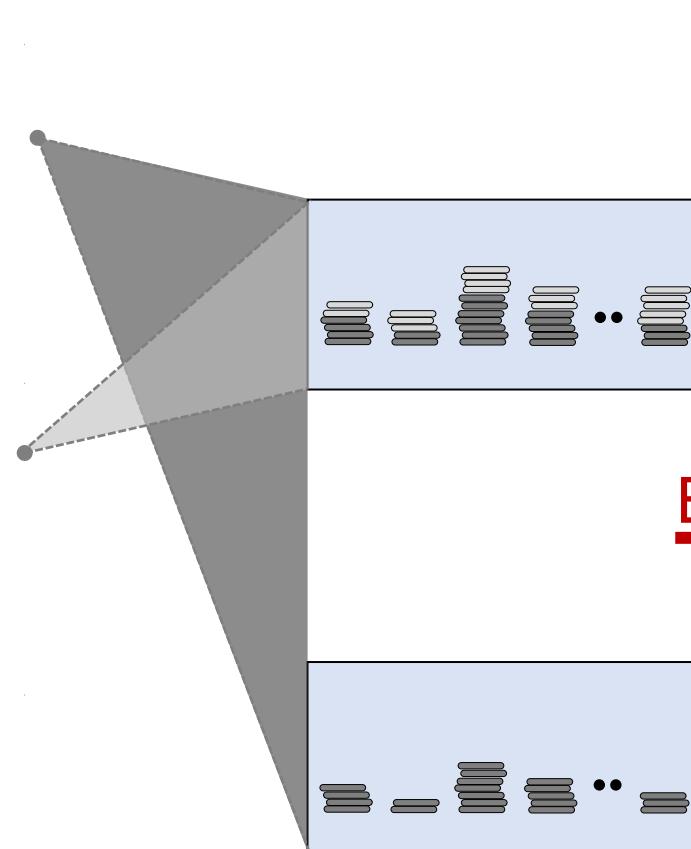
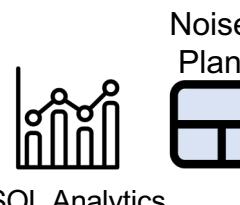
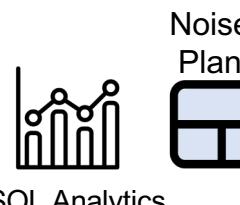
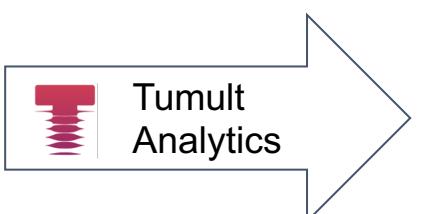
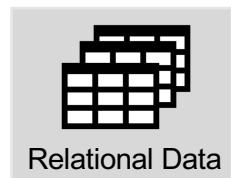
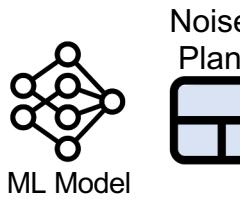
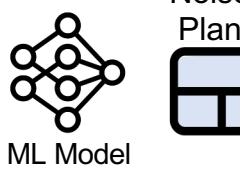
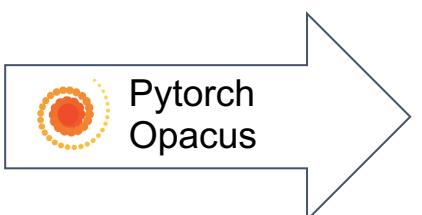
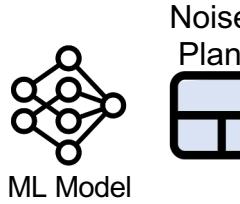
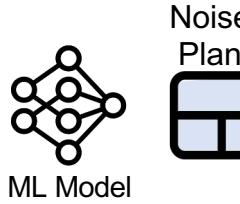
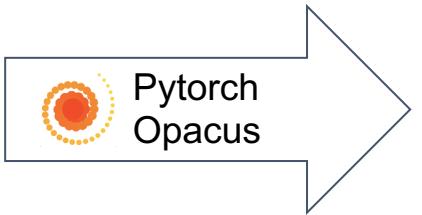
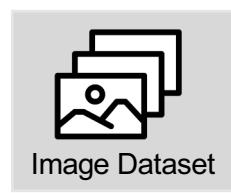
Total: $\max(\square, \square)$



Block Composition

[Lécuyer SOSP'19]

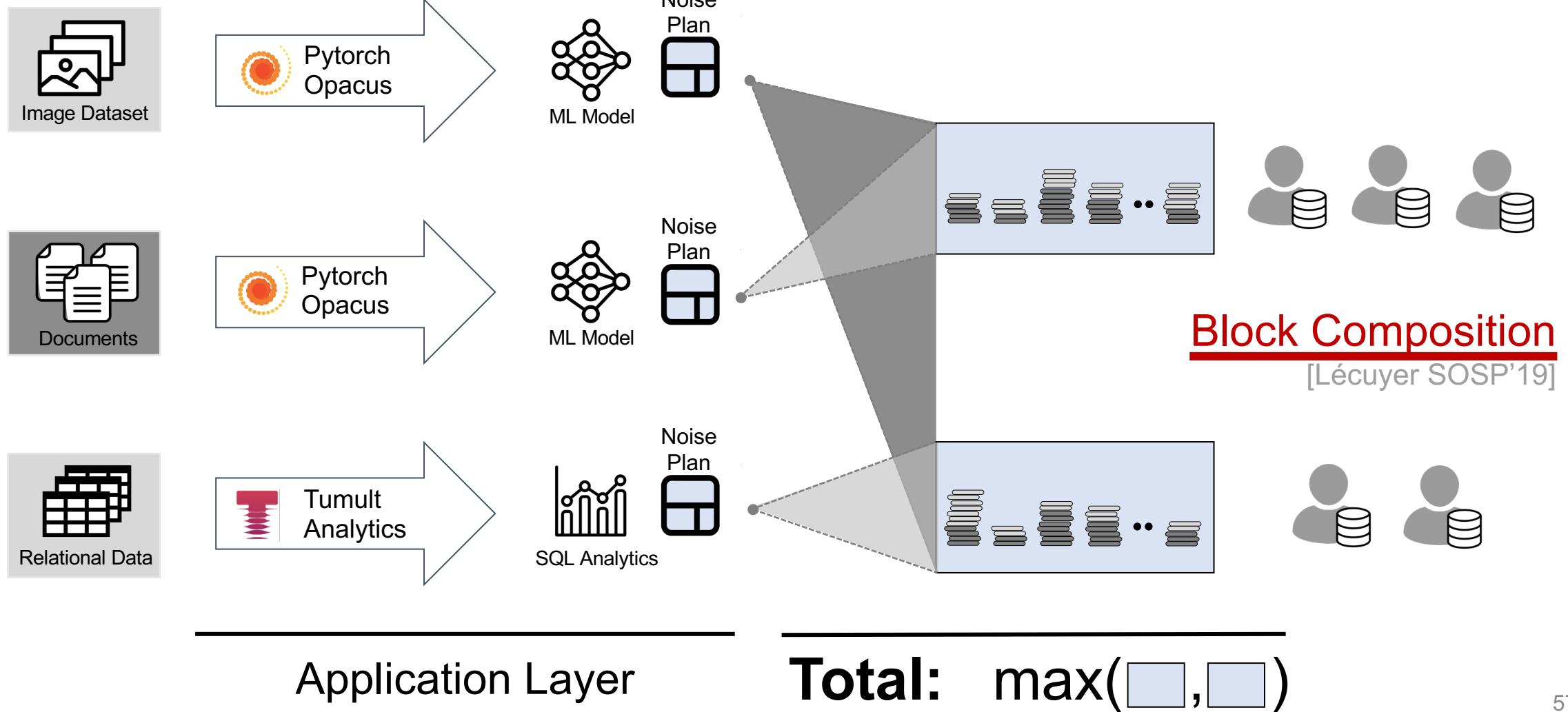
Benefiting from Data Access Patterns



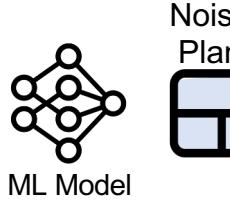
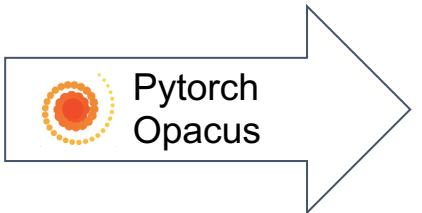
Application Layer

Total: $\max(\square, \square)$

Benefiting from Data Access Patterns



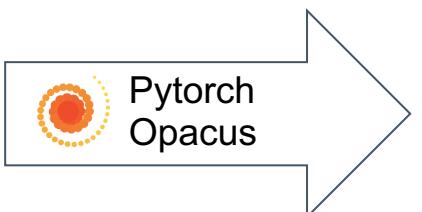
Benefiting from Data Access Patterns



Noise Plan



ML Model

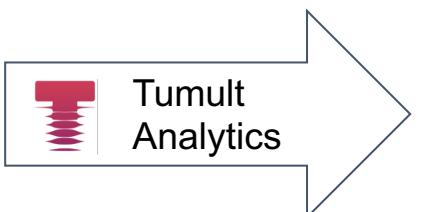


Noise Plan



ML Model

■ ■ - LLM



Noise Plan



SQL Analytics

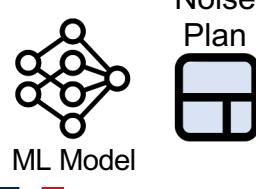
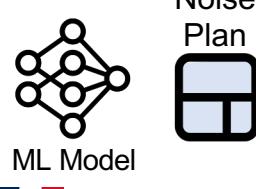
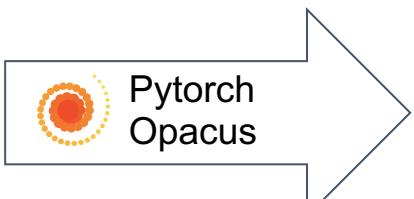
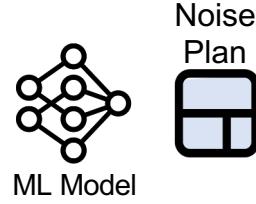
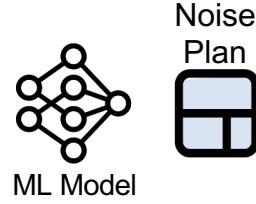
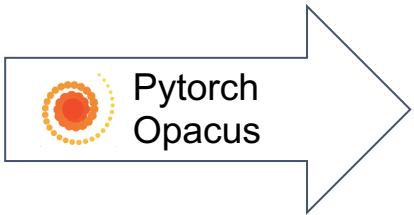
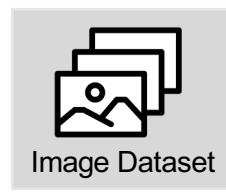
■ ■ - Statistics

Application Layer

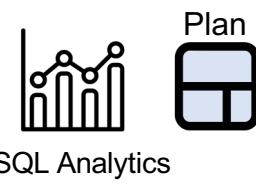
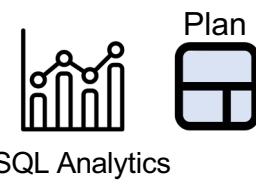
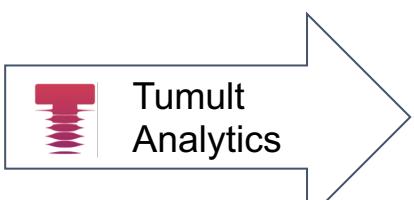
Total: $\max(\epsilon_{\text{■ ■}}, \epsilon_{\text{■ ■ - Statistics}})$



Fine-grained Privacy Analysis

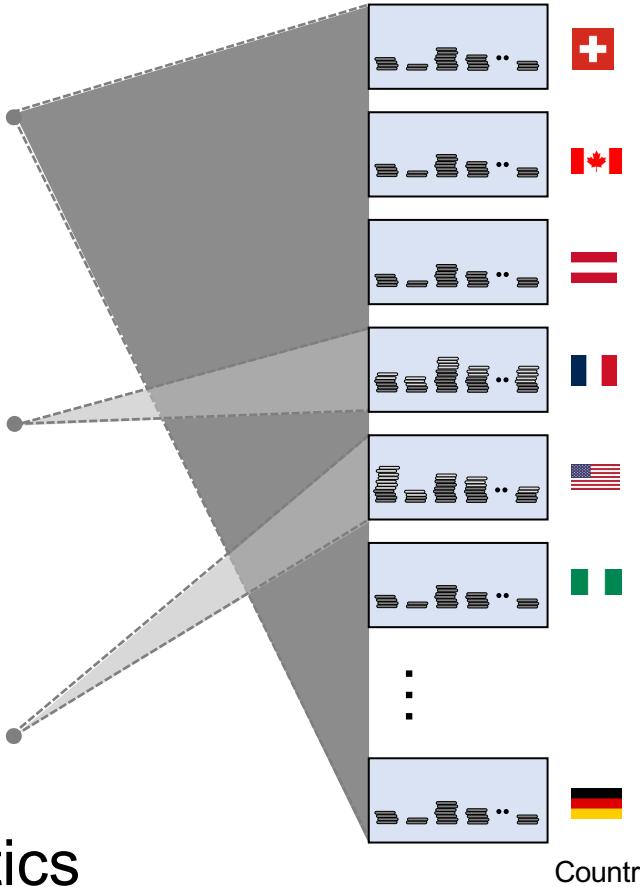


■ - LLM



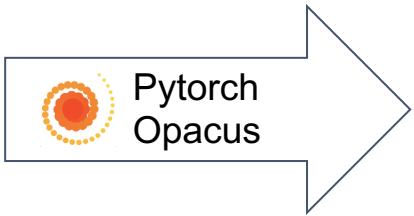
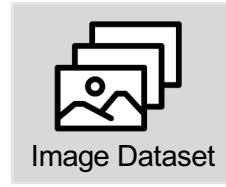
🇺🇸 - Statistics

Application Layer

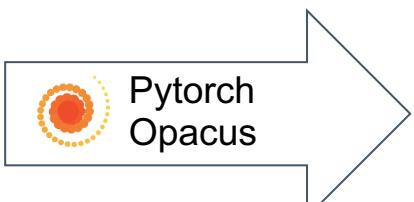


Partitioning Attributes
Schema must be known in advance

Fine-grained Privacy Analysis

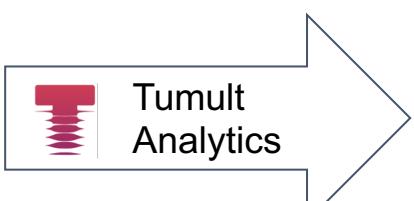


Noise Plan
ML Model



Noise Plan
ML Model

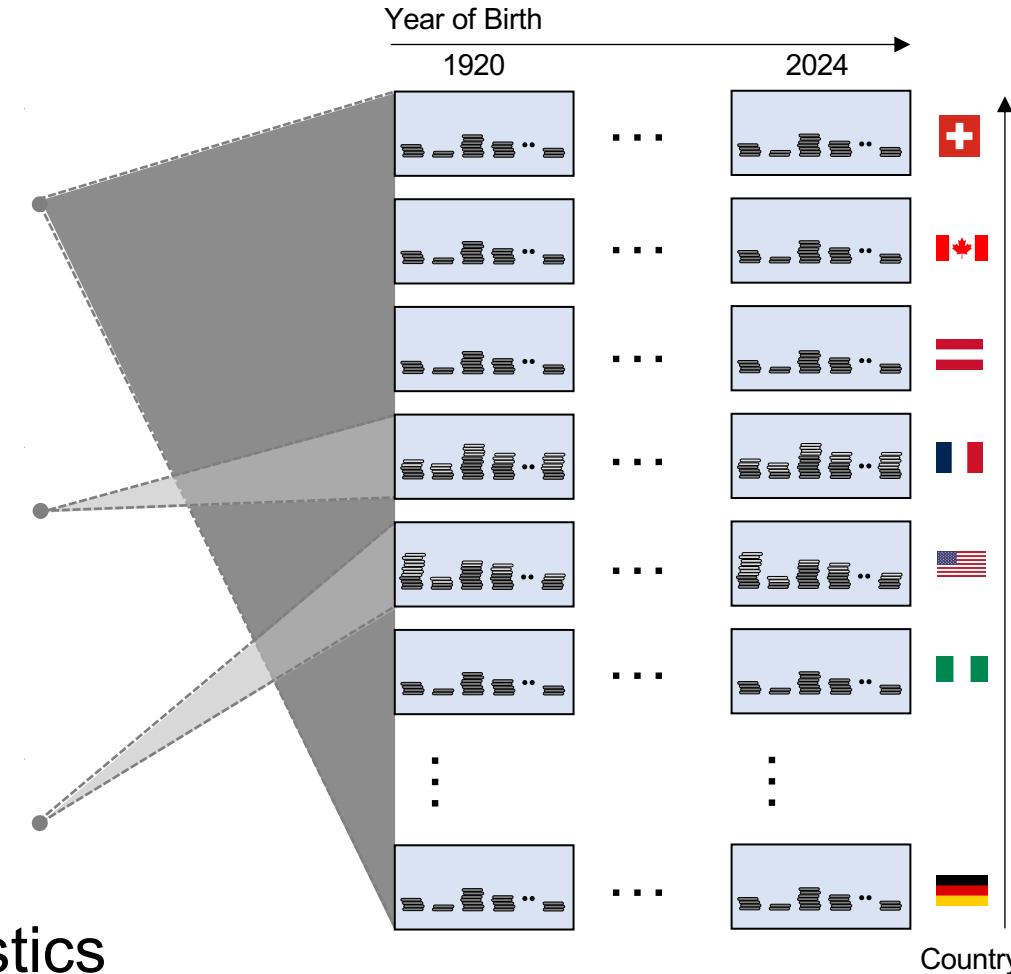
■ - LLM



Noise Plan
SQL Analytics

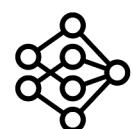
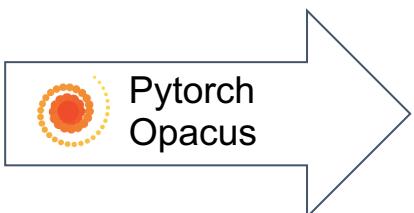
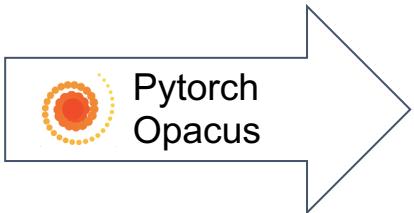
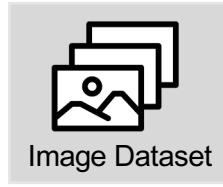
■ - Statistics

Application Layer

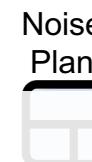


Partitioning Attributes
Schema must be known in advance

Fine-grained Privacy Analysis



■ - LLM



**Fine-grained Privacy Analysis allows
for a tighter Composition.**

Application Layer

Year of Birth

1920

2024



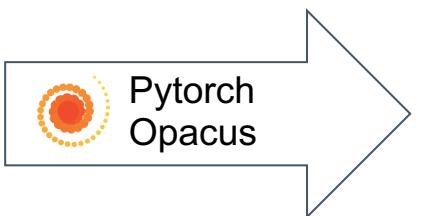
Country

Partitioning Attributes
Schema must be known in advance

Sampling: Random Subset Selection



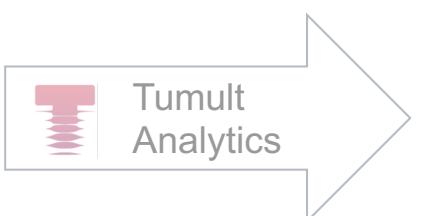
Noise Plan



Noise Plan

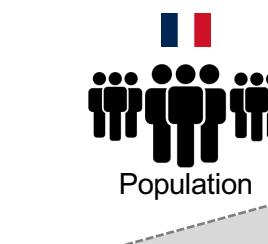
ML Model

— LLM

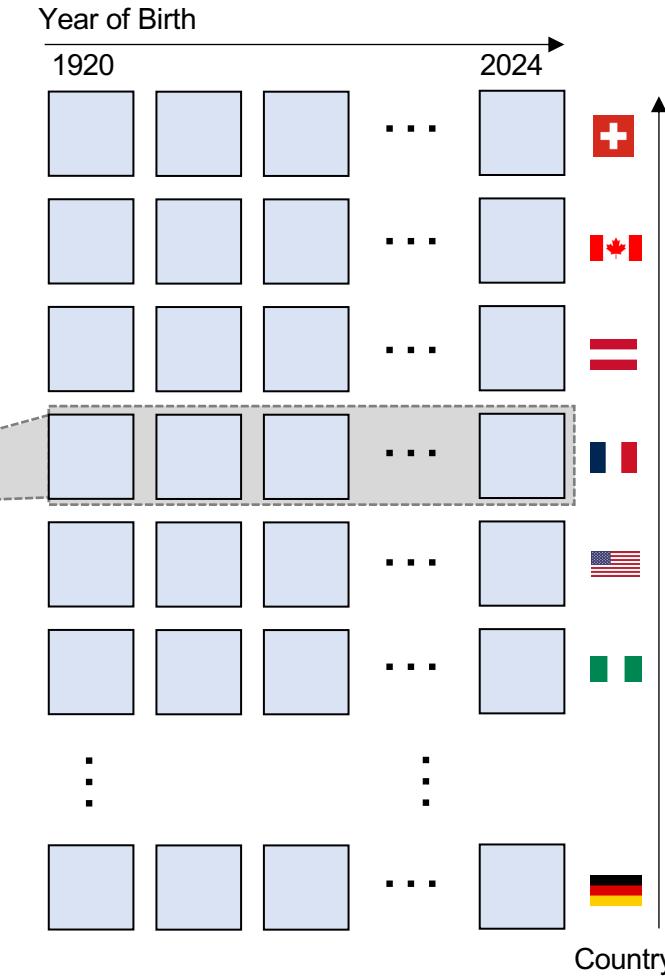


Noise Plan

— Statistics



Population

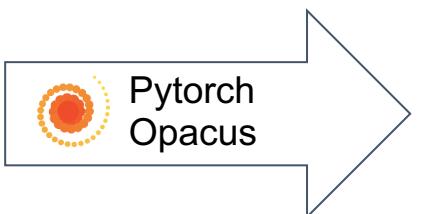


Application Layer

Sampling: Random Subset Selection



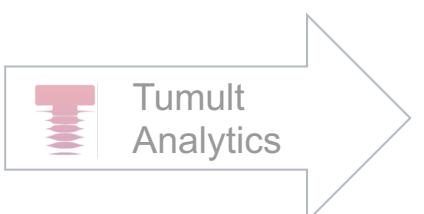
Noise Plan



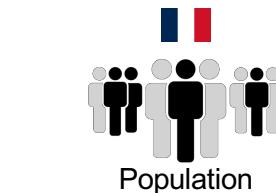
ML Model



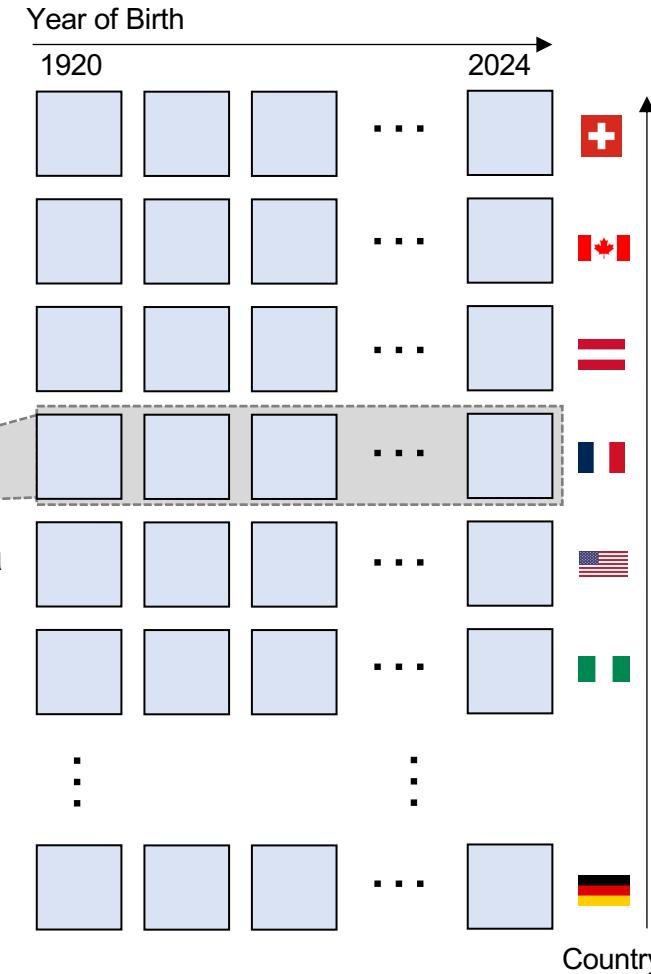
LLM



Statistics

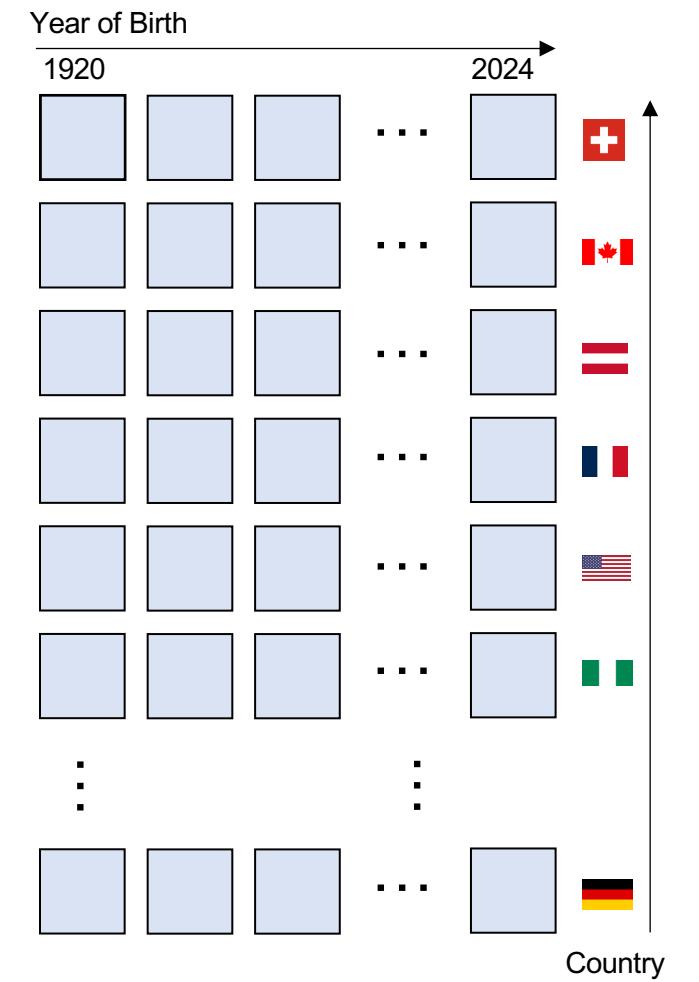
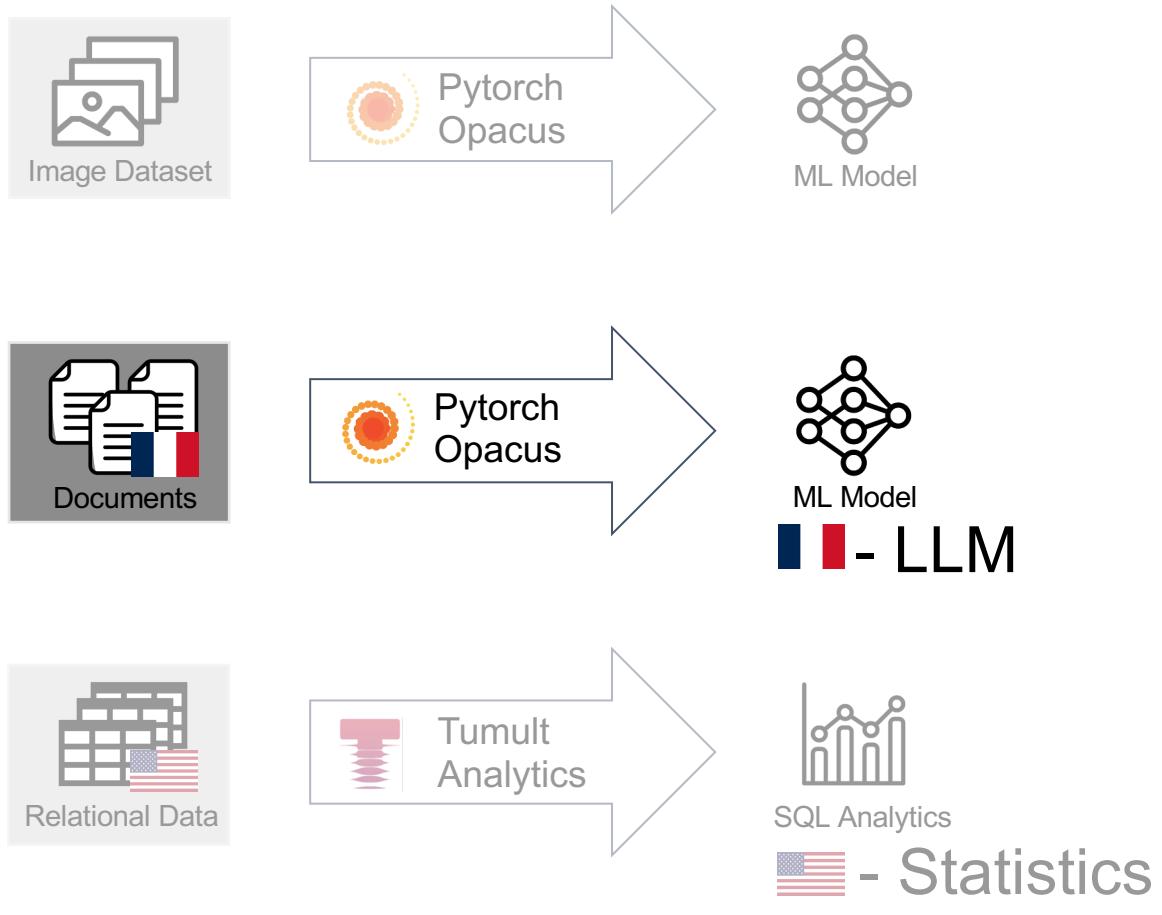


Amplification via
Subsampling



Application Layer

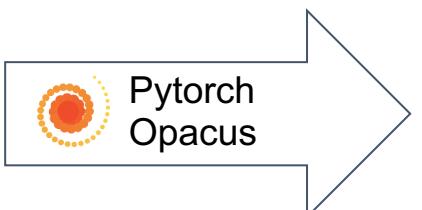
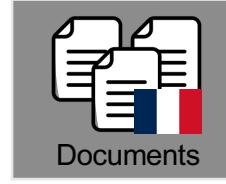
Scarce and Finite Resource



Application Layer

Management Layer

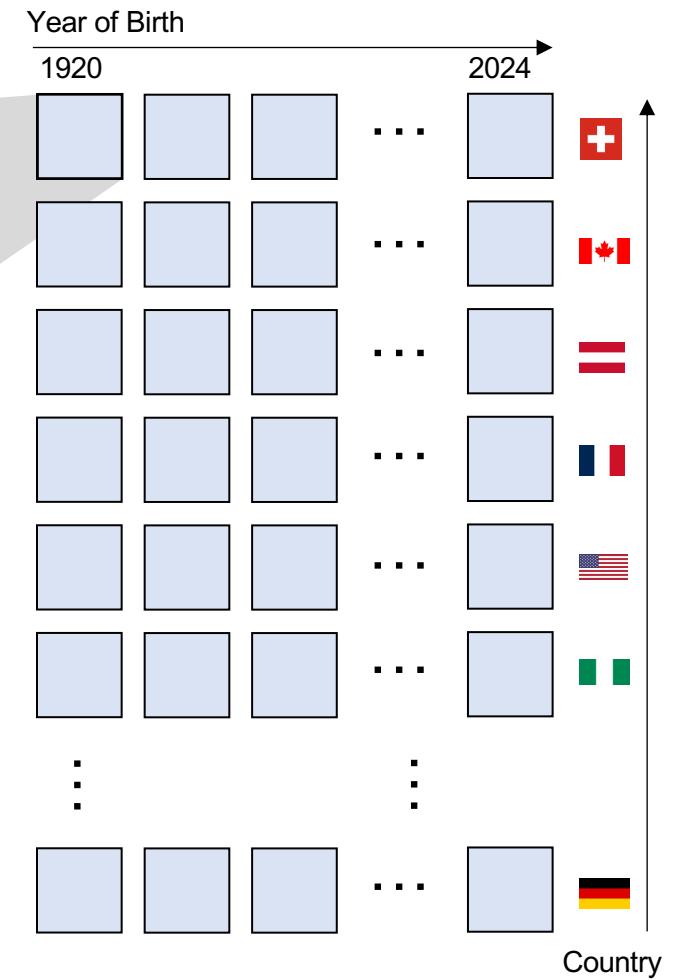
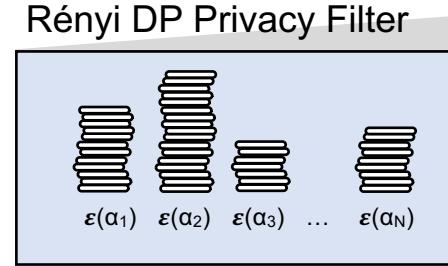
Scarce and Finite Resource



ML Model



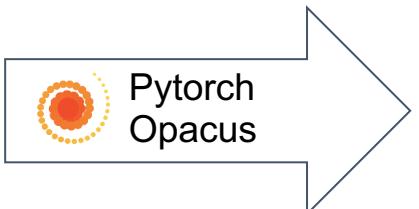
Statistics



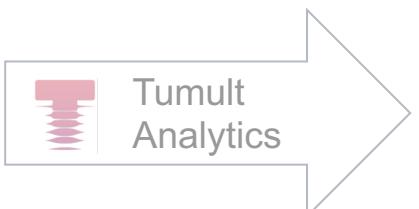
Application Layer

Management Layer

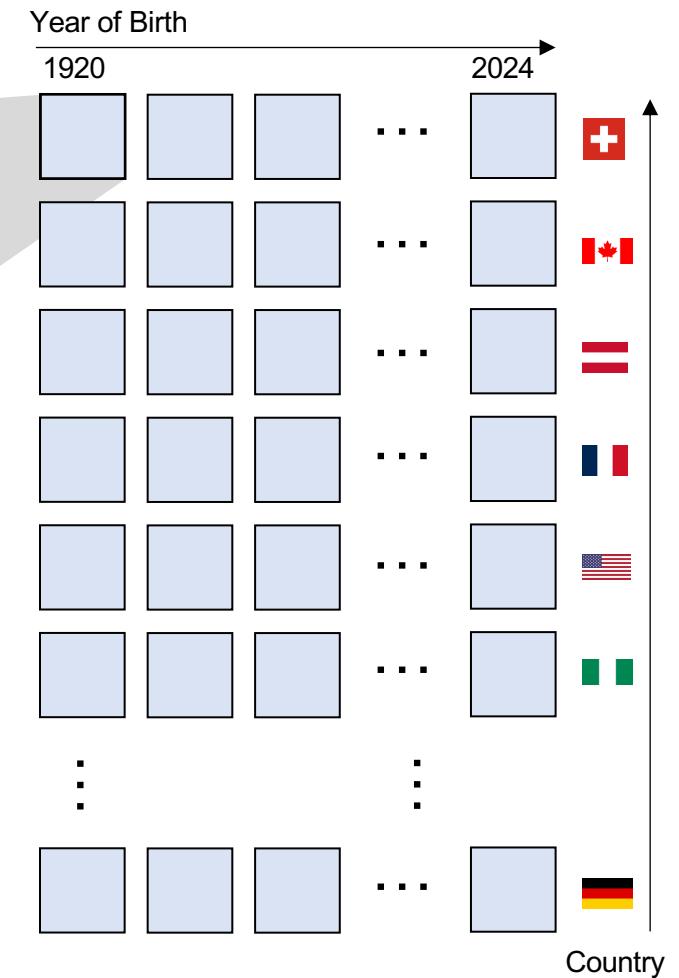
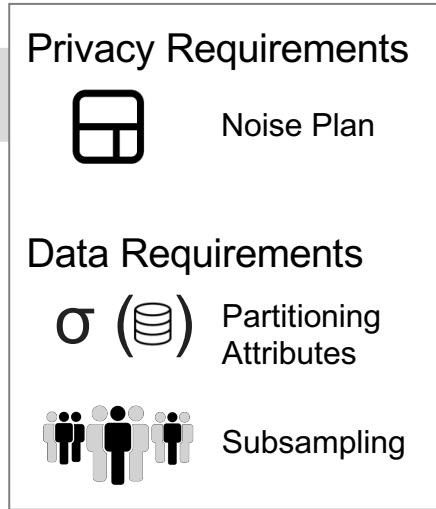
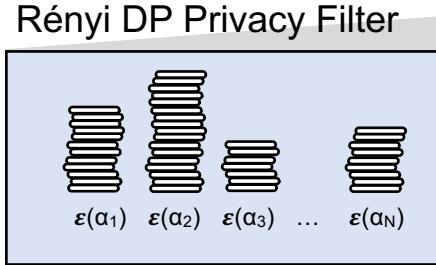
Scarce and Finite Resource



■ - LLM



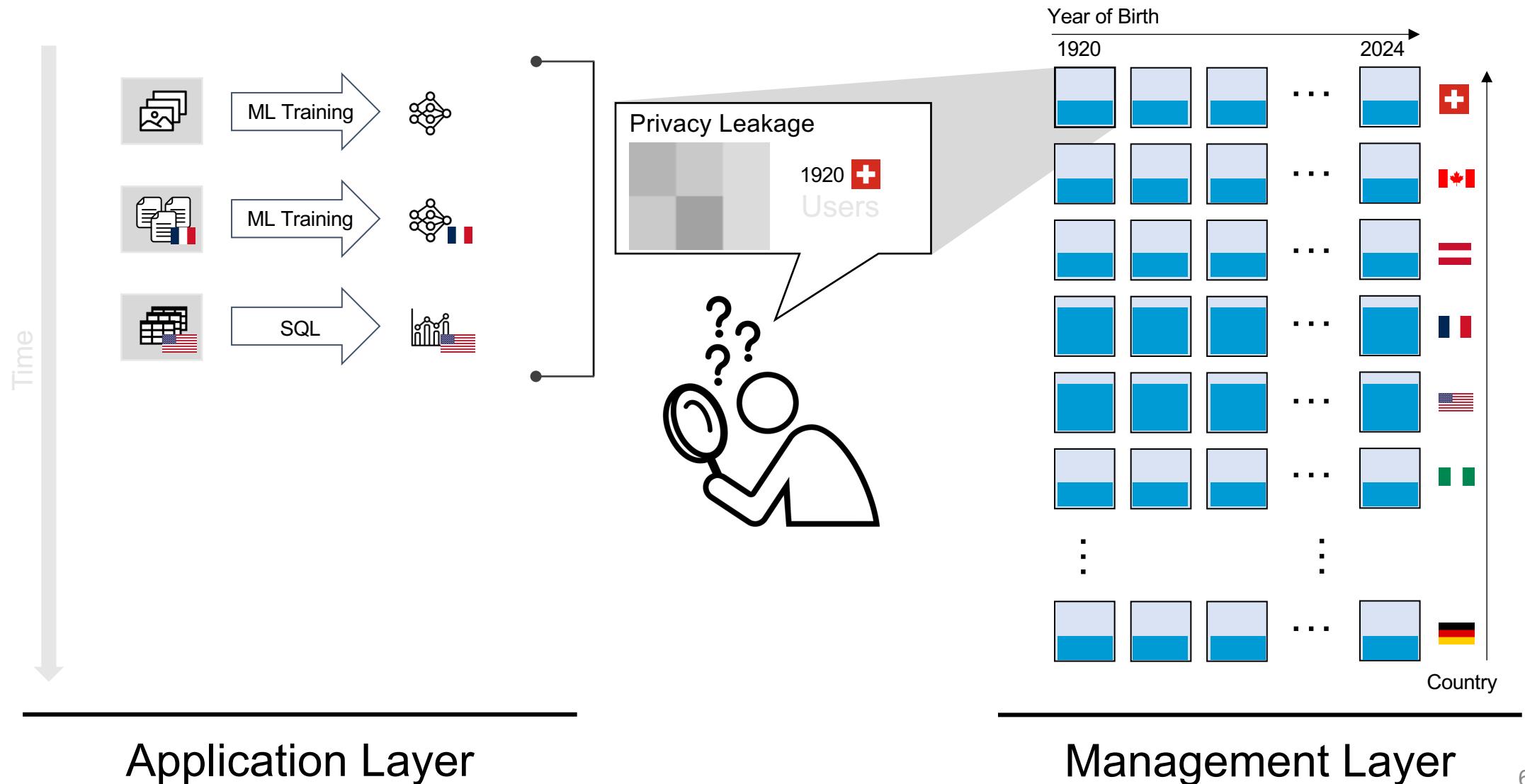
USA - Statistics



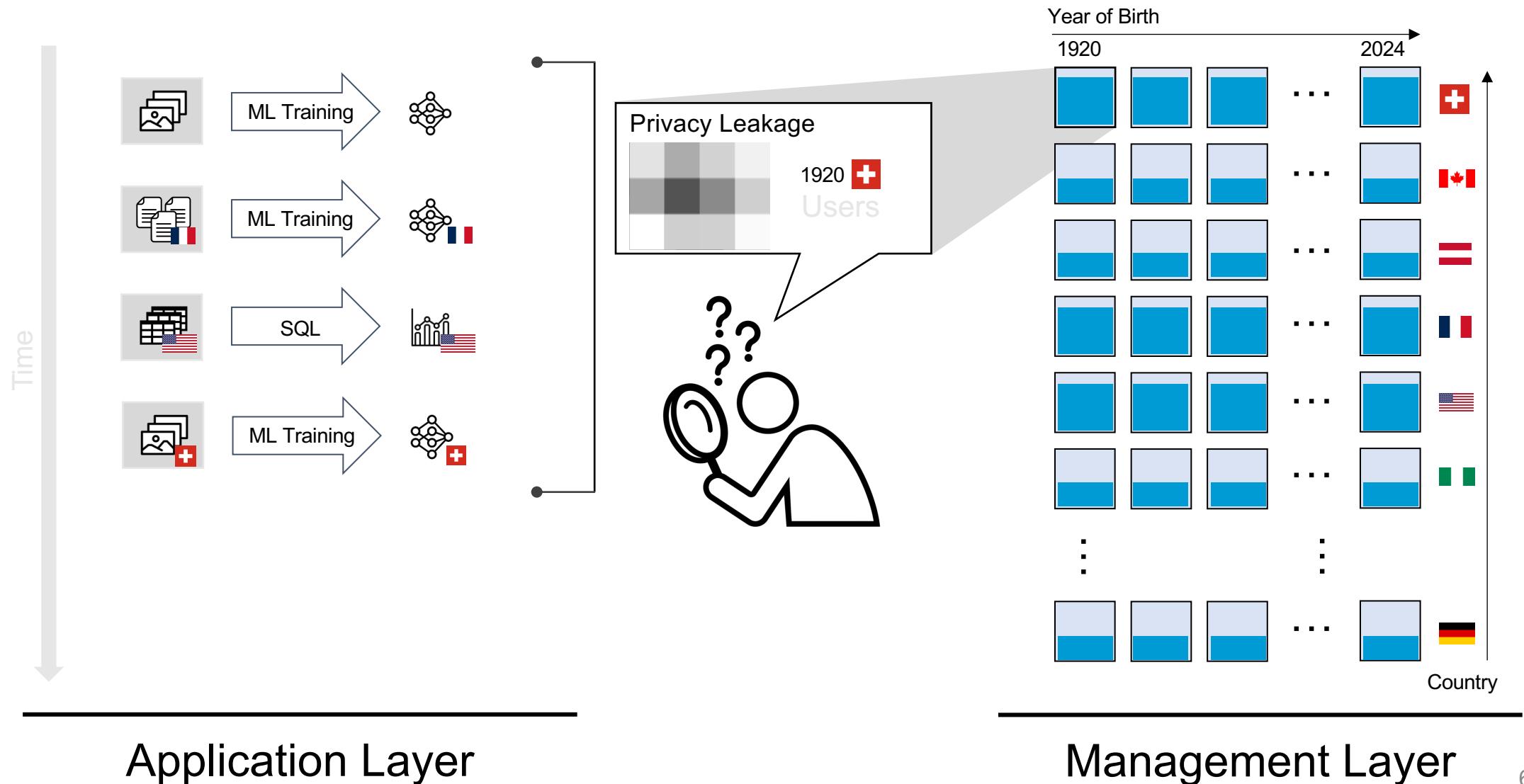
Application Layer

Management Layer

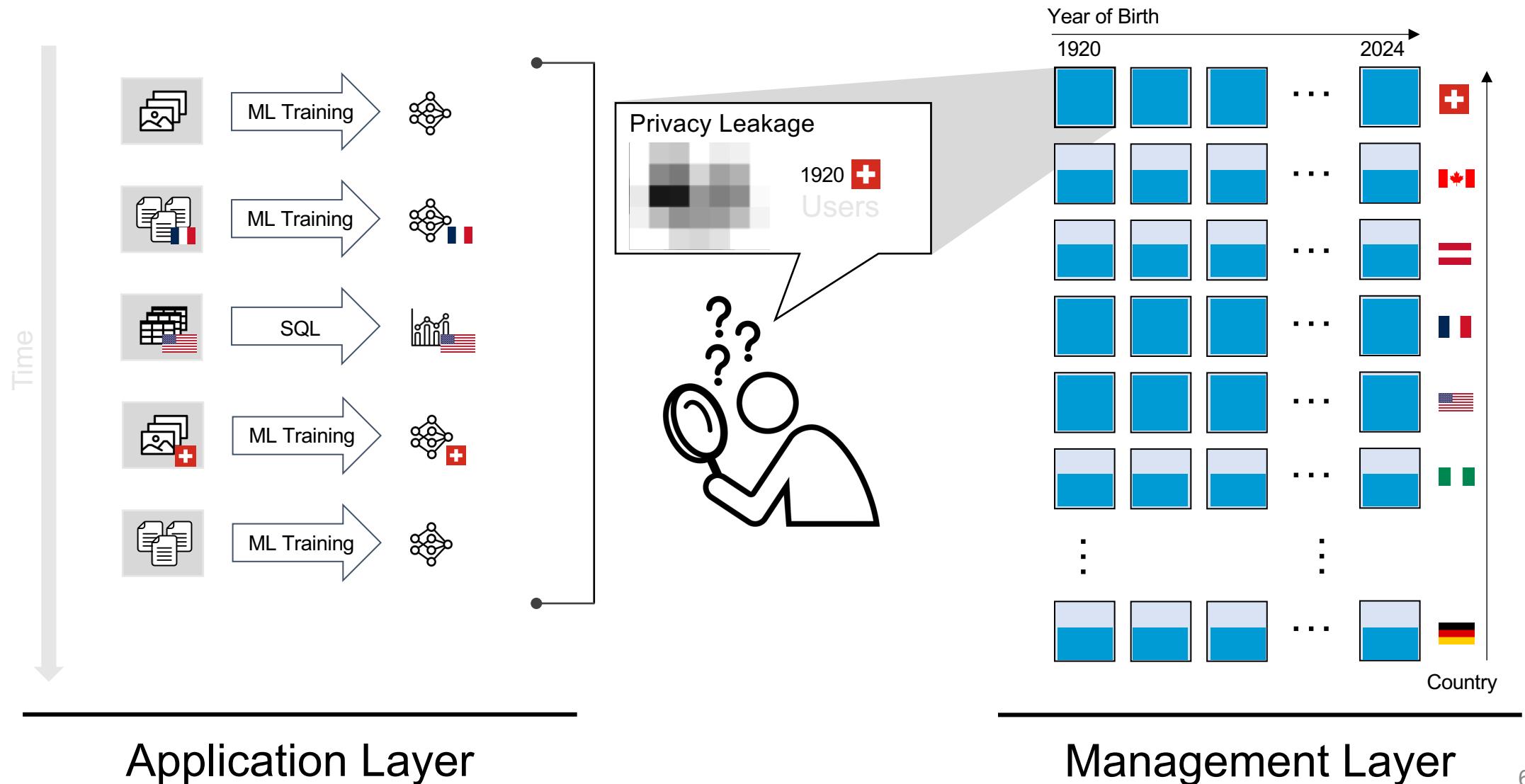
Scarce and Finite Resource



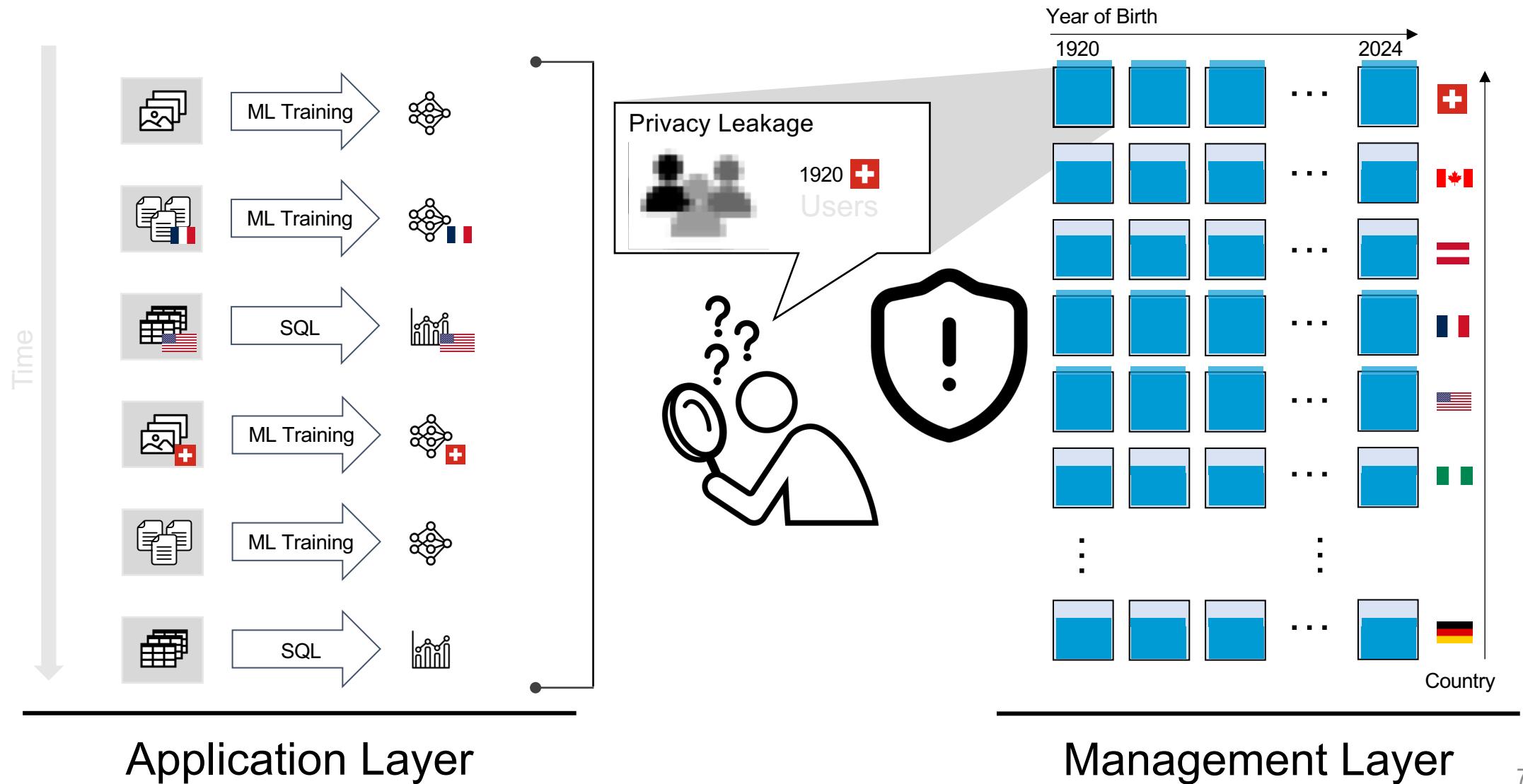
Scarce and Finite Resource



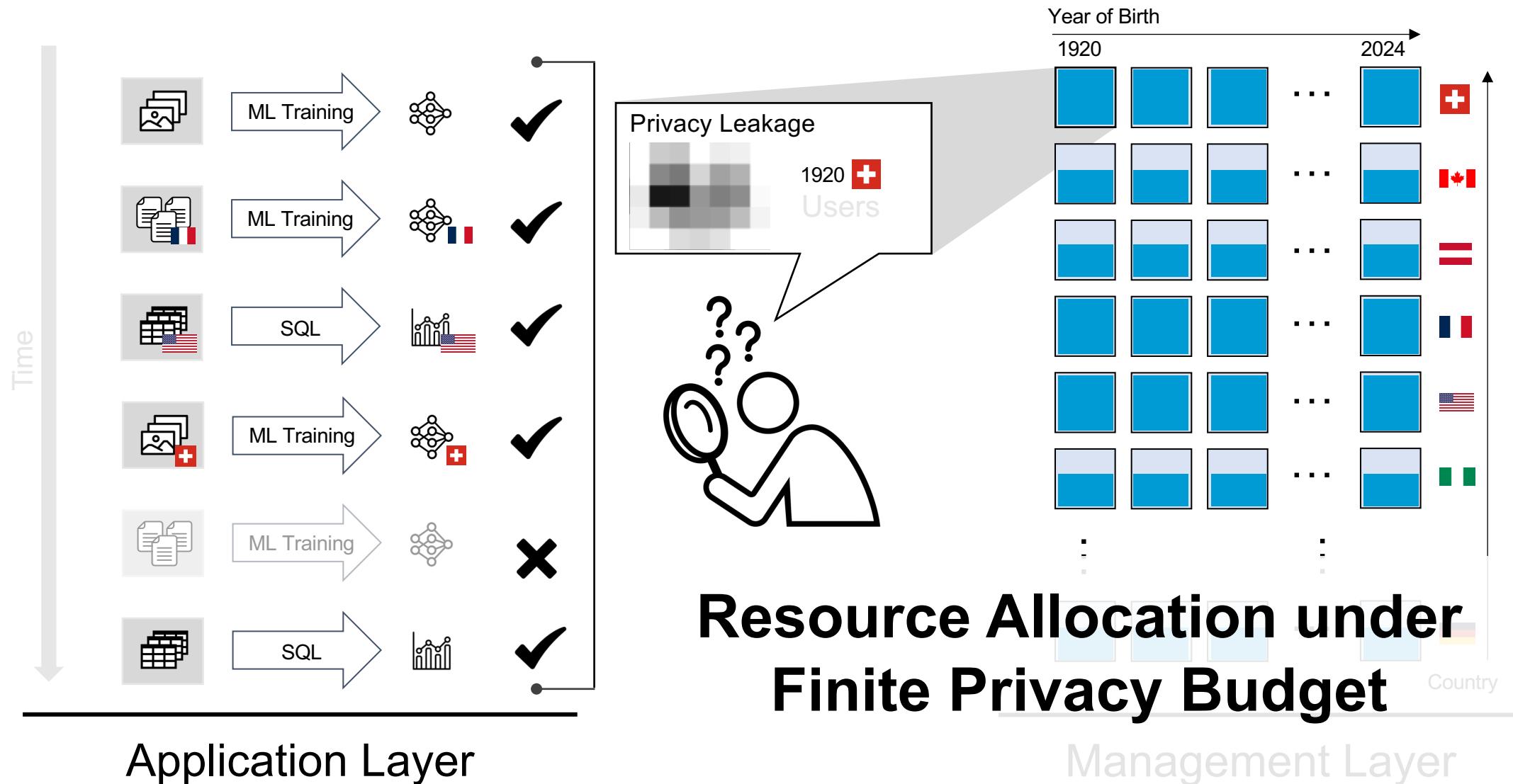
Scarce and Finite Resource



Scarce and Finite Resource



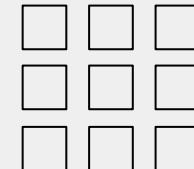
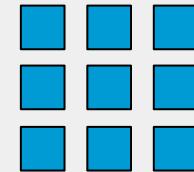
Scarce and Finite Resource



Continuity under a Finite Budget

Ensuring Sustained Budget Allocation Over Time

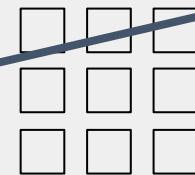
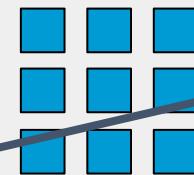
**Resetting
Budget**



Continuity under a Finite Budget

Ensuring Sustained Budget Allocation Over Time

**Resetting
Budget**

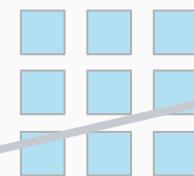


DP Violation

Continuity under a Finite Budget

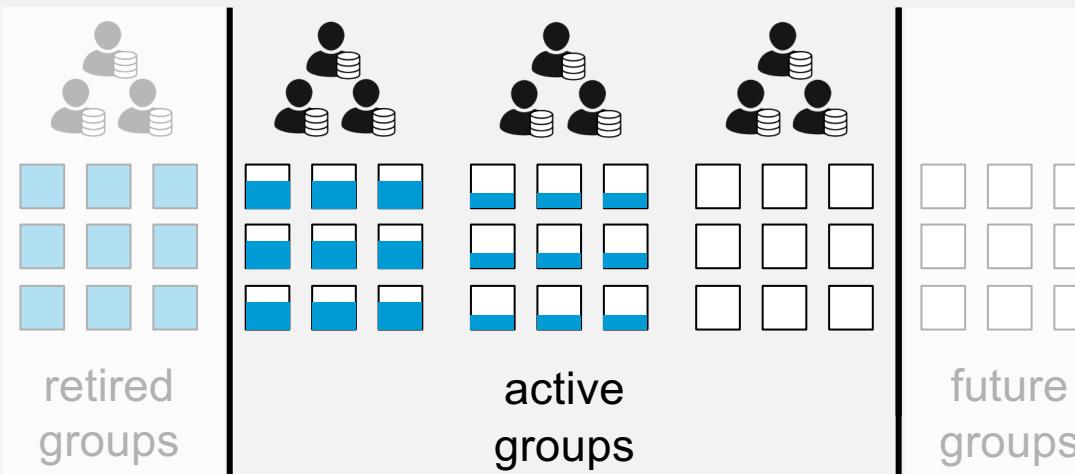
Ensuring Sustained Budget Allocation Over Time

Resetting
Budget



DP Violation

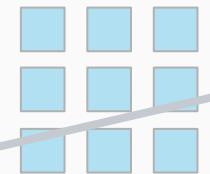
User
Rotation



Continuity under a Finite Budget

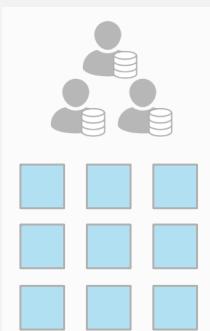
Ensuring Sustained Budget Allocation Over Time

Resetting
Budget

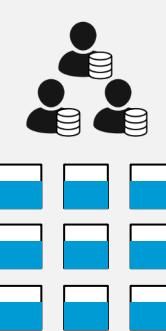


DP Violation

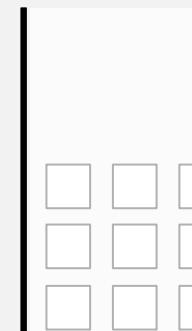
User
Rotation



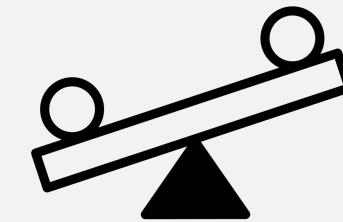
retired
groups



active
groups



future
groups

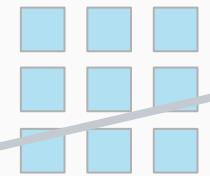


Biased Set of
Active Users

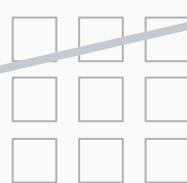
Continuity under a Finite Budget

Ensuring Sustained Budget Allocation Over Time

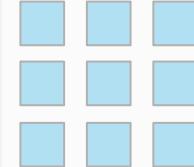
Resetting
Budget



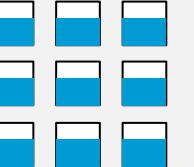
DP Violation



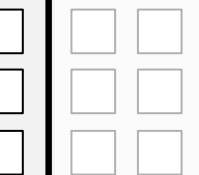
User
Rotation



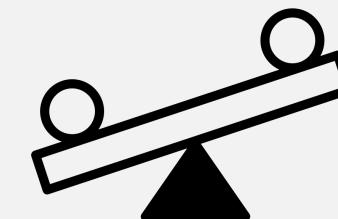
retired
groups



active
groups



future
groups



Biased Set of
Active Users

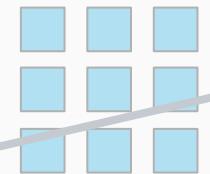


Budget Guarantees
with Unlocking

Continuity under a Finite Budget

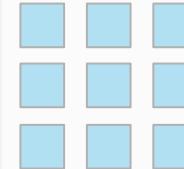
Ensuring Sustained Budget Allocation Over Time

Resetting
Budget

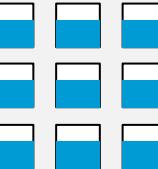


DP Violation

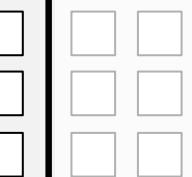
User
Rotation



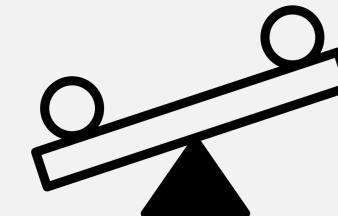
retired
groups



active
groups



future
groups



Biased Set of
Active Users

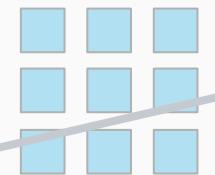


Budget Guarantees
with Unlocking

Continuity under a Finite Budget

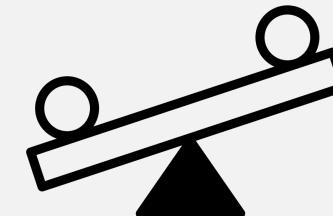
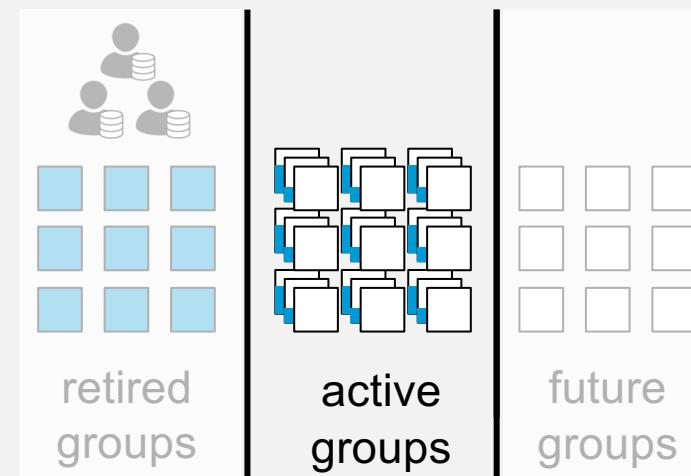
Ensuring Sustained Budget Allocation Over Time

Resetting
Budget



DP Violation

User
Rotation

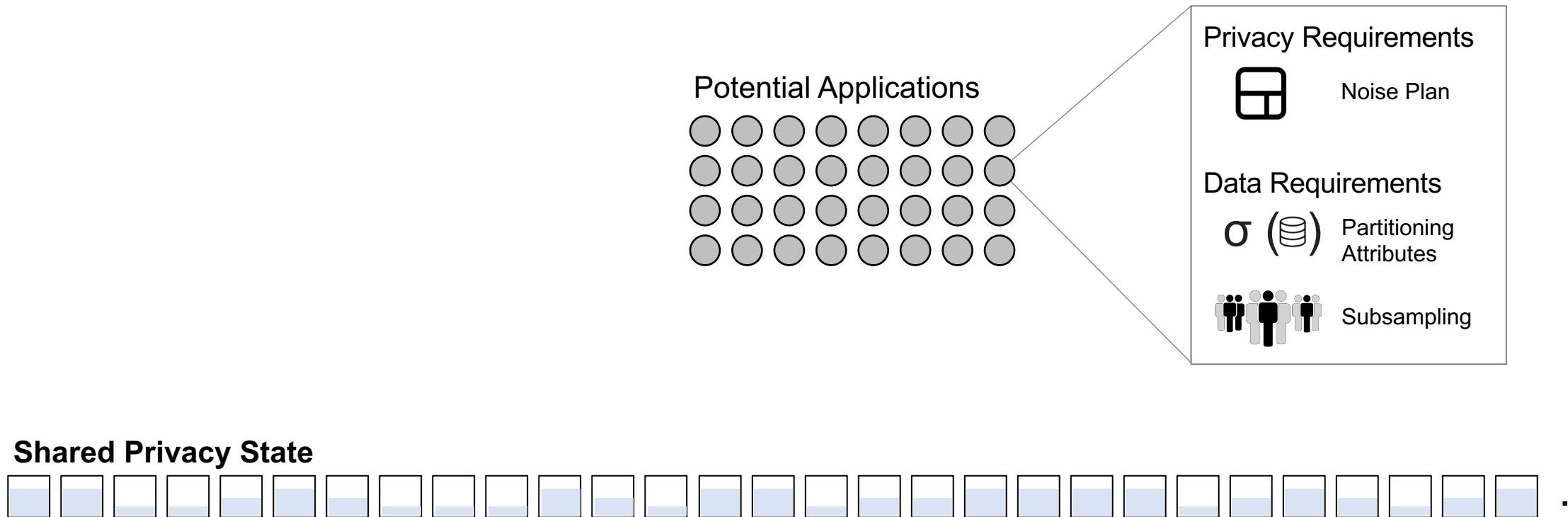


Biased Set of
Active Users

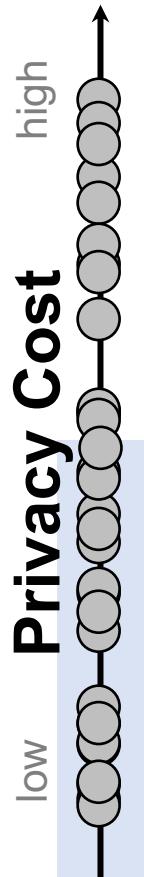


Budget Guarantees
with Unlocking

Privacy Resource Allocation

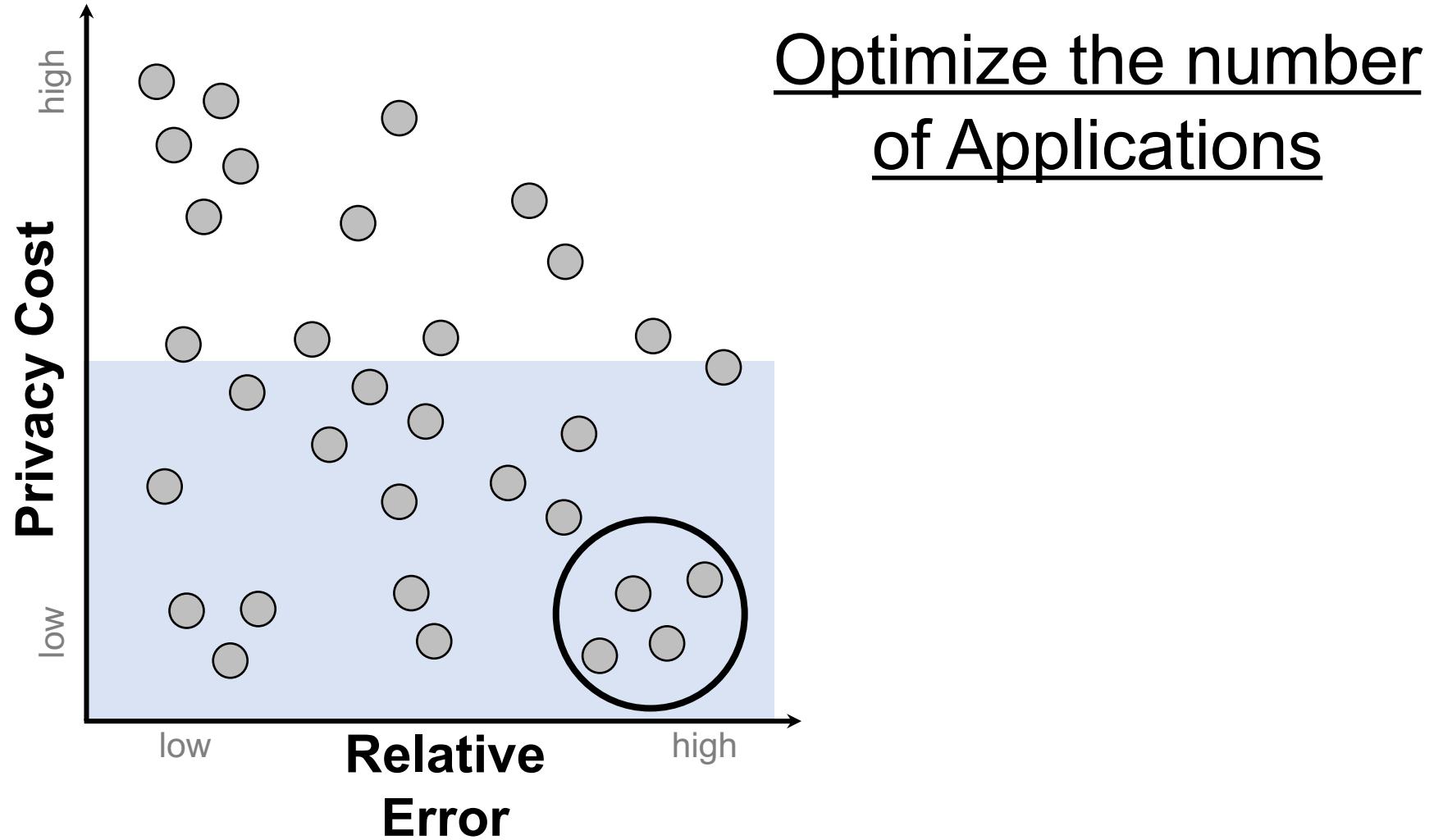


Privacy Resource Allocation

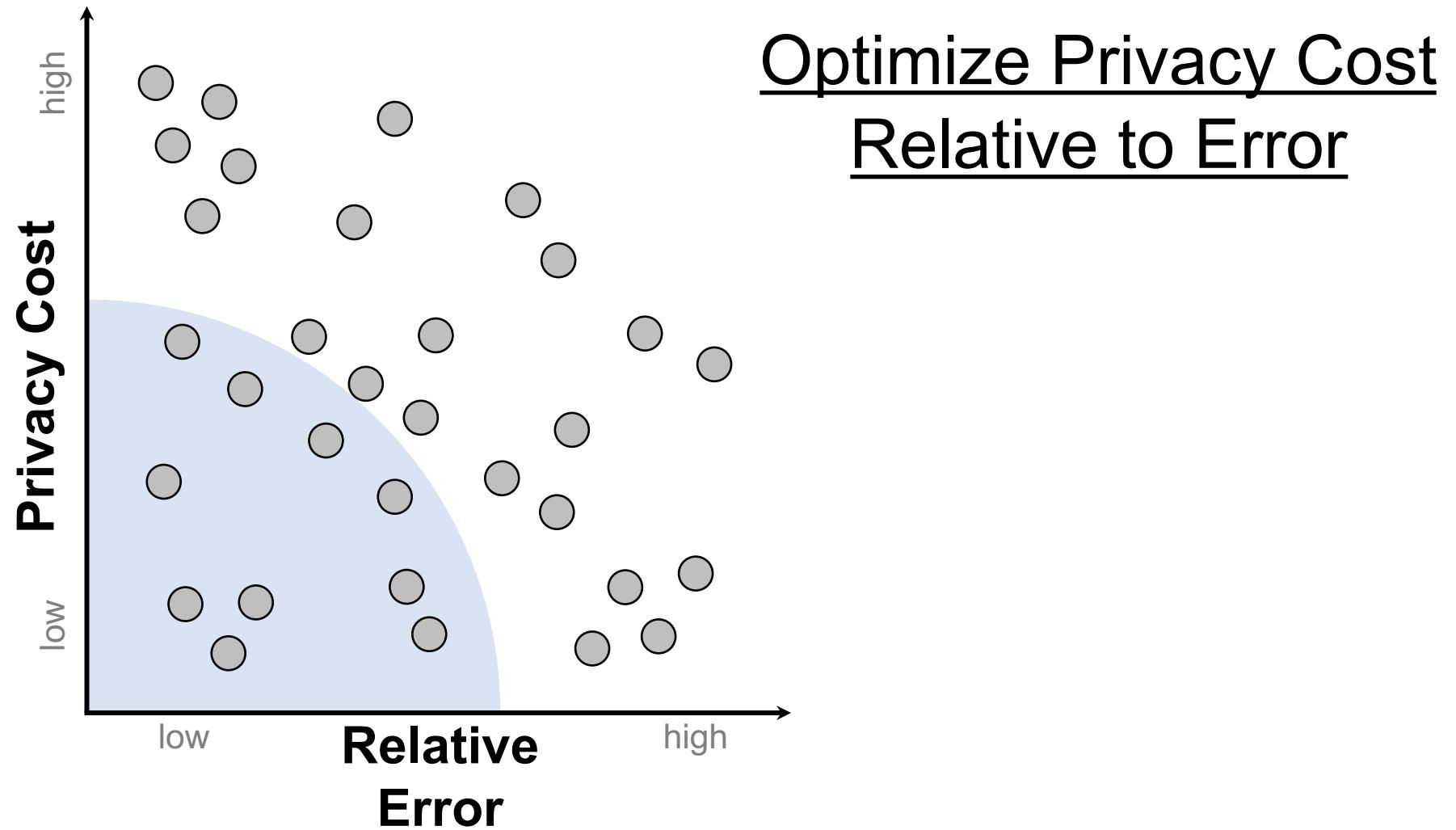


Optimize the number
of Applications

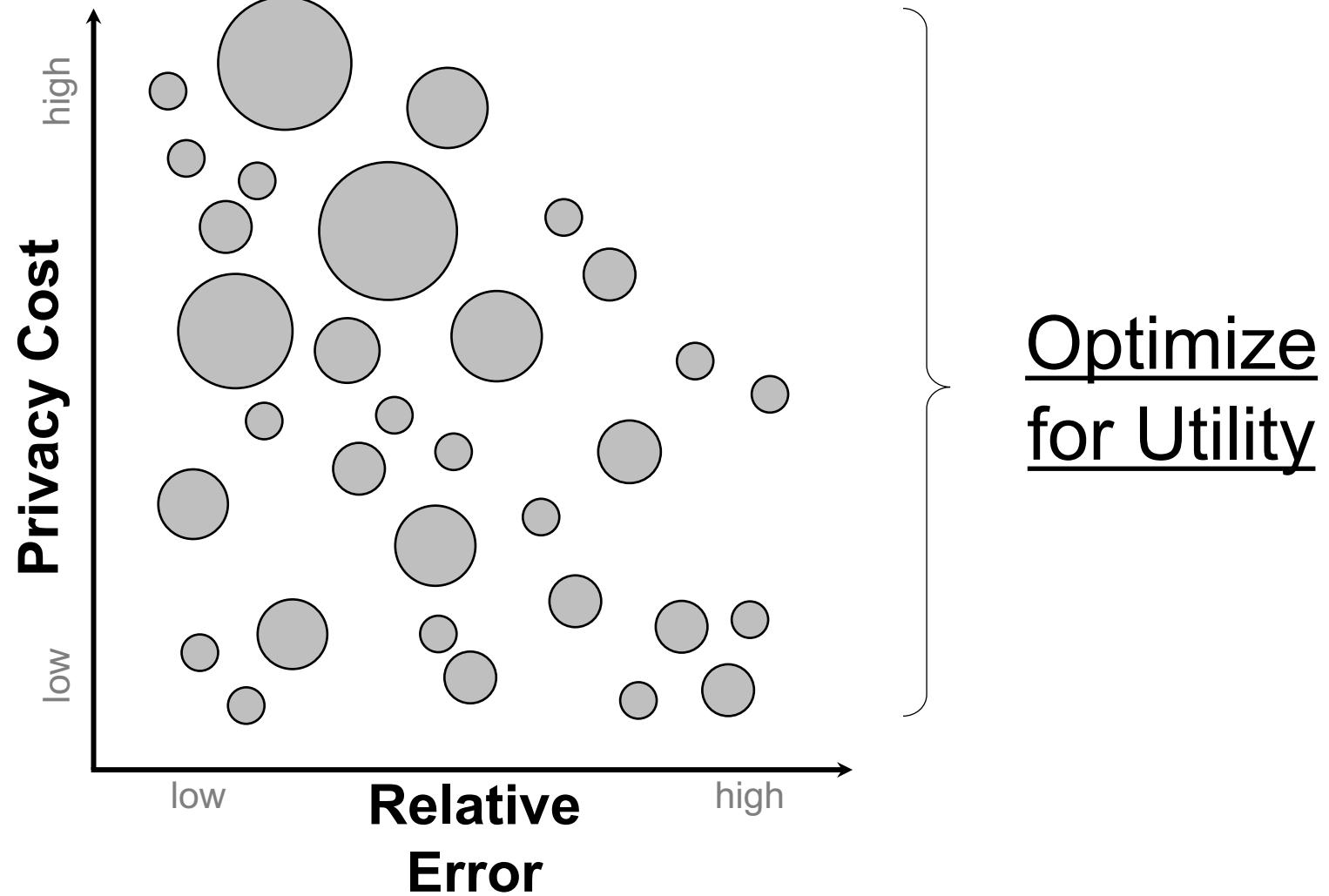
Privacy Resource Allocation



Privacy Resource Allocation

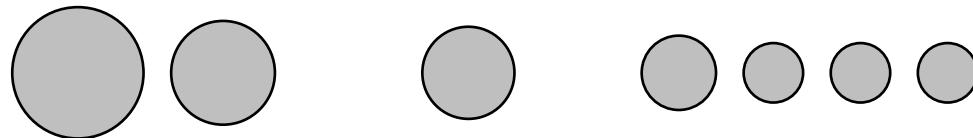


Privacy Resource Allocation



Privacy Resource Allocation

Potential Applications

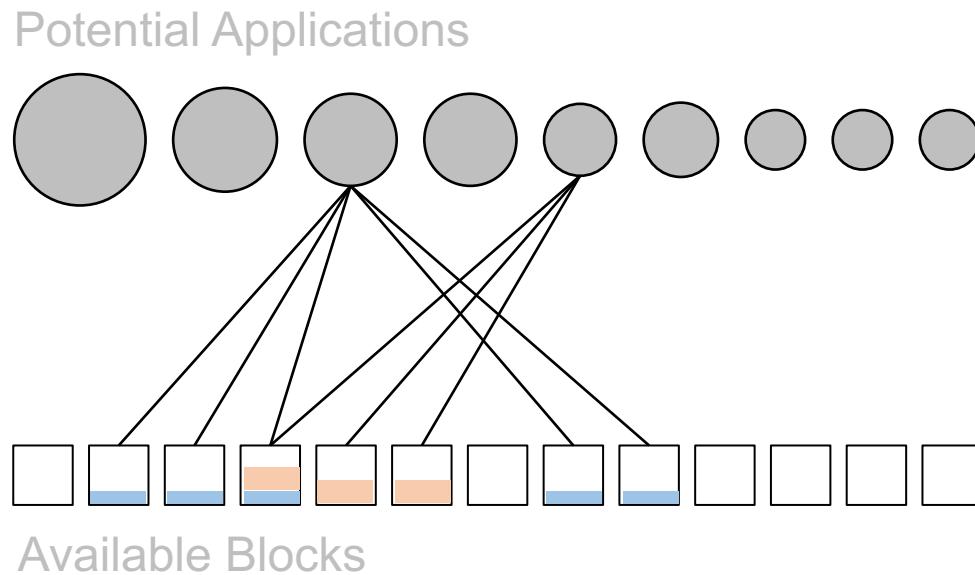


Objective:

$$\max \sum_{i \in Apps} Utility_i * y_i$$

$y_i = 1$ if application i is allocated, else 0

Privacy Resource Allocation



Multidimensional Knapsack Problem

Objective:

$$\max \sum_{i \in Apps} Utility_i * y_i$$

$y_i = 1$ if application i is allocated, else 0

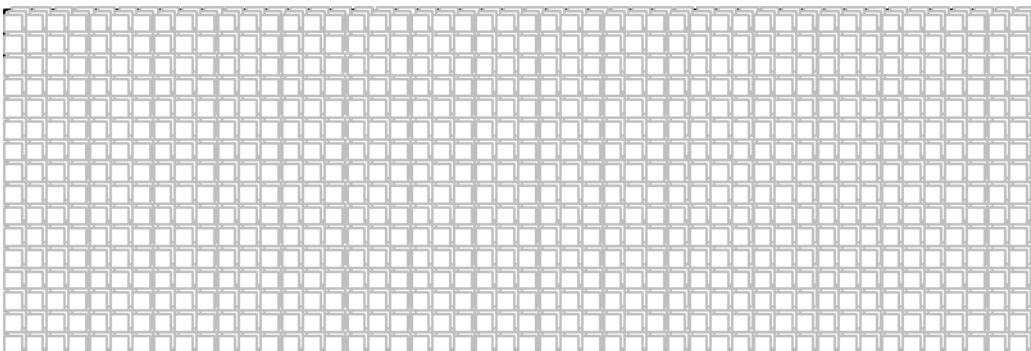
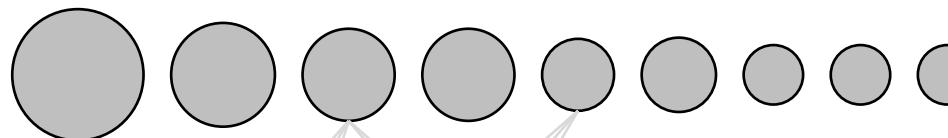
Budget Constraints:

$$s.t. \sum_{i \in Apps} \varepsilon_{ij} * y_i \leq Budget_j \quad \forall j \in Blocks$$

Privacy cost of application i for block j
* for simplicity we show the cost in ε -DP rather than RDP

Privacy Resource Allocation

Potential Applications



Available Blocks

Multidimensional Knapsack Problem

Objective:

$$\max \sum_{i \in Apps} Utility_i * y_i$$

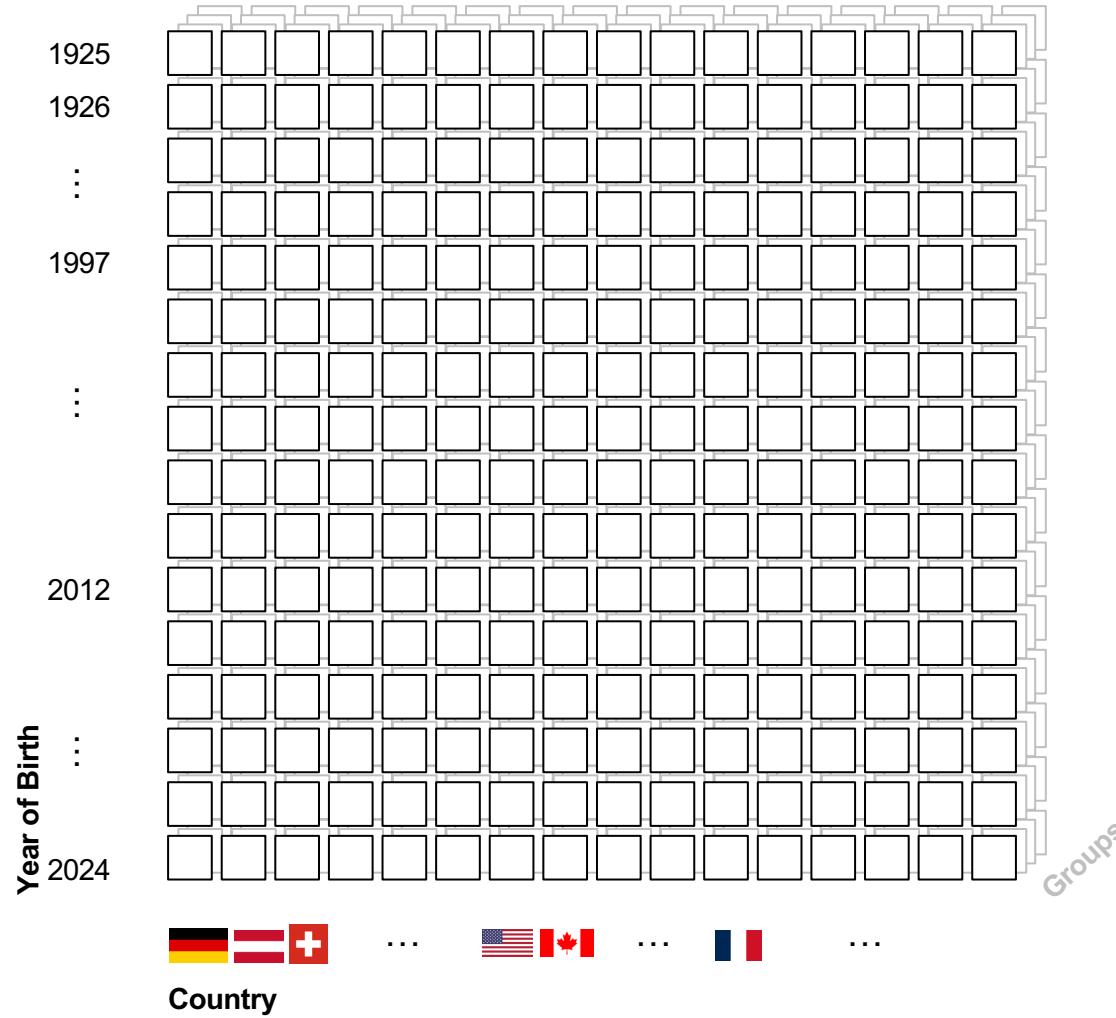
$y_i = 1$ if application i is allocated, else 0

Budget Constraints:

$$s.t. \sum_{i \in Apps} \epsilon_{ij} * y_i \leq Budget_j \quad \forall j \in Blocks$$

Privacy cost of application i for block j
* for simplicity we show the cost in ϵ -DP rather than RDP

Resource Allocation: Taming the Complexity



Resource Allocation: Taming the Complexity

Request 1



1925
1926

⋮

1997

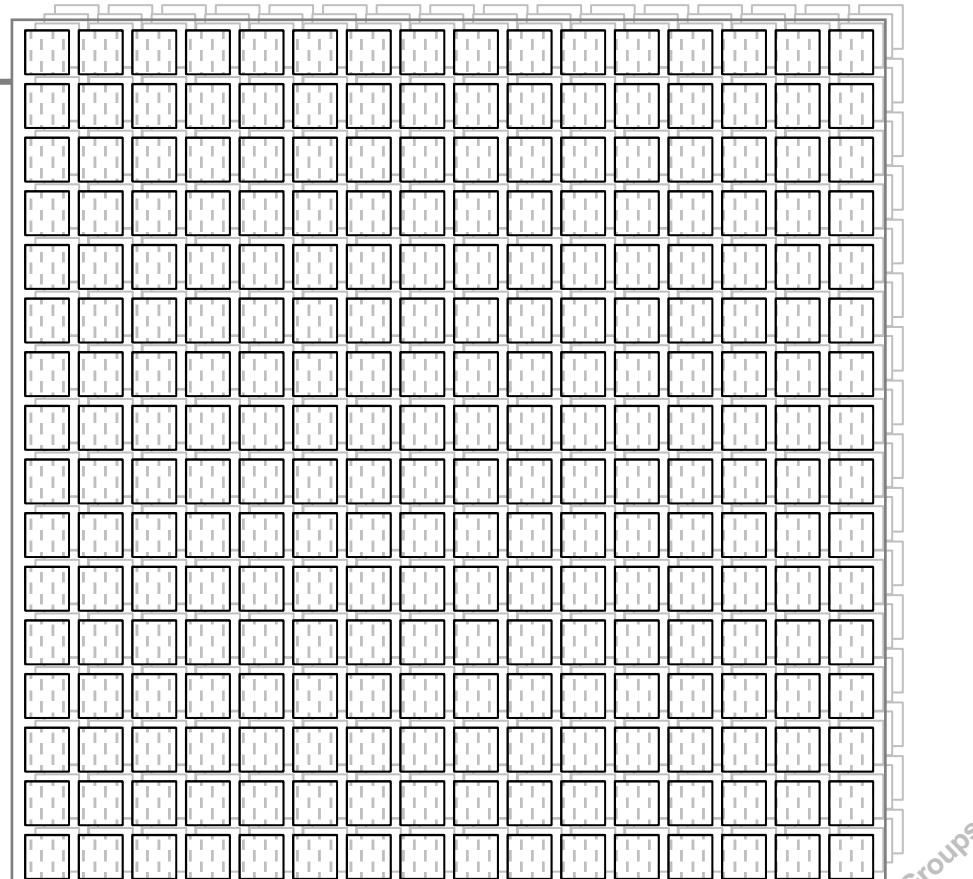
⋮

2012

⋮

2024

Year of Birth



Groups



...



...



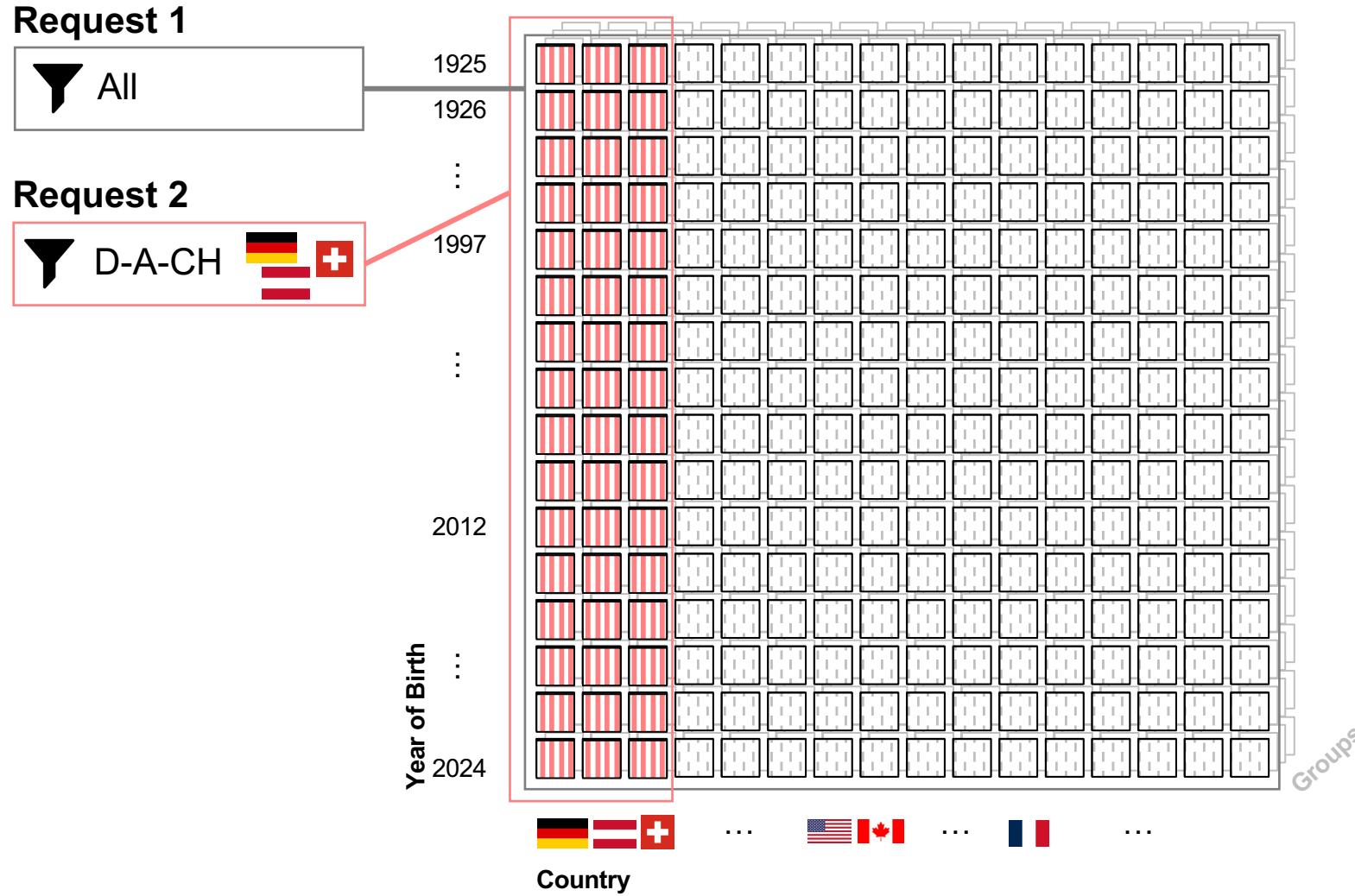
...



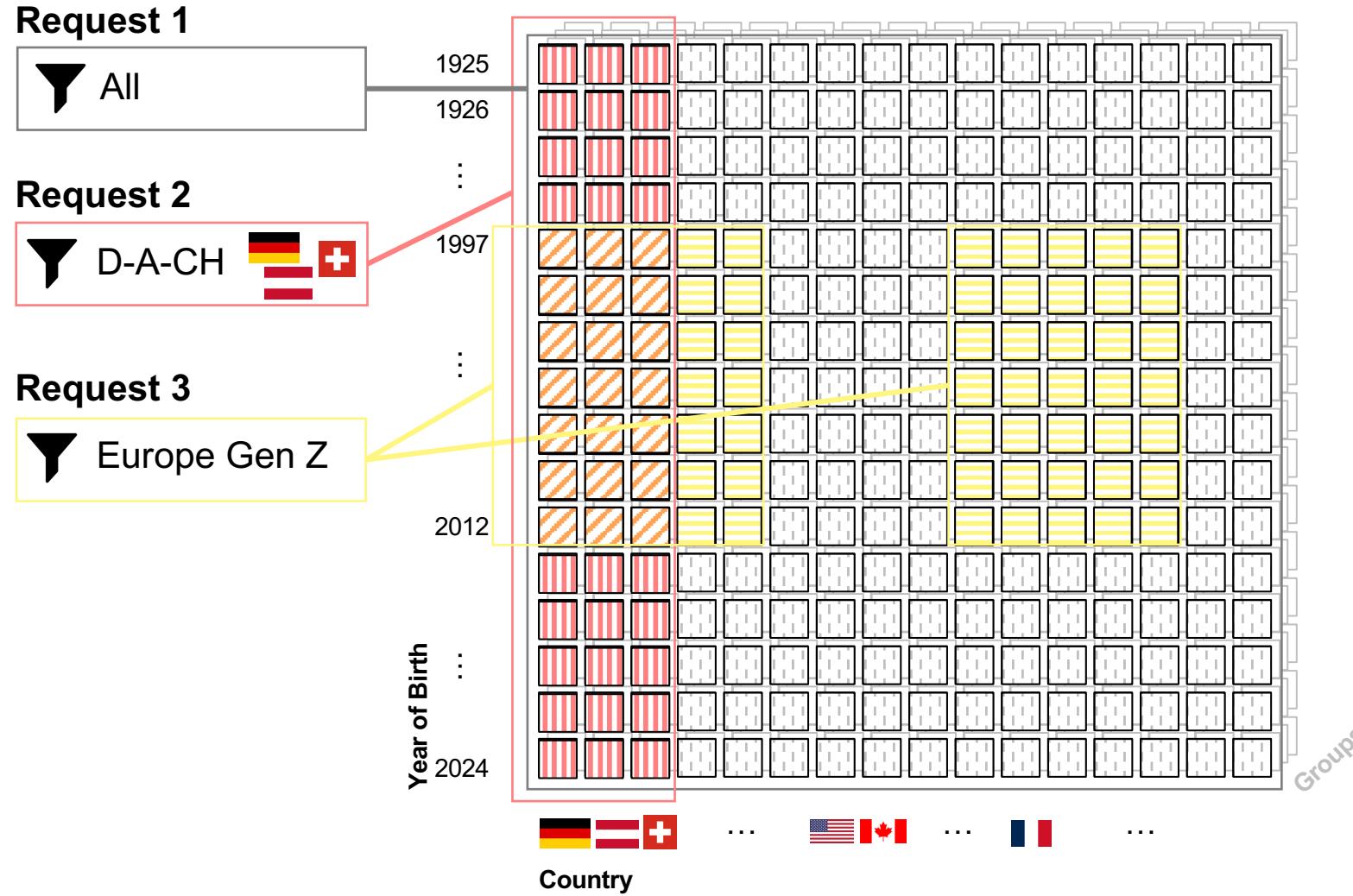
...

Country

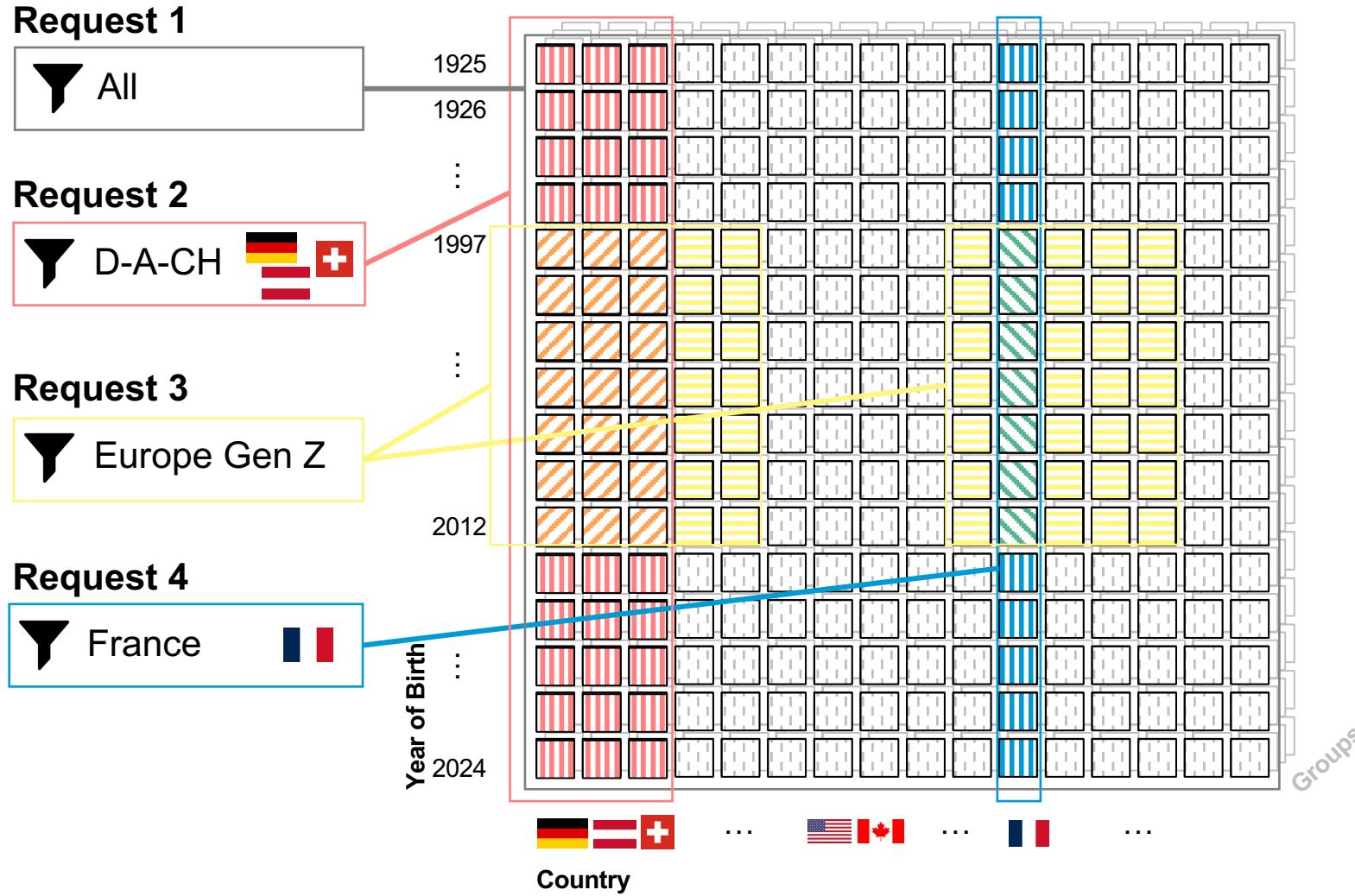
Resource Allocation: Taming the Complexity



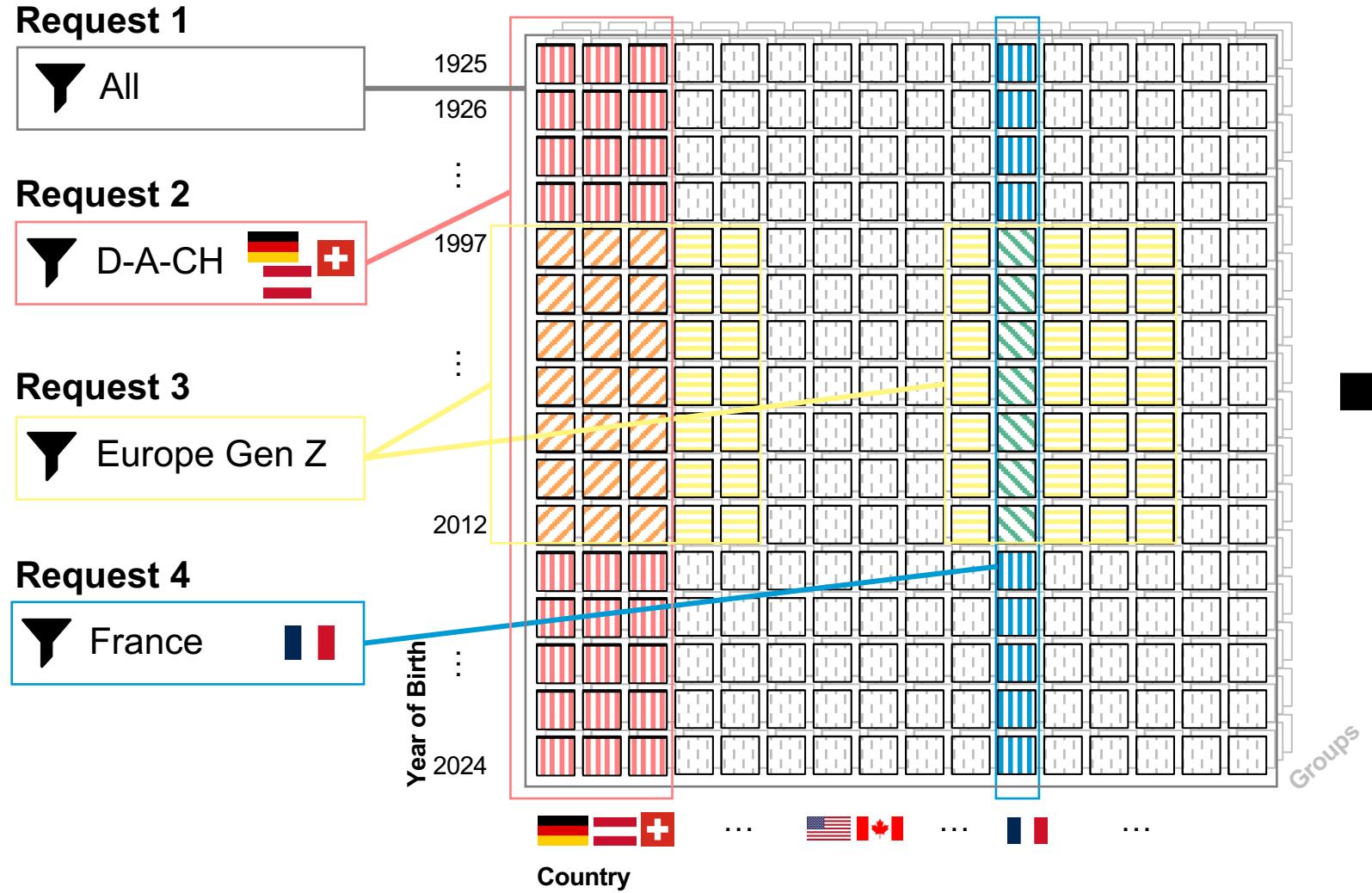
Resource Allocation: Taming the Complexity



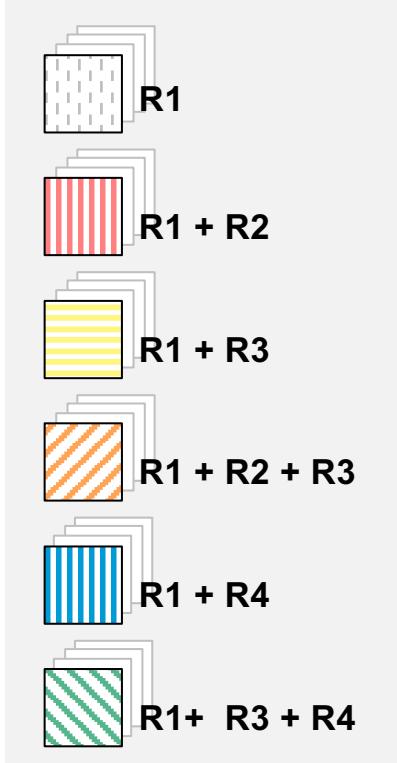
Resource Allocation: Taming the Complexity



Resource Allocation: Taming the Complexity



Contending Costs



Resource Allocation: Taming the Complexity

Request 1



Request 2



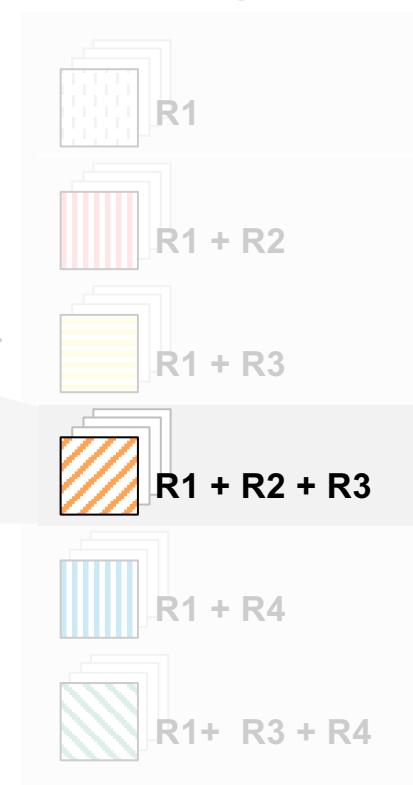
Request 3



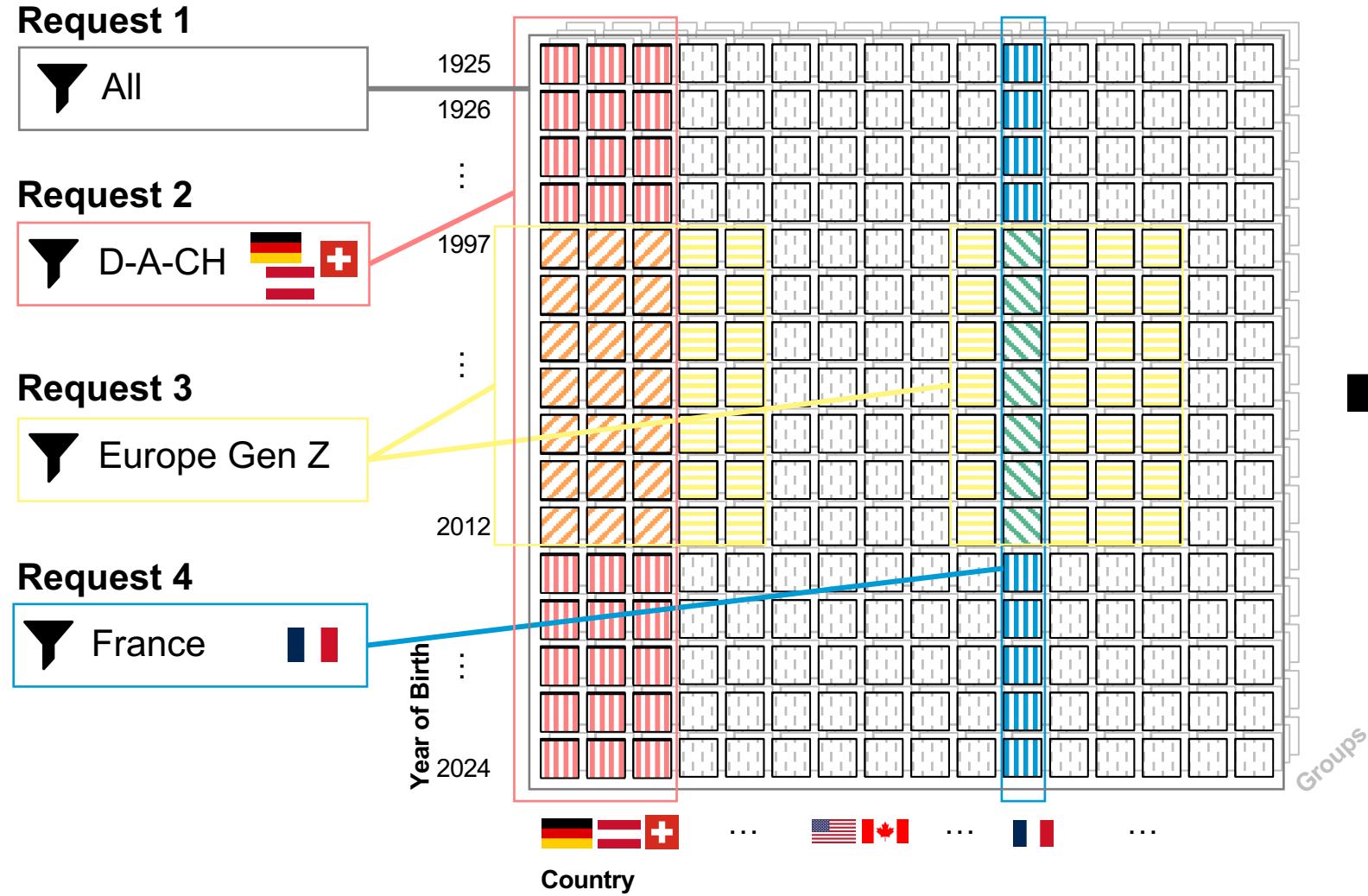
Request 4



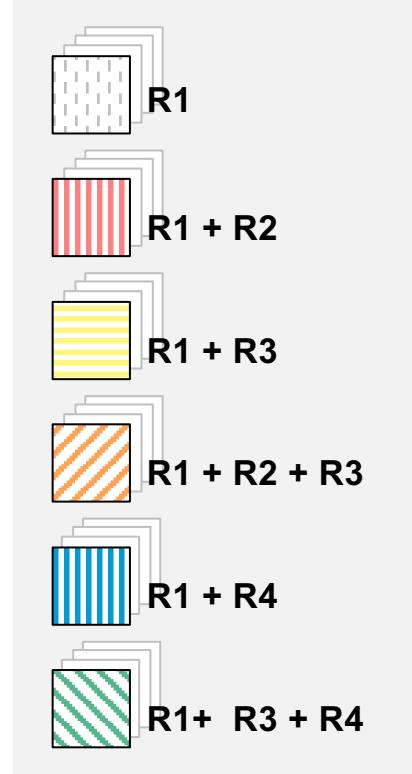
Contending Costs



Resource Allocation: Taming the Complexity



Contending Costs



Resource Allocation: Taming the Complexity

Request 1



All

1925

1926

⋮

Request 2



D-A-CH



1997

⋮

Request 3



Europe Gen Z

2012

⋮

Request 4



France



Year of Birth

⋮

2024

⋮



⋮



⋮



⋮

Groups



Application History
L

Budgets

Contending Costs



R1



R1 + R2



R1 + R3



R1 + R2 + R3



R1 + R4



R1 + R3 + R4

Dimensionality Reduction

Resource Allocation: Taming the Complexity

Request 1

All

1925

1926

Request 2

D-A-CH



1997

Request 3

Europe Gen Z

2012

Request 4

France



Years of Birth

2024

Country

rather than the domain size of the partitioning attributes

...

Country

1925

1926

1927

1928

1929

1930

1931

1932

1933

1934

1935

1936

1937

1938

1939

1940

1941

1942

1943

1944

1945

1946

1947

1948

1949

1950

1951

1952

1953

1954

1955

1956

1957

1958

1959

1960

1961

1962

1963

1964

1965

1966

1967

1968

1969

1970

1971

1972

1973

1974

1975

1976

1977

1978

1979

1980

1981

1982

1983

1984

1985

1986

1987

1988

1989

1990

1991

1992

1993

1994

1995

1996

1997

1998

1999

2000

2001

2002

2003

2004

2005

2006

2007

2008

2009

2010

2011

2012

2013

2014

2015

2016

2017

2018

2019

2020

2021

2022

2023

2024

2025

2026

2027

2028

2029

2030

2031

2032

2033

2034

2035

2036

2037

2038

2039

2040

2041

2042

2043

2044

2045

2046

2047

2048

2049

2050

2051

2052

2053

2054

2055

2056

2057

2058

2059

2060

2061

2062

2063

2064

2065

2066

2067

2068

2069

2070

2071

2072

2073

2074

2075

2076

2077

2078

2079

2080

2081

2082

2083

2084

2085

2086

2087

2088

2089

2090

2091

2092

2093

2094

2095

2096

2097

2098

2099

20100

20101

20102

20103

20104

20105

20106

20107

20108

20109

20110

20111

20112

20113

20114

20115

20116

20117

20118

20119

20120

20121

20122

20123

20124

20125

20126

20127

20128

20129

20130

20131

20132

20133

20134

20135

20136

20137

20138

20139

20140

20141

20142

20143

20144

20145

20146

20147

20148

20149

20150

20151

20152

20153

20154

20155

20156

20157

20158

20159

20160

20161

20162

20163

20164

20165

20166

20167

20168

20169

20170

20171

20172

20173

20174

20175

20176

20177

20178

20179

20180

20181

20182

20183

20184

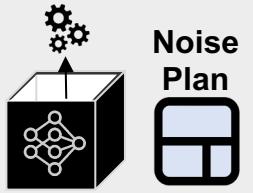
20185

20186

20187

1

Unified System Architecture



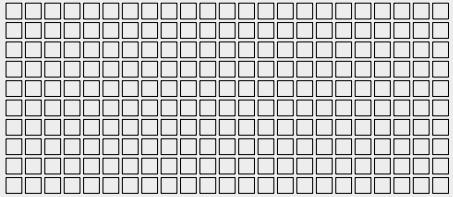
Unified Application Layer

Privacy ID with
Partitioning Attributes
 (a_1, a_2, a_3)

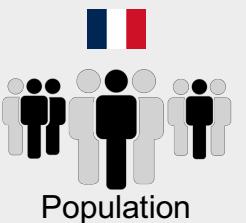
Unified Data Layer

2

Access Pattern for Privacy Analysis



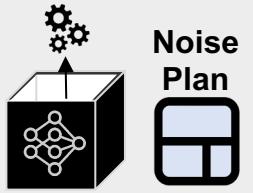
Fine-Grained Block Composition



Subsampling

1

Unified System Architecture



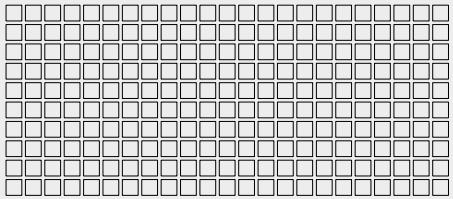
Unified Application Layer

Privacy ID with
Partitioning Attributes
(a₁, a₂, a₃)

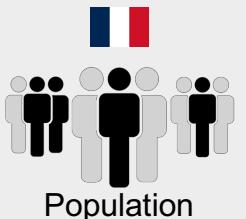
Unified Data Layer

2

Access Pattern for Privacy Analysis



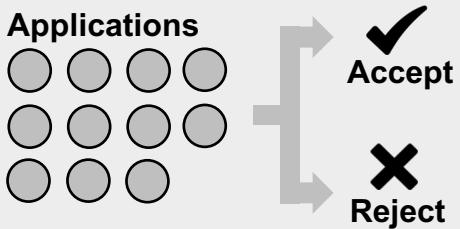
Fine-Grained Block Composition



Subsampling

3

System Continuity Guarantee



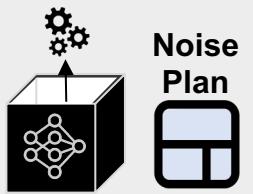
Periodic Planning



Budget Guarantee

1

Unified System Architecture



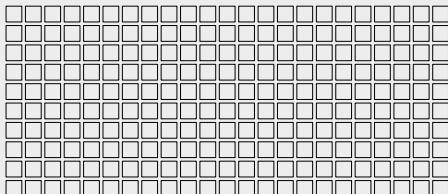
Unified Application Layer

Privacy ID with
Partitioning Attributes
 (a_1, a_2, a_3)

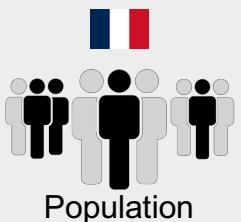
Unified Data Layer

2

Access Pattern for Privacy Analysis



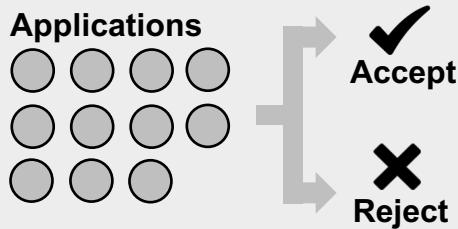
Fine-Grained Block Composition



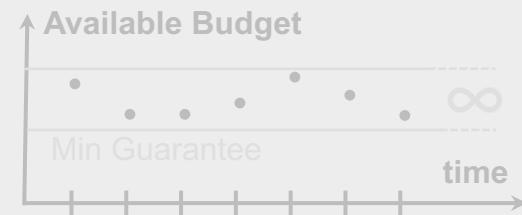
Subsampling

3

System Continuity Guarantee



Periodic Planning



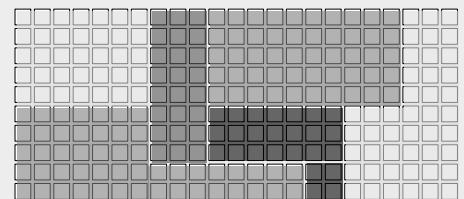
Budget Guarantee

4

Resource Allocation

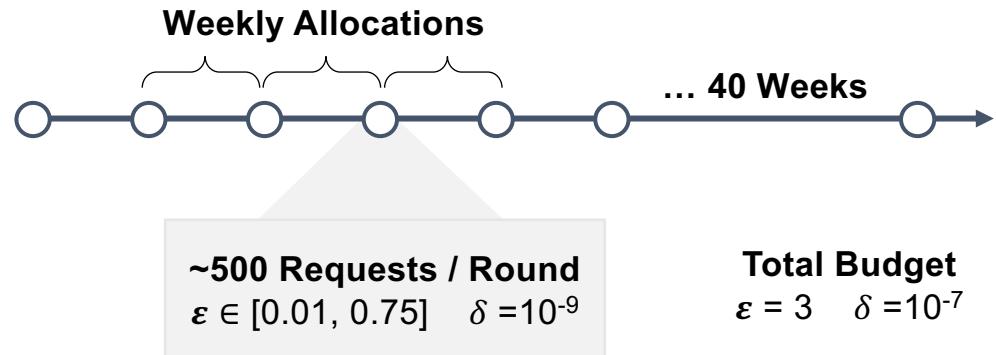
$$\begin{aligned} \max \quad & \sum_{i \in Apps} Utility_i * y_i \\ \text{s.t.} \quad & \sum_{i \in Apps} \varepsilon_{ij} * y_i \leq Budget_j \quad \forall j \\ & \text{for simplicity in } \varepsilon\text{-DP} \end{aligned}$$

Multidimensional Knapsack

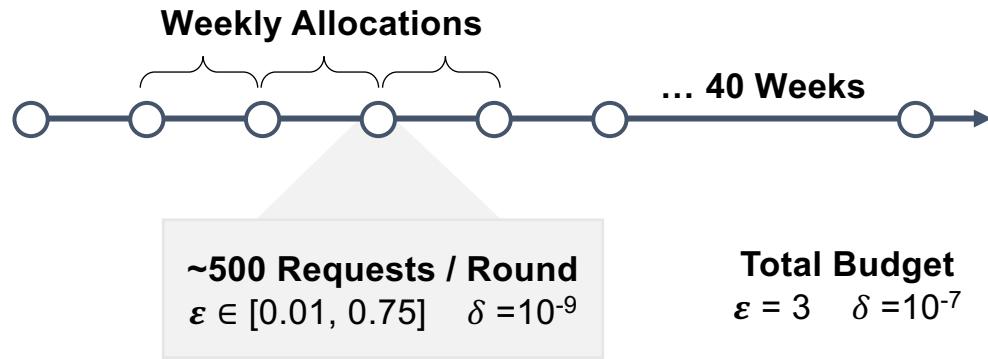


Dimensionality Reduction

Evaluation Scenario



Evaluation Scenario



Baseline

PrivateKube
[Luo et al. OSDI'21]

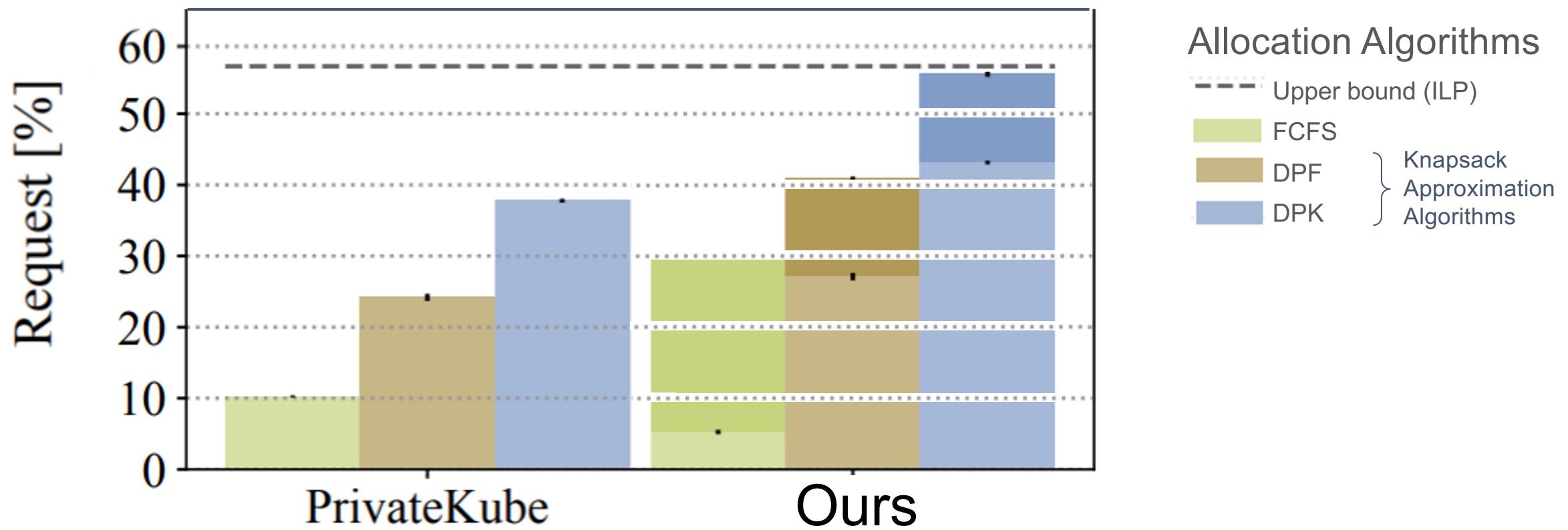


kubernetes

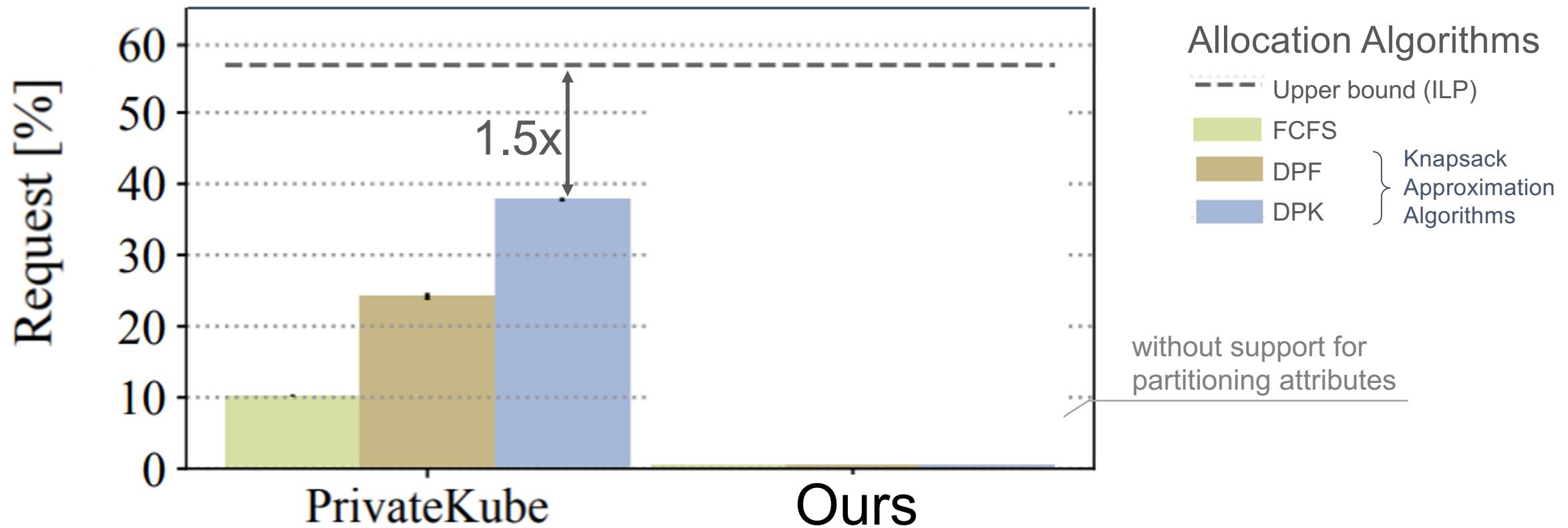


**Fixed Coarse-Grained
Privacy Analysis**

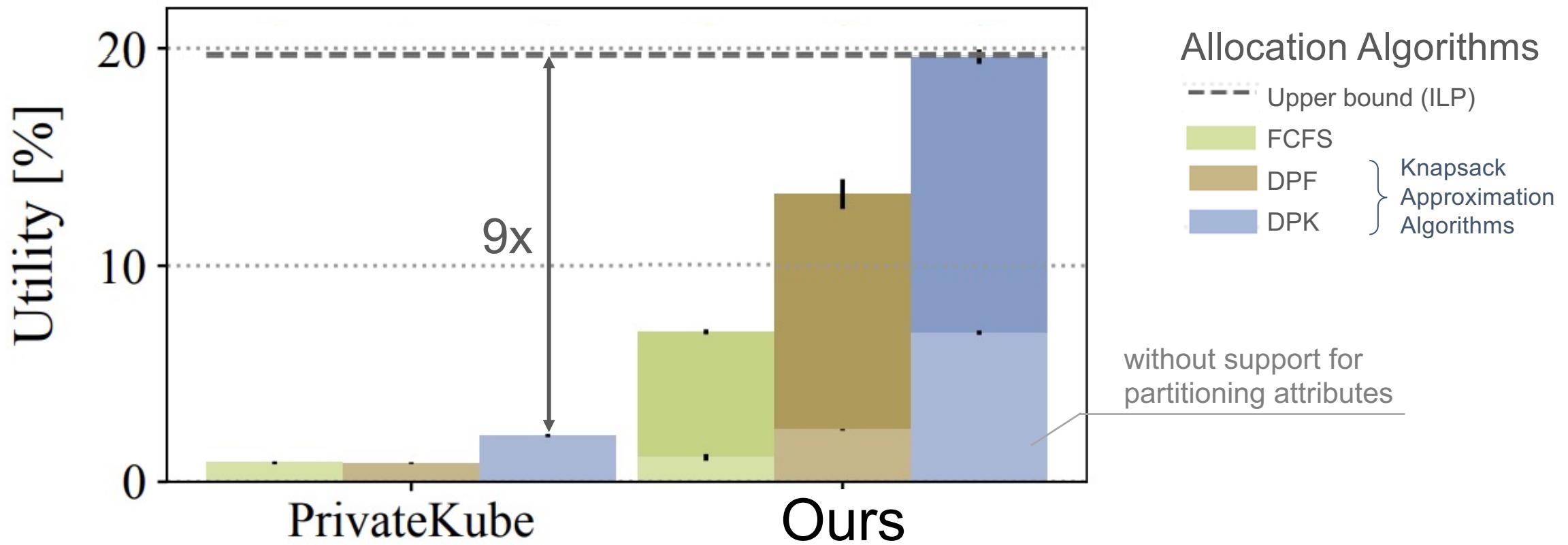
Workload: Mixture of Analytics and ML Tasks



Workload: Mixture of Analytics and ML Tasks



Workload: Mixture of Analytics and ML Tasks



Cohere: Managing DP in Large - Scale Systems

DP
Theory

SCAN ME



pps-lab/cohere

System-wide DP Guarantee
Cross-framework Compatibility and Efficient Privacy Analysis

Resource Allocation
Distributing Budget across various Applications

System Continuity
Ensuring Sustained Budget Allocation Over Time

Practice