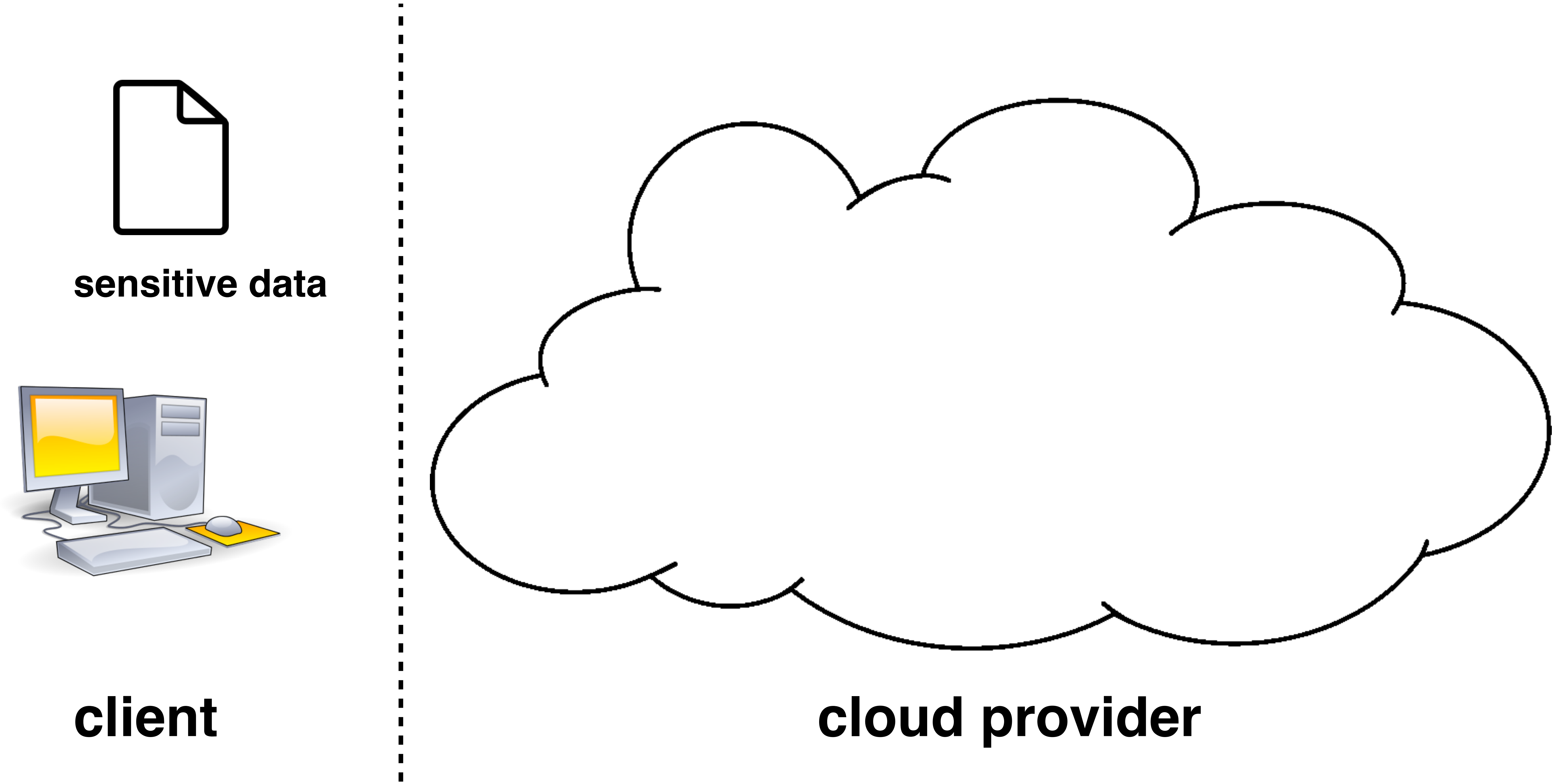
Opaque

An Oblivious and Encrypted Distributed Analytics Platform

Wenting Zheng, Ankur Dave, Jethro G. Beekman,
Raluca Ada Popa, Joseph E. Gonzalez, Ion Stoica

UC Berkeley

Complex analytics run on sensitive data

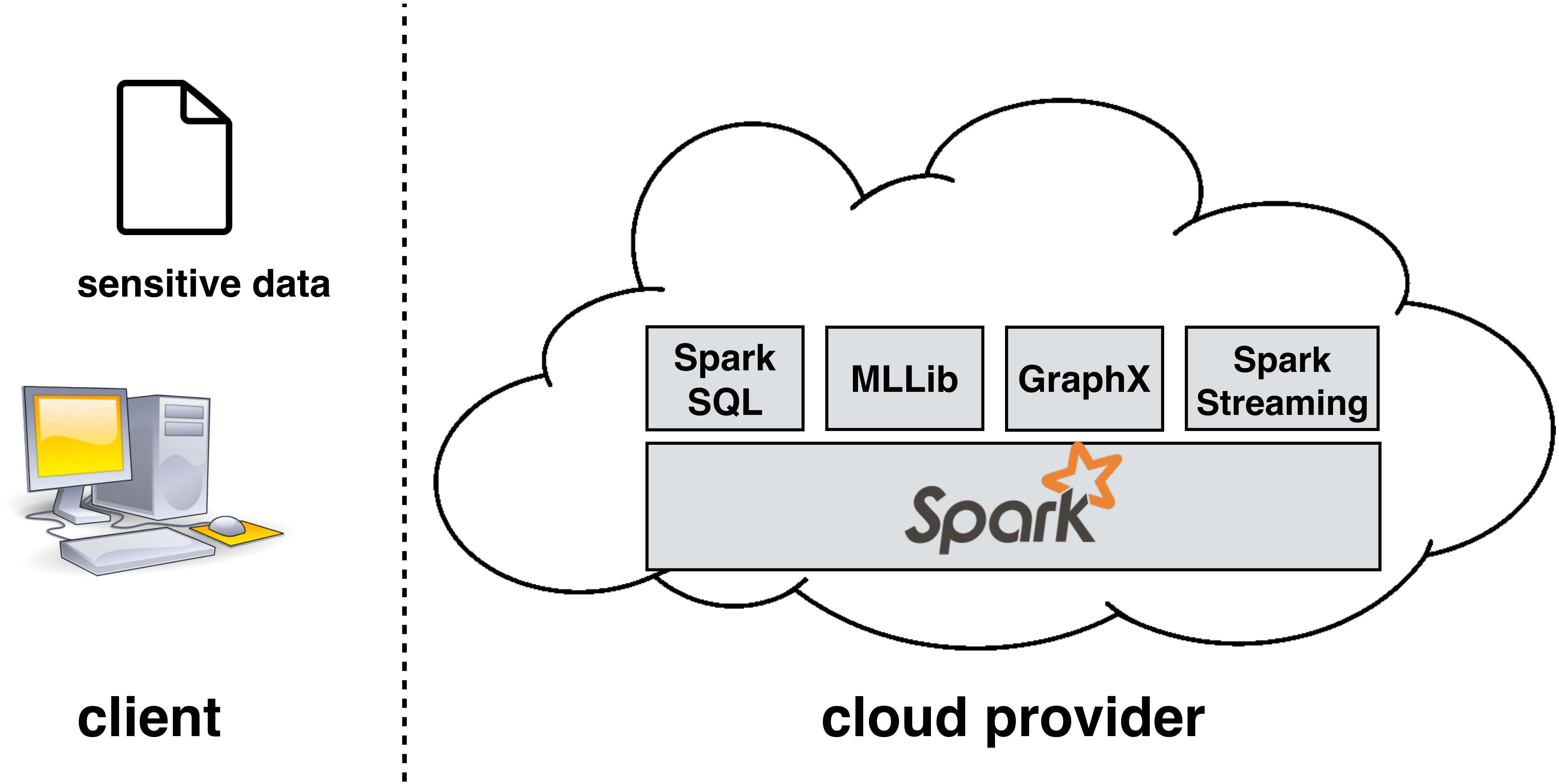


sensitive data

client

cloud provider

Complex analytics run on sensitive data



sensitive data

Spark SQL

MLLib

GraphX

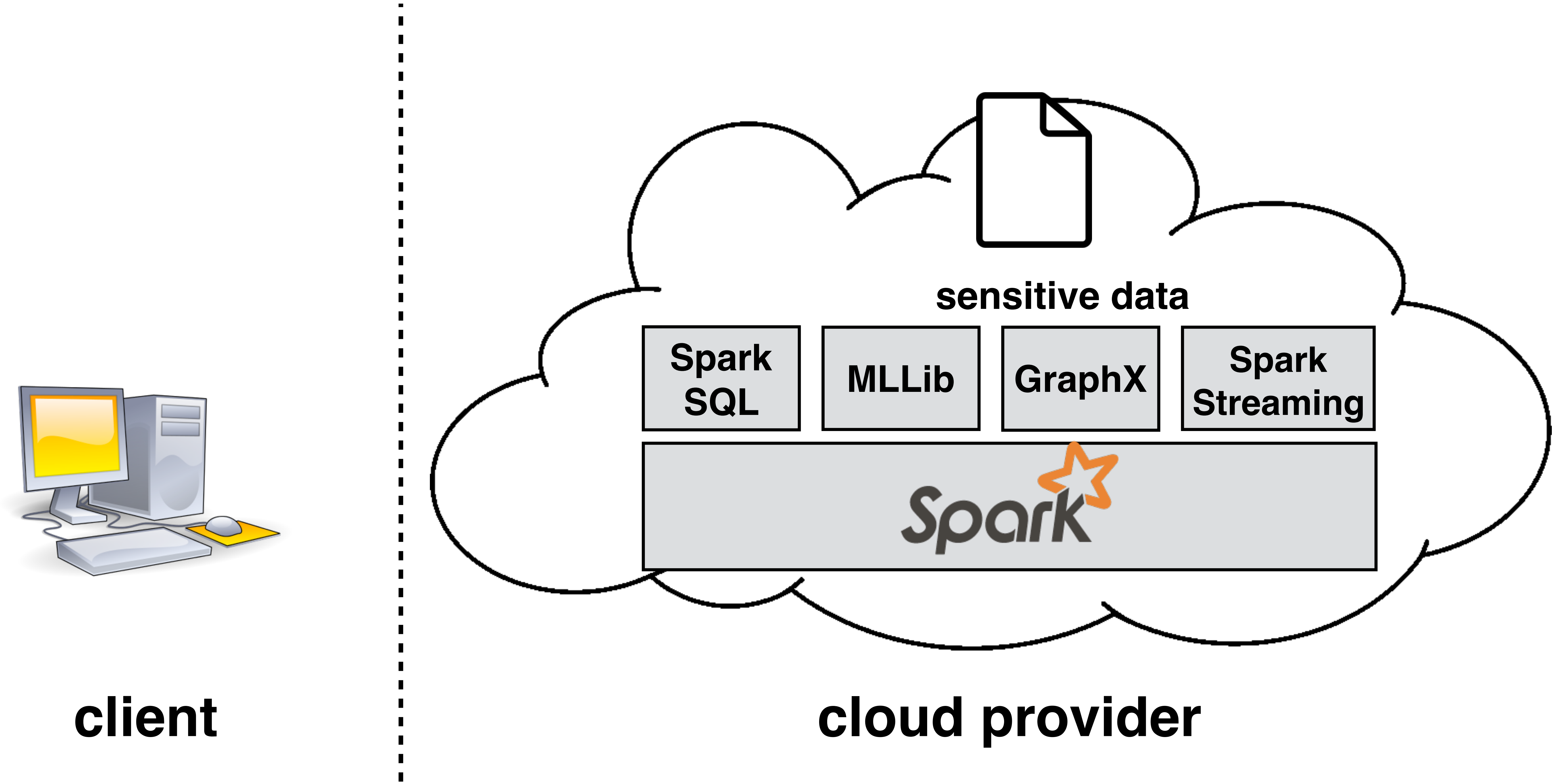
Spark Streaming

Spark

client

cloud provider

Complex analytics run on sensitive data



sensitive data

Spark SQL

MLLib

GraphX

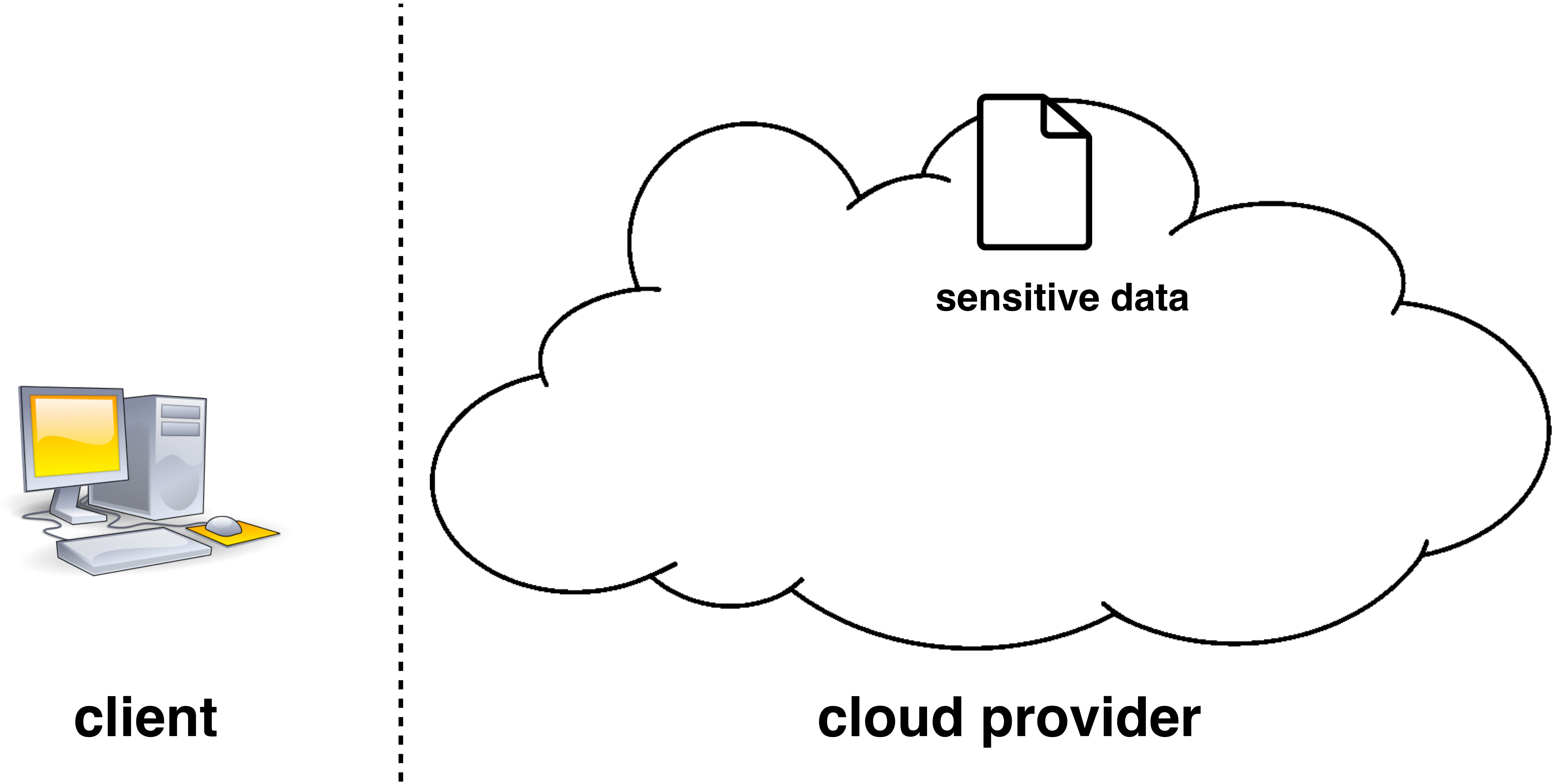
Spark Streaming

Spark

client

cloud provider

Cloud attackers

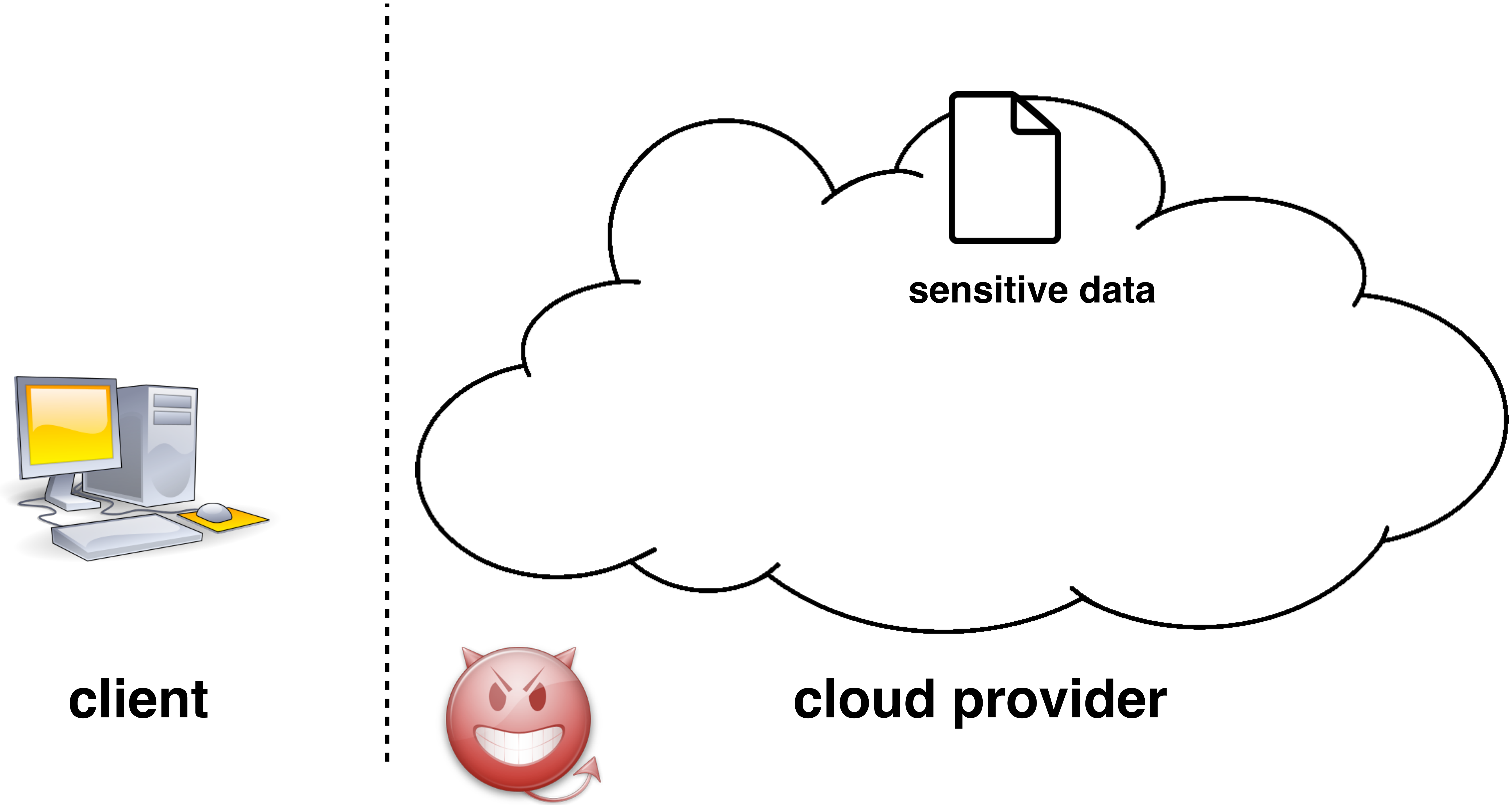


sensitive data

client

cloud provider

Cloud attackers

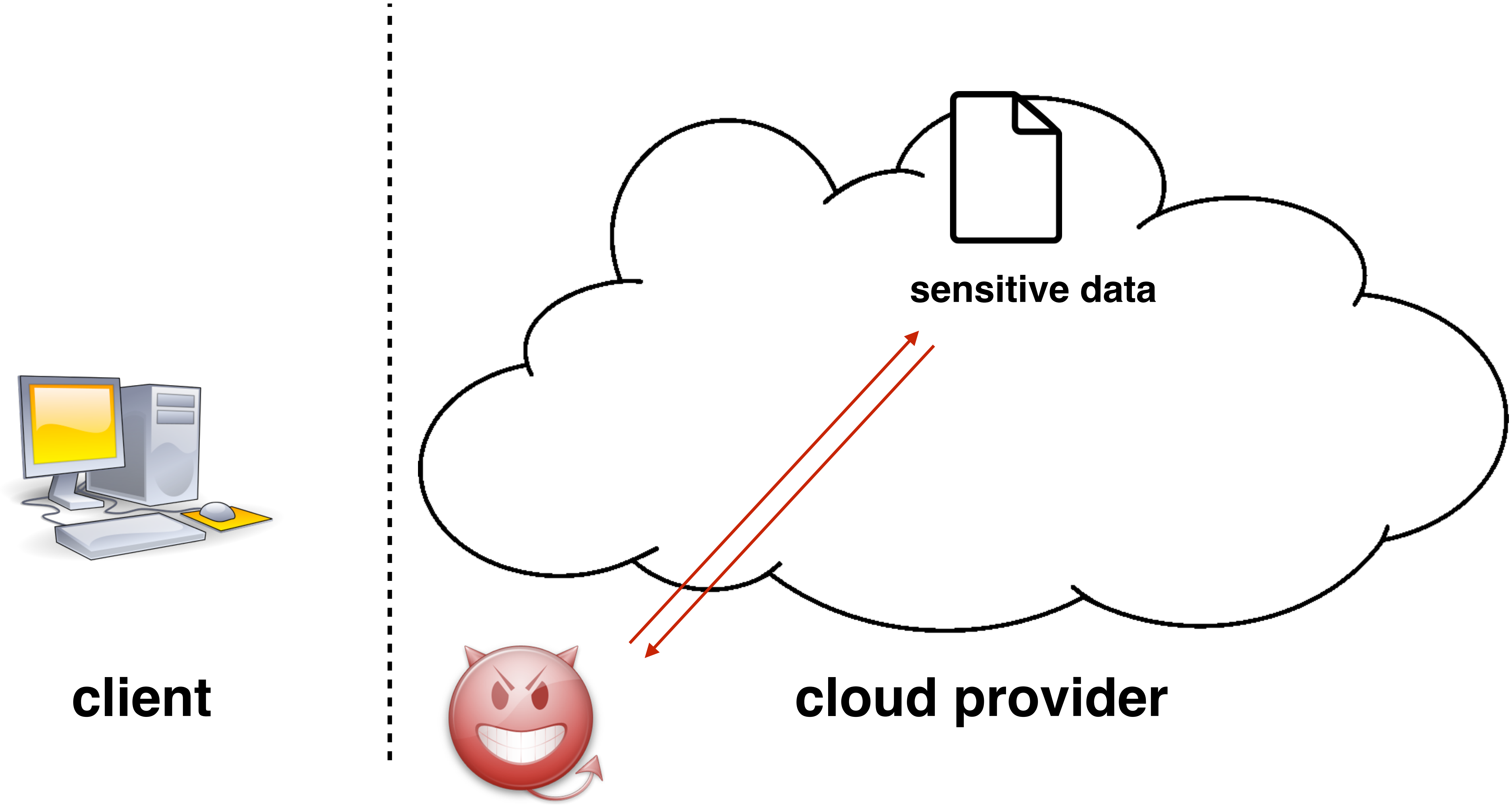


sensitive data

client

cloud provider

Cloud attackers

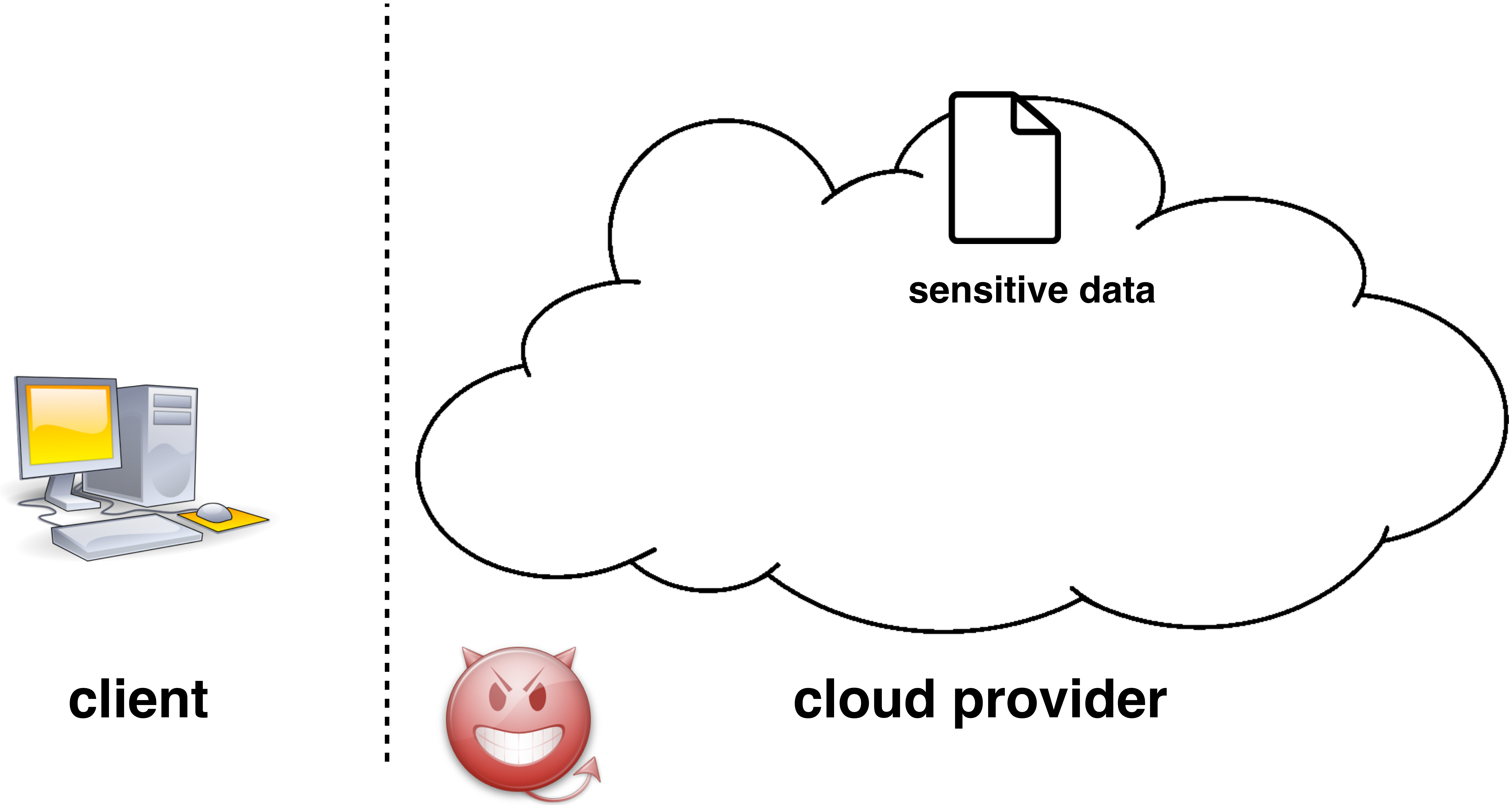


sensitive data

client

cloud provider

Cloud attackers

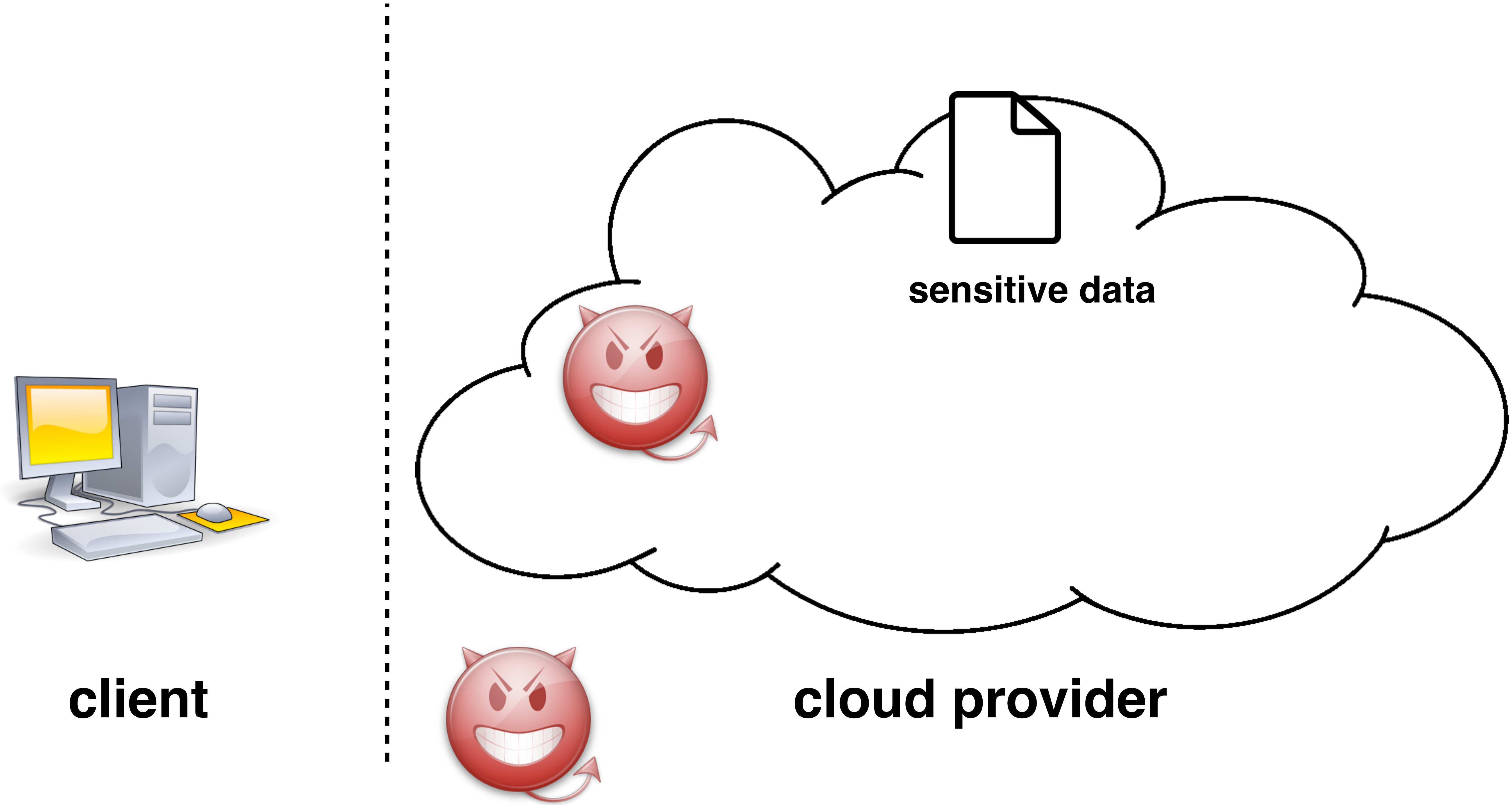


sensitive data

client

cloud provider

Cloud attackers

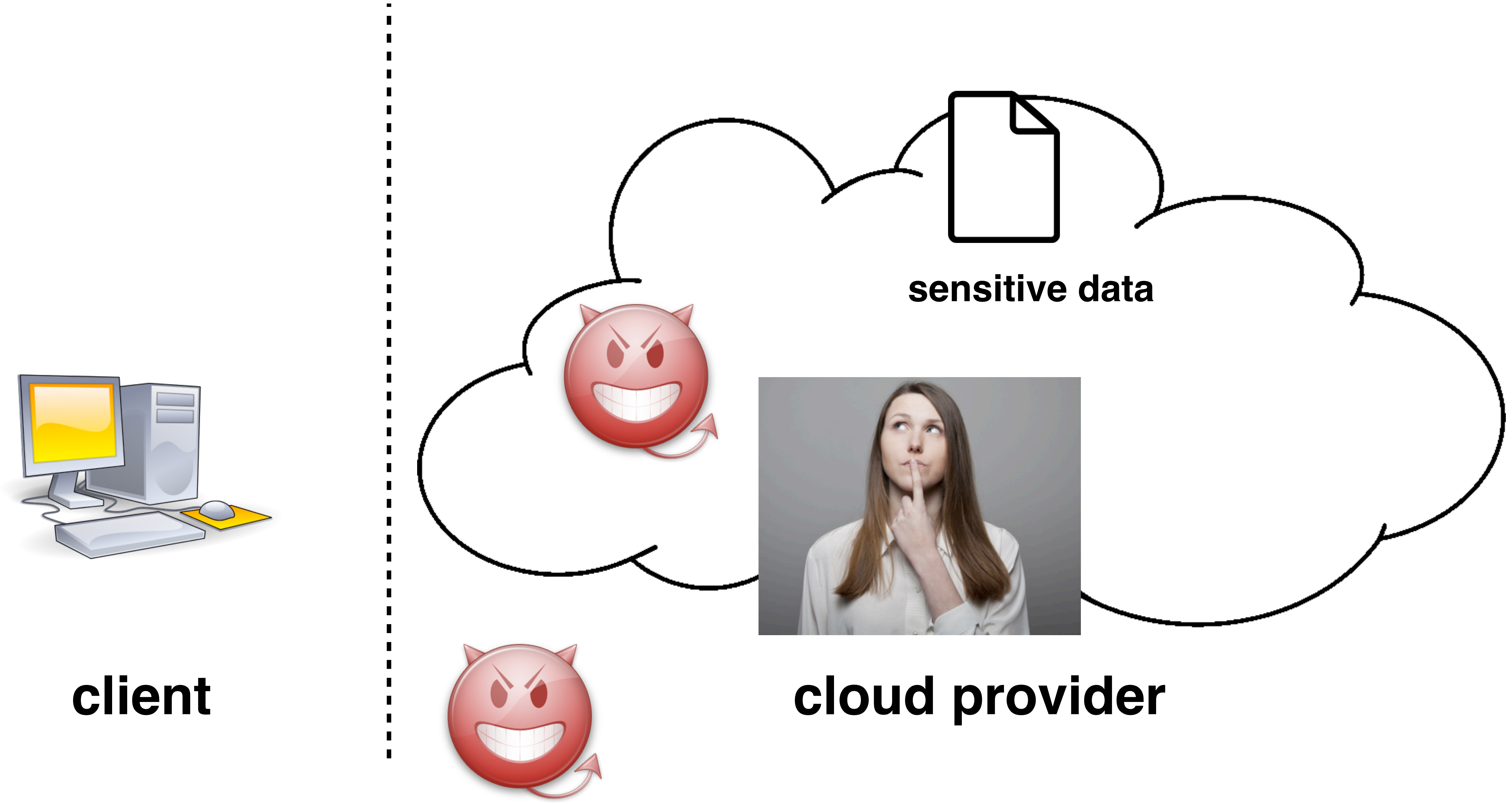


sensitive data

client

cloud provider

Cloud attackers

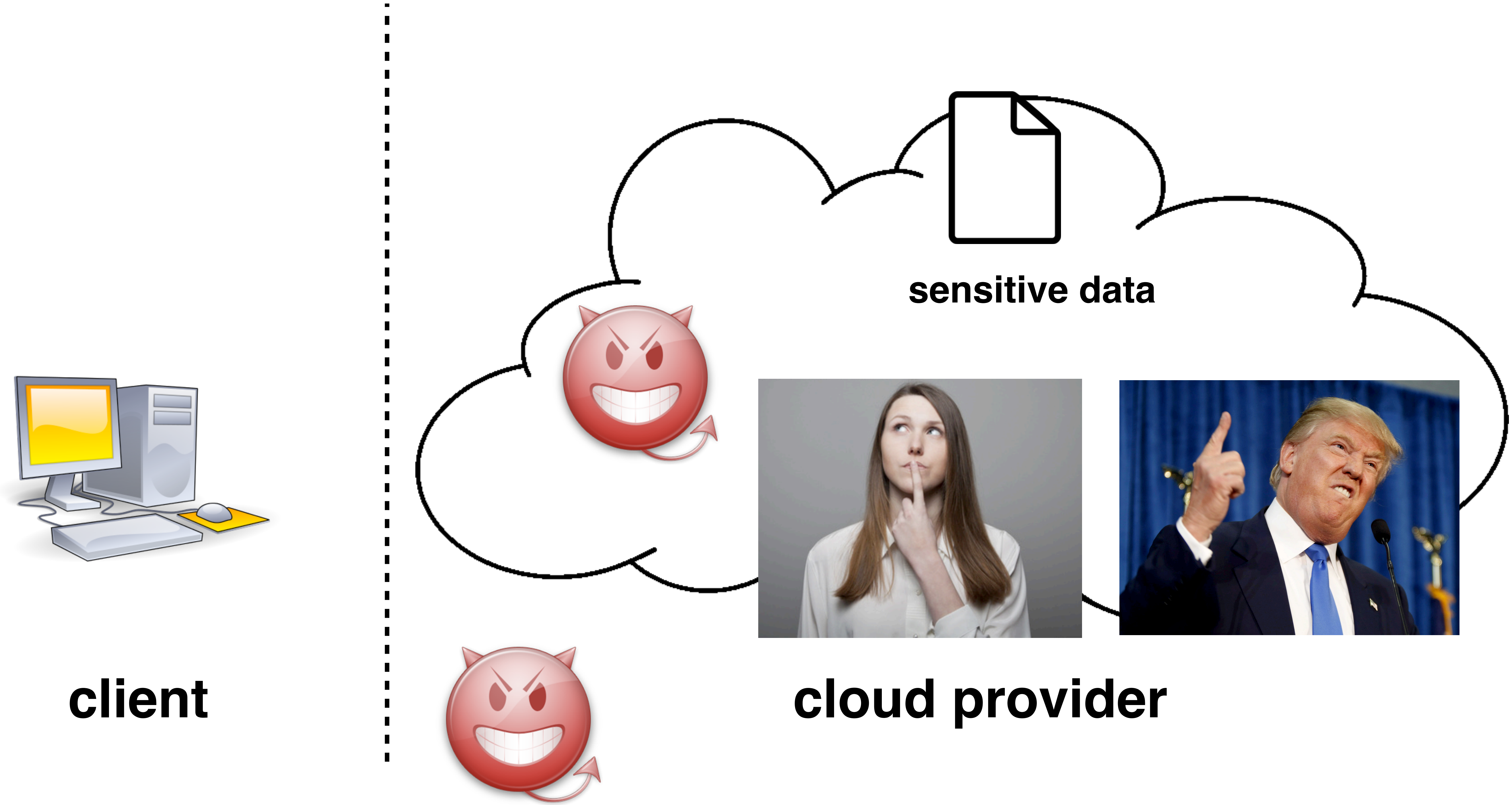


sensitive data

client

cloud provider

Cloud attackers



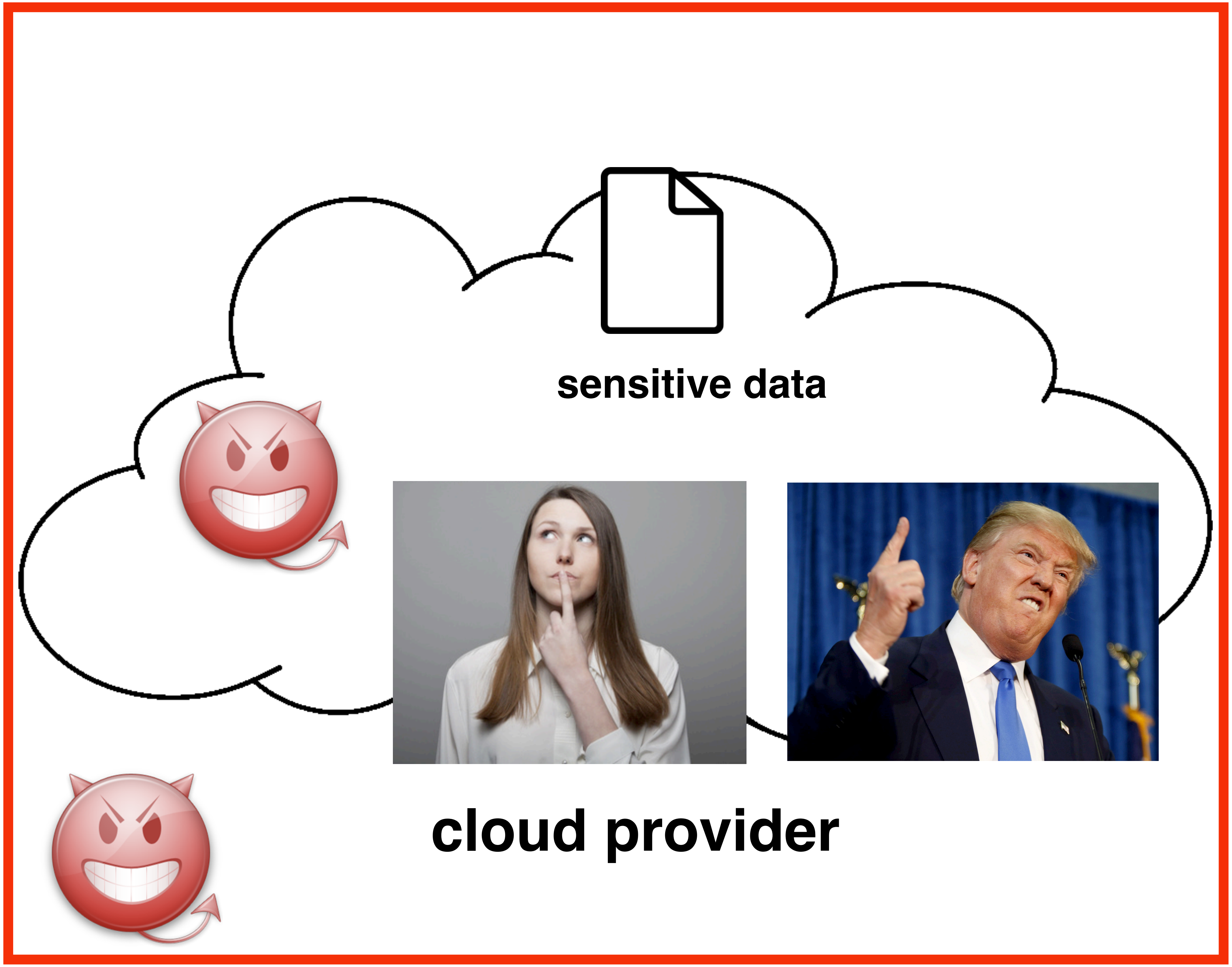
sensitive data

client

cloud provider

Cloud attackers

client



sensitive data

cloud provider

How to protect data
while preserving functionality?

Cryptographic approach?

Cryptographic approach?

- Fully homomorphic encryption [Gentry ’09]

Cryptographic approach?

- Fully homomorphic encryption [Gentry '09]
  - **fully functional**

Cryptographic approach?

- Fully homomorphic encryption [Gentry '09]
  - **fully functional**
  - **too slow**

Cryptographic approach?

- Fully homomorphic encryption [Gentry '09]
  - **fully functional**
  - **too slow**
- CryptDB [PRZB '11]

Cryptographic approach?

- Fully homomorphic encryption [Gentry '09]
  - **fully functional**
  - **too slow**
- CryptDB [PRZB '11]
  - **more practical performance**

Cryptographic approach?

- Fully homomorphic encryption [Gentry ’09]
  - **fully functional**
  - **too slow**
- CryptDB [PRZB ’11]
  - **more practical performance**
  - **limited functionality**

Cryptographic approach?

- Fully homomorphic encryption [Gentry '09]
  - **fully functional**
  - **too slow**
- CryptDB [PRZB '11]
  - **more practical performance**
  - **limited functionality**

Alternative: hardware enclaves

Hardware enclaves

(e.g., Intel SGX, AMD memory encryption)

Hardware enclaves

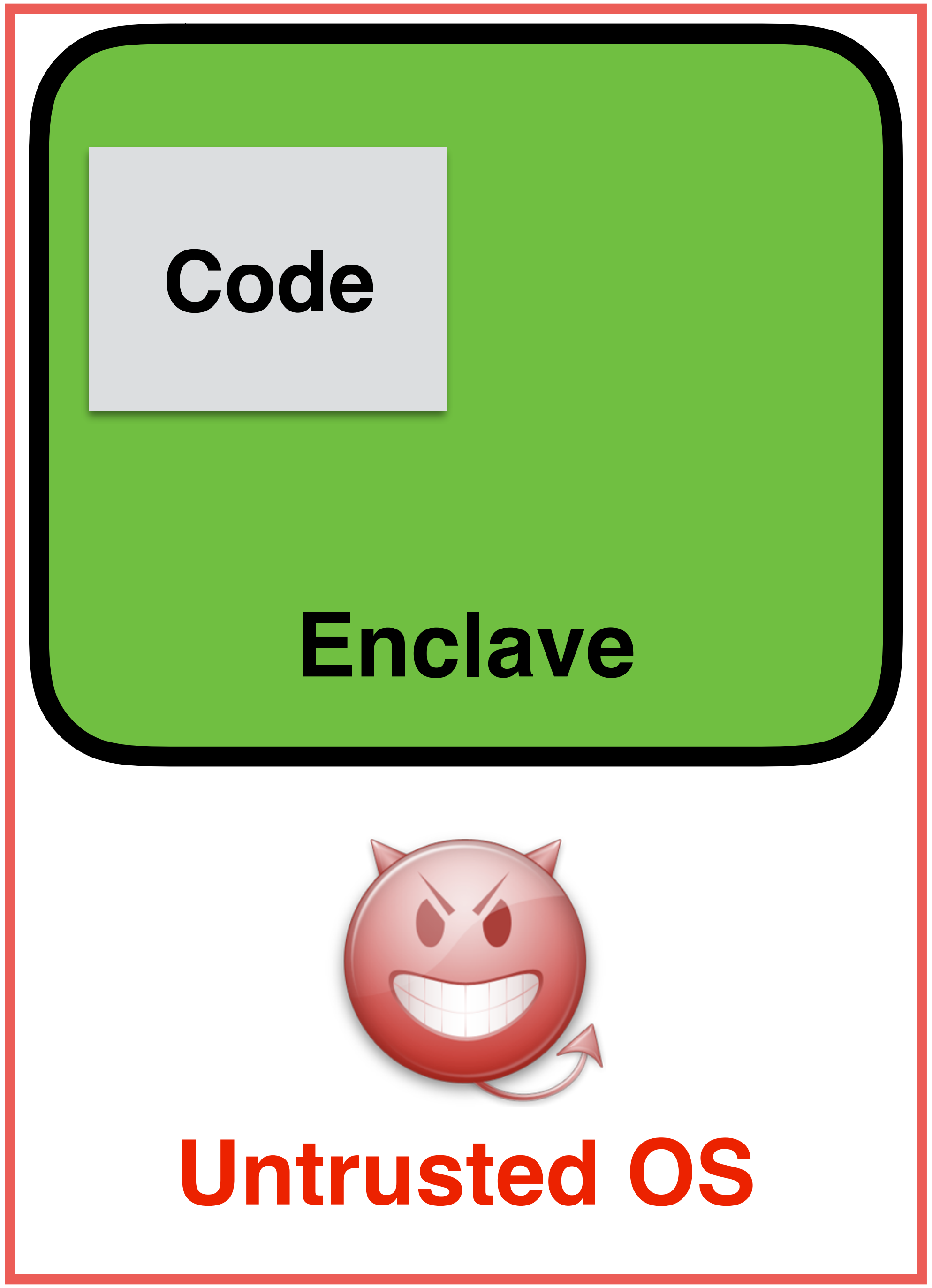
(e.g., Intel SGX, AMD memory encryption)

- Hardware-enforced secure execution environment

Hardware enclaves

(e.g., Intel SGX, AMD memory encryption)

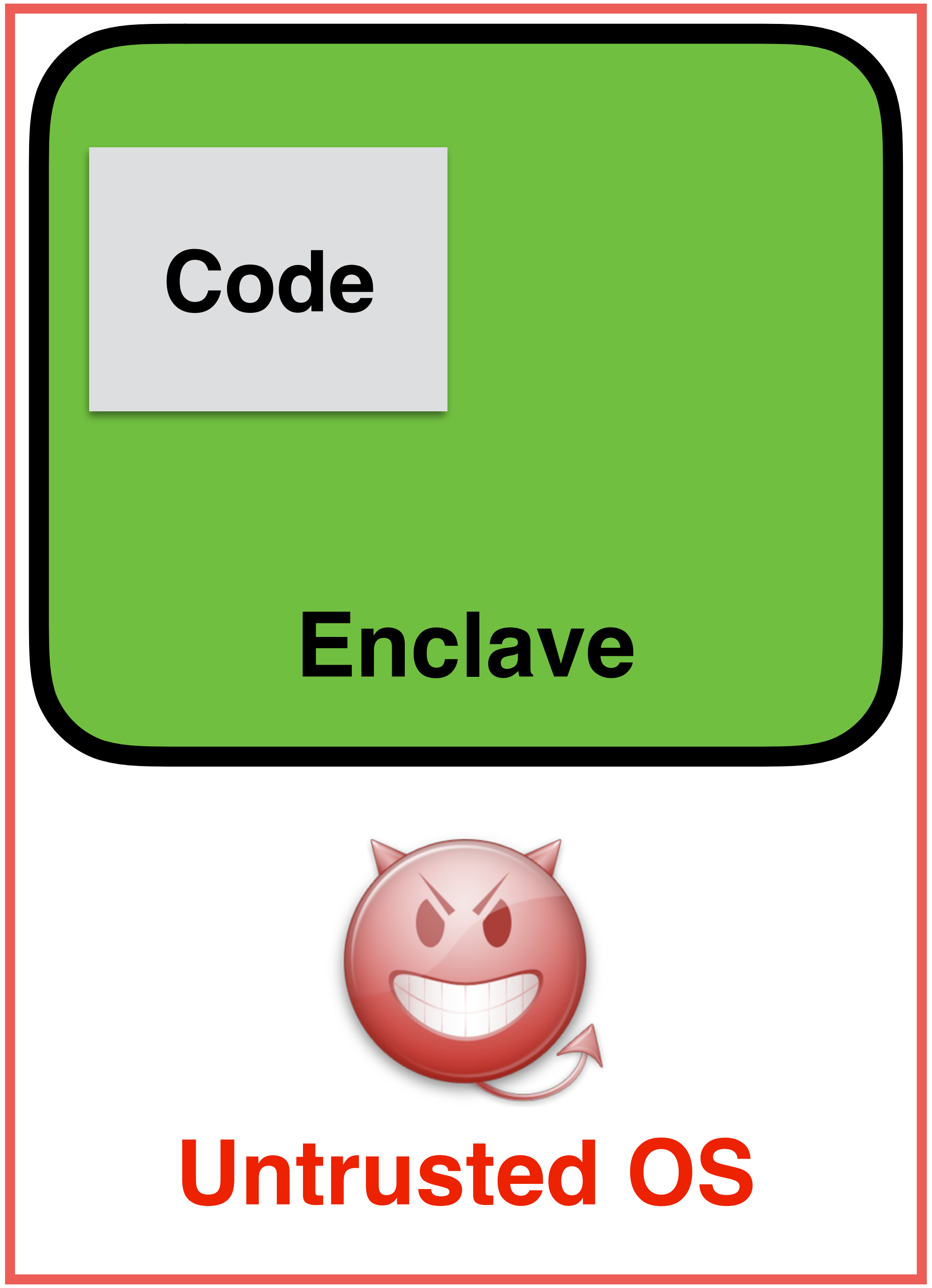
- Hardware-enforced secure execution environment



Hardware enclaves

(e.g., Intel SGX, AMD memory encryption)

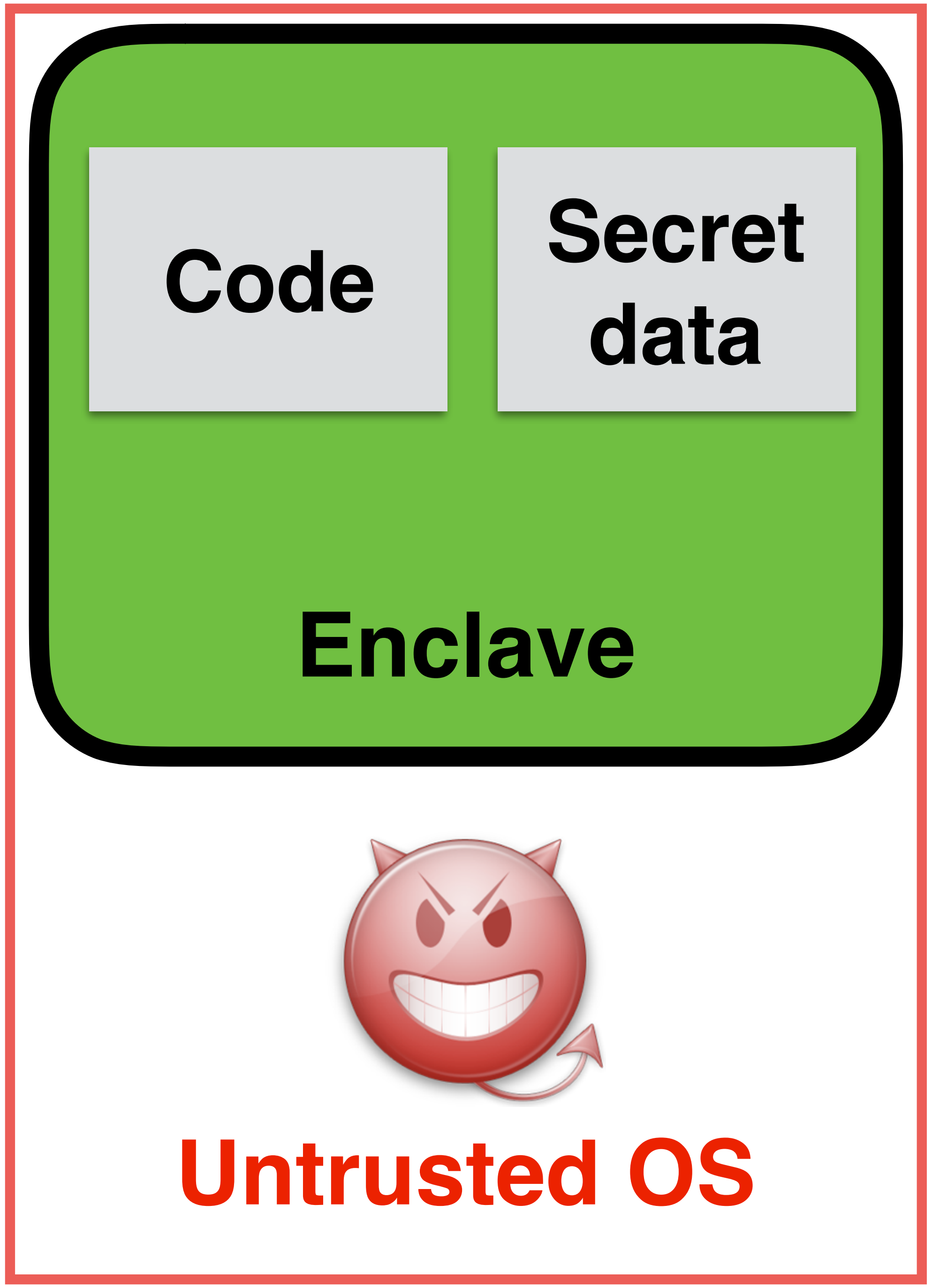
- Hardware-enforced secure execution environment
- EPC: encrypted enclave memory (accessible only from the enclave)



Hardware enclaves

(e.g., Intel SGX, AMD memory encryption)

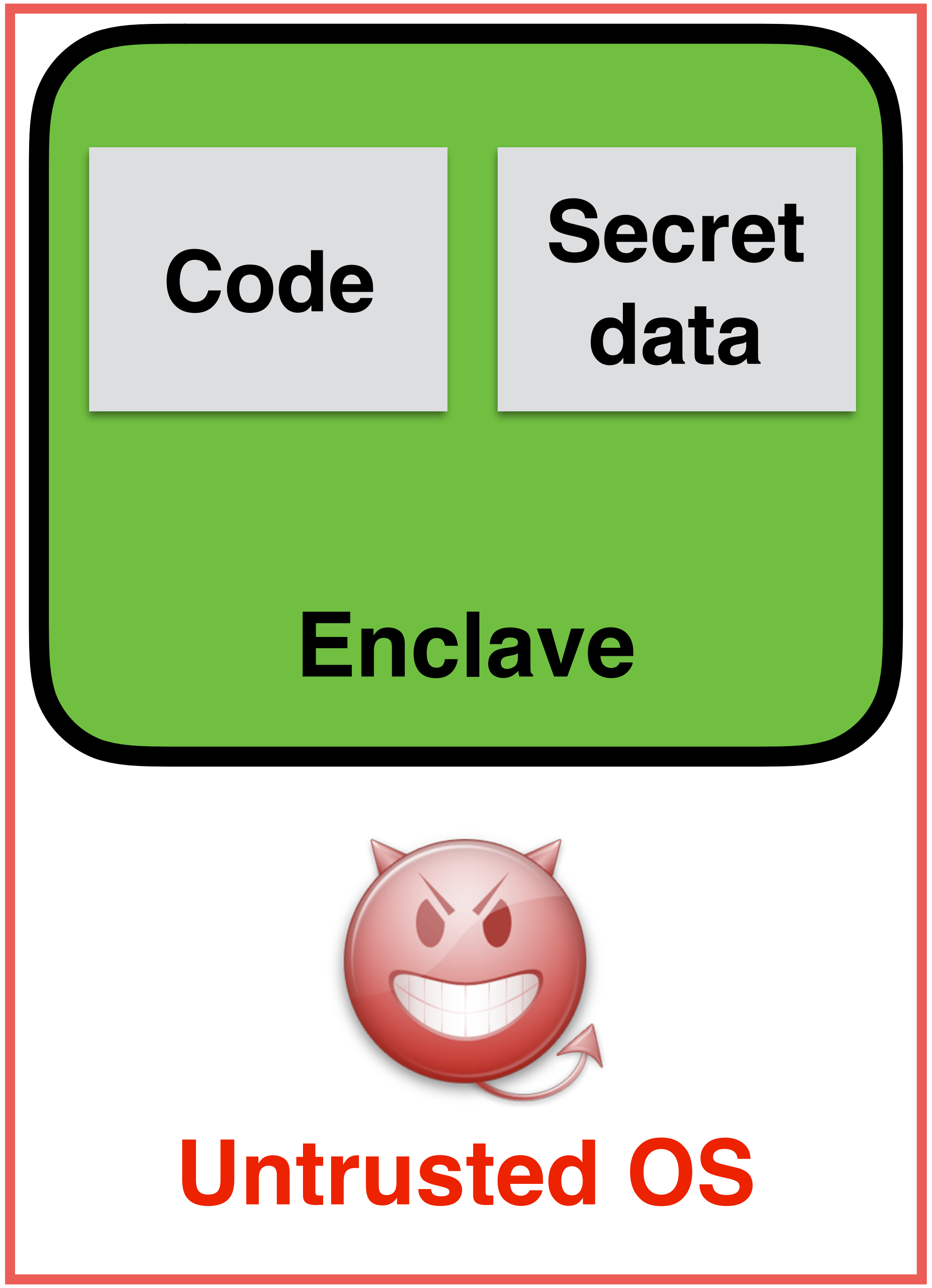
- Hardware-enforced secure execution environment
- EPC: encrypted enclave memory (accessible only from the enclave)



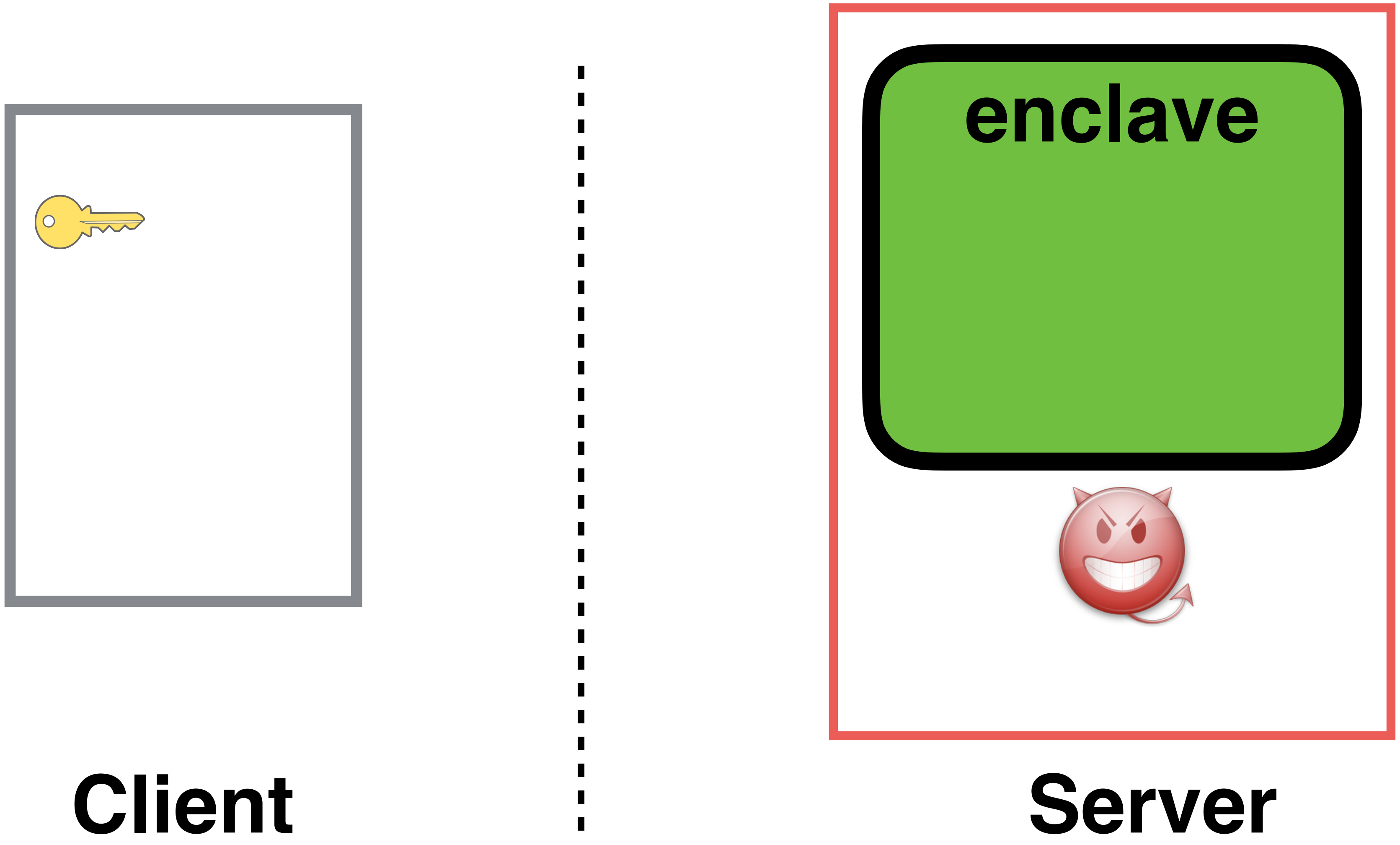
Hardware enclaves

(e.g., Intel SGX, AMD memory encryption)

- Hardware-enforced secure execution environment
- EPC: encrypted enclave memory (accessible only from the enclave)
- Protect against an attacker who has root access



Remote attestation



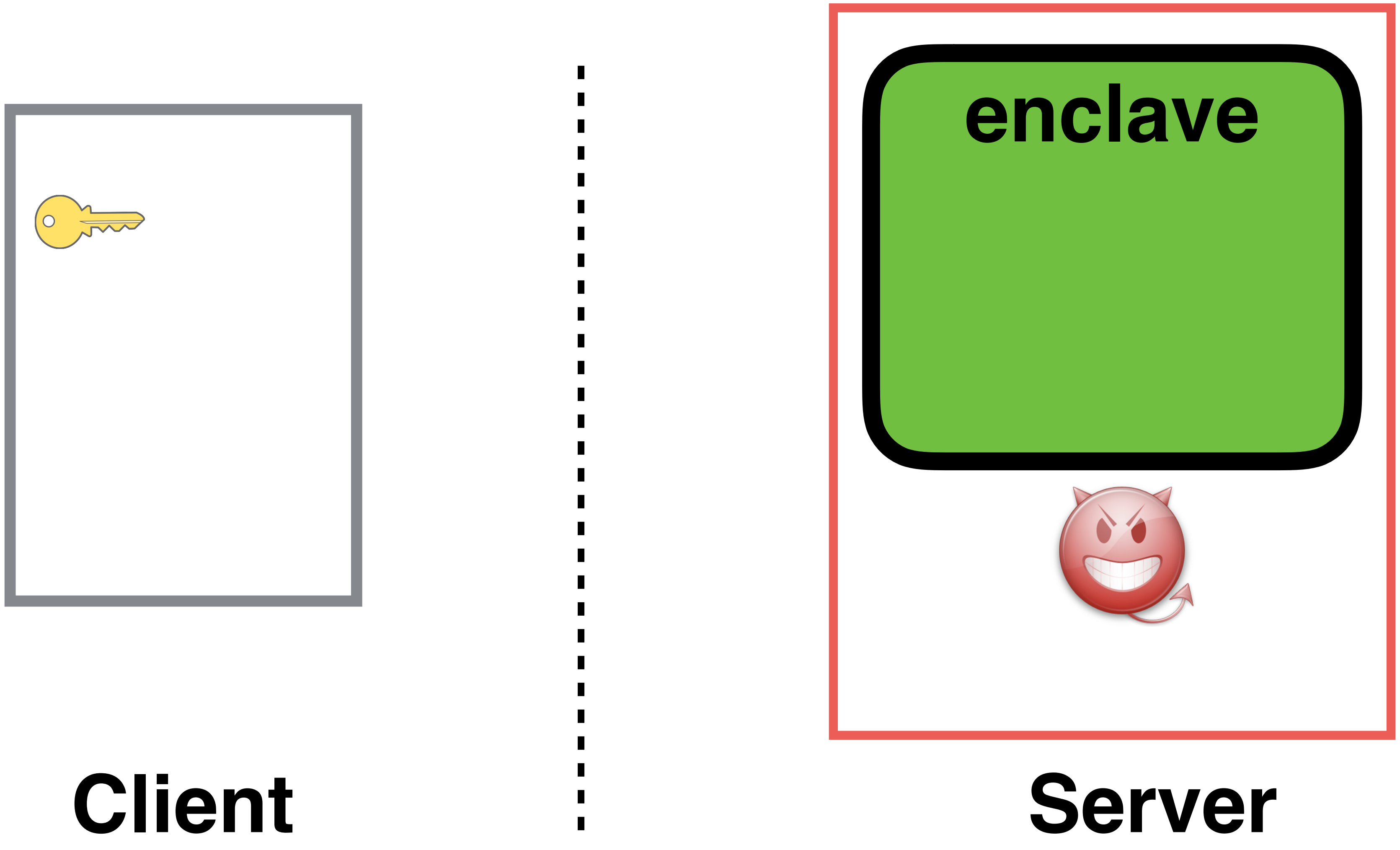
enclave

Client

Server

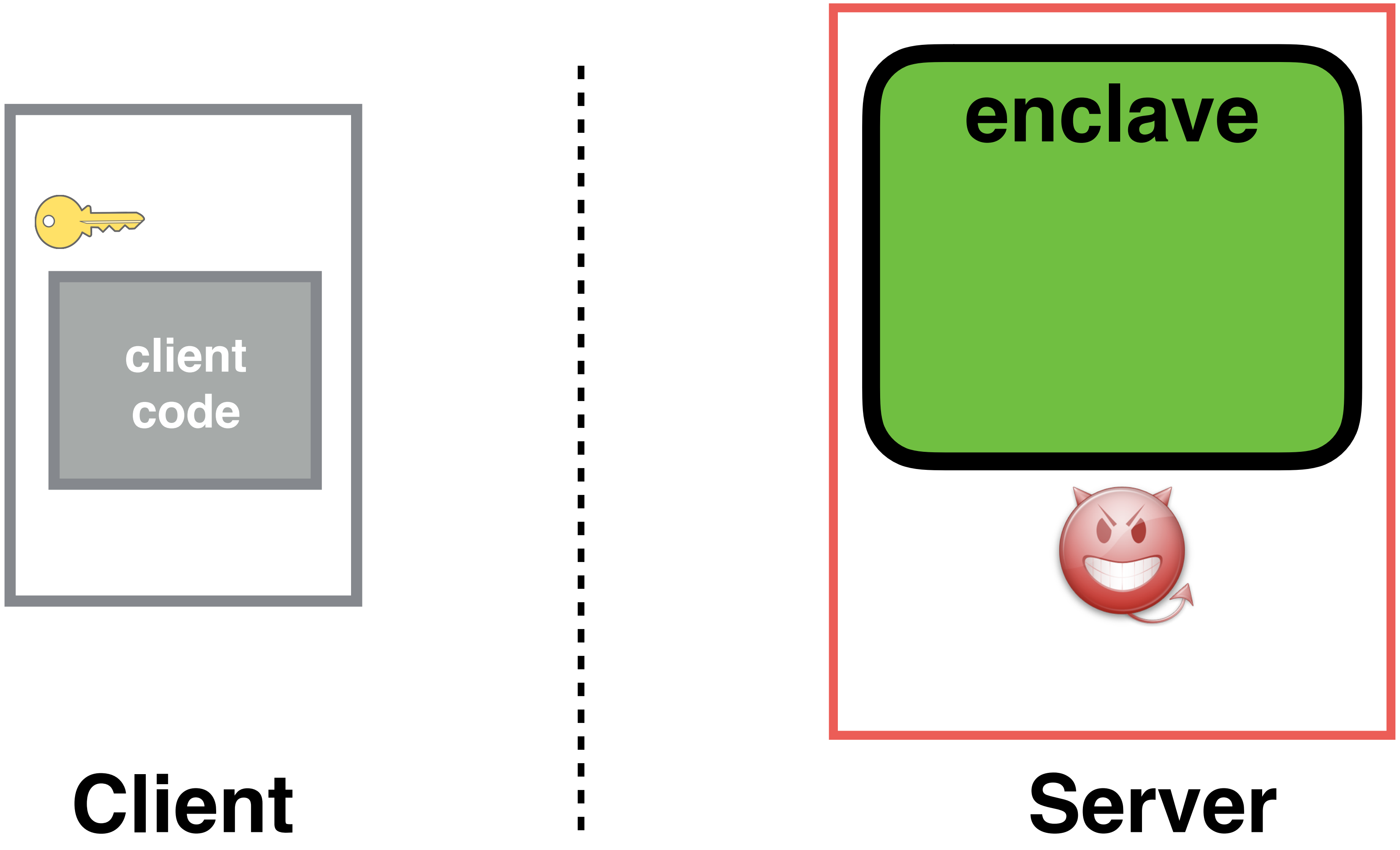
Remote attestation

Enables verifying which code runs in the enclave and performing key exchange



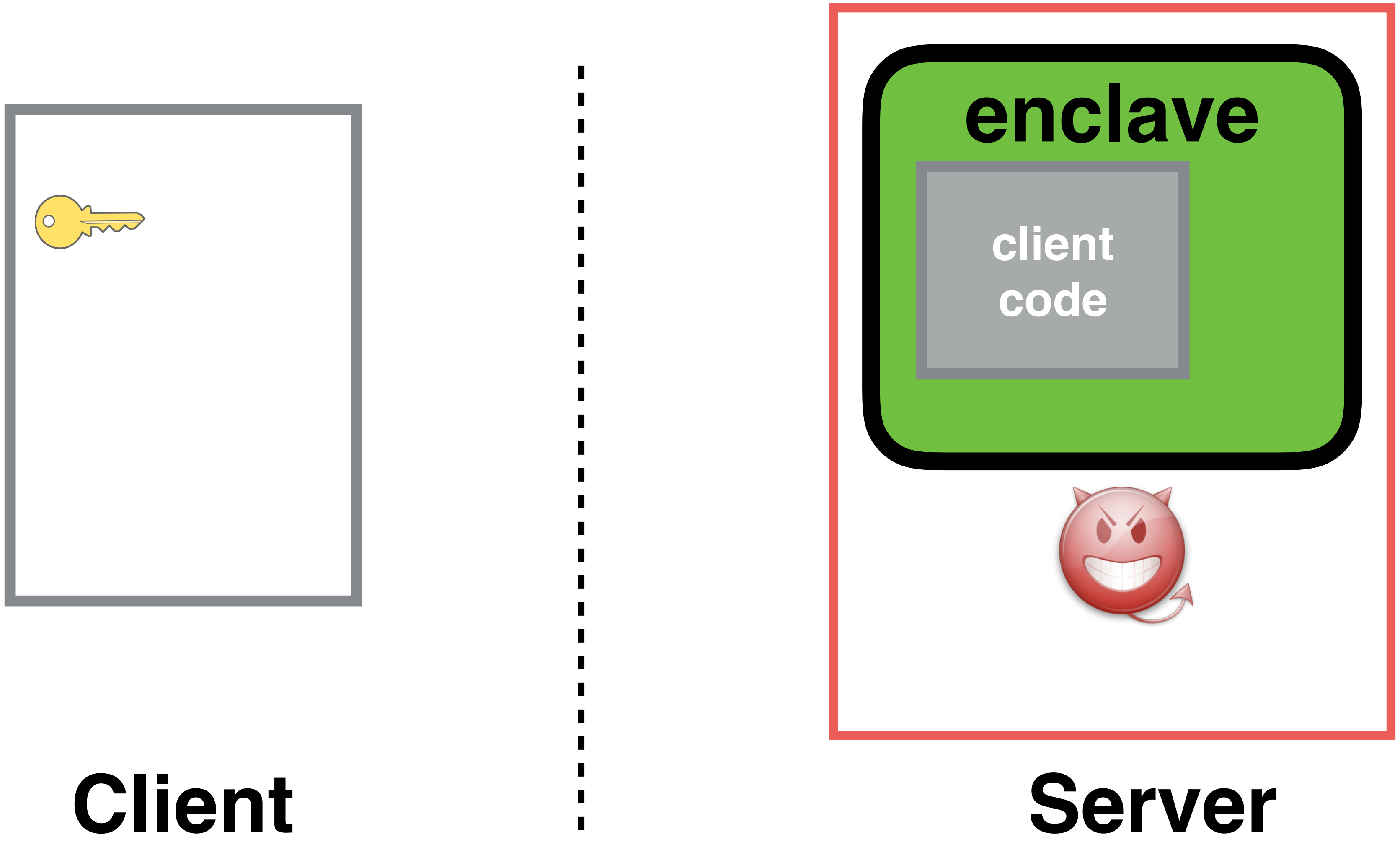
Remote attestation

Enables verifying which code runs in the enclave and performing key exchange



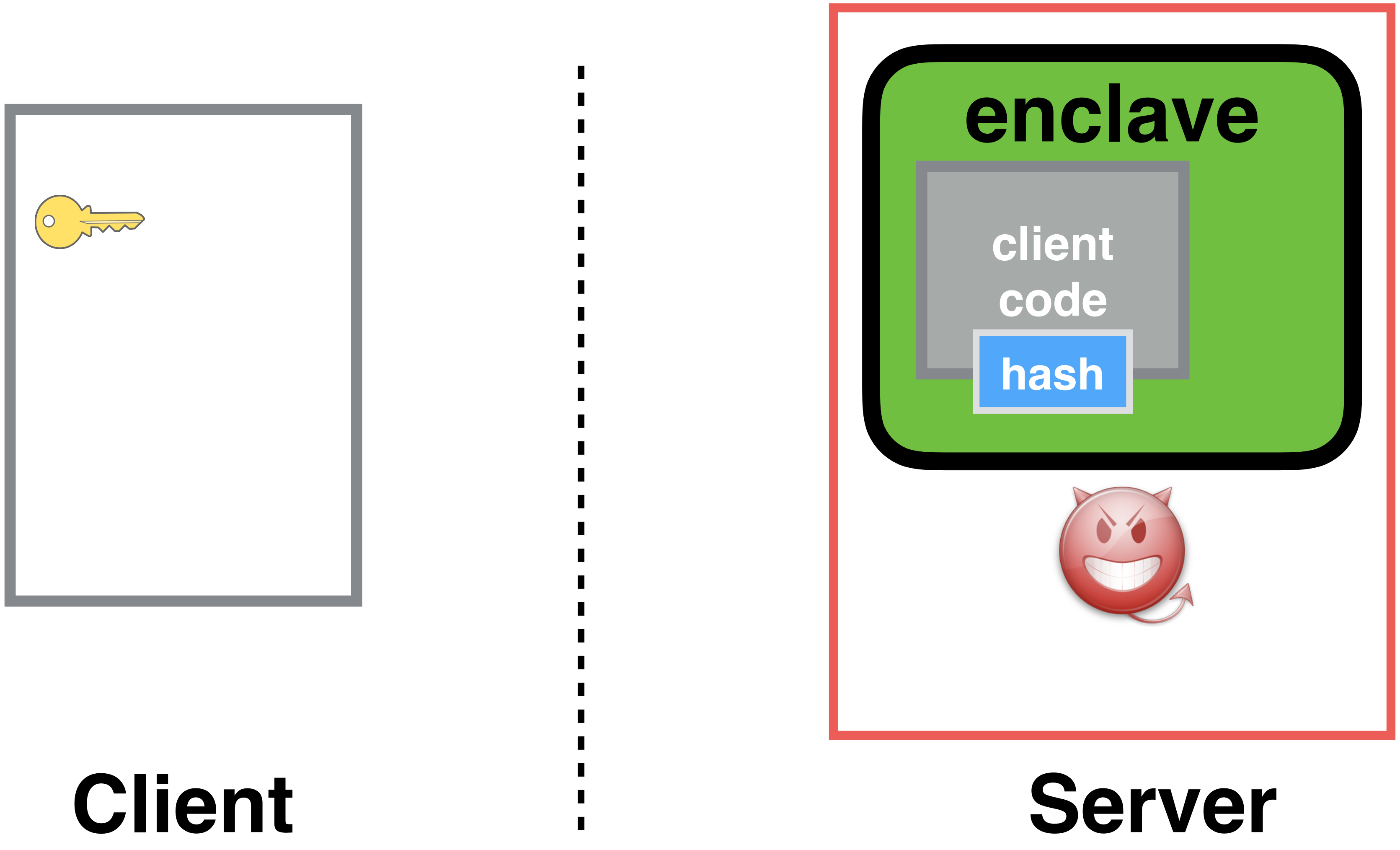
Remote attestation

Enables verifying which code runs in the enclave and performing key exchange



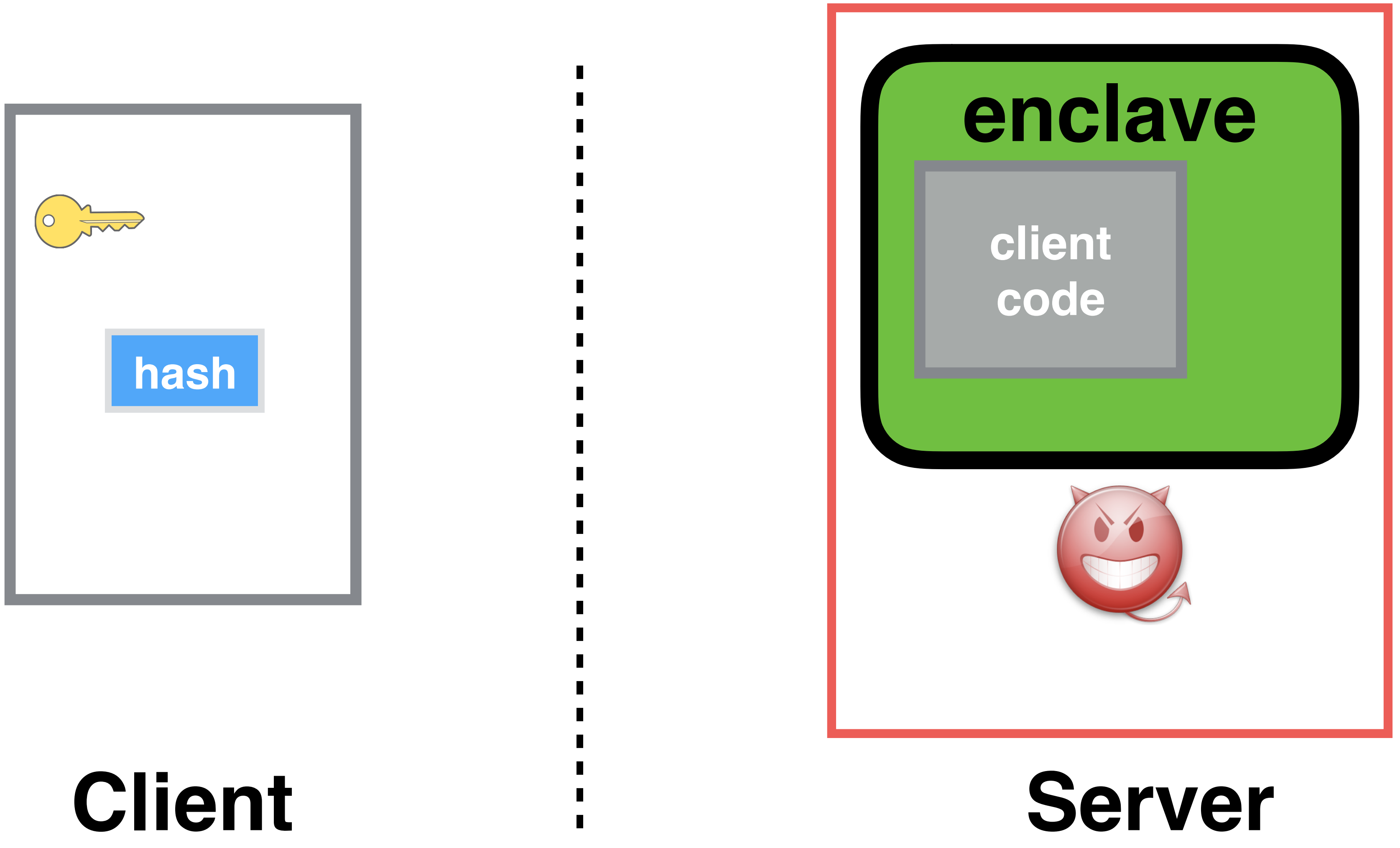
Remote attestation

Enables verifying which code runs in the enclave and performing key exchange



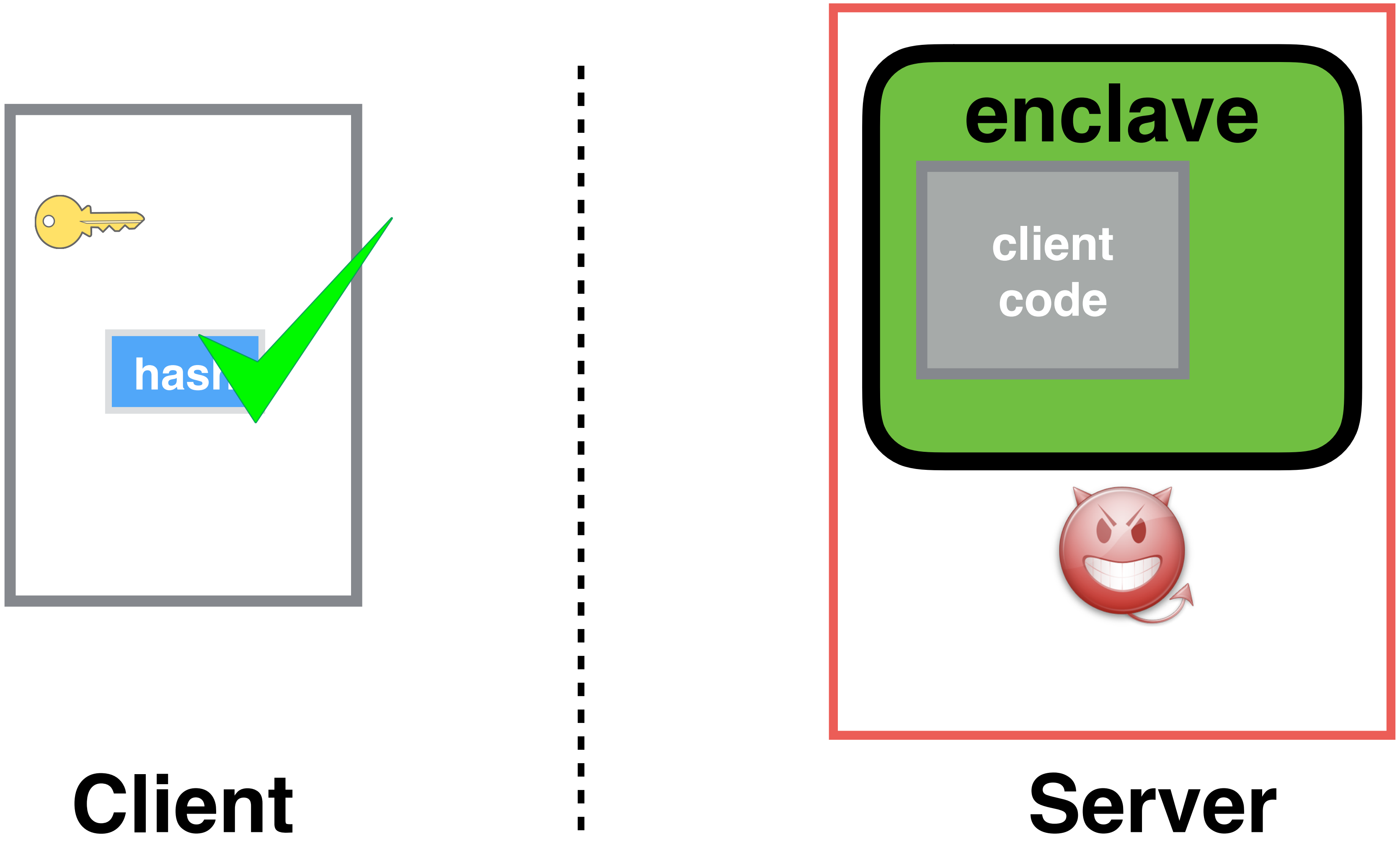
Remote attestation

Enables verifying which code runs in the enclave and performing key exchange



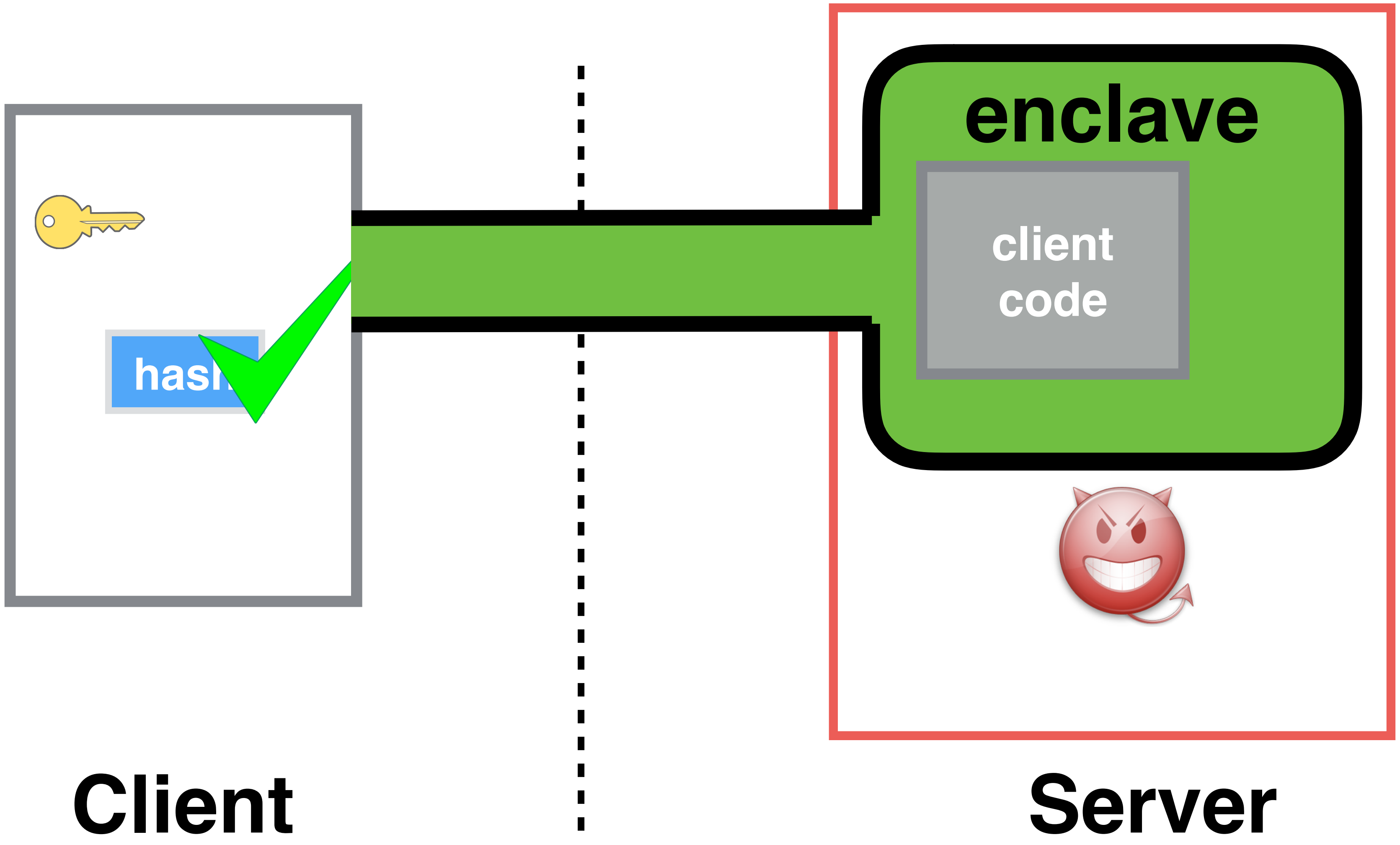
Remote attestation

Enables verifying which code runs in the enclave and performing key exchange



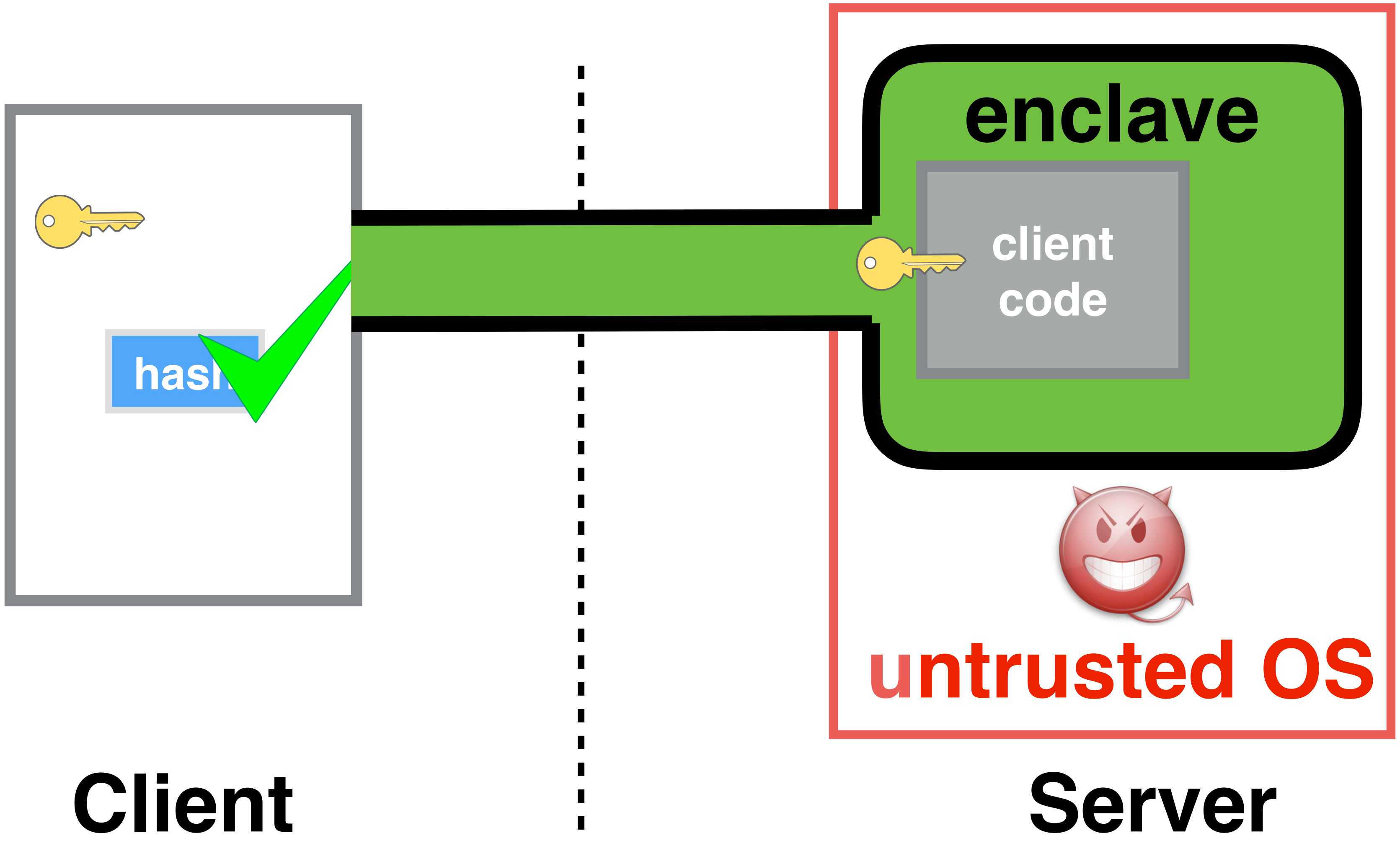
Remote attestation

Enables verifying which code runs in the enclave and performing key exchange



Remote attestation

Enables verifying which code runs in the enclave and performing key exchange



Enclave-based systems

Enclave-based systems

- Prior work: Haven [BMG ’14], VC3 [SCFGPMR ‘15]:

Enclave-based systems

- Prior work: Haven [BMG '14], VC3 [SCFGPMR '15]:
  - **full functionality**

Enclave-based systems

- Prior work: Haven [BMG '14], VC3 [SCFGPMR '15]:
  - **full functionality**
  - **great performance**

Enclave-based systems

- Prior work: Haven [BMG '14], VC3 [SCFGPMR '15]:
  - **full functionality**
  - **great performance**
  - **data access pattern leakage [XCP '15, OCFGKS '15]**

Problem: access pattern leakage [XCP ‘15, OCFGKS ‘15]

Problem: access pattern leakage [XCP ‘15, OCFGKS ‘15]

ID	Name	Age	Disease
🔒 12809	Amanda D. Edwards	40	Diabetes
🔒 29489	Robert R. McGowan	56	Diabetes
🔒 13744	Kimberly R. Seay	51	Cancer
🔒 18740	Dennis G. Bates	32	Diabetes
🔒 98329	Ronald S. Ogden	53	Cancer
32591	Donna R. Bridges	26	Diabetes

Problem: access pattern leakage [XCP '15, OCFGKS '15]

ID	Name	Age	Disease
🔒 12809	Amanda D. Edwards	40	Diabetes
🔒 29489	Robert R. McGowan	56	Diabetes
🔒 13744	Kimberly R. Seay	51	Cancer
🔒 18740	Dennis G. Bates	32	Diabetes
🔒 98329	Ronald S. Ogden	53	Cancer
🔒 32591	Donna R. Bridges	26	Diabetes

SELECT count(*) FROM medical
GROUP BY disease

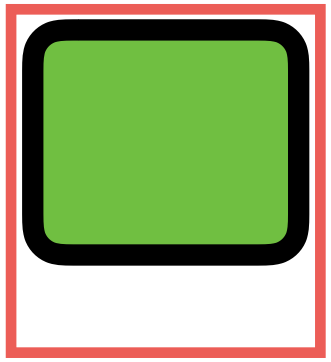
Problem: access pattern leakage [XCP ‘15, OCFGKS ‘15]

🔒 12809	…	Diabetes
🔒 29489	…	Diabetes
🔒 13744	…	Cancer

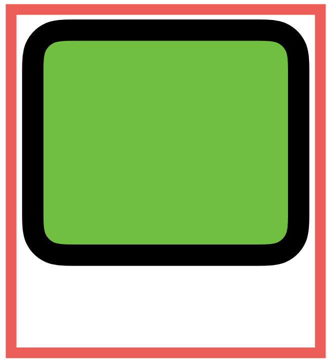
🔒 18740	…	Diabetes
🔒 98329	…	Cancer
🔒 32591	…	Diabetes

Problem: access pattern leakage [XCP '15, OCFGKS '15]

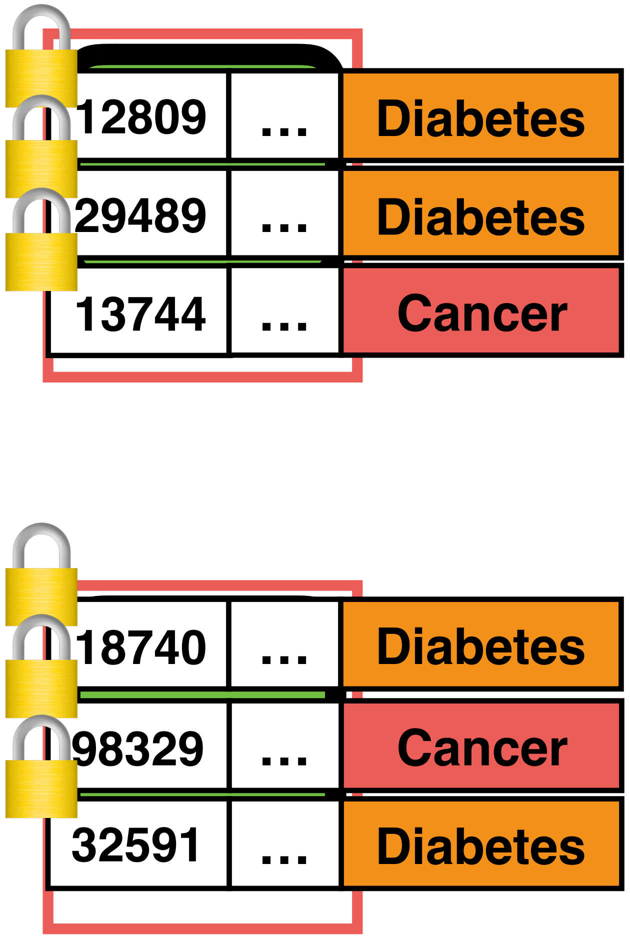
🔒	12809	…	Diabetes
🔒	29489	…	Diabetes
🔒	13744	…	Cancer



🔒	18740	…	Diabetes
🔒	98329	…	Cancer
🔒	32591	…	Diabetes



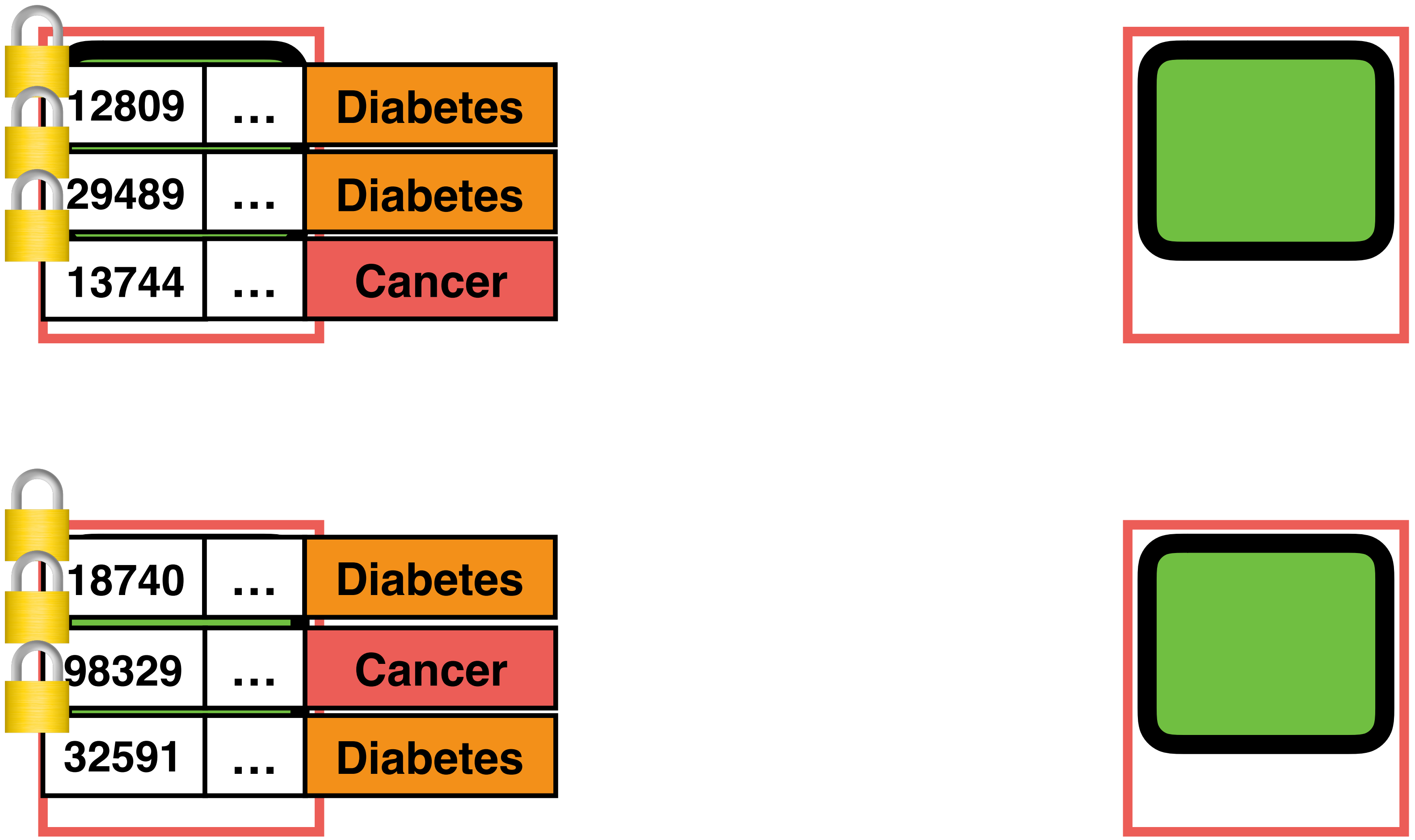
Problem: access pattern leakage [XCP '15, OCFGKS '15]



12809	…	Diabetes
29489	…	Diabetes
13744	…	Cancer

18740	…	Diabetes
98329	…	Cancer
32591	…	Diabetes

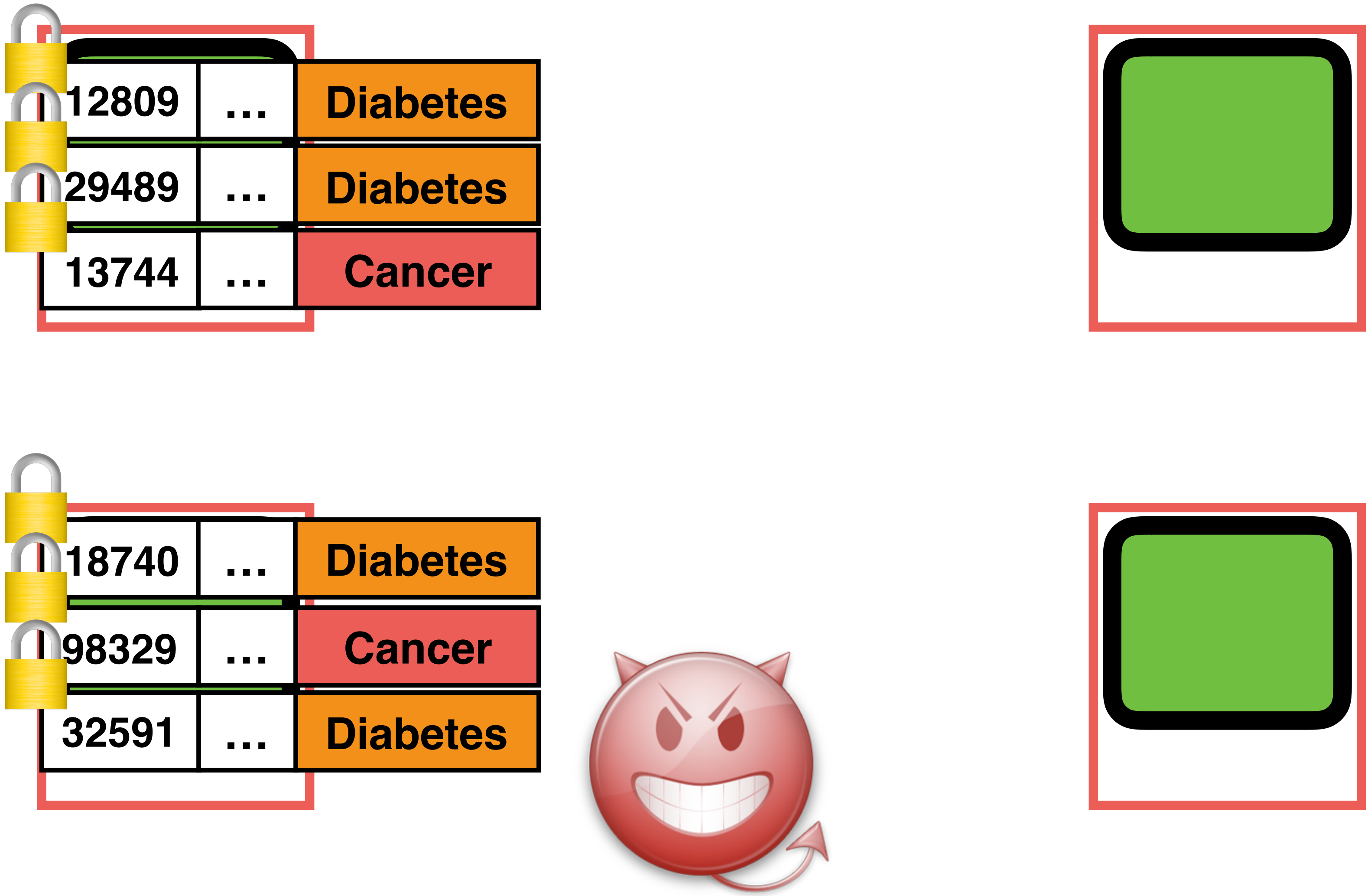
Problem: access pattern leakage [XCP ‘15, OCFGKS ‘15]



12809	…	Diabetes
29489	…	Diabetes
13744	…	Cancer

18740	…	Diabetes
98329	…	Cancer
32591	…	Diabetes

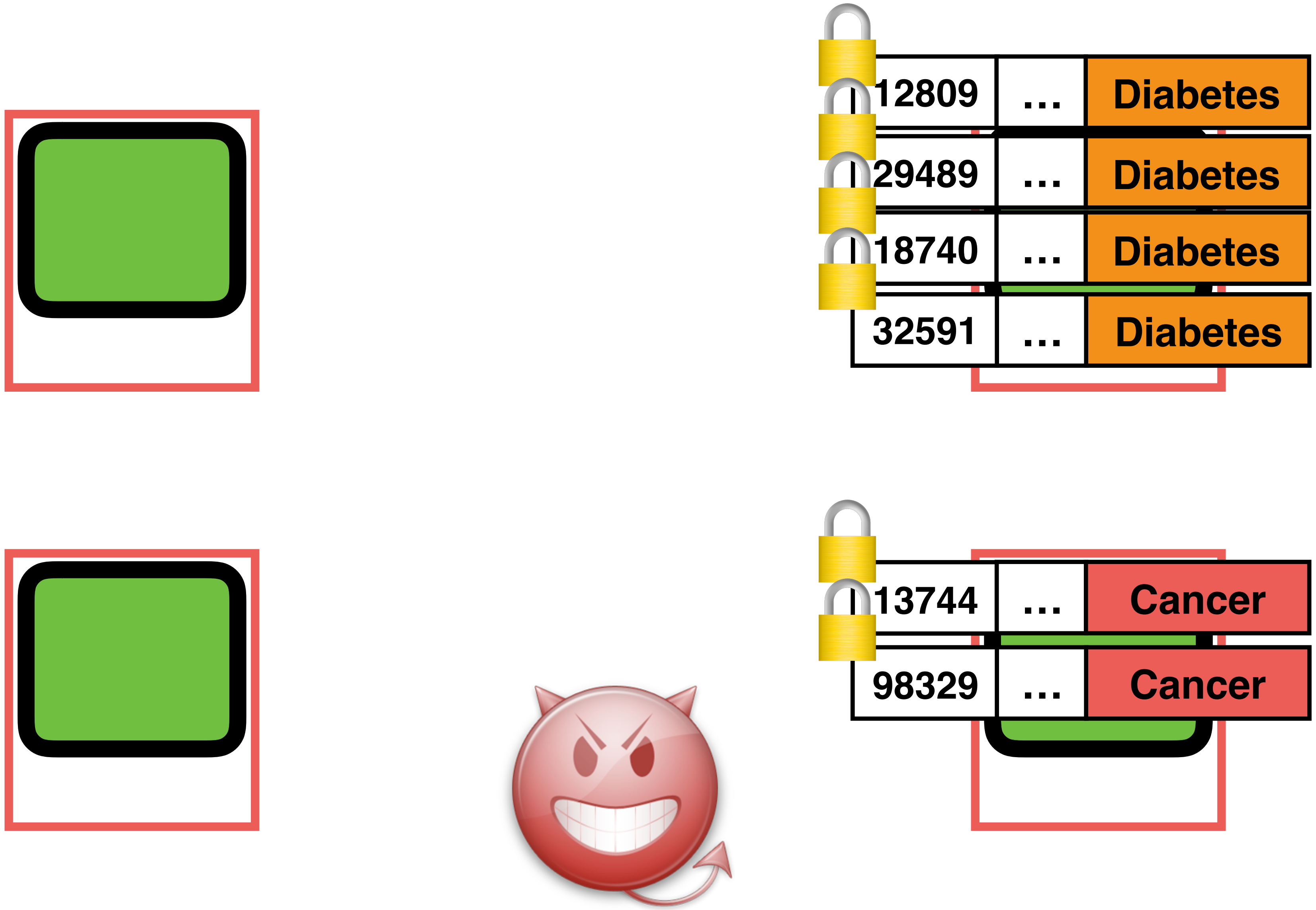
Problem: access pattern leakage [XCP ‘15, OCFGKS ‘15]



12809 … Diabetes
29489 … Diabetes
13744 … Cancer

18740 … Diabetes
98329 … Cancer
32591 … Diabetes

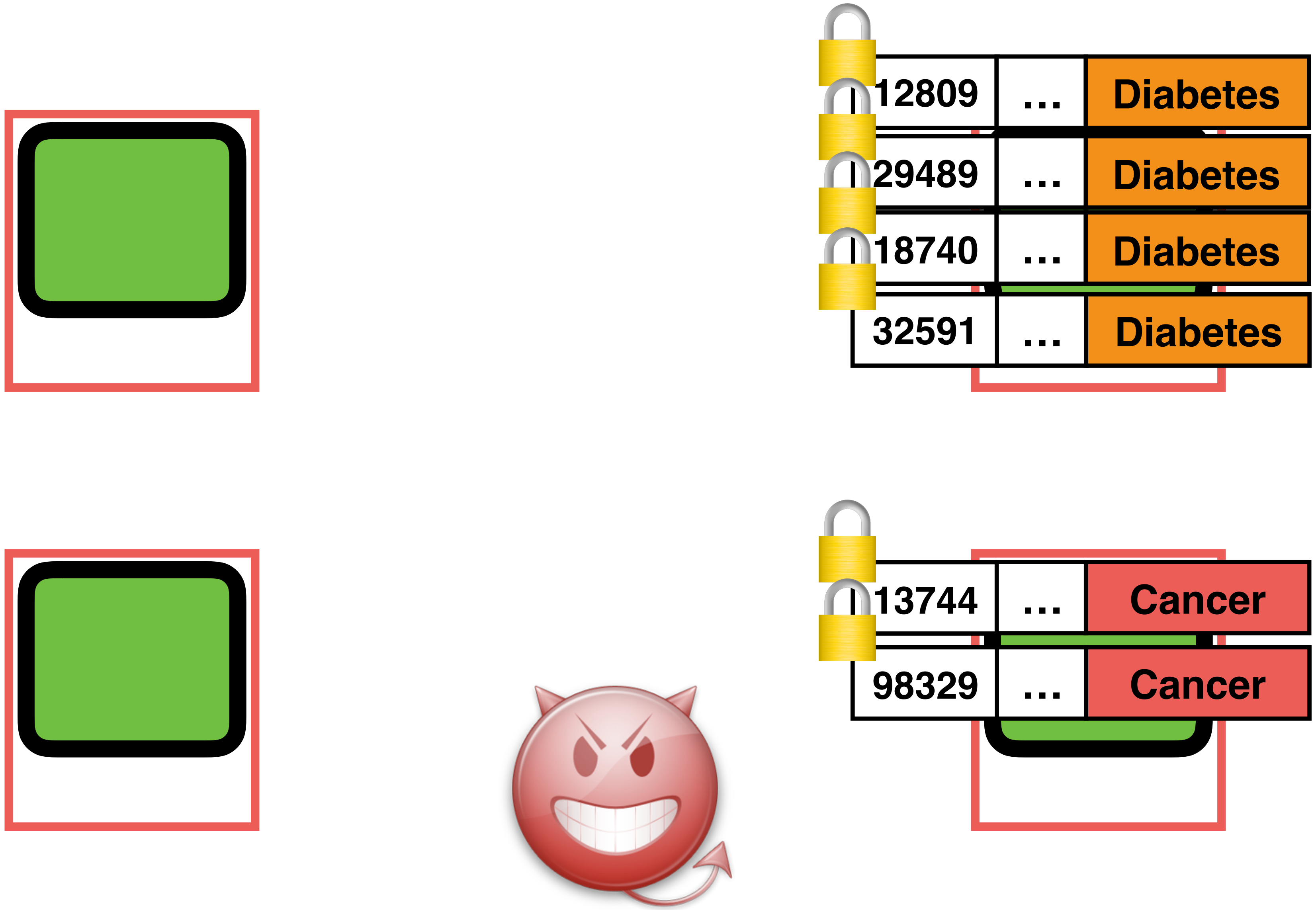
Problem: access pattern leakage [XCP '15, OCFGKS '15]



12809	…	Diabetes
29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

13744	…	Cancer
98329	…	Cancer

Problem: access pattern leakage [XCP ‘15, OCFGKS ‘15]

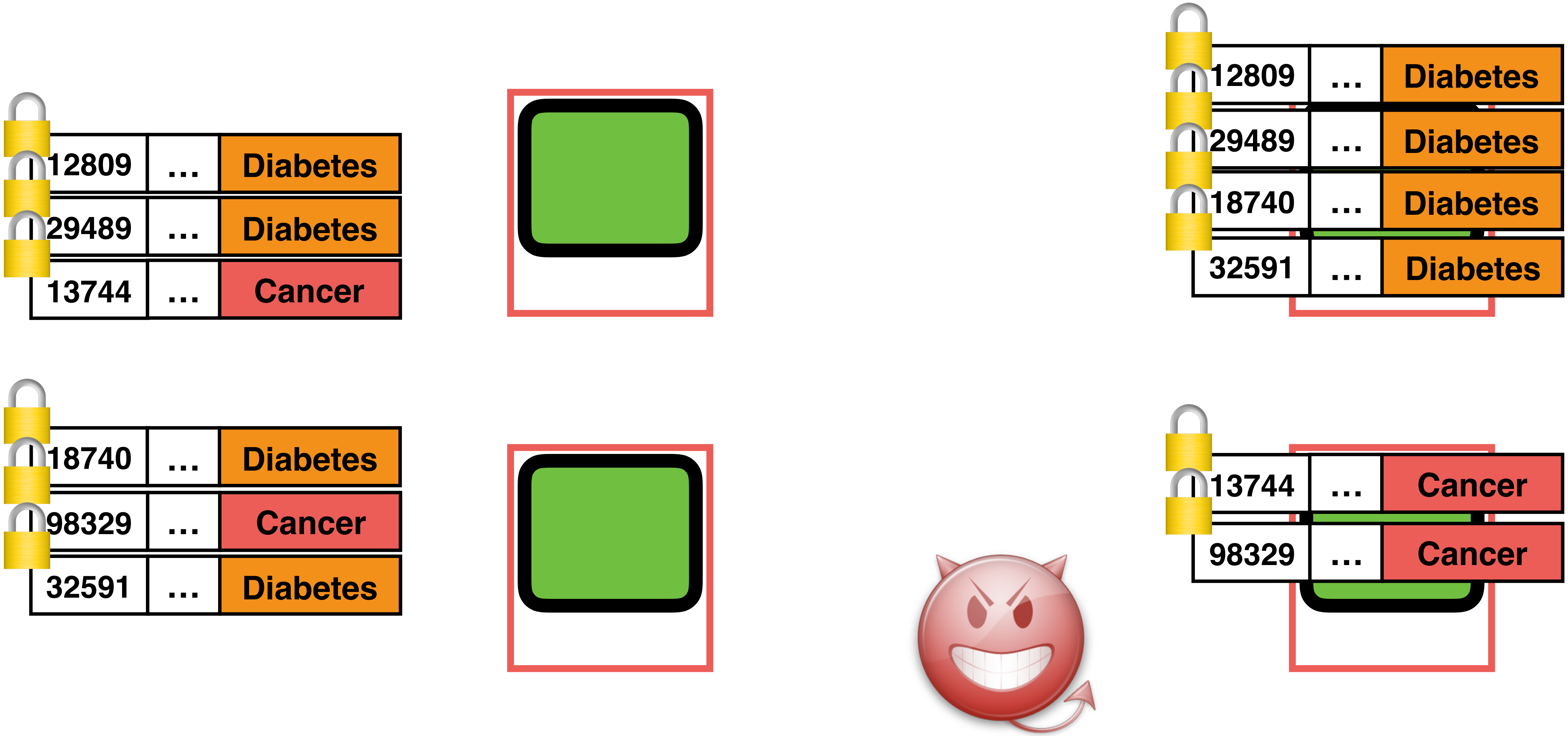


12809	…	Diabetes
29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

13744	…	Cancer
98329	…	Cancer

Public information:
Diabetes twice as
common
as cancer

Problem: access pattern leakage [XCP '15, OCFGKS '15]



12809	…	Diabetes
29489	…	Diabetes
13744	…	Cancer

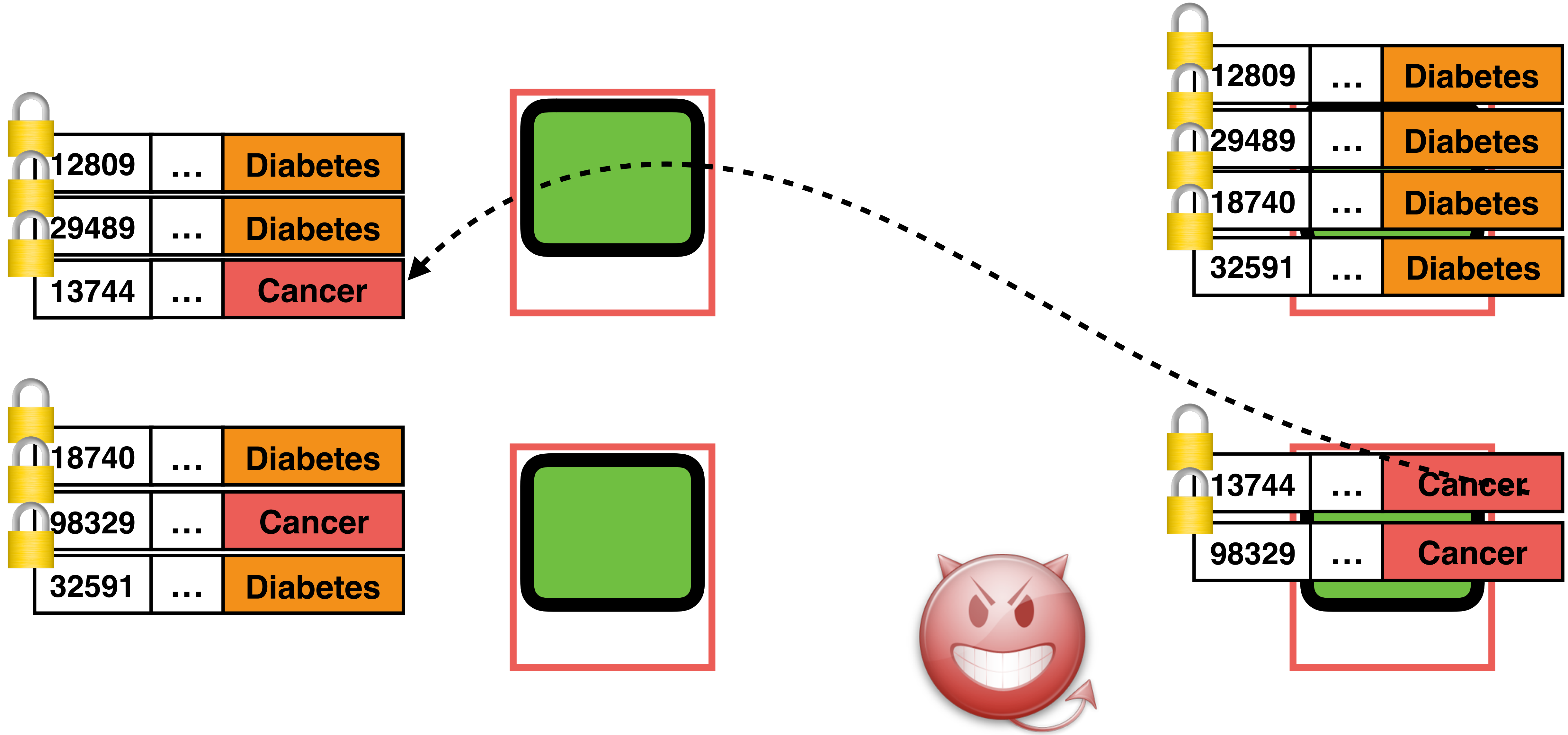
18740	…	Diabetes
98329	…	Cancer
32591	…	Diabetes

12809	…	Diabetes
29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

13744	…	Cancer
98329	…	Cancer

Public information: Diabetes twice as common as cancer

Problem: access pattern leakage [XCP ‘15, OCFGKS ‘15]



12809	…	Diabetes
29489	…	Diabetes
13744	…	Cancer

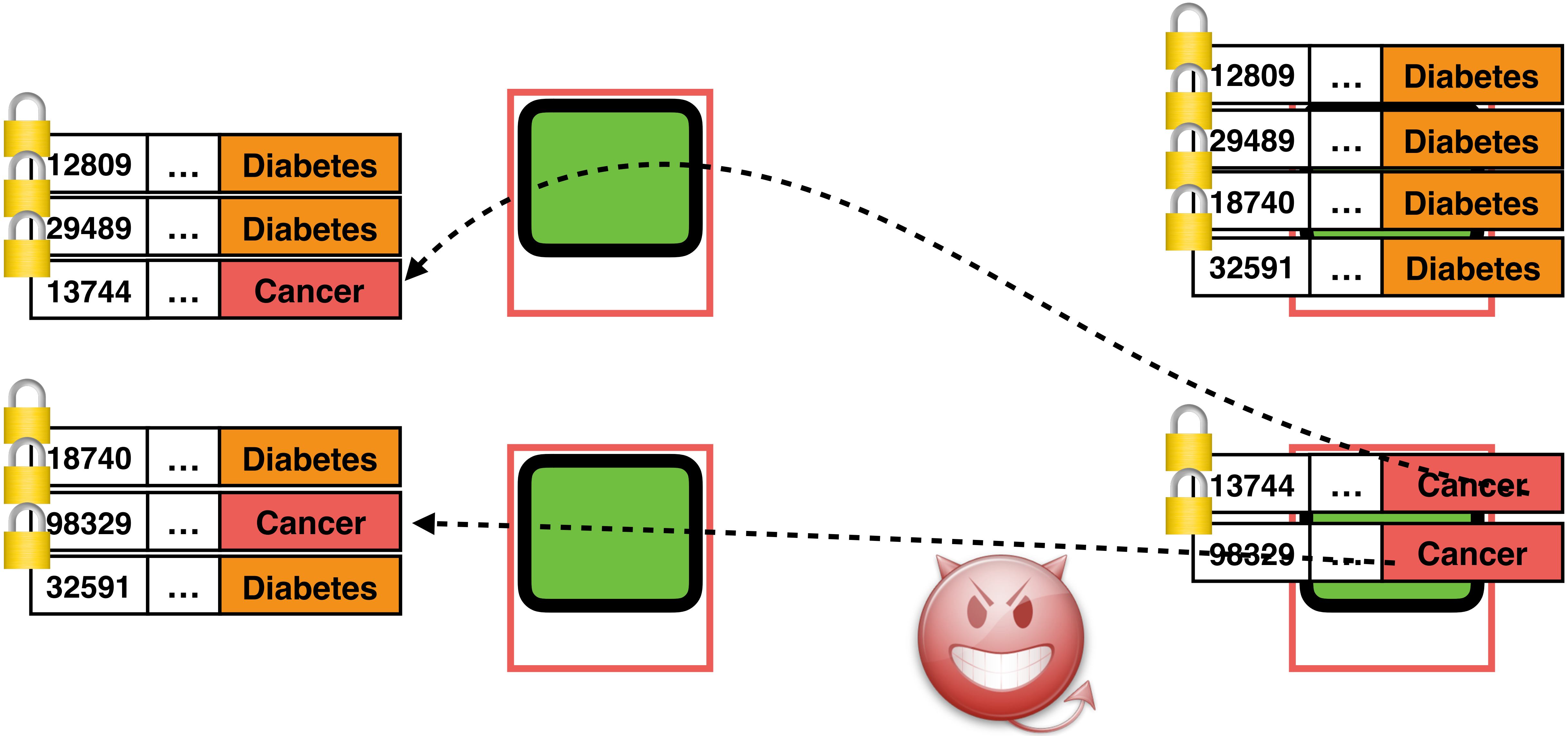
18740	…	Diabetes
98329	…	Cancer
32591	…	Diabetes

12809	…	Diabetes
29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

13744	…	Cancer
98329	…	Cancer

Public information:
Diabetes twice as common
as cancer

Problem: access pattern leakage [XCP ‘15, OCFGKS ‘15]



12809	…	Diabetes
29489	…	Diabetes
13744	…	Cancer

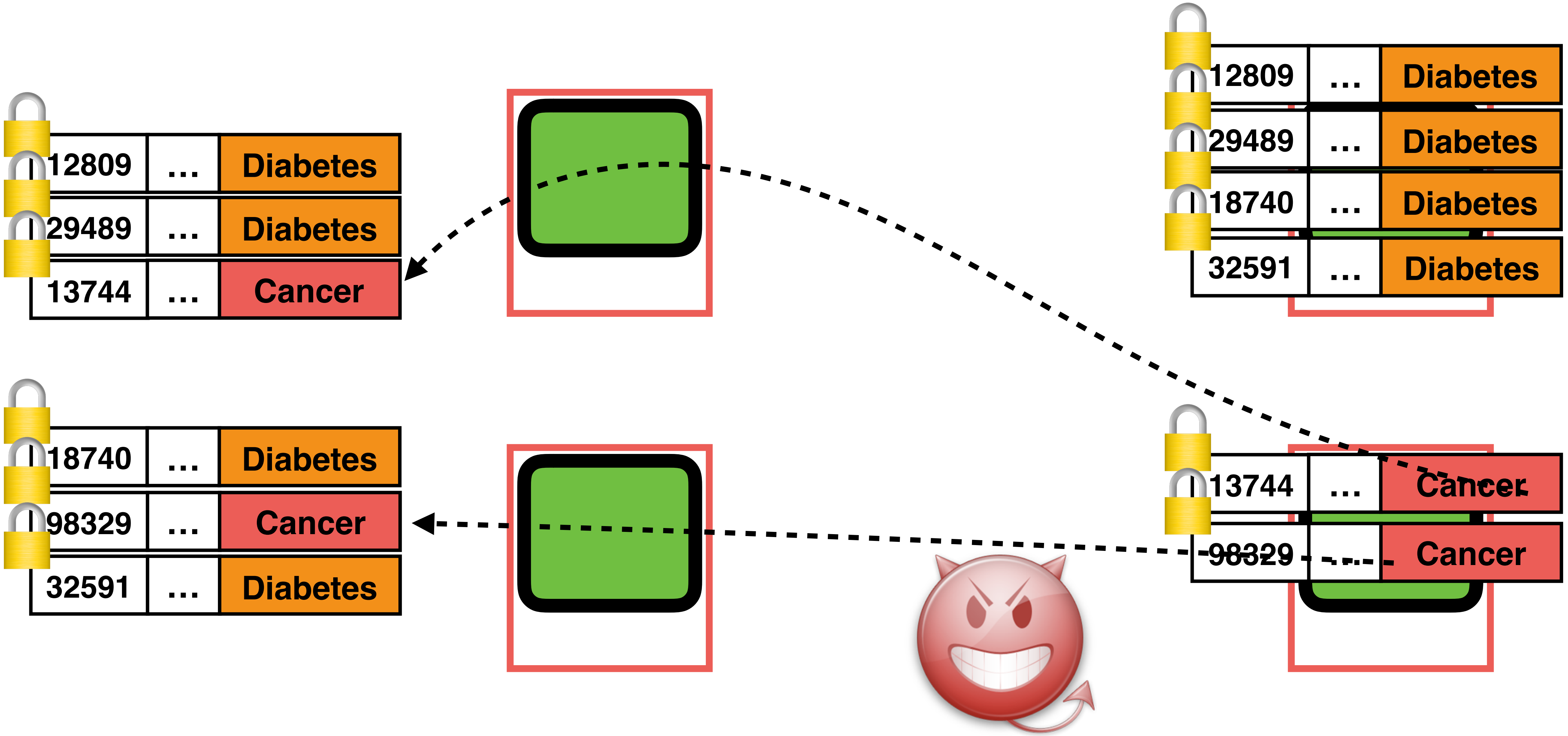
18740	…	Diabetes
98329	…	Cancer
32591	…	Diabetes

12809	…	Diabetes
29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

13744	…	Cancer
98329	…	Cancer

Public information: Diabetes twice as common as cancer

Problem: access pattern leakage [XCP '15, OCFGKS '15]



12809	…	Diabetes
29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

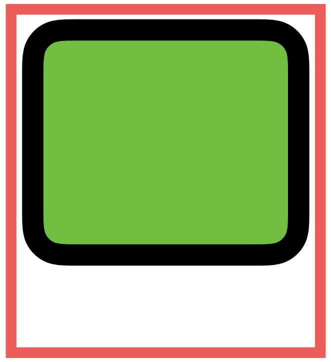
13744	…	Cancer
98329	…	Cancer

12809	…	Diabetes
29489	…	Diabetes
13744	…	Cancer

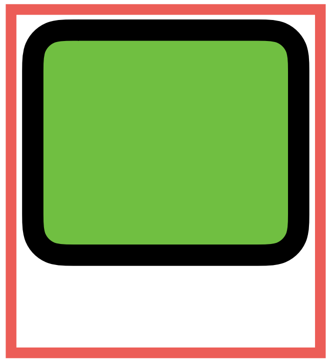
18740	…	Diabetes
98329	…	Cancer
32591	…	Diabetes

Problem: access pattern leakage [XCP '15, OCFGKS '15]

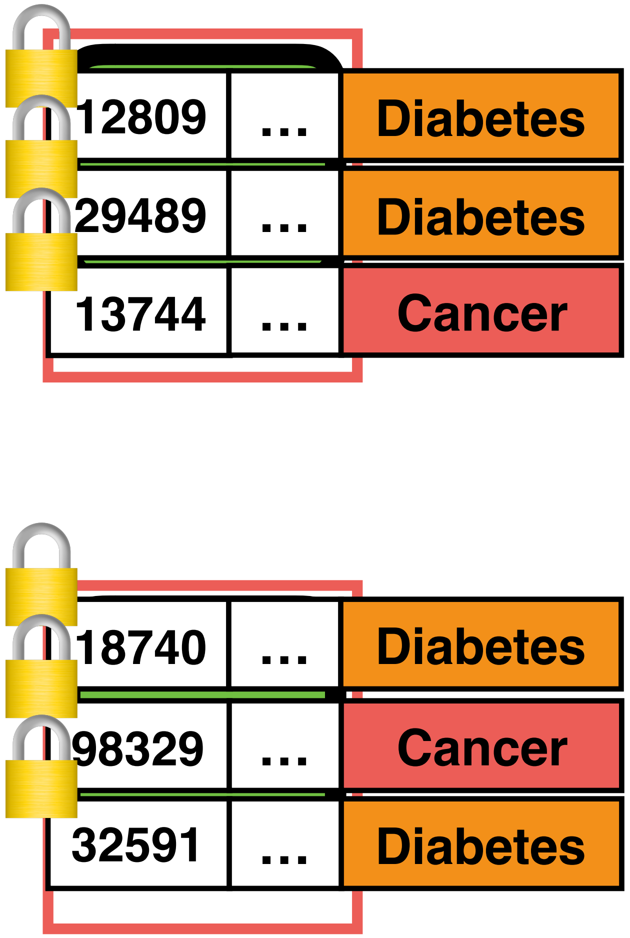
12809	…	Diabetes
29489	…	Diabetes
13744	…	Cancer



18740	…	Diabetes
98329	…	Cancer
32591	…	Diabetes



Problem: access pattern leakage [XCP ‘15, OCFGKS ‘15]



12809	…	Diabetes
29489	…	Diabetes
13744	…	Cancer

18740	…	Diabetes
98329	…	Cancer
32591	…	Diabetes

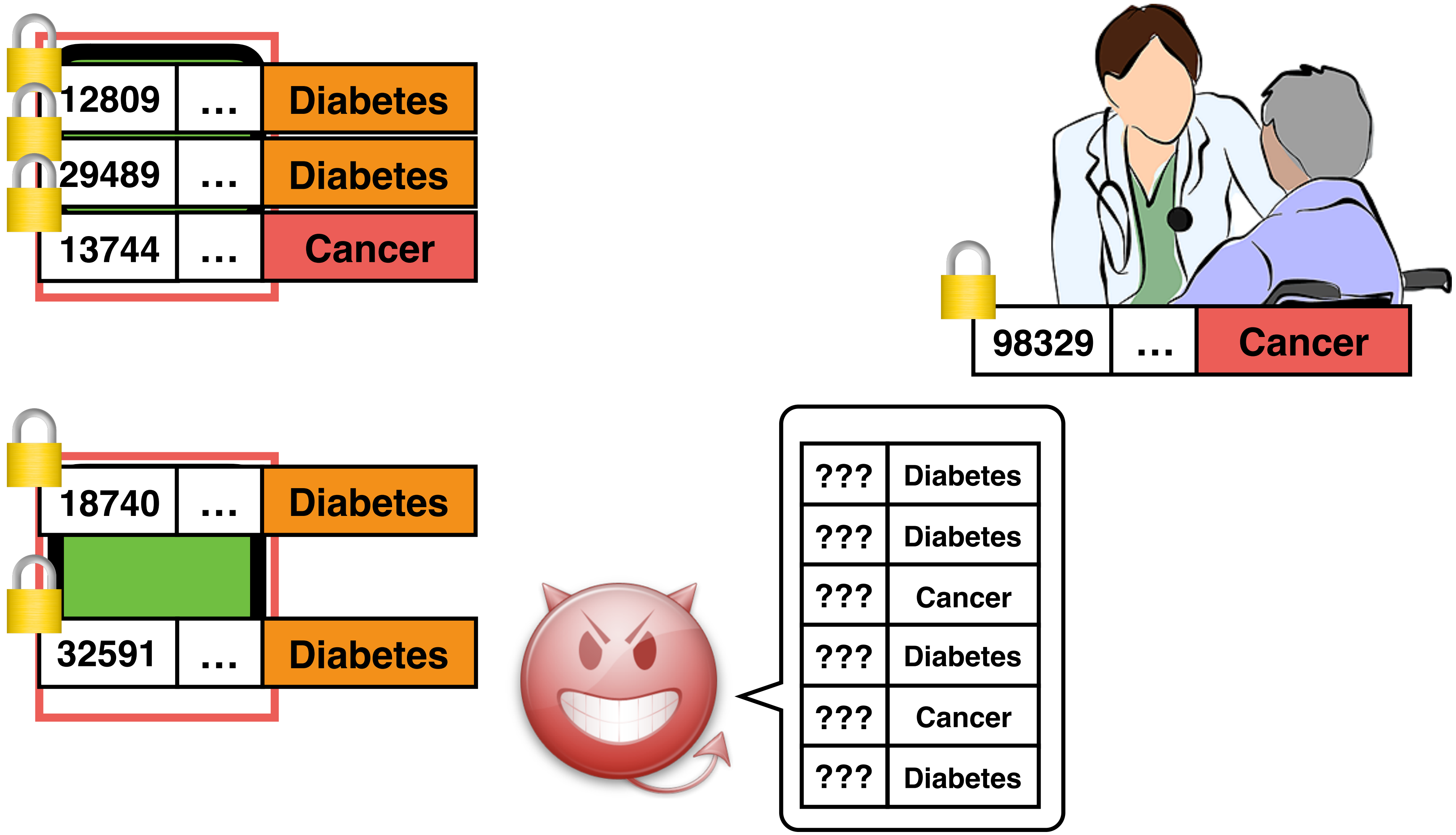
Problem: access pattern leakage [XCP '15, OCFGKS '15]

12809	…	Diabetes
29489	…	Diabetes
13744	…	Cancer

98329	…	Cancer

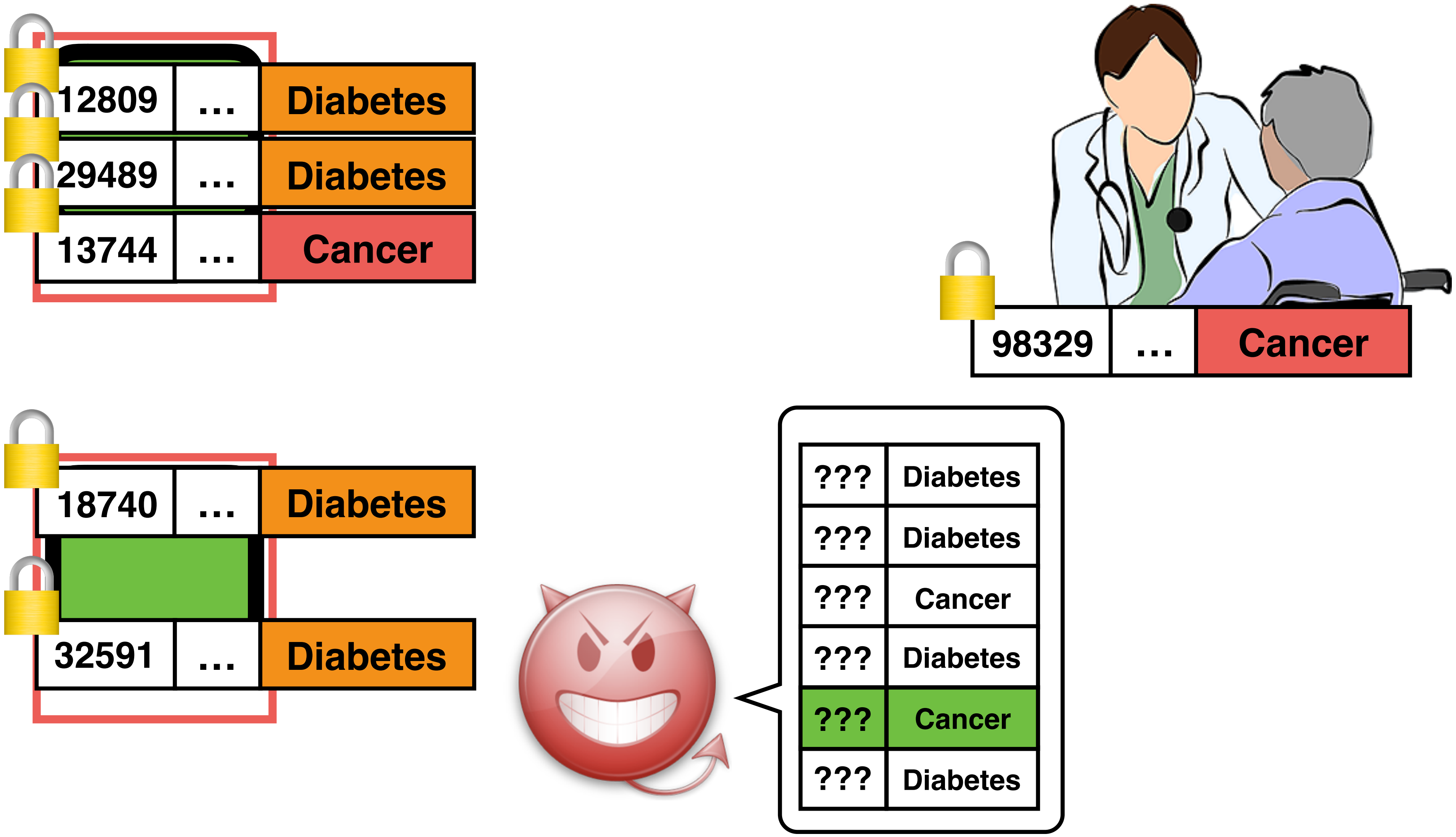
18740	…	Diabetes
🟩		
32591	…	Diabetes

Problem: access pattern leakage [XCP '15, OCFGKS '15]



???	Diabetes
???	Diabetes
???	Cancer
???	Diabetes
???	Cancer
???	Diabetes

Problem: access pattern leakage [XCP '15, OCFGKS '15]



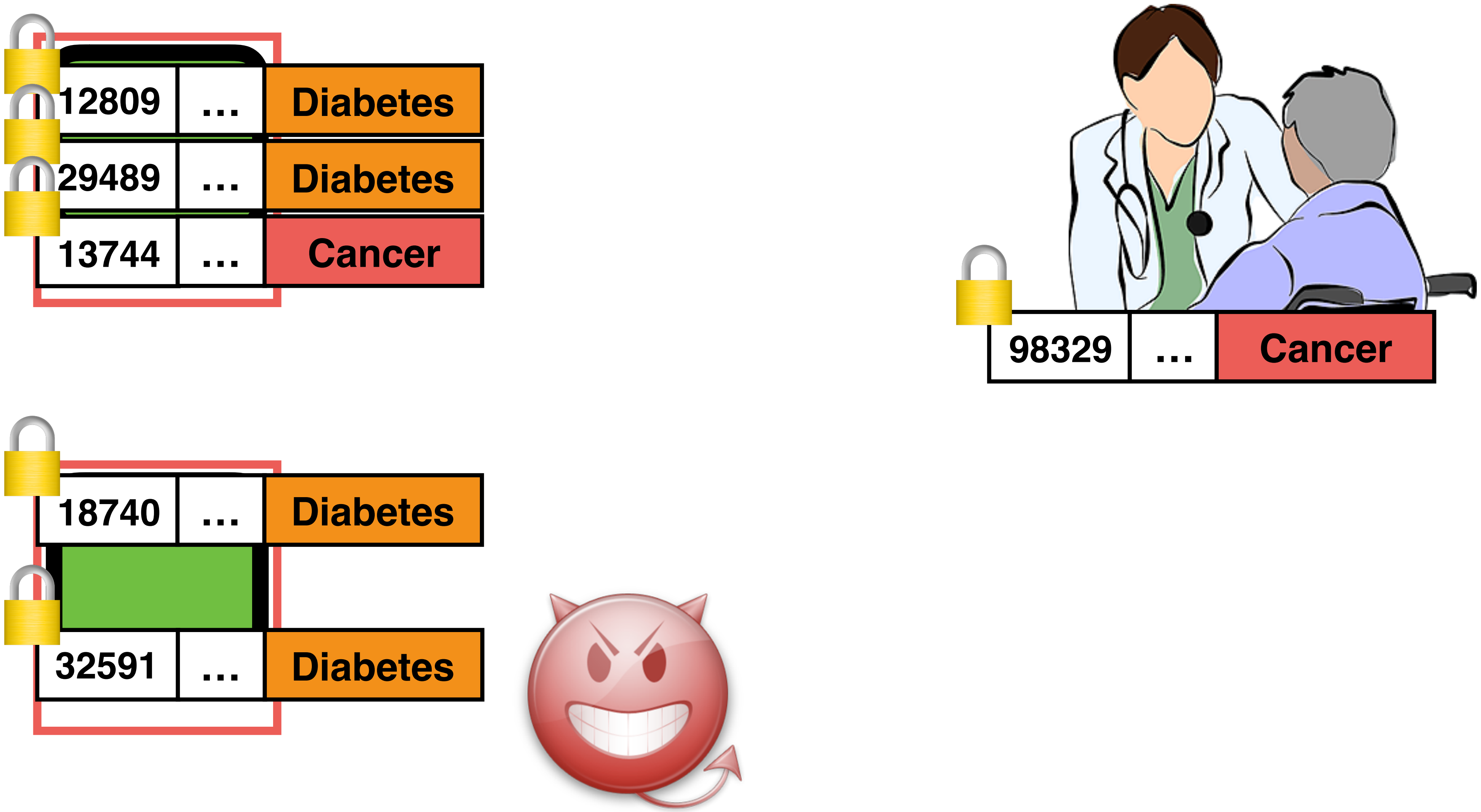
12809	…	Diabetes
29489	…	Diabetes
13744	…	Cancer

98329	…	Cancer

18740	…	Diabetes
32591	…	Diabetes

???	Diabetes
???	Diabetes
???	Cancer
???	Diabetes
???	Cancer
???	Diabetes

Problem: access pattern leakage [XCP ‘15, OCFGKS ‘15]



12809 … Diabetes

29489 … Diabetes

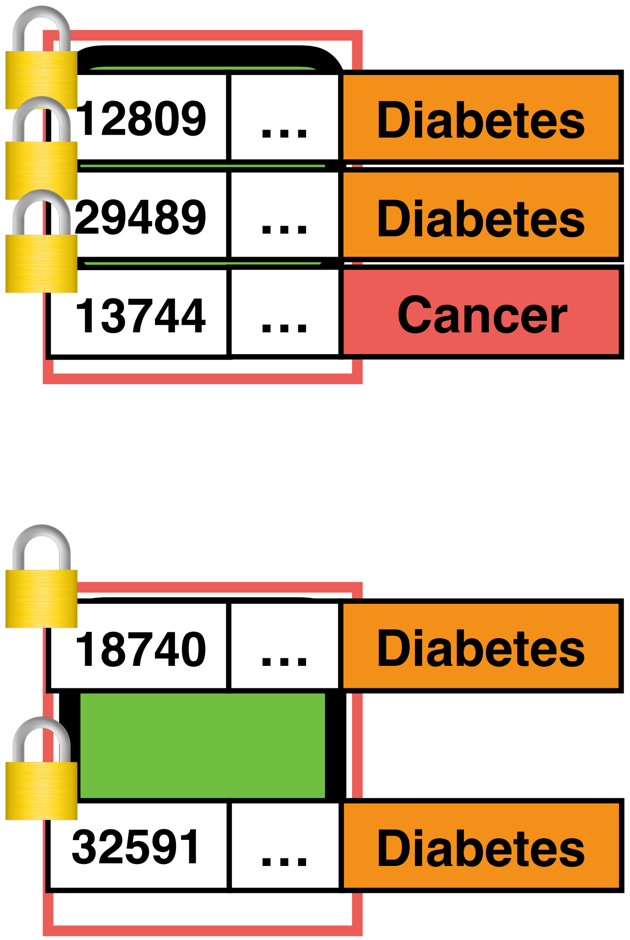
13744 … Cancer

98329 … Cancer

18740 … Diabetes

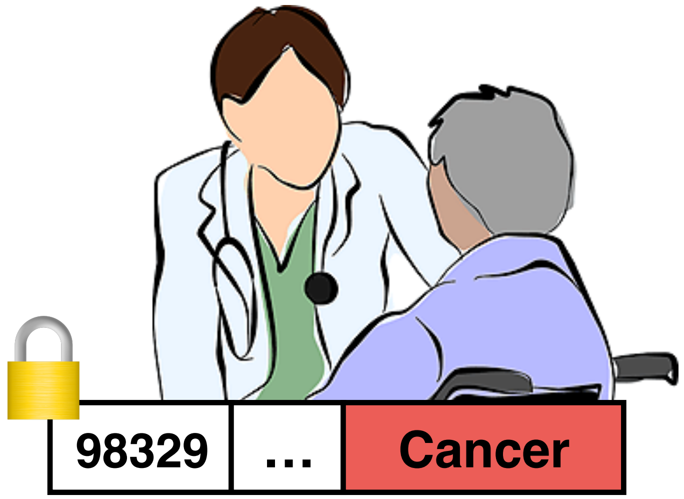
32591 … Diabetes

Problem: access pattern leakage [XCP '15, OCFGKS '15]



12809 … Diabetes
29489 … Diabetes
13744 … Cancer

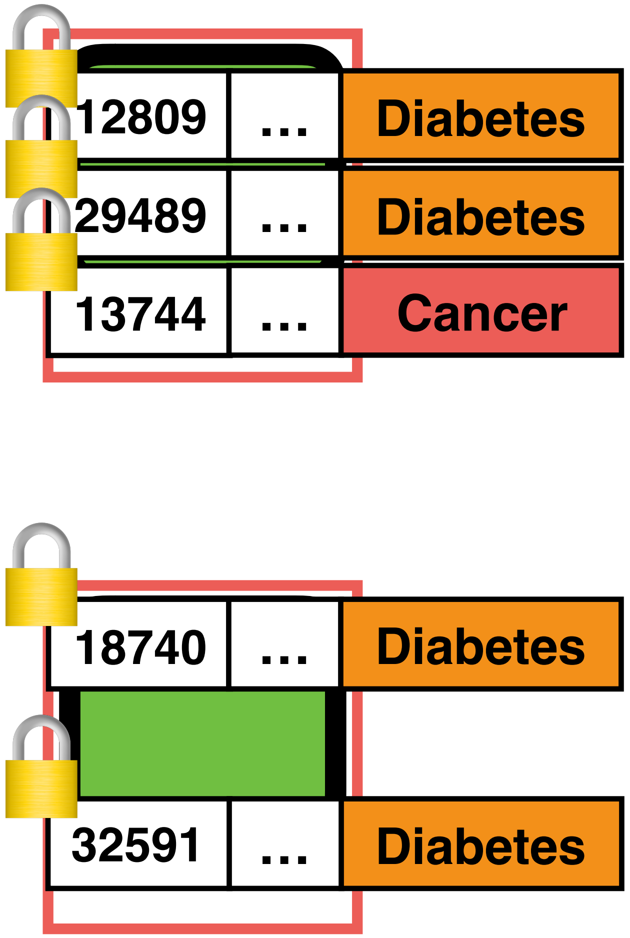
18740 … Diabetes
32591 … Diabetes



98329 … Cancer

Learns that the patient has cancer

Problem: access pattern leakage [XCP '15, OCFGKS '15]



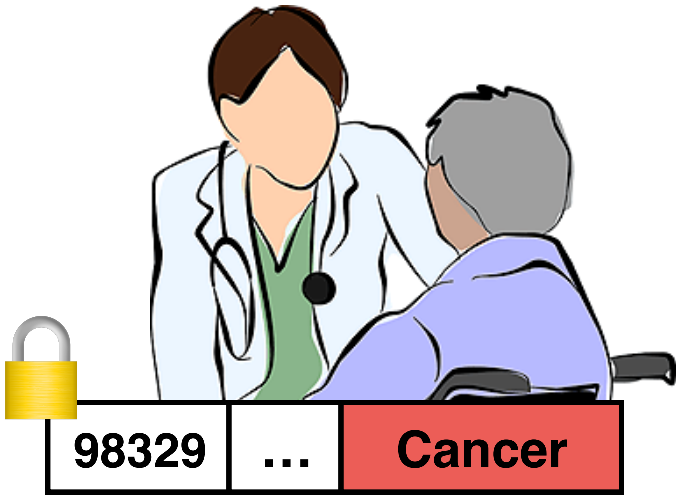
12809 … Diabetes

29489 … Diabetes

13744 … Cancer

18740 … Diabetes

32591 … Diabetes

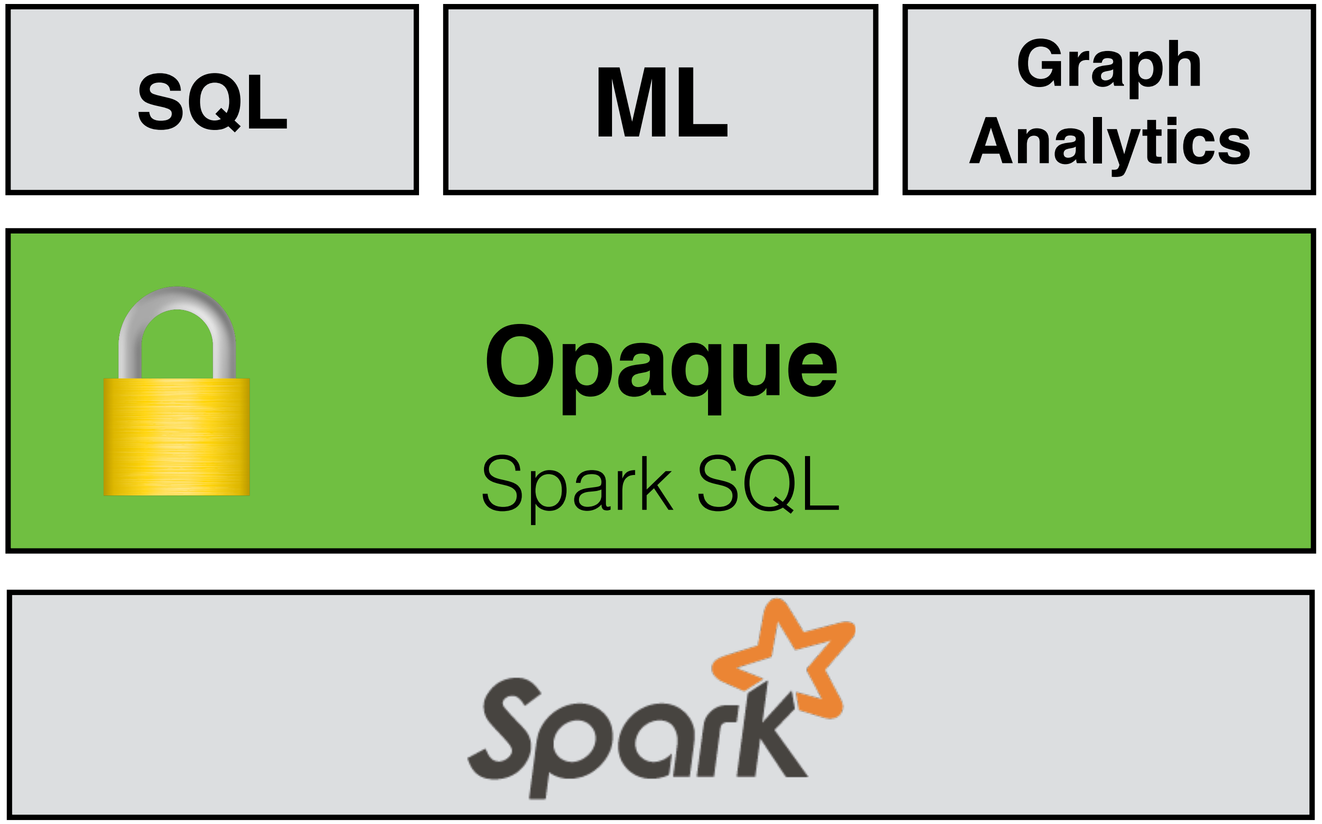


98329 … Cancer

Learns that the patient has cancer

Attack viable by observing both memory and network accesses!

Opaque*: secure distributed analytics



* Oblivious Platform for Analytic QUEries

Security guarantees

Security guarantees

- Data encryption and authentication

Security guarantees

- Data encryption and authentication
- Computation integrity: a check enforcing that the computation is executed correctly

Security guarantees

- Data encryption and authentication
- Computation integrity: a check enforcing that the computation is executed correctly
  - see paper for more!

Security guarantees

- Data encryption and authentication
- Computation integrity: a check enforcing that the computation is executed correctly
  - see paper for more!
- **Obliviousness = hiding access patterns**

Security guarantees

- Data encryption and authentication
- Computation integrity: a check enforcing that the computation is executed correctly
  - see paper for more!
- **Obliviousness = hiding access patterns**
  - *Informal statement*

Security guarantees

- Data encryption and authentication
- Computation integrity: a check enforcing that the computation is executed correctly
  - see paper for more!
- **Obliviousness = hiding access patterns**
  - *Informal statement*
    - *The memory and network accesses of the computation is the same for any input of the same size*

Challenge: obliviousness is expensive

Challenge: obliviousness is expensive

Two-part solution:

Challenge: obliviousness is expensive

Two-part solution:

Distributed oblivious SQL operators

Challenge: obliviousness is expensive

Two-part solution:

Distributed oblivious SQL operators

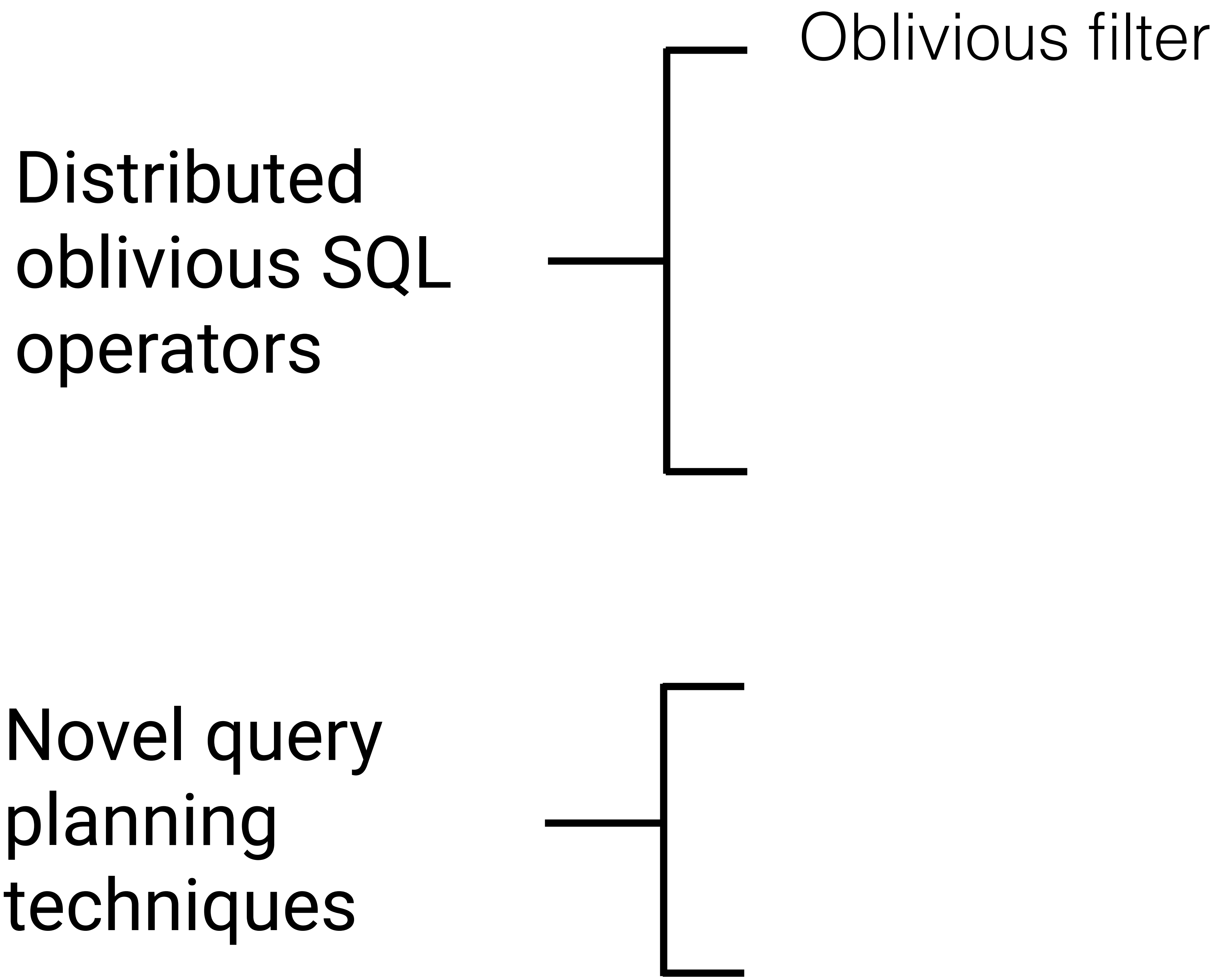
Novel query planning techniques

Two-part solution:

Distributed oblivious SQL operators

Novel query planning techniques

Two-part solution:



Distributed oblivious SQL operators

Oblivious filter

Novel query planning techniques

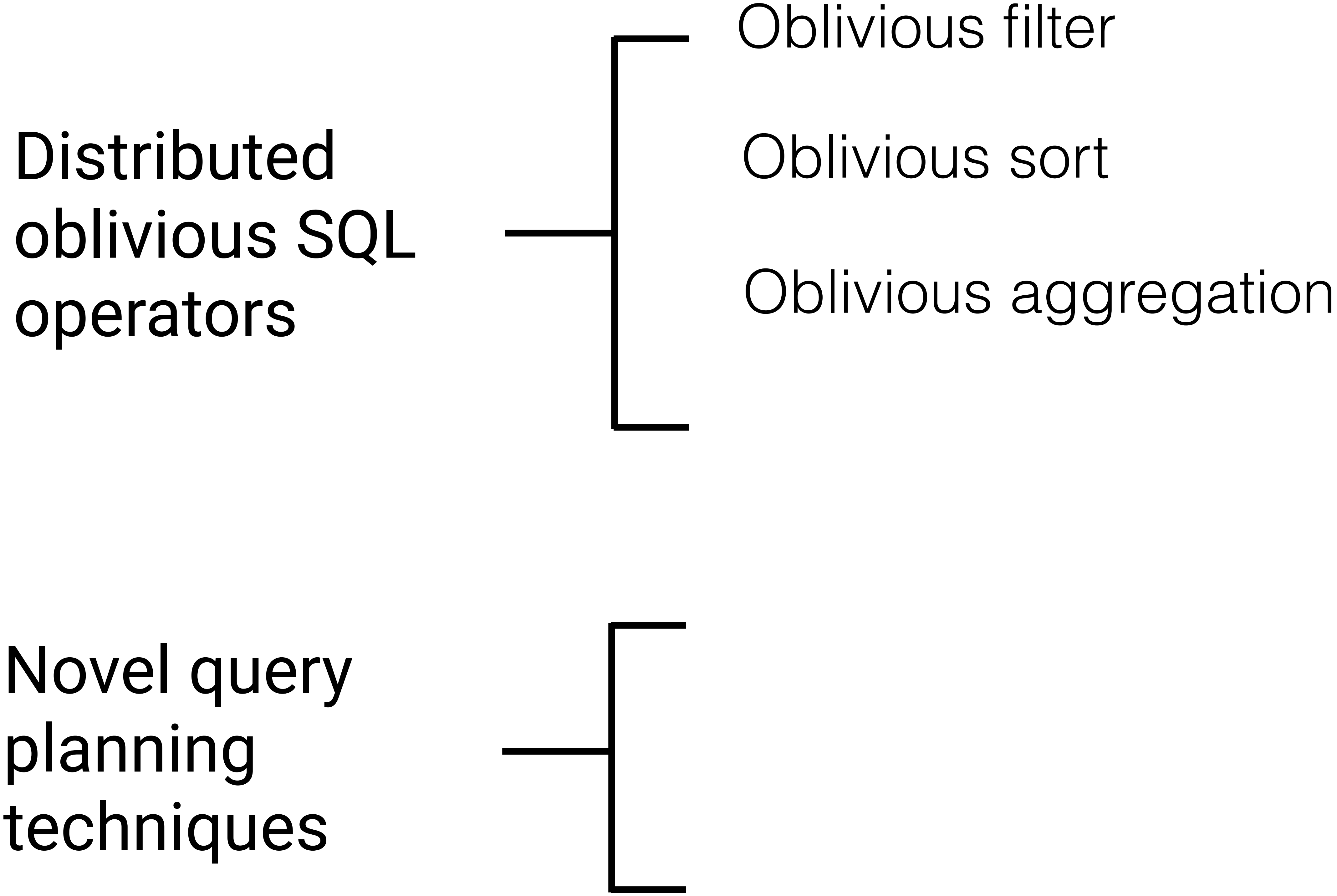
Two-part solution:

Distributed oblivious SQL operators

- Oblivious filter
- Oblivious sort

Novel query planning techniques

Two-part solution:



Distributed oblivious SQL operators

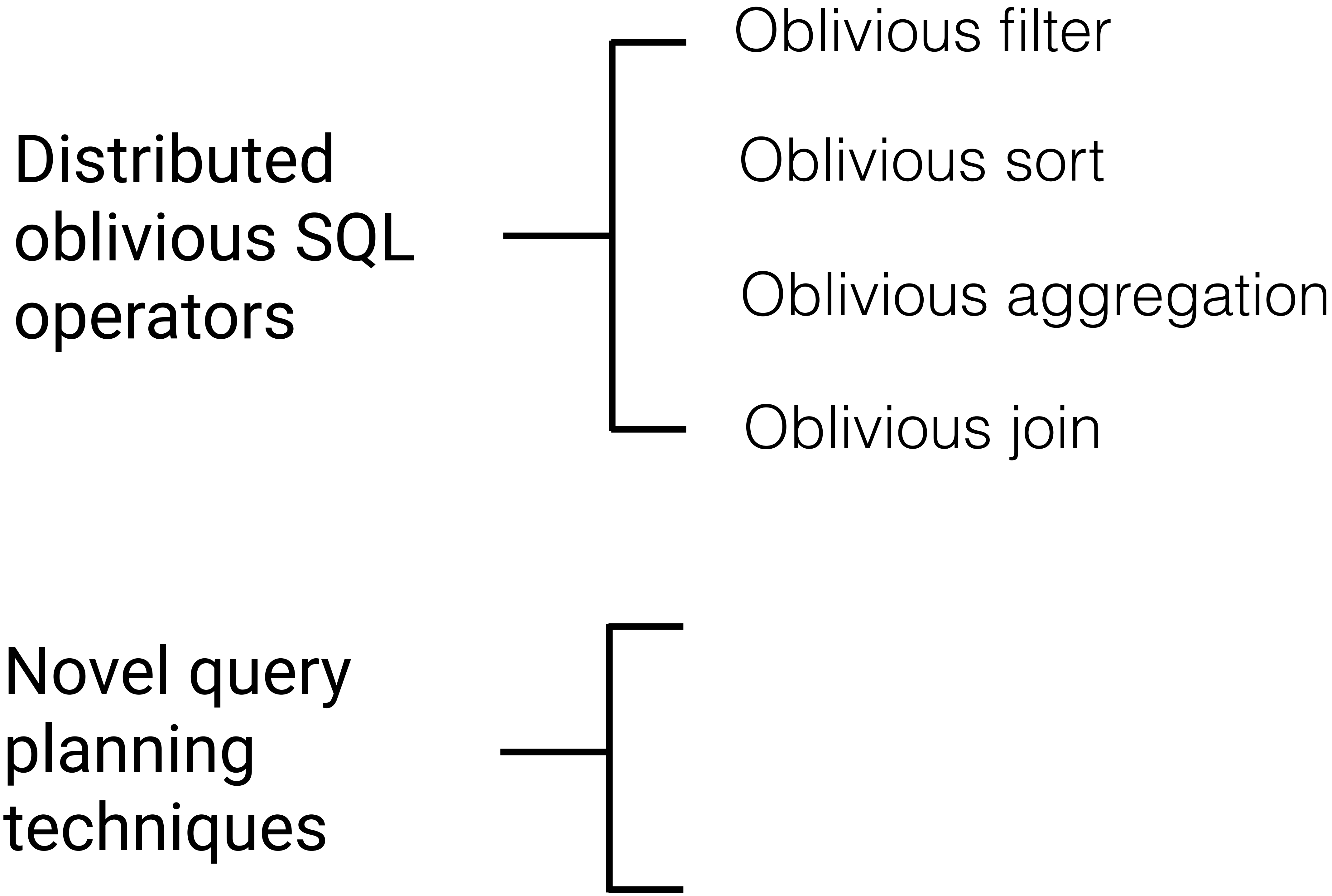
Oblivious filter

Oblivious sort

Oblivious aggregation

Novel query planning techniques

Two-part solution:



Distributed oblivious SQL operators

Oblivious filter

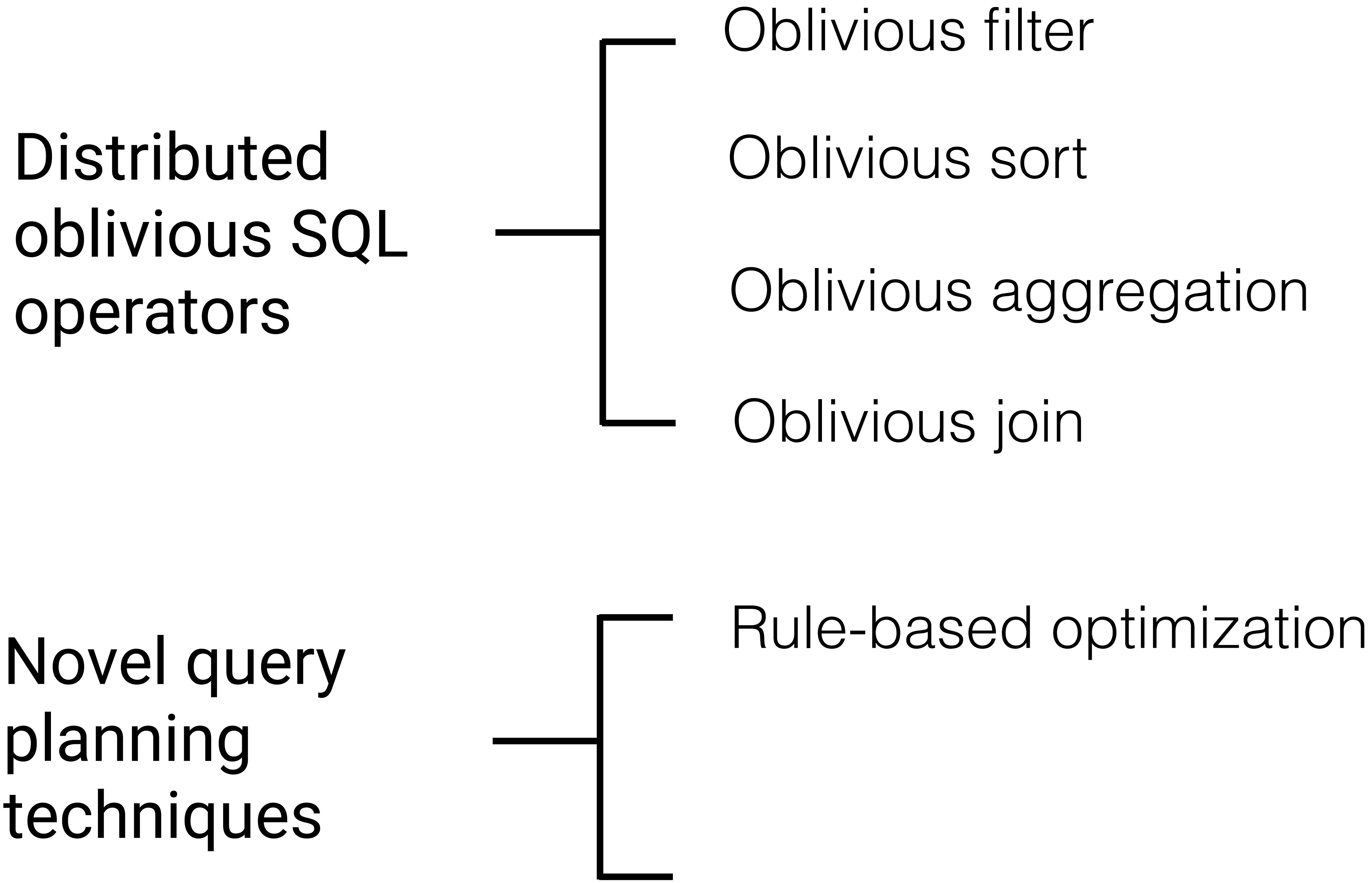
Oblivious sort

Oblivious aggregation

Oblivious join

Novel query planning techniques

Two-part solution:



Distributed oblivious SQL operators

Oblivious filter

Oblivious sort

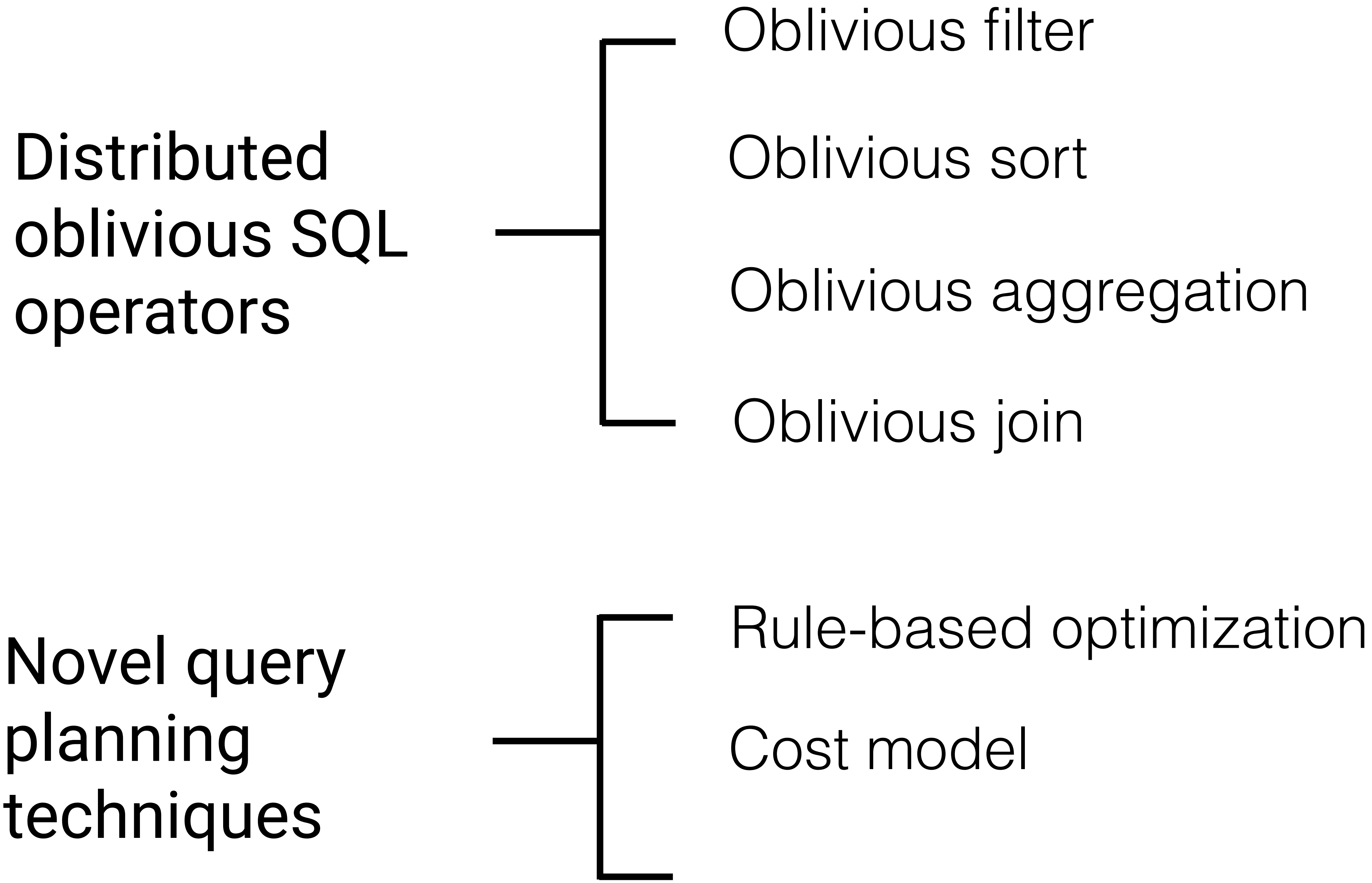
Oblivious aggregation

Oblivious join

Novel query planning techniques

Rule-based optimization

Two-part solution:



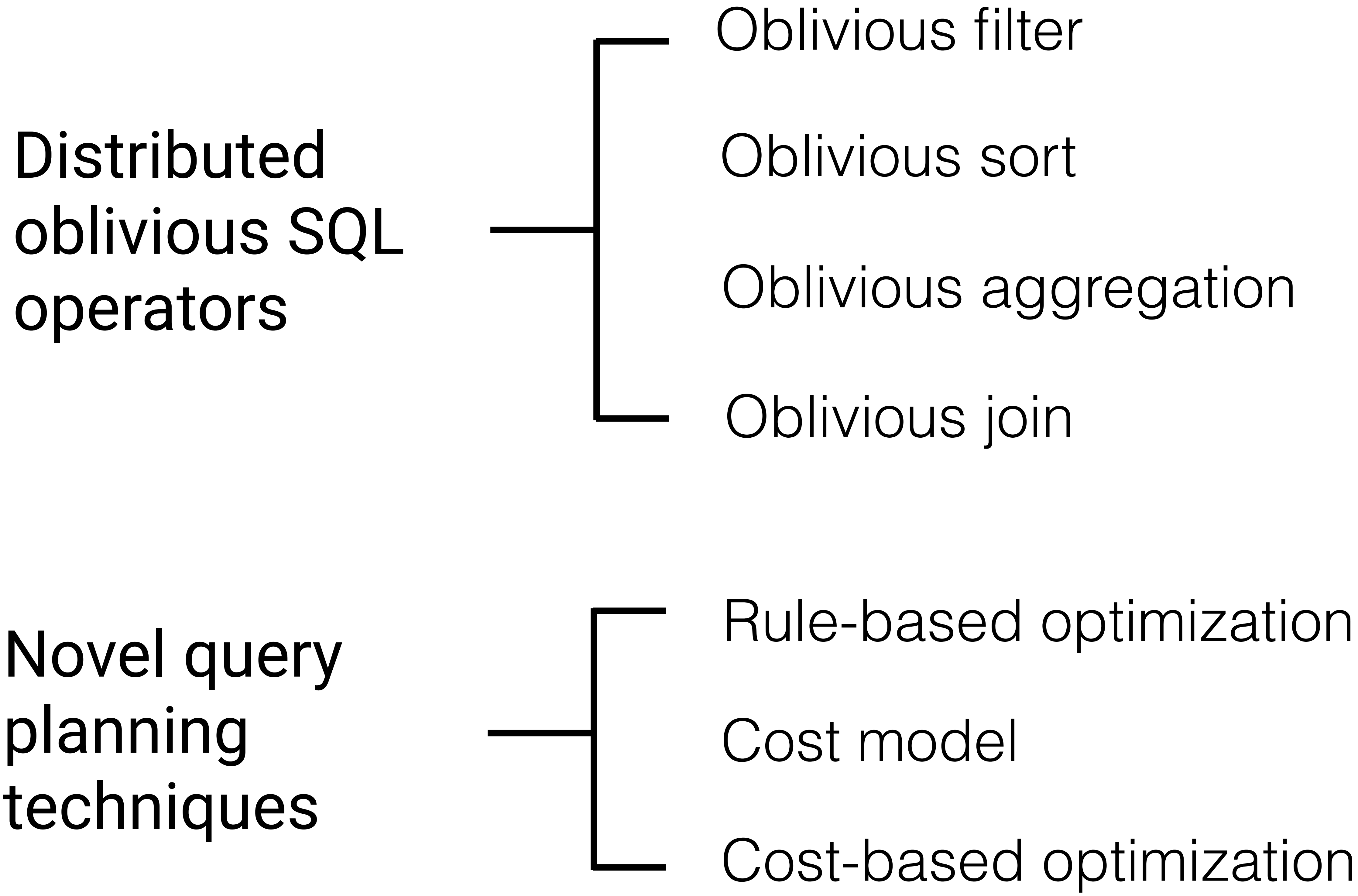
Distributed oblivious SQL operators

- Oblivious filter
- Oblivious sort
- Oblivious aggregation
- Oblivious join

Novel query planning techniques

- Rule-based optimization
- Cost model

Two-part solution:



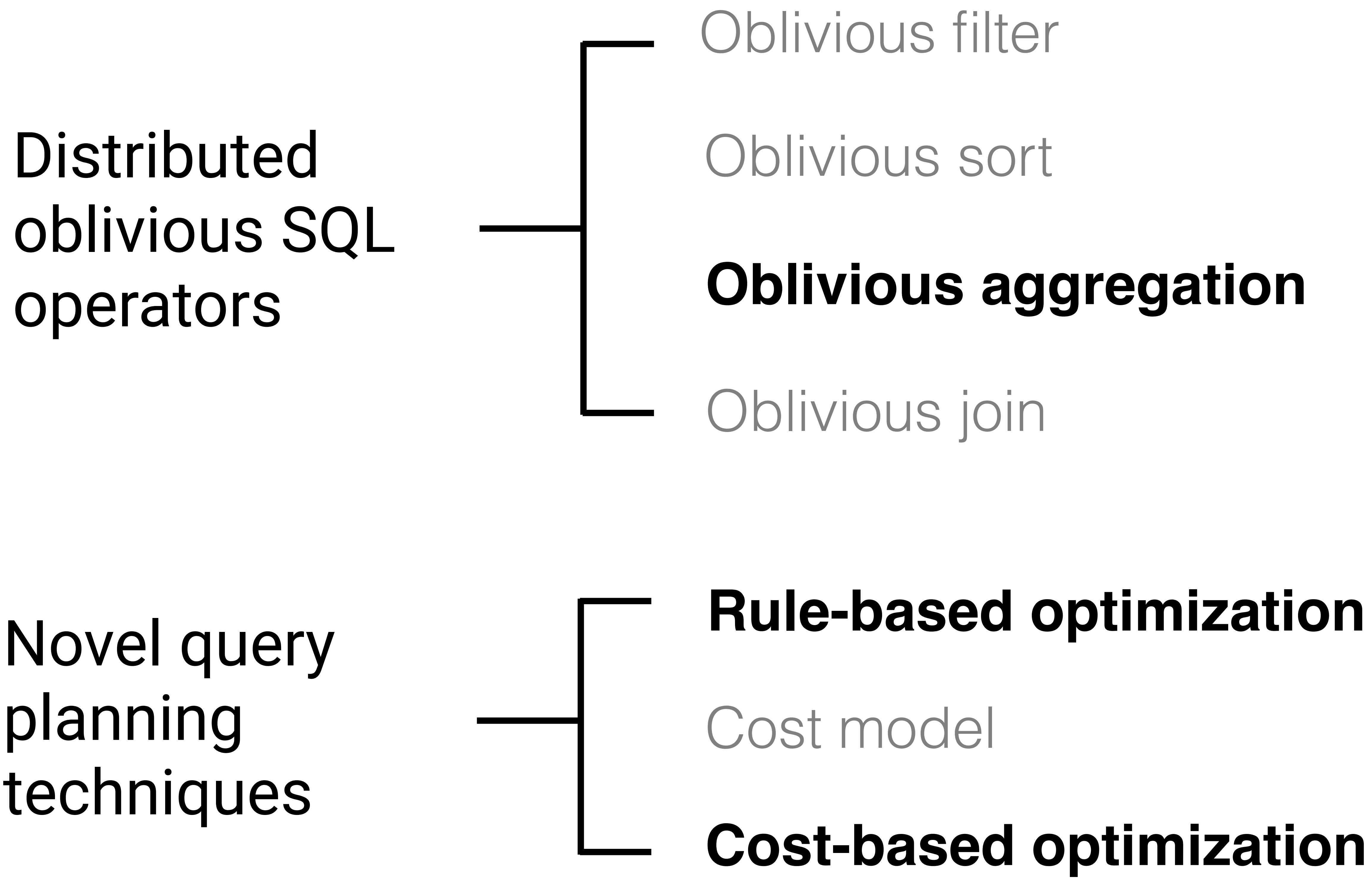
Distributed oblivious SQL operators

- Oblivious filter
- Oblivious sort
- Oblivious aggregation
- Oblivious join

Novel query planning techniques

- Rule-based optimization
- Cost model
- Cost-based optimization

Two-part solution:



Distributed oblivious SQL operators

- Oblivious filter
- Oblivious sort
- **Oblivious aggregation**
- Oblivious join

Novel query planning techniques

- **Rule-based optimization**
- Cost model
- **Cost-based optimization**

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease

🔒	12809	…	Diabetes
🔒	29489	…	Diabetes
🔒	13744	…	Cancer

🔒	18740	…	Diabetes
🔒	98329	…	Cancer
🔒	32591	…	Diabetes

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease

1

🔒	12809	…	Diabetes
🔒	29489	…	Diabetes
🔒	13744	…	Cancer

2

🔒	18740	…	Diabetes
🔒	98329	…	Cancer
🔒	32591	…	Diabetes

Oblivious aggregation

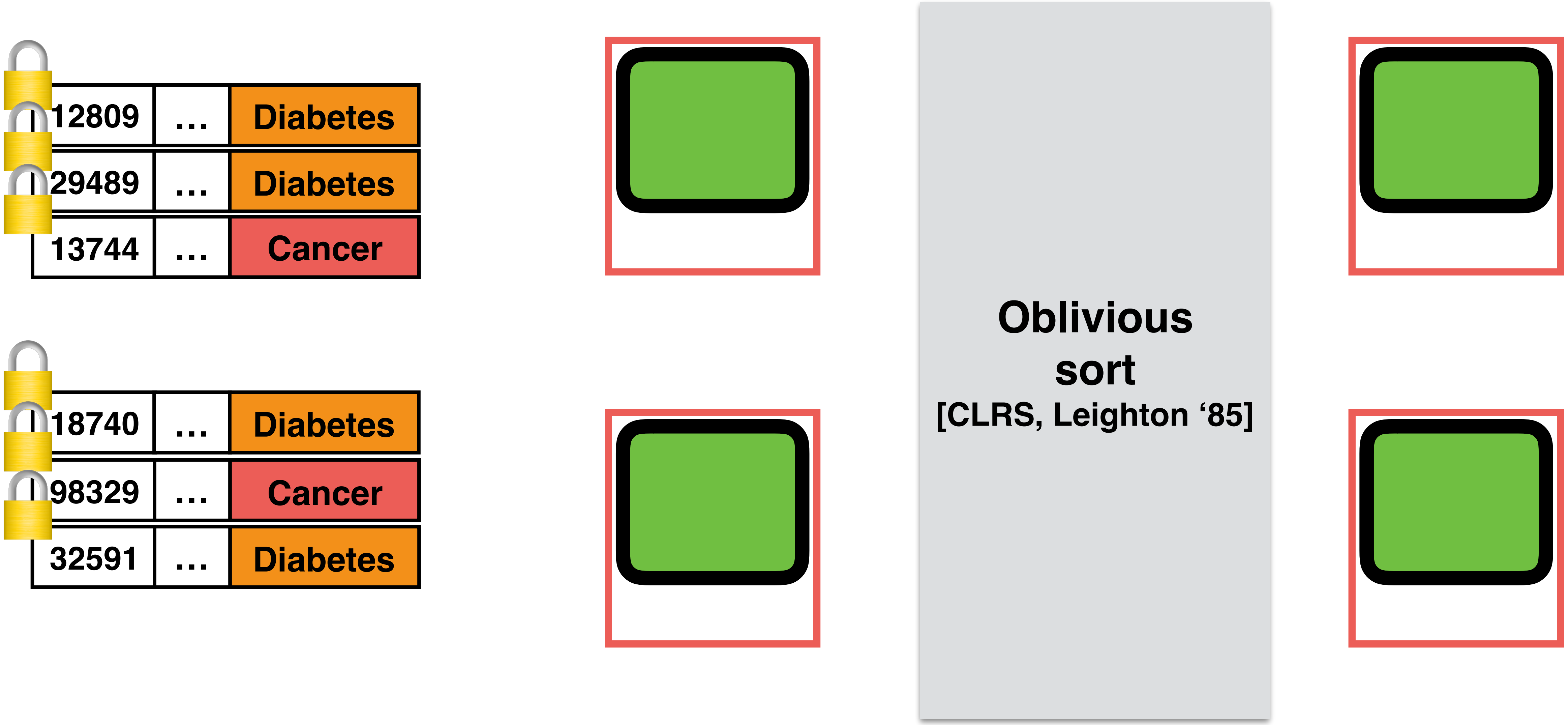
SELECT count(*) FROM medical GROUP BY disease

🔒	12809	…	Diabetes
🔒	29489	…	Diabetes
🔒	13744	…	Cancer

🔒	18740	…	Diabetes
🔒	98329	…	Cancer
🔒	32591	…	Diabetes

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



12809	…	Diabetes
29489	…	Diabetes
13744	…	Cancer

18740	…	Diabetes
98329	…	Cancer
32591	…	Diabetes

Oblivious sort

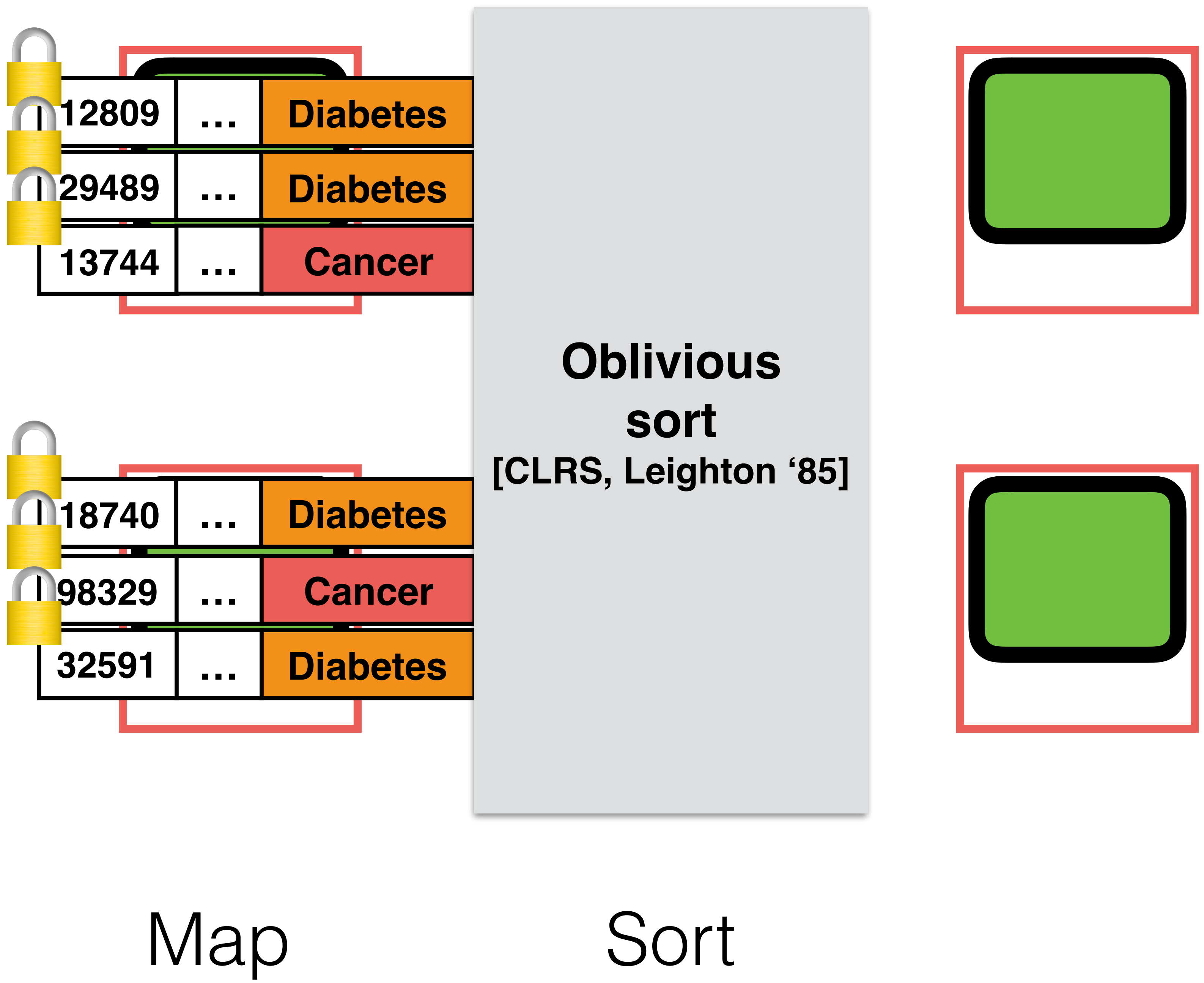
[CLRS, Leighton ‘85]

Map

Sort

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



12809	…	Diabetes
29489	…	Diabetes
13744	…	Cancer

18740	…	Diabetes
98329	…	Cancer
32591	…	Diabetes

Oblivious sort

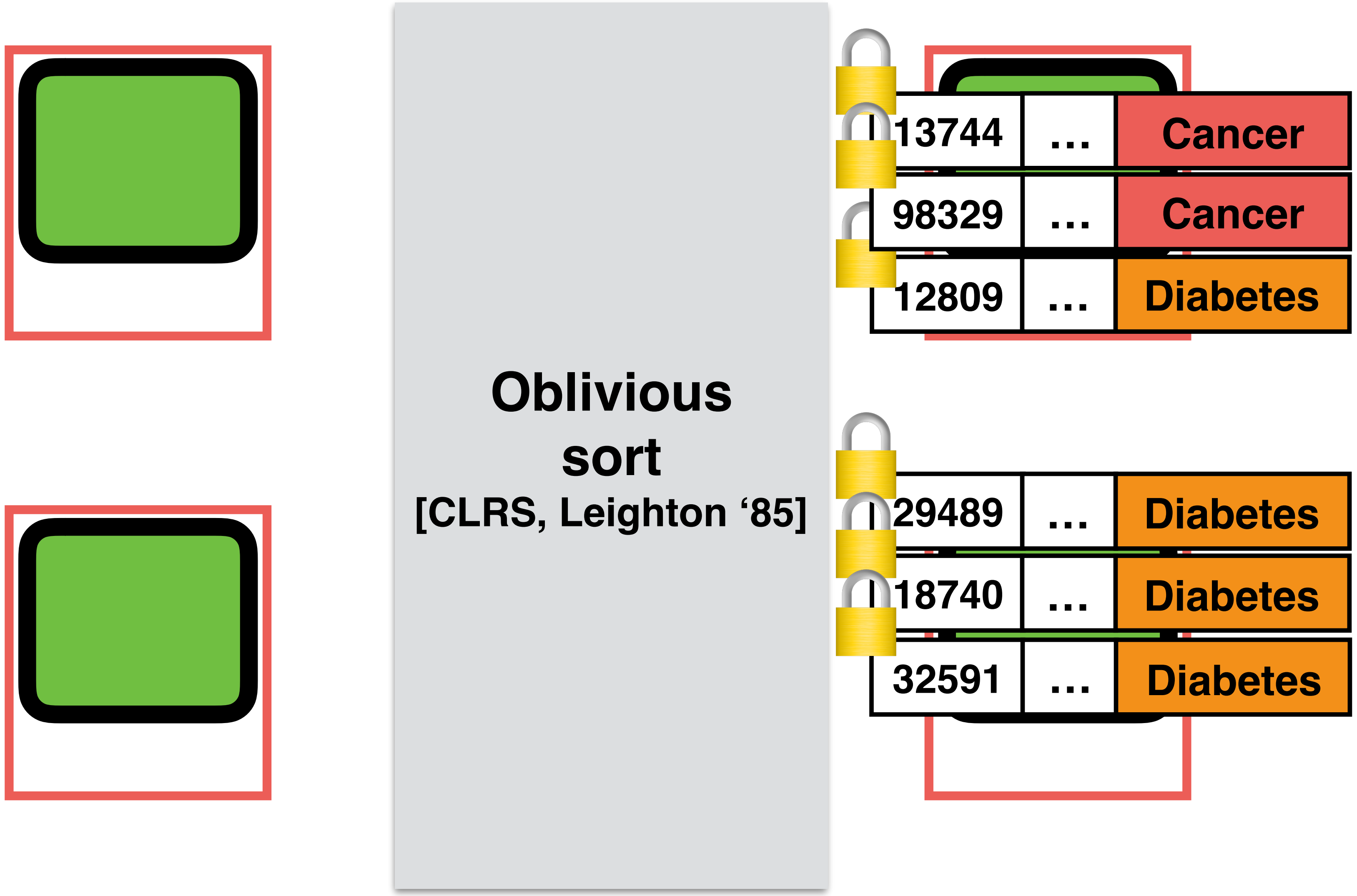
[CLRS, Leighton ‘85]

Map

Sort

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



Oblivious sort
[CLRS, Leighton '85]

13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

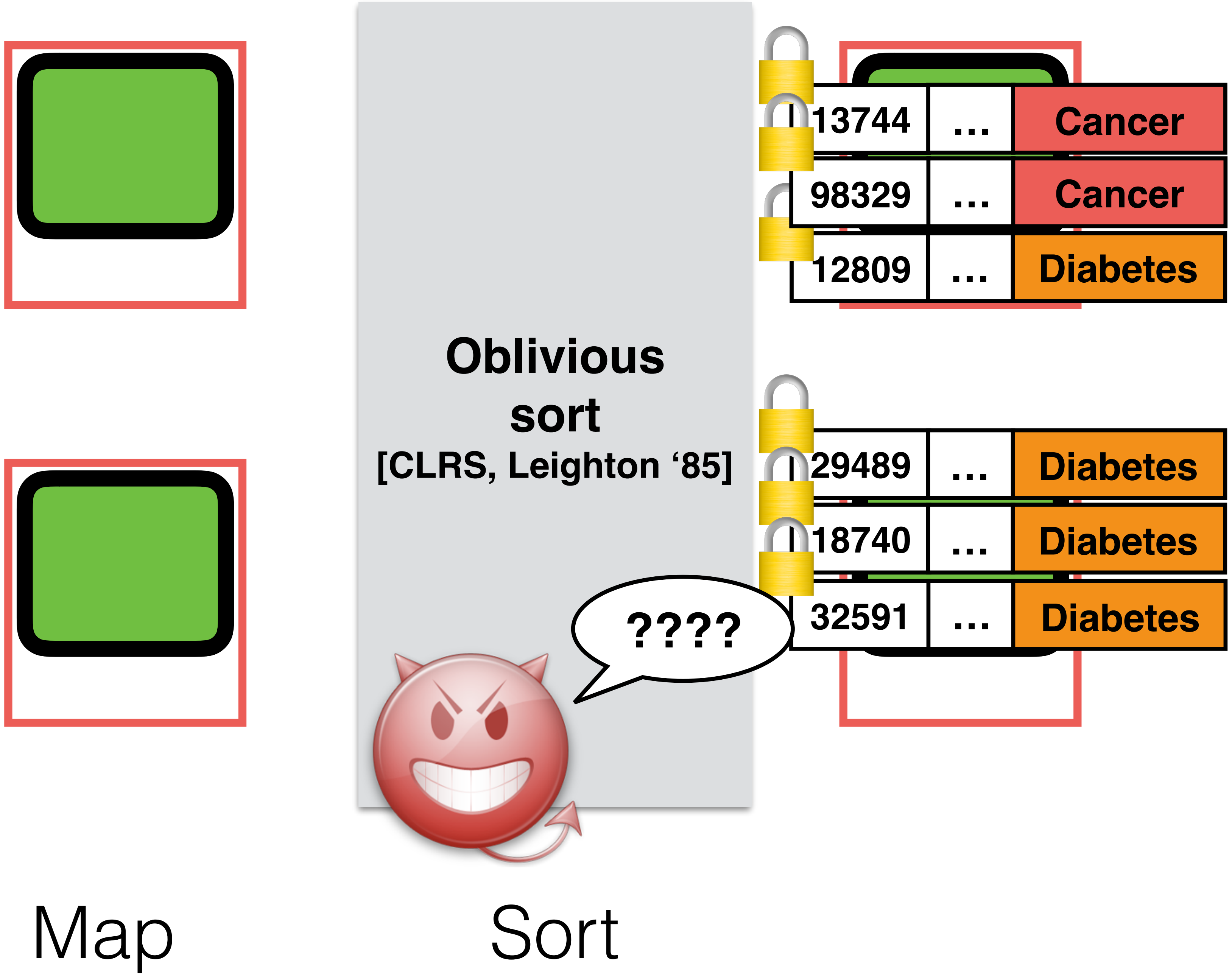
29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

Map

Sort

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease

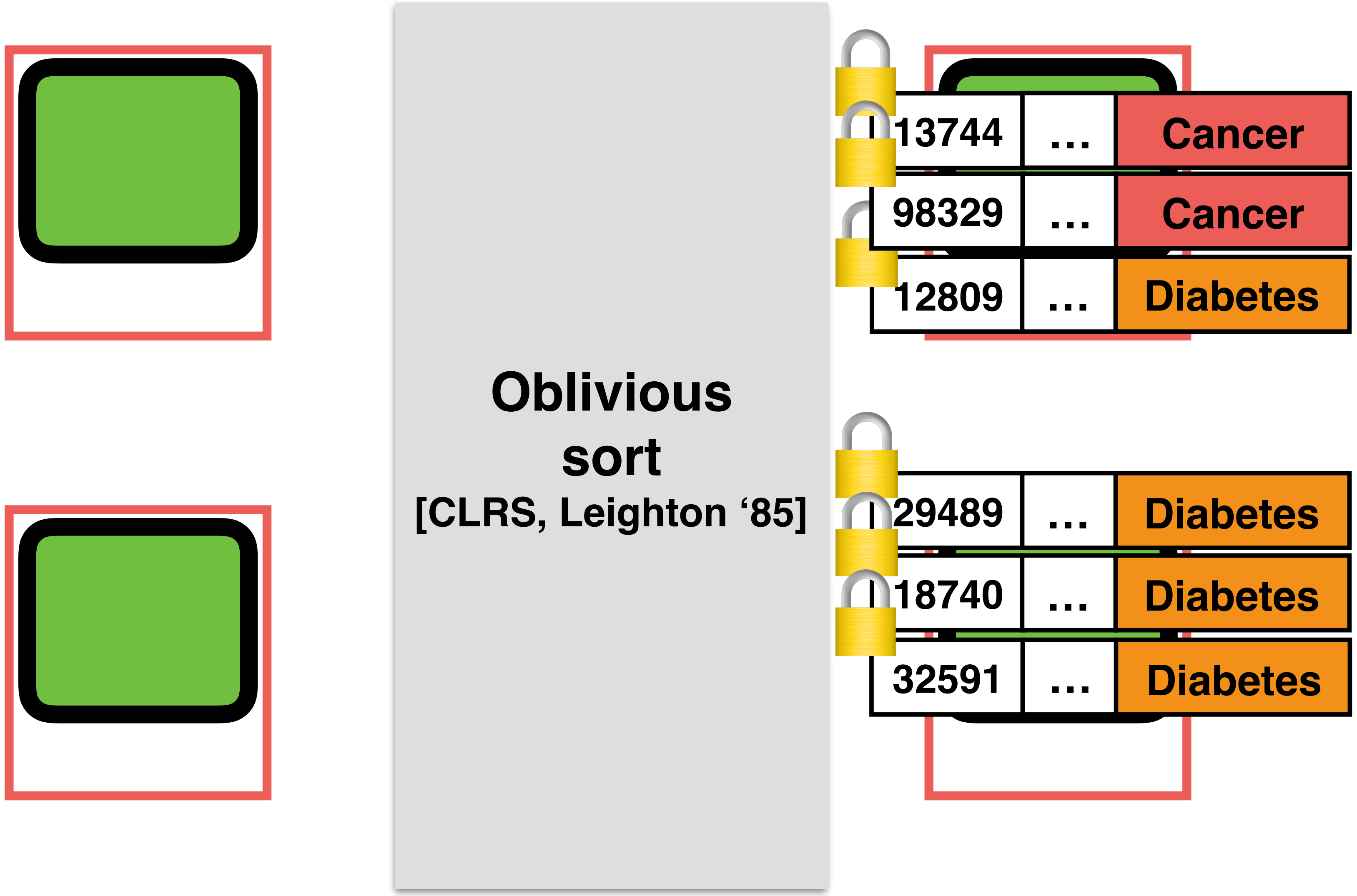


13744	...	Cancer
98329	...	Cancer
12809	...	Diabetes

29489	...	Diabetes
18740	...	Diabetes
32591	...	Diabetes

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



Oblivious sort
[CLRS, Leighton ‘85]

13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

Map

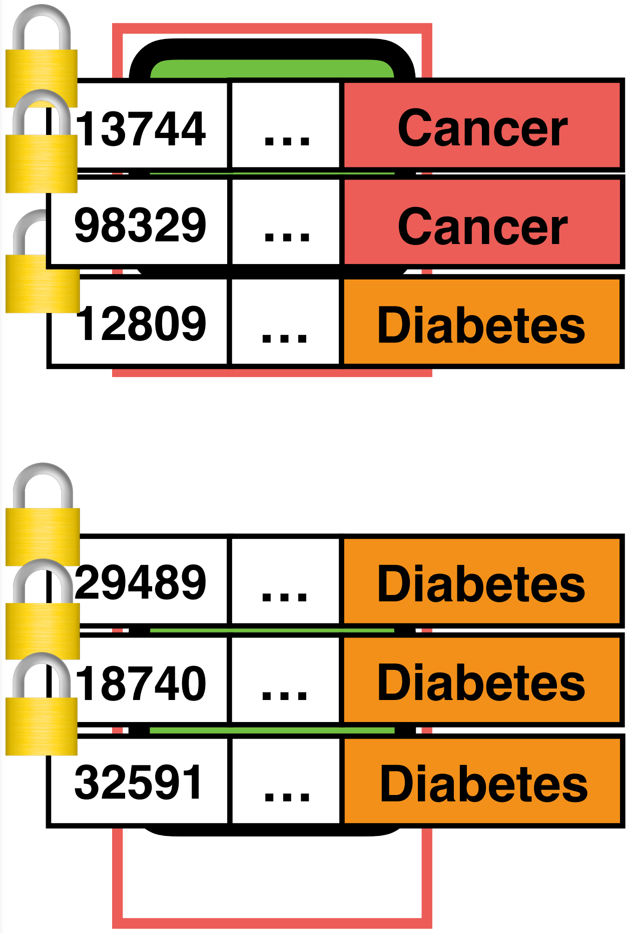
Sort

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease

s

n ‘85]

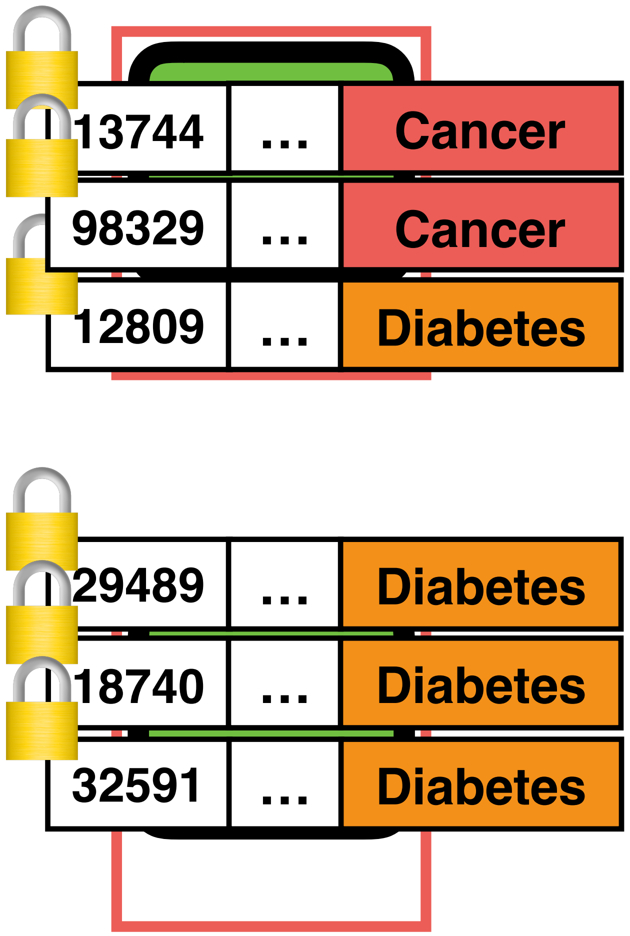


13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease

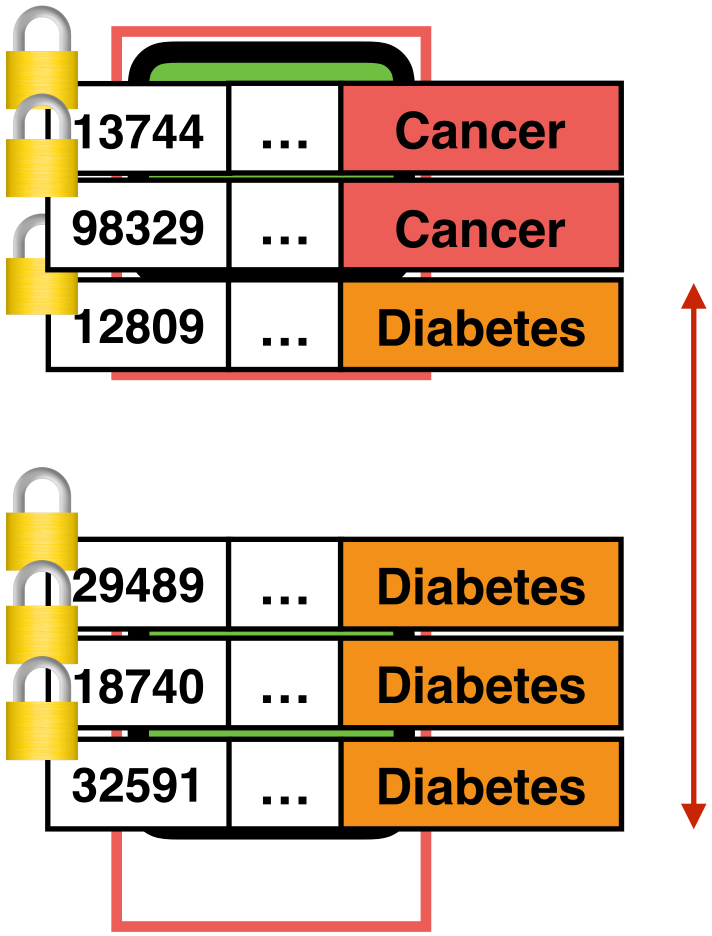


13744	...	Cancer
98329	...	Cancer
12809	...	Diabetes

29489	...	Diabetes
18740	...	Diabetes
32591	...	Diabetes

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



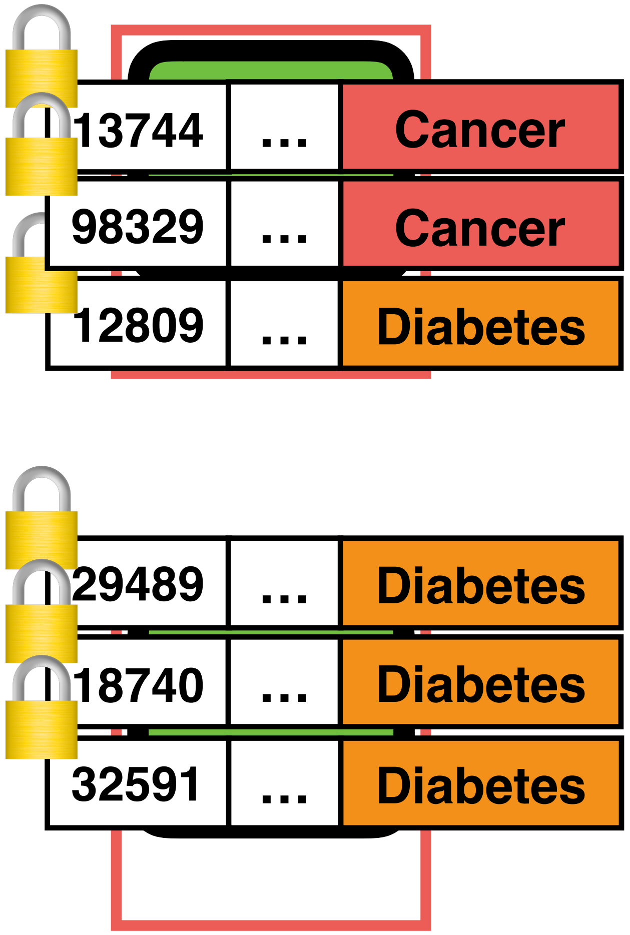
13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

The “Diabetes” group
is split!

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease

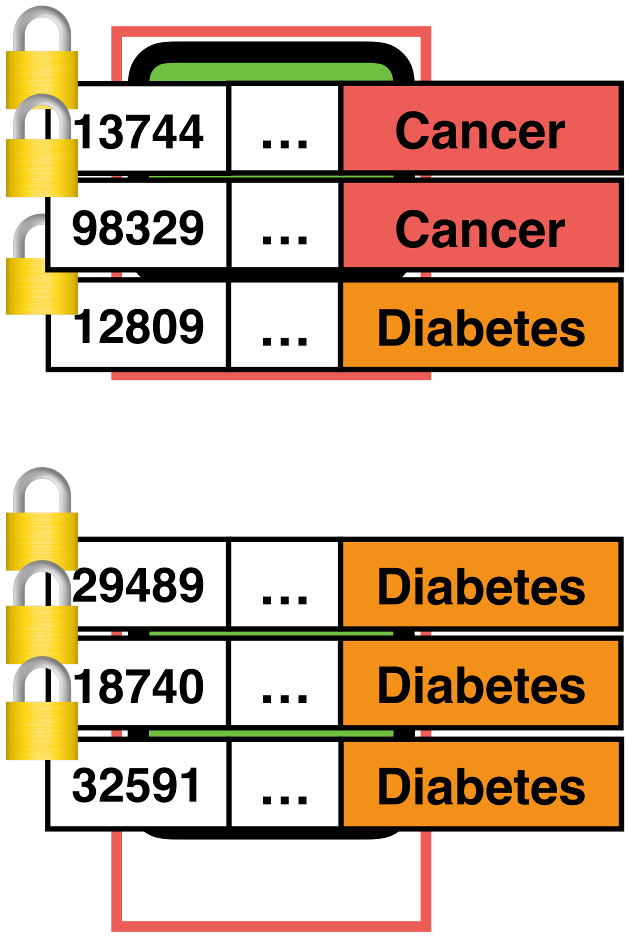


13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



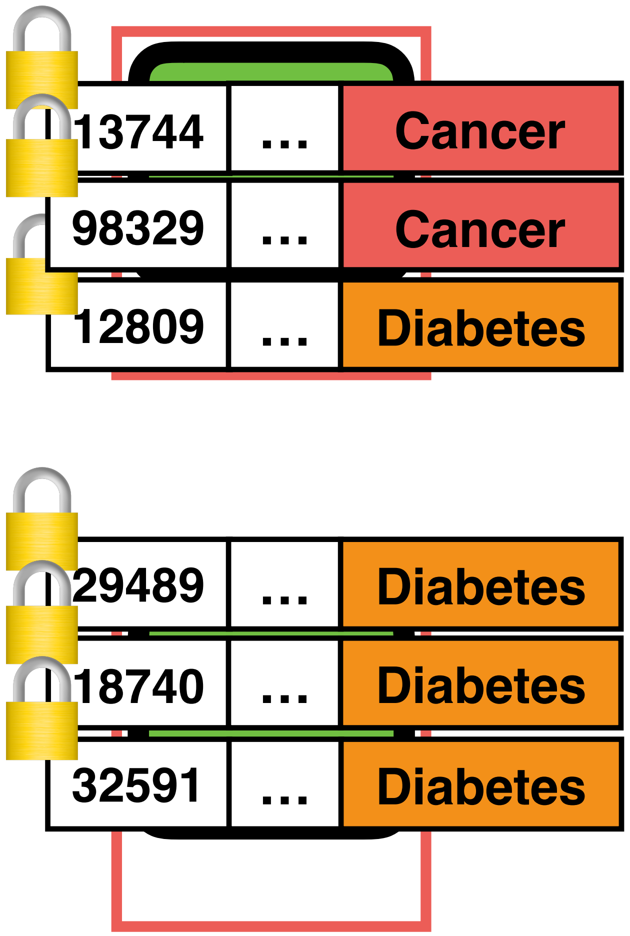
13744	...	Cancer
98329	...	Cancer
12809	...	Diabetes

29489	...	Diabetes
18740	...	Diabetes
32591	...	Diabetes

How to aggregate *obliviously* and in *parallel*?

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease

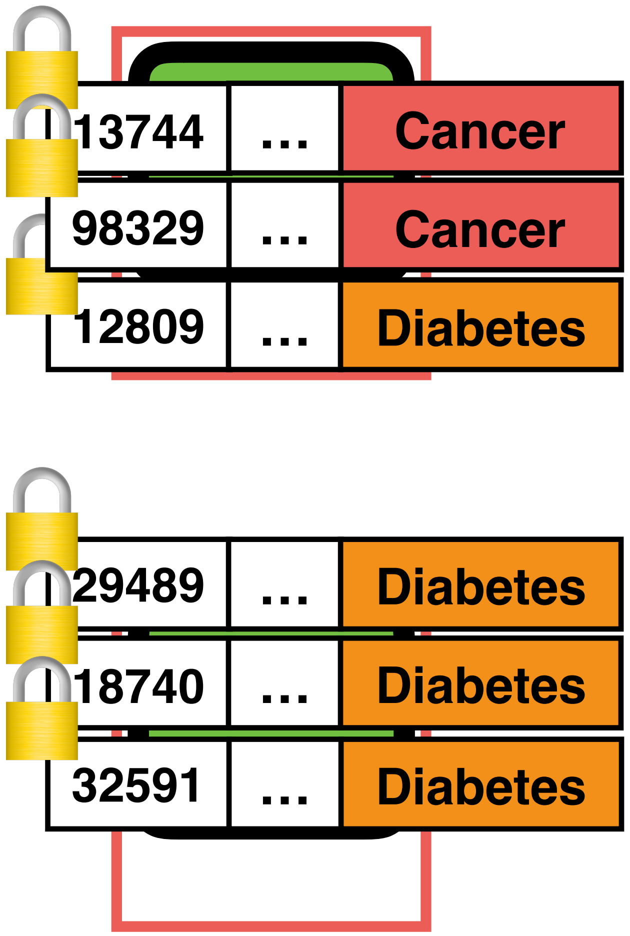


13744	...	Cancer
98329	...	Cancer
12809	...	Diabetes

29489	...	Diabetes
18740	...	Diabetes
32591	...	Diabetes

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



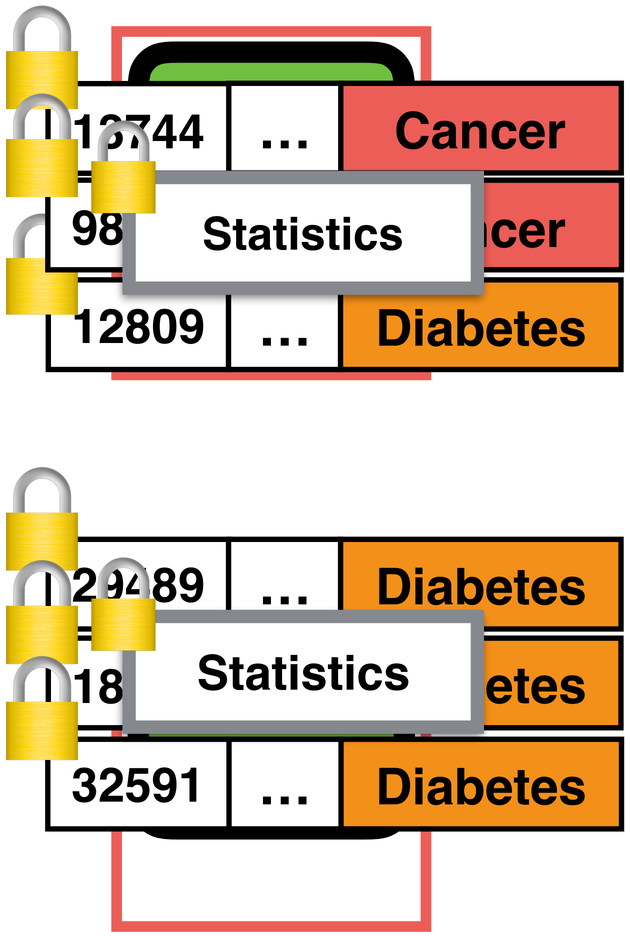
13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

Scan

Oblivious aggregation

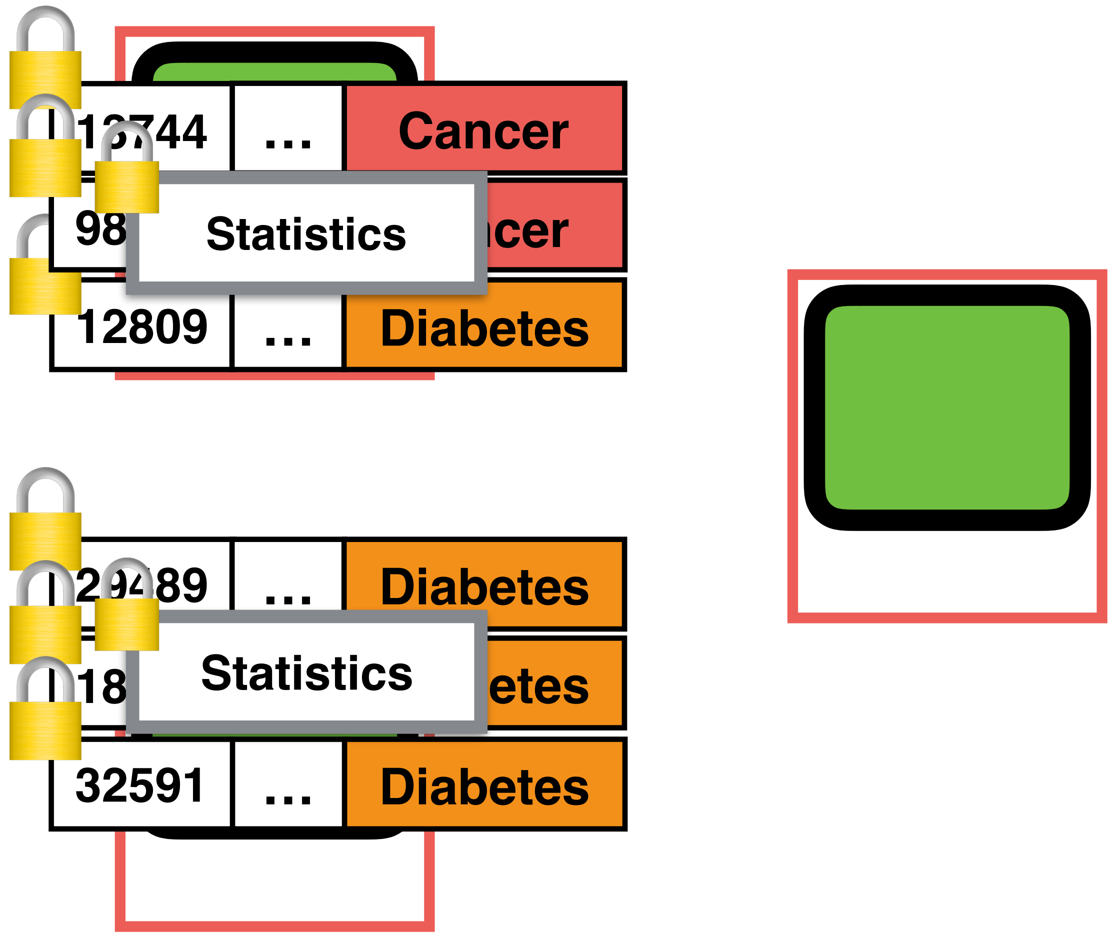
SELECT count(*) FROM medical GROUP BY disease



Scan

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease

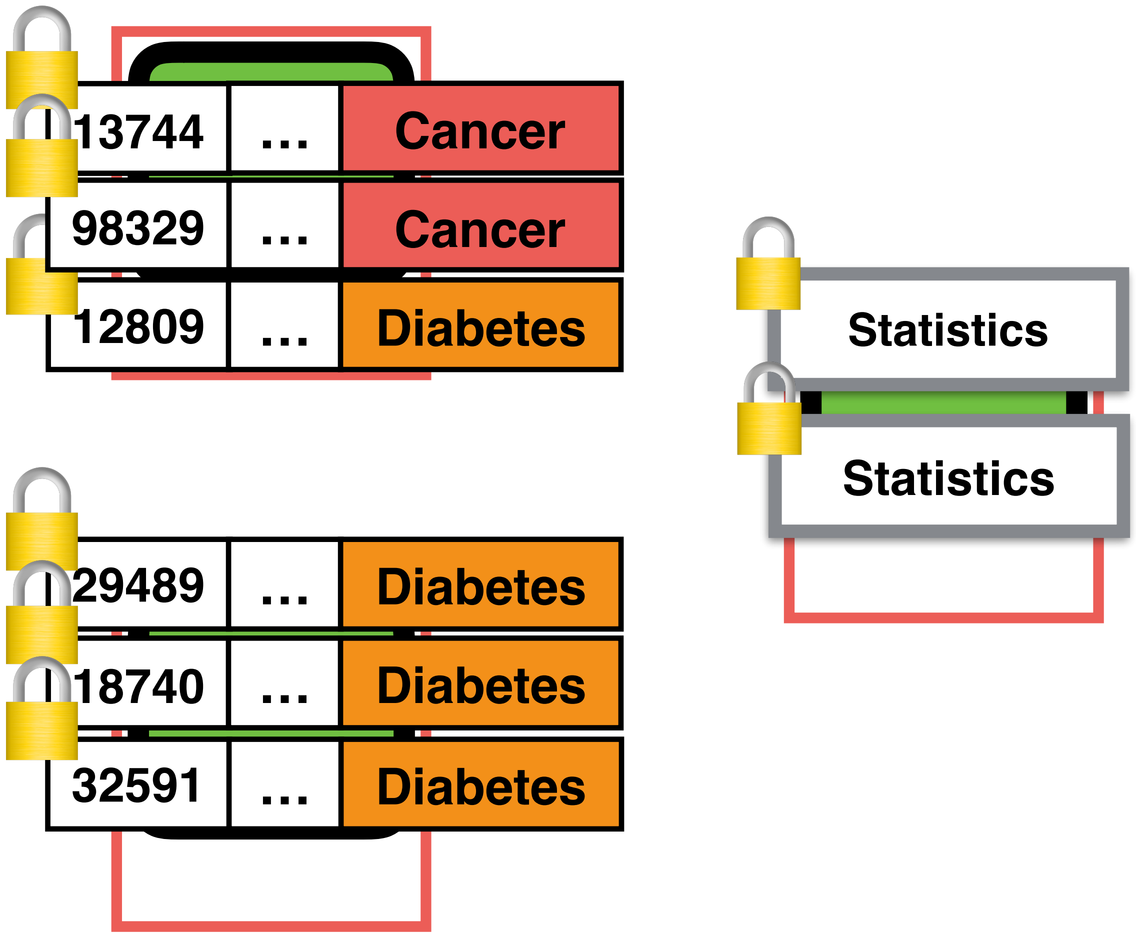


Scan

Boundary processing

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

Statistics

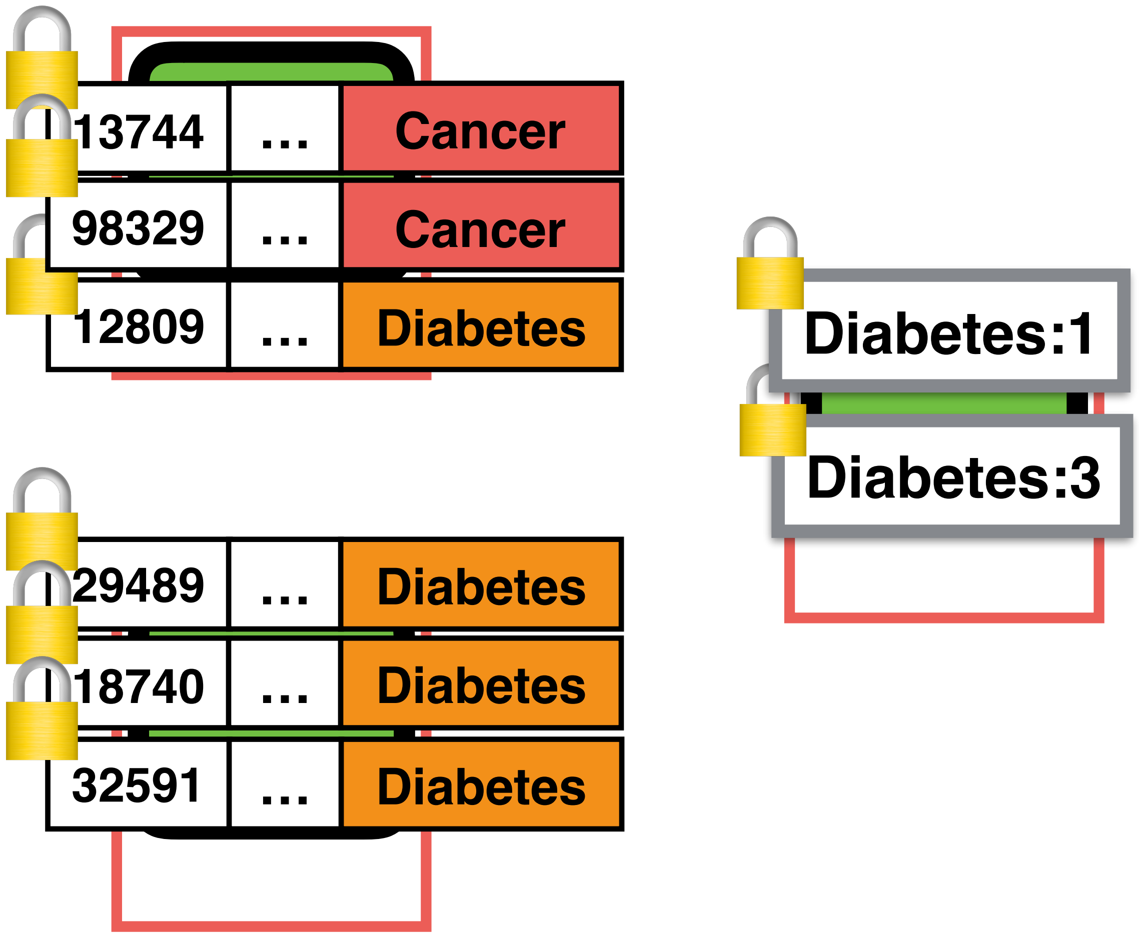
Statistics

Scan

Boundary processing

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

Diabetes:1

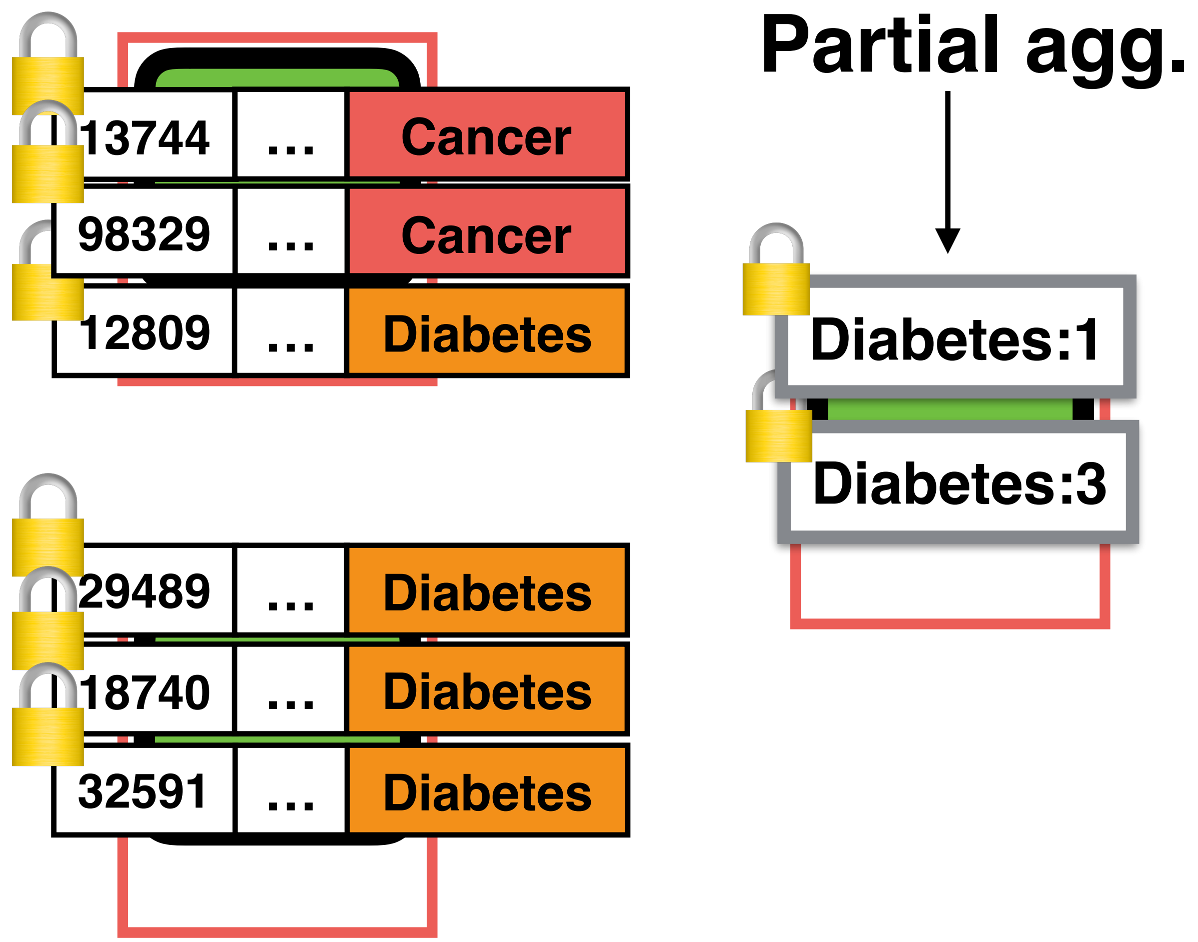
Diabetes:3

Scan

Boundary processing

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

Partial agg.

Diabetes:1

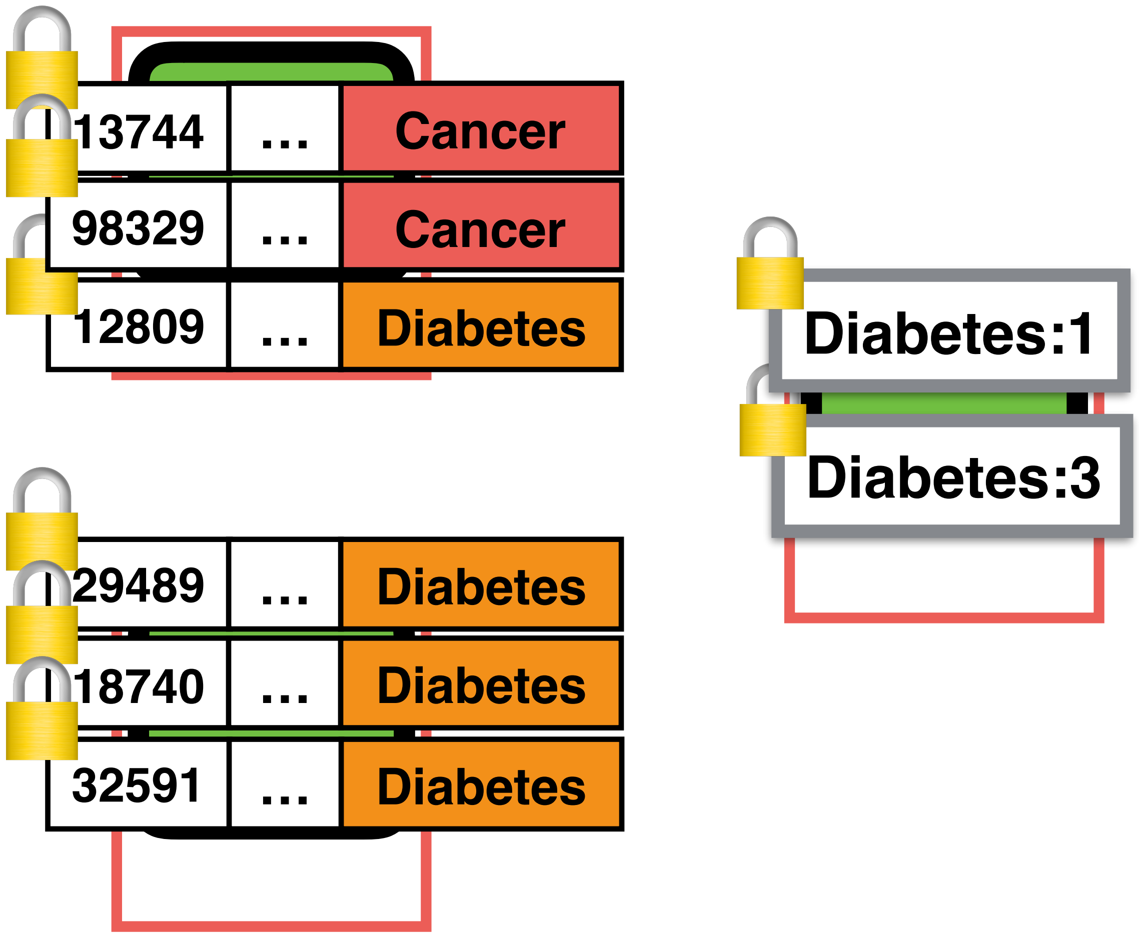
Diabetes:3

Scan

Boundary processing

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

Diabetes:1

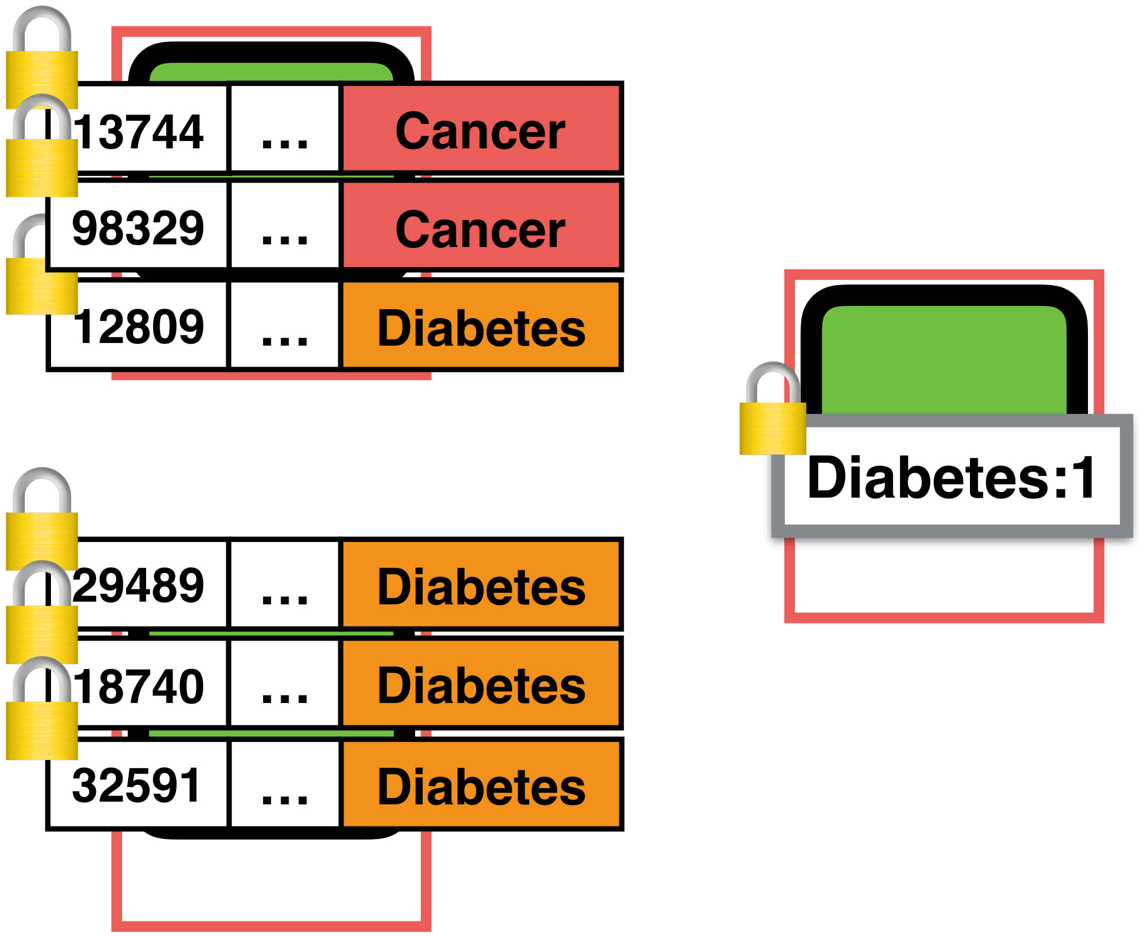
Diabetes:3

Scan

Boundary processing

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

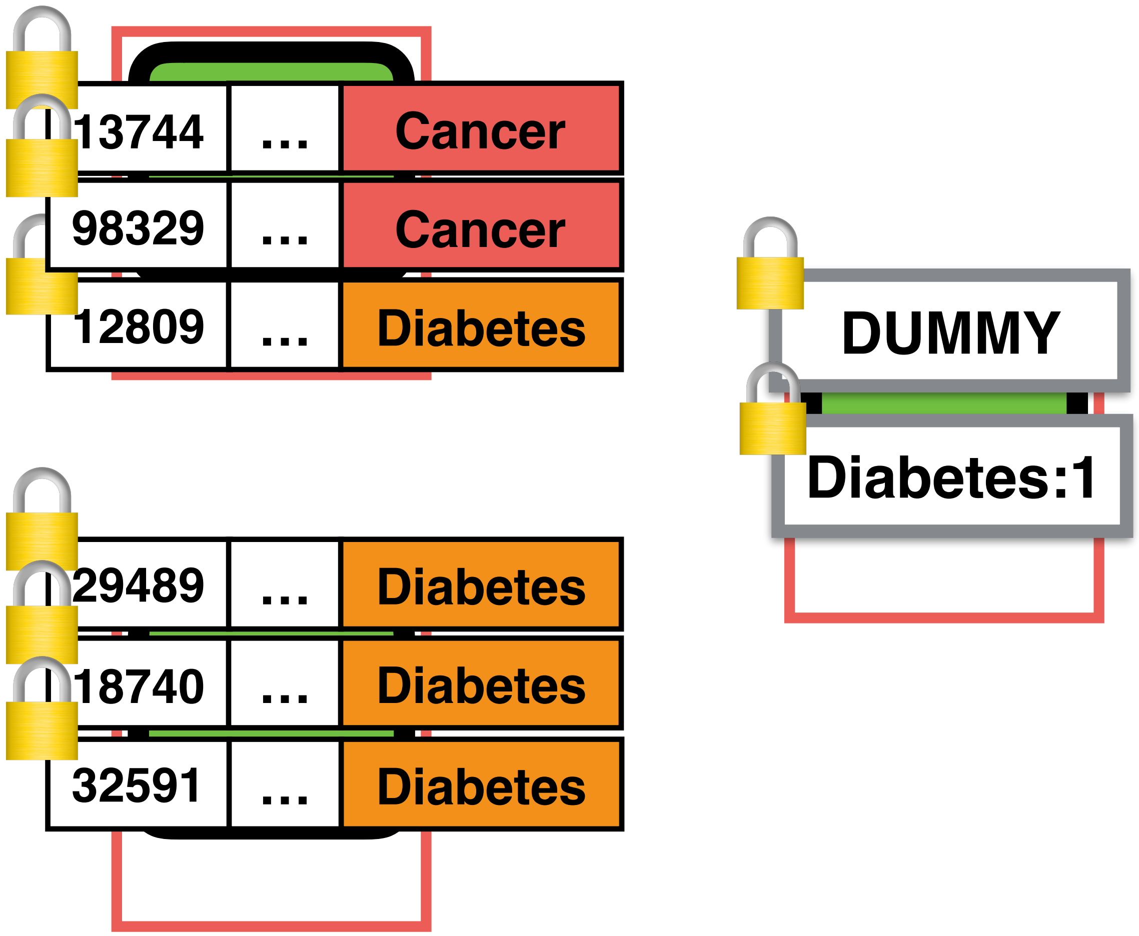
29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

Scan

Boundary processing

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease

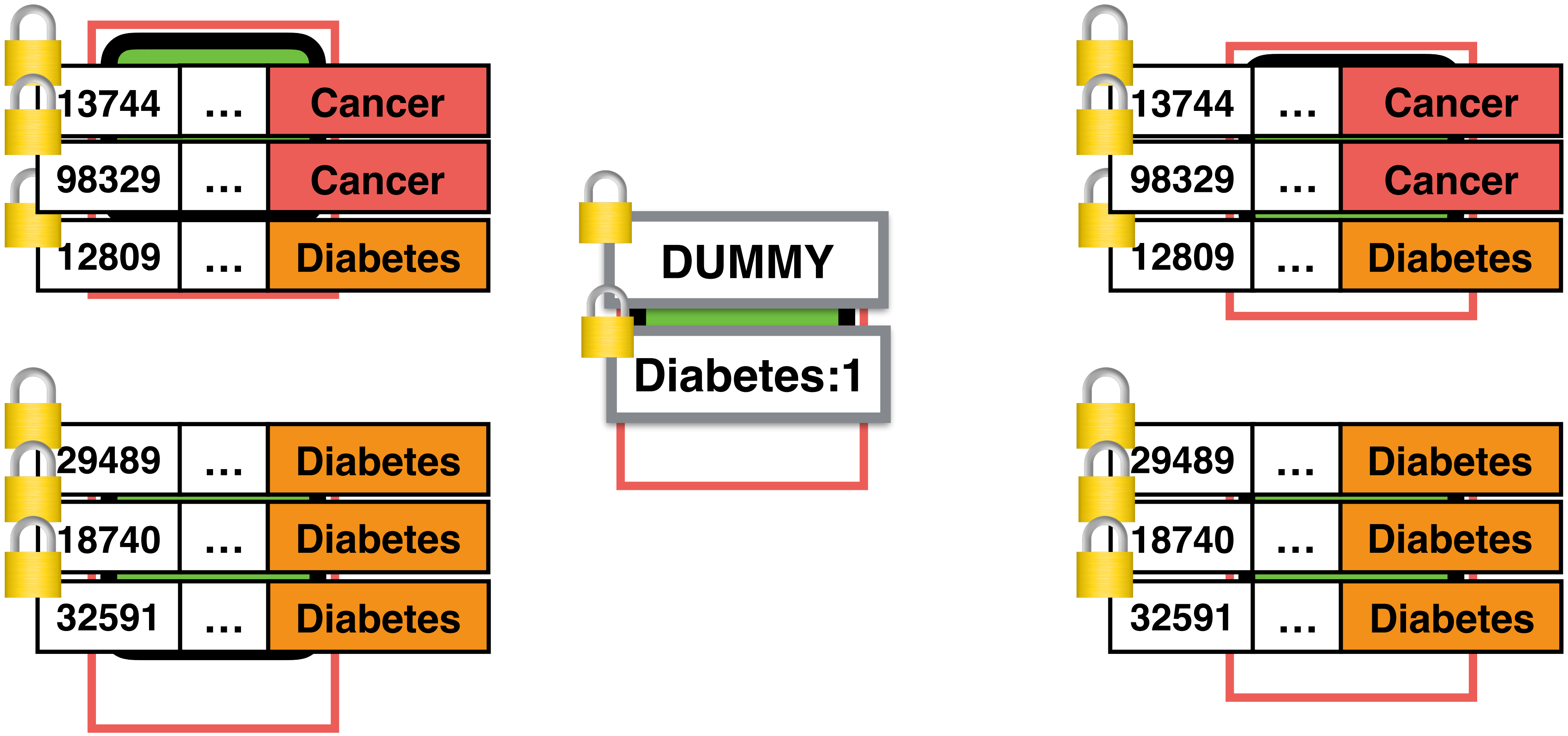


Scan

Boundary processing

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

DUMMY

Diabetes:1

13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

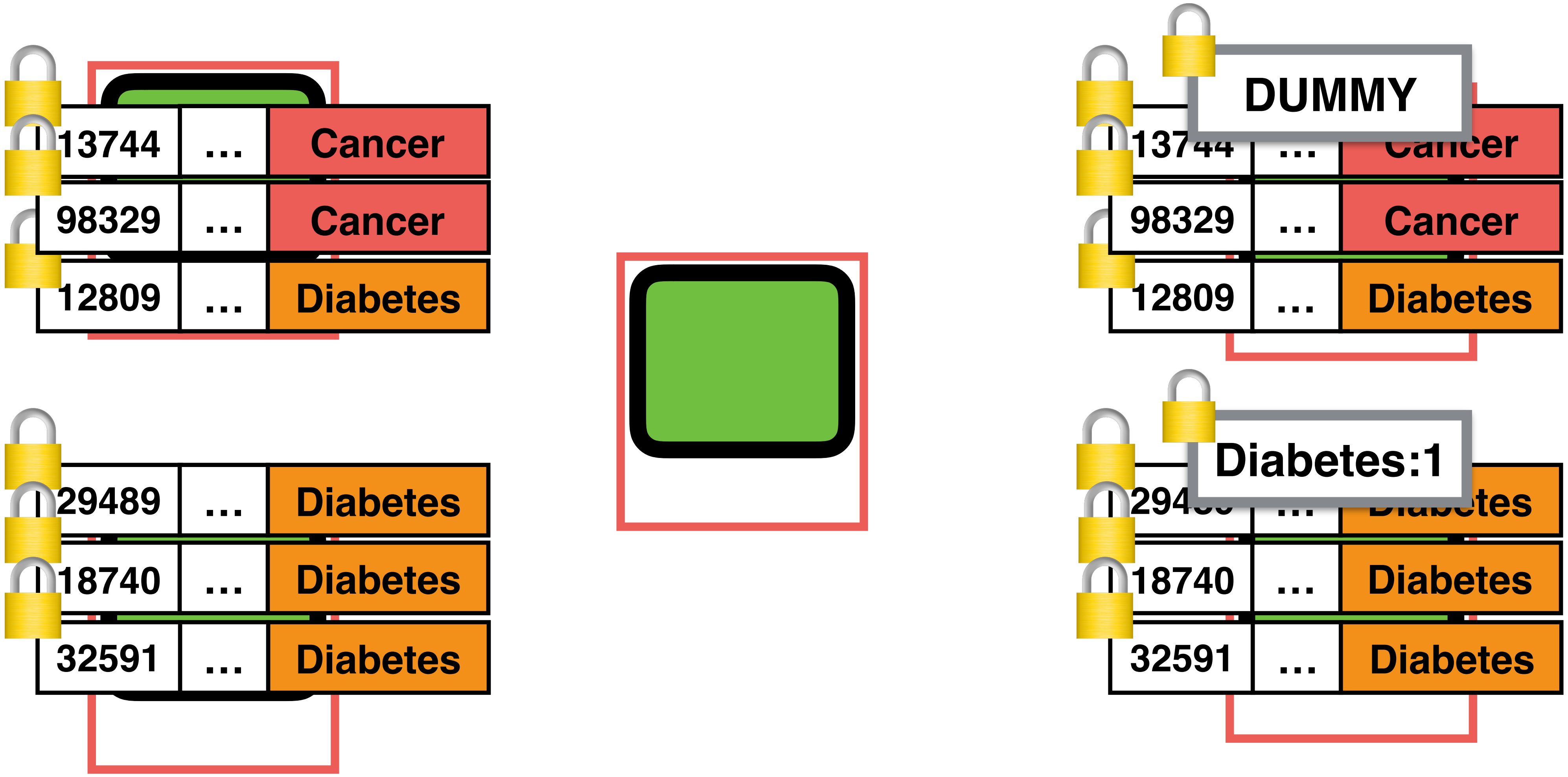
29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

Scan

Boundary processing

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

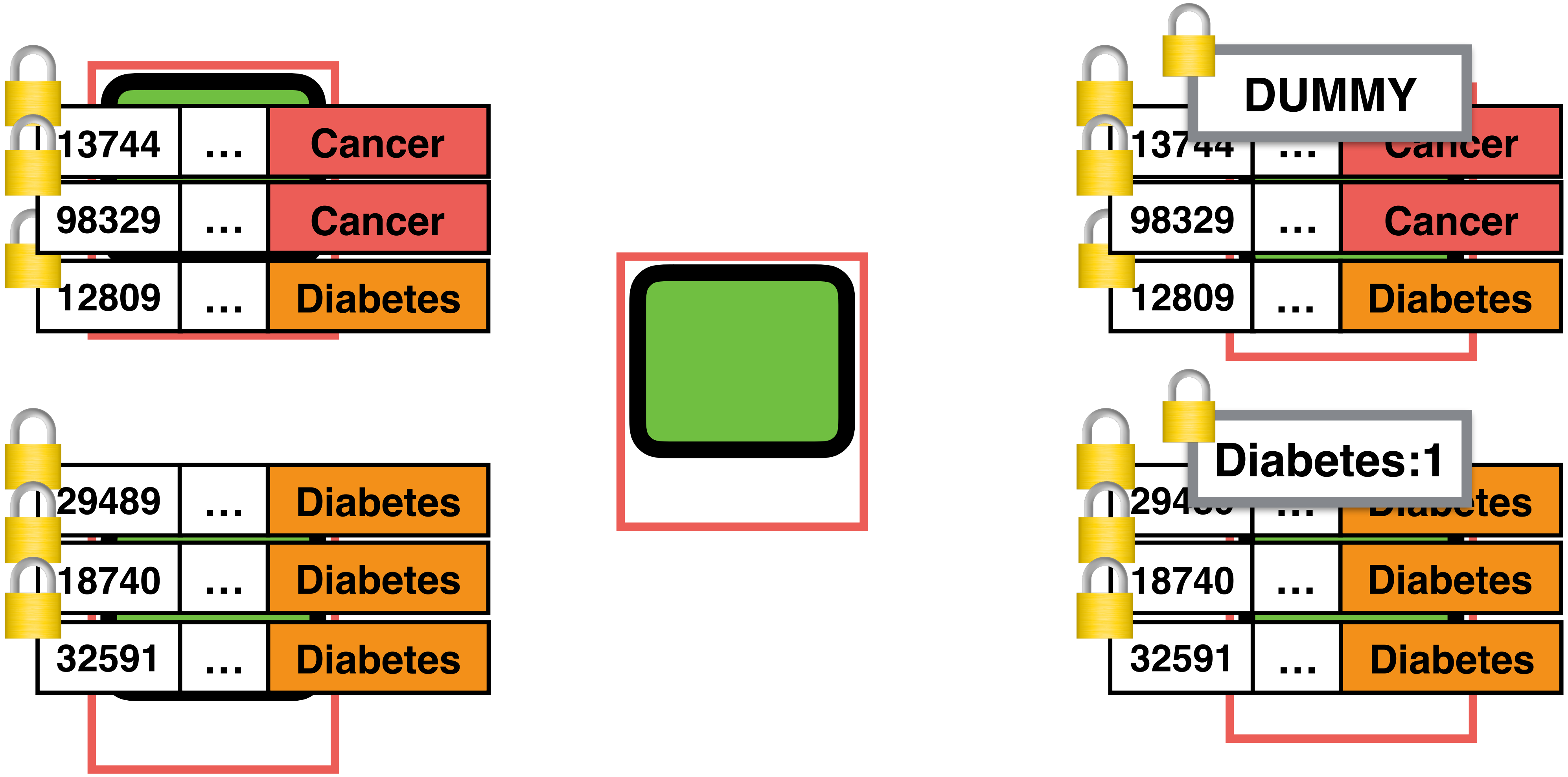
29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

Scan

Boundary processing

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

DUMMY

13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

Diabetes:1

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

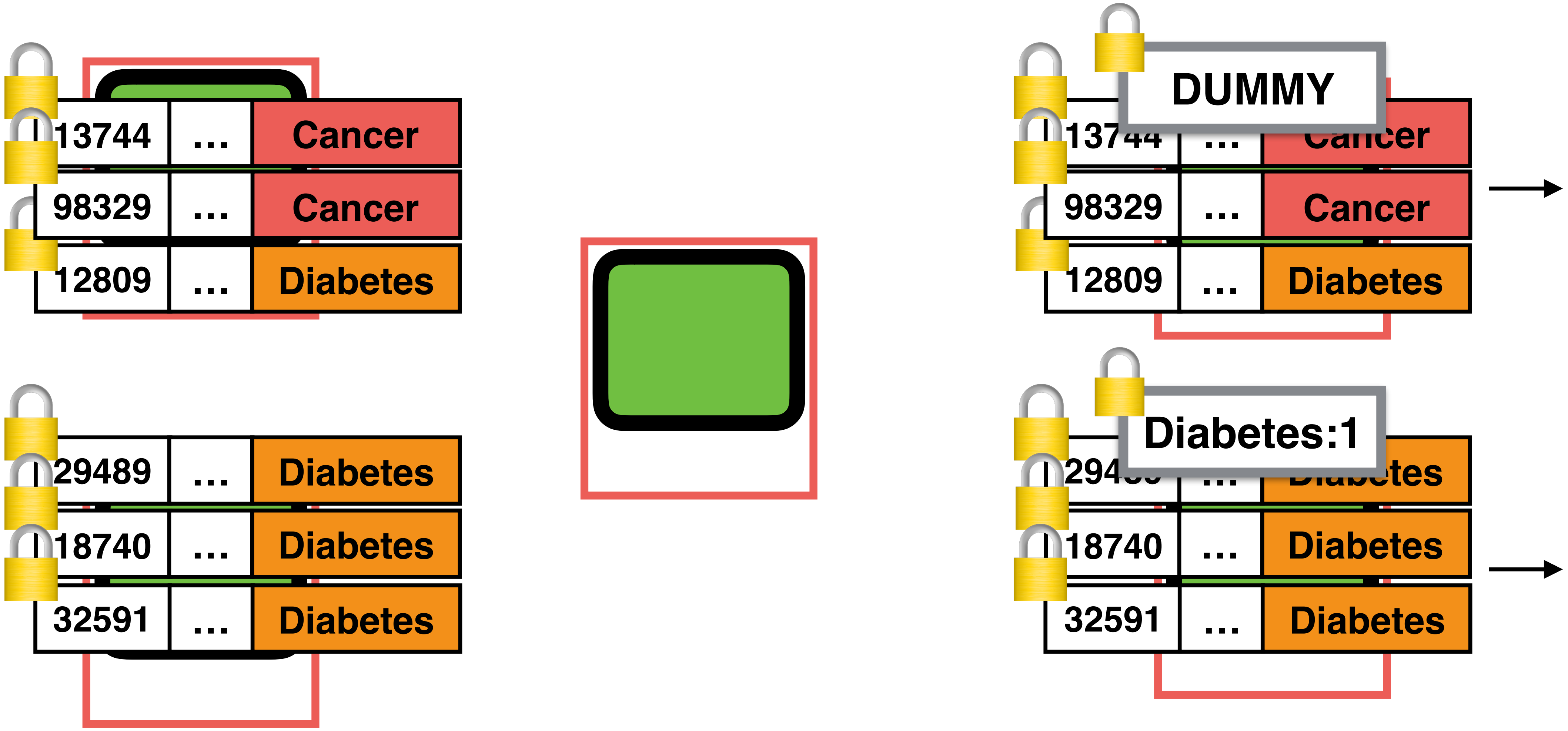
Scan

Boundary processing

Scan

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

DUMMY

13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

Diabetes:1

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

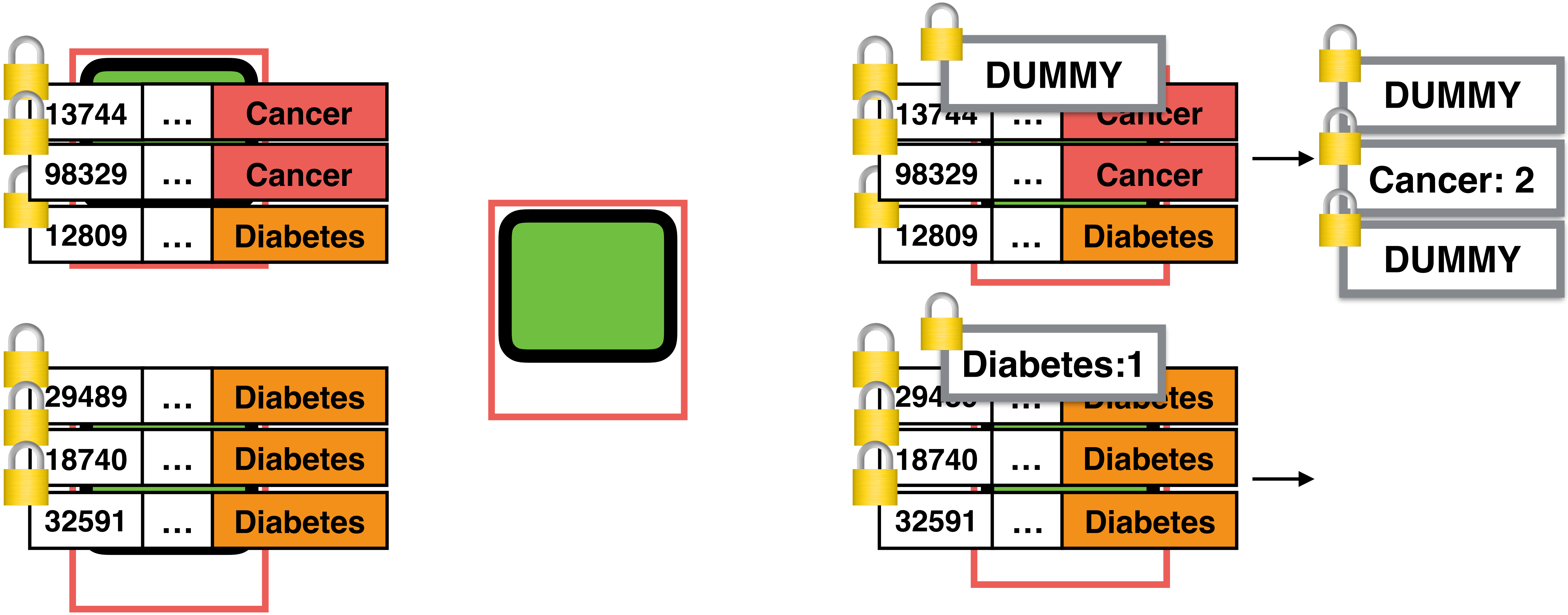
Scan

Boundary processing

Scan

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

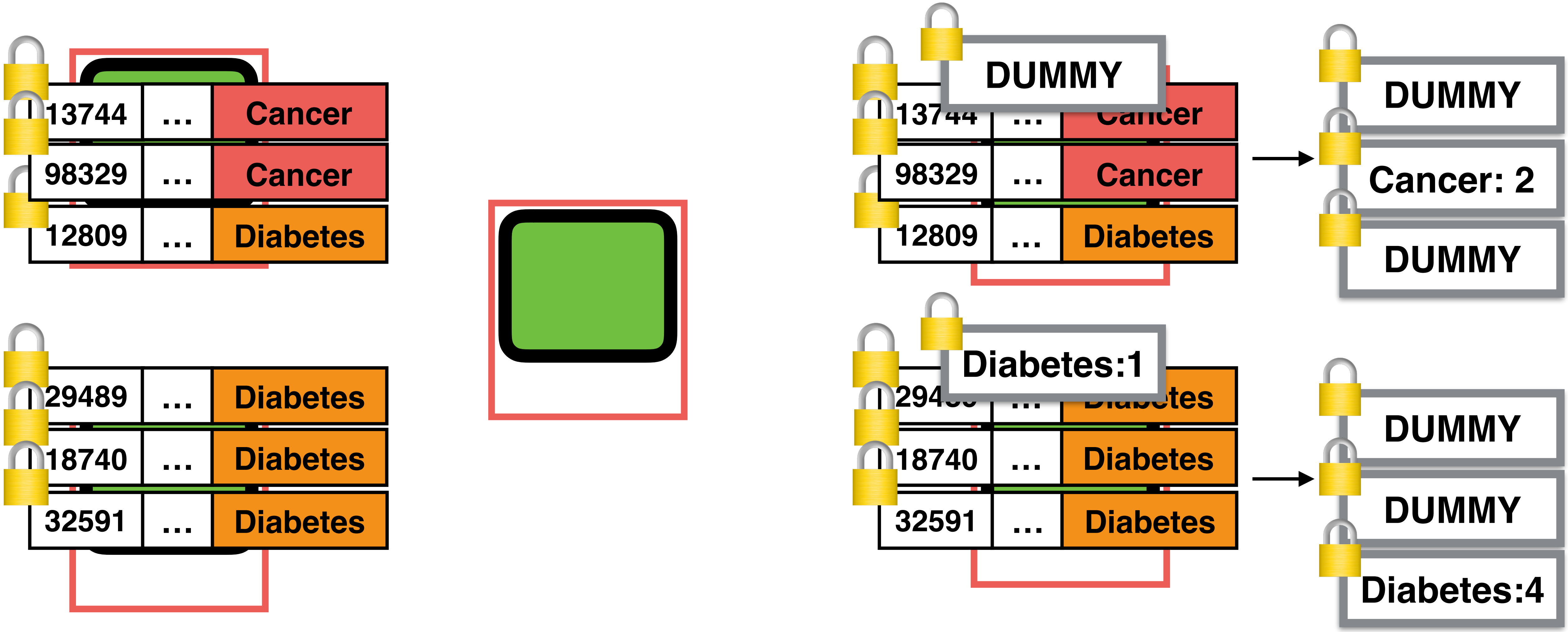
Scan

Boundary processing

Scan

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



13744 … Cancer

98329 … Cancer

12809 … Diabetes

29489 … Diabetes

18740 … Diabetes

32591 … Diabetes

DUMMY

13744 … Cancer

98329 … Cancer

12809 … Diabetes

DUMMY

Cancer: 2

DUMMY

Diabetes:1

29489 … Diabetes

18740 … Diabetes

32591 … Diabetes

DUMMY

DUMMY

Diabetes:4

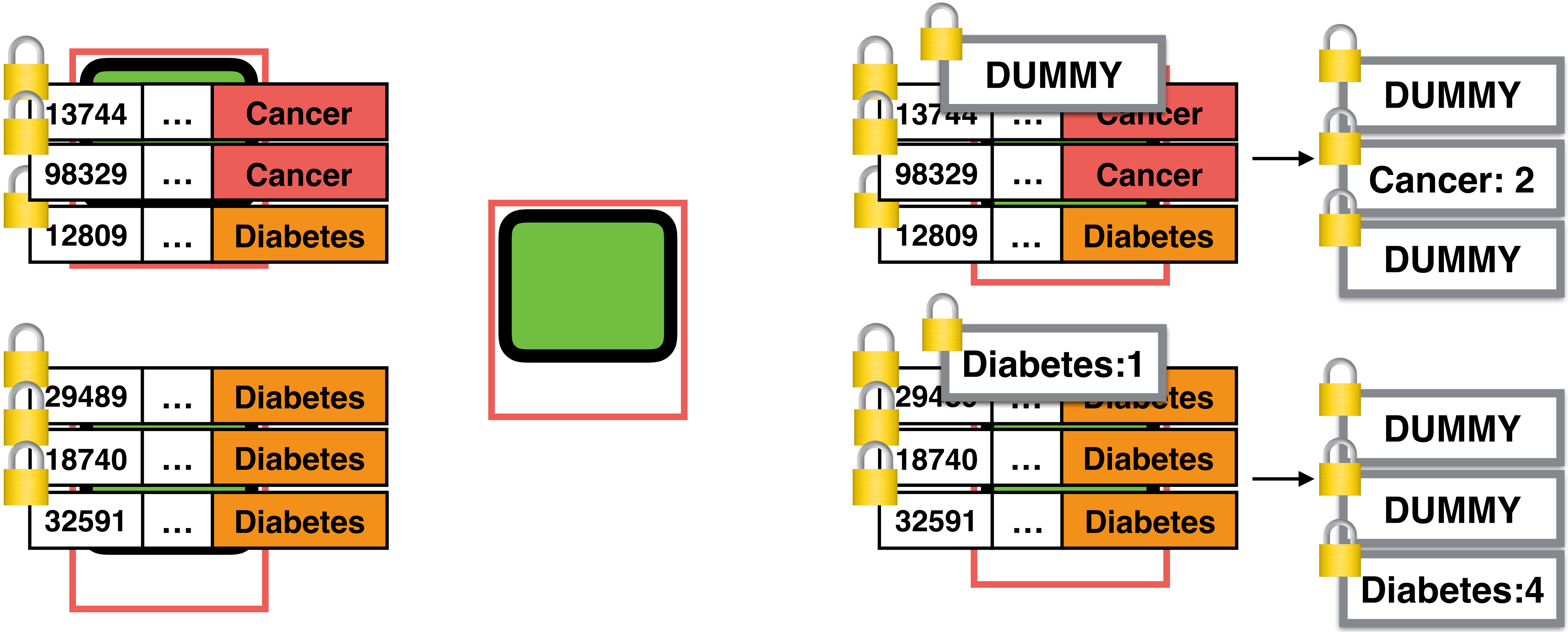
Scan

Boundary processing

Scan

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

29489	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

DUMMY

13744	…	Cancer
98329	…	Cancer
12809	…	Diabetes

DUMMY
Cancer: 2
DUMMY

Diabetes:1

294	…	Diabetes
18740	…	Diabetes
32591	…	Diabetes

DUMMY
DUMMY
Diabetes:4

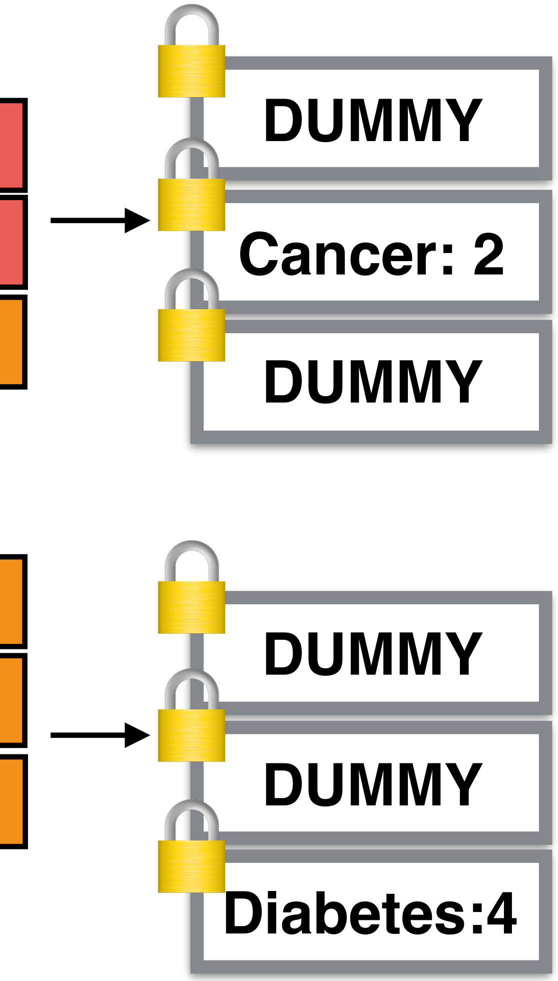
Scan

Boundary processing

Scan

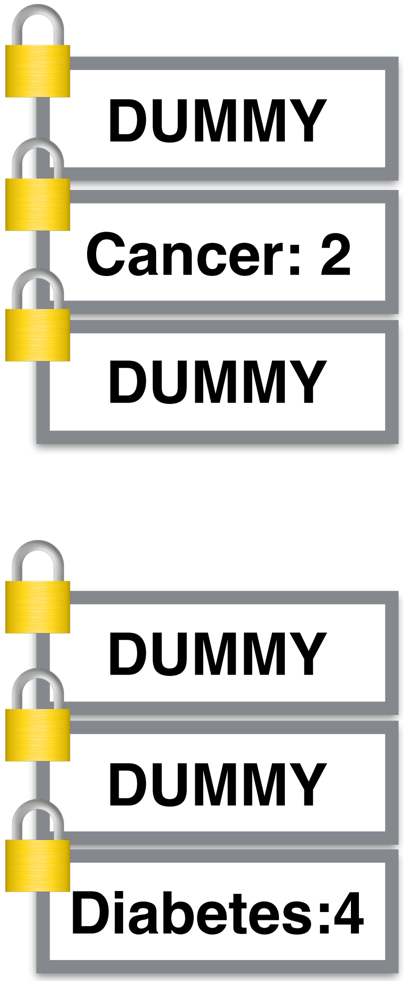
Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



DUMMY

Cancer: 2

DUMMY

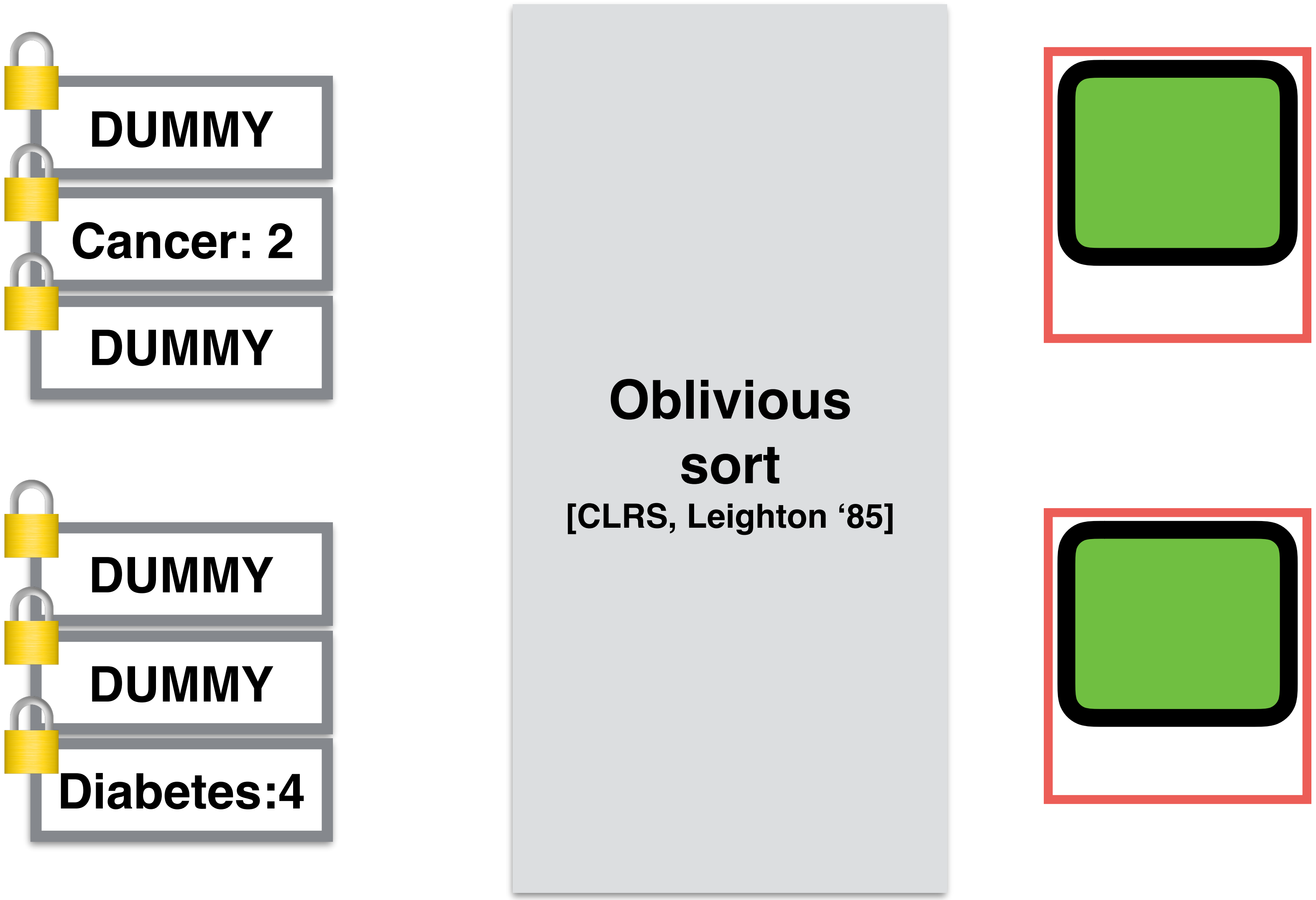
DUMMY

DUMMY

Diabetes:4

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



DUMMY

Cancer: 2

DUMMY

DUMMY

DUMMY

Diabetes:4

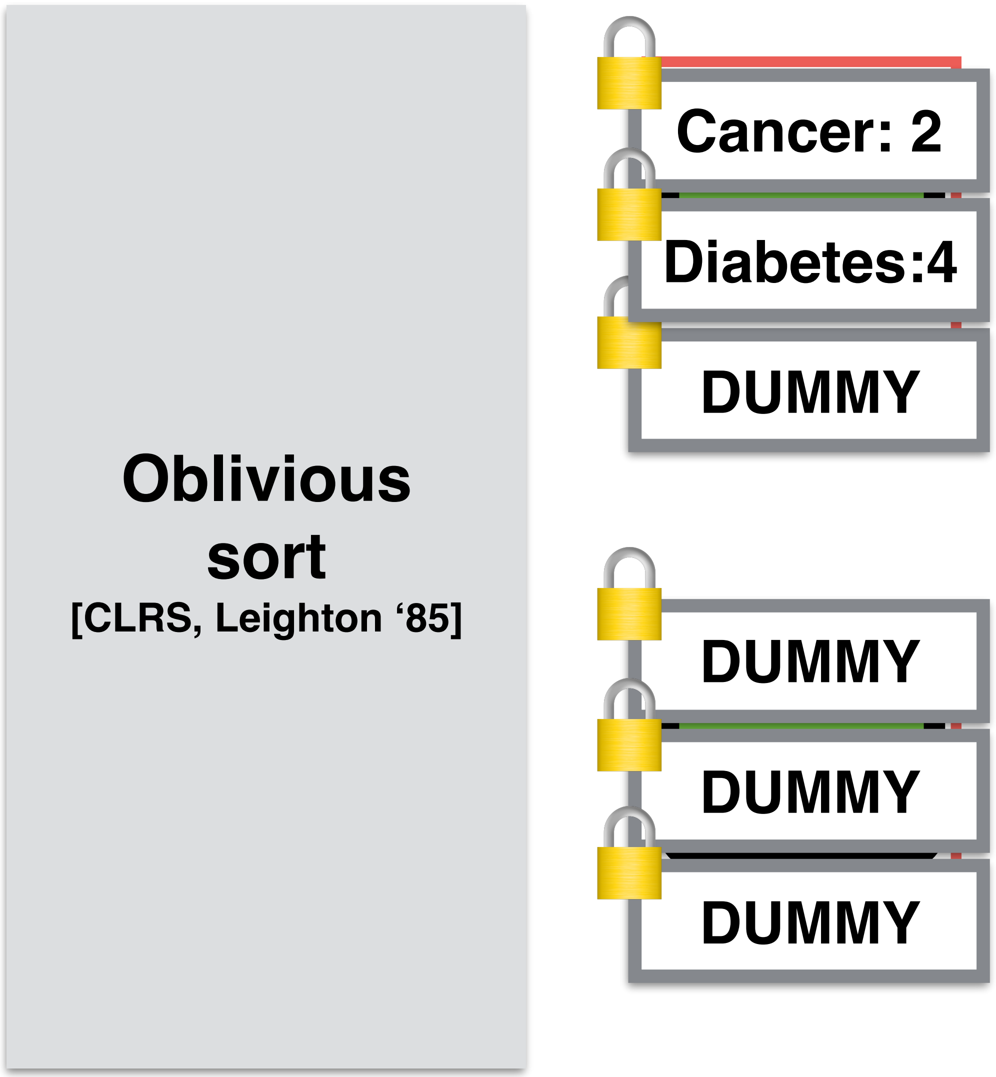
Oblivious sort

[CLRS, Leighton '85]

Sort

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



Oblivious sort

[CLRS, Leighton ‘85]

Cancer: 2

Diabetes:4

DUMMY

DUMMY

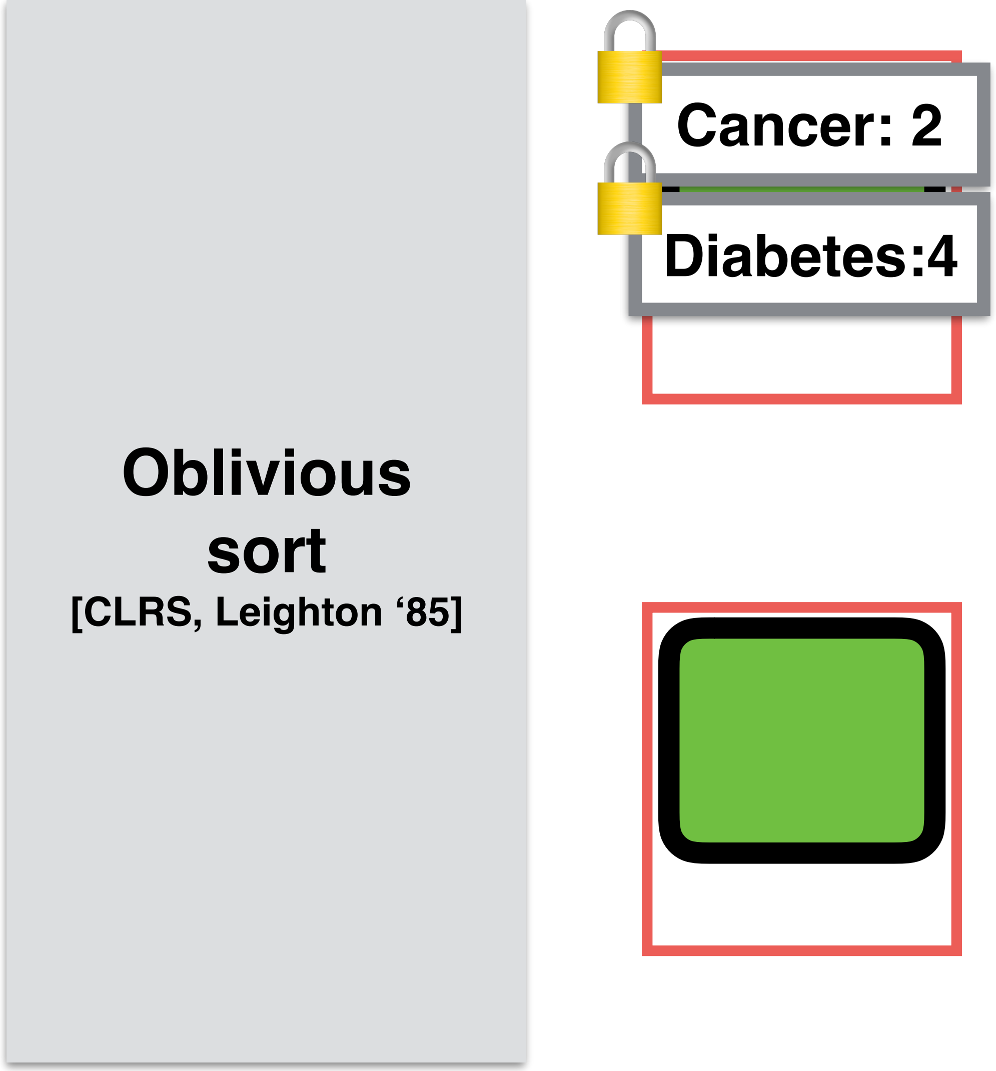
DUMMY

DUMMY

Sort

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



Oblivious sort

[CLRS, Leighton '85]

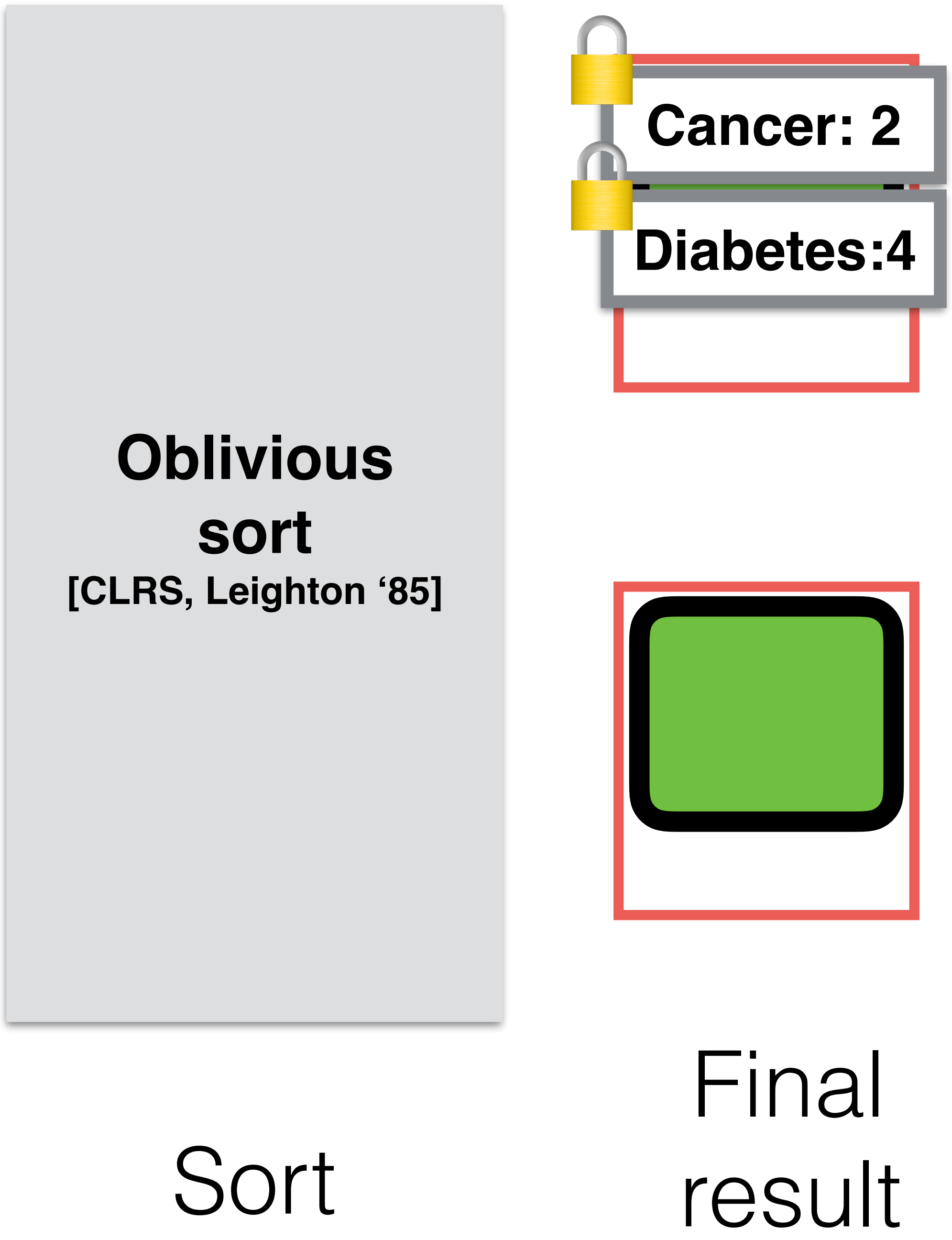
Cancer: 2

Diabetes:4

Sort

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



Oblivious sort
[CLRS, Leighton '85]

Cancer: 2

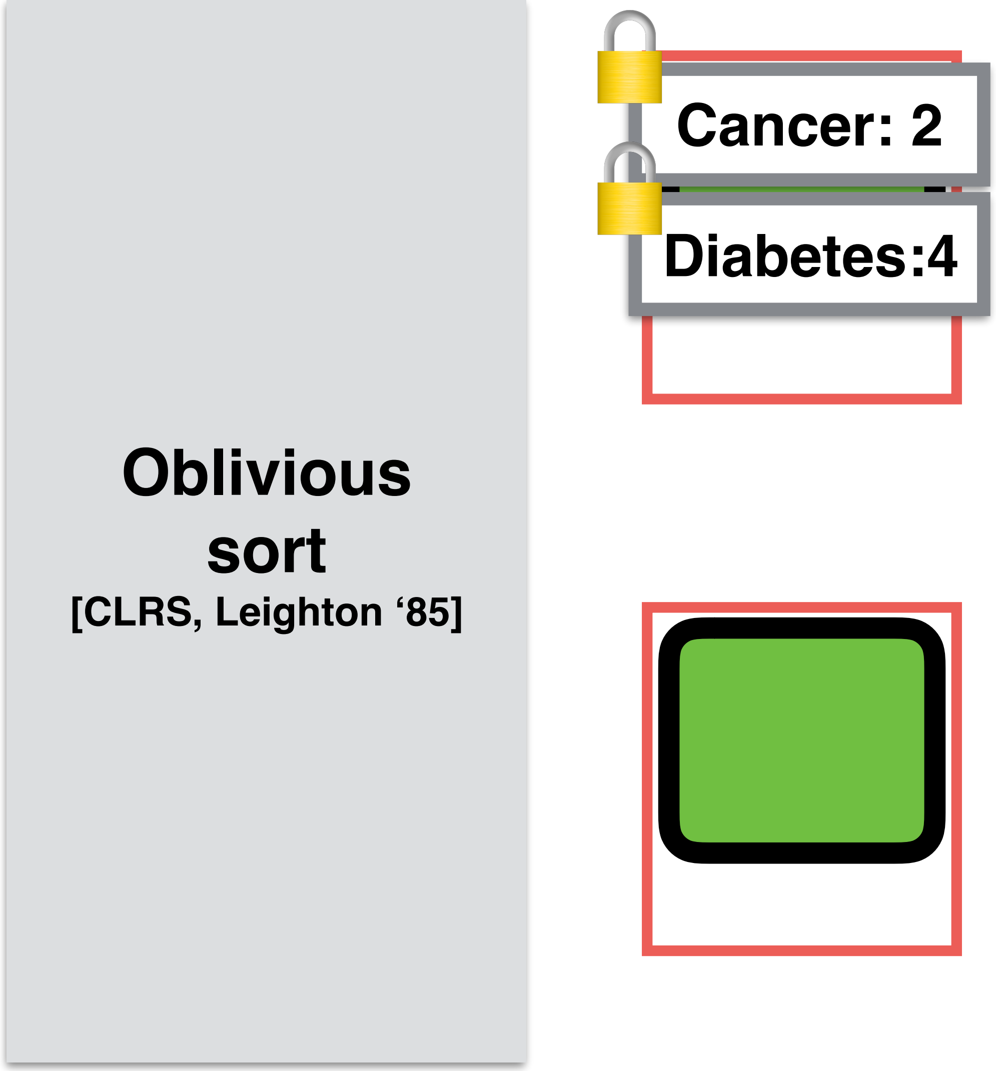
Diabetes:4

Sort

Final result

Oblivious aggregation

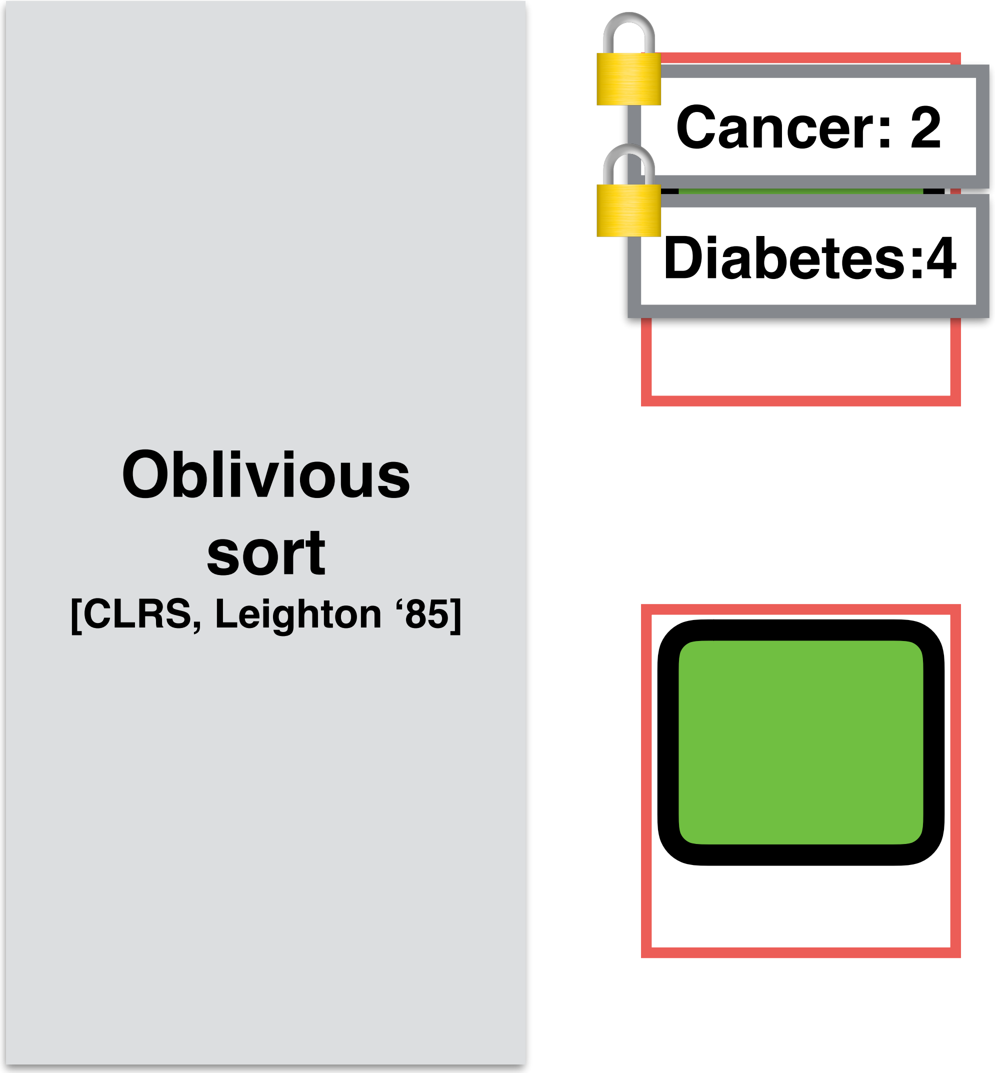
SELECT count(*) FROM medical GROUP BY disease



Aggregation has two sorts…

Oblivious aggregation

SELECT count(*) FROM medical GROUP BY disease



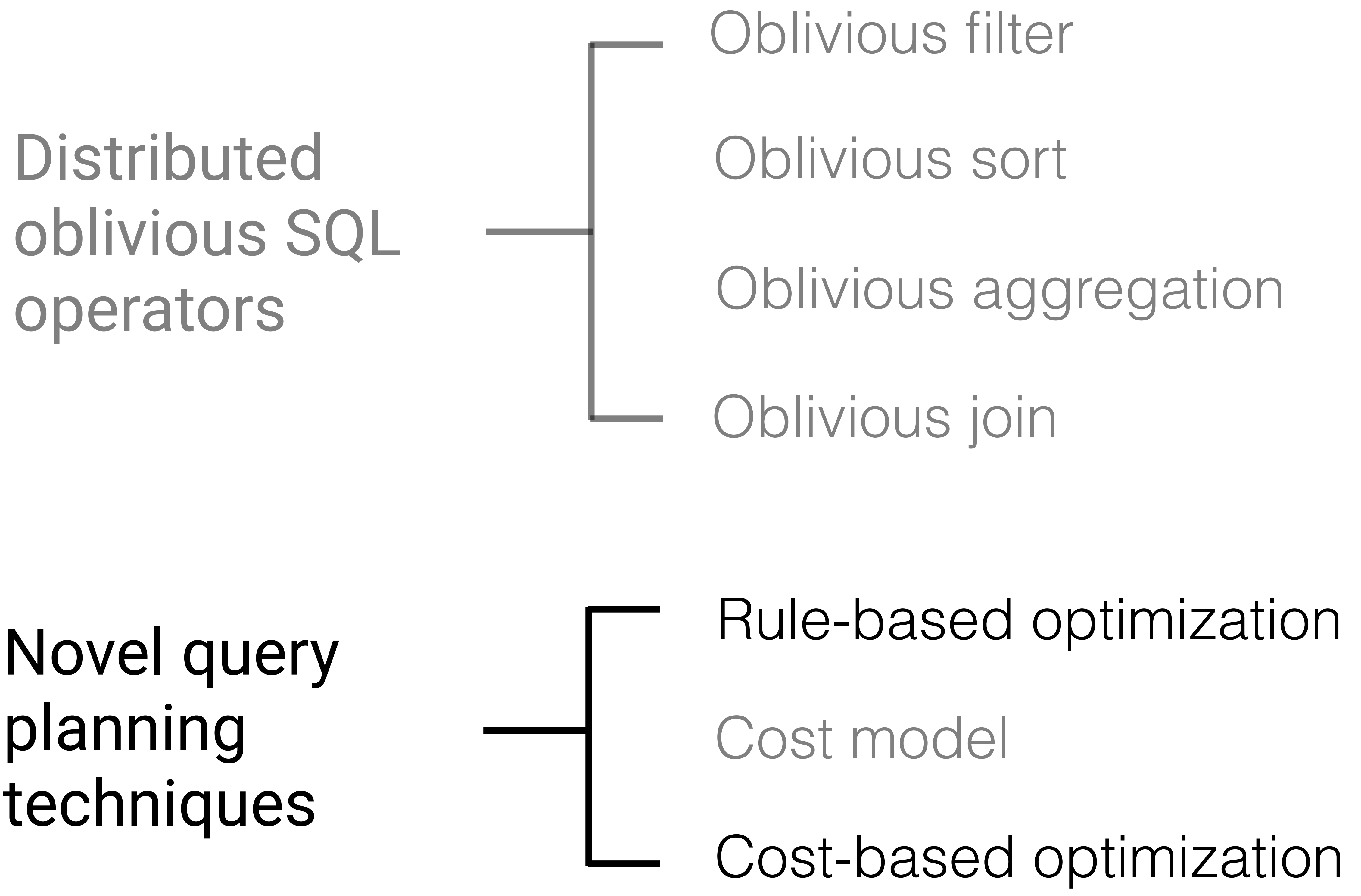
Sort

Final result

Aggregation has two sorts…

Can we do better?

Two-part solution:



Distributed oblivious SQL operators

- Oblivious filter
- Oblivious sort
- Oblivious aggregation
- Oblivious join

Novel query planning techniques

- Rule-based optimization
- Cost model
- Cost-based optimization

Two-part solution:

- Distributed oblivious SQL operators
  - Oblivious filter
  - Oblivious sort
  - Oblivious aggregation
  - Oblivious join
- Novel query planning techniques
  - **Rule-based optimization**
  - Cost model
  - Cost-based optimization

Rule-based optimization

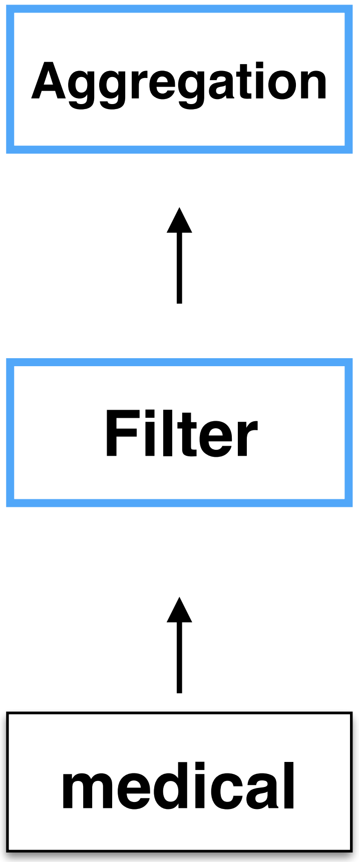
Rule-based optimization

```
SELECT count(*)
FROM medical
WHERE age > 30
GROUP BY disease
```

Rule-based optimization

```
SELECT count(*)
FROM medical
WHERE age > 30
GROUP BY disease
```

Logical op.



Insight 1

Insight 1

1. Split each logical operator into smaller Opaque operators

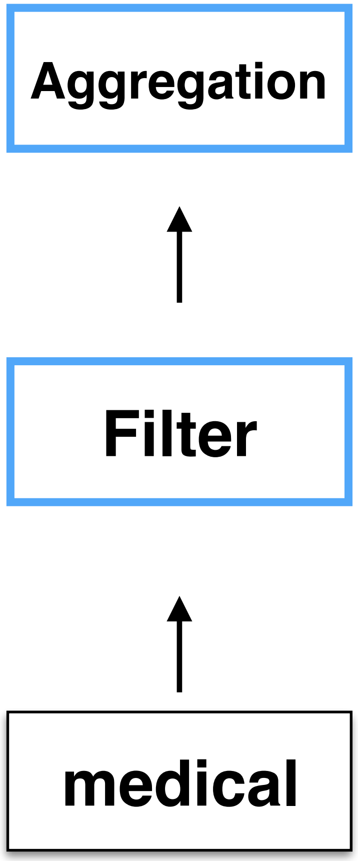
Insight 1

1. Split each logical operator into smaller Opaque operators

2. Take a global view across the plan to remove some Opaque operators

Rule-based optimization

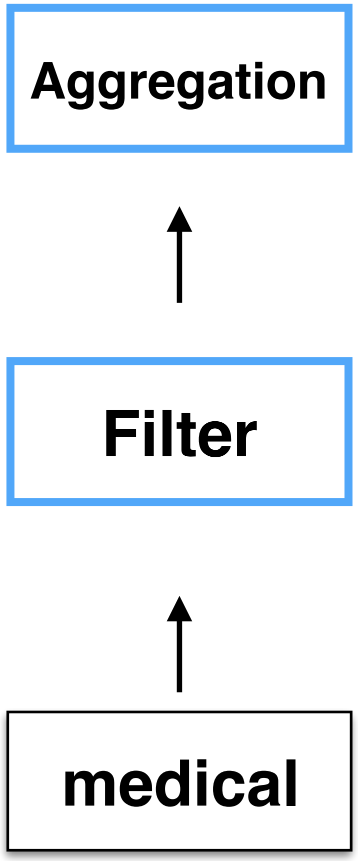
Logical op.



Rule-based optimization

Opaque op.

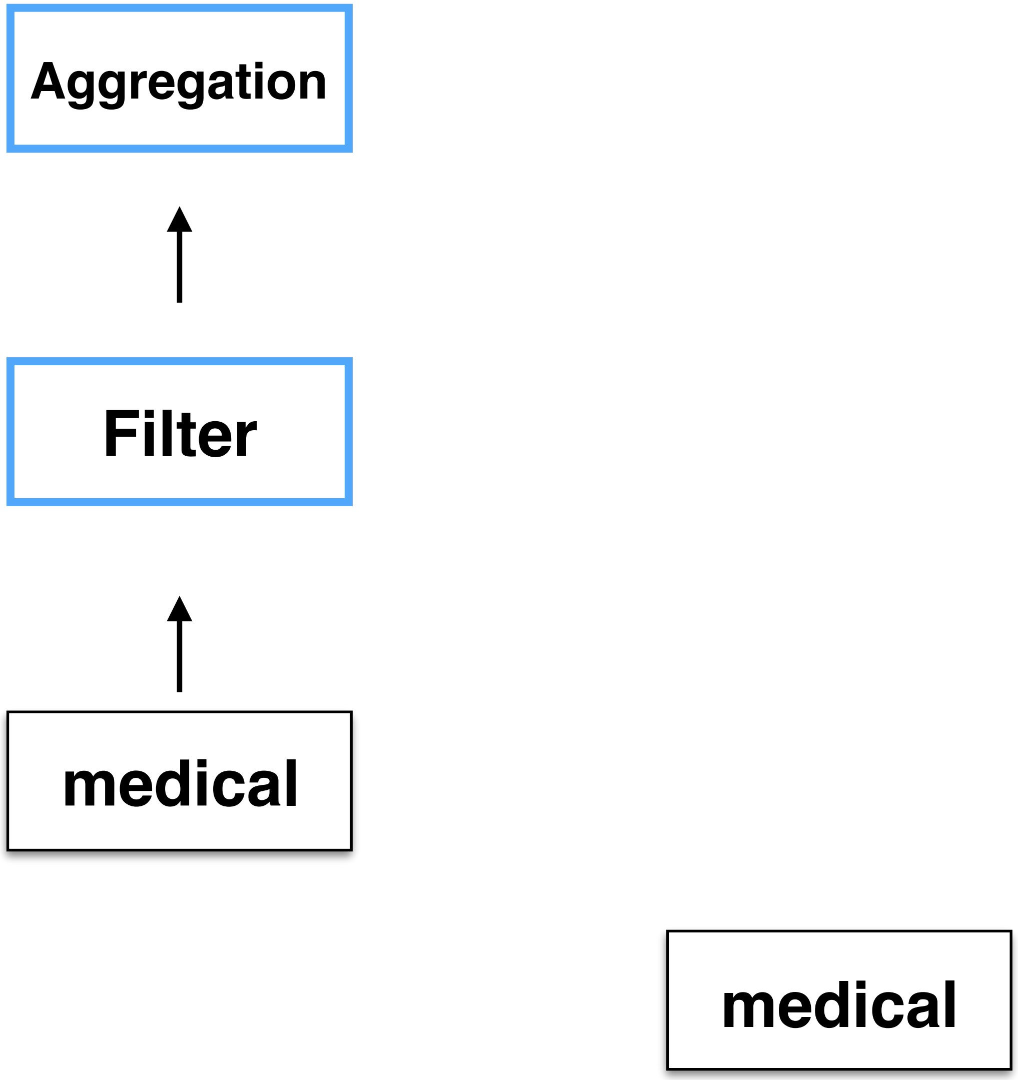
Logical op.



Rule-based optimization

Opaque op.

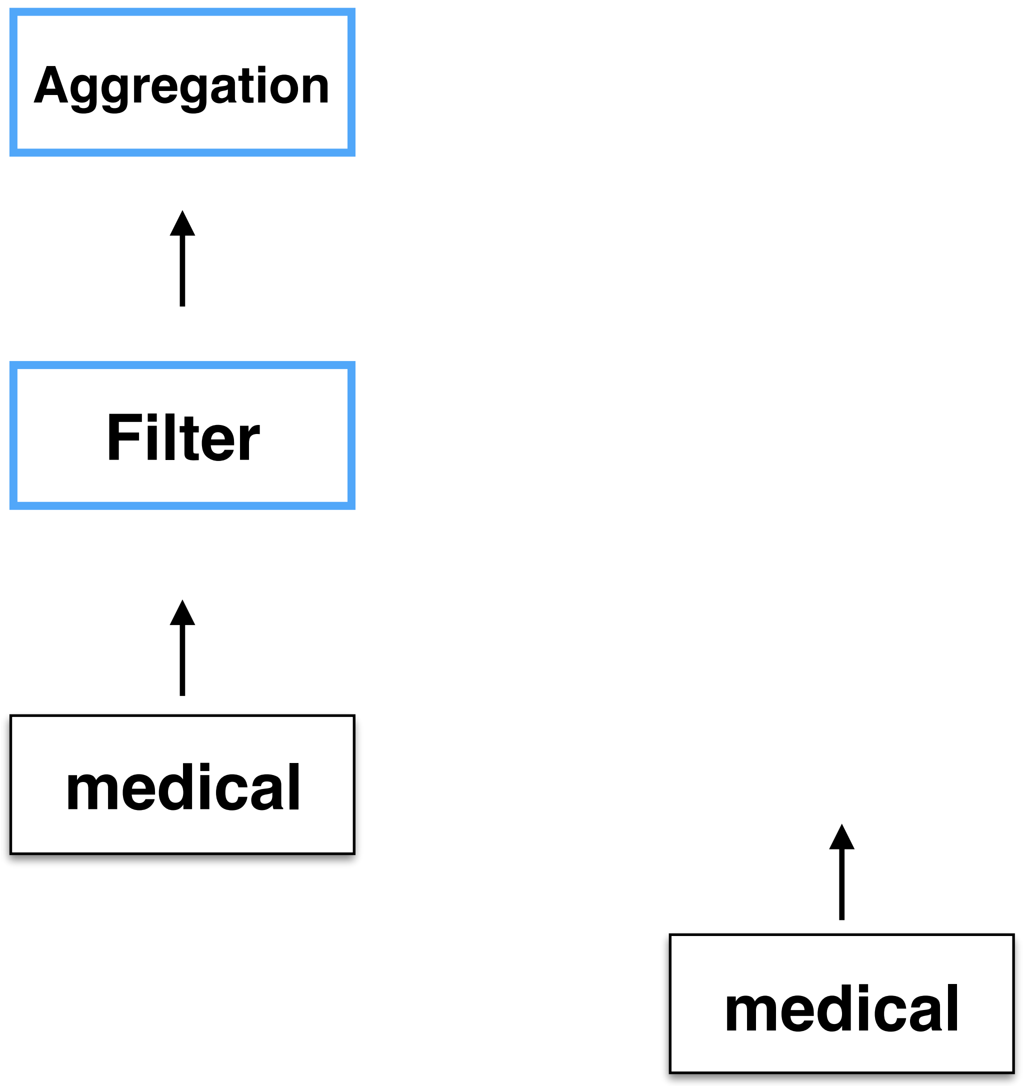
Logical op.



Rule-based optimization

Opaque op.

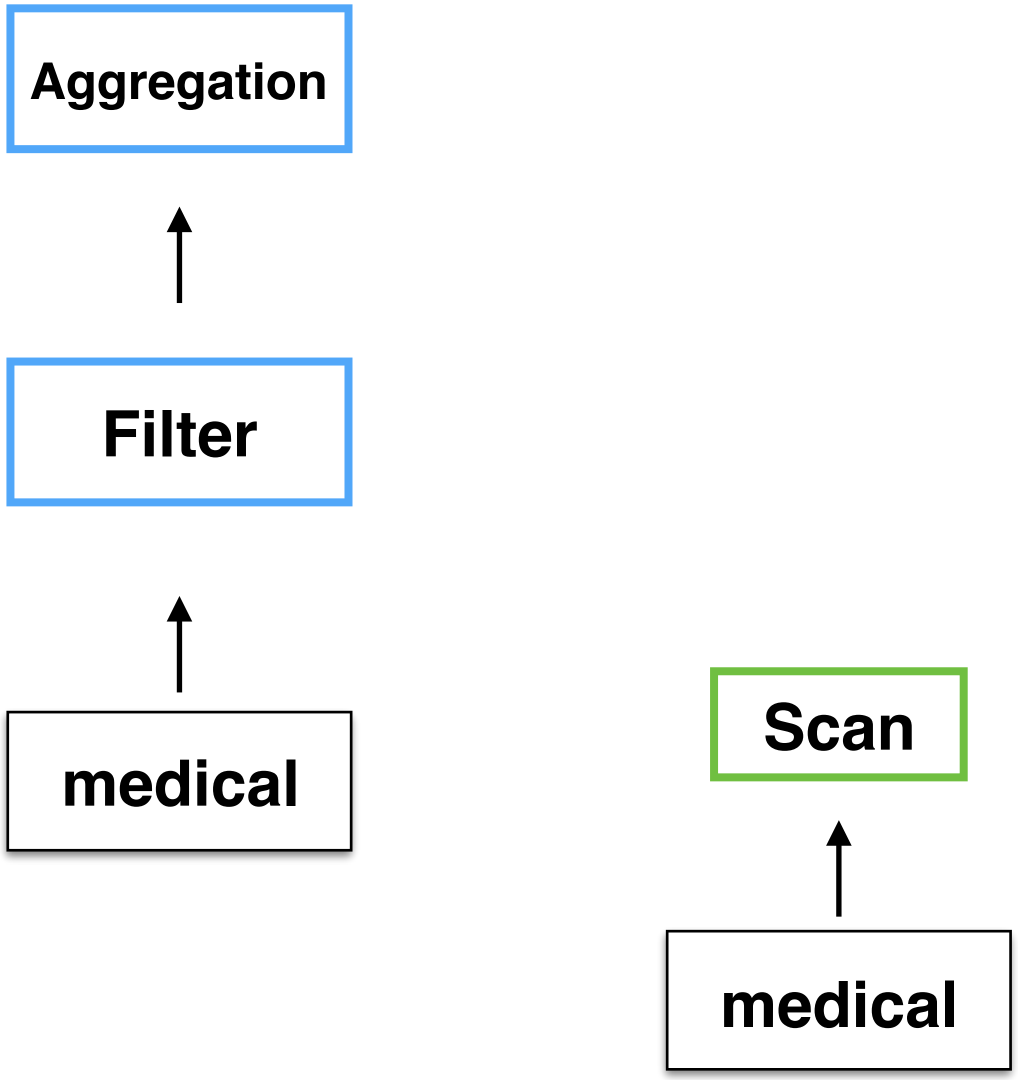
Logical op.



Rule-based optimization

Opaque op.

Logical op.



Aggregation

Filter

medical

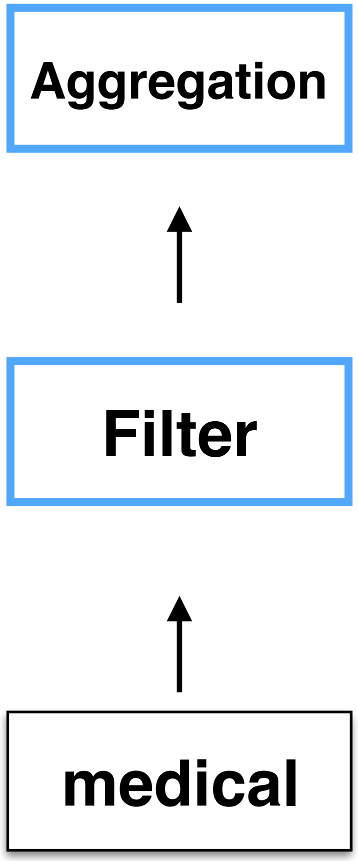
Scan

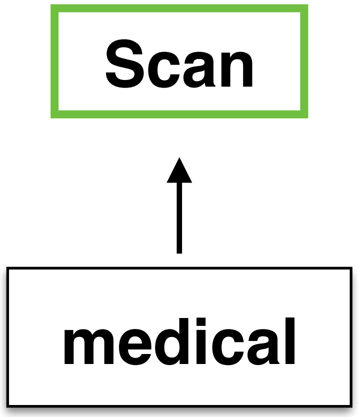
medical

Rule-based optimization

Opaque op.

Logical op.



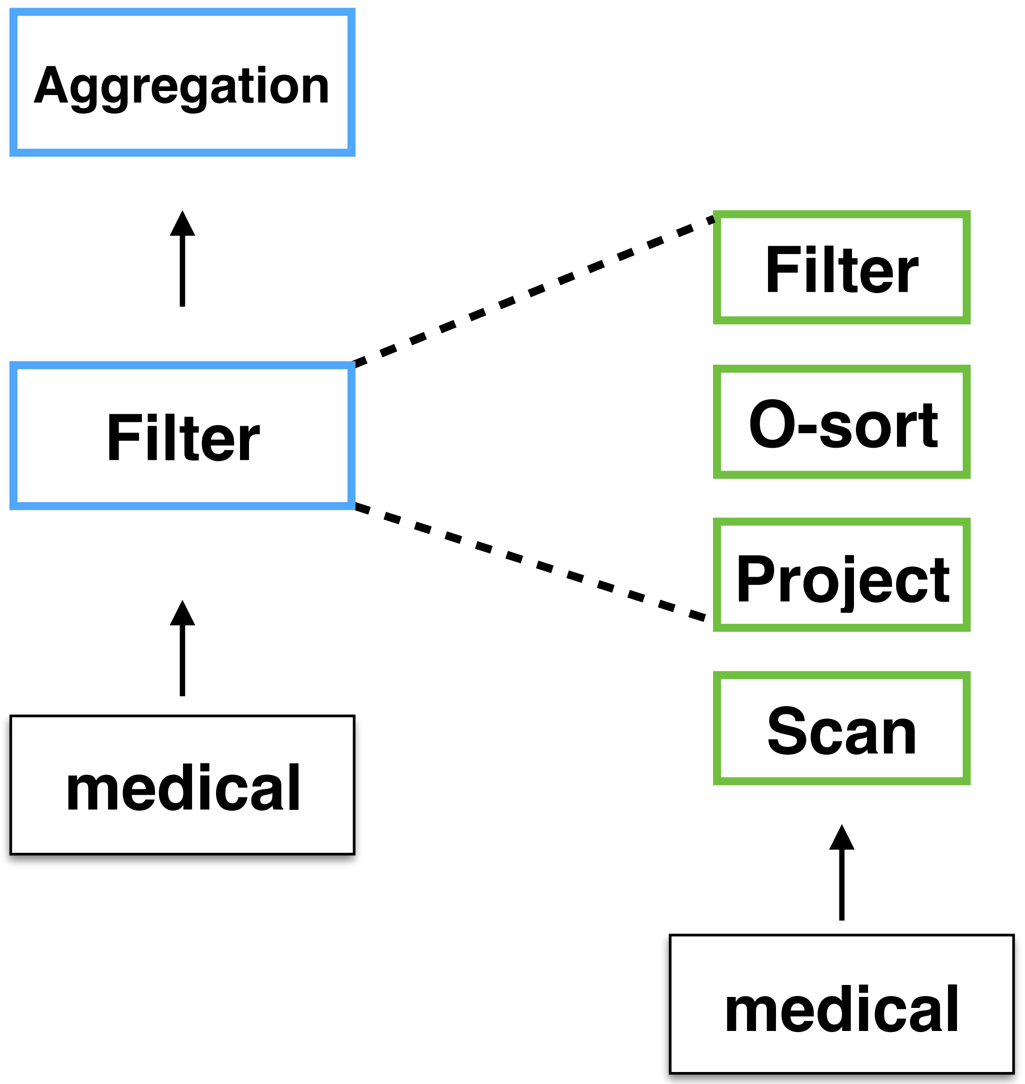


🔒 12809	Amanda D. Edwards	40	Diabetes
🔒 29489	Robert R. McGowan	56	Diabetes
🔒 13744	Kimberly R. Seay	51	Cancer
🔒 18740	Dennis G. Bates	32	Diabetes
🔒 32591	Donna R. Bridges	26	Diabetes
98329	Ronald S. Ogden	53	Cancer

Rule-based optimization

Opaque op.

Logical op.



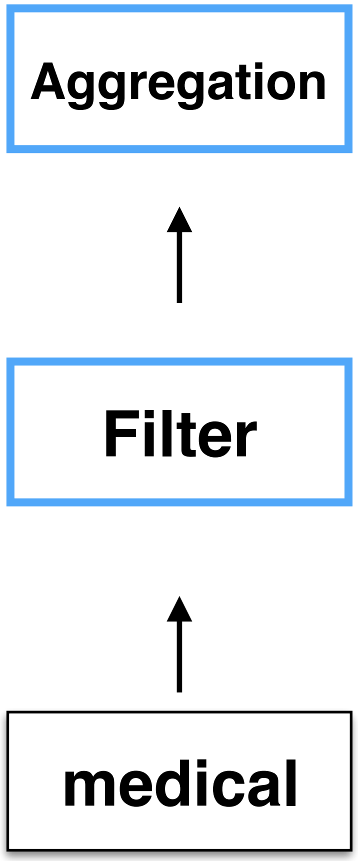
🔒 12809	Amanda D. Edwards	40	Diabetes
🔒 29489	Robert R. McGowan	56	Diabetes
🔒 13744	Kimberly R. Seay	51	Cancer
🔒 18740	Dennis G. Bates	32	Diabetes
🔒 32591	Donna R. Bridges	26	Diabetes
98329	Ronald S. Ogden	53	Cancer

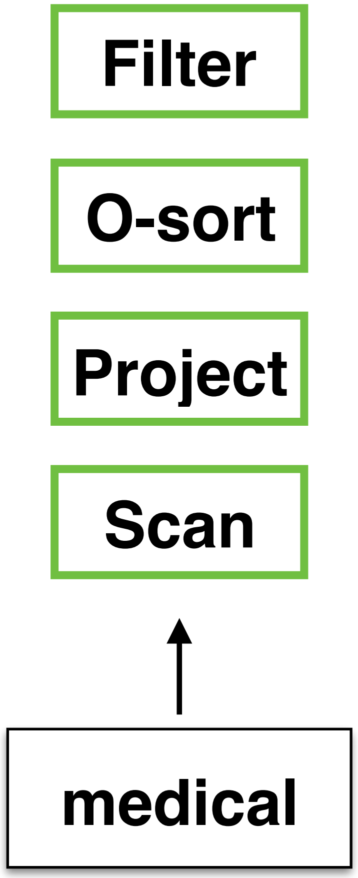
Rule-based optimization

Opaque op.

Logical op.

🔒	12809	Amanda D. Edwards	40	Diabetes
🔒	29489	Robert R. McGowan	56	Diabetes
🔒	13744	Kimberly R. Seay	51	Cancer
🔒	18740	Dennis G. Bates	32	Diabetes
🔒	32591	Donna R. Bridges	26	Diabetes
	98329	Ronald S. Ogden	53	Cancer

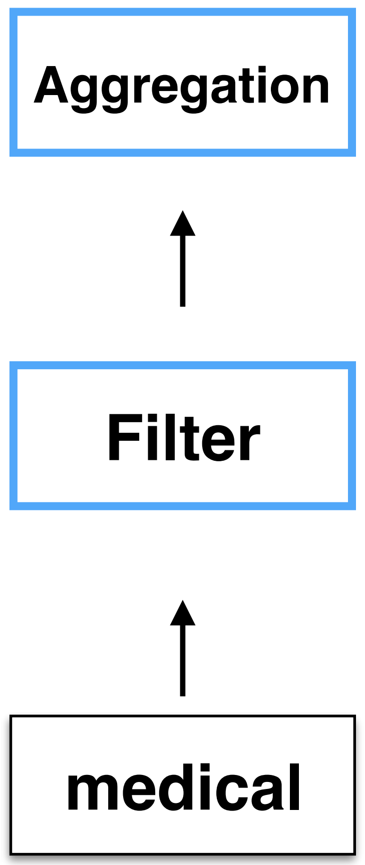


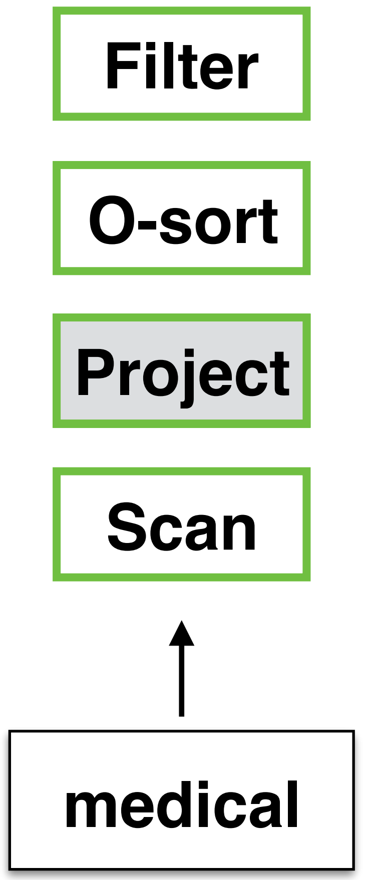


Rule-based optimization

Opaque op.

Logical op.

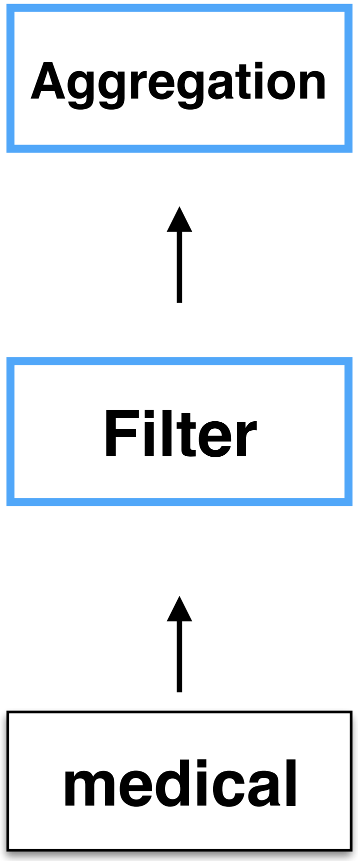




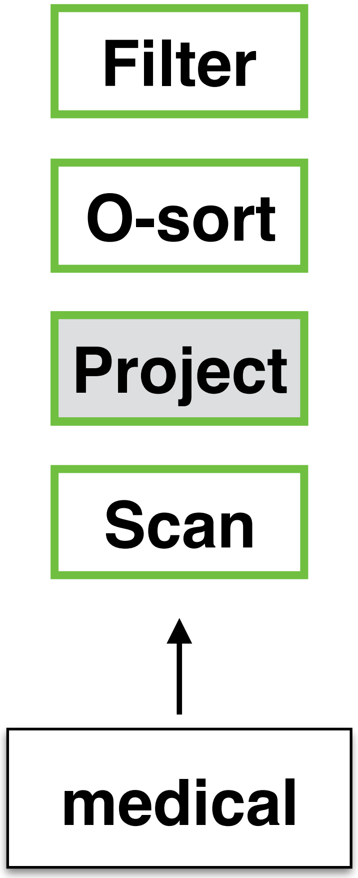
🔒 12809	Amanda D. Edwards	40	Diabetes
🔒 29489	Robert R. McGowan	56	Diabetes
🔒 13744	Kimberly R. Seay	51	Cancer
🔒 18740	Dennis G. Bates	32	Diabetes
🔒 32591	Donna R. Bridges	26	Diabetes
98329	Ronald S. Ogden	53	Cancer

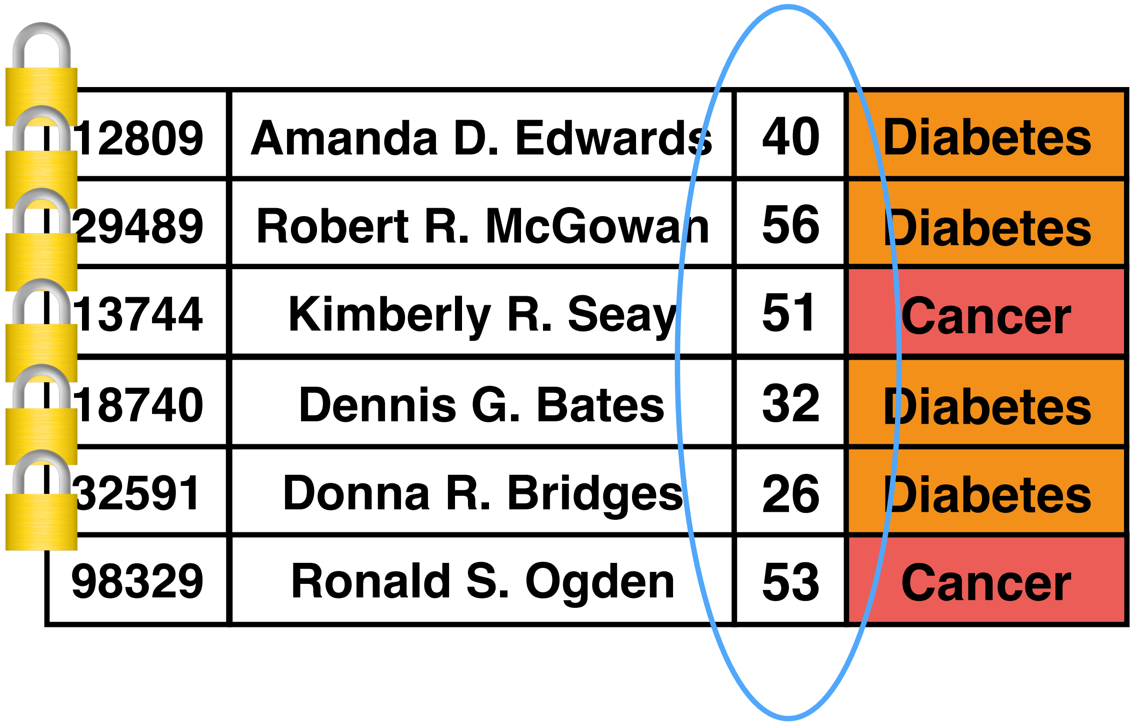
Rule-based optimization

Logical op.



Opaque op.



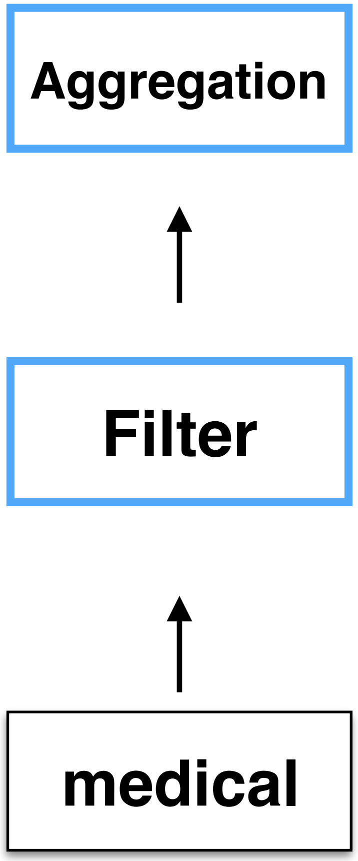


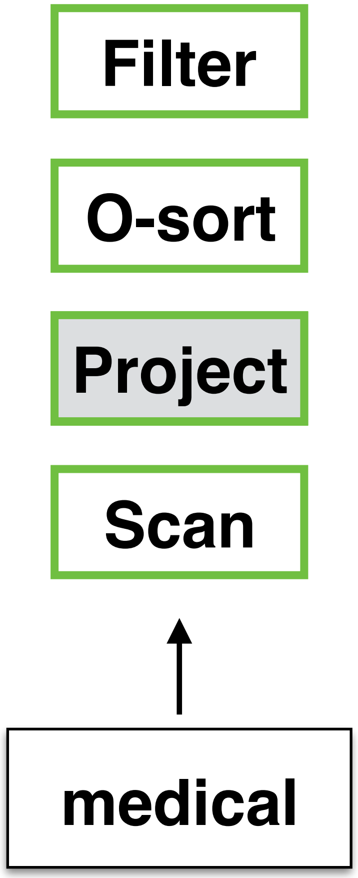
12809	Amanda D. Edwards	40	Diabetes
29489	Robert R. McGowan	56	Diabetes
13744	Kimberly R. Seay	51	Cancer
18740	Dennis G. Bates	32	Diabetes
32591	Donna R. Bridges	26	Diabetes
98329	Ronald S. Ogden	53	Cancer

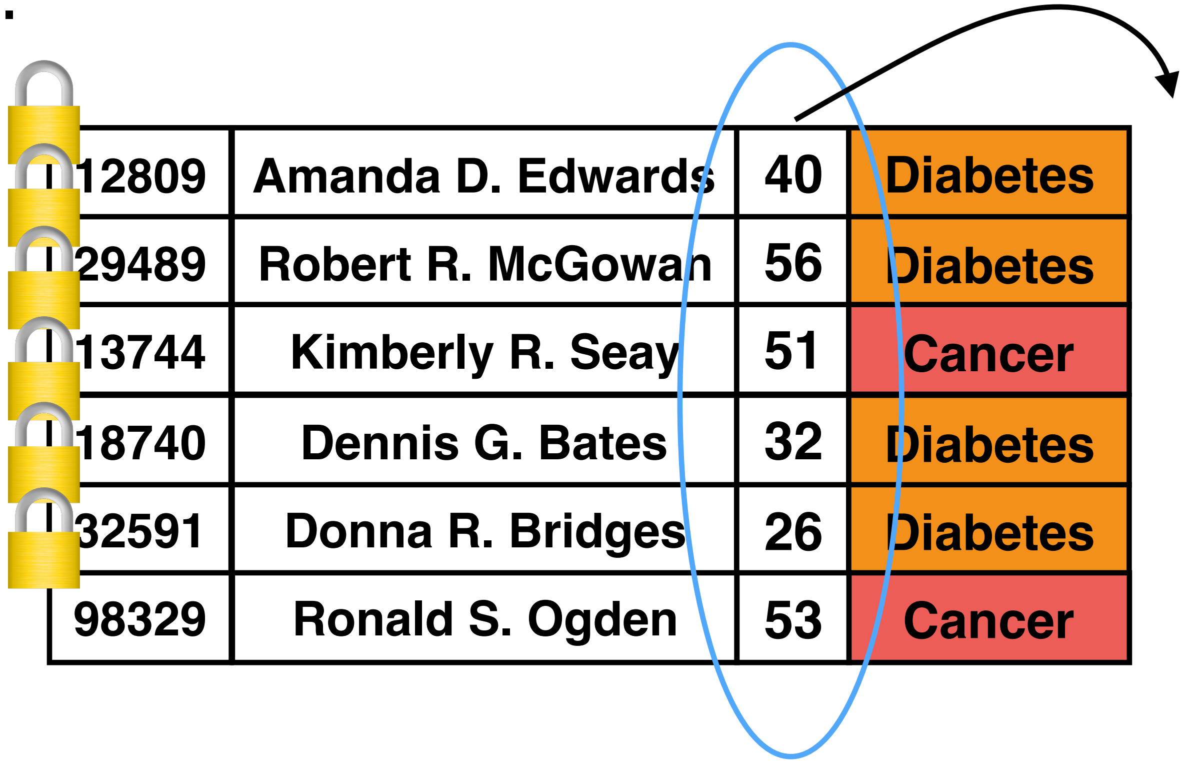
Rule-based optimization

Opaque op.

Logical op.





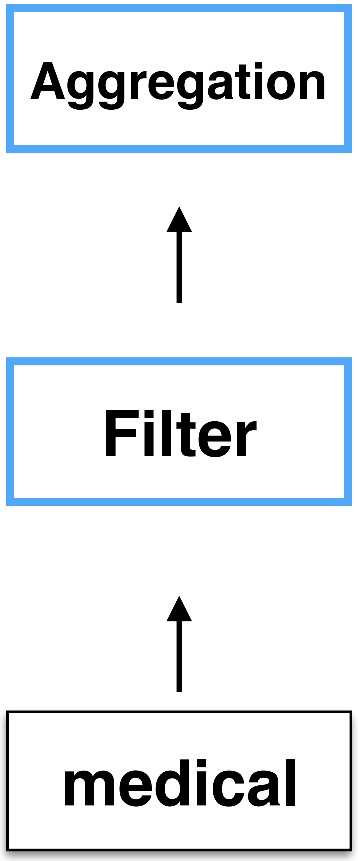


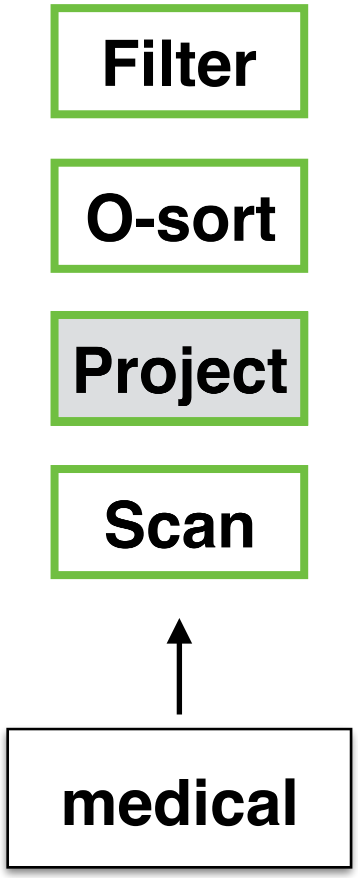
12809	Amanda D. Edwards	40	Diabetes
29489	Robert R. McGowan	56	Diabetes
13744	Kimberly R. Seay	51	Cancer
18740	Dennis G. Bates	32	Diabetes
32591	Donna R. Bridges	26	Diabetes
98329	Ronald S. Ogden	53	Cancer

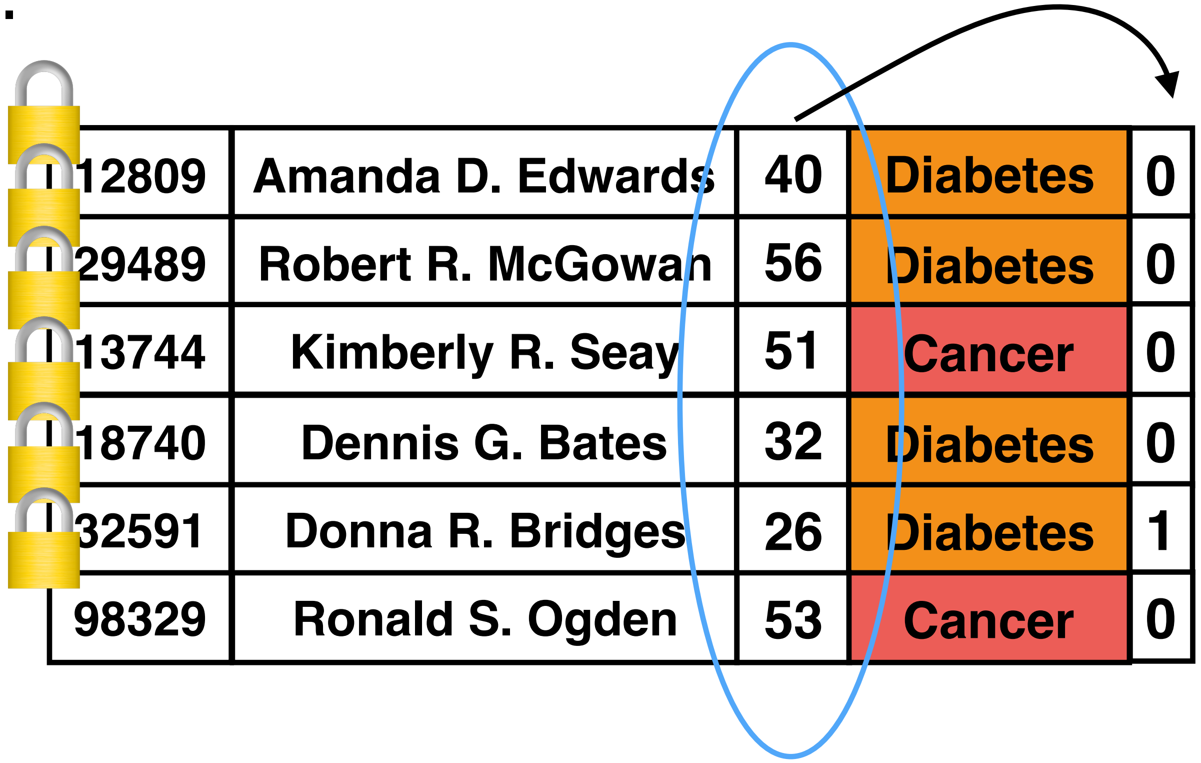
Rule-based optimization

Opaque op.

Logical op.





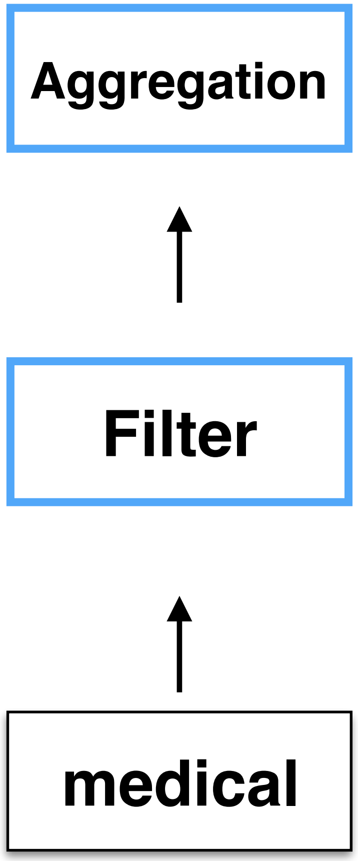


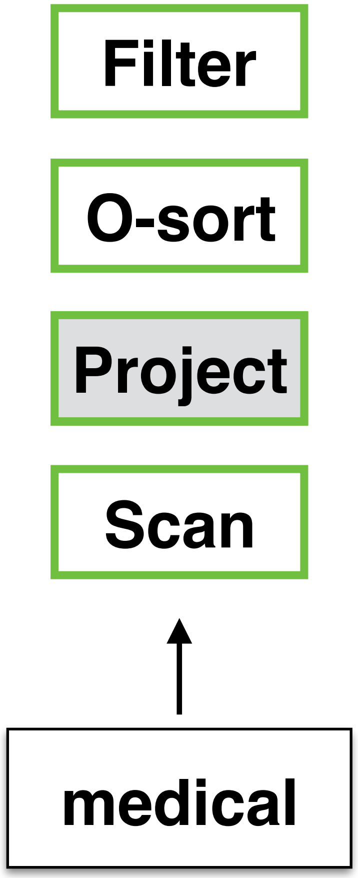
12809	Amanda D. Edwards	40	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
13744	Kimberly R. Seay	51	Cancer	0
18740	Dennis G. Bates	32	Diabetes	0
32591	Donna R. Bridges	26	Diabetes	1
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization

Opaque op.

Logical op.



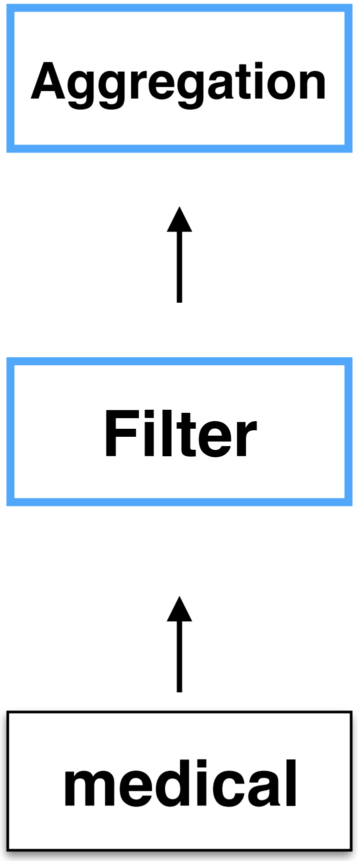


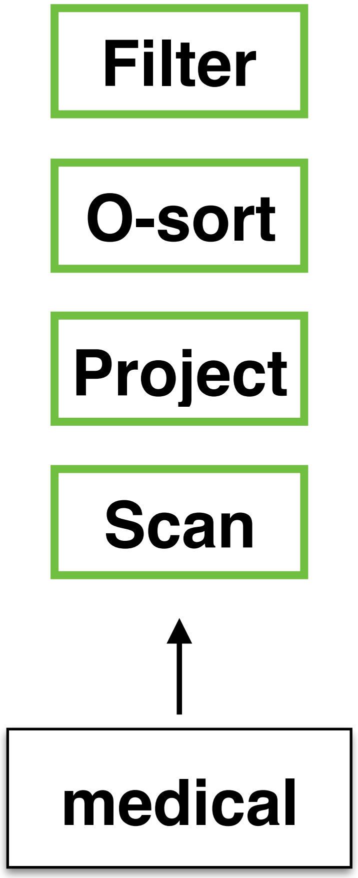
12809	Amanda D. Edwards	40	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
13744	Kimberly R. Seay	51	Cancer	0
18740	Dennis G. Bates	32	Diabetes	0
32591	Donna R. Bridges	26	Diabetes	1
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization

Opaque op.

Logical op.



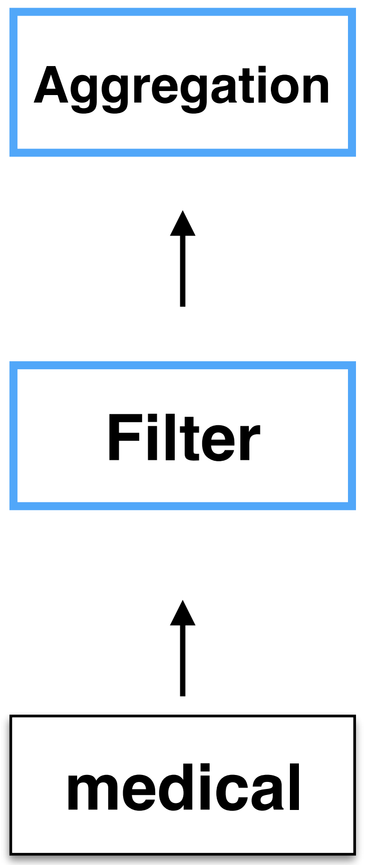


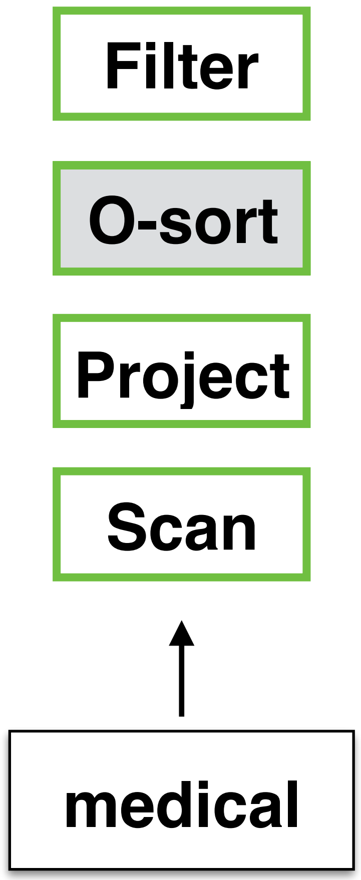
12809	Amanda D. Edwards	40	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
13744	Kimberly R. Seay	51	Cancer	0
18740	Dennis G. Bates	32	Diabetes	0
32591	Donna R. Bridges	26	Diabetes	1
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization

Opaque op.

Logical op.



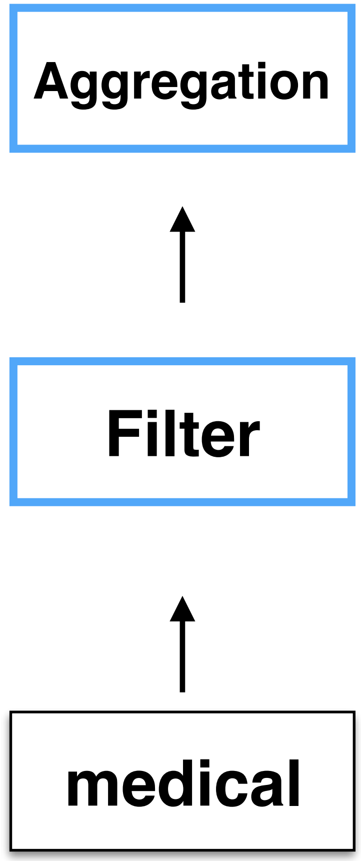


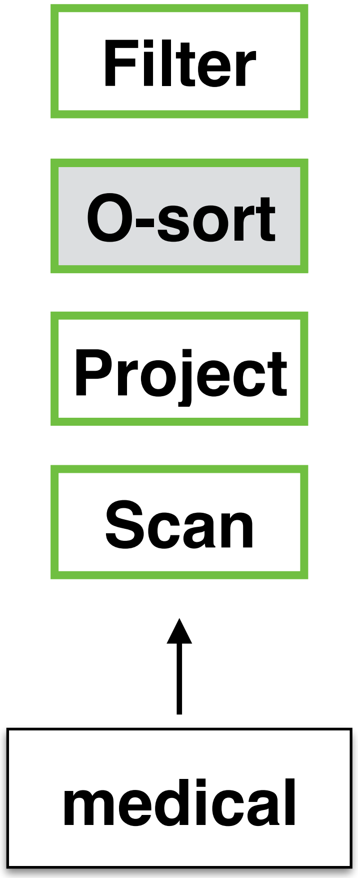
12809	Amanda D. Edwards	40	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
13744	Kimberly R. Seay	51	Cancer	0
18740	Dennis G. Bates	32	Diabetes	0
32591	Donna R. Bridges	26	Diabetes	1
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization

Opaque op.

Logical op.



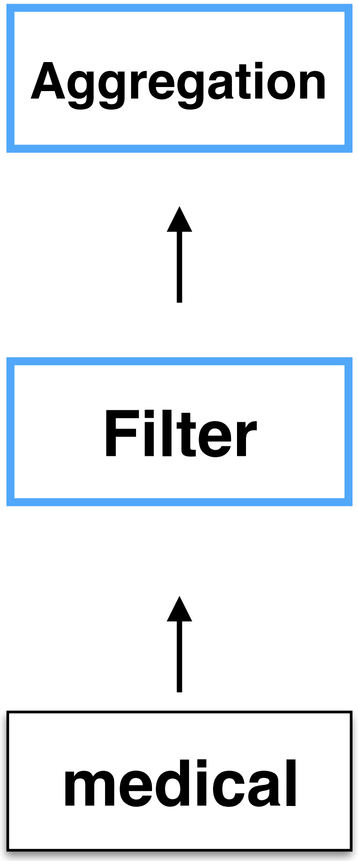


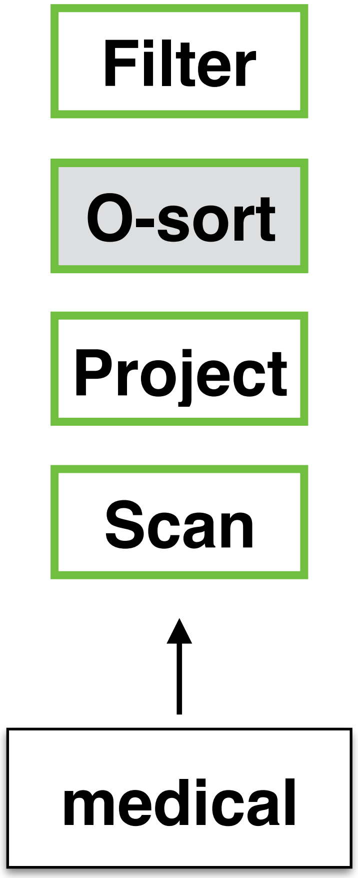
12809	Amanda D. Edwards	40	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
13744	Kimberly R. Seay	51	Cancer	0
18740	Dennis G. Bates	32	Diabetes	0
32591	Donna R. Bridges	26	Diabetes	1
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization

Opaque op.

Logical op.



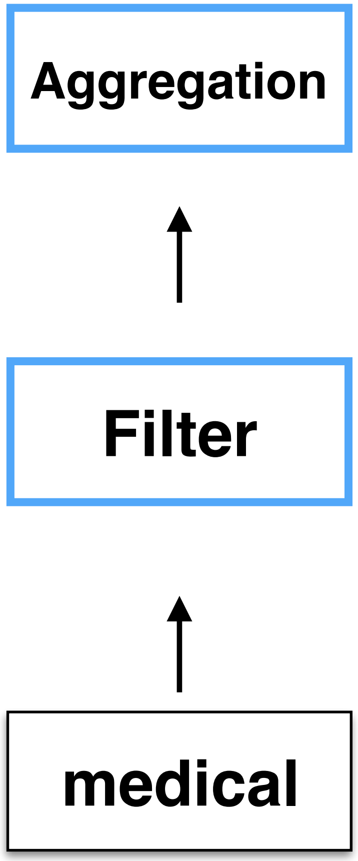


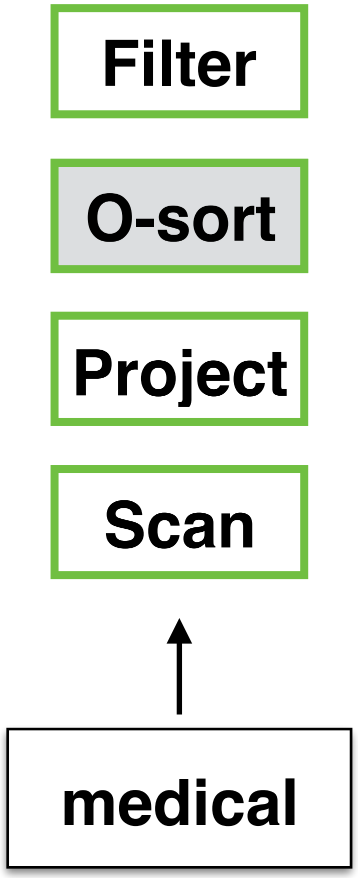
12809	Amanda D. Edwards	40	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
13744	Kimberly R. Seay	51	Cancer	0
18740	Dennis G. Bates	32	Diabetes	0
98329	Ronald S. Ogden	53	Cancer	0
32591	Donna R. Bridges	26	Diabetes	1

Rule-based optimization

Opaque op.

Logical op.



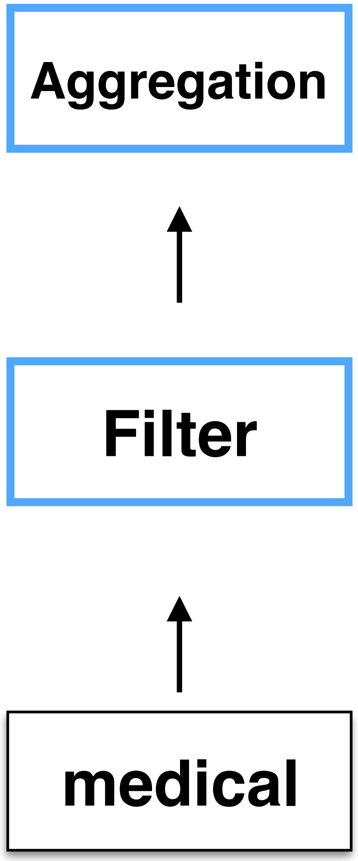


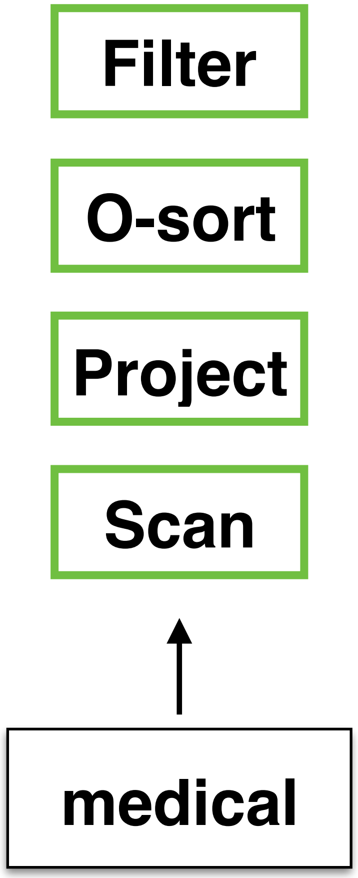
🔒 12809	Amanda D. Edwards	40	Diabetes	0
🔒 29489	Robert R. McGowan	56	Diabetes	0
🔒 13744	Kimberly R. Seay	51	Cancer	0
🔒 18740	Dennis G. Bates	32	Diabetes	0
🔒 98329	Ronald S. Ogden	53	Cancer	0
32591	Donna R. Bridges	26	Diabetes	1

Rule-based optimization

Opaque op.

Logical op.



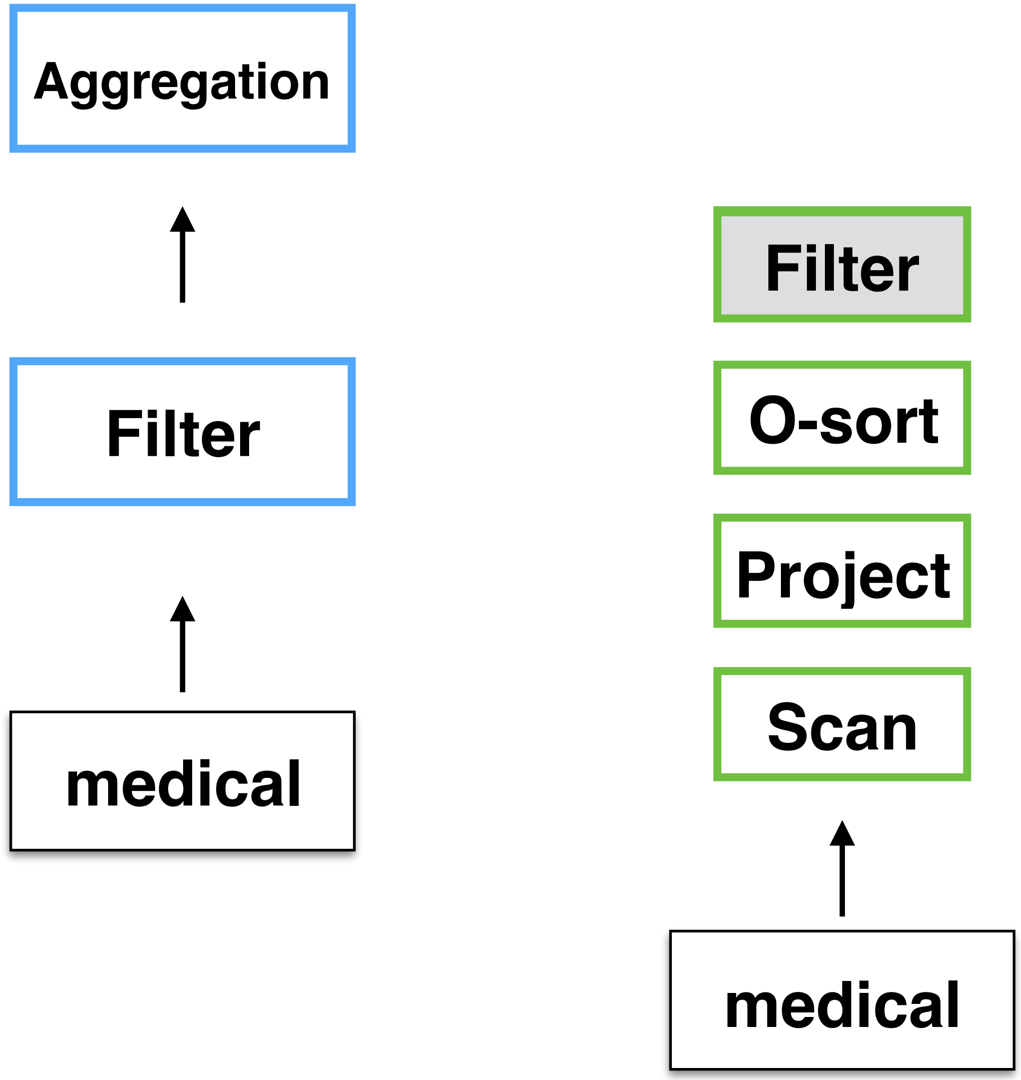


12809	Amanda D. Edwards	40	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
13744	Kimberly R. Seay	51	Cancer	0
18740	Dennis G. Bates	32	Diabetes	0
98329	Ronald S. Ogden	53	Cancer	0
32591	Donna R. Bridges	26	Diabetes	1

Rule-based optimization

Opaque op.

Logical op.

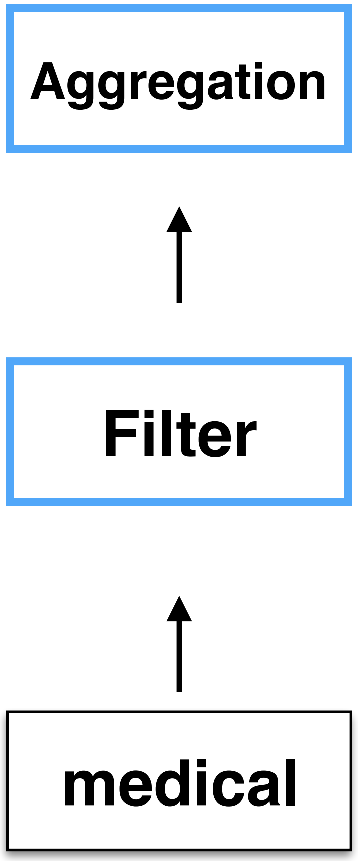


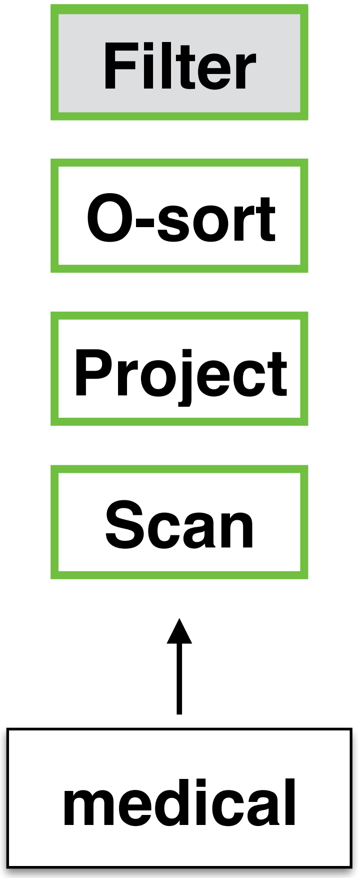
🔒	12809	Amanda D. Edwards	40	Diabetes	0
🔒	29489	Robert R. McGowan	56	Diabetes	0
🔒	13744	Kimberly R. Seay	51	Cancer	0
🔒	18740	Dennis G. Bates	32	Diabetes	0
🔒	98329	Ronald S. Ogden	53	Cancer	0
	32591	Donna R. Bridges	26	Diabetes	1

Rule-based optimization

Opaque op.

Logical op.

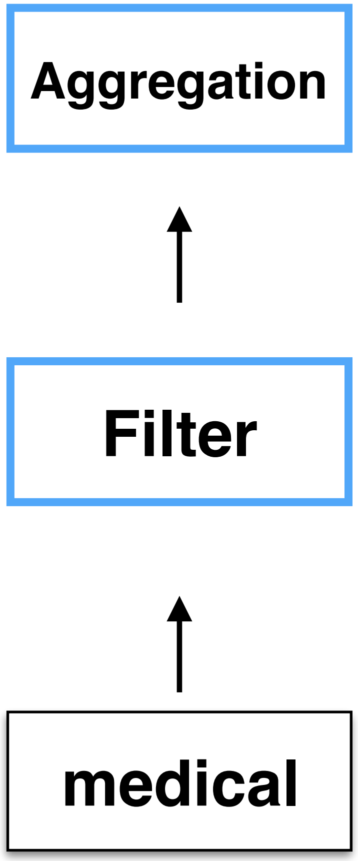




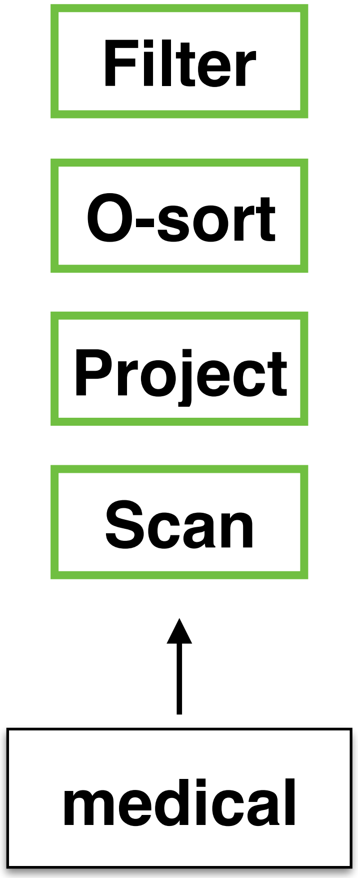
🔒 12809	Amanda D. Edwards	40	Diabetes	0
🔒 29489	Robert R. McGowan	56	Diabetes	0
🔒 13744	Kimberly R. Seay	51	Cancer	0
🔒 18740	Dennis G. Bates	32	Diabetes	0
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization

Logical op.

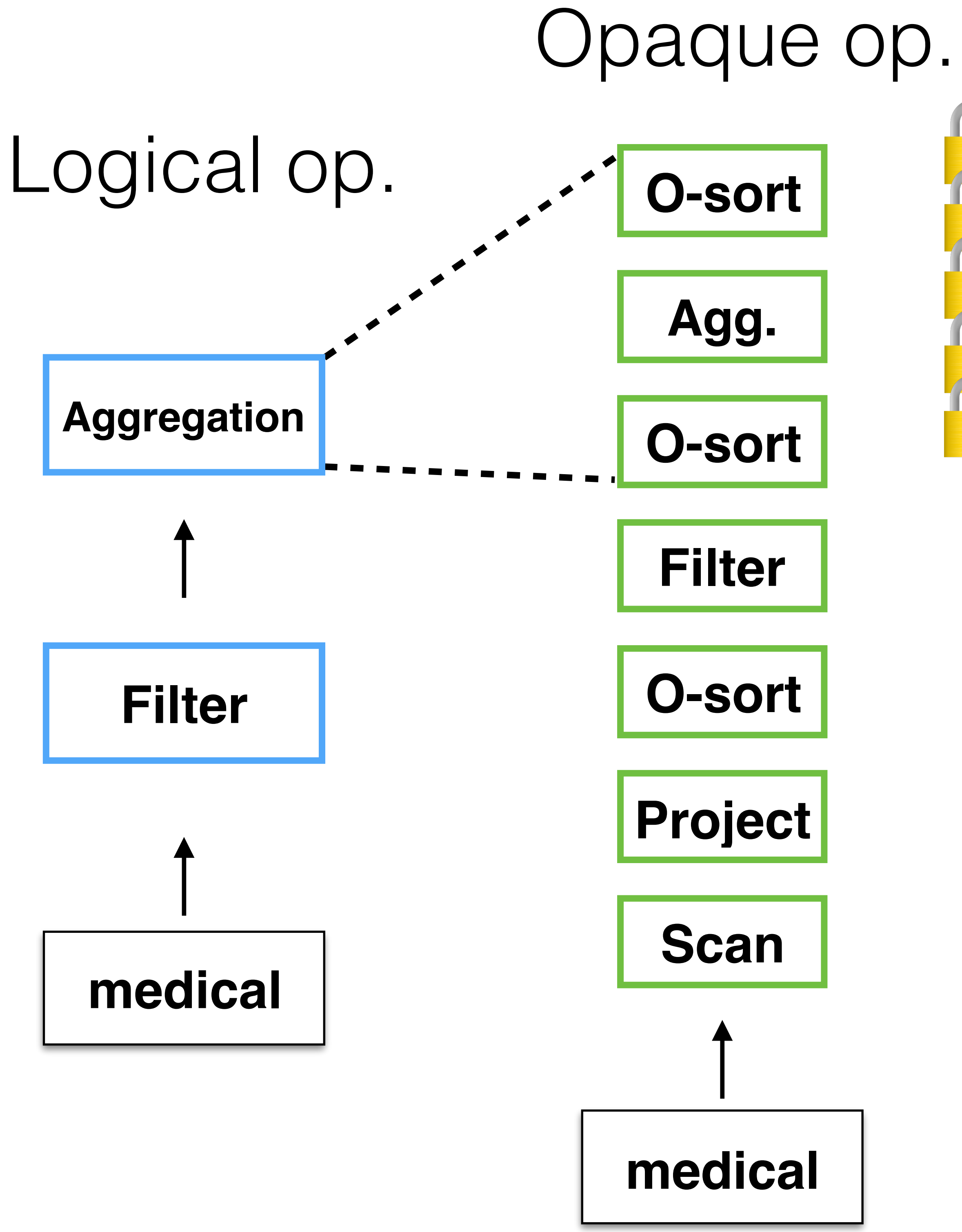


Opaque op.



🔒	12809	Amanda D. Edwards	40	Diabetes	0
🔒	29489	Robert R. McGowan	56	Diabetes	0
🔒	13744	Kimberly R. Seay	51	Cancer	0
🔒	18740	Dennis G. Bates	32	Diabetes	0
	98329	Ronald S. Ogden	53	Cancer	0

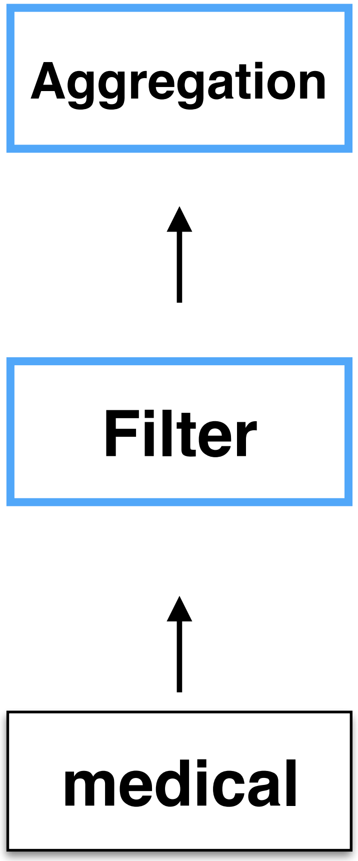
Rule-based optimization



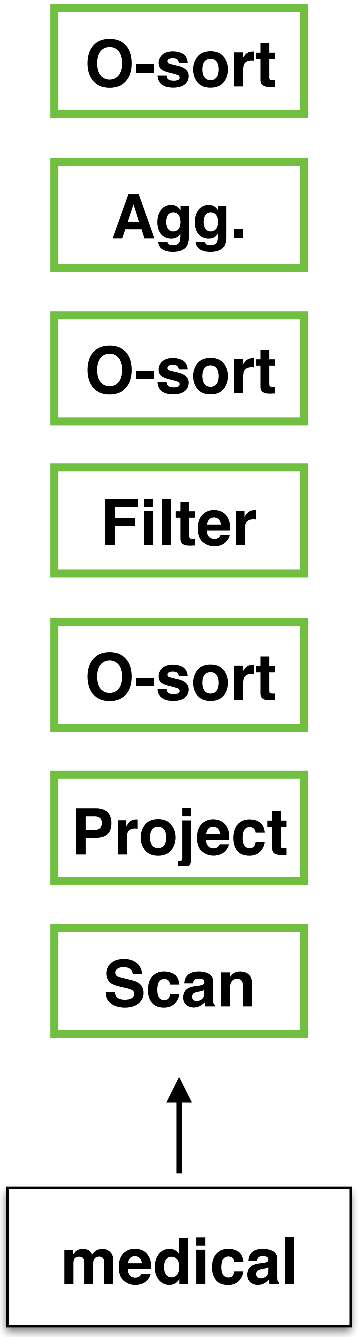
🔒 12809	Amanda D. Edwards	40	Diabetes	0
🔒 29489	Robert R. McGowan	56	Diabetes	0
🔒 13744	Kimberly R. Seay	51	Cancer	0
🔒 18740	Dennis G. Bates	32	Diabetes	0
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization

Logical op.



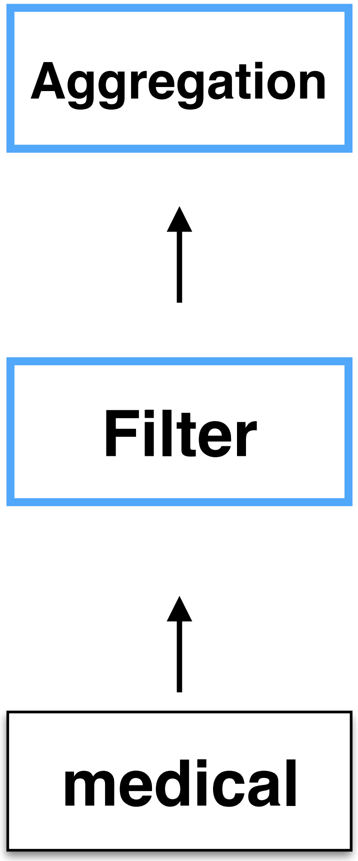
Opaque op.



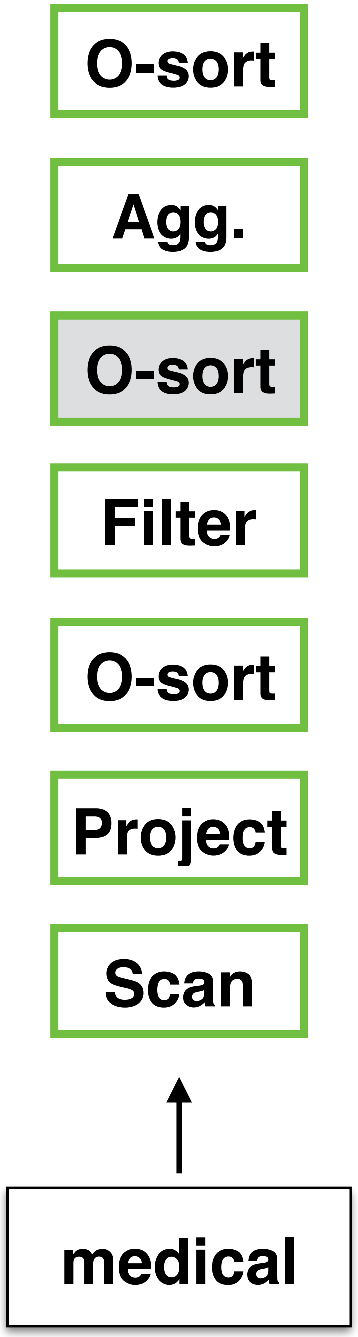
🔒 12809	Amanda D. Edwards	40	Diabetes	0
🔒 29489	Robert R. McGowan	56	Diabetes	0
🔒 13744	Kimberly R. Seay	51	Cancer	0
🔒 18740	Dennis G. Bates	32	Diabetes	0
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization

Logical op.



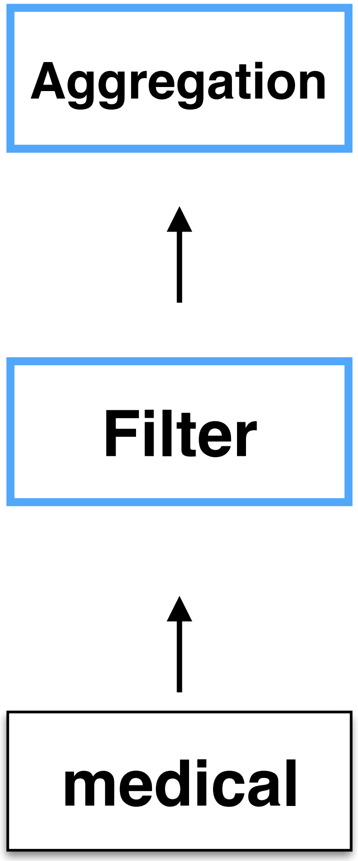
Opaque op.



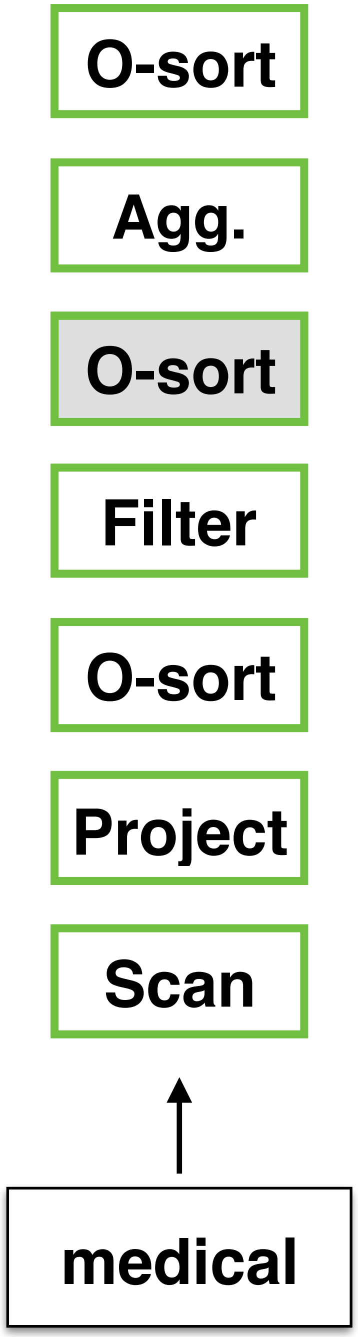
12809	Amanda D. Edwards	40	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
13744	Kimberly R. Seay	51	Cancer	0
18740	Dennis G. Bates	32	Diabetes	0
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization

Logical op.



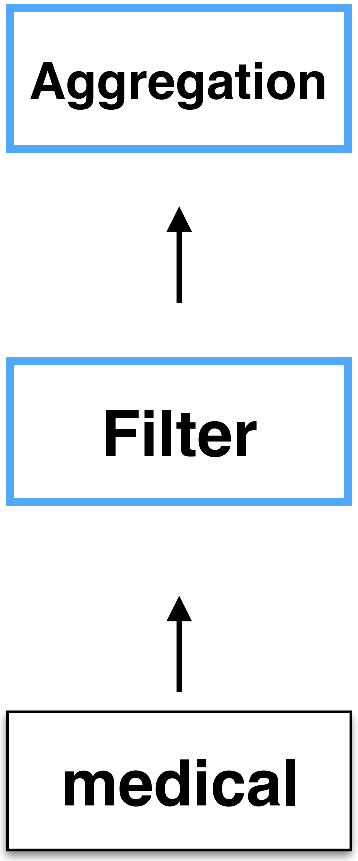
Opaque op.



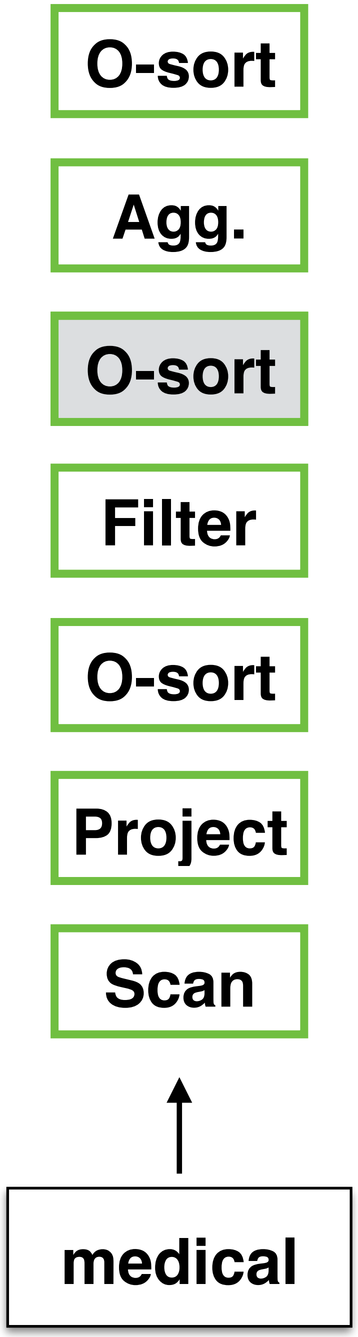
13744	Kimberly R. Seay	51	Cancer	0
98329	Ronald S. Ogden	53	Cancer	0
29489	Robert R. McGowan	56	Diabetes	0
18740	Dennis G. Bates	32	Diabetes	0
12809	Amanda D. Edwards	40	Diabetes	0

Rule-based optimization

Logical op.



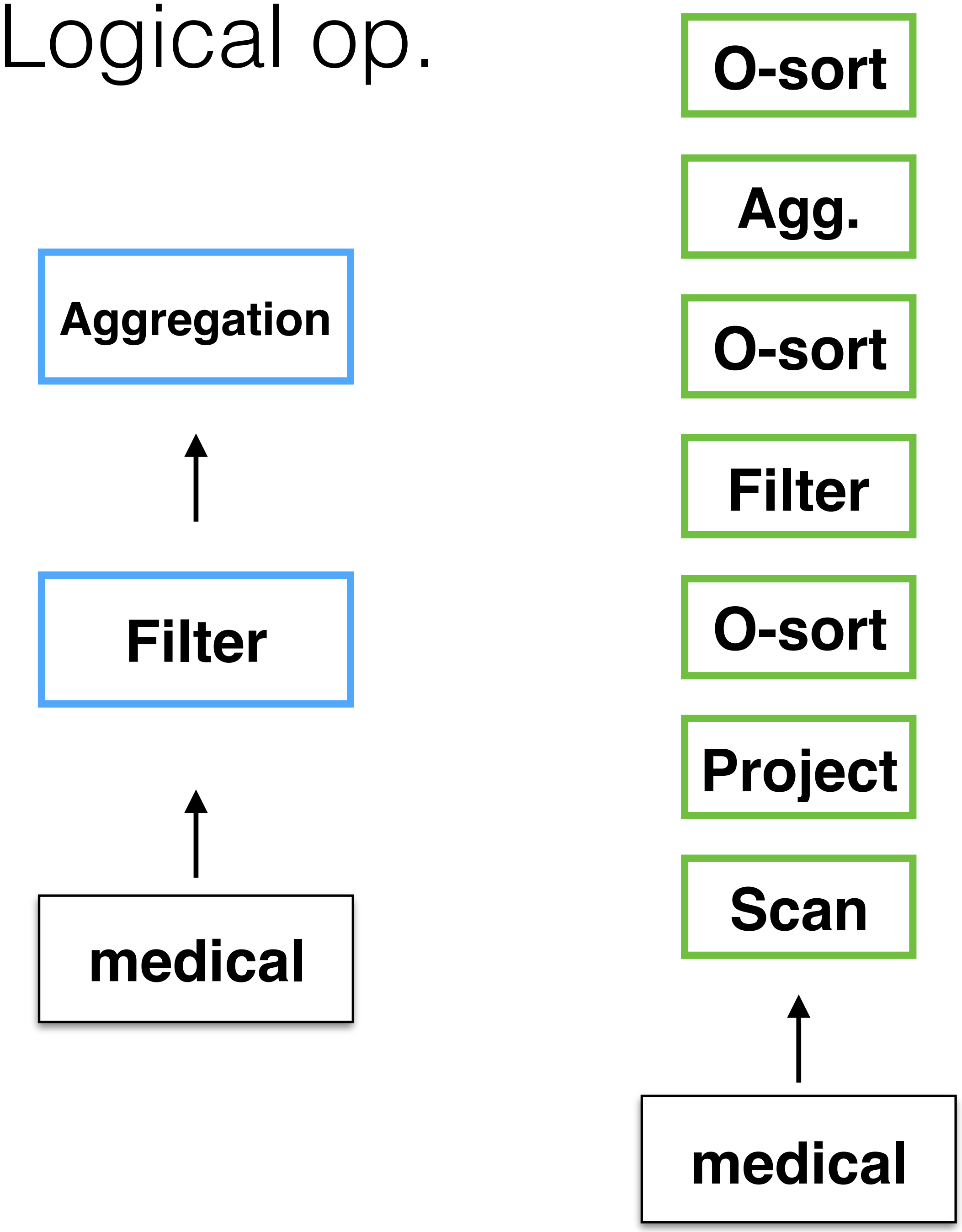
Opaque op.



🔒 13744	Kimberly R. Seay	51	Cancer	0
🔒 98329	Ronald S. Ogden	53	Cancer	0
🔒 29489	Robert R. McGowan	56	Diabetes	0
🔒 18740	Dennis G. Bates	32	Diabetes	0
🔒 12809	Amanda D. Edwards	40	Diabetes	0

Can we remove any sort?

Rule-based optimization



Logical op.

Aggregation

Filter

medical

Opaque op.

O-sort

Agg.

O-sort

Filter

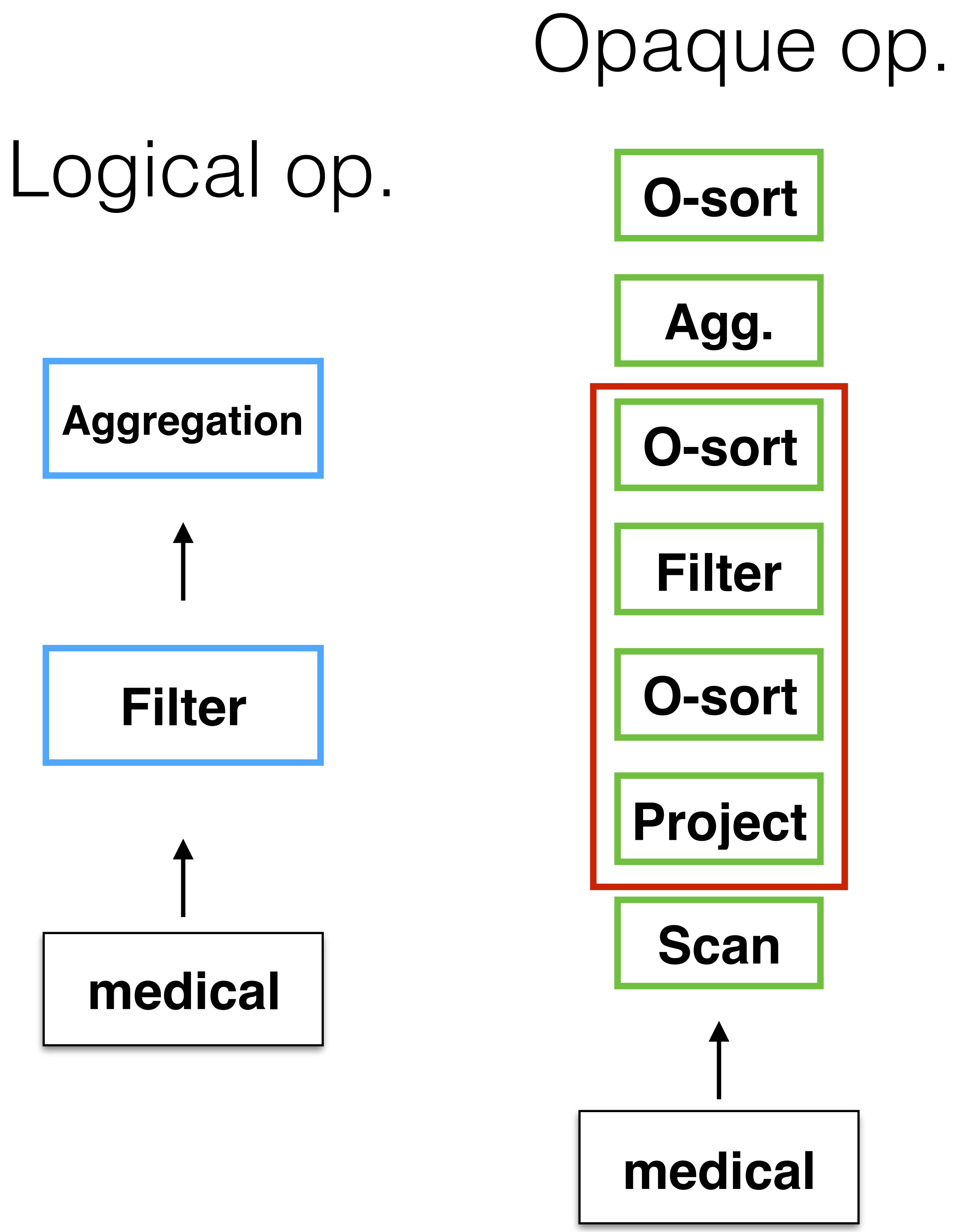
O-sort

Project

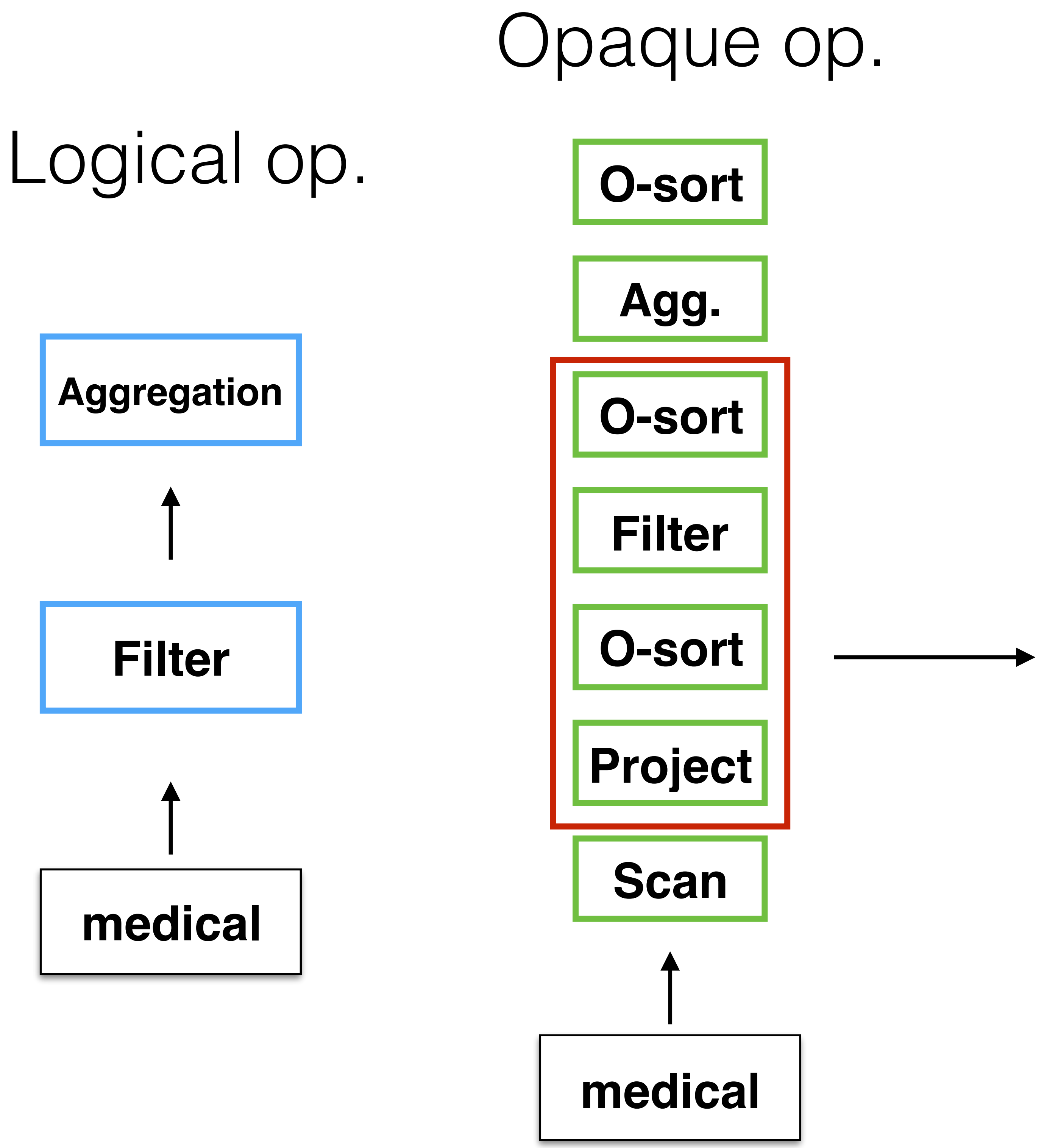
Scan

medical

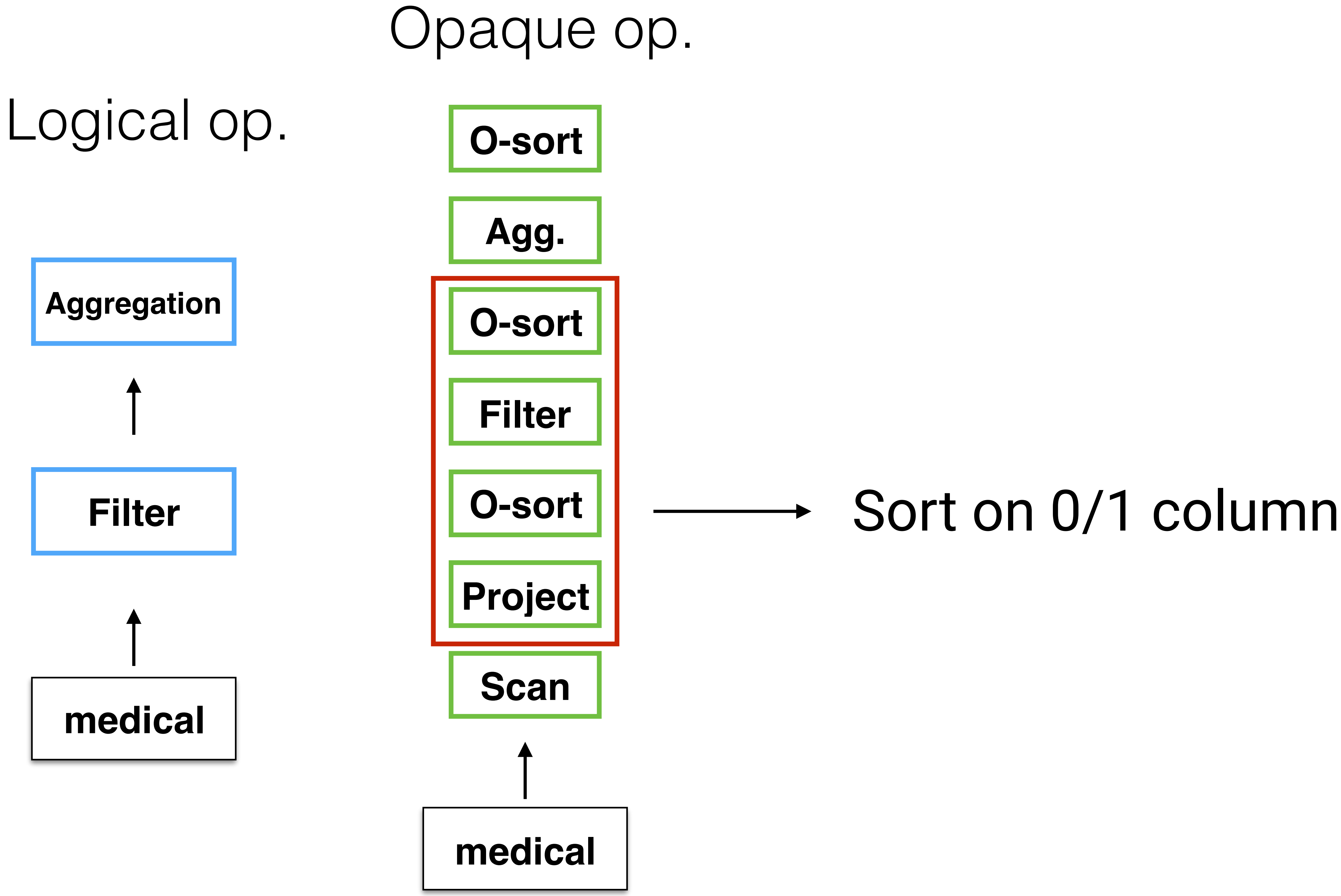
Rule-based optimization



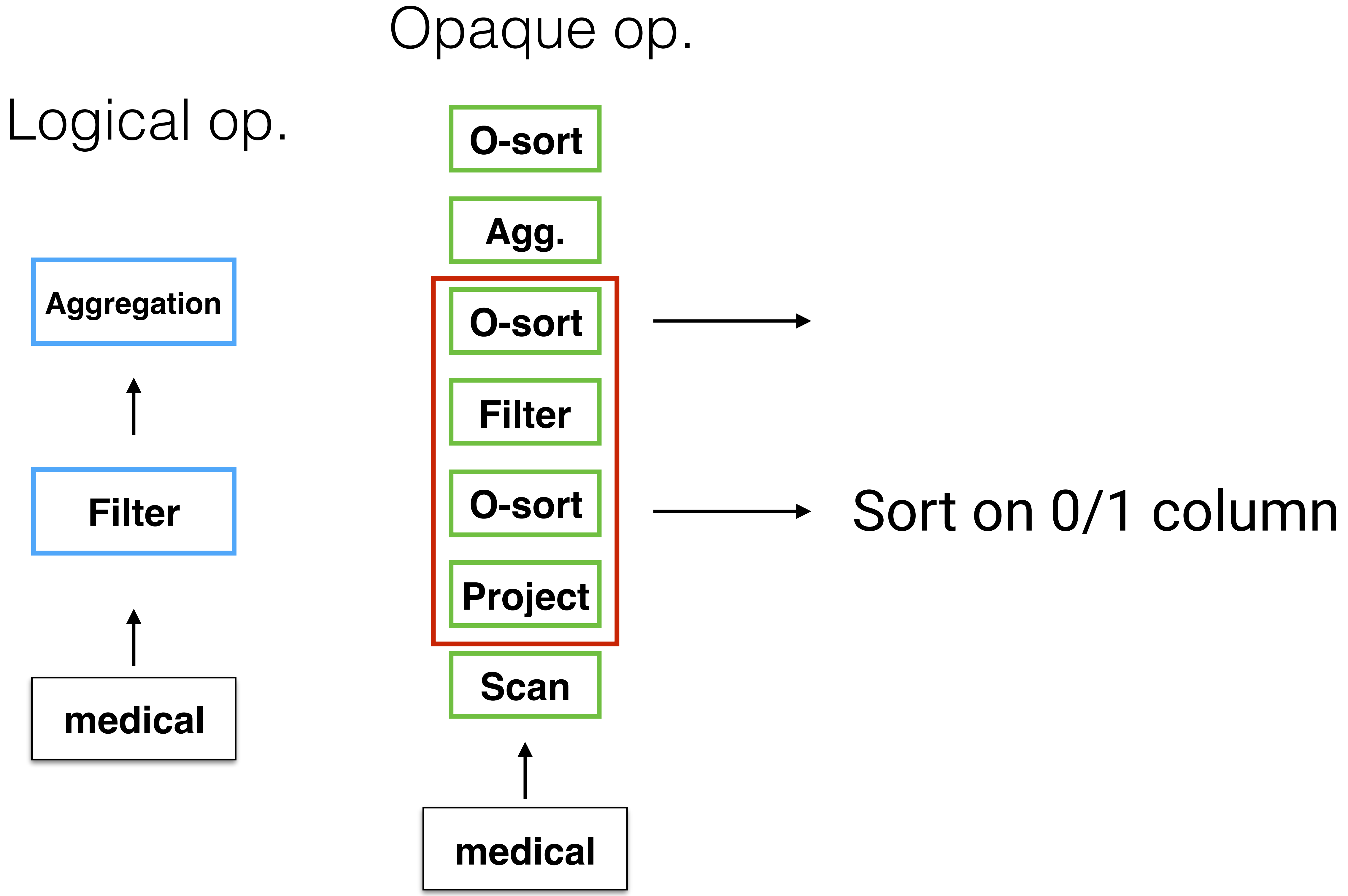
Rule-based optimization



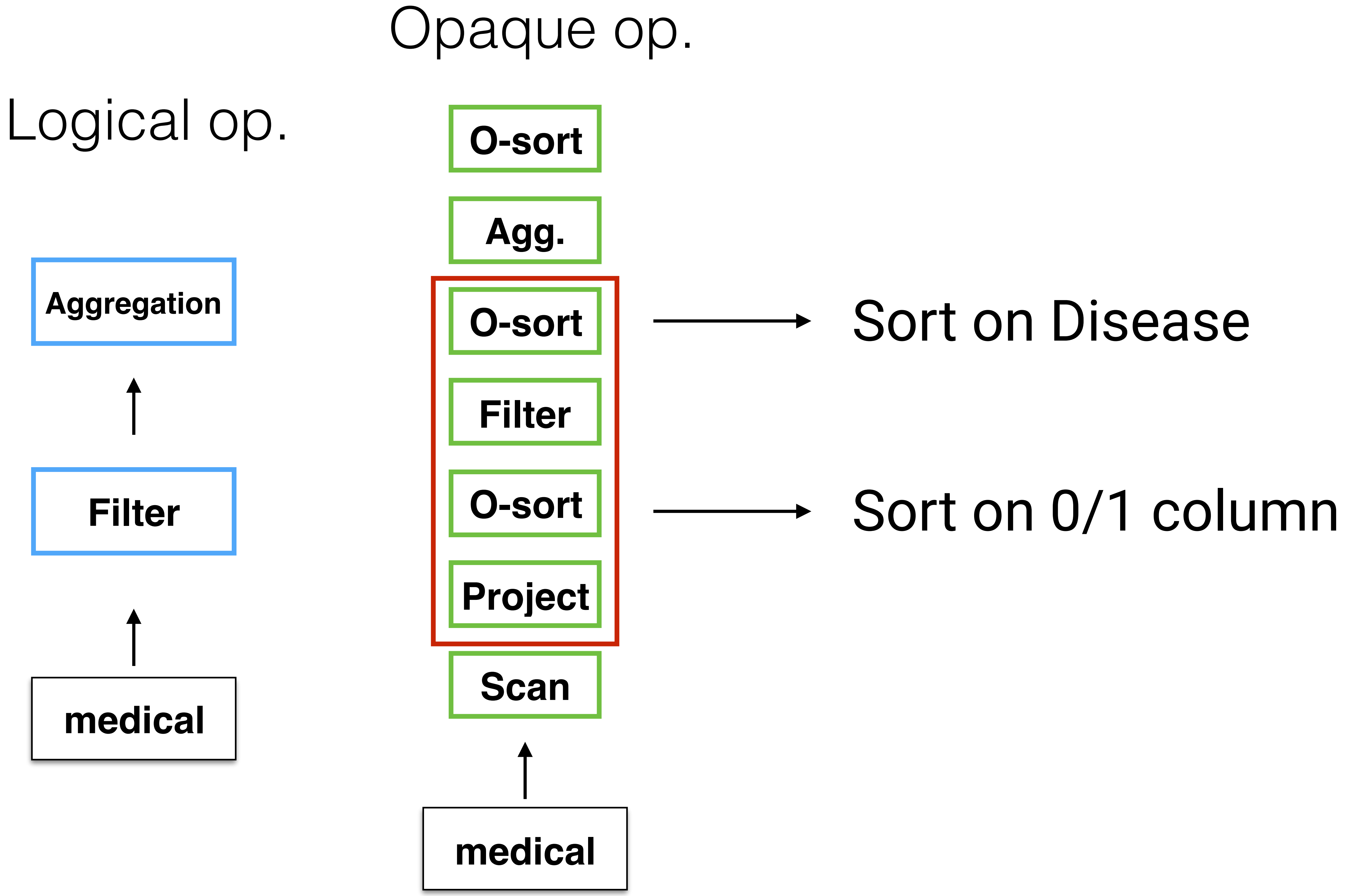
Rule-based optimization



Rule-based optimization



Rule-based optimization



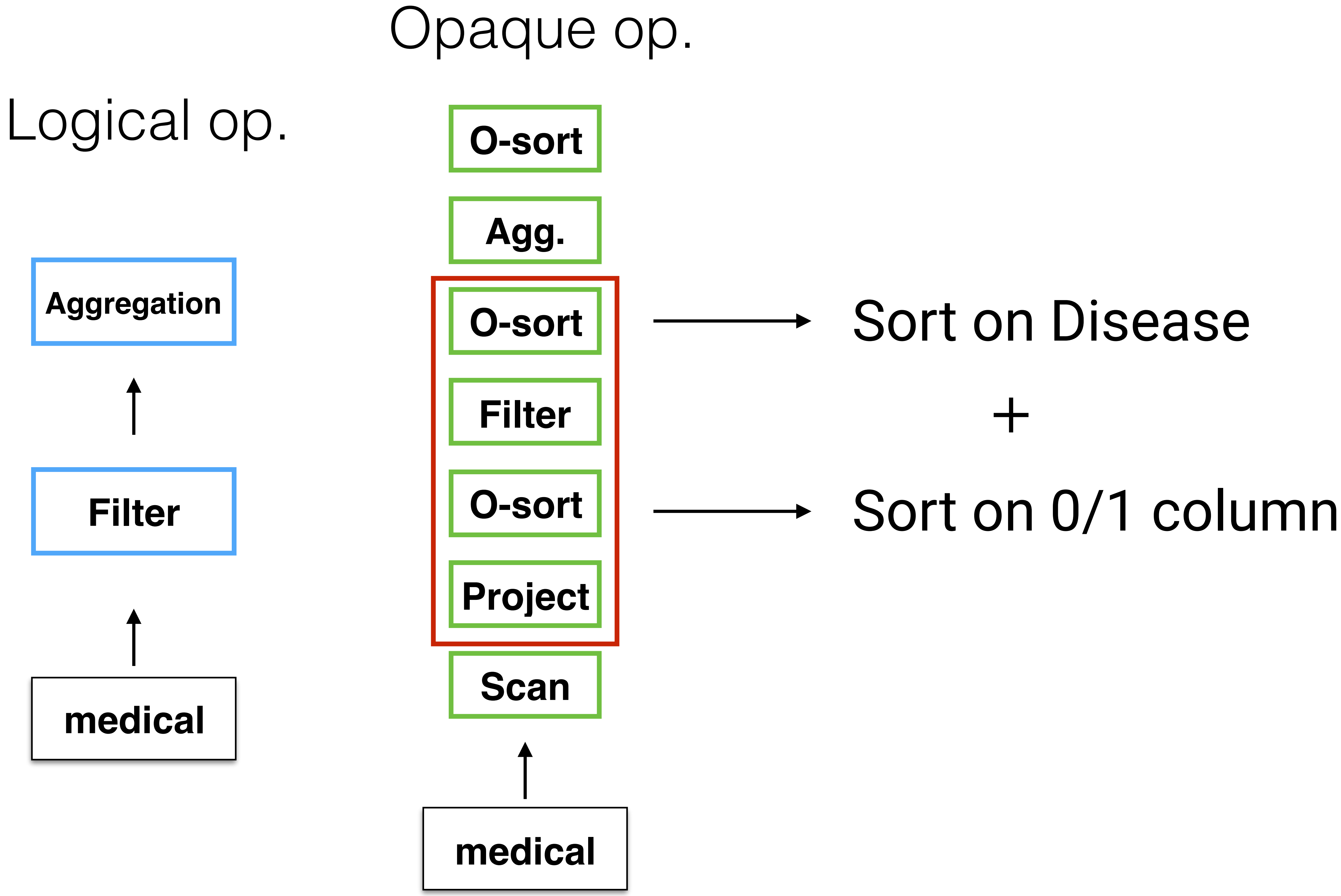
Opaque op.

Logical op.

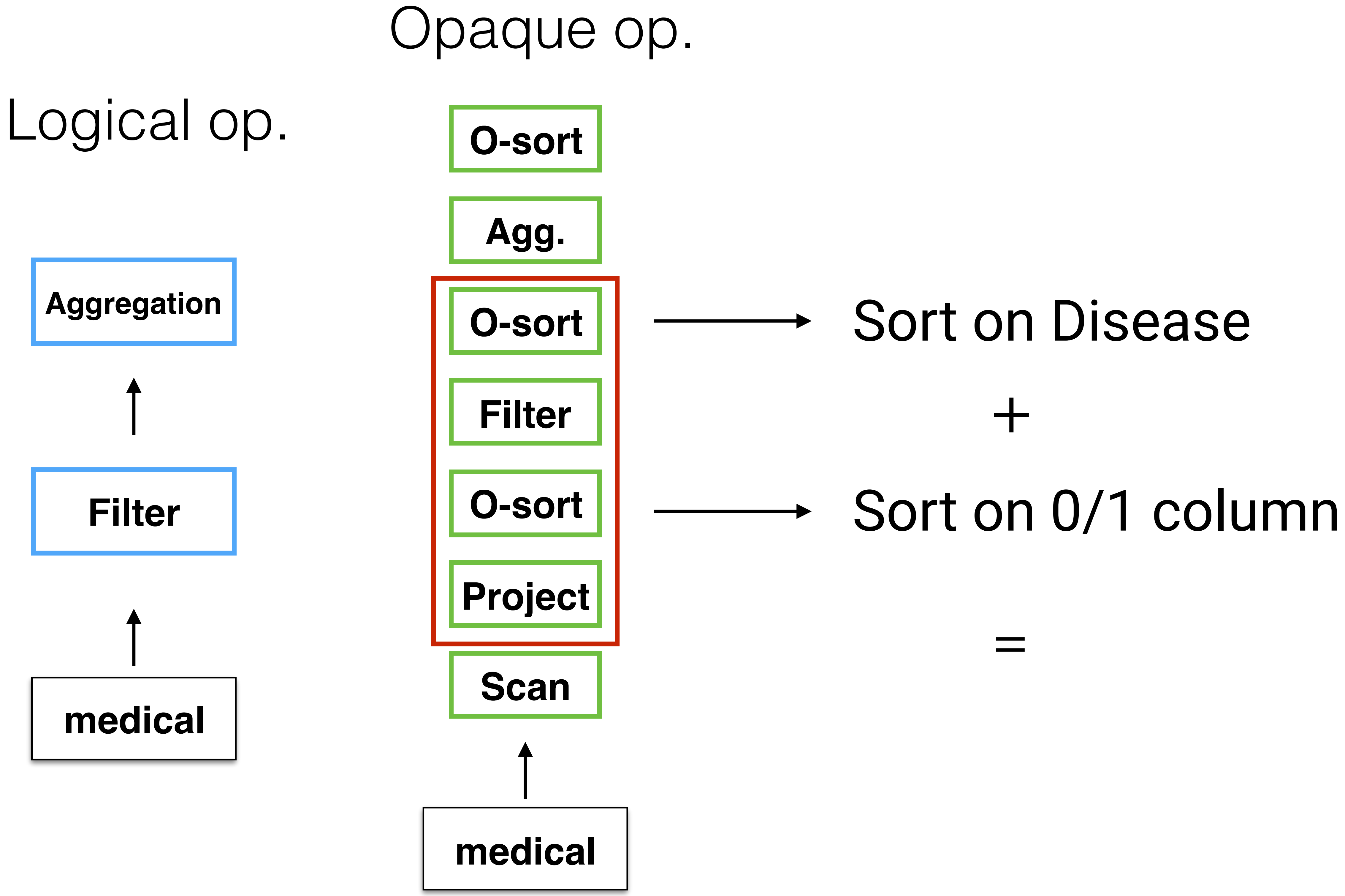
Sort on Disease

Sort on 0/1 column

Rule-based optimization



Rule-based optimization



Opaque op.

Logical op.

Aggregation

Filter

medical

O-sort

Agg.

O-sort

Filter

O-sort

Project

Scan

medical

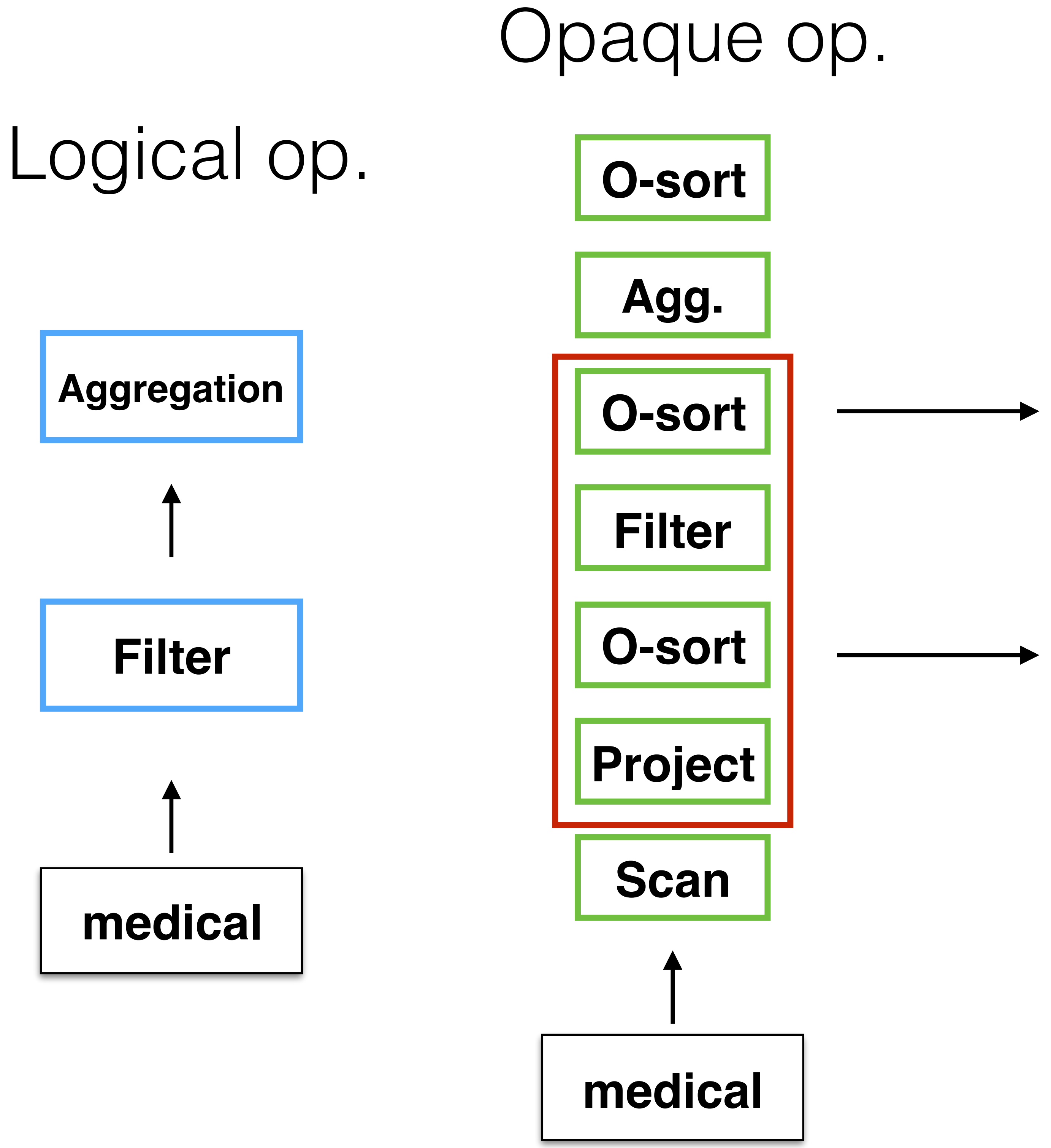
Sort on Disease

+

Sort on 0/1 column

=

Rule-based optimization



Sort on Disease

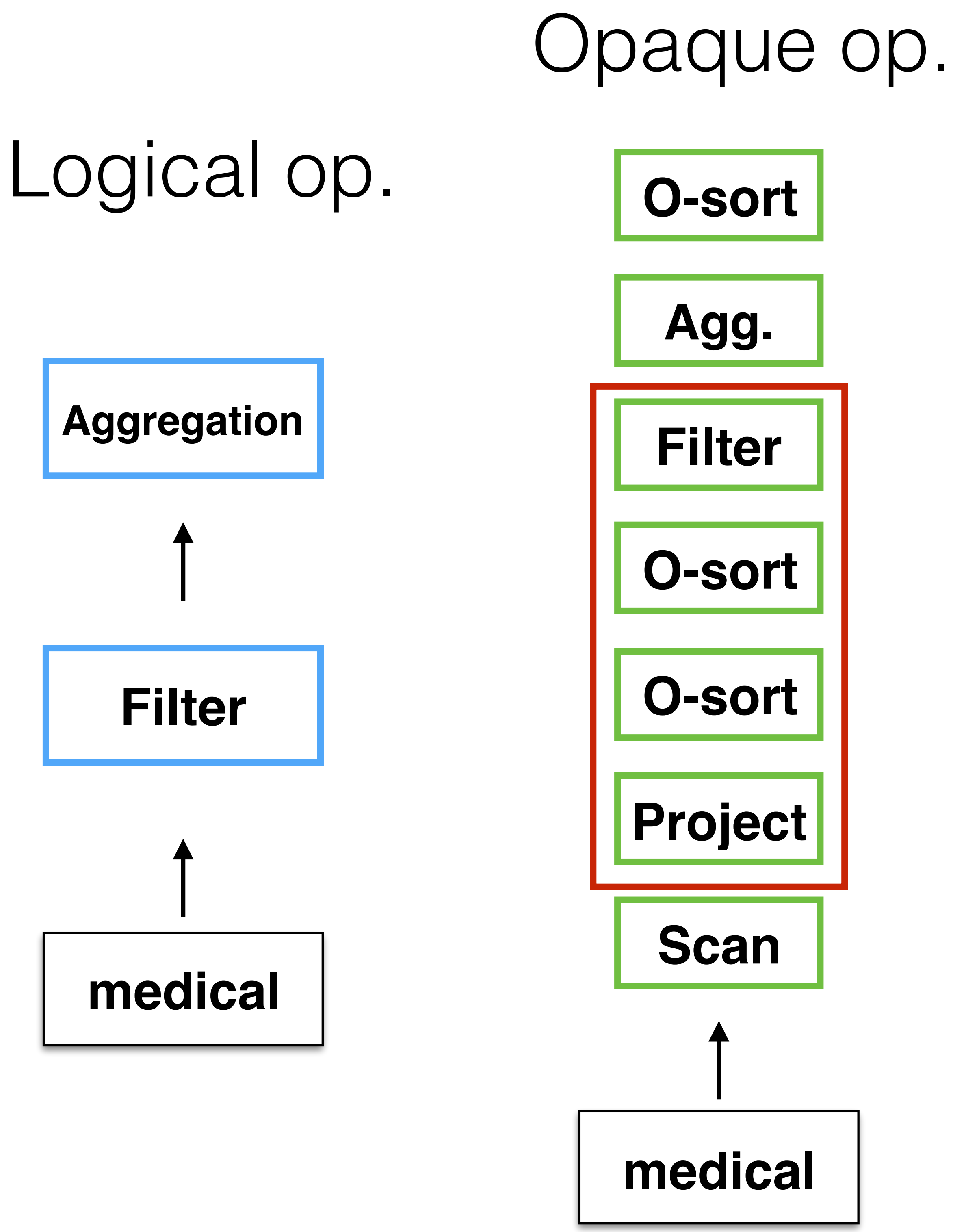
+

Sort on 0/1 column

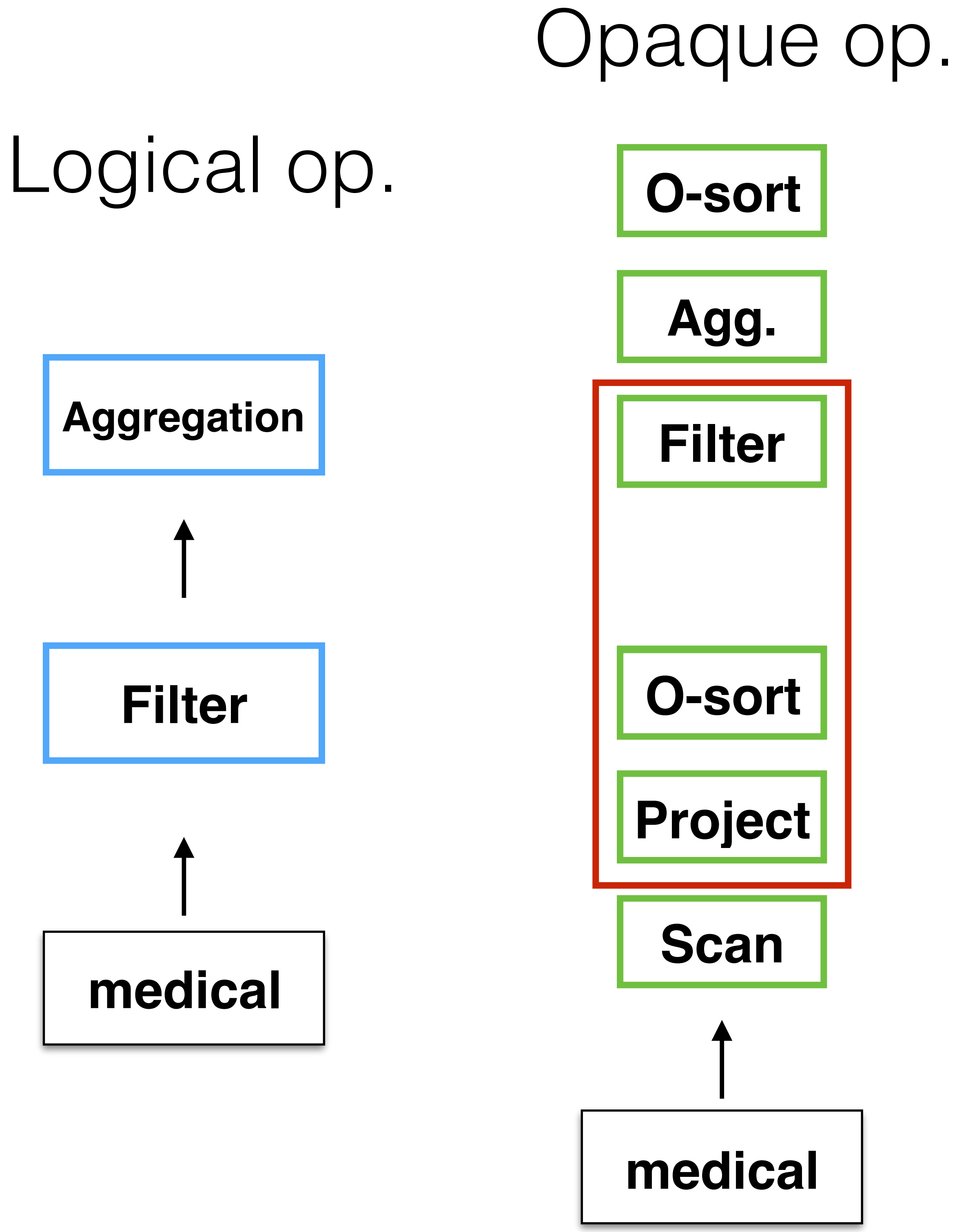
=

Sort on (0/1, Disease)

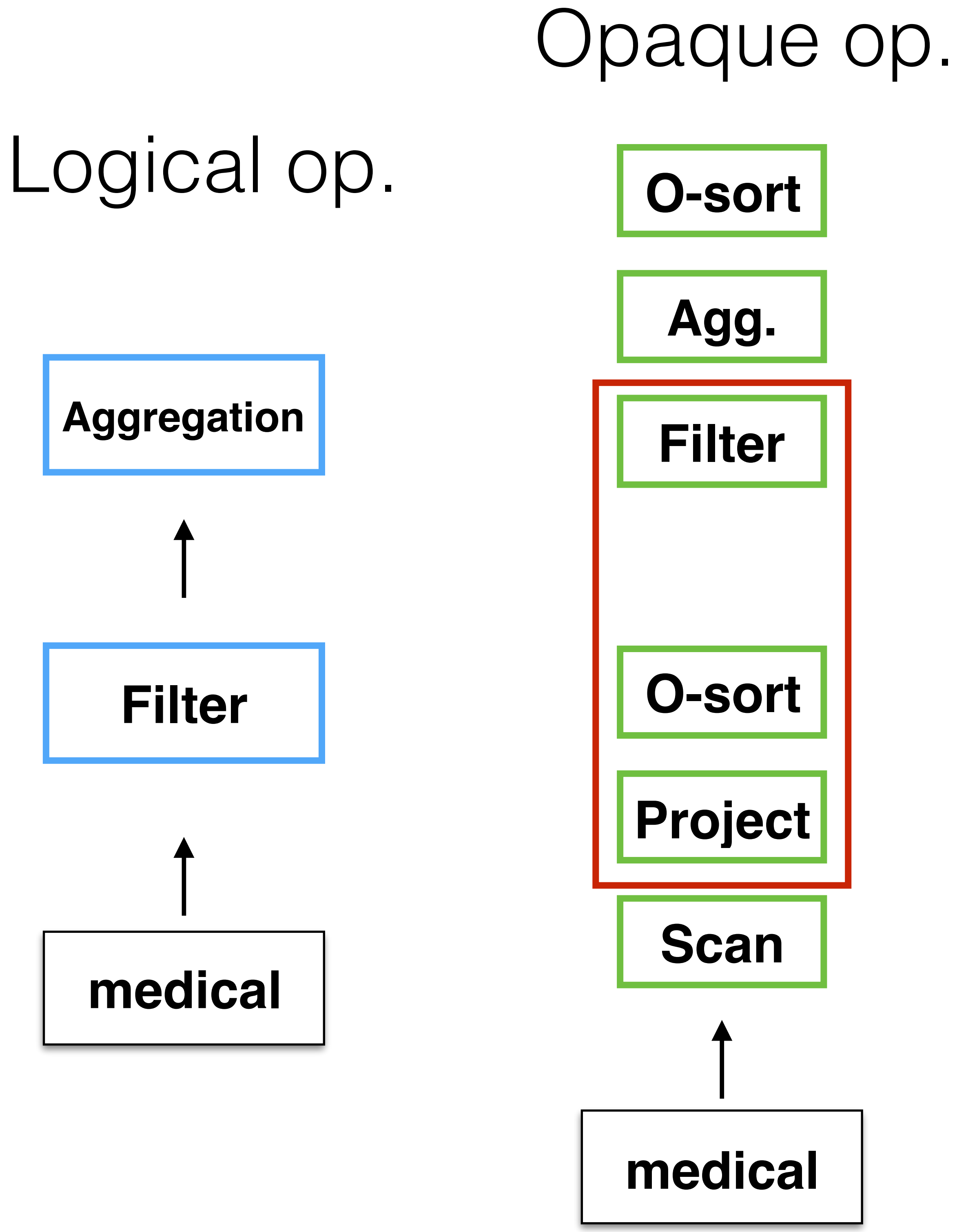
Rule-based optimization



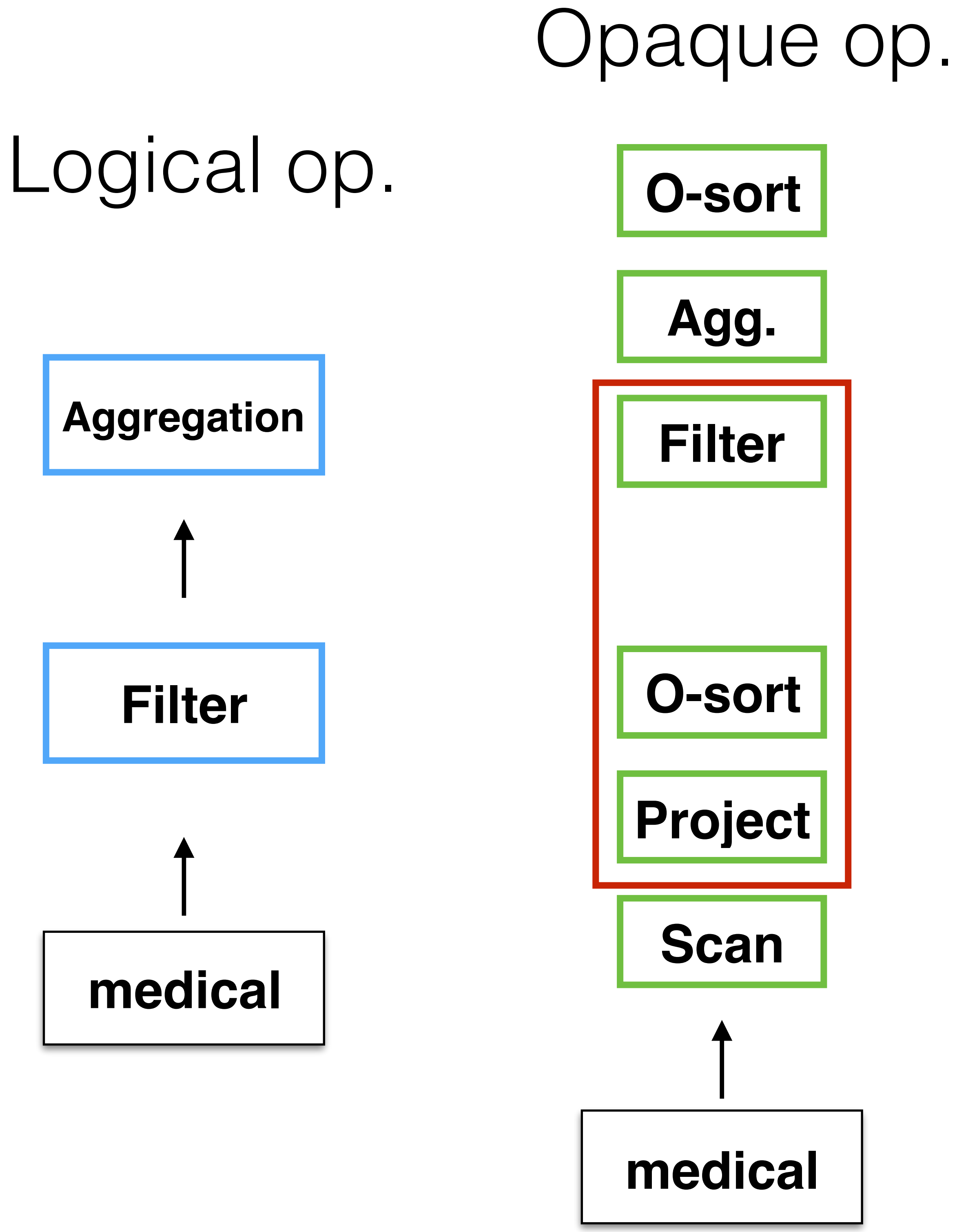
Rule-based optimization



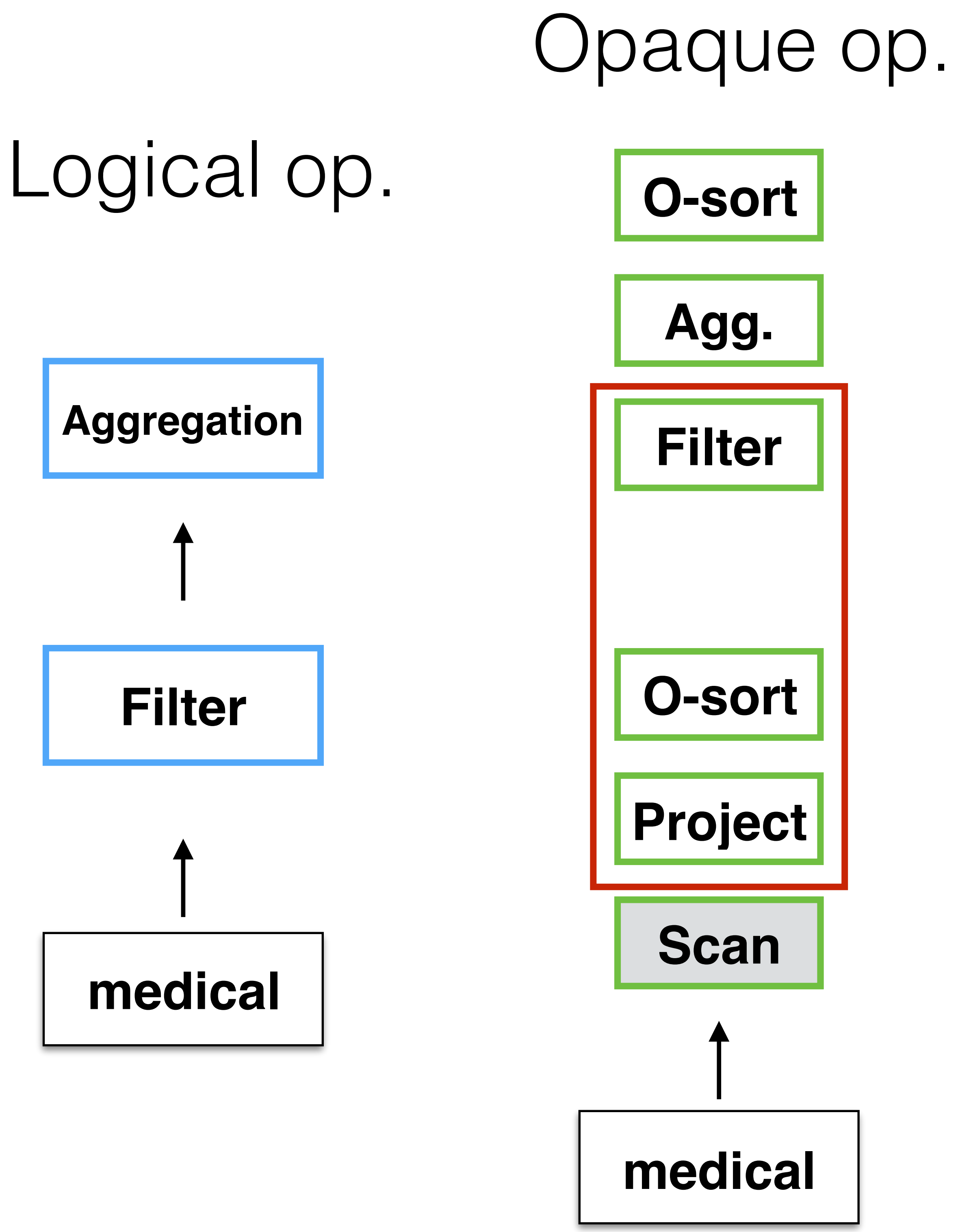
Rule-based optimization



Rule-based optimization



Rule-based optimization



Opaque op.

Logical op.

Aggregation

Filter

medical

O-sort

Agg.

Filter

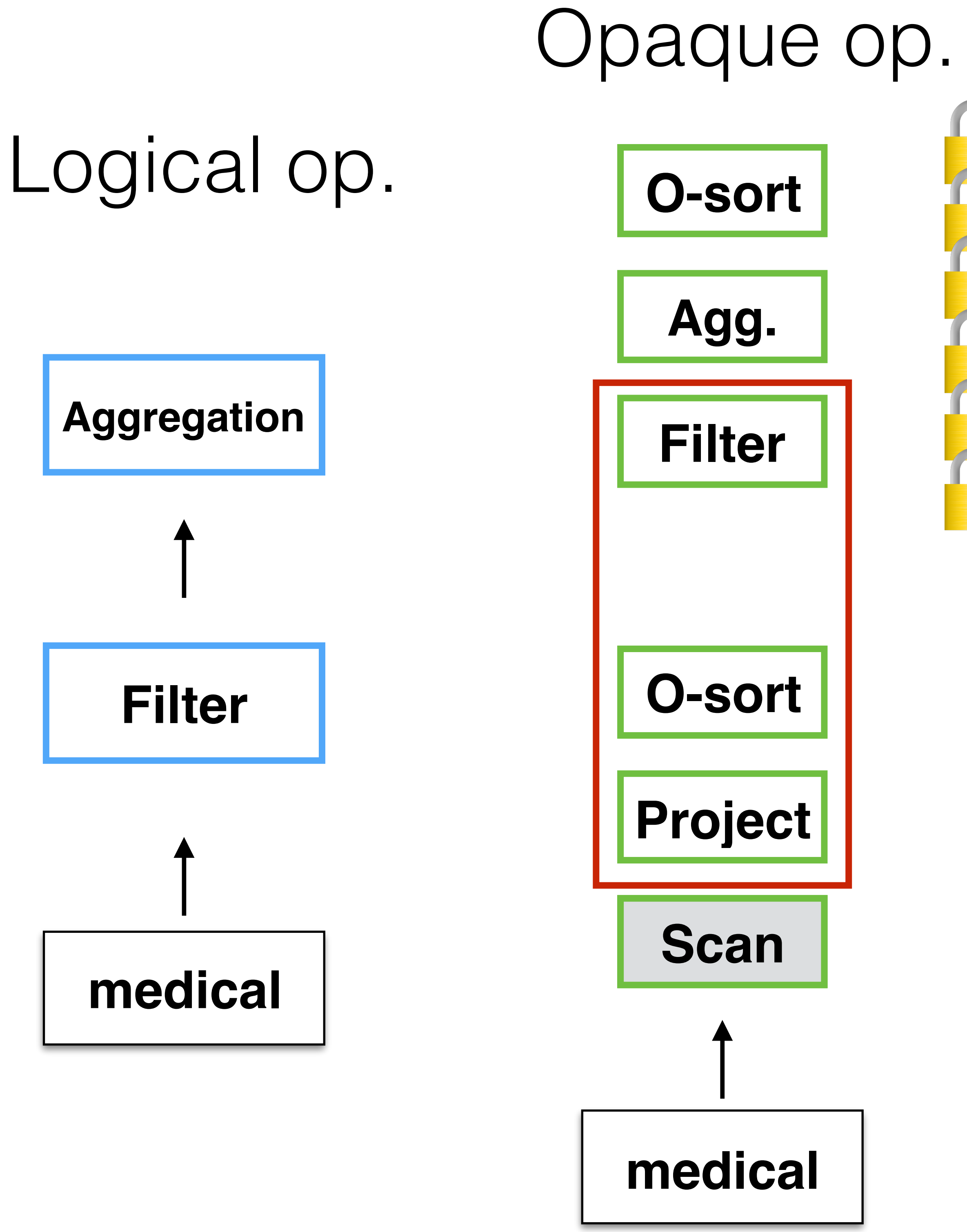
O-sort

Project

Scan

medical

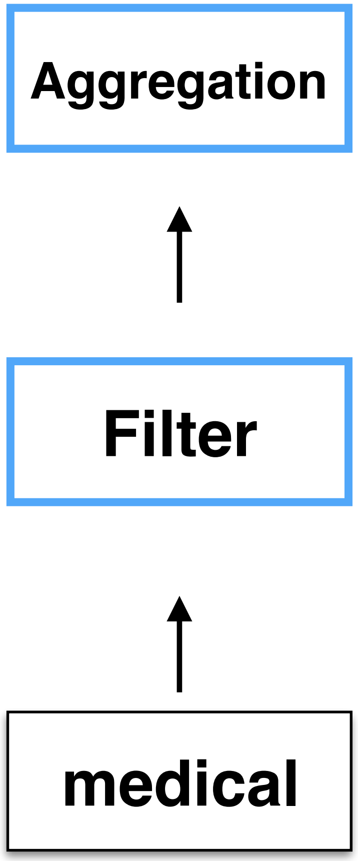
Rule-based optimization



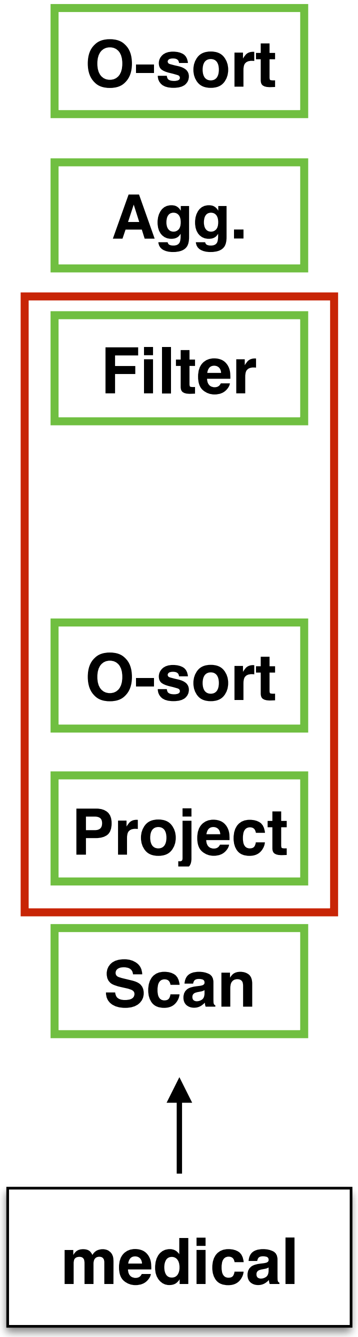
🔒 12809	Amanda D. Edwards	40	Diabetes
🔒 29489	Robert R. McGowan	56	Diabetes
🔒 13744	Kimberly R. Seay	51	Cancer
🔒 18740	Dennis G. Bates	32	Diabetes
🔒 32591	Donna R. Bridges	26	Diabetes
98329	Ronald S. Ogden	53	Cancer

Rule-based optimization

Logical op.



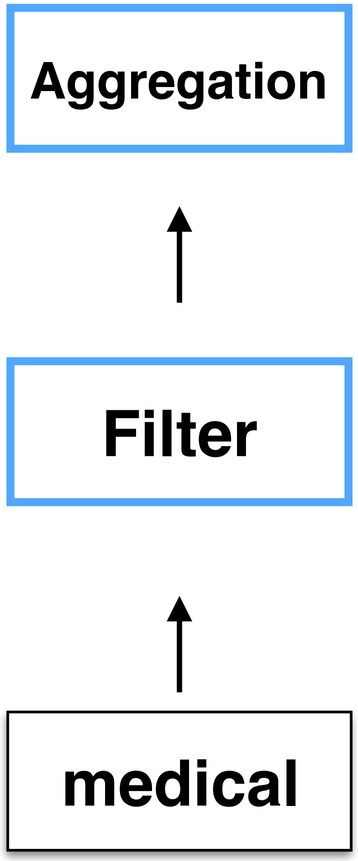
Opaque op.



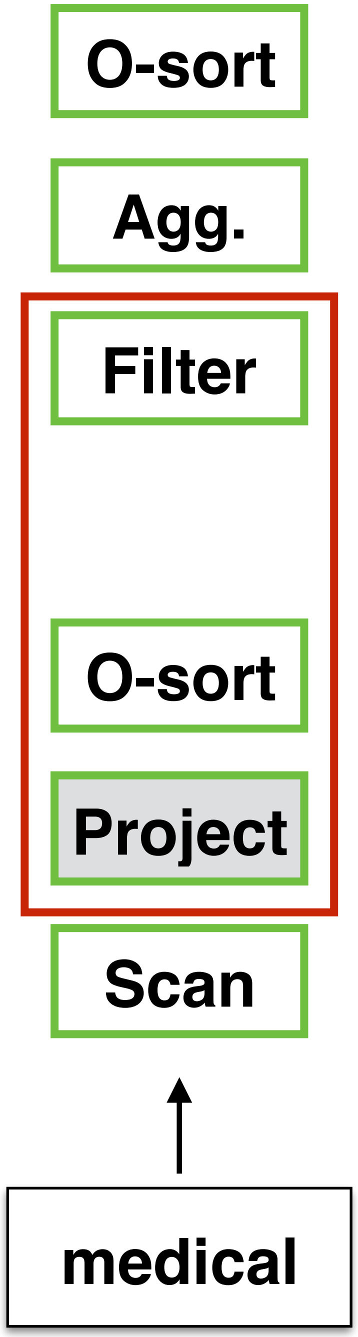
🔒 12809	Amanda D. Edwards	40	Diabetes
🔒 29489	Robert R. McGowan	56	Diabetes
🔒 13744	Kimberly R. Seay	51	Cancer
🔒 18740	Dennis G. Bates	32	Diabetes
🔒 32591	Donna R. Bridges	26	Diabetes
98329	Ronald S. Ogden	53	Cancer

Rule-based optimization

Logical op.



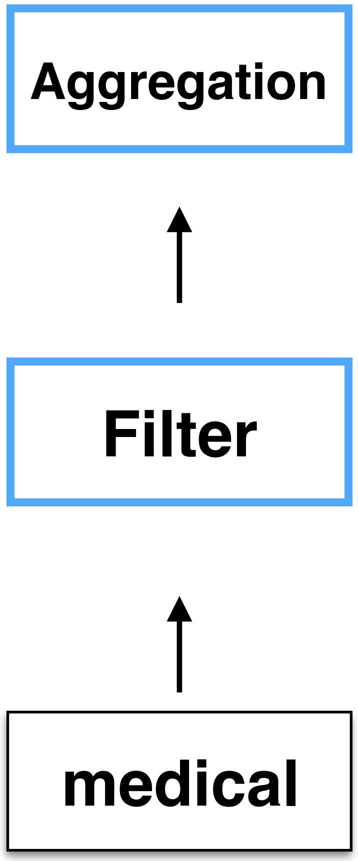
Opaque op.



🔒 12809	Amanda D. Edwards	40	Diabetes
🔒 29489	Robert R. McGowan	56	Diabetes
🔒 13744	Kimberly R. Seay	51	Cancer
🔒 18740	Dennis G. Bates	32	Diabetes
🔒 32591	Donna R. Bridges	26	Diabetes
98329	Ronald S. Ogden	53	Cancer

Rule-based optimization

Logical op.

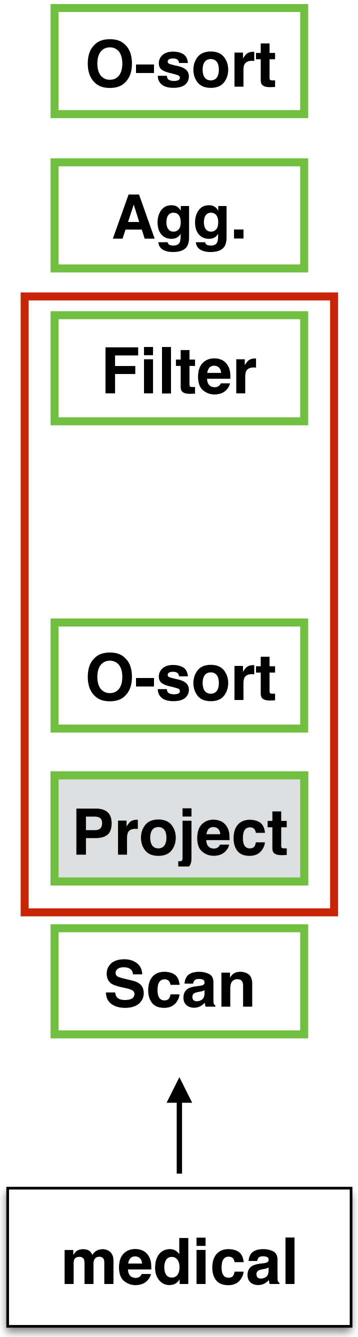


Aggregation

Filter

medical

Opaque op.



O-sort

Agg.

Filter

O-sort

Project

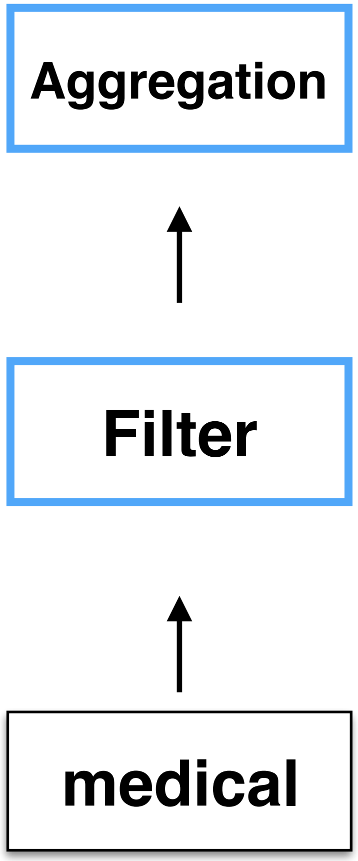
Scan

medical

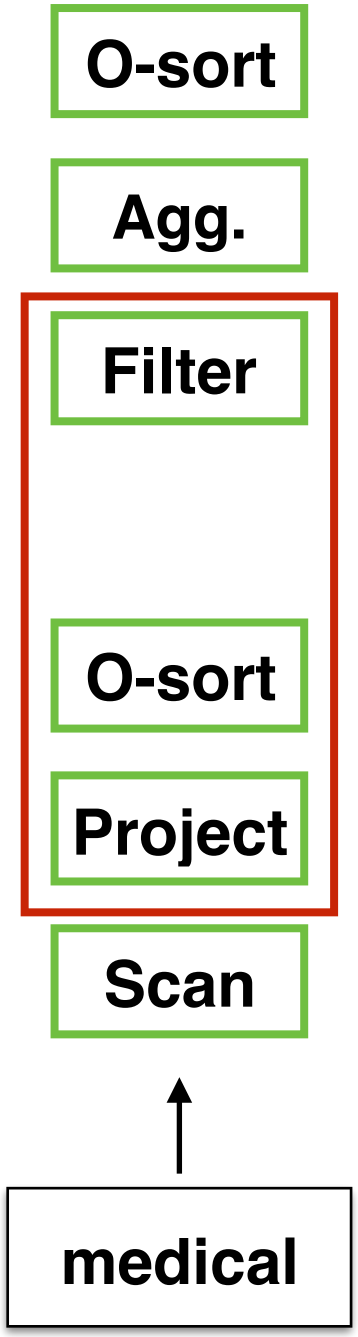
🔒 12809	Amanda D. Edwards	40	Diabetes	0
🔒 29489	Robert R. McGowan	56	Diabetes	0
🔒 13744	Kimberly R. Seay	51	Cancer	0
🔒 18740	Dennis G. Bates	32	Diabetes	0
🔒 32591	Donna R. Bridges	26	Diabetes	1
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization

Logical op.



Opaque op.

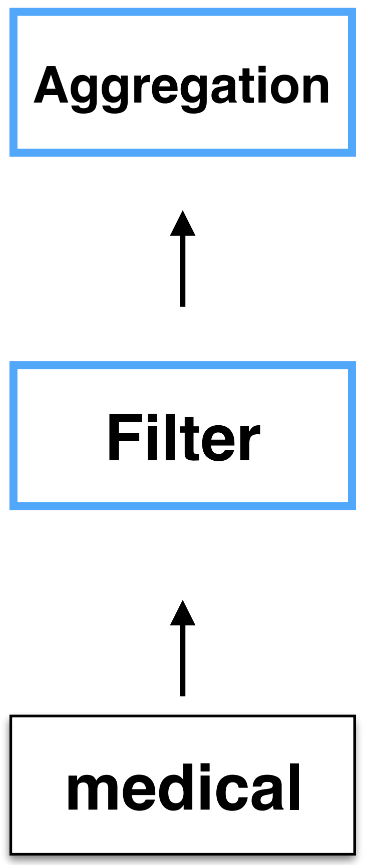


🔒 12809	Amanda D. Edwards	40	Diabetes	0
🔒 29489	Robert R. McGowan	56	Diabetes	0
🔒 13744	Kimberly R. Seay	51	Cancer	0
🔒 18740	Dennis G. Bates	32	Diabetes	0
🔒 32591	Donna R. Bridges	26	Diabetes	1
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization

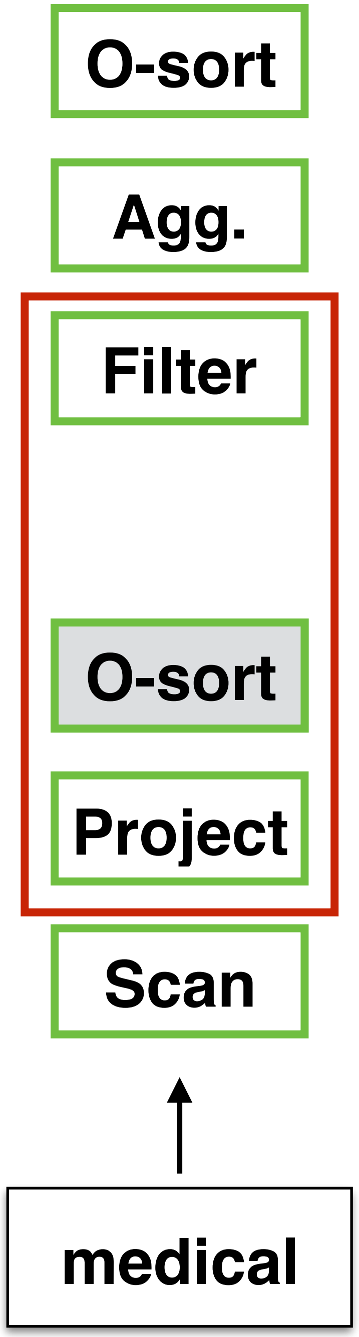
Opaque op.

Logical op.


Aggregation

Filter

medical


O-sort

Agg.

Filter

O-sort

Project

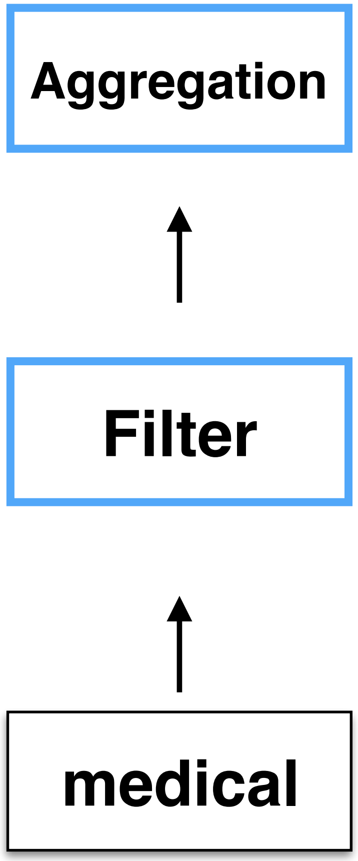
Scan

medical

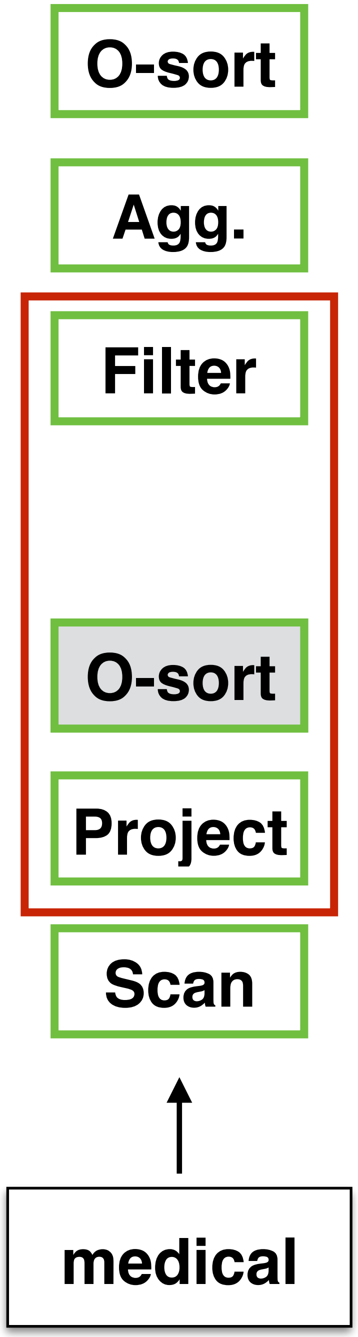
12809	Amanda D. Edwards	40	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
13744	Kimberly R. Seay	51	Cancer	0
18740	Dennis G. Bates	32	Diabetes	0
32591	Donna R. Bridges	26	Diabetes	1
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization

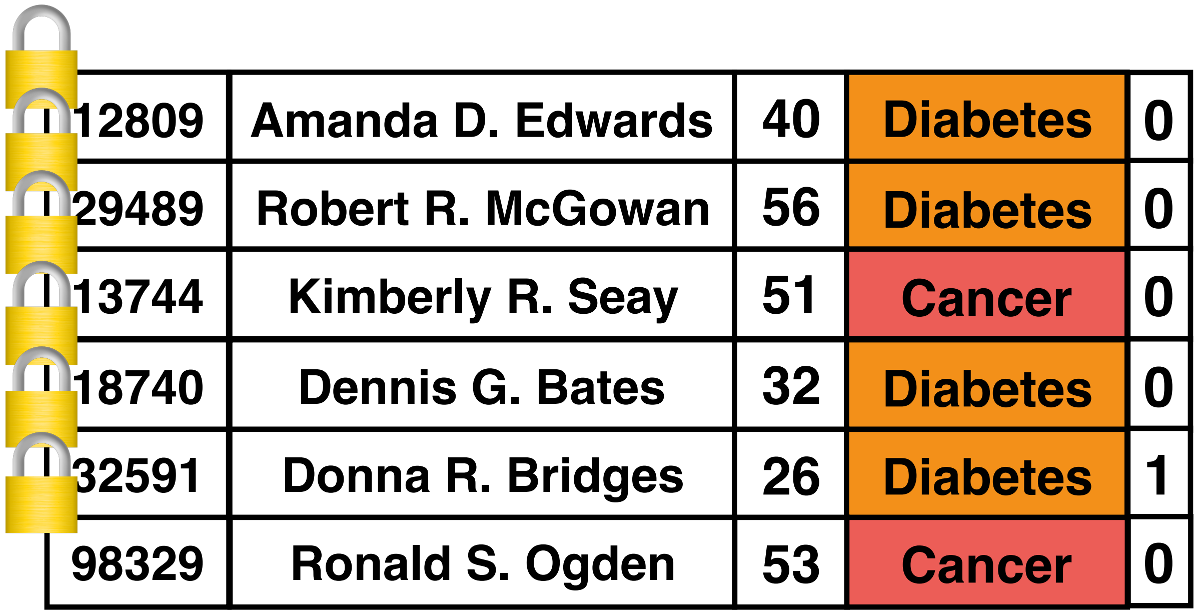
Logical op.



Opaque op.

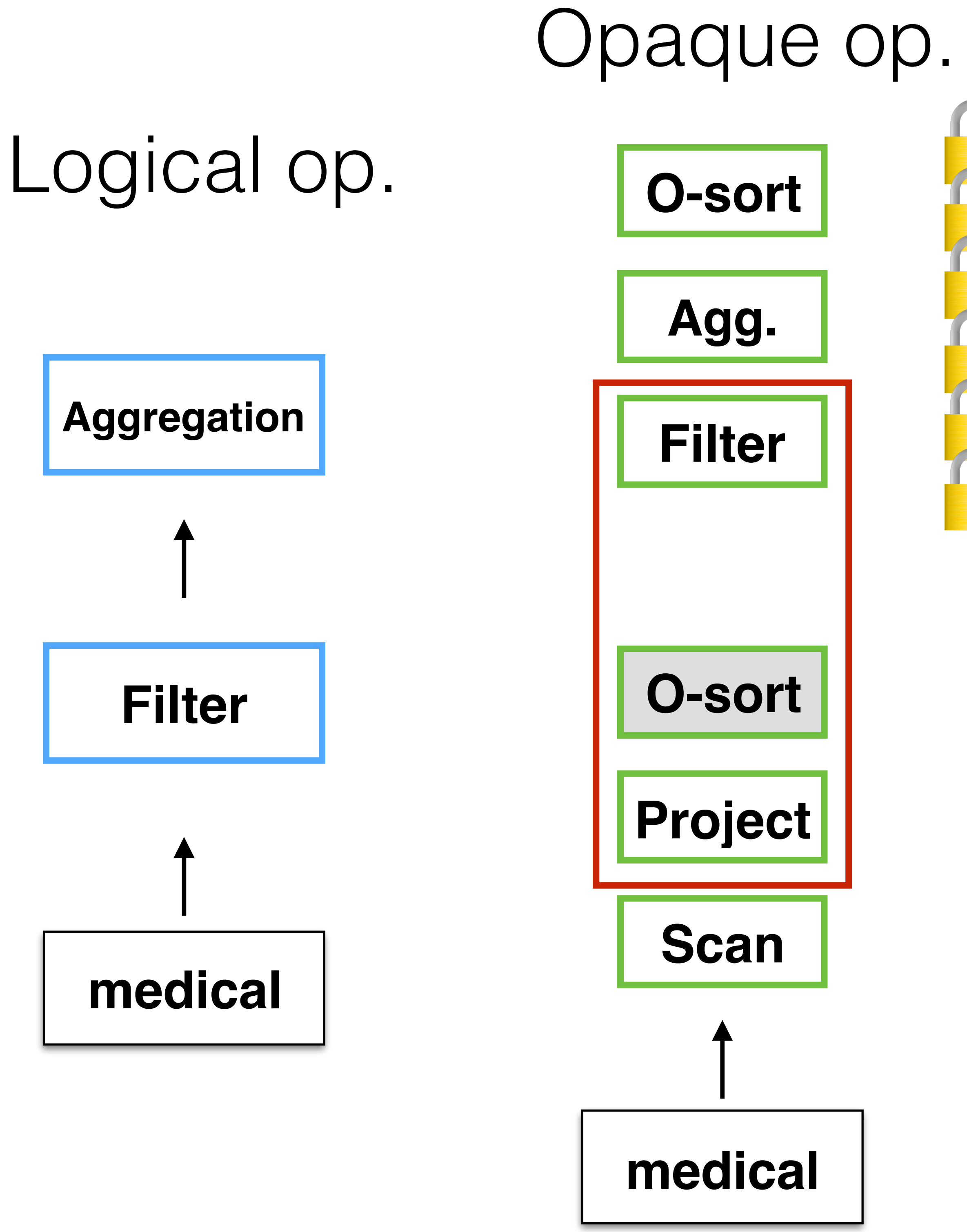


multi-column sort



12809	Amanda D. Edwards	40	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
13744	Kimberly R. Seay	51	Cancer	0
18740	Dennis G. Bates	32	Diabetes	0
32591	Donna R. Bridges	26	Diabetes	1
98329	Ronald S. Ogden	53	Cancer	0

Rule-based optimization



Logical op.

Aggregation

Filter

medical

Opaque op.

O-sort

Agg.

Filter

O-sort

Project

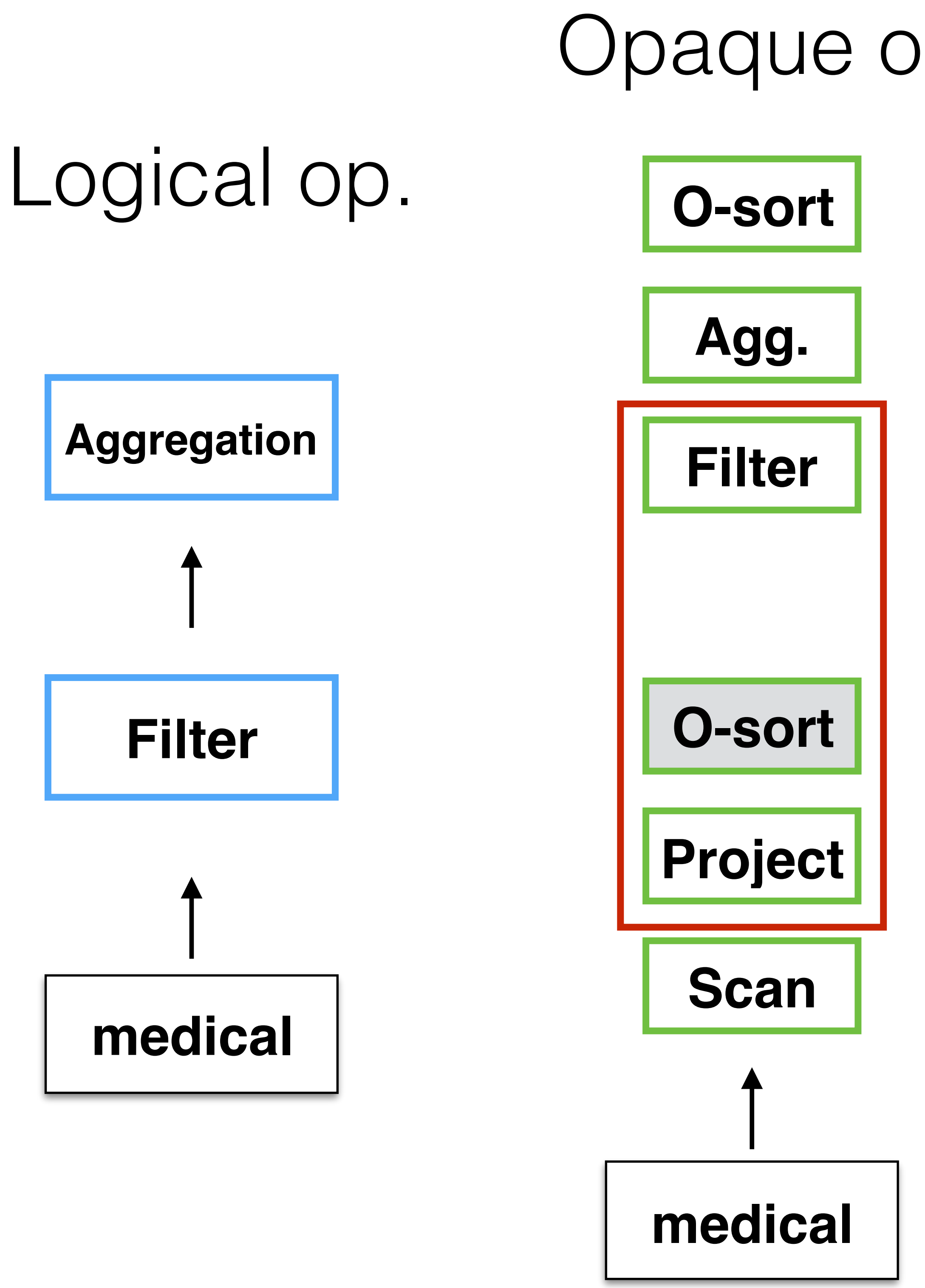
Scan

medical

multi-column sort

12809	Amanda D. Edwards	40	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
13744	Kimberly R. Seay	51	Cancer	0
18740	Dennis G. Bates	32	Diabetes	0
32591	Donna R. Bridges	26	Diabetes	1
98329	Ronald S. Ogden	53	Cancer	0

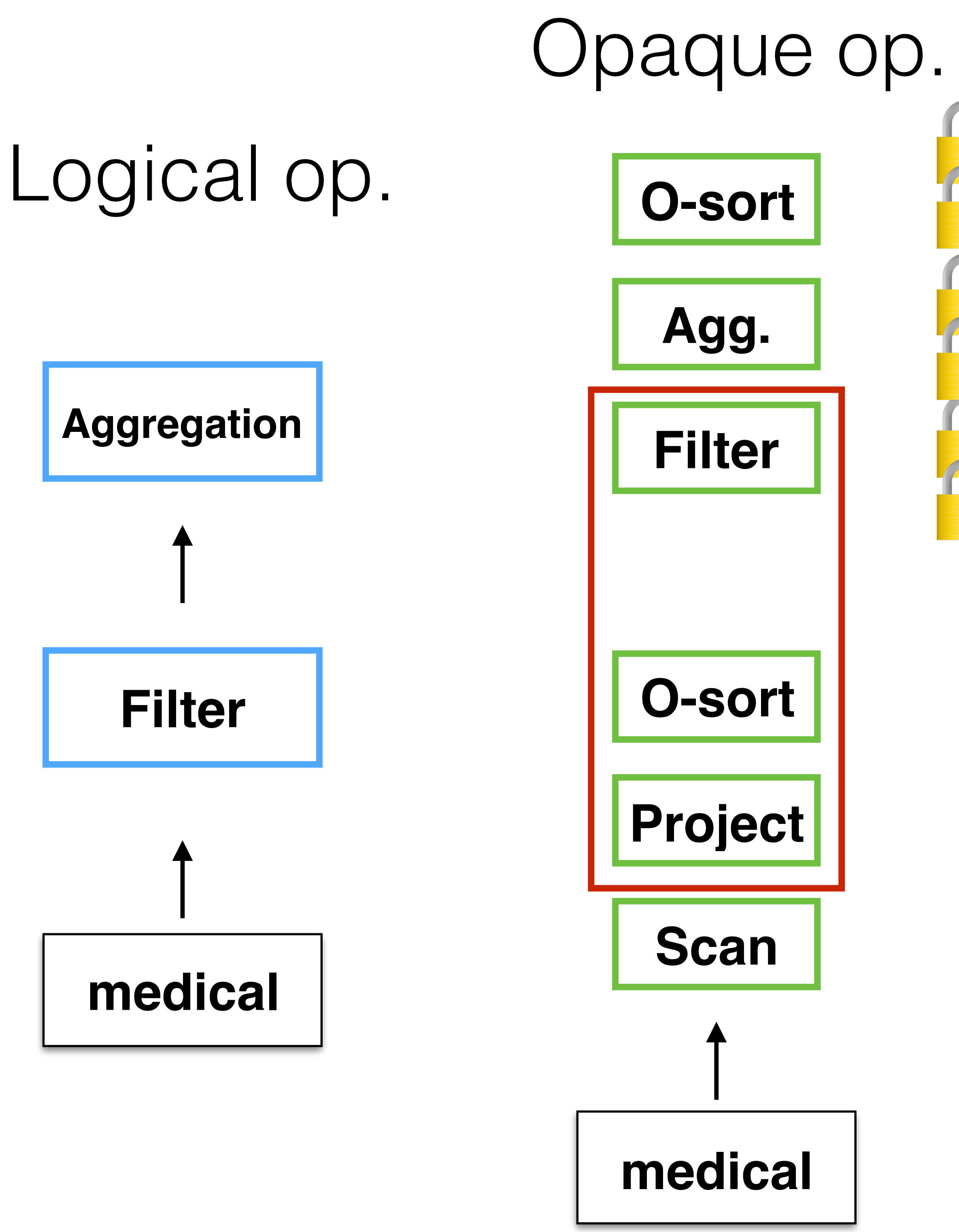
Rule-based optimization



multi-column sort

13744	Kimberly R. Seay	51	Cancer	0
98329	Ronald S. Ogden	53	Cancer	0
12809	Amanda D. Edwards	40	Diabetes	0
18740	Dennis G. Bates	32	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
32591	Donna R. Bridges	26	Diabetes	1

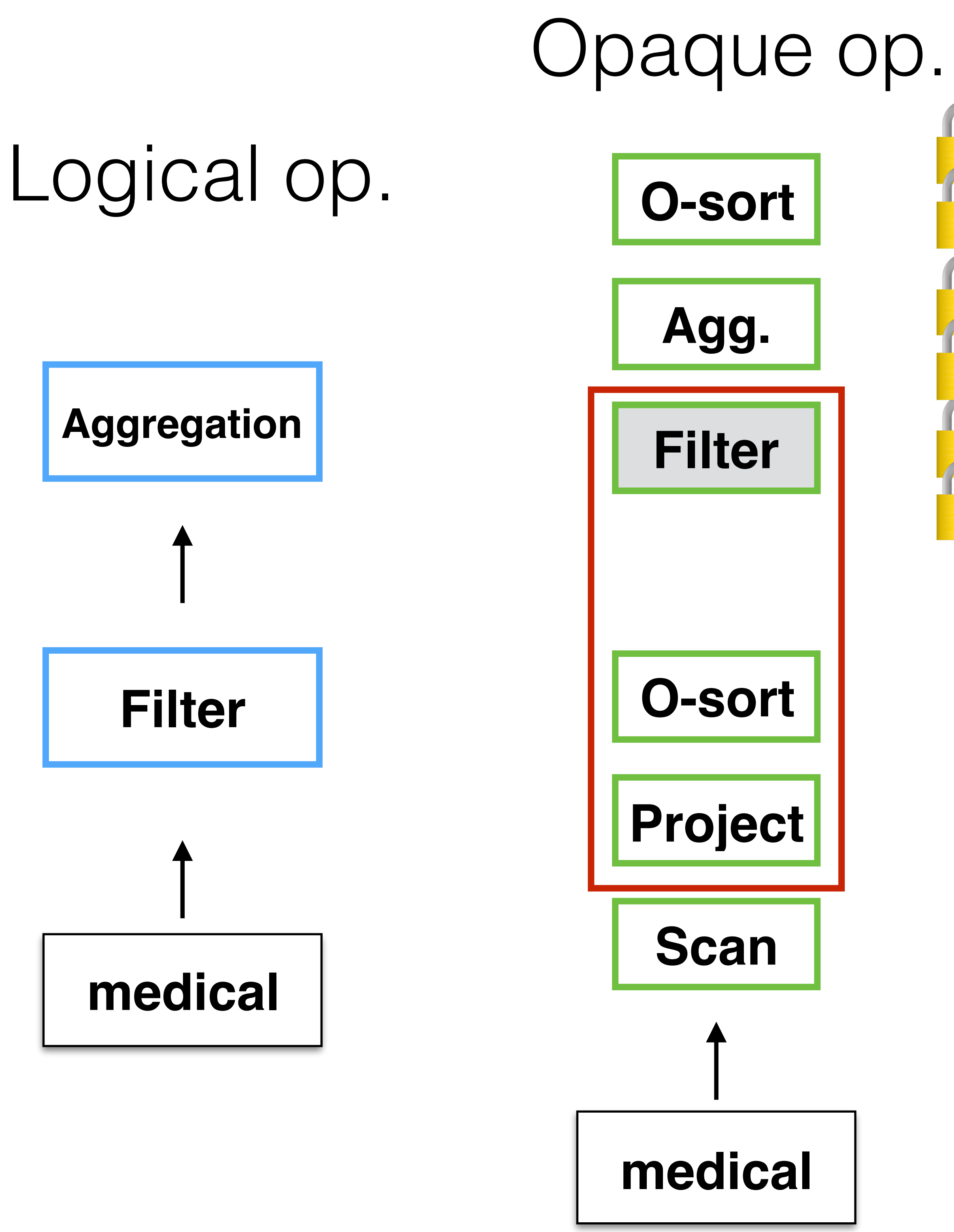
Rule-based optimization



multi-column sort

13744	Kimberly R. Seay	51	Cancer	0
98329	Ronald S. Ogden	53	Cancer	0
12809	Amanda D. Edwards	40	Diabetes	0
18740	Dennis G. Bates	32	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
32591	Donna R. Bridges	26	Diabetes	1

Rule-based optimization

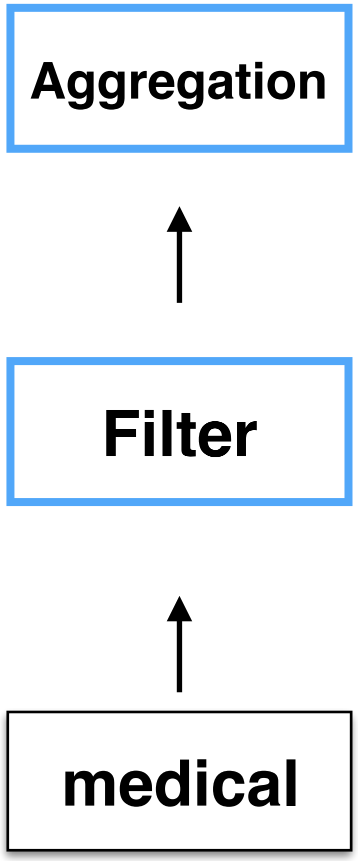


multi-column sort

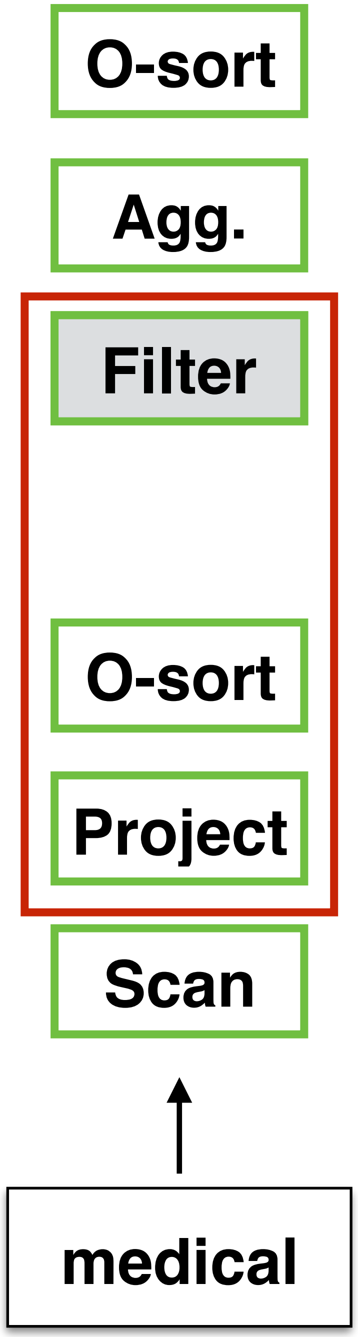
13744	Kimberly R. Seay	51	Cancer	0
98329	Ronald S. Ogden	53	Cancer	0
12809	Amanda D. Edwards	40	Diabetes	0
18740	Dennis G. Bates	32	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0
32591	Donna R. Bridges	26	Diabetes	1

Rule-based optimization

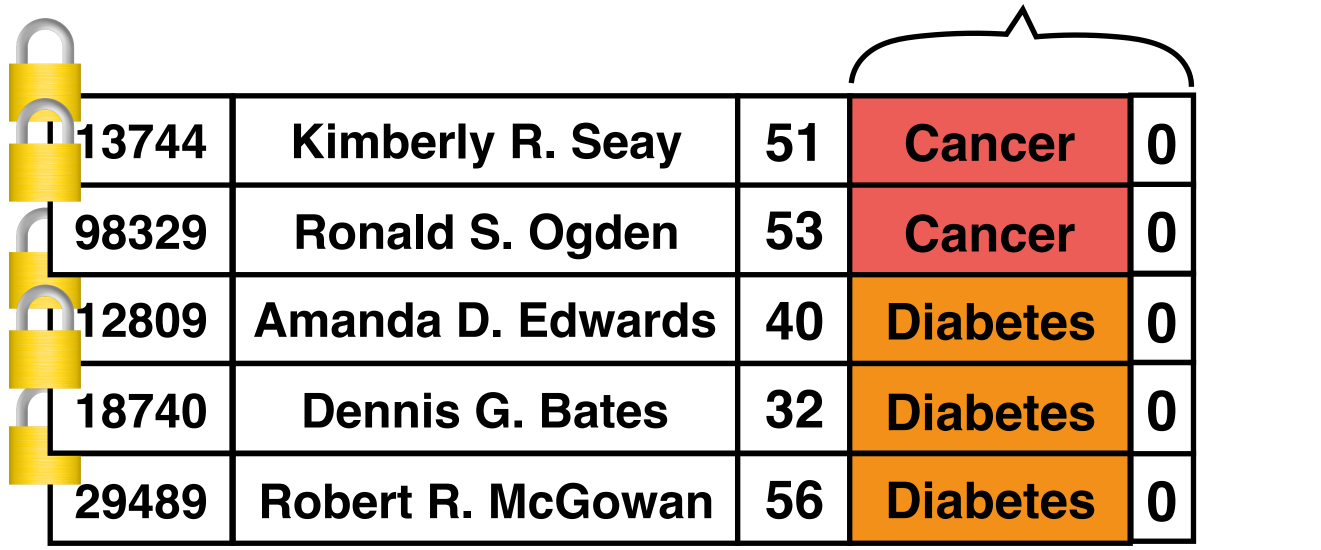
Logical op.



Opaque op.

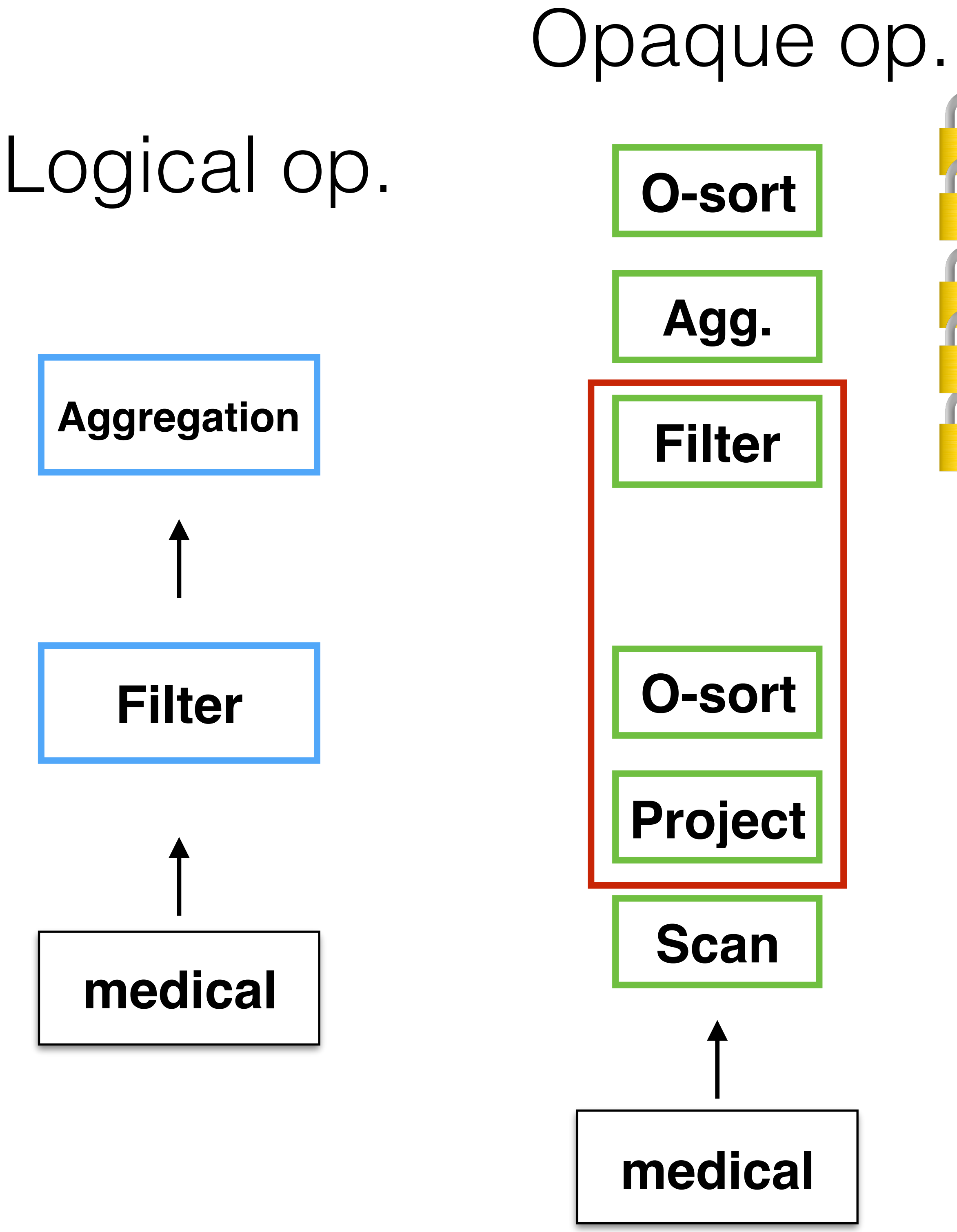


multi-column sort



13744	Kimberly R. Seay	51	Cancer	0
98329	Ronald S. Ogden	53	Cancer	0
12809	Amanda D. Edwards	40	Diabetes	0
18740	Dennis G. Bates	32	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0

Rule-based optimization

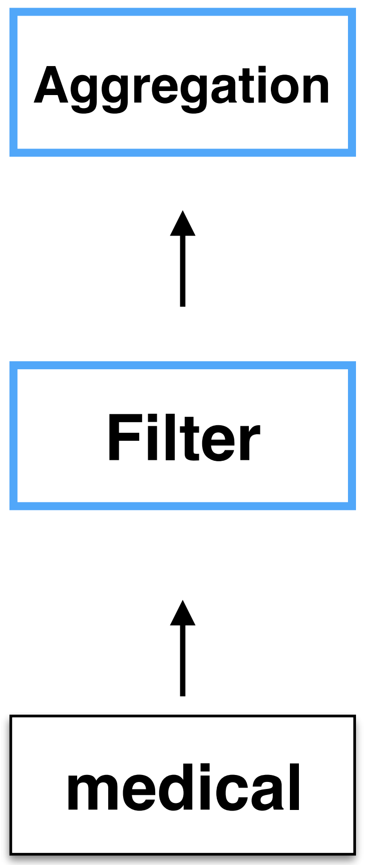


multi-column sort

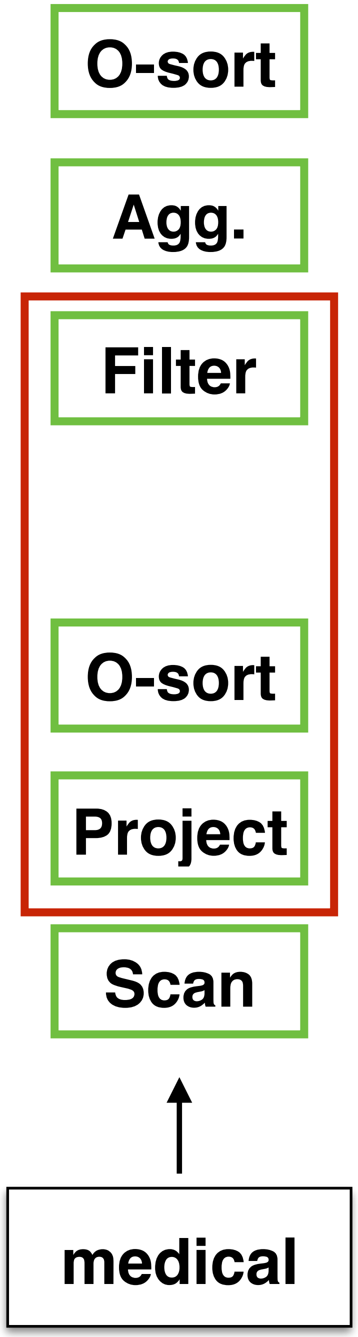
13744	Kimberly R. Seay	51	Cancer	0
98329	Ronald S. Ogden	53	Cancer	0
12809	Amanda D. Edwards	40	Diabetes	0
18740	Dennis G. Bates	32	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0

Rule-based optimization

Logical op.



Opaque op.



multi-column sort

13744	Kimberly R. Seay	51	Cancer	0
98329	Ronald S. Ogden	53	Cancer	0
12809	Amanda D. Edwards	40	Diabetes	0
18740	Dennis G. Bates	32	Diabetes	0
29489	Robert R. McGowan	56	Diabetes	0

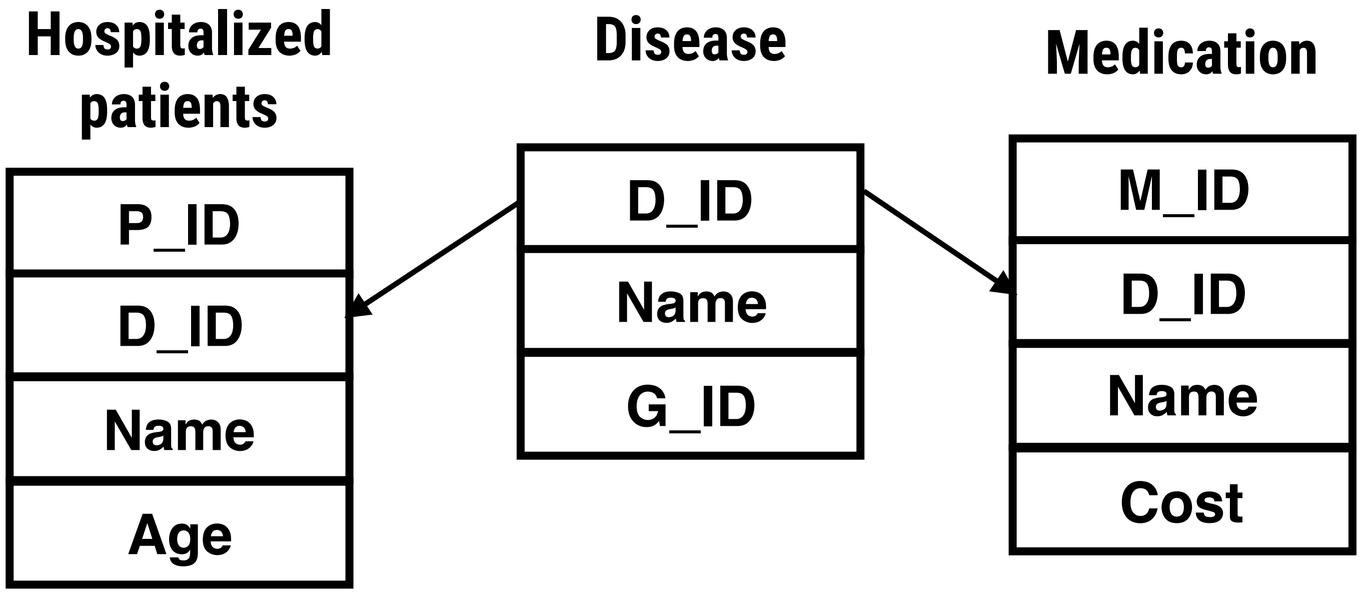
Eliminated one oblivious sort!

Two-part solution:

- Distributed oblivious SQL operators
  - Oblivious filter
  - Oblivious sort
  - Oblivious aggregation
  - Oblivious join
- Novel query planning techniques
  - Rule-based optimization
  - Cost model
  - **Cost-based optimization**

Observation: not all tables are sensitive

Observation: not all tables are sensitive



Hospitalized patients

P_ID

D_ID

Name

Age

Disease

D_ID

Name

G_ID

Medication

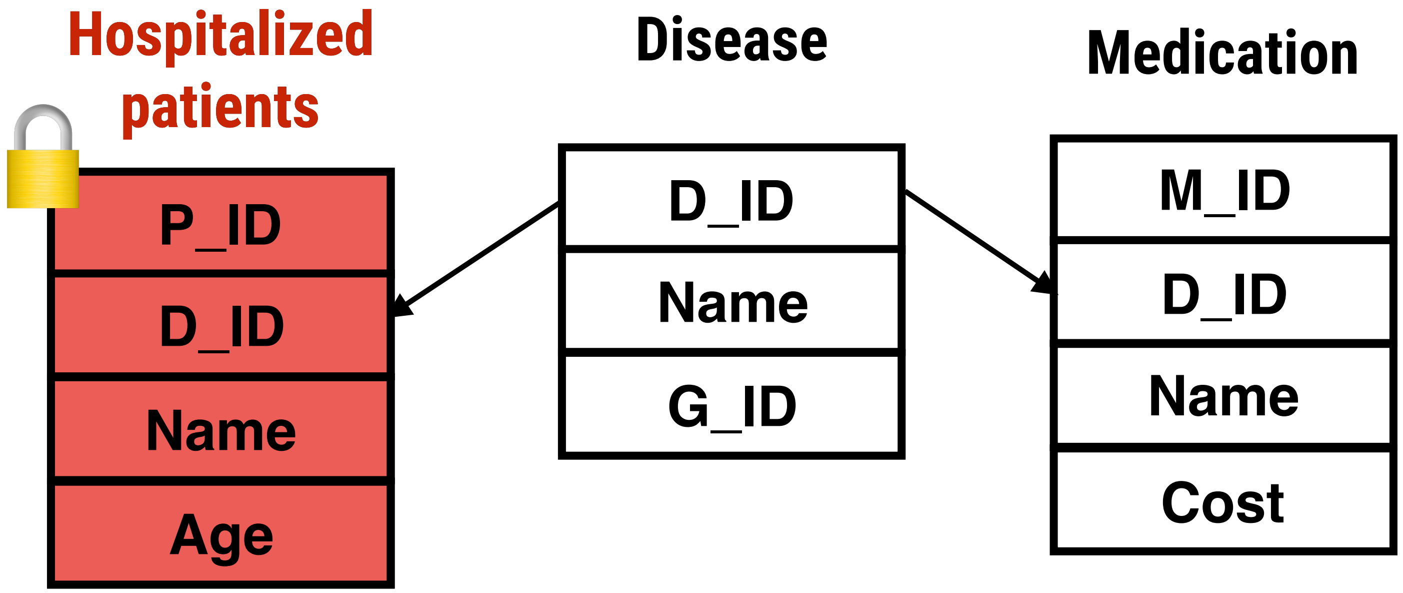
M_ID

D_ID

Name

Cost

Observation: not all tables are sensitive



Hospitalized patients

P_ID

D_ID

Name

Age

Disease

D_ID

Name

G_ID

Medication

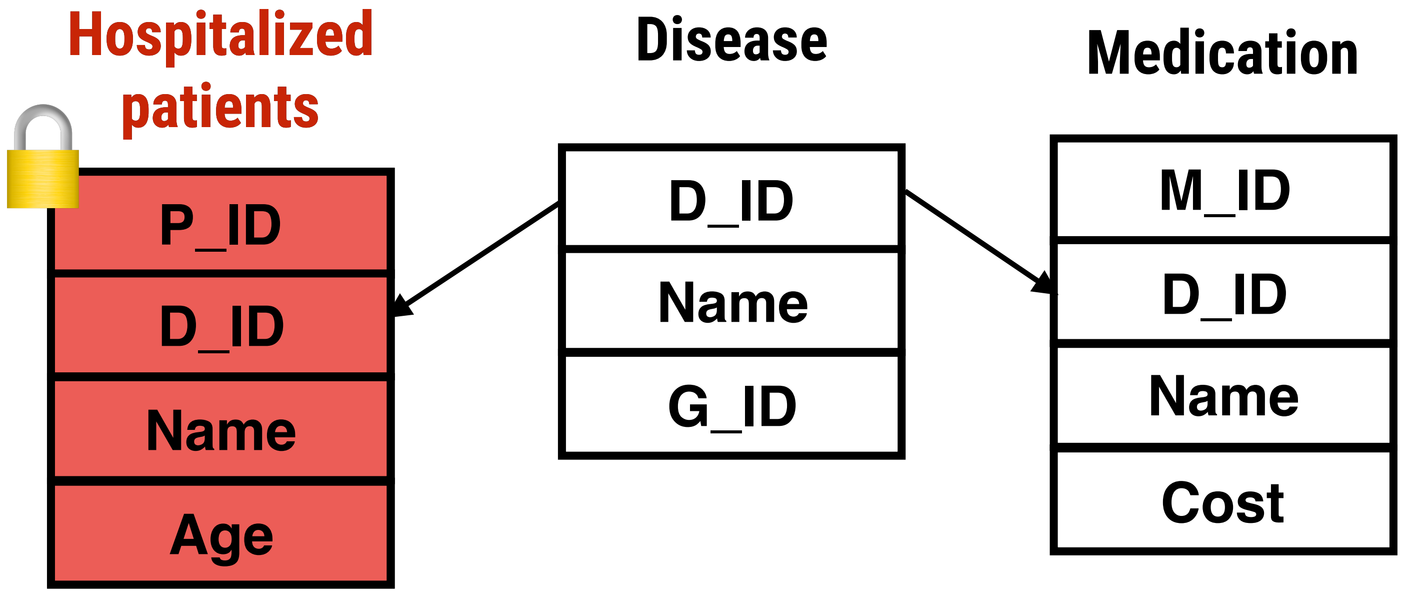
M_ID

D_ID

Name

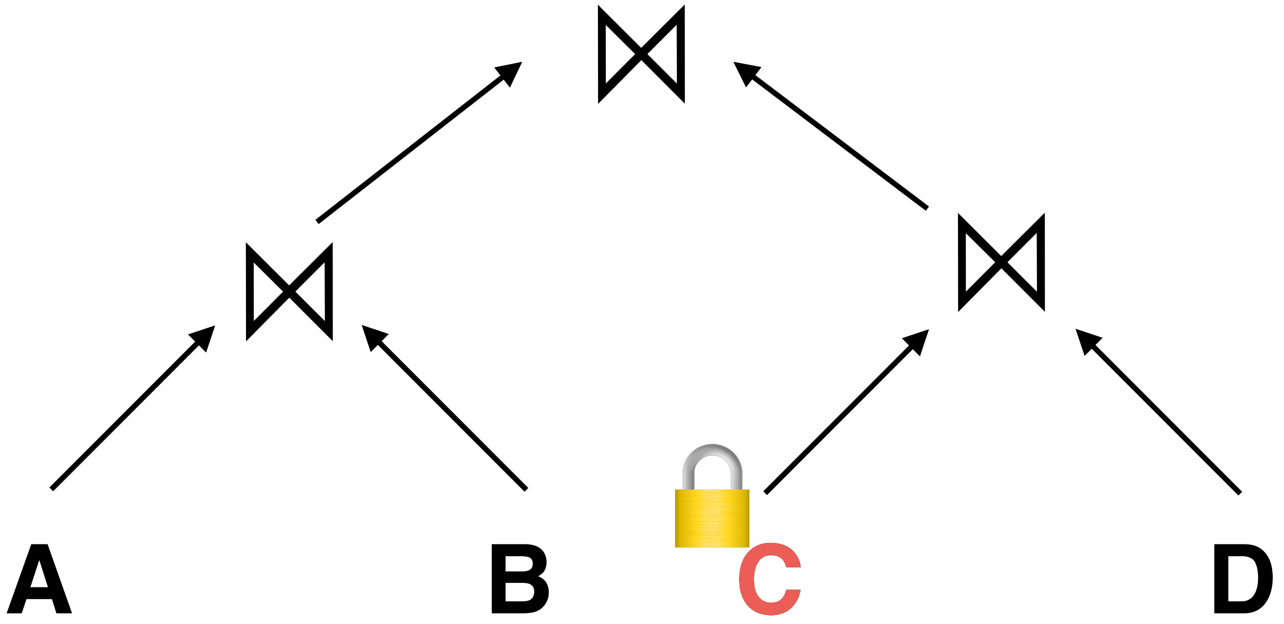
Cost

Observation: not all tables are sensitive

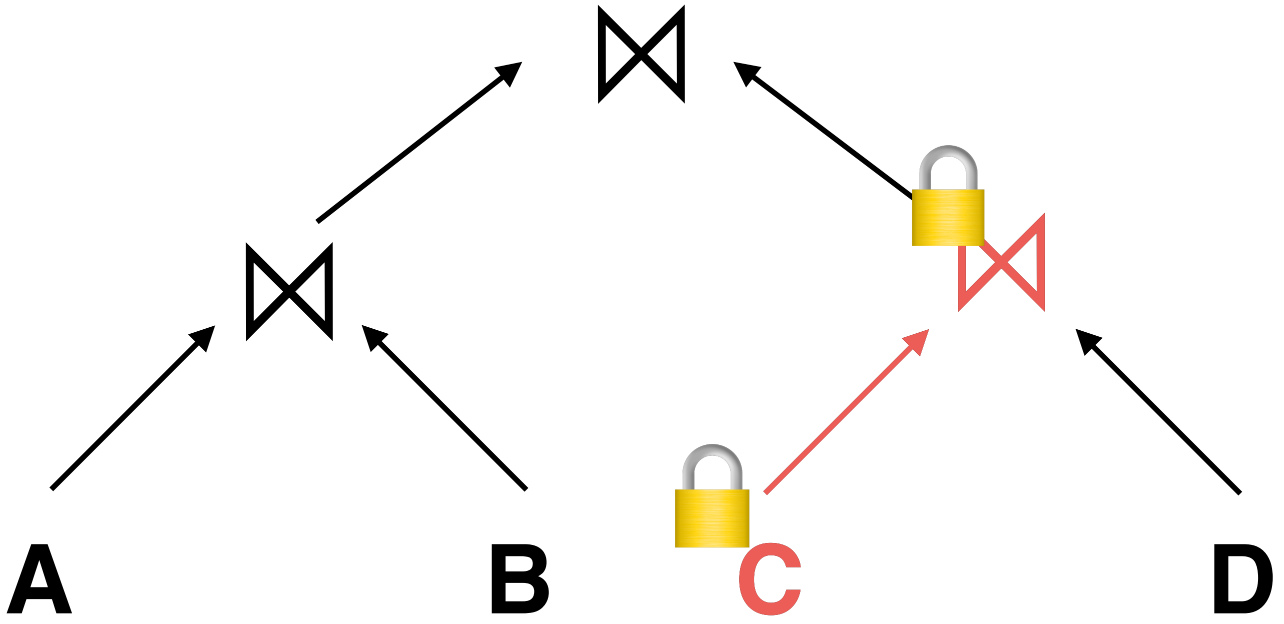


Opaque can operate in *mixed sensitivity*:
sensitive tables are run with oblivious operators

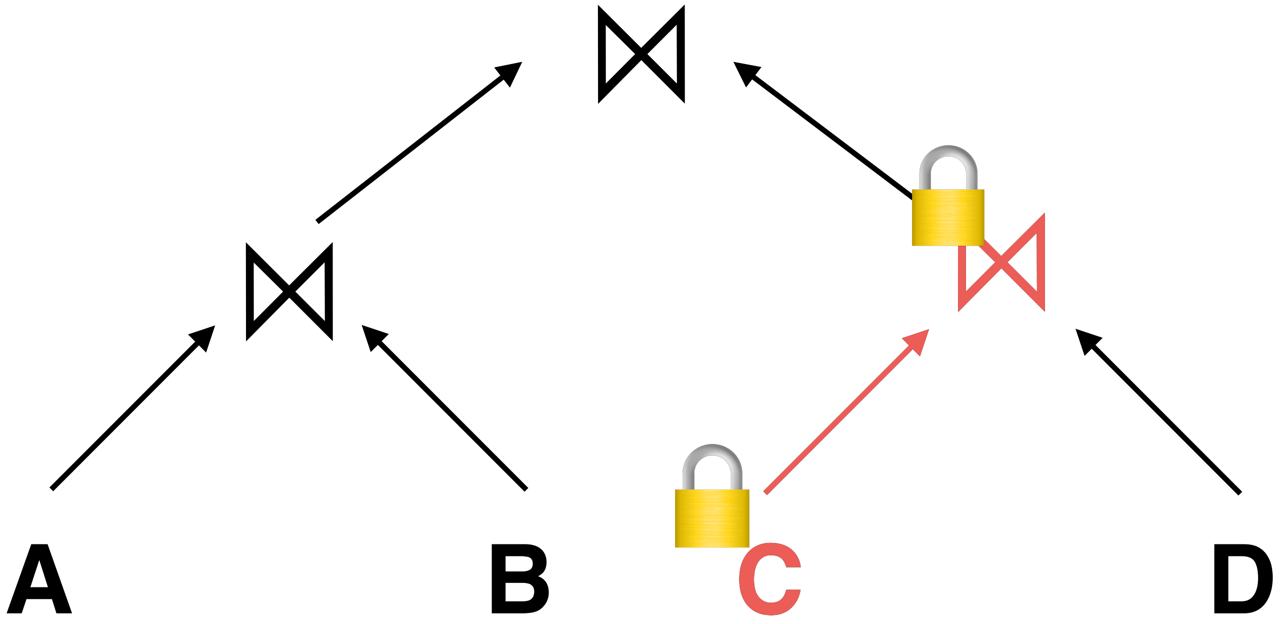
Observation: not all tables are sensitive



Observation: not all tables are sensitive

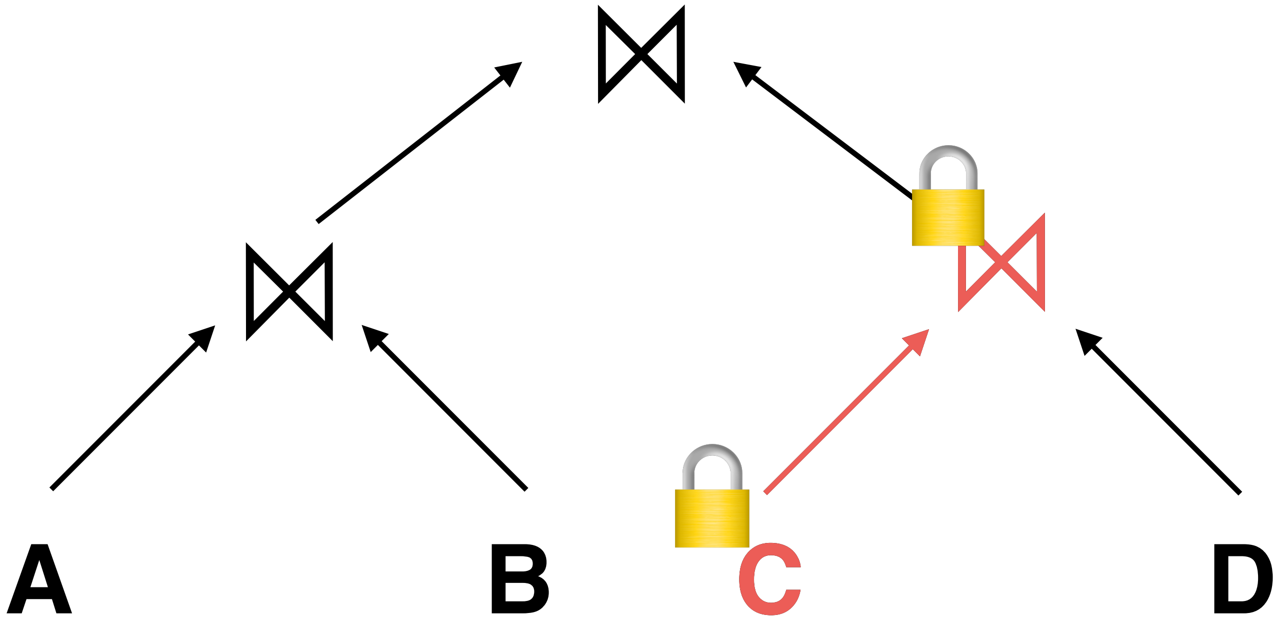


Observation: not all tables are sensitive

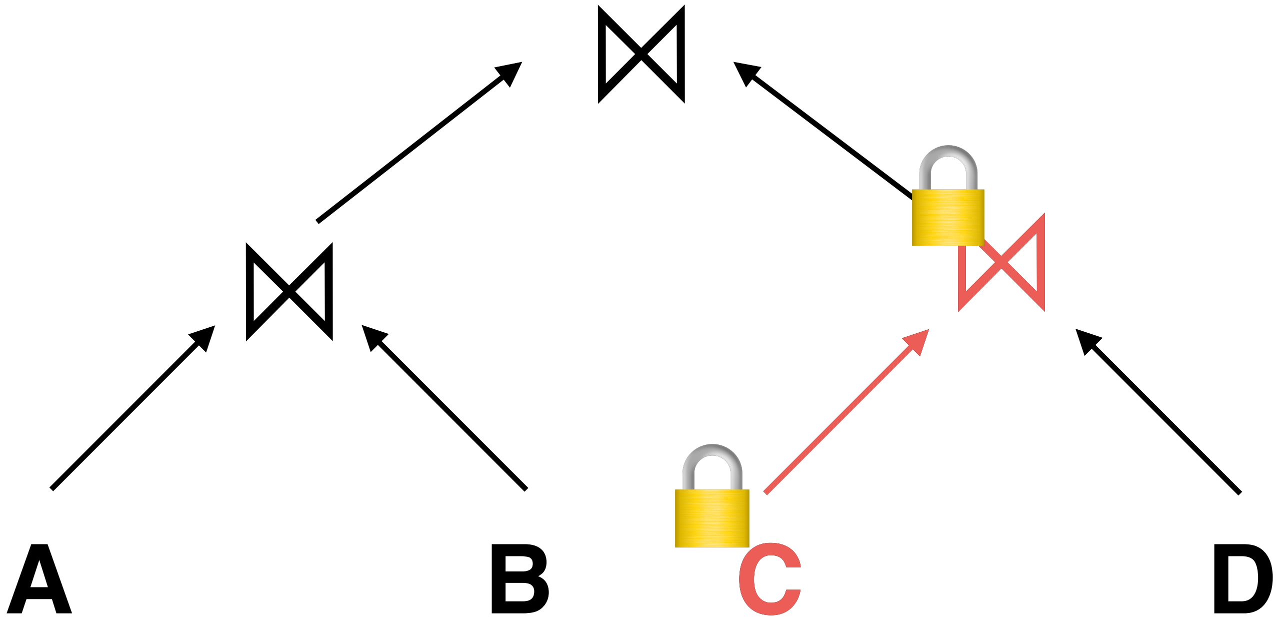


Not oblivious!

Observation: not all tables are sensitive

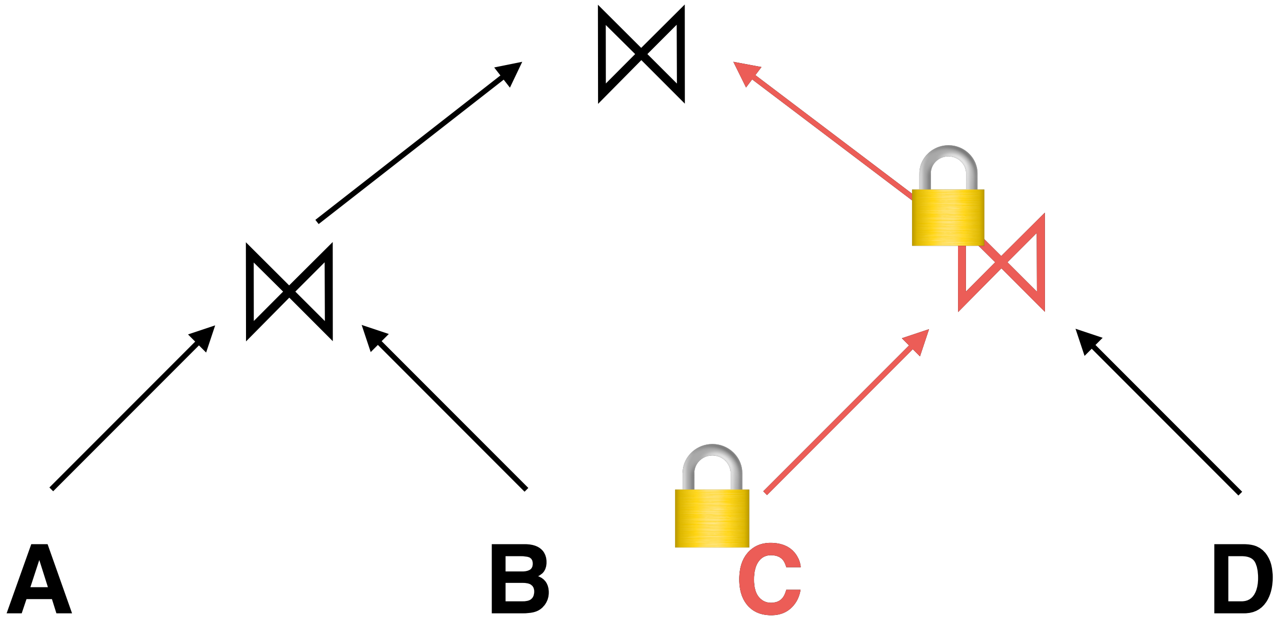


Observation: not all tables are sensitive



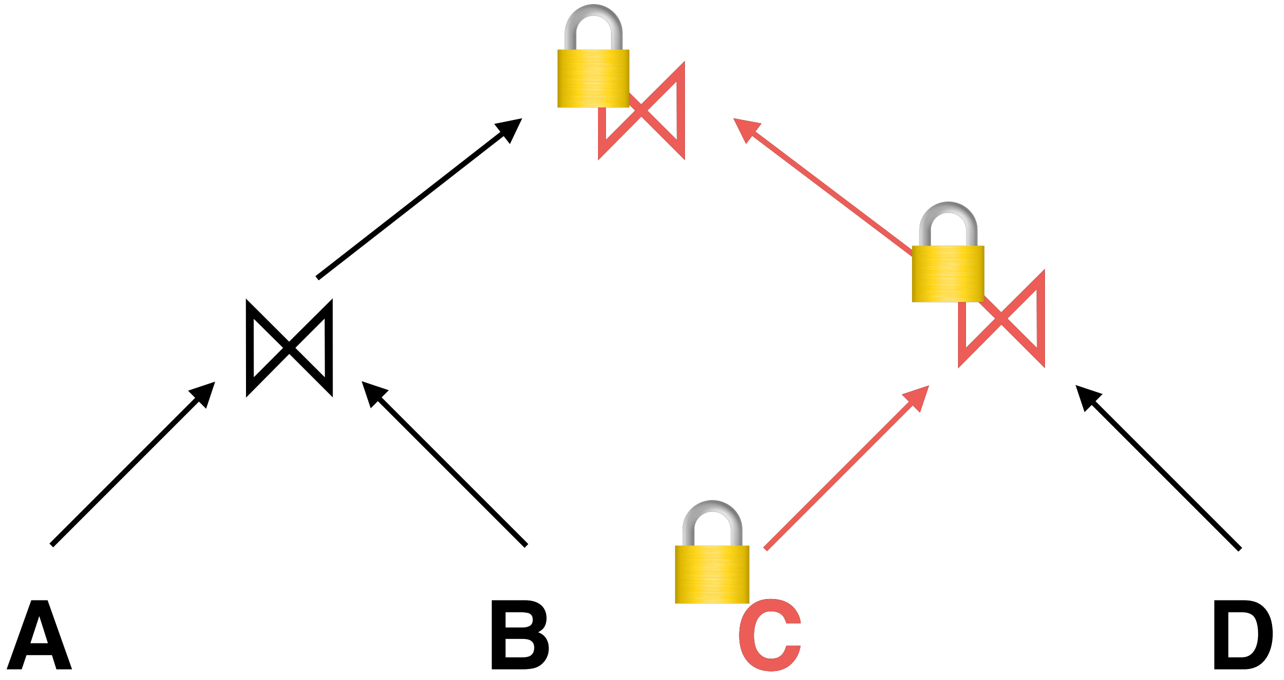
Sensitivity propagation: propagate obliviousness from leaf to root

Observation: not all tables are sensitive



Sensitivity propagation: propagate obliviousness from leaf to root

Observation: not all tables are sensitive

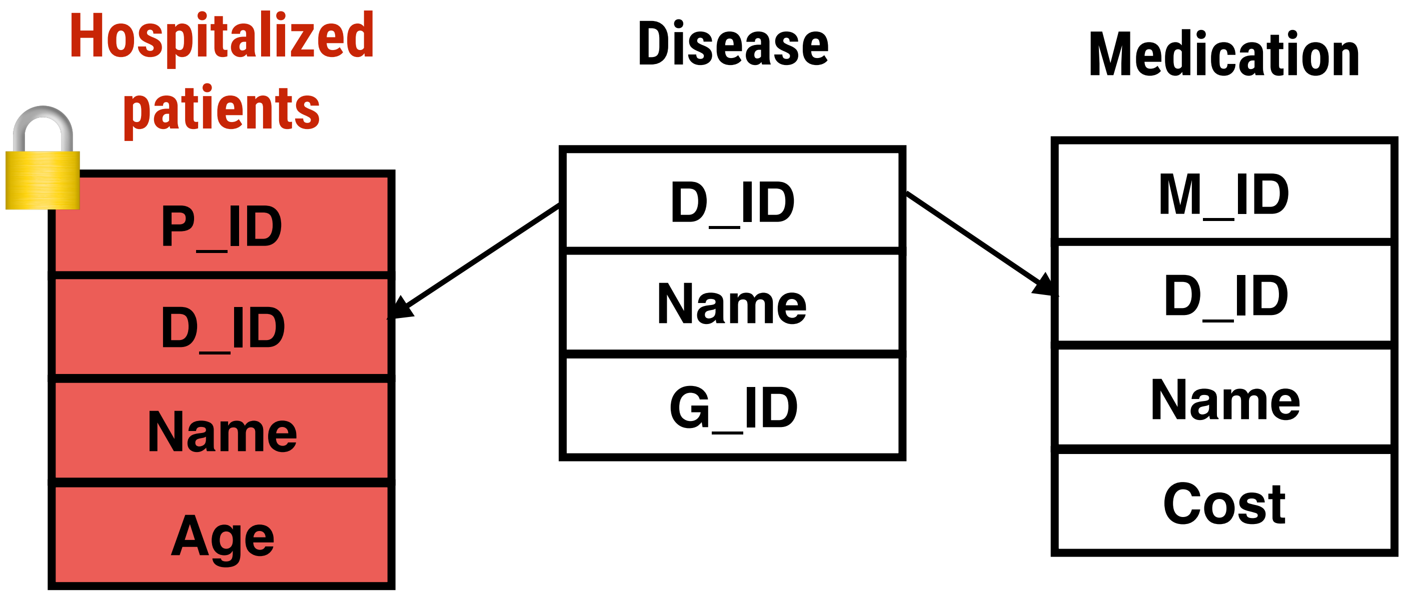


Sensitivity propagation: propagate obliviousness from leaf to root

Insight 2

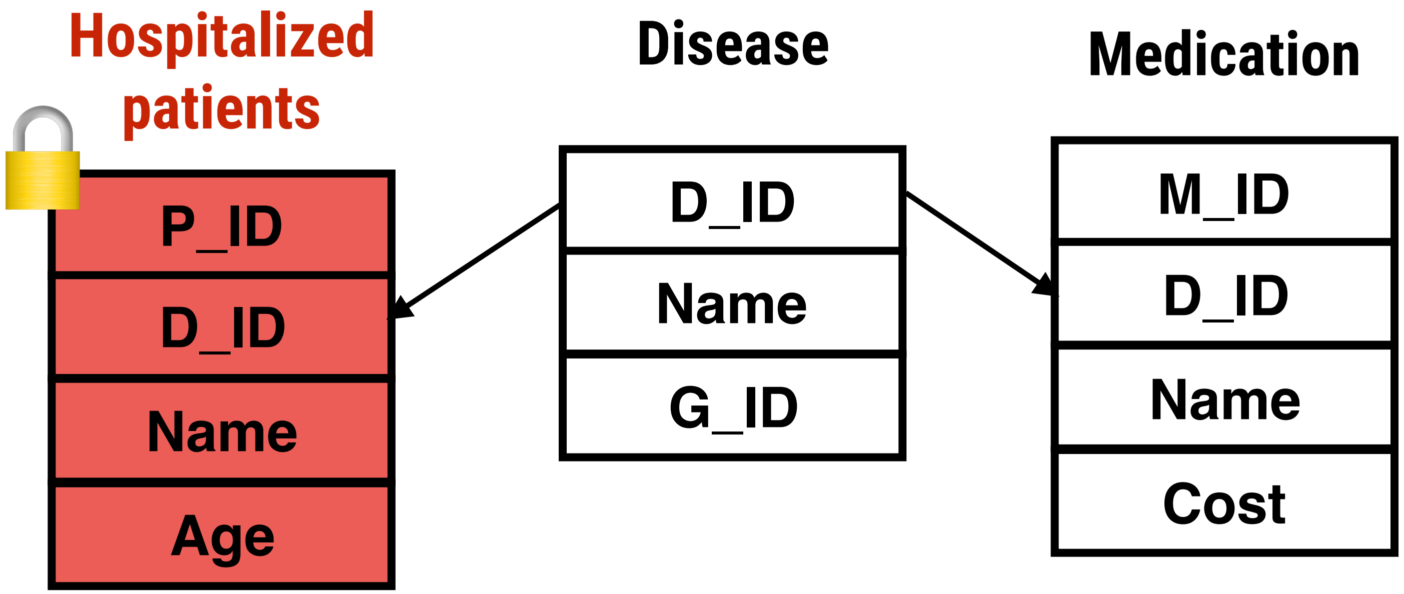
Sensitivity propagation
introduces a new dimension to
query optimization

Cost-based optimization



Find the least costly medication for each patient

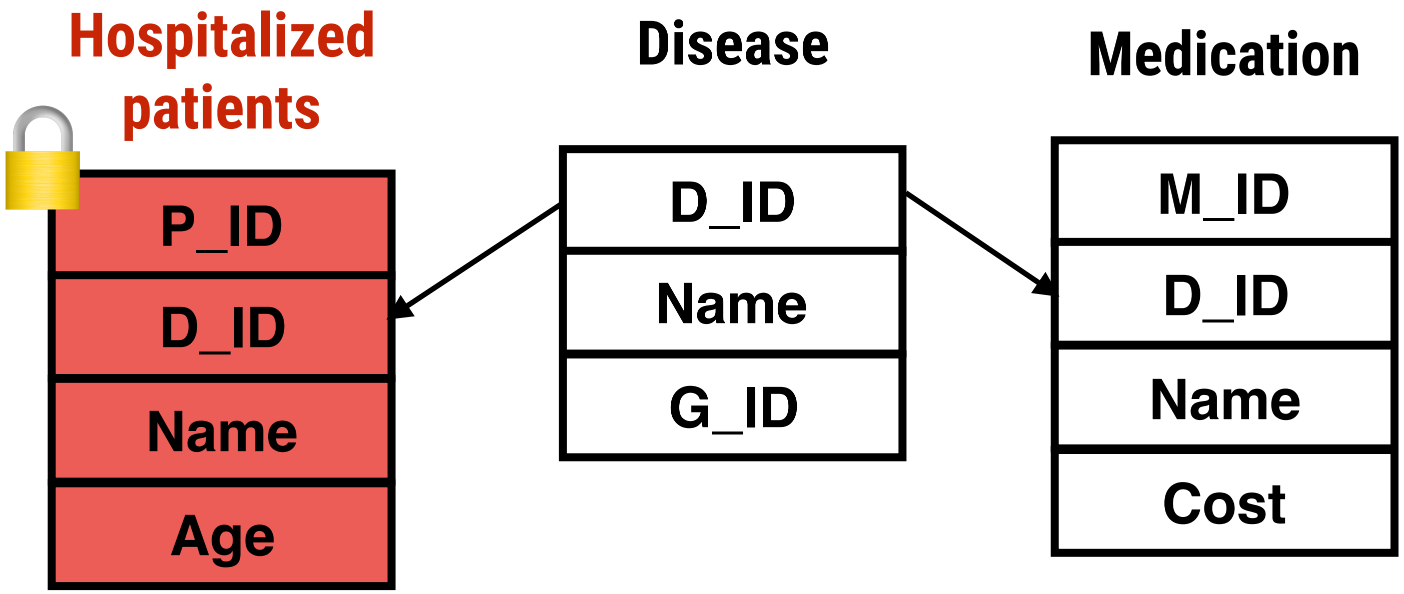
Cost-based optimization



Find the least costly medication for each patient

Assumption: |P| < |D| < |M|

Cost-based optimization

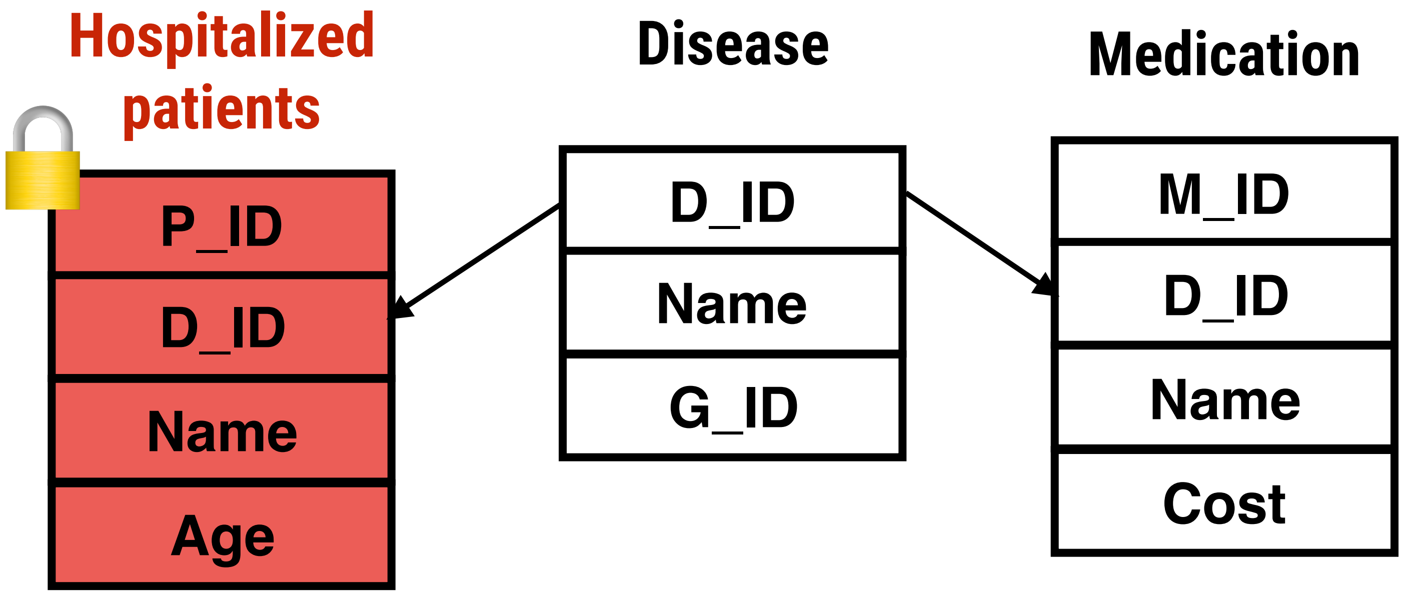


Find the least costly medication for each patient

Assumption: |P| < |D| < |M|

```
SELECT p_name, d_name, med_cost
FROM patient, disease,
     (SELECT d_id, min(cost) AS med_cost
        FROM medication
        GROUP BY d_id) AS med
WHERE disease.d_id = patient.d_id
      AND disease.d_id = med.d_id
```

Cost-based optimization



Hospitalized patients

P_ID

D_ID

Name

Age

Disease

D_ID

Name

G_ID

Medication

M_ID

D_ID

Name

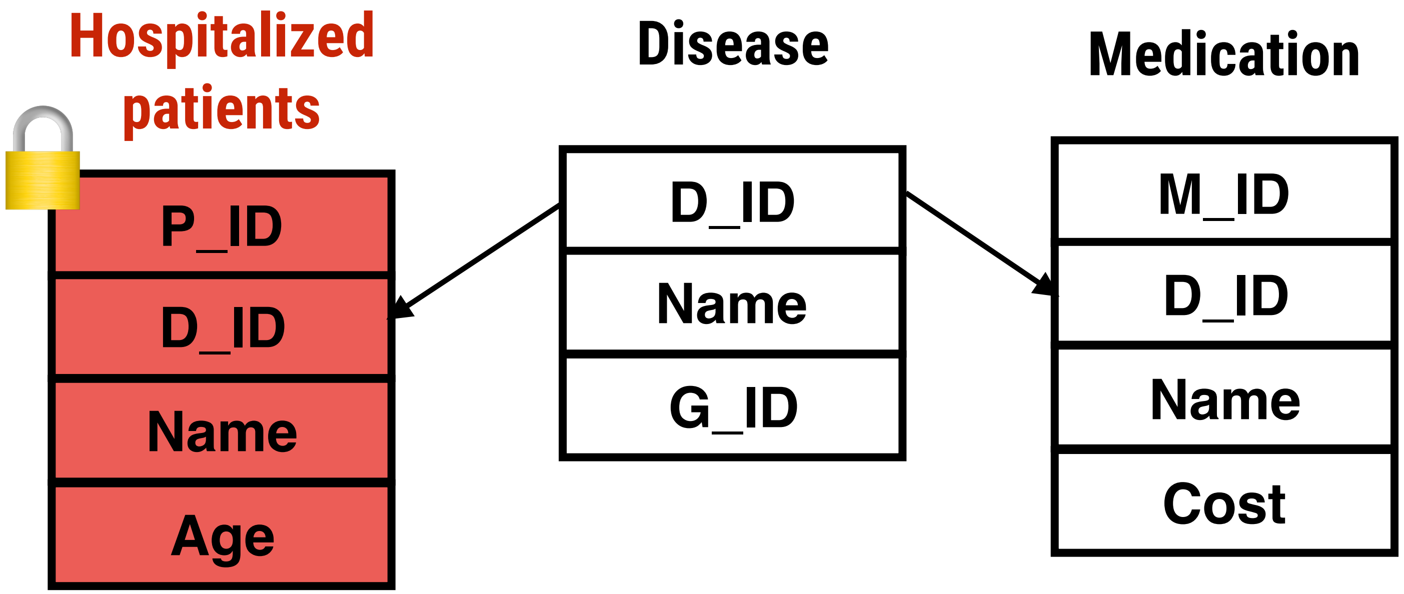
Cost

Find the least costly medication for each patient

Assumption: |P| < |D| < |M|

```
SELECT p_name, d_name, med_cost
FROM patient, disease,
     (SELECT d_id, min(cost) AS med_cost
        FROM medication
        GROUP BY d_id) AS med
WHERE disease.d_id = patient.d_id
      AND disease.d_id = med.d_id
```

Cost-based optimization



Hospitalized patients

P_ID

D_ID

Name

Age

Disease

D_ID

Name

G_ID

Medication

M_ID

D_ID

Name

Cost

Find the least costly medication for each patient

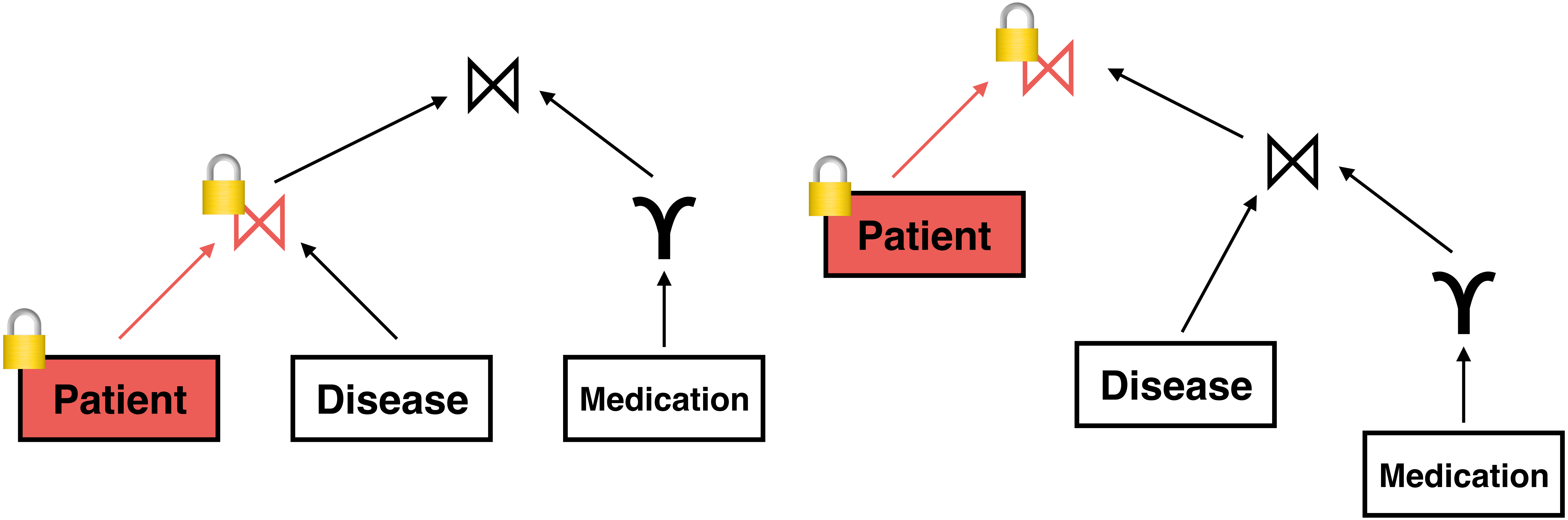
Assumption: |P| < |D| < |M|

```
SELECT p_name, d_name, med_cost
FROM patient, disease,
     (SELECT d_id, min(cost) AS med_cost
        FROM medication
        GROUP BY d_id) AS med
WHERE disease.d_id = patient.d_id
      AND disease.d_id = med.d_id
```

3-way join

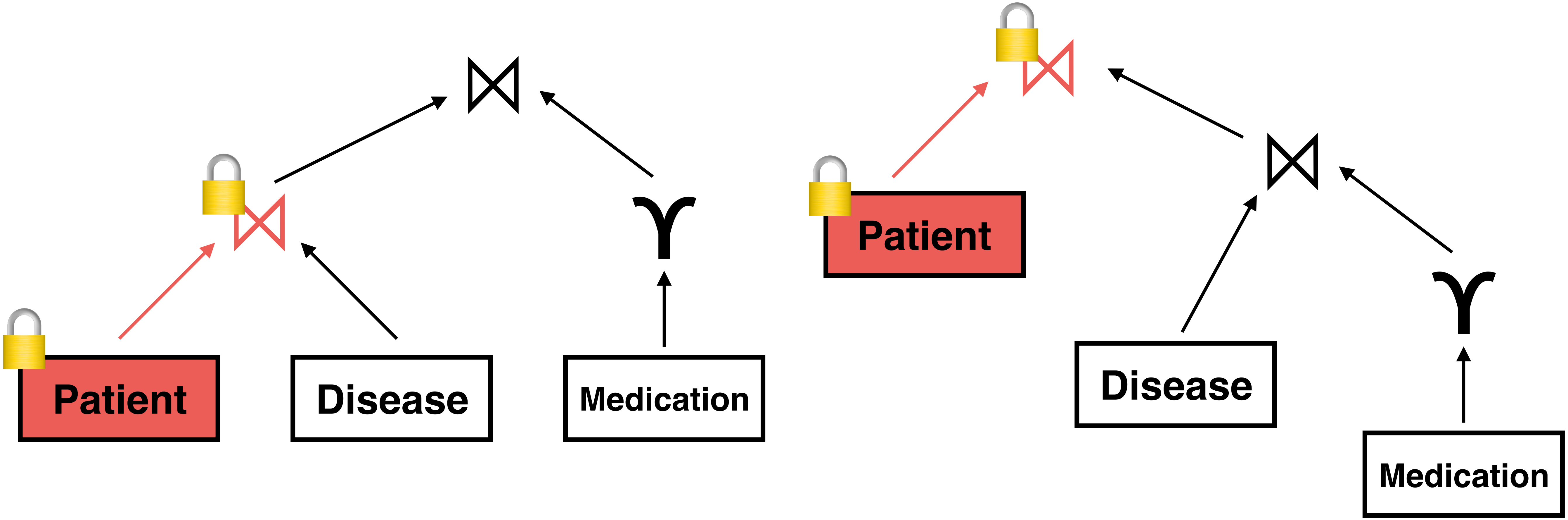
Cost-based optimization

SQL optimizer with new cost:



Cost-based optimization

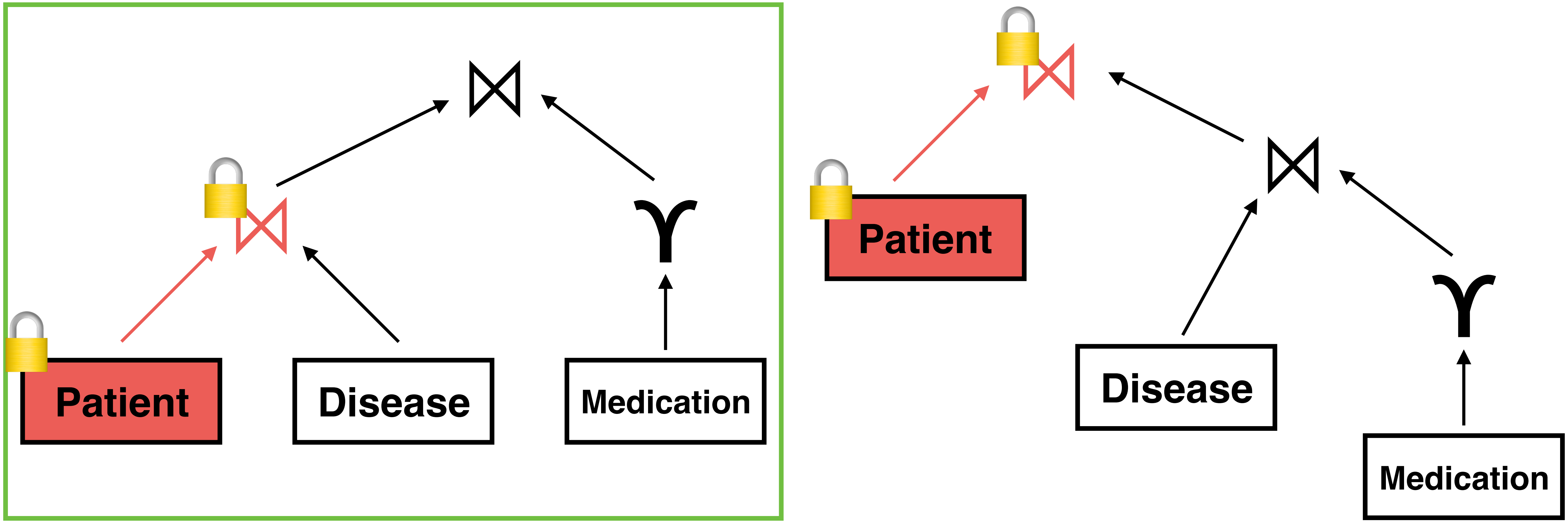
SQL optimizer with new cost:



More selective
non-oblivious join

Cost-based optimization

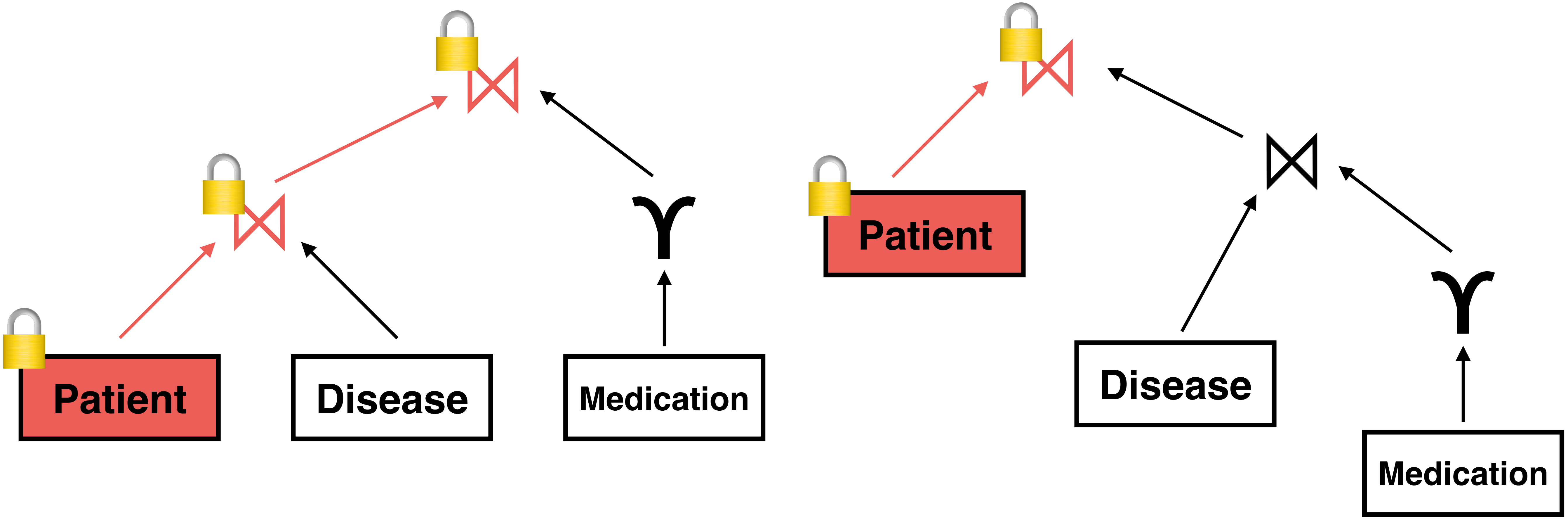
SQL optimizer with new cost:



More selective
non-oblivious join

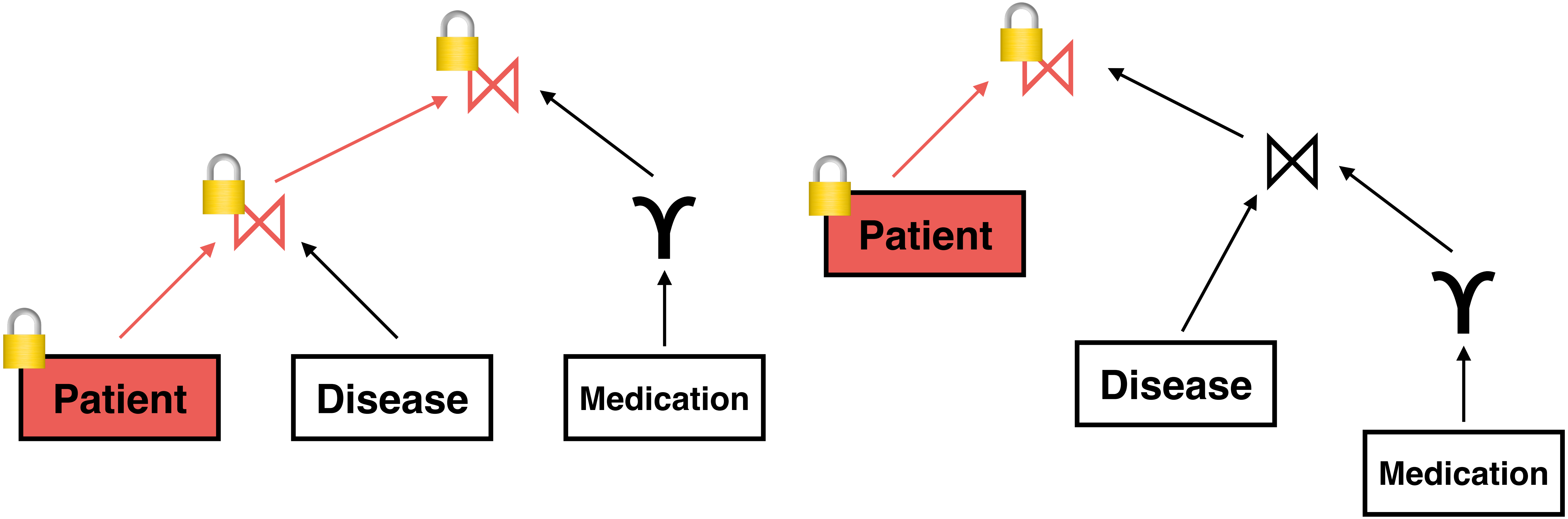
Cost-based optimization

SQL optimizer with new cost and **sensitivity propagation**:



Cost-based optimization

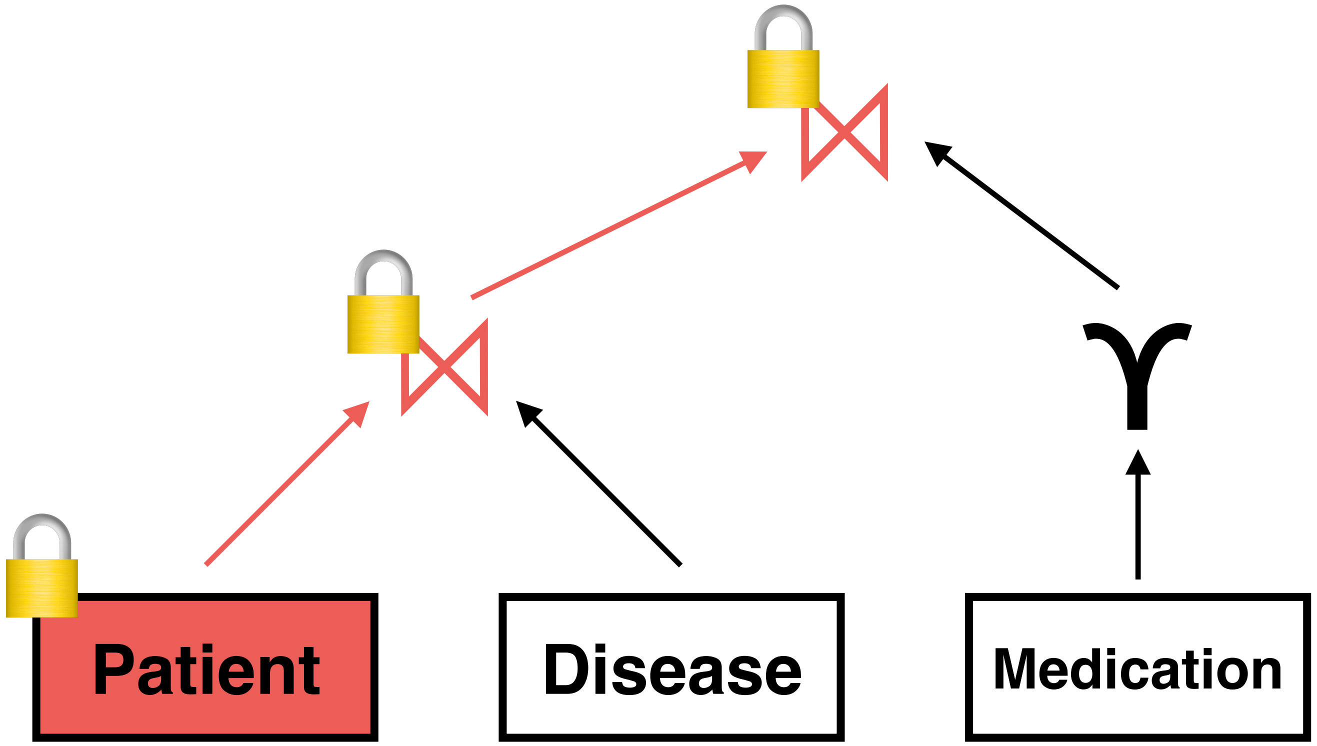
SQL optimizer with new cost and **sensitivity propagation**:

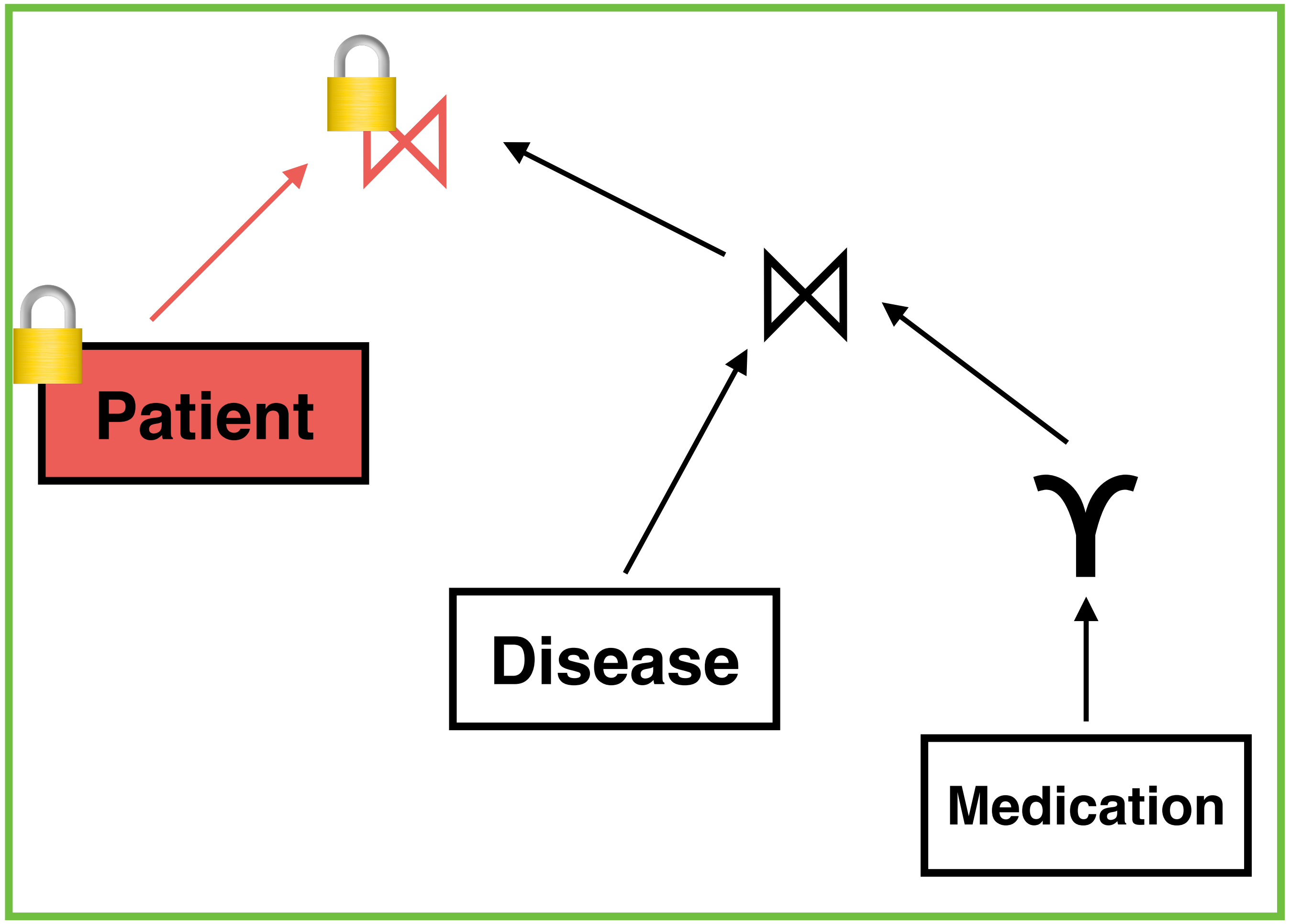


Fewer oblivious joins

Cost-based optimization

SQL optimizer with new cost and **sensitivity propagation**:





Fewer oblivious joins

Evaluation setup

Evaluation setup

- Single machine experiments:
  - Intel Xeon E3-1280 v5, 4 cores, 64 GB RAM
  - Intel SGX: 128 MB of enclave page cache (EPC)

Evaluation setup

- Single machine experiments:
    - Intel Xeon E3-1280 v5, 4 cores, 64 GB RAM
    - Intel SGX: 128 MB of enclave page cache (EPC)
- Distributed experiments
    - A cluster of 5 SGX machines

Evaluation

Evaluation

- **How does Opaque compare to Spark SQL?**

Evaluation

- **How does Opaque compare to Spark SQL?**
    - Big Data Benchmark (BDB); 4 queries total

Evaluation

- **How does Opaque compare to Spark SQL?**
  - Big Data Benchmark (BDB); 4 queries total
    - Queries 1, 2, 3: filter, aggregation, join

Evaluation

- **How does Opaque compare to Spark SQL?**
  - Big Data Benchmark (BDB); 4 queries total
    - Queries 1, 2, 3: filter, aggregation, join
    - 1 million records

Evaluation

- **How does Opaque compare to Spark SQL?**
  - Big Data Benchmark (BDB); 4 queries total
    - Queries 1, 2, 3: filter, aggregation, join
    - 1 million records
- **How does Opaque compare to state-of-the-art oblivious systems?**

Evaluation

- **How does Opaque compare to Spark SQL?**
  - Big Data Benchmark (BDB); 4 queries total
    - Queries 1, 2, 3: filter, aggregation, join
    - 1 million records
- **How does Opaque compare to state-of-the-art oblivious systems?**
  - GraphSC (oblivious graph analytics)

Evaluation

- **How does Opaque compare to Spark SQL?**
  - Big Data Benchmark (BDB); 4 queries total
    - Queries 1, 2, 3: filter, aggregation, join
    - 1 million records
- **How does Opaque compare to state-of-the-art oblivious systems?**
  - GraphSC (oblivious graph analytics)
    - PageRank

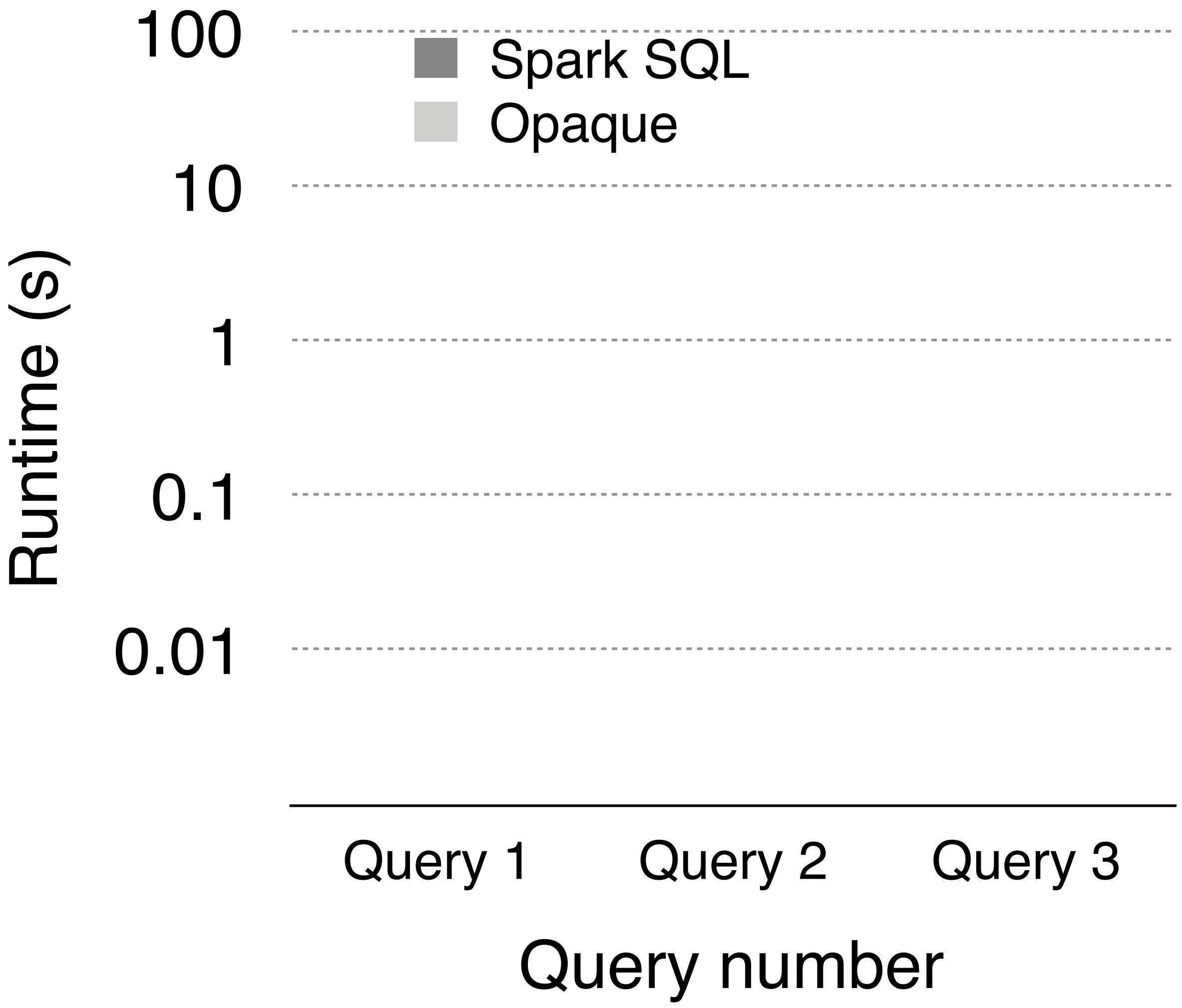
Big Data Benchmark (distributed)

Big Data Benchmark (distributed)

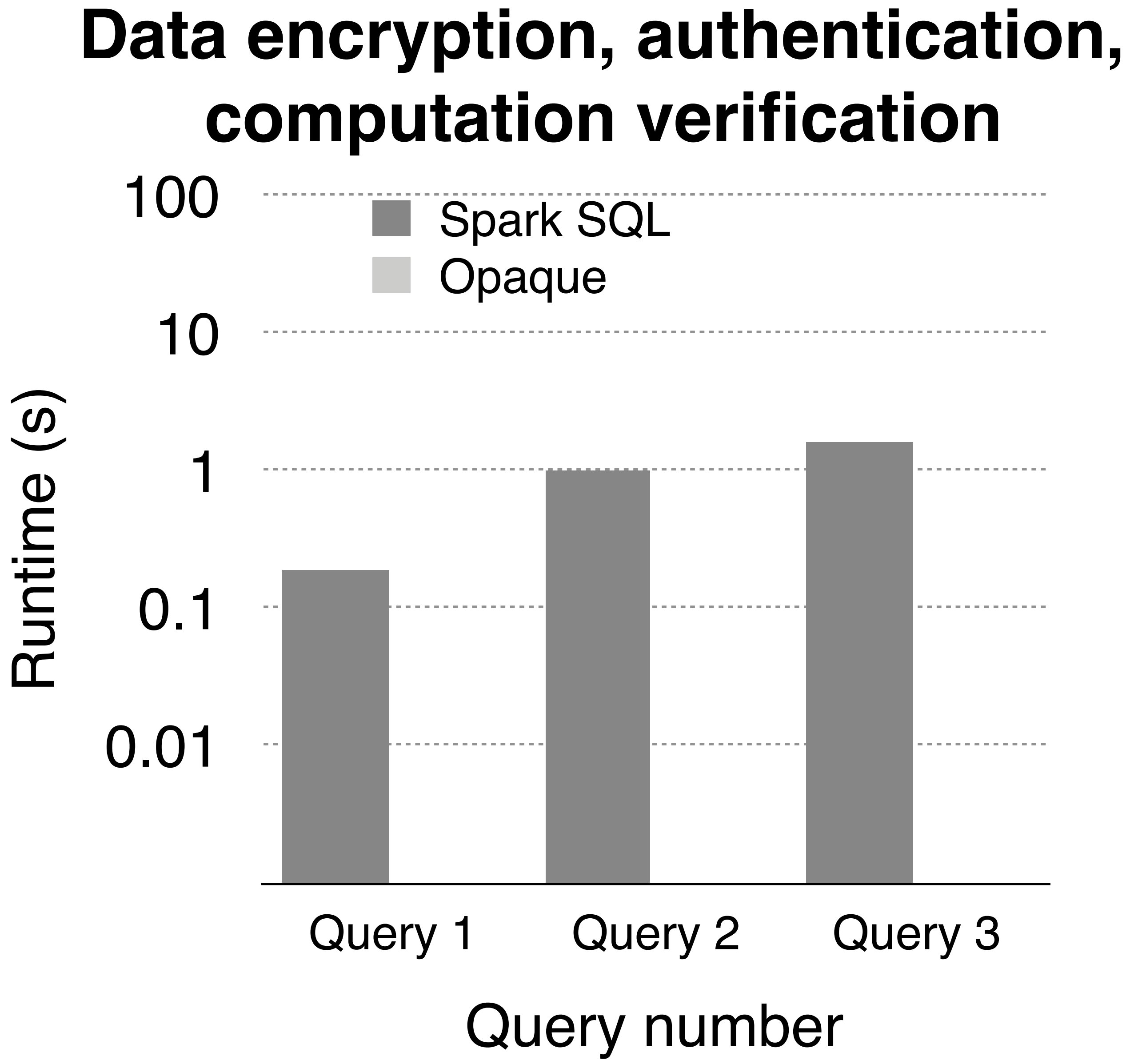
Data encryption, authentication, computation verification

Big Data Benchmark (distributed)

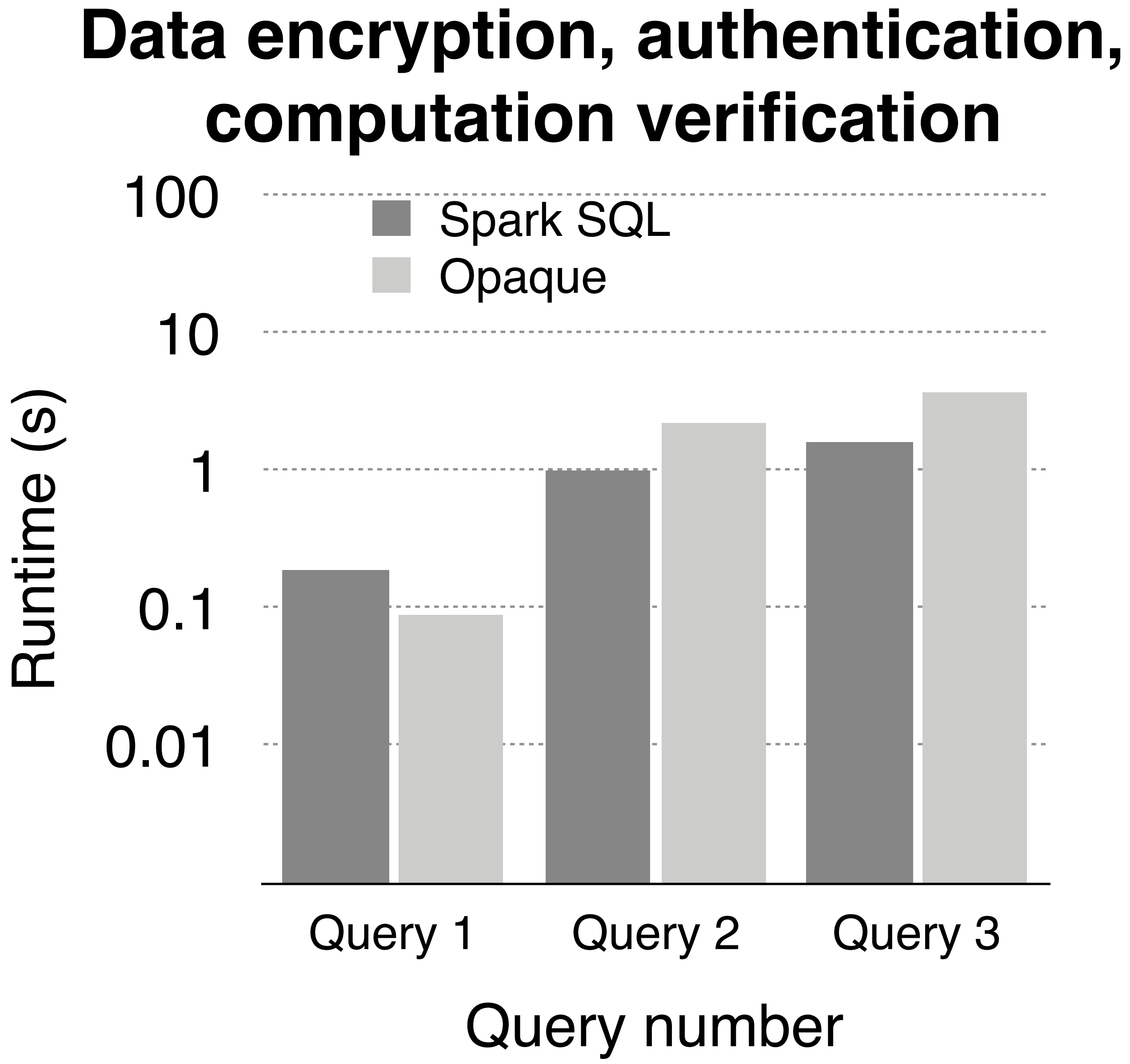
Data encryption, authentication, computation verification



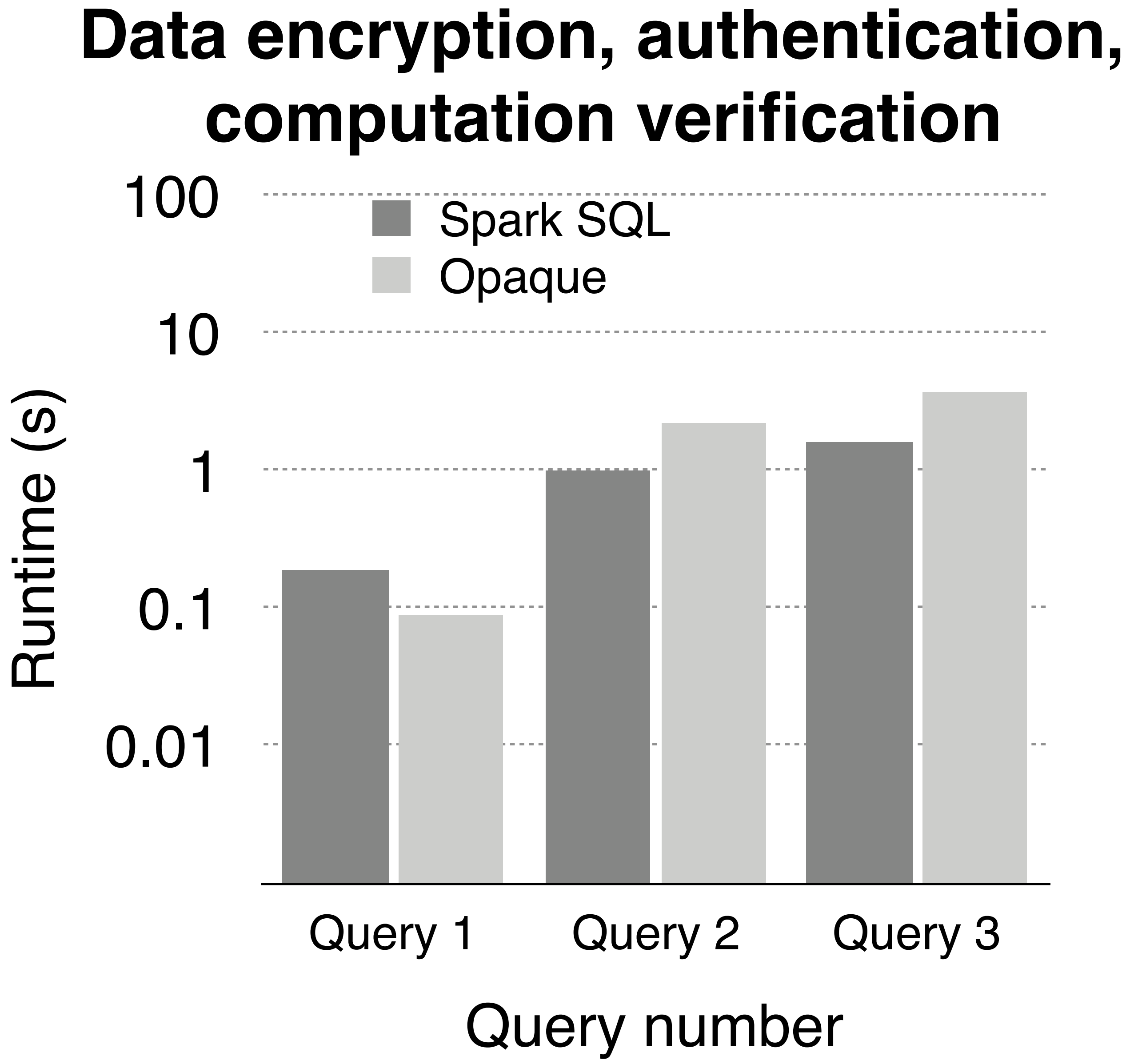
Big Data Benchmark (distributed)



Big Data Benchmark (distributed)



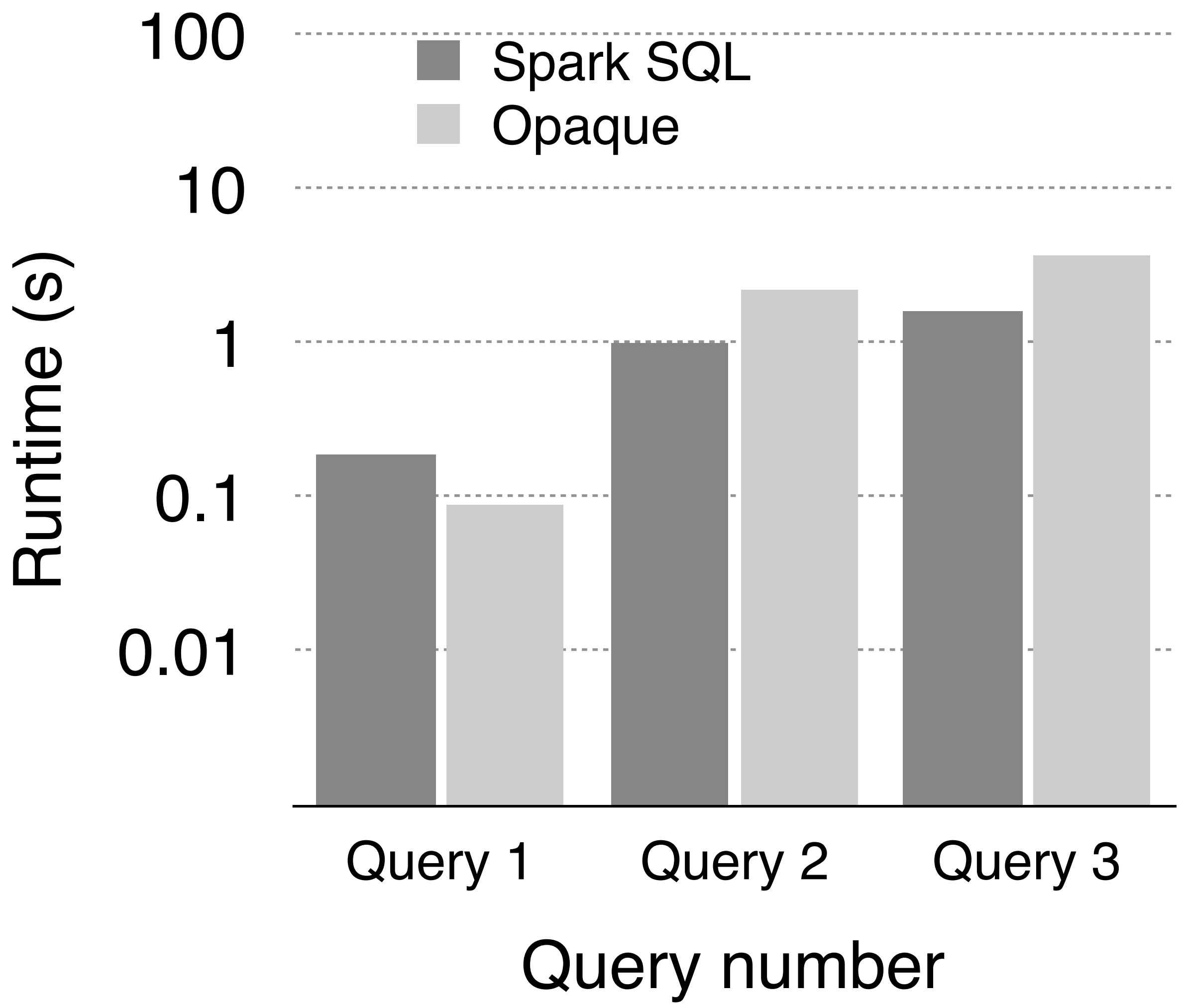
Big Data Benchmark (distributed)



Overhead: 0.47x to 2.3x

Big Data Benchmark (distributed)

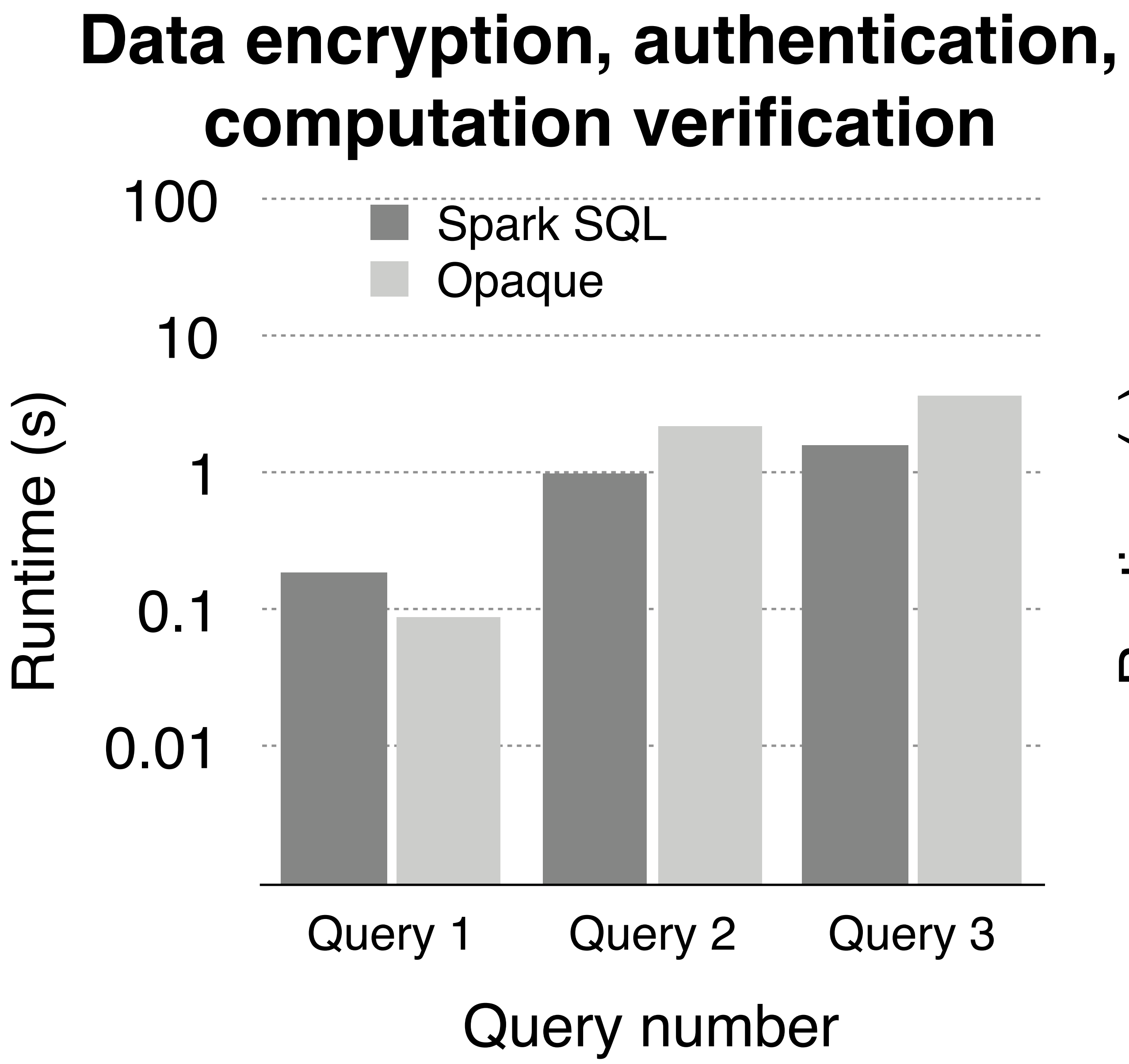
Data encryption, authentication, computation verification

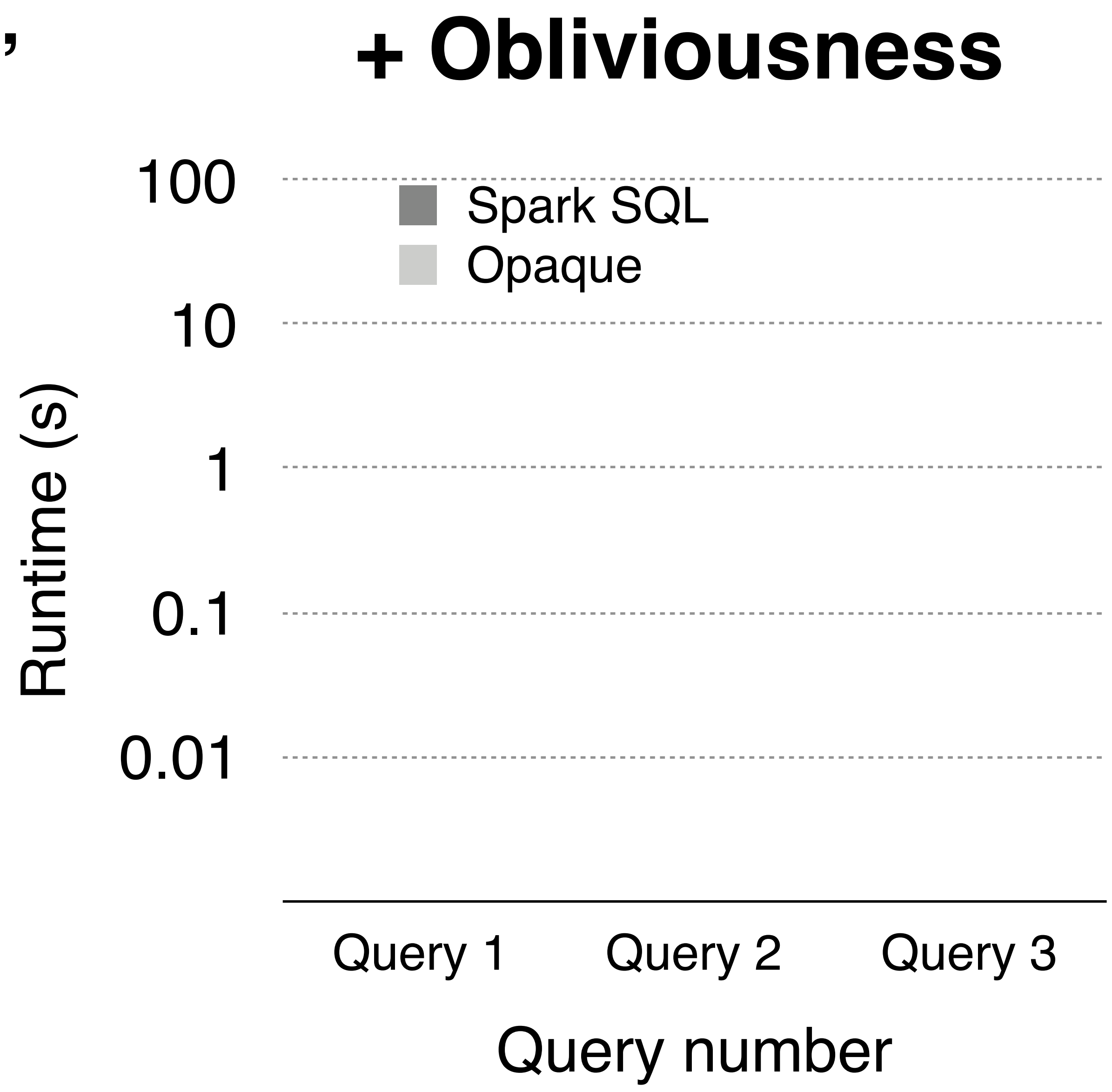


+ Obliviousness

Overhead: 0.47x to 2.3x

Big Data Benchmark (distributed)

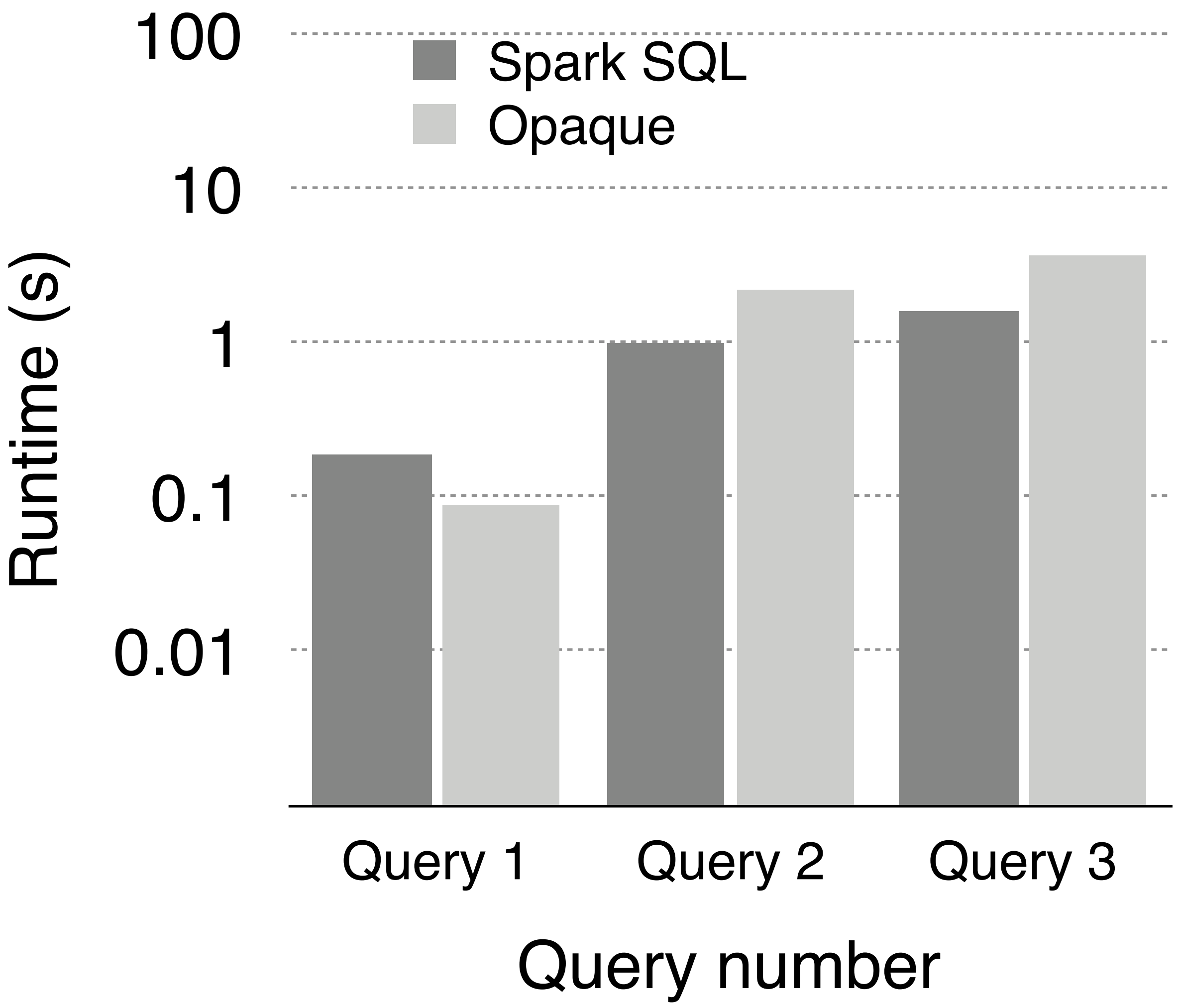




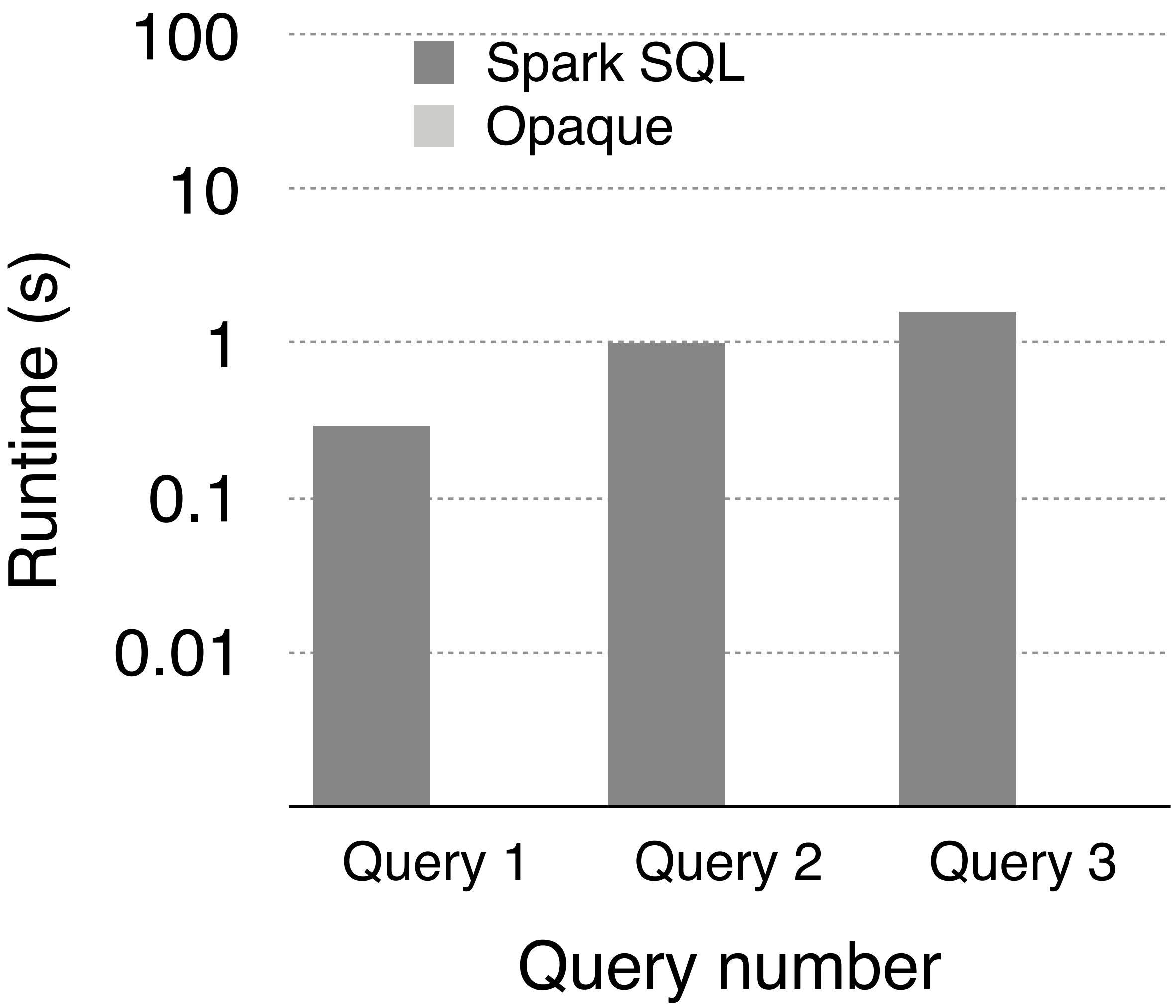
Overhead: 0.47x to 2.3x

Big Data Benchmark (distributed)

Data encryption, authentication, computation verification



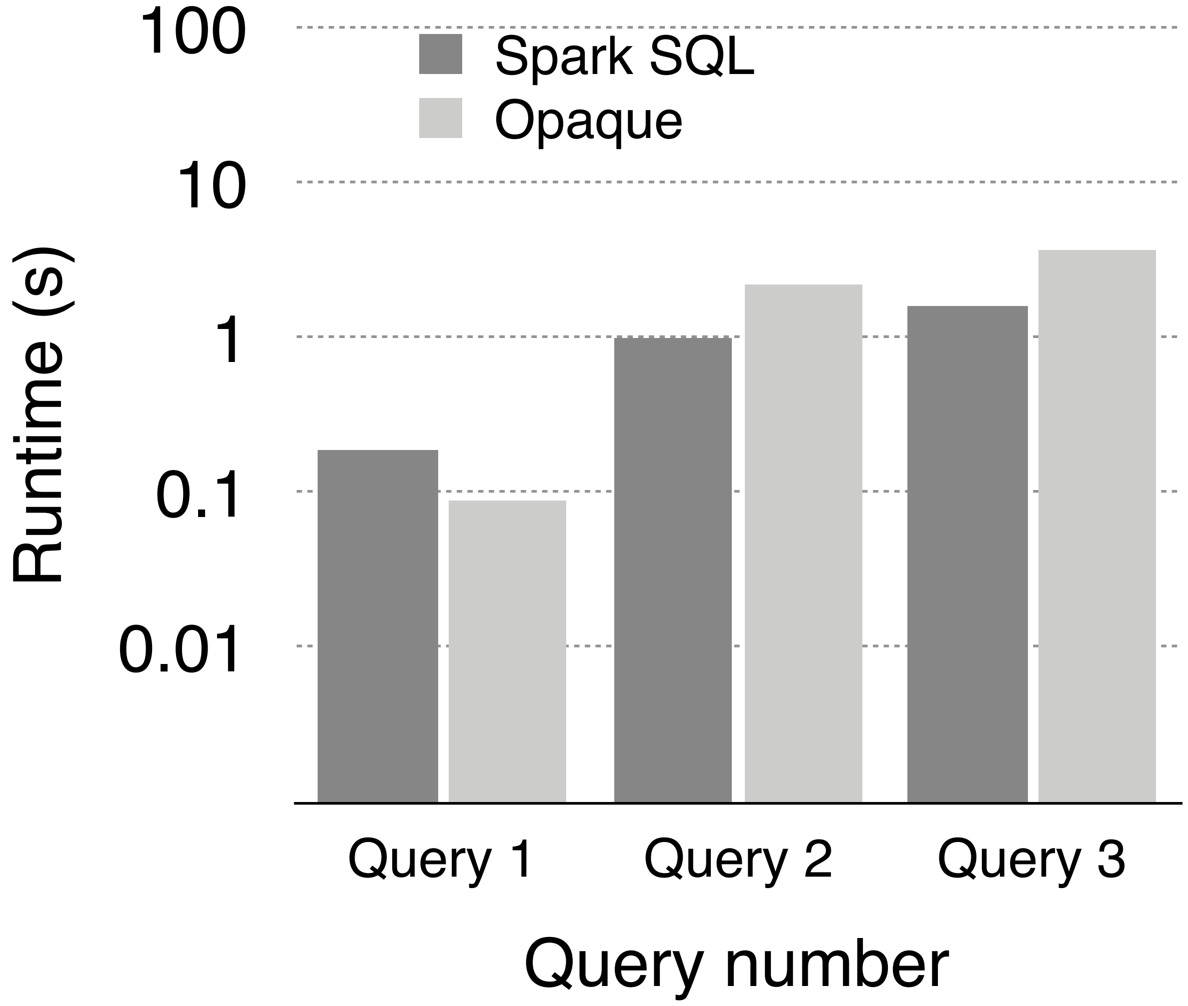
+ Obliviousness



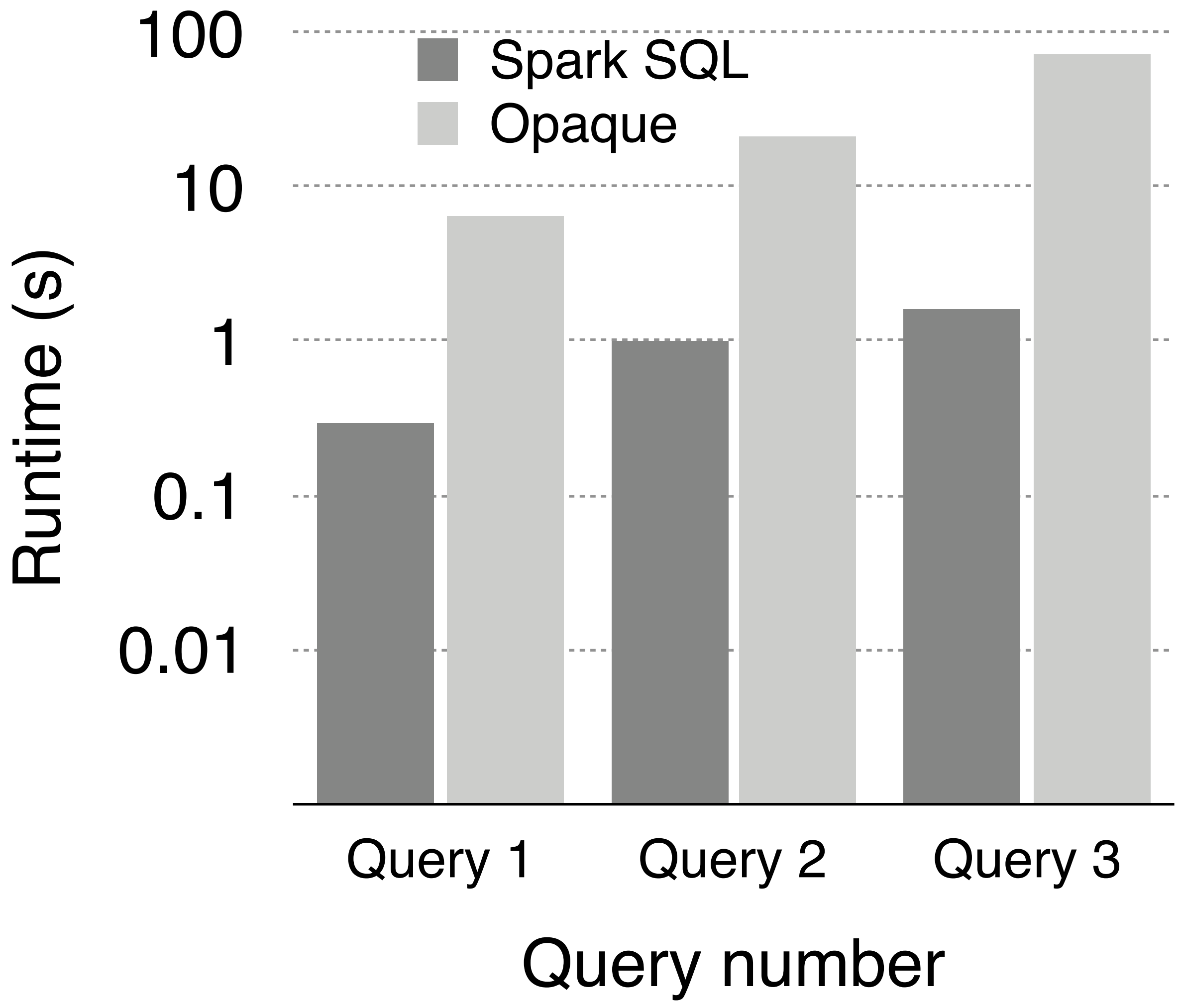
Overhead: 0.47x to 2.3x

Big Data Benchmark (distributed)

Data encryption, authentication, computation verification



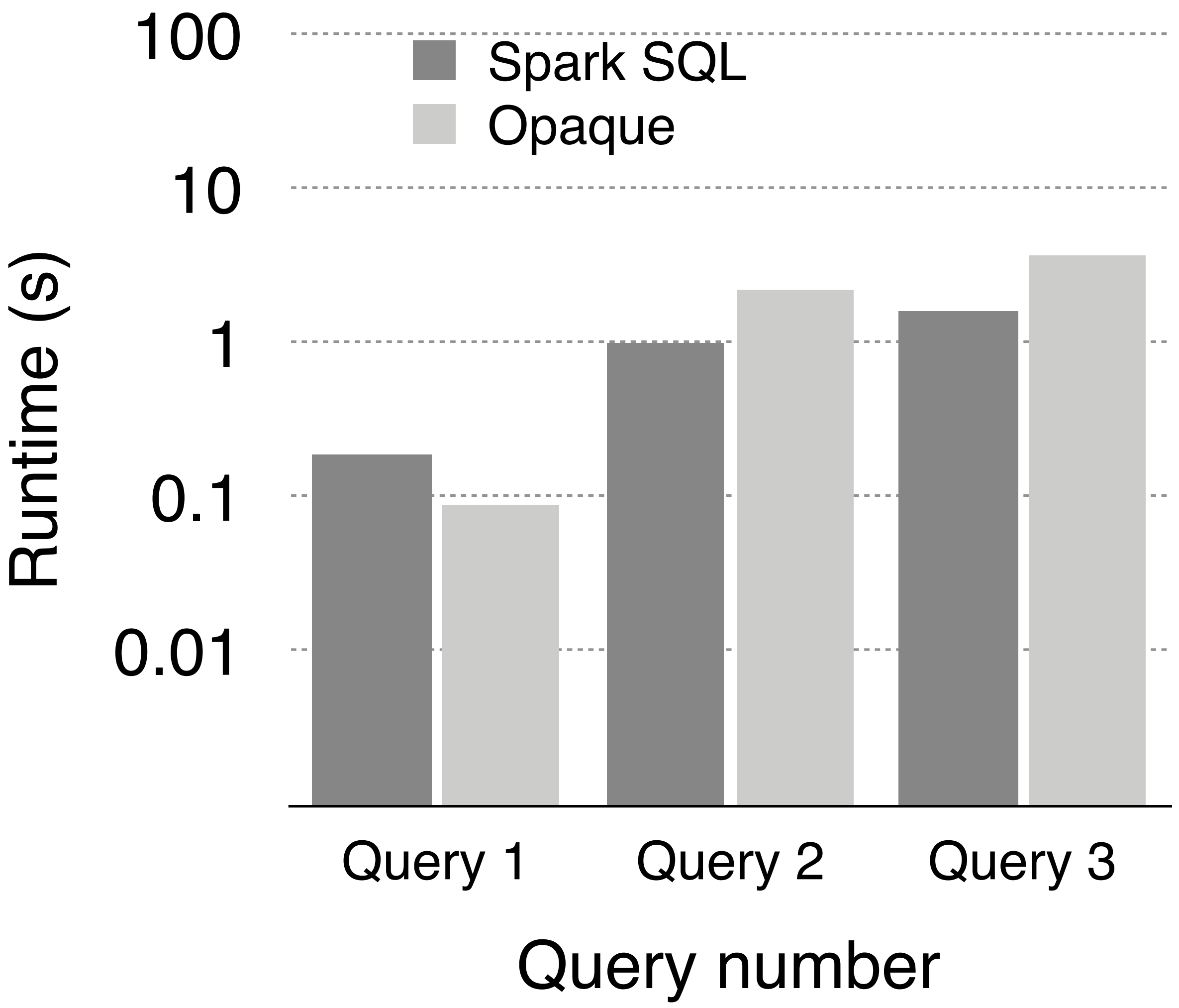
+ Obliviousness



Overhead: 0.47x to 2.3x

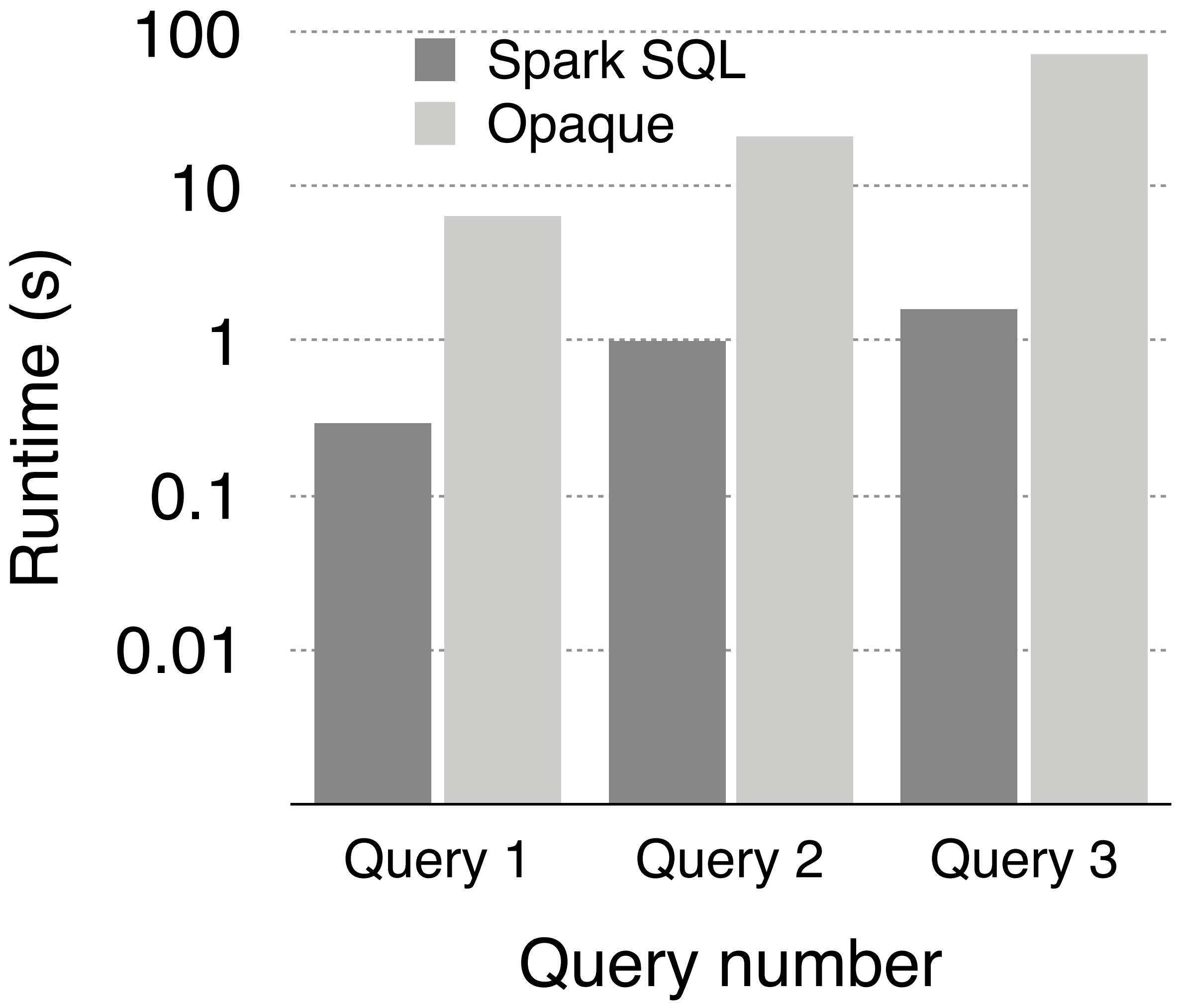
Big Data Benchmark (distributed)

Data encryption, authentication, computation verification



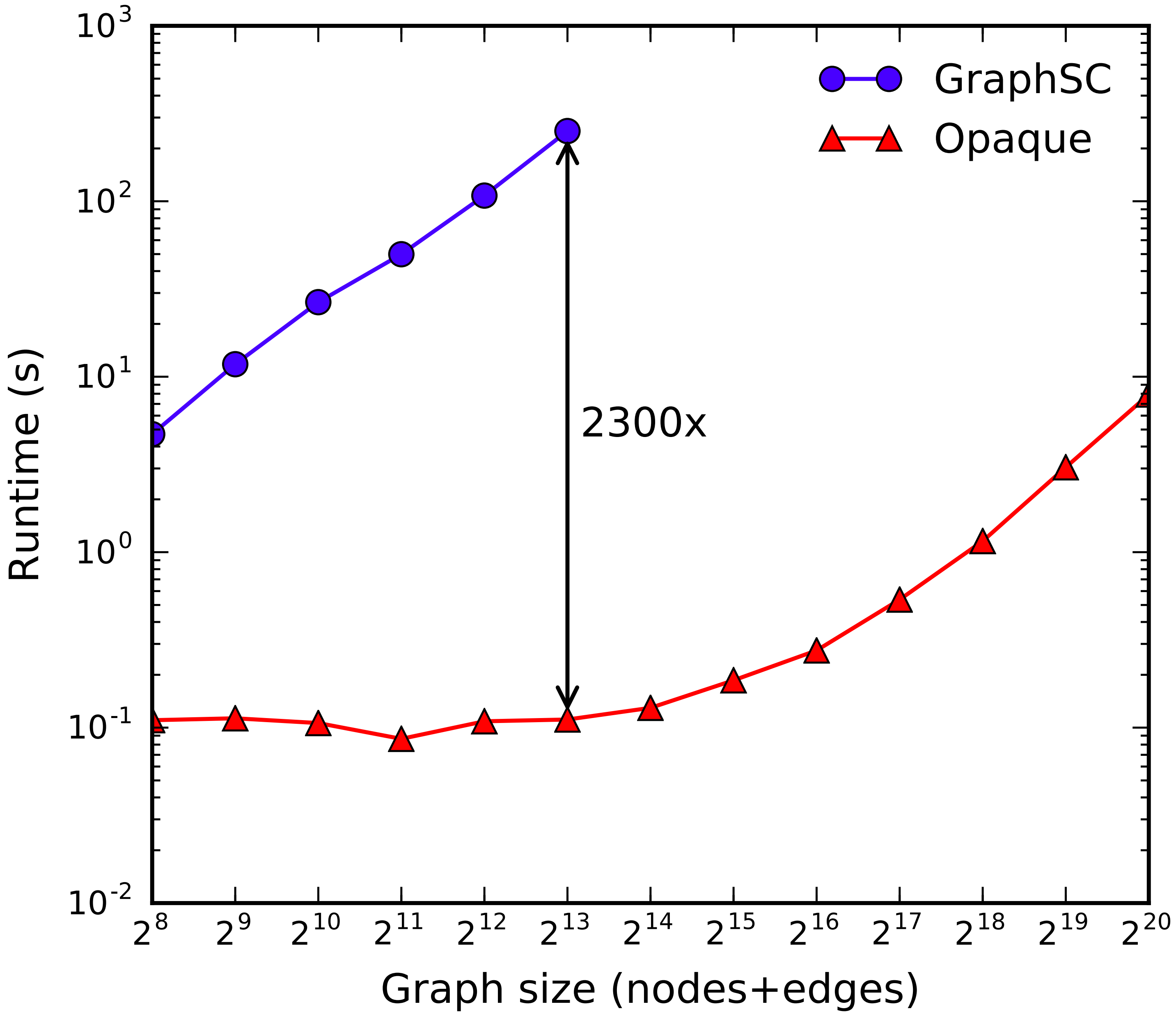
Overhead: 0.47x to 2.3x

+ Obliviousness



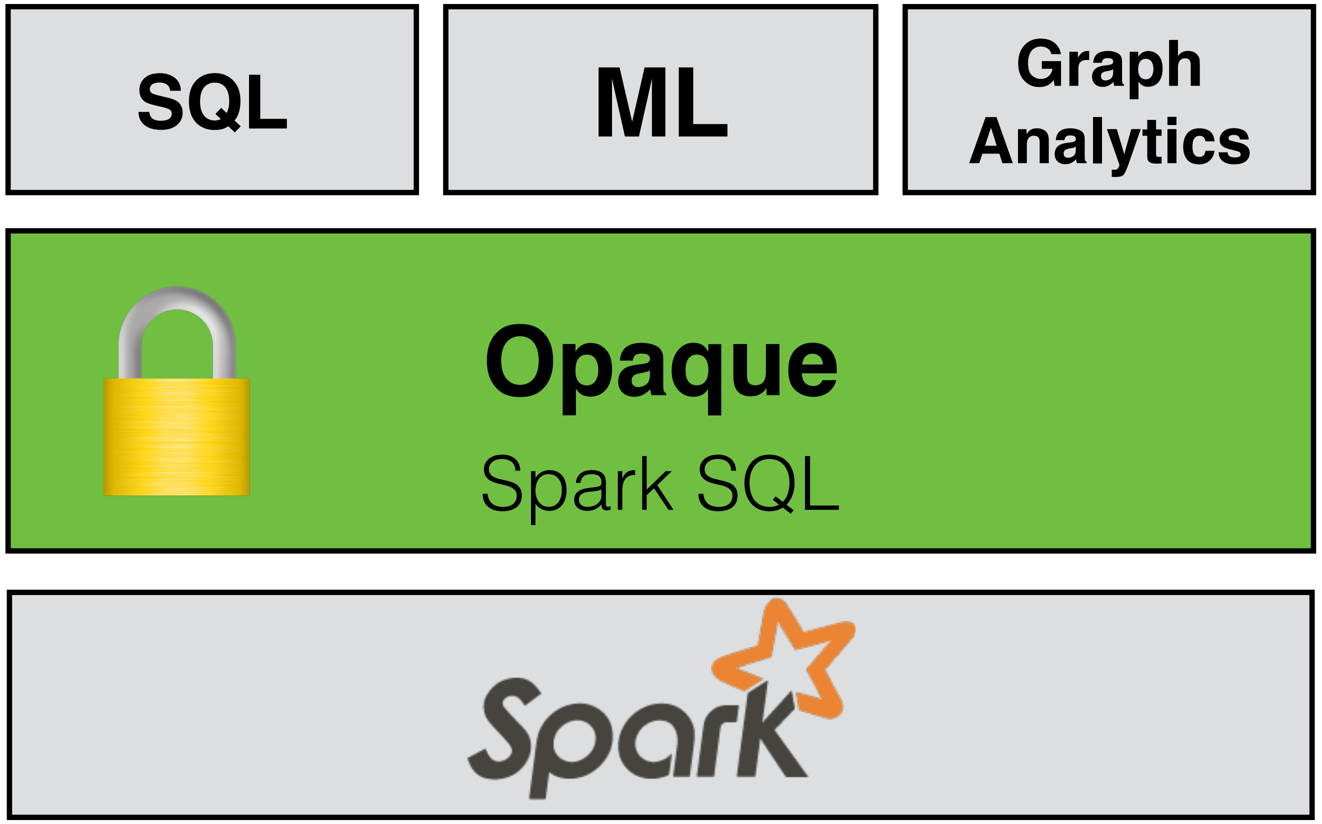
Overhead: 21x to 45x

PageRank: comparison with GraphSC (single machine)



Conclusion

Opaque is an oblivious and encrypted distributed analytics platform



Open source: github.com/ucbrise/opaque