پایگاه داده های بقراطی

Masood Niazi Torshiz
www.mniazi.ir

فهرست مطالب

- حریم خصوصی(Privacy)
- رویکردهای حفظ حریم خصوصی
- پایگاه داده های بقراطی(Hippocratic Databases)
  - اِعمال فعال (Active Enforcement)
  - بررسی اجابت(Compliance Auditing)

حریم خصوصی

- هر نوع اطلاعات شخصی
- حفظ حریم خصوصی
  - حق افراد برای مشخص کردن اینکه اطلاعات شخصی خودشان چه زمانی، چگونه و به چه میزانی میتواند به دیگران منتقل گردد.
- مشکلات حفظ حریم خصوصی
  - ممکن است قسمتی از حریم خصوصی را برای بدست آوردن آنچه می خواهیم، از دست بدهیم.
  - هنگامی که داده را ارائه دادیم، دیگر کنترلی روی داده هایمان نخواهیم داشت.
  - سوال مهم: آیا داده را می توانیم واگذار کنیم، بدون اینکه برای حریم خصوصی خدشه ای وارد شود؟

رویکردهای مختلف

- رویکردهای مختلف به حفظ حریم خصوصی
  - درشت دانه: حفظ حریم خصوصی بیرون از پایگاه داده ها
    - حفظ حریم خصوصی پس از استخراج داده از پایگاه داده ها در یک برنامه کاربردی
    - IDEMIX
  - ریز دانه: حفظ حریم خصوصی در درون پایگاه داده ها
    - حفظ حریم خصوصی قبل از استخراج داده از پایگاه داده ها در DBMS
    - Hippocratic DBs

پایگاه داده های بقراطی

- Hippocratic Databases
- توسط IBM از سال ۲۰۰۲
- به صورت یک Middleware روی DB2
- ویژگیهای بارز
  - اعمال کنترل دسترسی ریزدانه
  - تبدیل پرس و جوی ورودی به پرس و جوی حافظ حریم خصوصی
  - استفاده از فراداده ها برای حفظ حریم خصوصی

مفهوم کلیدی: هدف

- داده ها برای اهداف خاص جمع آوری می شوند.
- هدف یا اهداف باید به همراه داده ذخیره شوند
- هدف، نحوه استفاده از داده را محدود می کند.

اصول پایه ای در HDB-۱

- توصیف هدف(Purpose Specification)
  - هدف جمع آوری اطلاعات شخصی باید در کنار داده ها نگه داشته شوند
  - مثال: فروشنده کتاب، اطلاعات شخصی را برای آمار خریدها، توصیه کتاب به مشتری و ... نگه می دارد.
- توافق(Consent)
  - در مورد هدف بایستی توافق همه کسانی که مصادیق اطلاعات هستند وجود داشته باشد
  - مثال: خریدار می تواند برای خرید توافق خود را اعلام نماید ولی از سرویس توصیه کردن کتاب اجتناب نماید
- مجموعه محدود(Limited Collection)
  - مینیمم اطلاعات و در حد نیاز باید جمع آوری گردد
  - مثال: خریدار برای استفاده از سرویس توصیه کردن کتاب نیازی به ارائه شماره کارت اعتباری اش ندارد
- استفاده محدود(Limited Use)
  - تنها باید پرس و جوهایی توسط پایگاه داده اجرا گردد که سازگار با هدف باشد
  - مثال: یک پرس و جو برای توصیه کردن کتاب نمی تواند به آدرس تحویل کتاب دسترسی داشته باشد
- افشای محدود(Limited Disclosure)
  - افشای اطلاعات باید تنها در حدی باشد که توافق شده و سازگار با هدف است.
  - مثال: شرکت حامل کتاب نیازی به دانستن شماره کارت اعتباری خریدار ندارد.

اصول پایه ای در HDB-۲

- نگهداری محدود(Limited Retention)
  - اطلاعات باید در تنها تا زمانی نیاز نگه داشته شوند که مورد نیاز می باشند.
  - مثال: زمانی که خرید کتاب انجام شد، دیگر نیازی به نگه داشتن شماره کارت اعتباری نمی باشد.
- دقت(Accuracy)
  - اطلاعات شخصی نگه داشته شده در پایگاه داده ها باید دقیق و به روز باشد .
  - مثال: آدرس ارسال کتاب باید در پایگاه داده ها معتبر و دقیق باشد.
- امنیت(Safety)
  - در مقابل هر نوع تخطی مانند دزدی اطلاعات باید ایمن باشد.
  - مثال: داده ها را باید رمز شده نگه دارد.
- باز بودن(Openness)
  - صاحب اطلاعات باید به تمام اطلاعات ذخیره شده در پایگاه داده ها مربوط به خودش دسترسی داشته باشد
  - مثال:کاربران می توانند به تمام سوابق خرید خود و اطلاعات شخصی خود دسترسی برای دیدن داشته باشند
- ایجاب(Compliance)
  - صاحب اطلاعات باید از اجابت شدن تمام توافقاتش، مطمئن باشد
  - مثال: ثبت تمام دسترسی ها به اطلاعات شخصی بطوریکه زمان و قلم داده مورد دسترسی را بتوان بدست آورد

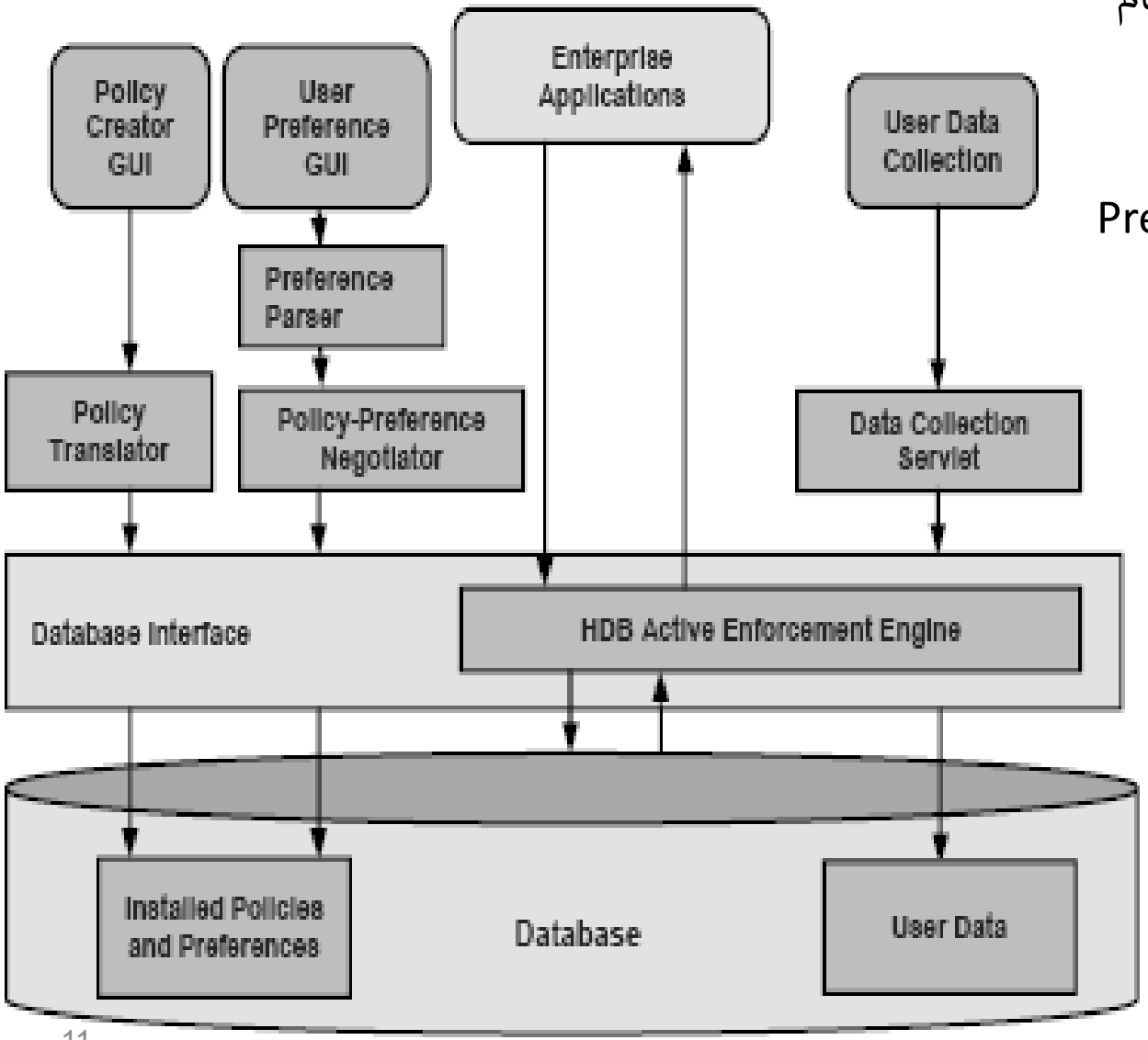
فناوری های Hippocratic DBs

- اِعمال فعال(Active Enforcement)
  - توصیف هدف(Purpose Specification)
  - توافق(Consent)
  - استفاده محدود(Limited Use)
  - افشای محدود(Limited Disclosure)
- بررسی اجابت(Compliance Auditing)
  - ایجاب(Compliance)

سیستم اِعمال فعالانه-۱

- Active Enforcement System
- اولین گام برای پیاده سازی HDB
- سیستمی است که دسترسی به و افشای اطلاعات شخصی را با استفاده از خط مشی های حریم خصوصی ریزدانه، قوانین قابل اعمال، و ترجیحات افراد محدود می کند.
- خط مشی های حریم خصوصی سازمان و ترجیحات کاربران را در جداول پایگاه داده ها نگه می دارد
- با گرفتن پرس و جوی کاربرآن را به پرس و جوهایی مطابق با خط مشی های حریم خصوصی و ترجیحات کاربران تبدیل، و تضمین می کند افراد دارای مجوز به داده های مجاز برای اهداف مناسب دسترسی پیدا کنند.

معماری HDB AE



- پیاده سازی کنونی در سه گام این کار انجام می دهد:
  - Policy Creation
  - Preference Negotiation
  - Application Data Retrieval

سیستم اِعمال فعالانه-۳

- Policy Creation:
  - یک سازمان حافظ حریم خصوصی باید خط مشی های حریم خصوصی خود را مشخص نماید
  - سازمان ها این خط مشی ها را بوسیله یک زبان حریم خصوصی و از طریق یک واسط کاربری توصیف خط مشی بیان می کنند.
  - خطمشی ها توسط AE پویش شده و در پایگاه داده ها به عنوان فراداده نگه داشته می شوند.
- Preference Negotiation:
  - افراد از خط مشی های حریم خصوصی سازمان در این گام مطلع می شوند.
  - افراد ترجیحات حریم خصوصی خود را تنظیم و آنها را بوسیله یک زبان توصیف ترجیحات در سمت مشتری بیان می کنند.
  - قبل از هر گونه افشای اطلاعات شخصی، سیستم این ترجیحات را با خط مشی های سازمان منطق ساخته و افراد را از هرگونه تناقض و ناسازگاری مطلع می کند

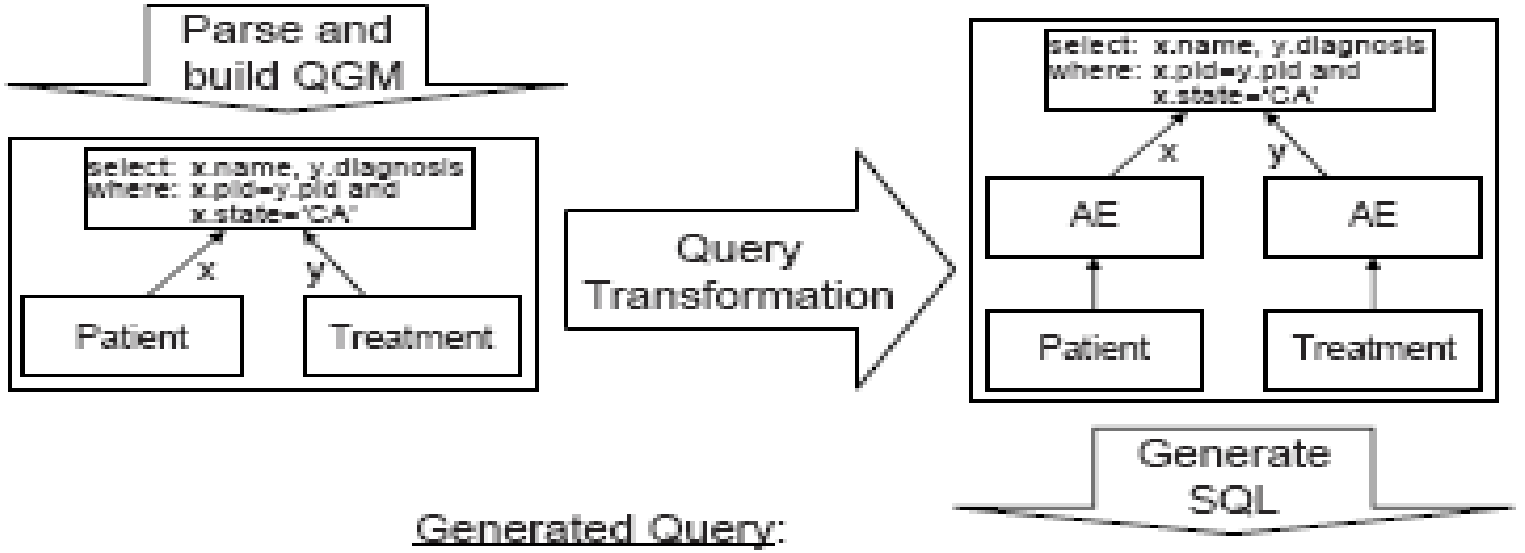
سیستم اِعمال فعالانه-۴

- Application Data Retrieval:
  - پرس و جوهای ورودی کاربر برای برآورده شدن خط مشی های حریم خصوصی ترجمه می شوند.
  - پایگاه داده ها آنرا اجرا کرده و تنها اطلاعاتی را بر می گرداند که مطابق با خط مشی ها باشد.
  - به این ترتیب سیستم به صورت کاملاً شفاف، کنترل افشای اطلاعات را در سطح سلول، بر اساس مجازشناسی درخواست کننده، هدف دسترسی و دریافت کننده نهایی اطلاعات و همین طور ترجیحات افراد اِعمال می کند.

ترجمه پرس و جو در AE HDB

Application Query:

```
select    x.name, y.diagnosis
from      Patient x, Treatment y
where     x.pid=y.pid and x.state='CA'
```



Generated Query:

```
select    x.name, y.diagnosis
from      Patient x,Treatment y
where     x.pid=y.pid and x.state='CA' and
          exists (   select * from Patient_choice c
                     where x.pid=c.pid and
                     c.choice='research' and c.value='opt-in')
```

مثالی از AE HDB

شمای فراداده حریم خصوصی

table	attributes
privacy-policies	purpose, table, attribute, { external-recipients }, retention
privacy-authorizations	purpose, table, attribute, { authorized-users }

شمای پایگاه داده

table	attributes
customer	purpose, customer-id, name, shipping-address, email, credit-card-info
order	purpose, customer-id, transaction-id, book-info, status

مثالی از AE HDB-۲

جدول خط مشی های حریم خصوصی

purpose	table	attribute	external-recipients	retention
purchase	customer	name	{ delivery-company, credit-card-company }	1 month
purchase	customer	shipping-address	{ delivery-company }	1 month
purchase	customer	email	*empty*	1 month
purchase	customer	credit-card-info	{ credit-card-company}	1 month
purchase	order	book-info	*empty*	1 month
registration	customer	name	*empty*	3 years
registration	customer	shipping-address	*empty*	3 years
registration	customer	email	*empty*	3 years
recommendations	order	book-info	*empty*	10 years
purchase-circles	customer	shipping-address	*empty*	1 year
purchase-circles	order	book-info	{ aggregated-all }	1 year

مثالی از ۳-AE HDB

جدول مجازشماریهای های حریم خصوصی

purpose	table	attribute	authorized-users
purchase	customer	customer-id	*all*
purchase	customer	name	{ shipping, charge, customer-service }
purchase	customer	shipping-address	{ shipping }
purchase	customer	email	{ shipping, customer-service }
purchase	customer	credit-card-info	{ charge }
purchase	order	customer-id	*all*
purchase	order	transaction-id	*all*
purchase	order	book-info	{ shipping }
purchase	order	status	{ shipping, customer-service }
registration	customer	customer-id	*all*
registration	customer	name	{ registration, customer-service }
registration	customer	shipping-address	{ registration }
registration	customer	email	{ registration, customer-service }
recommendations	order	customer-id	{ mining }
recommendations	order	transaction-id	{ mining }
recommendations	order	book-info	{ mining }
purchase-circles	customer	customer-id	{ olap }
purchase-circles	customer	shipping-address	{ olap }
purchase-circles	order	customer-id	{ olap }
purchase-circles	order	book-info	{ olap }

۴-AE HDB مثالی از

مثالی از یک پرس و جو

User: Customer Service
Query Purpose Tag: purchase

Select status
From order-info as oi, personal-info as p, order as o
Where email='bob@yale.edu' and order-number=12345 and p.customer-id=o.customer-id and o.order-id=oi.order-id

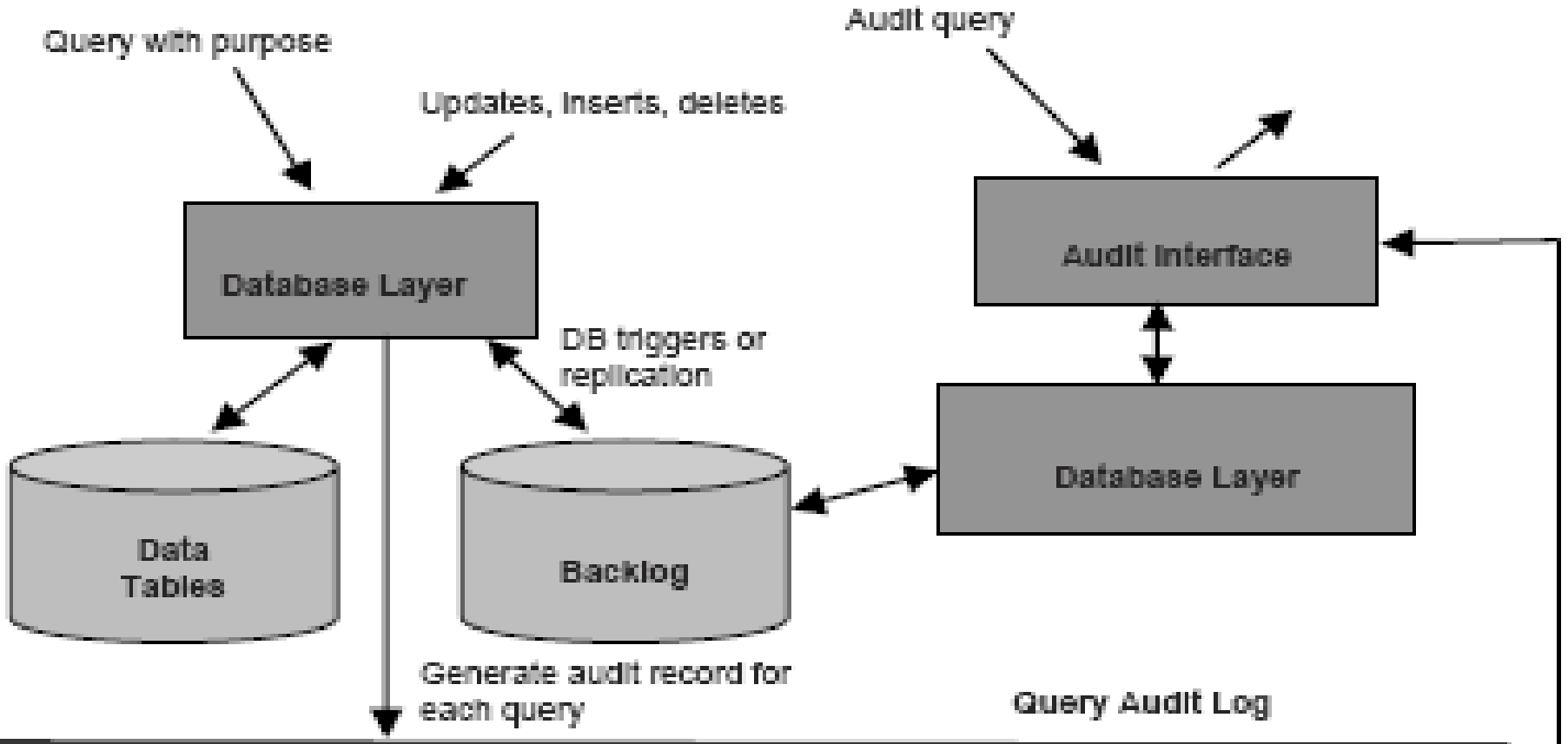
سیستم بررسی اجابت-۱

- Compliance Audit System
- سیستم CA این امکان را می دهد که افشاهای اطلاعات قبلی قابل ردگیری باشد و بتوانیم افشاهای مشکوک را بیابیم
- این جزء به سازمانها این امکان را می دهد که تاریخ و زمان هر پرس و جو، هدف دسترسی، دریافت کننده نهایی اطلاعات، و اطلاعات دقیق در مورد قلم داده هایی که افشا شده اند را استخراج نمایند.
- به این ترتیب این جزء می تواند اصل Compliance(ایجاب) را برآورده نماید.

سیستم بررسی اجابت-۲

- دو نوع Log خواهیم داشت:
  - Query Log: رشته پرس و جو و اطلاعات زمینه ای دیگر مرتبط با آن را نگه می دارد.
    - اطلاعات زمینه ای: شناسه، زمان، هدف، و گیرنده
  - Back Log: تمام updateها، insertها، deleteها روی جداول مبدا را در backlog نگه می دارد.
    - این کار با استفاده از فناوری های موجود مانند trigger و دیگر ویژگیهای replication قابل انجام است.
- این دو نوع جدول کافی هستند که افشاهای گذشته را با ساختن وضعیت پایگاه داده ها استخراج نماییم.

CA HDB معماری



ID	Timestamp	Query	User	Purpose	Recipient
1	2006-02...	Select ...	Dr. Jones	Treatment	Dr. Jones
2	2006-02...	Select ...	Dr. Roberts	Treatment	Dr. Roberts

سیستم بررسی اجابت-۴

- برای بازرسی یک زبان مانندSQL تعریف شده است :

```
audit <audit-list>
from <audited-tables>
where <condition>
```

- در هنگام بازرسی، HDB یک آنالیز ایستا بروی پرس و جو ها (Query Log) انجام می دهد تا بتواند بعضی پرس و جوهای کاندیدا برای تحلیل بیشتر تولید نماید
  - پرس و جو های کاندیدا مشکوک خواهند بود اگر یک تاپل حتمی(Indispensable Tuple) مشترک با عبارت بازرسی داشته باشند.
- سیستم با ترکیب این پرس و جو ها به یک پرس و جوی بازرسی واحد که از نوع SQL است، می رسد که روی جداول Back Log اجرا می شود.
- به این ترتیب اطلاعات موجود در مورد افشاء شده ها را استخراج خواهیم کرد

سیستم بررسی اجابت-۵

- سناریوی بازرسی

I. پس از دادن یک عبارت بازرسی به عنوان ورودی سیستم CA، یک آنالیز ایستا برای جداسازی مجموعه ای از پرس و جو های کاندیدا انجام می شود.

- پرس و جو های کاندیدا(Candidate Queries): تمام پرس و جو هایی هستند که به تمام فیلدهای مشخص شده در audit-list دسترسی پیدا کرده اند.

II. سیستم پرس و جو های کاندیدا را با عبارت بازرسی ترکیب کرده و پرس و جو های مشکوک(Suspicious Queries) را که ممکن است به اطلاعات audit-list دسترسی پیدا کرده باشند، استخراج می کند.

- پرس و جو های مشکوک(Suspicious Queries): پرس و جو هایی که یک تاپل حتمی(Indispensable Tuple) را با عبارت بازرسی مشترک دارند
- تاپل حتمی(Indispensable Tuple): تاپل t از جدول T را حتمی می نامیم اگر یک پرس و جوی Select-Project-Join Q روی پایگاه داده ها داشته باشیم که نتایج Q روی پایگاه داده ها با زمانی که t را از T حذف کنیم، یکسان نباشد.

سیستم بررسی اجابت-۶

- سناریوی بازرسی (ادامه)

III. سیستم CA، پرس و جوهای مشکوک را با هم ترکیب کرده و تشکیل یک پرس و جوی بازرسی واحد را داده و آنرا روی جداول Back Log اجرا می کند. به این ترتیب، اطلاعات دقیق دسترسی یافته بوسیله این پرس و جو را می توان استخراج نماید.

IV. در نهایت، به عنوان نتایج بازرسی، پرس و جو های دسترسی یافته به اطلاعات مشخص شده در عبارت بازرسی را می توان به همراه تاریخ و زمان پرس و جو، و هدف پرس و جو، استخراج کرد.

سیستم بررسی اجابت-۷

Application Query #11 @ T15:

```
select   x.name, y.diagnosis
from     Patient x, Treatment y
where    x.pid=y.pid and x.state='CA'
```

Audit Expression:

```
audit    y.diagnosis
from     Patient x, Treatment y
where    x.pid=y.pid and x.name = 'Adam'
```