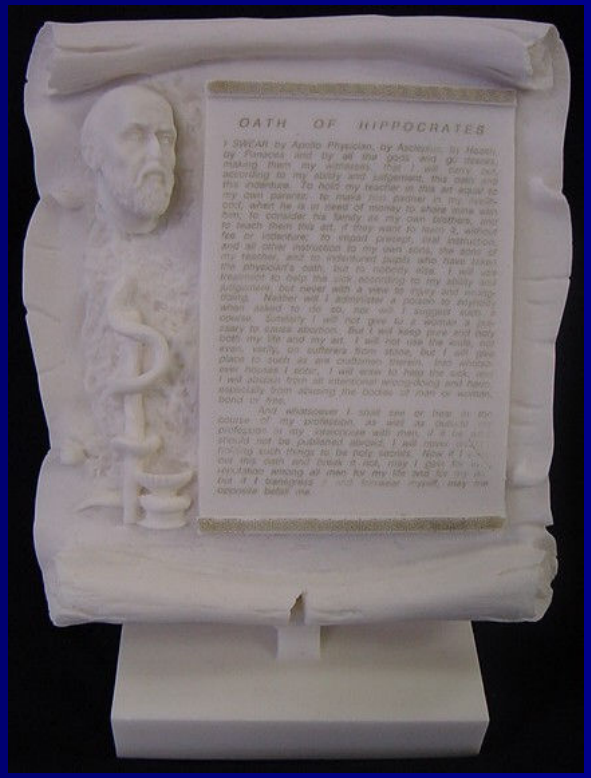
Hippocratic Databases



Presenter: Tyrone Grandison

Team: Rakesh Agrawal, Jerry Kiernan, Ameet Kini, Kristen LeFevre, Ramakrishnan Srikant, Amy Wang, Yirong Xu, Diana Zhou

Our Motivation

New regulations requiring companies to protect personal information

Privacy is a major concern for On-demand businesses

Lack of technology for efficient privacy enforcement and data handling

Loss of revenue

Dilution of brand image

Audit failure

Basics

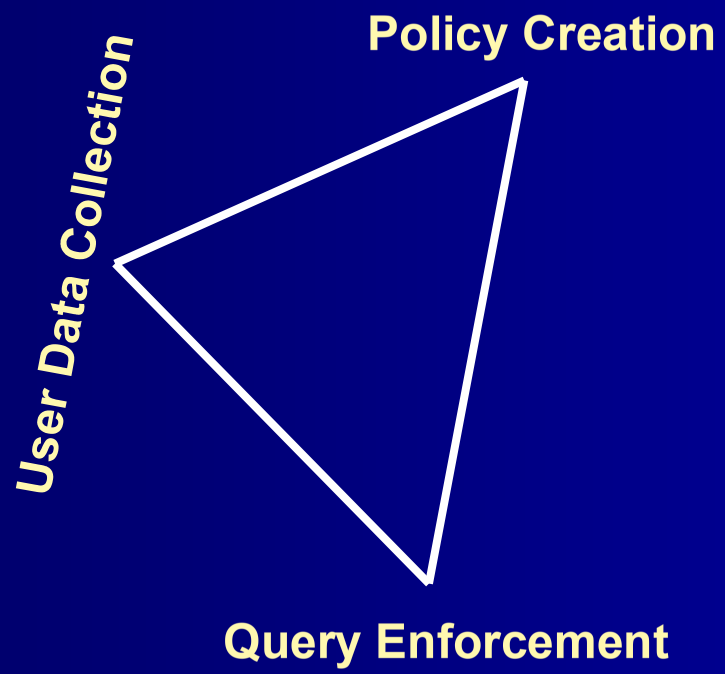
- Founding tenet

  Database systems that take responsibility for the privacy of data they manage, while not impeding the flow of information

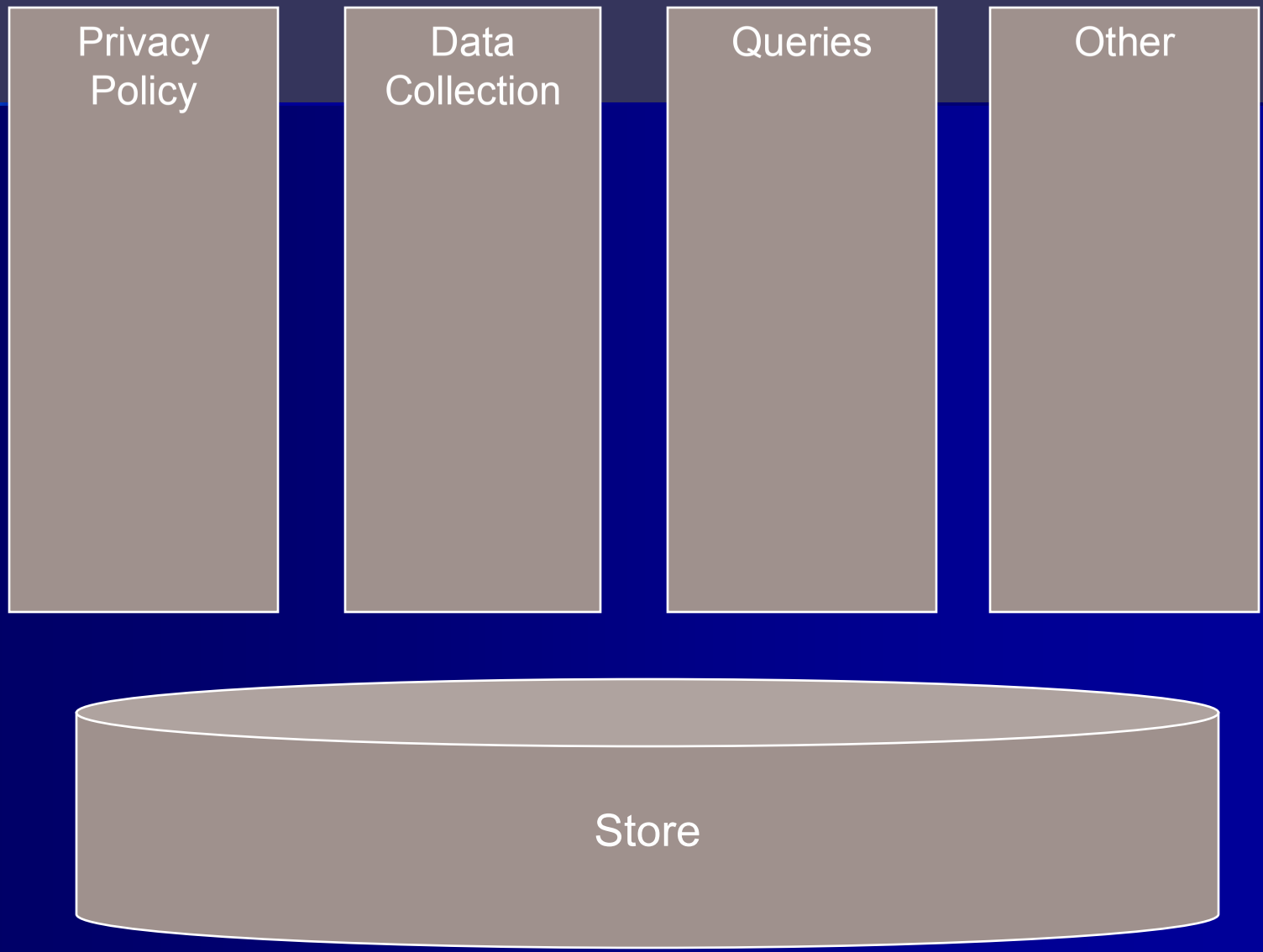
- Principles
  - Collection Group:
    - Purpose Specification, Consent, Limited Collection
  - Use Group:
    - Limited Use, Limited Disclosure, Limited Retention, Accuracy
  - Security & Openness Group:
    - Safety, Openness, Compliance
- Driven by current privacy legislation

  US (FIPA, 1974), Europe (OECD , 1980), Canada (1995), Australia (2000), Japan (2003)

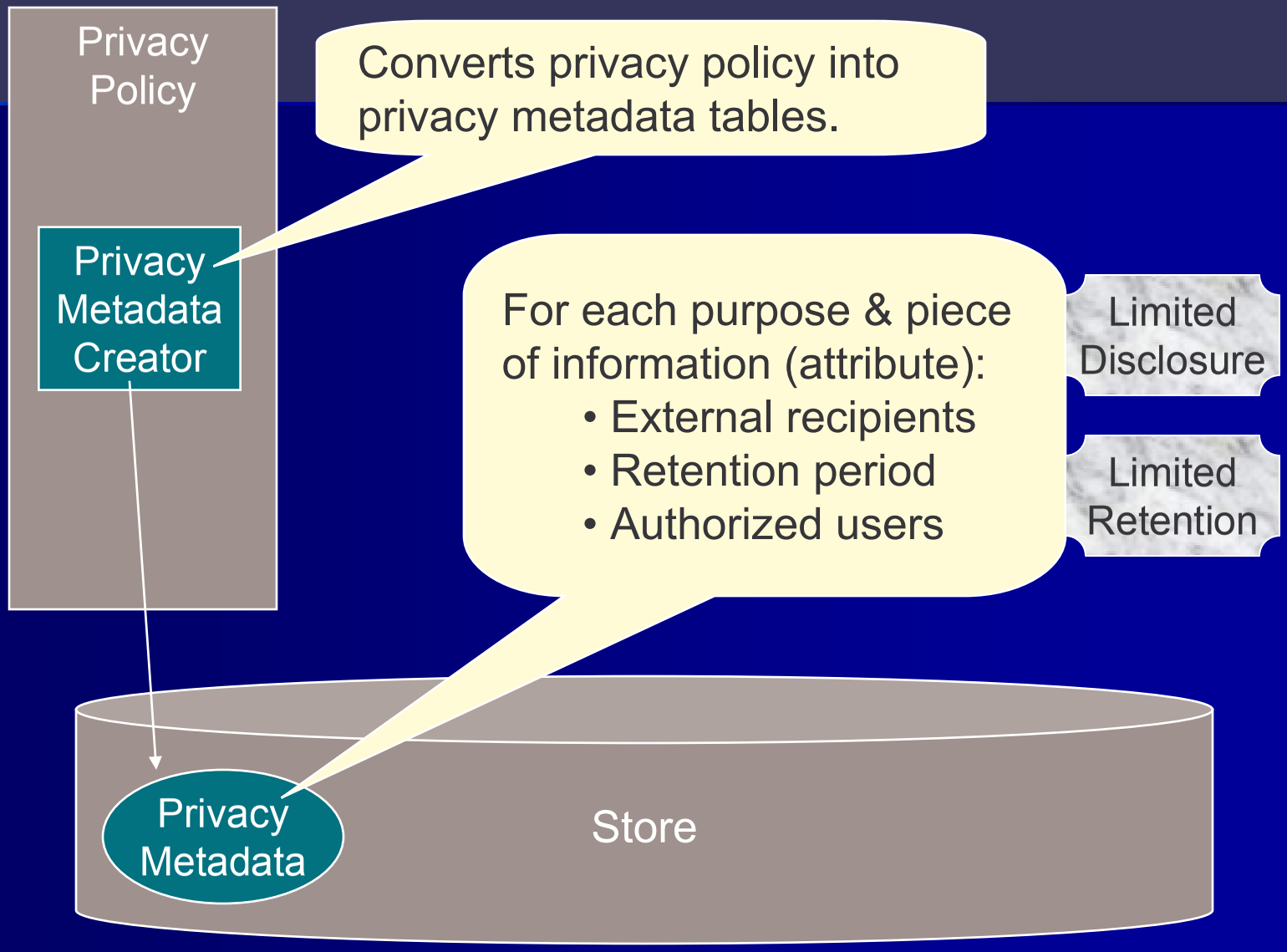
Privacy Enablers: The Triad



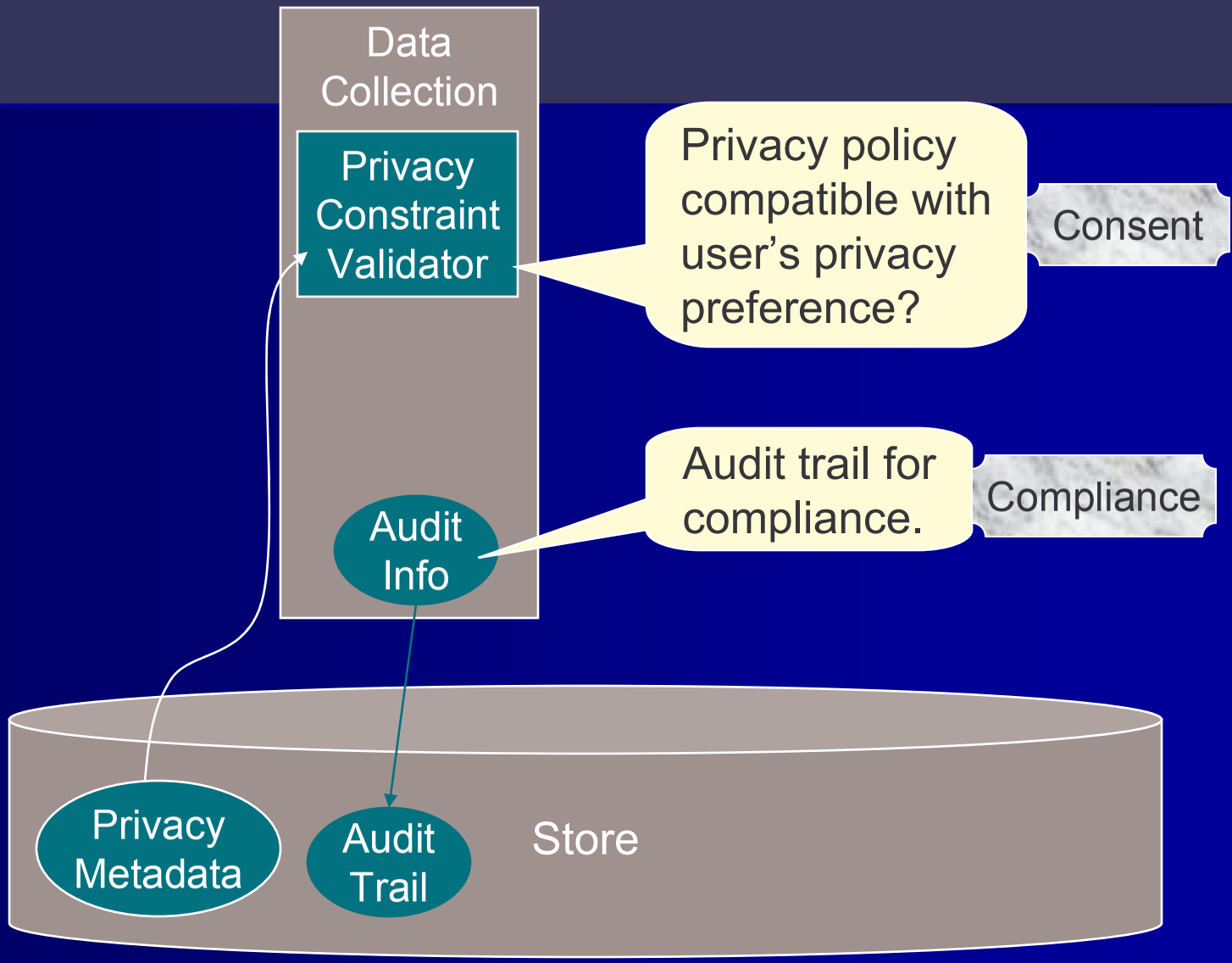
Basic Architecture



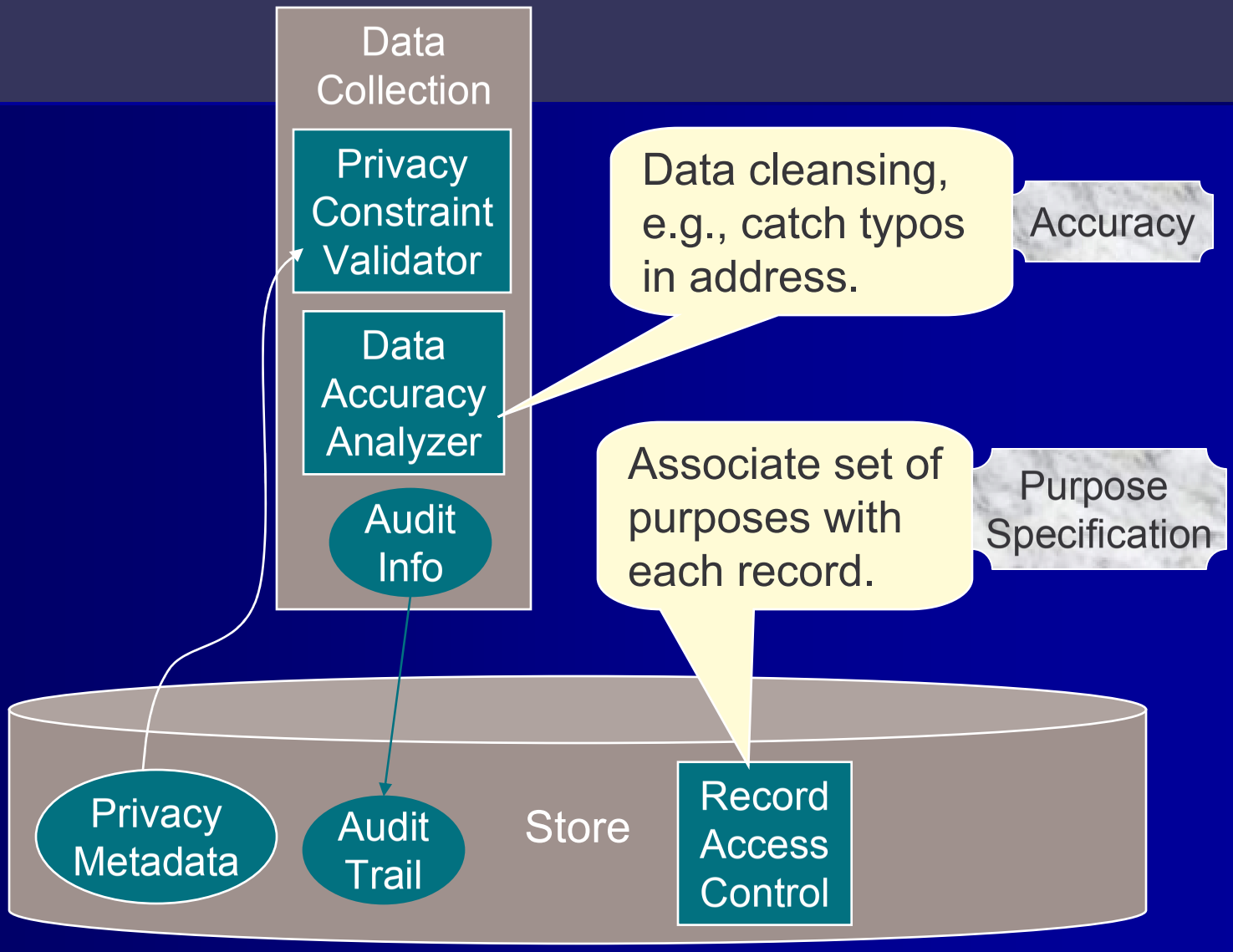
Architecture: Policy



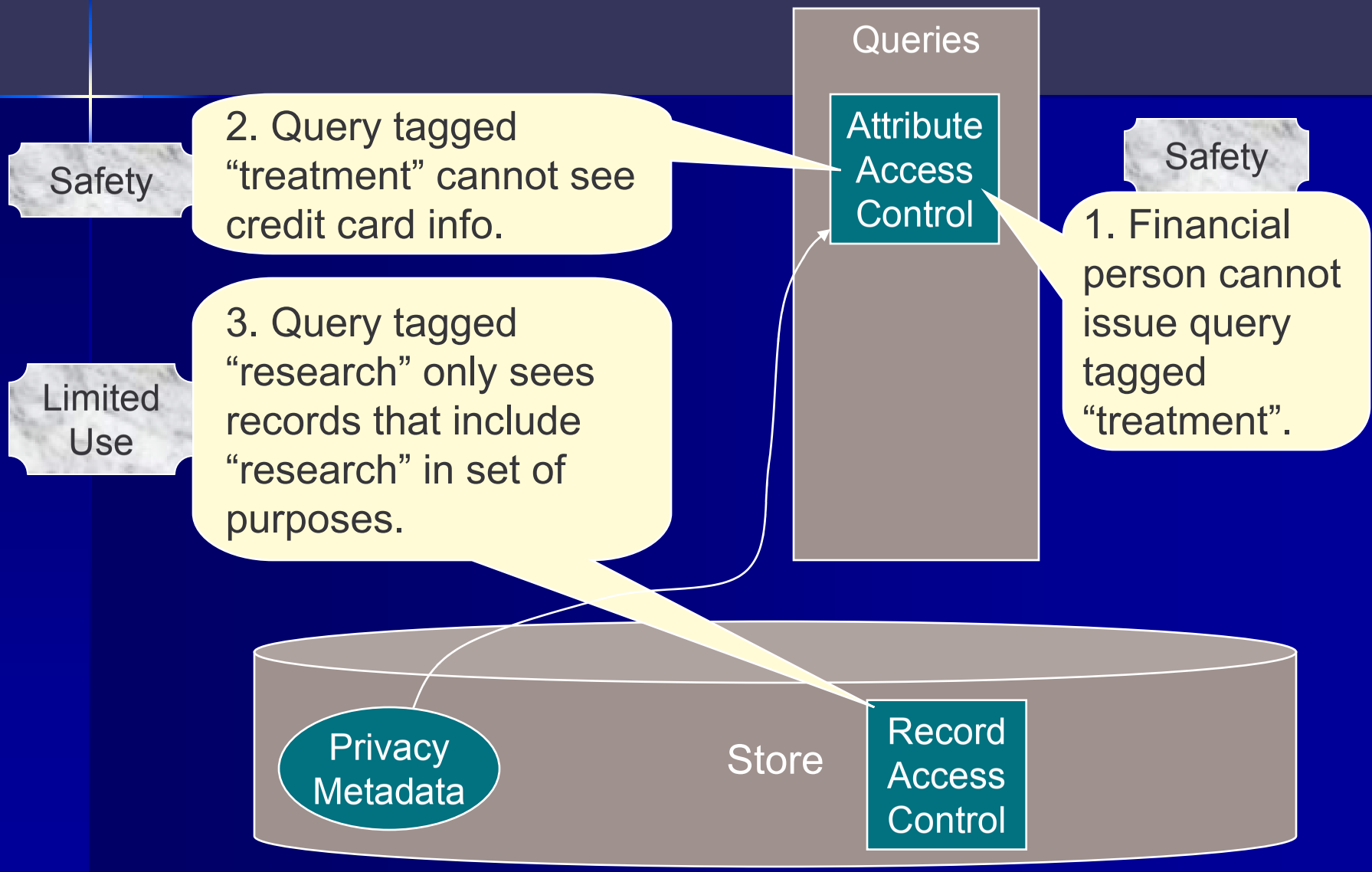
Architecture: Data Collection



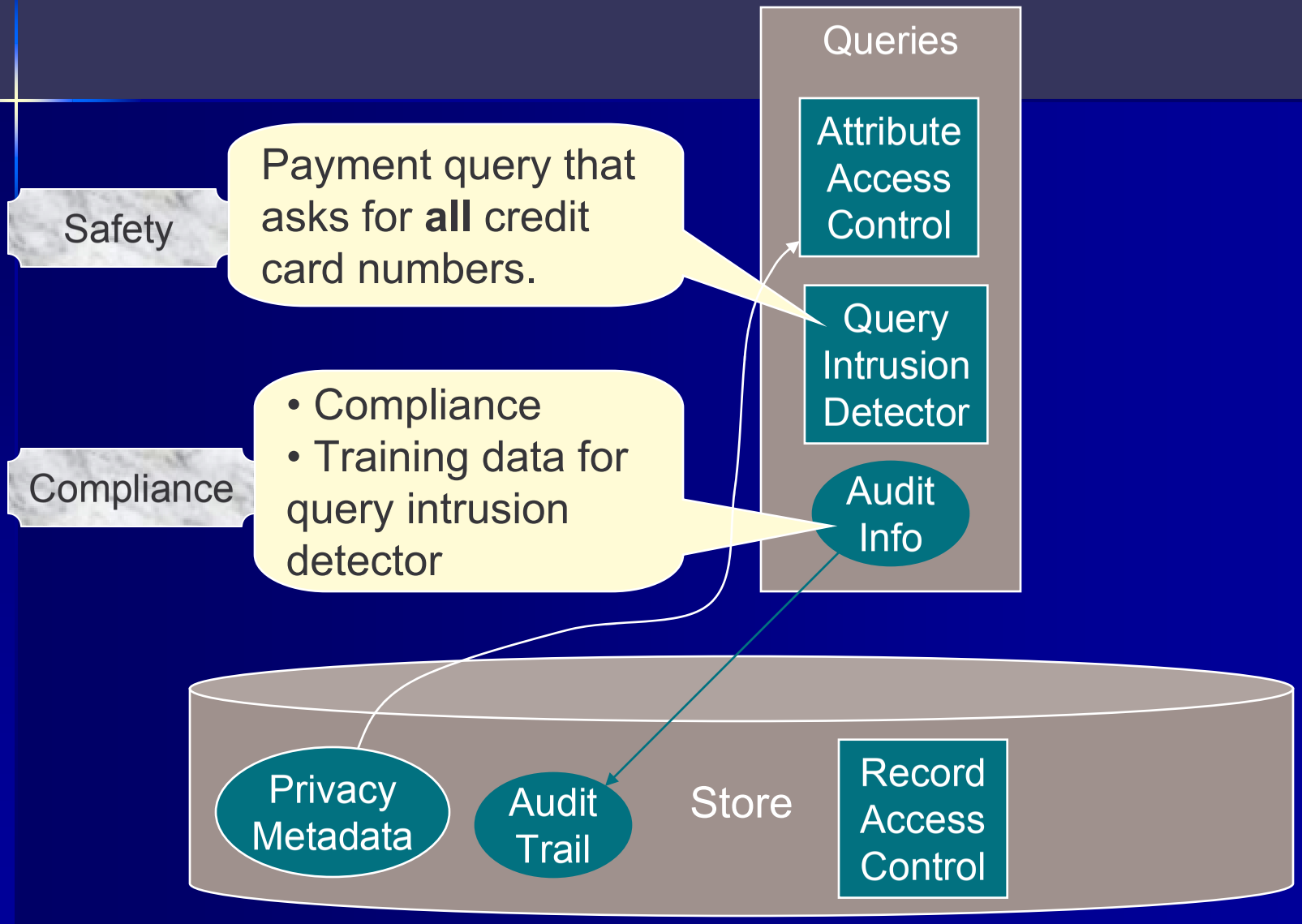
Architecture: Data Collection



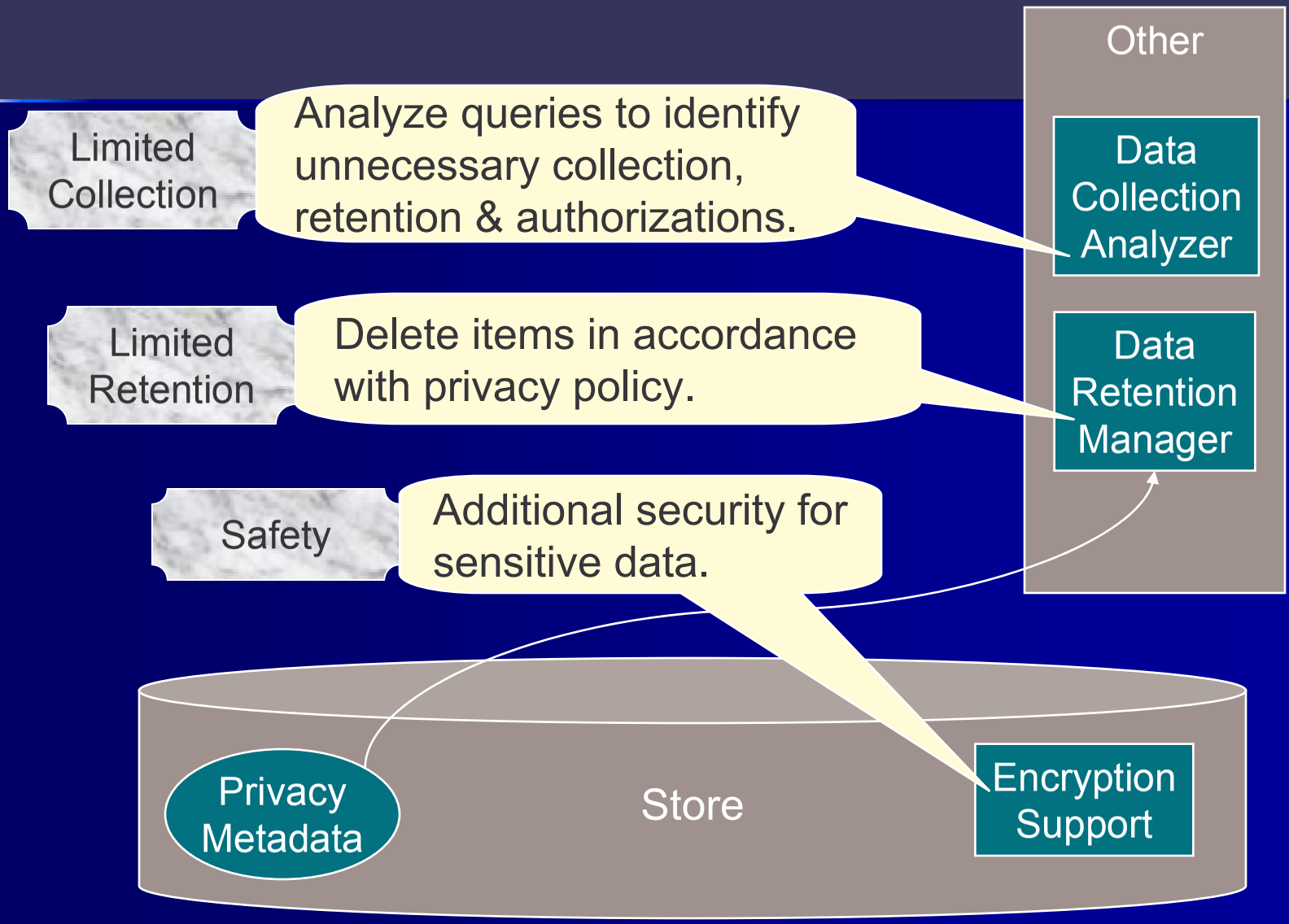
Architecture: Queries



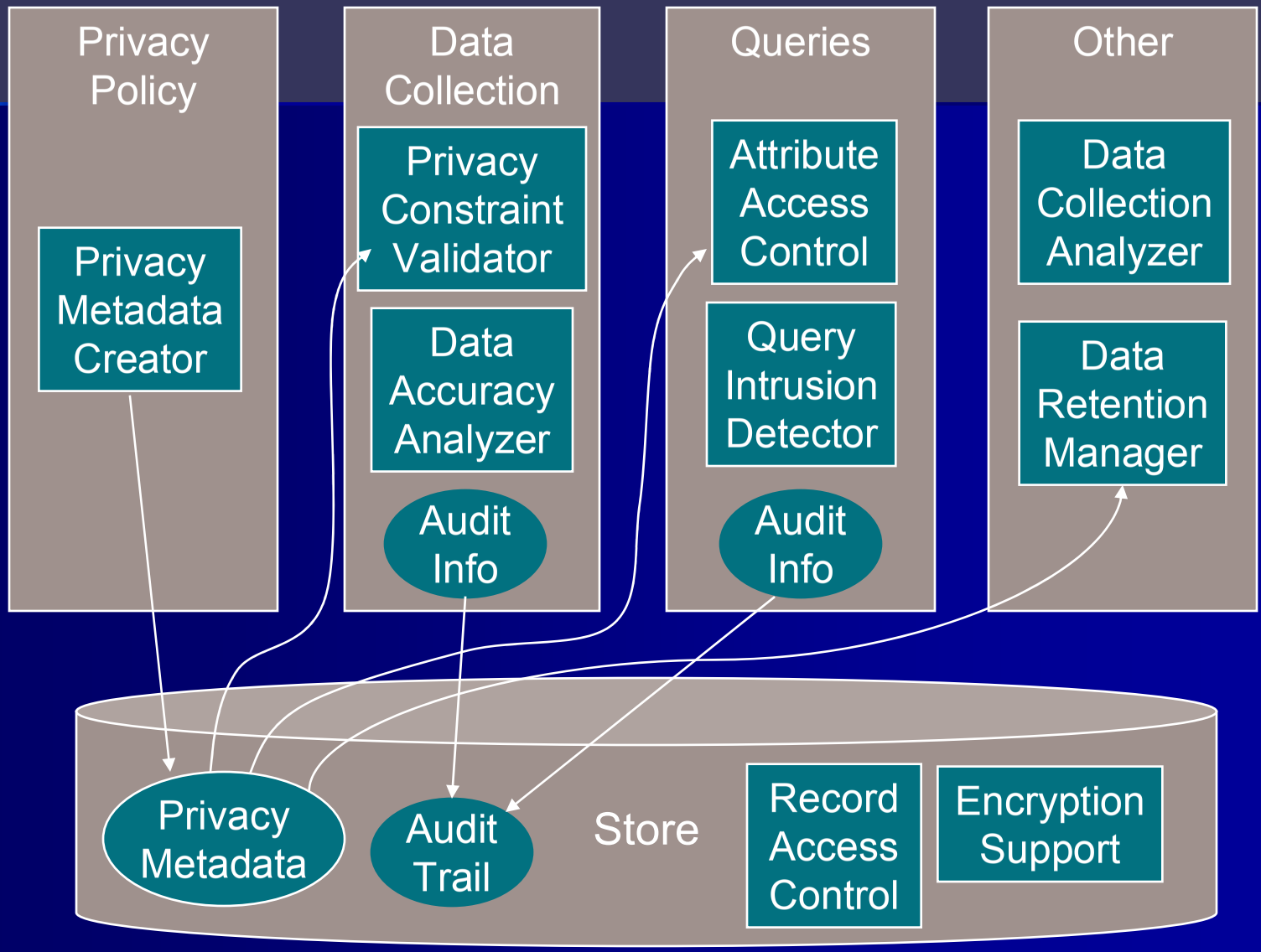
Architecture: Queries



Architecture: Other

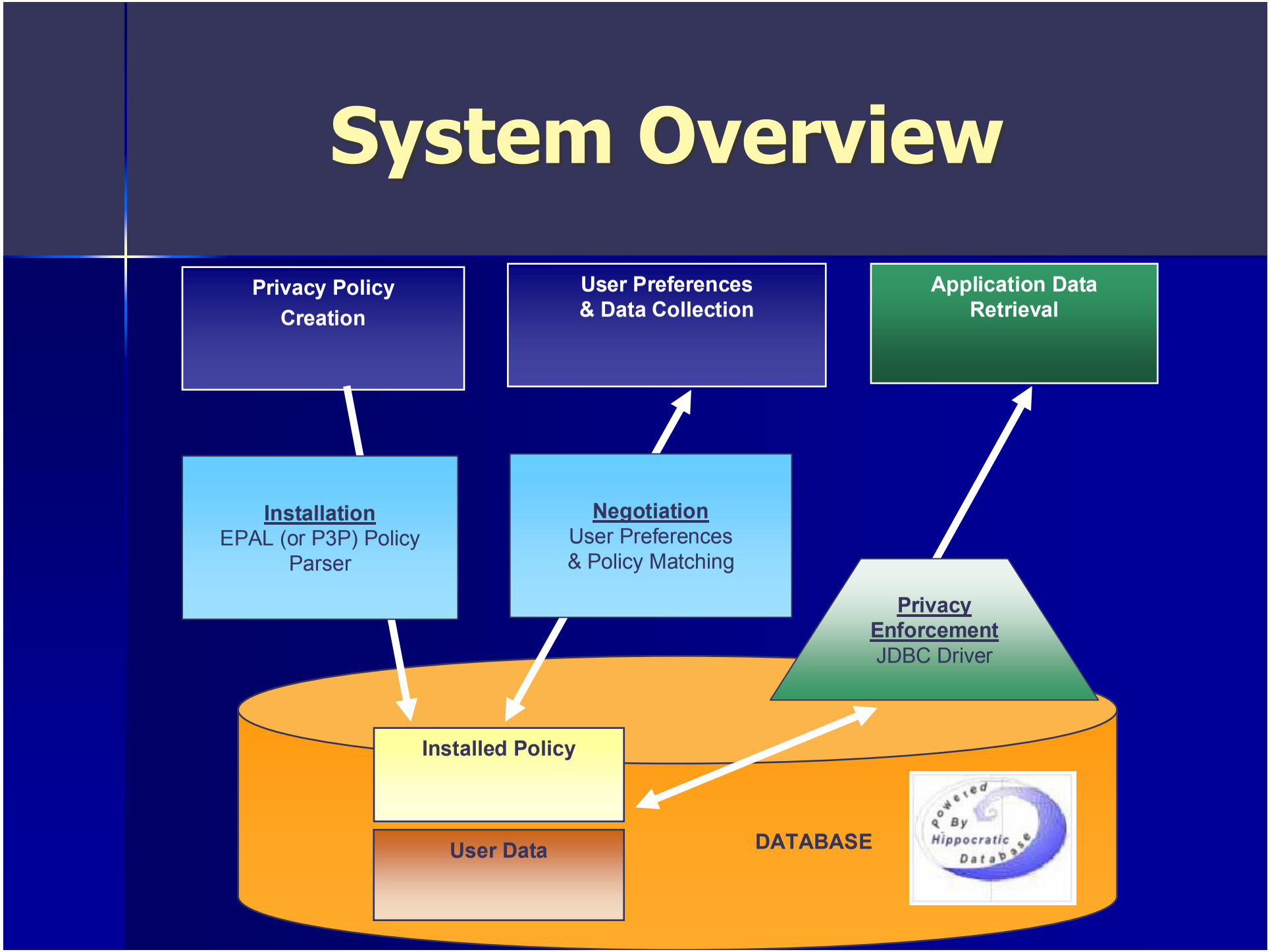


Architecture: All Together



Demo

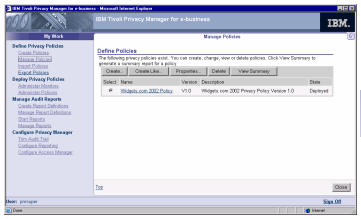
System Overview



Hippocratic Databases
NetCare Healthcare Business Scenario

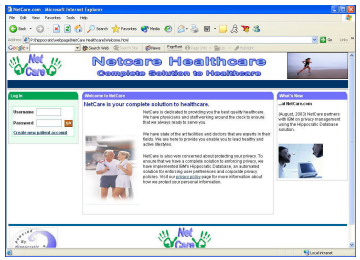
- *John Cane, Chief Privacy Officer, NetCare Healthcare*
- *Jane Smith, New Patient, NetCare Healthcare*
- *Dr. Young, Physician, NetCare Healthcare*
- *Christine Jones, Lab Technician, NetCare Healthcare*
- *Phil Crew, Drug Researcher, Innovative Drug Research*

Hippocratic Databases
NetCare Healthcare Business Scenario



John Cane, CPO installs corporate privacy policy

Jane, a new patient, defines her privacy preferences



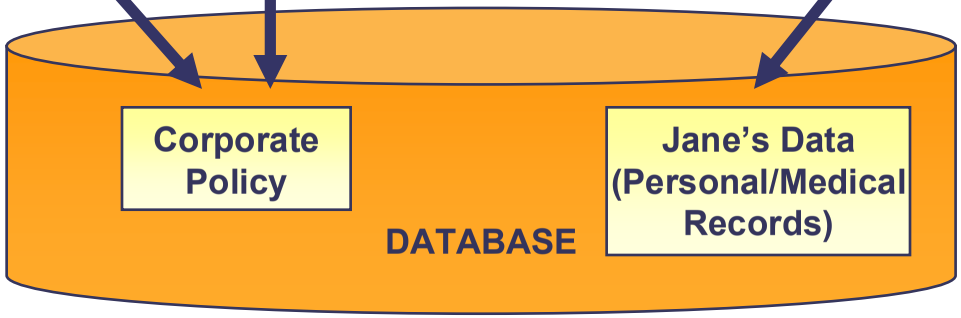
Jane visits NetCare's website to setup patient account

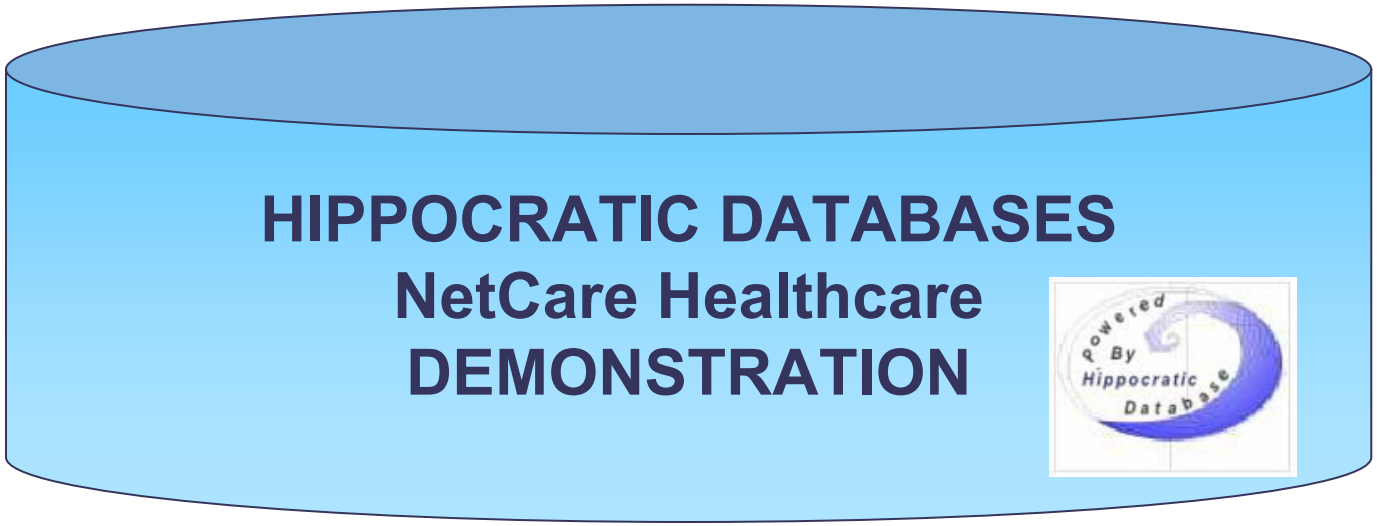
Jane submits her personal information

Installation

Negotiation

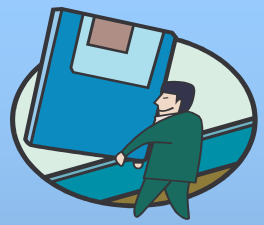
- Name, Address, SSN#, Email
- Opt-in to sharing data for research
- Opt-out of sharing full medical records to lab technicians



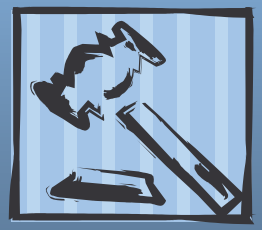

HIPPOCRATIC DATABASES
NetCare Healthcare
DEMONSTRATION
Powered By Hippocratic Database

Conclusion

Increase customer trust and business opportunities



Minimal modification of existing applications



Help mitigate legal risks

More efficient than competing privacy solutions.

Thank You

BACKUP SLIDES

Founding Tenets

Collection Group

1. Purpose Specification
   - For personal information stored in the database, the purposes for which the information has been collected shall be associated with that information.
2. Consent
   - The purposes associated with personal information shall have consent of the donor (person whose information is being stored).
3. Limited Collection
   - The information collected shall be limited to the minimum necessary for accomplishing the specified purposes.

Use Group

4. Limited Use
   - The database shall run only those queries that are consistent with the purposes for which the information has been collected.
5. Limited Disclosure
   - Personal information shall not be communicated outside the database for purposes other than those for which there is consent from the donor of the information.

Use Group (2)

6. Limited Retention
   - Personal information shall be retained only as long as necessary for the fulfillment of the purposes for which it has been collected.
7. Accuracy
   - Personal information stored in the database shall be accurate and up-to-date.

Security & Openness Group

8. Safety
   - Personal information shall be protected by security safeguards against theft and other misappropriations.
9. Openness
   - A donor shall be able to access all information about the donor stored in the database.
10. Compliance
   - A donor shall be able to verify compliance with the above principles. Similarly, the database shall be able to address a challenge concerning compliance.

The Triad

Privacy Policy

- Who can access which types of data for what purposes, e.g.
  - Allows a physician to access patients' names and disease records for treatment purpose
  - Allows a public-affair person to disclose anonymous disease records for research purposes. However patients can opt-out this disclosure.

User Data Collection

- How does the user want personal information to be used? e.g.
  - Allow insurance companies to access my medical data
  - Restrict disclosure of personally identifiable information
- Does the corporate privacy policy conflict with my personal preferences?

Query Enforcement

- Automatically enforce the privacy policy
  - Shred the policy into metadata
  - Analyze queries with respect to the policy, and either
    - Allow the query to run as-is, or
    - Return a subset of the records/cells that reflects individual persons' opt-in or opt-out preferences
    - Block the query if it is in violation of the policy

Query Enforcement through Rewriting

Query Rewrite for Privacy Enforcement

Consider a Simple Example…

ID	NAME	PHONE	SALARY
1	Alice	111-1111	10,000
2	Bob	222-2222	20,000
3	Carl	333-3333	30,000

ID	PhoneChoice
1	0
2	1
3	0

For a certain data accessor/purpose, Name is allowed under the privacy policy, Salary is prohibited, and Phone is allowed on an opt-in basis.

Query Rewrite for Privacy Enforcement

Original Query:

```
SELECT Name, Phone, Salary

FROM Patient
```

Rewritten Query:

```
SELECT Name, Phone, Salary
FROM ( SELECT Name,
            CASE WHEN EXISTS (SELECT 1 FROM Choices
                                                WHERE Choices.PhoneChoice = 1
                                                AND Choices.ID = Patient.ID)
                        THEN Patient.Phone
                         ELSE null
                        END AS Phone,
          CASE WHEN (0 = 1)
                        THEN Patient.Salary
                        ELSE null
                        END AS Salary )
      FROM Patient
      WHERE  ((Name is not null OR Phone is not null) OR Salary is not null) )
```

Query Rewrite for Privacy Enforcement

Results of query…

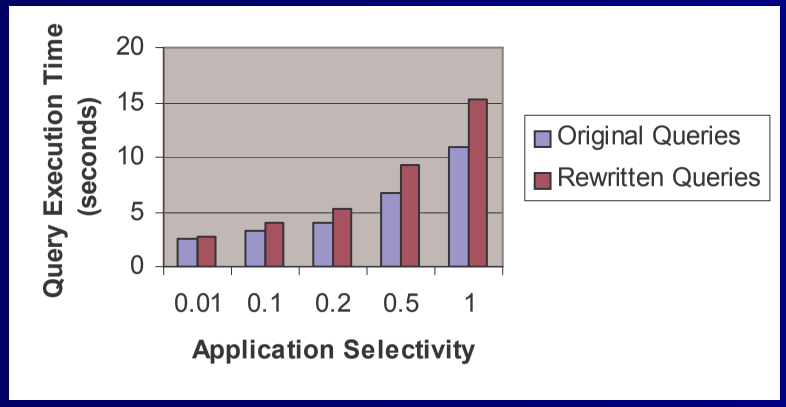
NAME	PHONE	SALARY
Alice	-	-
Bob	222-2222	-
Carl	-	-

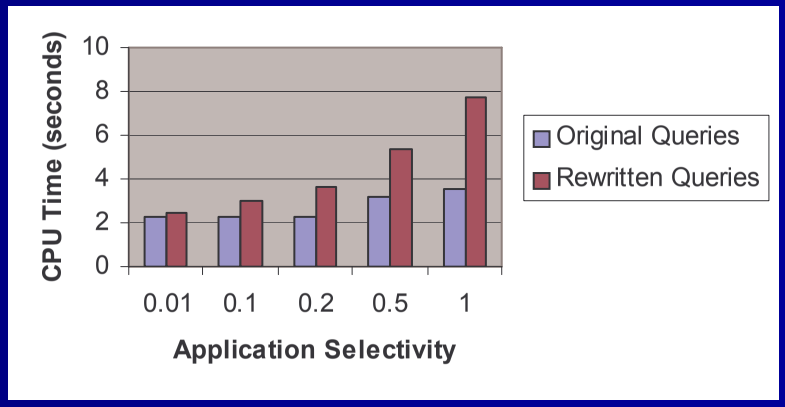
- Forbidden values covered by null values in resulting tables
- Entirely null rows filtered from the result set

Performance Tests

Scenario 1

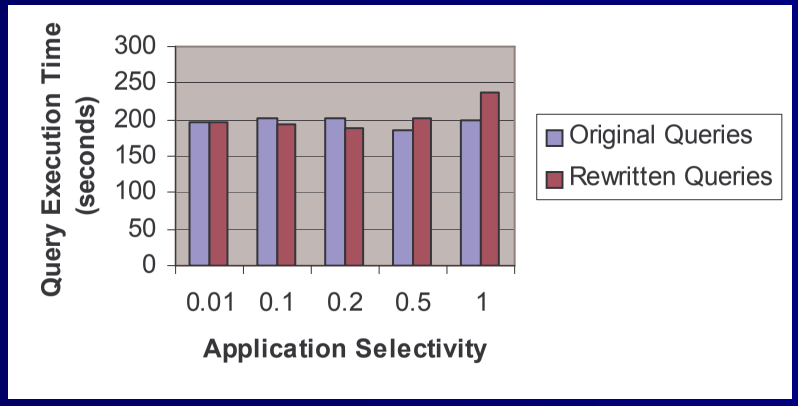
Table Size: 1 million, no index

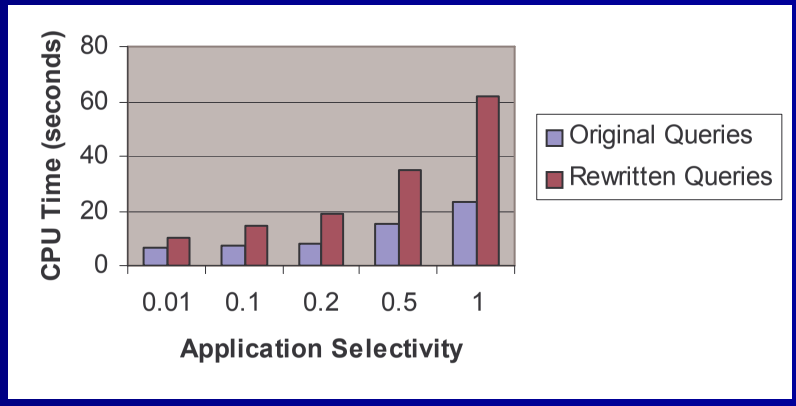




Scenario 2

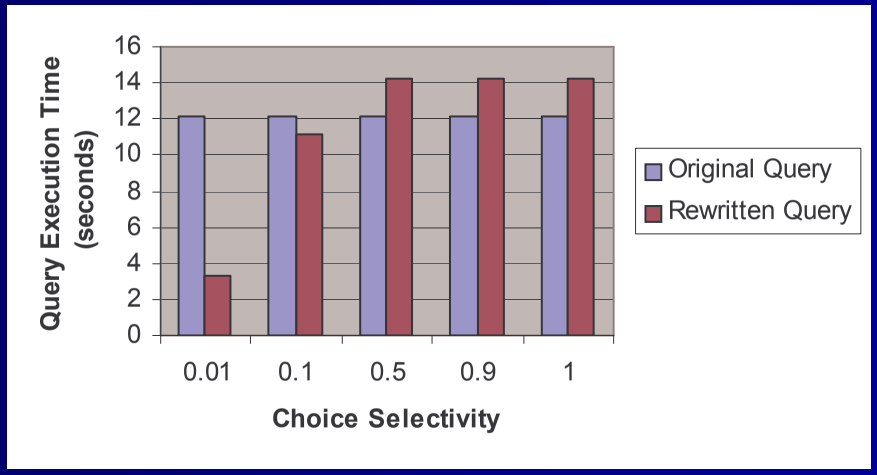
Table Size: 10 million, no index





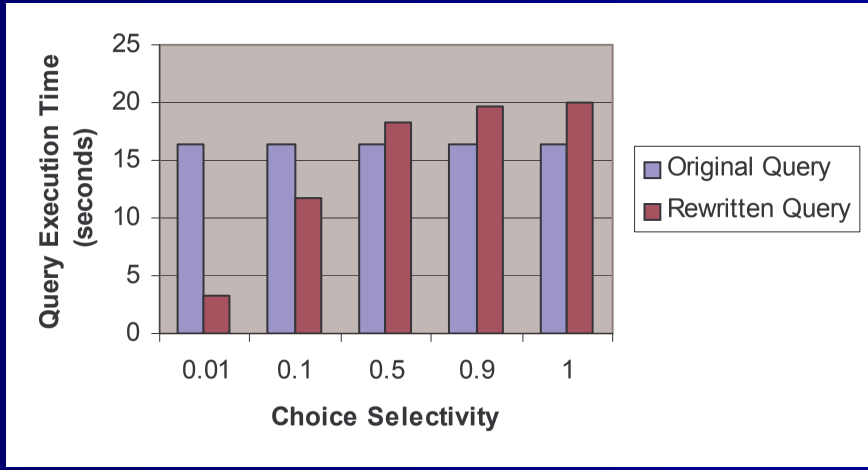
Scenario 3

App Selectivity = .01



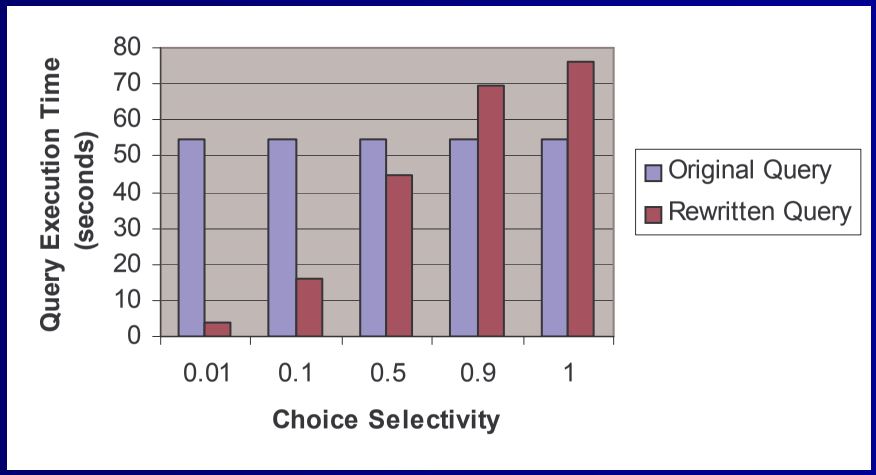
Scenario 4

App Selectivity = .1



Scenario 5

App Selectivity = 1.0



Sources of Information

References

- R. Agrawal, J. Kiernan, R. Srikant, Y. Xu. Order-Preserving Encryption: Opportunities and Limitations. Feb 2003.
- R. Agrawal, R. Srikant. Information Sharing Across Private Databases. ACM Int'l Conf. On Management of Data (SIGMOD), San Diego, California, June 2003.
- R. Agrawal, J. Kiernan, R. Srikant, Y. Xu. An Xpath Based Preference Language for P3P. 12th Int'l World Wide Web Conf. (WWW), Budapest, Hungary, May 2003.
- R. Agrawal, J. Kiernan, R. Srikant, Y. Xu. Implementing P3P Using Database Technology. 19th Int'l Conf.on Data Engineering(ICDE), Bangalore, India, March 2003.
- R. Agrawal, J. Kiernan, R. Srikant, Y. Xu. Server Centric P3P. W3C Workshop on the Future of P3P, Dulles, Virginia, Nov. 2002.
- R. Agrawal, J. Kiernan, R. Srikant, Y. Xu. Hippocratic Databases. 28th Int'l Conf. on Very Large Databases (VLDB), Hong Kong, August 2002.
- R. Agrawal, J. Kiernan. Watermarking Relational Databases. 28th Int'l Conf. on Very Large Databases (VLDB), Hong Kong, August 2002.
- A. Evfimievski, R. Srikant, R. Agrawal, J. Gehrke. Mining Association Rules Over Privacy Preserving Data. 8th Int'l Conf. on Knowledge Discovery in Databases and Data Mining (KDD), Edmonton, Canada, July 2002.
- R. Agrawal, R. Srikant. Privacy Preserving Data Mining. ACM Int'l Conf. On Management of Data (SIGMOD), Dallas, Texas, May 2000.