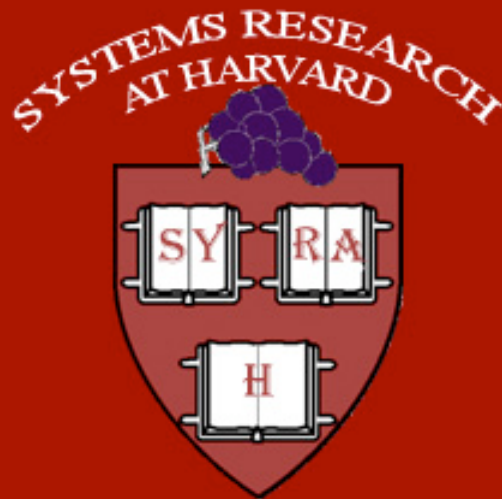


PASS

Provenance-Aware Storage System

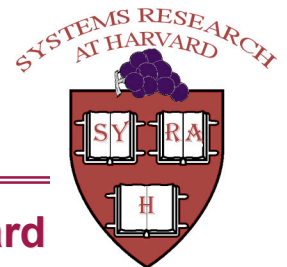


Margo Seltzer, David Holland,
Kiran-Kumar Muniswamy-Reddy,
Uri Braun, and Jonathan Ledlie

Harvard University

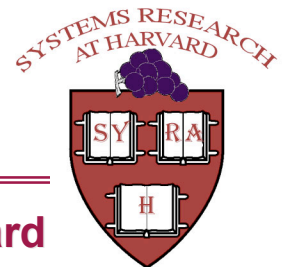
What is Provenance?

- What did the President know and when did he know it?
- What the President knew – data
- When he knew it – provenance
- Provenance is metadata about the history of an object

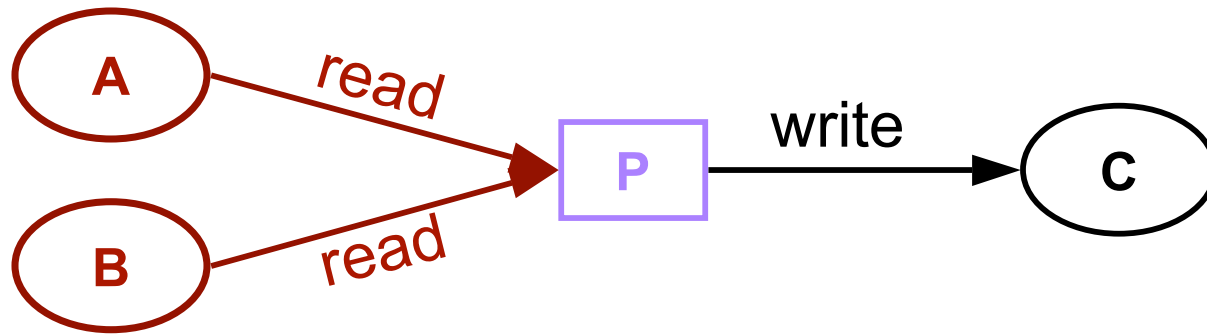


What is Provenance? (contd)

- For computer objects, provenance is the complete history or lineage of a object
 - On what is this object based?
 - How was this object created?
 - How can it be re-created?



Example

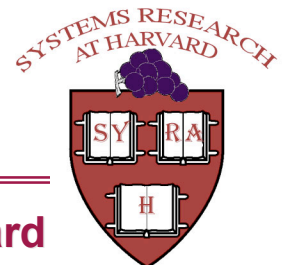


Provenance of C

- Input Files A, B
- Application P
- Command line Args
- Environment
- Processor type, OS, etc

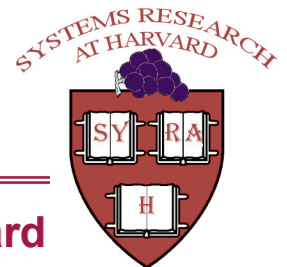
Sample Applications

- Science: how did I (or they) get this result?
- ILM: tweak ILM policies for data belonging to a particular application
- Homeland Security: from what sources did I derive this conclusion?



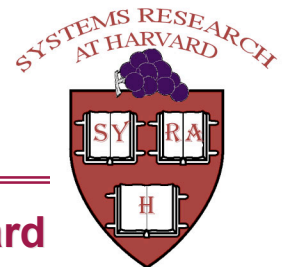
The State of Provenance Today

- Many provenance systems are domain-specific.
- Most provenance is entered manually.
- In many fields, provenance support is simply lacking.



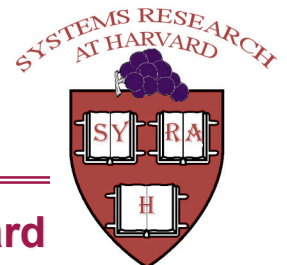
Provenance-Aware Storage Systems (PASS)

- Storage systems (e.g., file systems) in which provenance is a first class entity.
- Provenance:
 - is generated and maintained as transparently as possible.
 - can be indexed and queried.



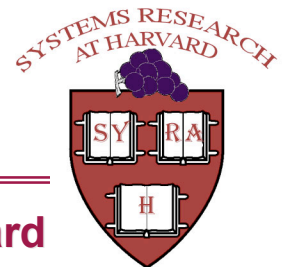
Research Questions:

- Storing provenance: What is the most appropriate way to represent provenance?
- Security: what is the right security model for provenance?
- The wire: how do we implement a distributed PASS?
- Evaluation: how do we evaluate PASS?



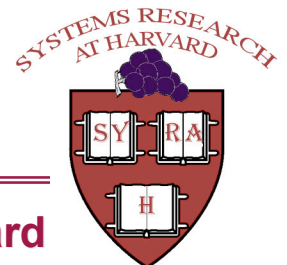
Research Questions (contd):

- What is the most appropriate query interface?
- Search: can we do better than general-purpose search?
- Pruning: when do you delete provenance (or change history)

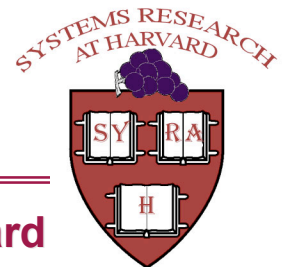
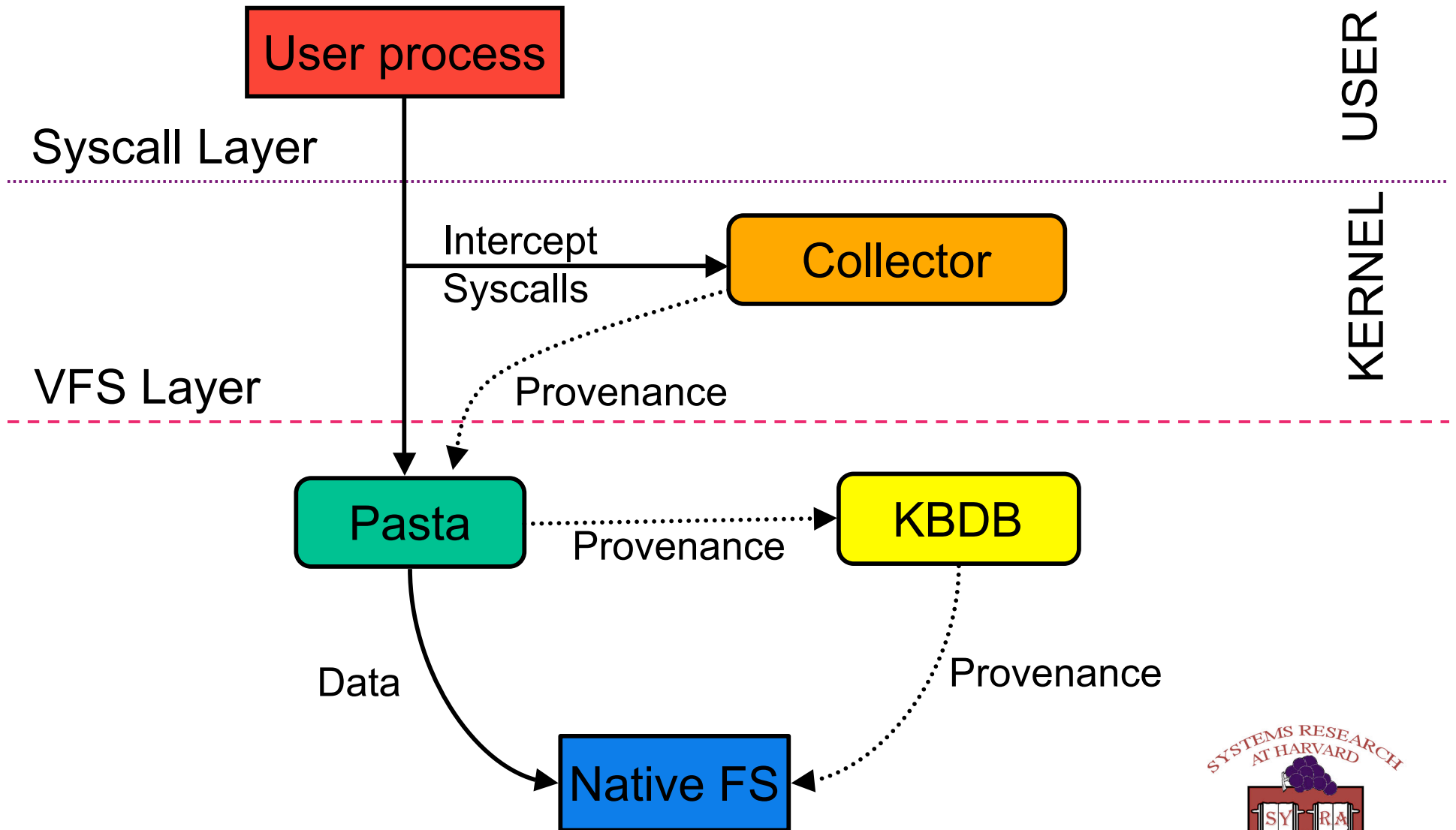


PASS Prototype

- Linux 2.4.29, RedHat 7.3
- In-kernel transactional data store
 - Port of Berkeley DB into the kernel
 - Provided by SUNY Stony Brook
- Provenance And Storage Layer: PASTA
 - Stacked file system
 - Constructed using FiST



PASS Architecture



Questions?

Contact:

pass@eecs.harvard.edu

www.eecs.harvard.edu/syrah/pass

Prototype Available in January

Thanks to our Sponsors:

