Jacqueline

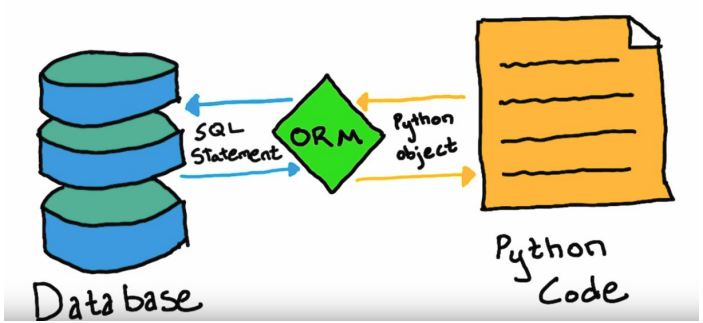
Precise, Dynamic Information Flow for Database-Backed Systems

Jacqueline is built on top of an ORM
What is an ORM?

Written in a object oriented language, some examples: Django (python), CakePHP (PHP), Apache Cayenne (Java)

Allows us to make SQL queries using the object oriented paradigm: Python, PHP, Java, etc.

Maps python object to SQL tables



Jacqueline: similar to Resin
What are the differences?

Jacqueline

- Built on top of ORM
- Policy agnostic
  - What does this mean?
- Separates the policy and execution
  - Executed at computation sinks
  - What does this mean?

Resin

- Assertions in the code
- Policy is enforced at execution

What are the implications of these differences in practice?

What does it Mean to be Policy Agnostic?

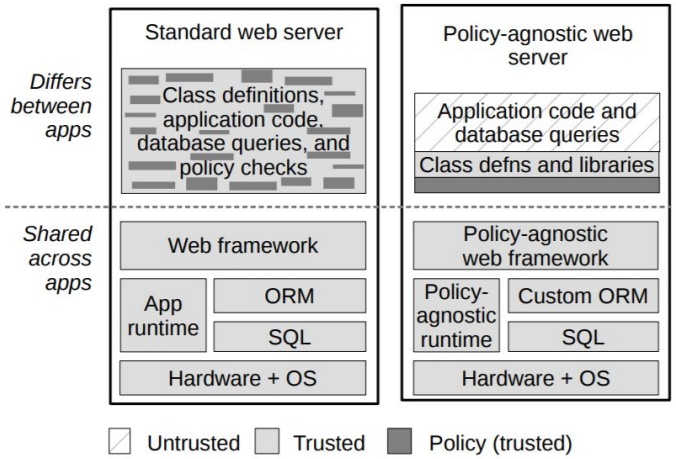


Figure 1. Application architecture in a standard web server compared to a policy-agnostic web server.

Who do we trust in a policy agnostic web server?

Who do we not trust?

How are the components different between a policy-agnostic web server and a standard one?

Faceted Data: What is it?

Faceted data sits in a table filled with variables that describe what this data is in various contexts

At the computation sink, the faceted data is evaluated to be public or private (SAT problem)

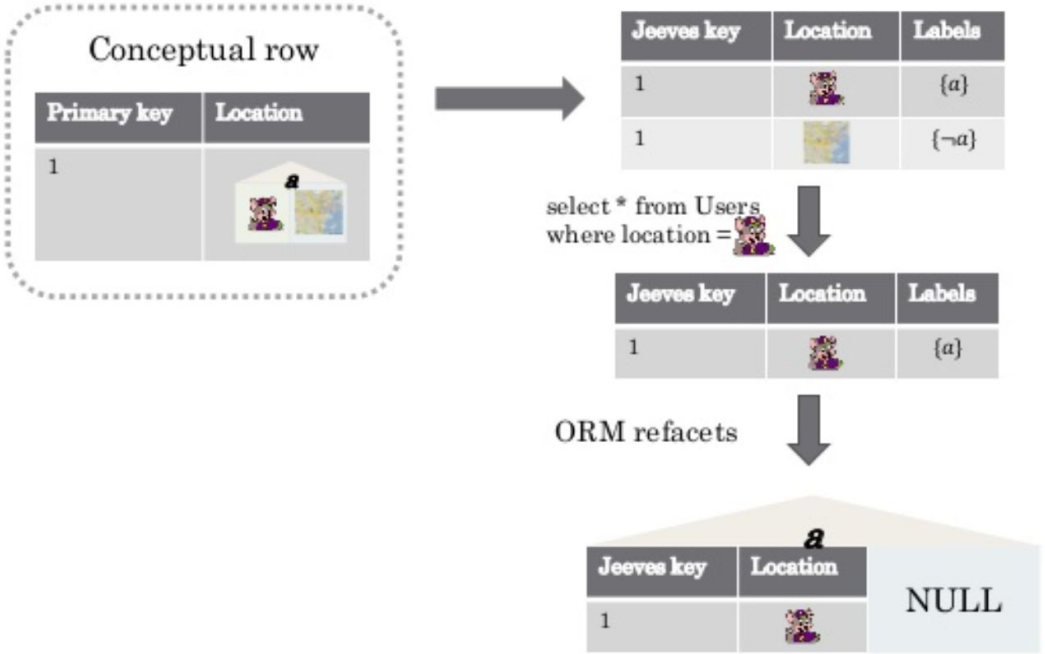
Say we are throwing a party, but we only want those invited to know where it is:

Id	location	jid	labels
1	Chuck E cheese	1	{a}
2	Providence County	1	{not a}

Python Semantics

where: a = (user in guest list)

The Faceted Data in Faceted Rows



Why is the ORM important?
It maps the facets!

Location can evaluate to either “Chuck E. Cheese” or to “Providence county” depending on the value of a

This seems like it takes up a lot of space How can we make it smaller? Pruning!

Intuition: If we are given a boolean expression, we can simplify it to a simpler, equivalent expression

This means after pruning, given the viewing context, (the user) the policy relevant state will not change before output

So we can do less and get the same outcome!

But can we always prune?

Stress test: How does Jacqueline affect performance?

Courses	Time w/o pruning	Time w/ pruning
4	0.377s	0.185s
8	64.024s	0.192s
16	–	0.248s
32	–	0.337s
64	–	0.522s
128	–	0.886s
256	–	1.630s
512	–	3.691s
1024	–	6.233s

Exponential!

Better...

ouch

Table 5. Showing all courses, with and without Early Pruning.

What do we think?
Do you find this acceptable?

Time to view single paper			Time to view single user		
Papers	Jacq.	Django	Users	Jacq.	Django
8	0.160s	0.177s	8	0.164s	0.158s
16	0.165s	0.175s	16	0.164s	0.159s
32	0.160s	0.177s	32	0.164s	0.159s
64	0.159s	0.173s	64	0.164s	0.159s
128	0.160s	0.173s	128	0.167s	0.158s
256	0.159s	0.173s	256	0.163s	0.159s
512	0.159s	0.178s	512	0.169s	0.162s
1024	0.161s	0.173s	1024	0.163s	0.159s

Table 4. Times to view profiles for a single paper and single user, in Jacqueline and Django.

Time to view all papers			Time to view all users		
# P	Jacq.	Django	# U	Jacq.	Django
8	0.241s	0.201s	8	0.172s	0.163s
16	0.299s	0.241s	16	0.249s	0.234s
32	0.542s	0.388s	32	0.279s	0.254s
64	0.855s	0.554s	64	0.358s	0.341s
128	1.551s	0.931s	128	0.510s	0.541s
256	2.810s	1.633s	256	0.769s	0.820s
512	5.717s	3.265s	512	1.352s	1.269s
1024	10.729s	6.055s	1024	2.305s	1.538s

Table 3. Times to view a list of summary information for conference manager stress tests, in Jacqueline and Django.

Discussion

- Say you were designing something like Canvas, would you use Resin or Jacqueline? Why?
- Is this more of a proof of concept such that future languages would incorporate these features? Or the beginnings of a potential proposal to actually add them to Python/Django? Or is this not the goal at all and it's more pure research? -Sam
- What role should a system like Resin or Jacqueline play in systems design today? Should they be common-place or only used in sensitive data contexts? Are frameworks like these necessary in order to be compliant in a GDPR world?
- Could this lead to a lot of overconstraint? If we were to use Jacqueline, do we still need checks? Would that be too much redundancy? Is it possible to have an approach that does both static as well as dynamic information flow control so as to eliminate such problem? -Zhiyuan
- How flexible is this framework to potential needs to change the schema? -Nam