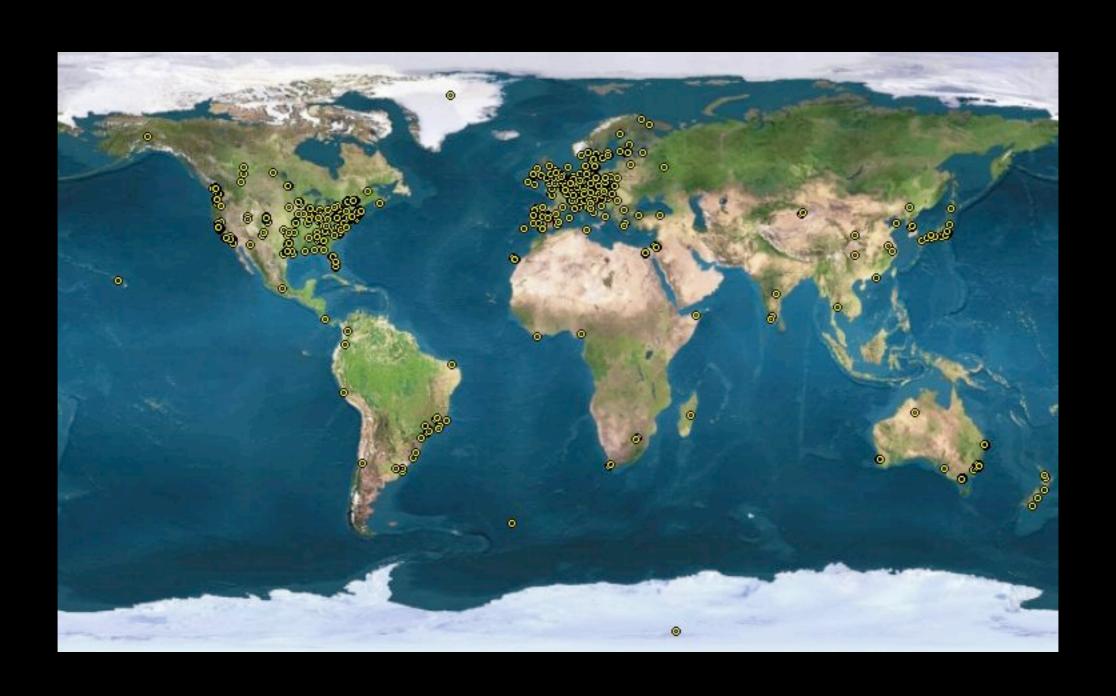
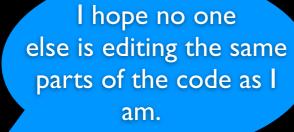


Lecture 2
Pre-processing
Concurrent Versions System Archives

Debian Developer Coordinates





Oh god! I missed my meeting and don't have access to the files.

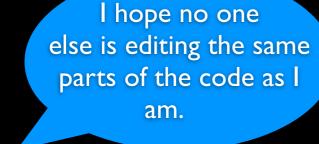


Hmmm... I think what I wrote yesterday was much better!



OK, I have now edited my code and added new classes. How can the others access it?





Oh god! I missed my meeting and don't have access to the files.



Hmmm... I think what I wrote yesterday was much better!



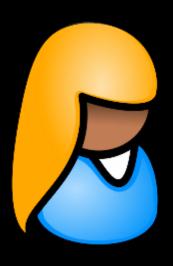
OK, I have now edited my code and added new classes. How can the others access it?

Am I the only one working on a weekend?













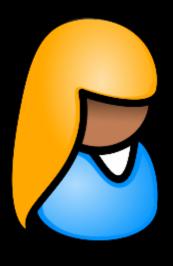
































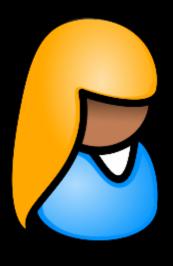




































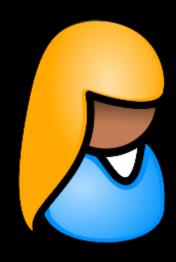
















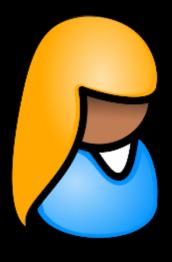






































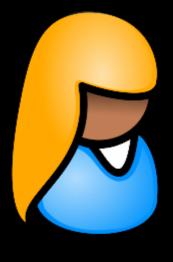
 \sum

CVS Server



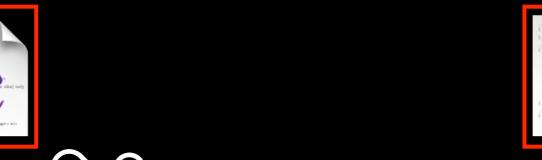






























Concurrent Versions System



...a mechanism for storing multiple versions of files.

Concurrent Versions System



...a mechanism for storing multiple versions of files.

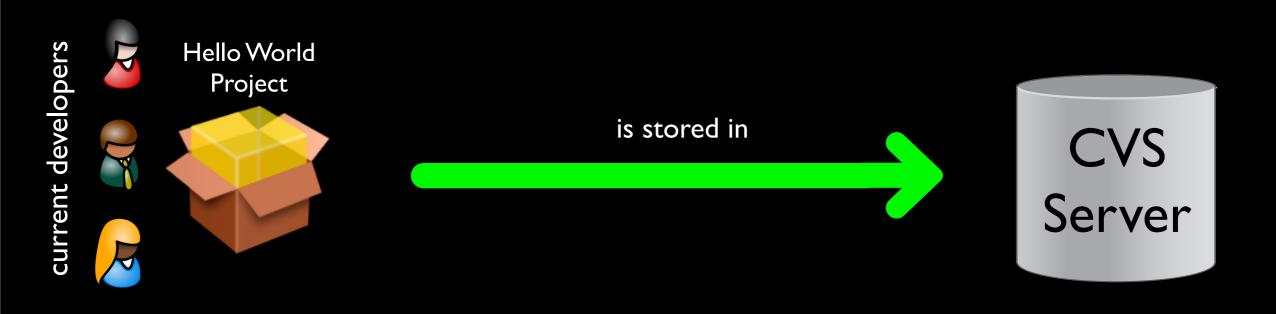


Not only for groups, but individuals too! - Back-up - Version - Roll-back - Branches

Concurrent Versions System

- Record the history of all files in the repository.
- Share the files amongst a group of people.
- Have multiple people edit the same files.
- Merge changes from different people.
- Time portal.

Hello World Project



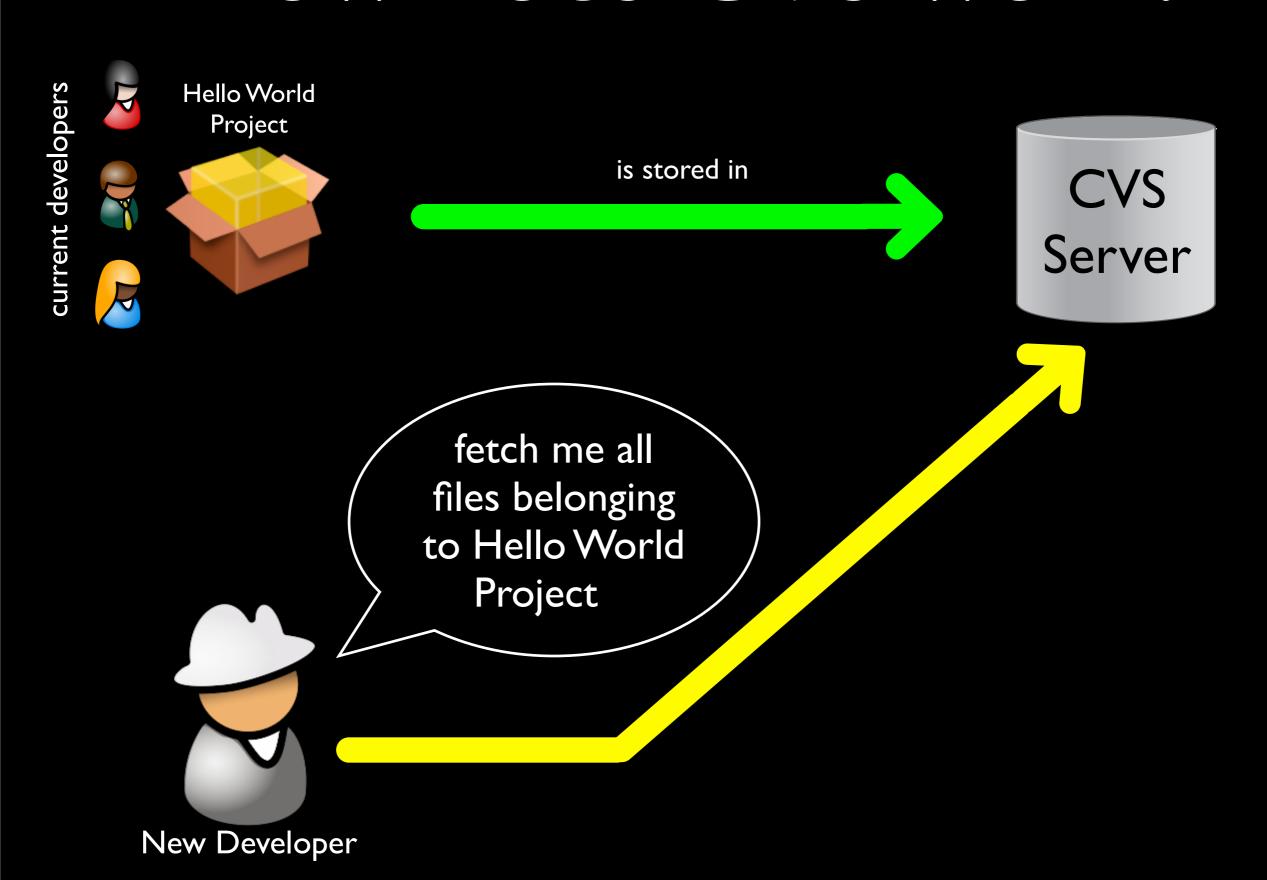
Hello World Project

is stored in

CVS

Server





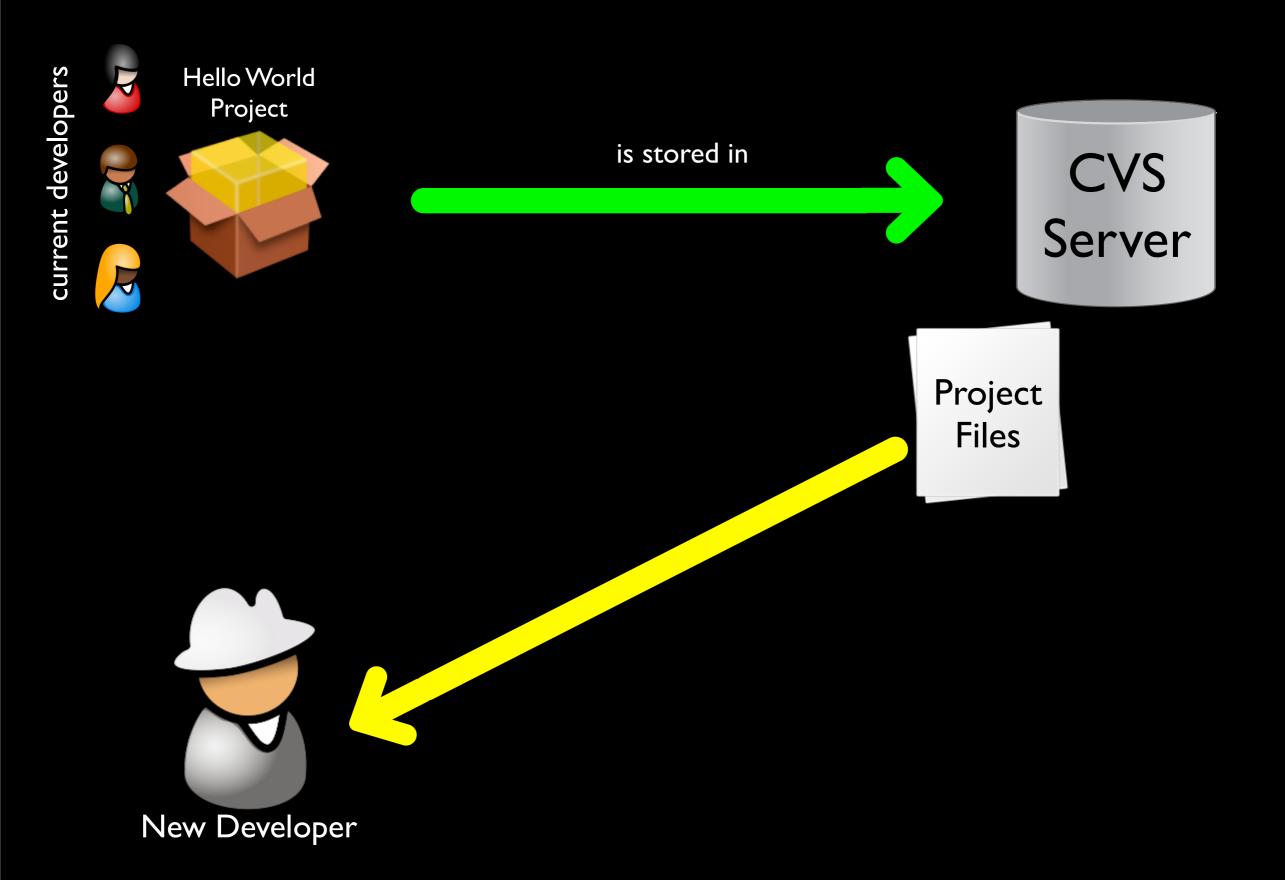
Hello World Project

is stored in

CVS

Server





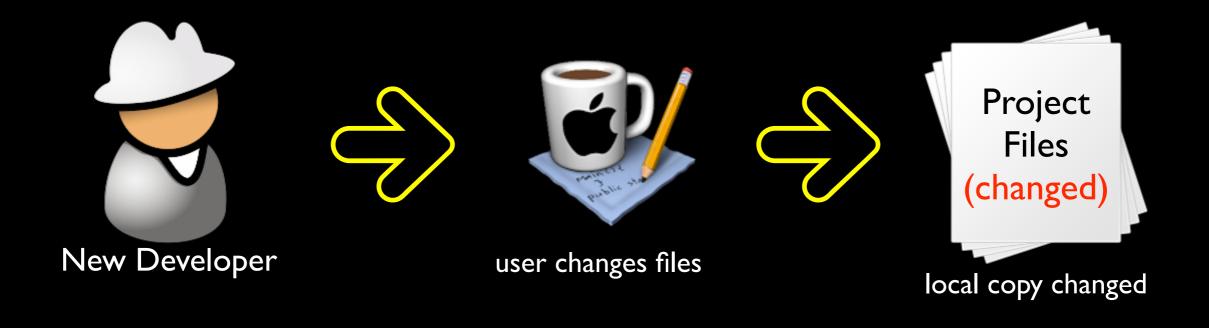


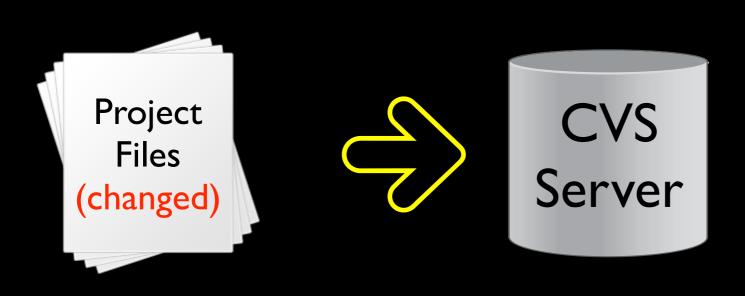




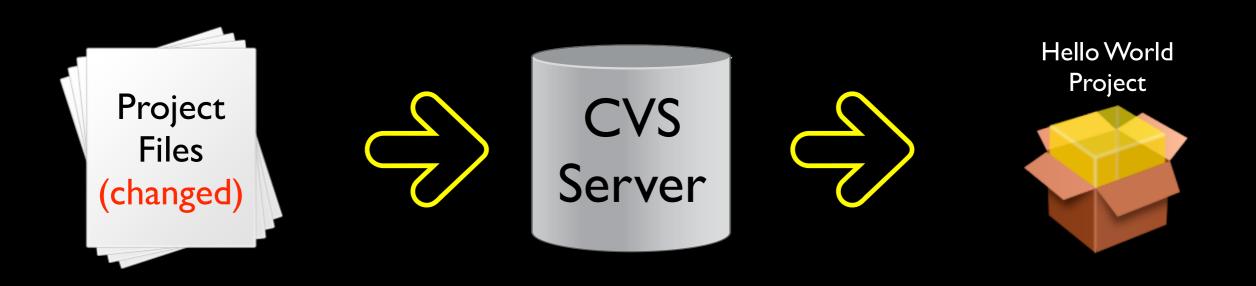


user changes files





local copy changed



server updates project repository

local copy changed





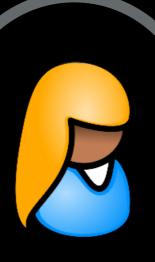
















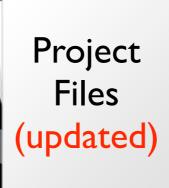






How does CVS work?











Revision Example

Revision Example

Revisions to Foo.java

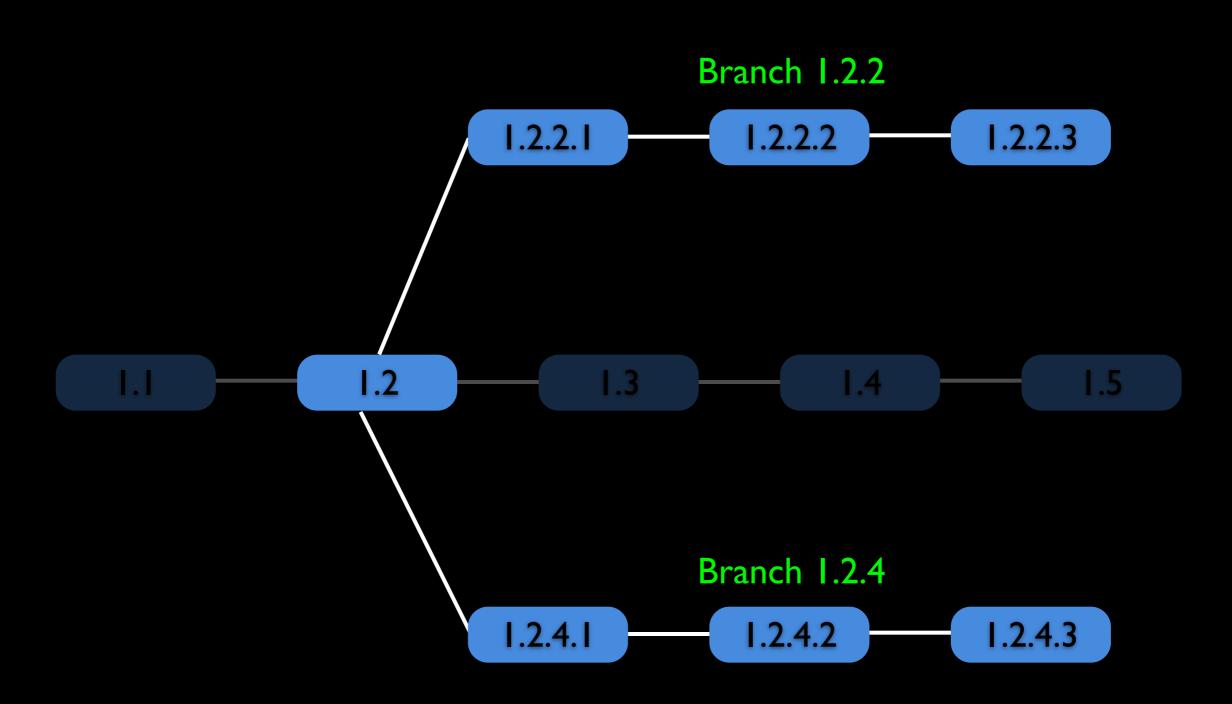
Revision Example

Revisions to Foo.java

1.3

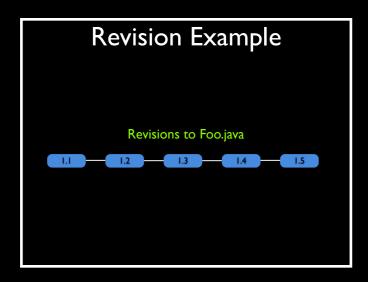
1.4

Branches Example



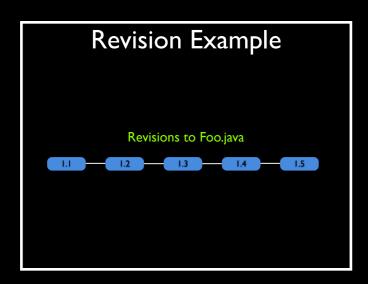
Revisions & Branches

Revisions & Branches

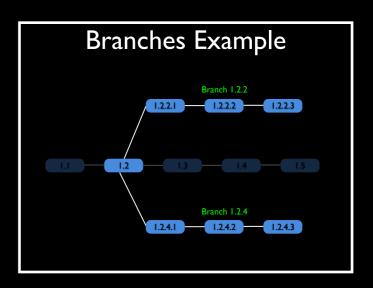


- Each version of a file has a unique revision number.
- Revision number has even number of dot separated integers.
- I.I is usually the first revision number.
- Successive revision numbers are incremented by I.

Revisions & Branches



- Each version of a file has a unique revision number.
- Revision number has even number of dot separated integers.
- I.I is usually the first revision number.
- Successive revision numbers are incremented by I.



- Each branch has a unique branch number.
- Branch number has odd number of dot separated integers.
- All revisions on a branch have revision numbers formed by appending an ordinal number to the branch number.

cvs checkout

cvs checkout

cvs update

cvs checkout

cvs update

cvs add

cvs checkout

cvs update

cvs remove

cvs add

cvs checkout

cvs update

cvs release

cvs remove

cvs add

cvs checkout

cvs update

cvs release

cvs remove

cvs add

cvs log

cvs checkout

cvs update

cvs diff

cvs release

cvs remove

cvs add

cvs log

cvs checkout

cvs update

cvs diff

cvs release

cvs remove

cvs add

cvs log

cvs tag

cvs checkout

cvs update

cvs diff

cvs release

cvs remove

cvs add

cvs rtag

cvs log

cvs tag

cvs checkout

cvs update

cvs diff

cvs release

cvs remove

cvs add

cvs rtag

cvs log

cvs history

cvs tag

Tom Ball et al. "If your version control system could talk..."

Tom Ball et al. "If your version control system could talk..."
So, why is my CVS so silent?

Tom Ball et al. "If your version control system could talk..."

So, why is my CVS so silent?

- CVS has limited quer y functionality and is slow.
 - ⇒ Copy CVS into a database
- CVS splits up changes on multiple files.
 - ⇒ Infer transactions
- CVS knows only files—but what about functions?
 - ⇒ Detect fine-grained changes
- CVS contains unreliable data which is noise.
 - ⇒ Clean data.

Tom Ball et al. "If your version control system could talk..."

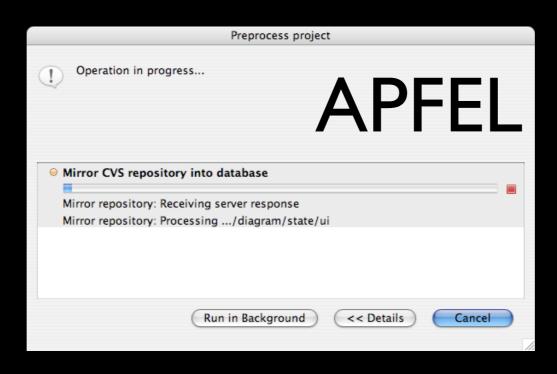
So, why is my CVS so silent?

- CVS has limited quer y functionality and is slow.
 - ⇒ Copy CVS into a database
- CVS splits up changes on multiple files.
 - ⇒ Infer transactions
- CVS knows only files—but what about functions?
 - ⇒ Detect fine-grained changes
- CVS contains unreliable data which is noise.
 - ⇒ Clean data.

Preprocessing is the key to a talkative version control system.

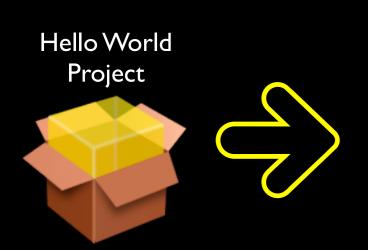
APFEL collects CVS data

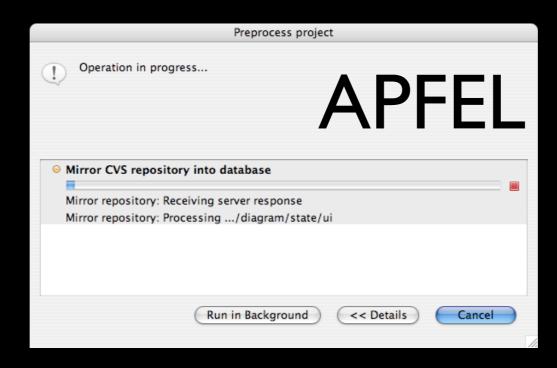






APFEL collects CVS data





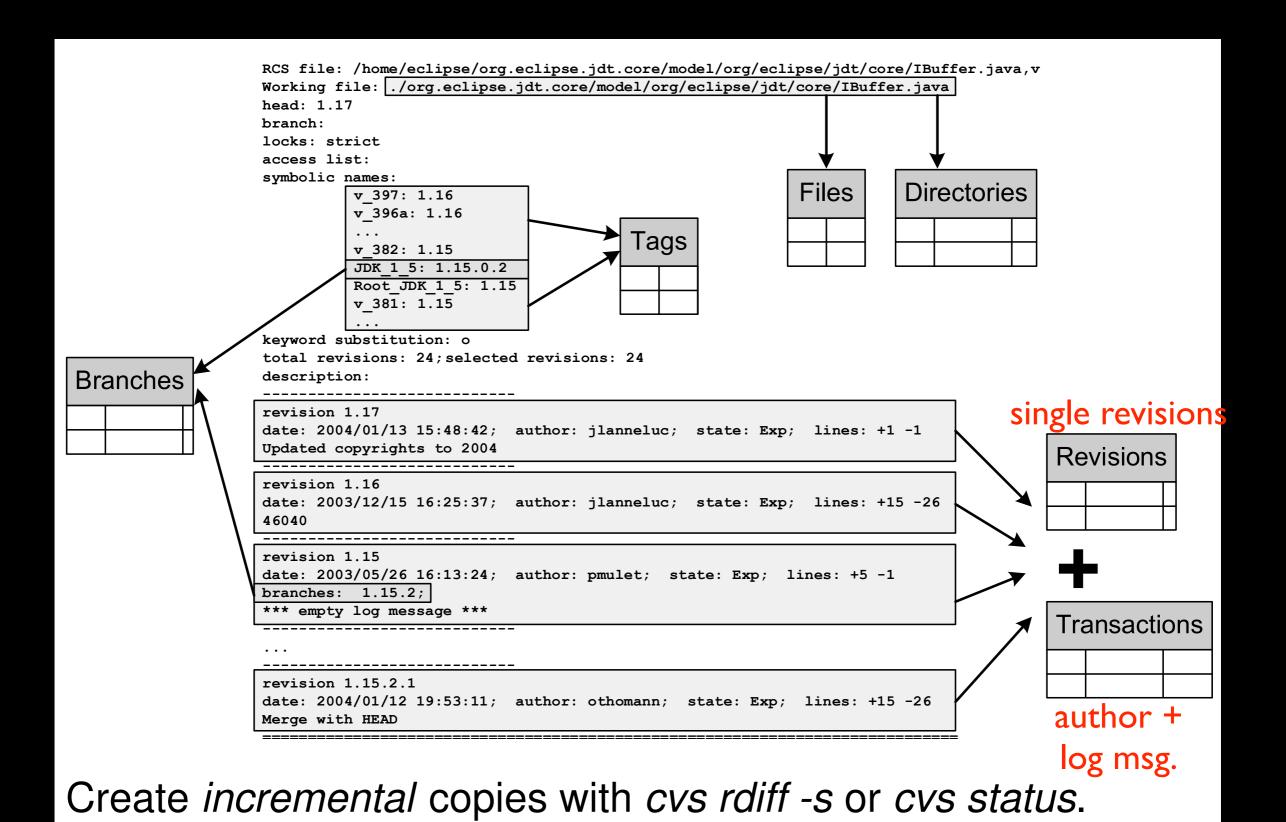




Data Extraction

- What data should be extracted?
- "Filtering should be performed within the analysis, and not during extraction."

Data Extraction



Commit Mails

All changes listed in a commit mail belong to one transaction.

```
CVSROOT: /cvs/gcc
Module name: gcc
Changes by: zack@gcc.gnu.org 2004-05-01 19:12:47

Modified files:
gcc/cp : ChangeLog decl.c

Log message:
* decl.c (reshape_init): Do not apply TYPE_DOMAIN to a VECTOR_TYPE.
Instead, dig into the representation type to find the array bound.

Patches:
http://.../cvsweb.cgi/gcc/gcc/cp/ChangeLog.diff?...&r2=1.4042
http://.../cvsweb.cgi/gcc/gcc/cp/decl.c.diff?...&r2=1.1204
```

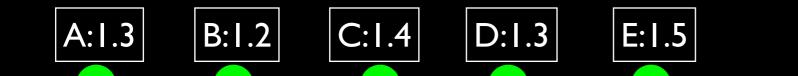
Commit mails for GCC: http://gcc.gnu.org/ml/gcc-cvs/

Not every project provides useful commit mails.

All changes by the same developer, with the same message, made at the same time belong to one transaction.

All changes by the same developer, with the same message, made at the same time belong to one transaction.

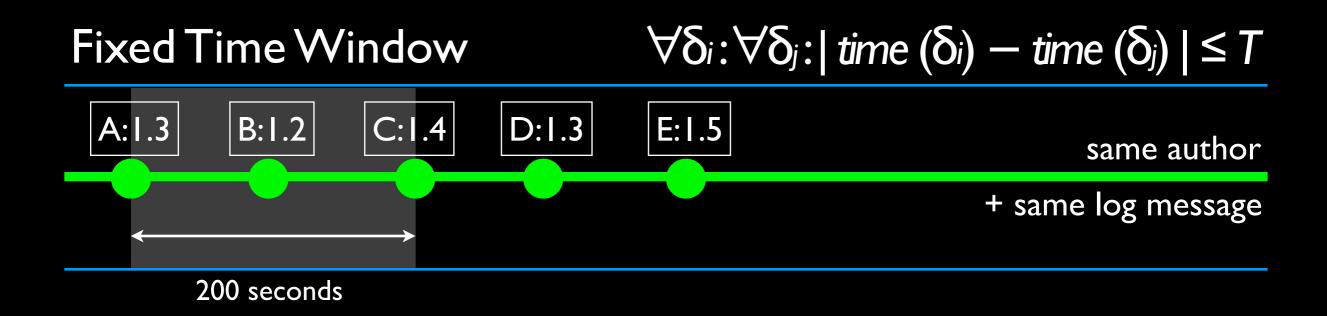
Fixed Time Window



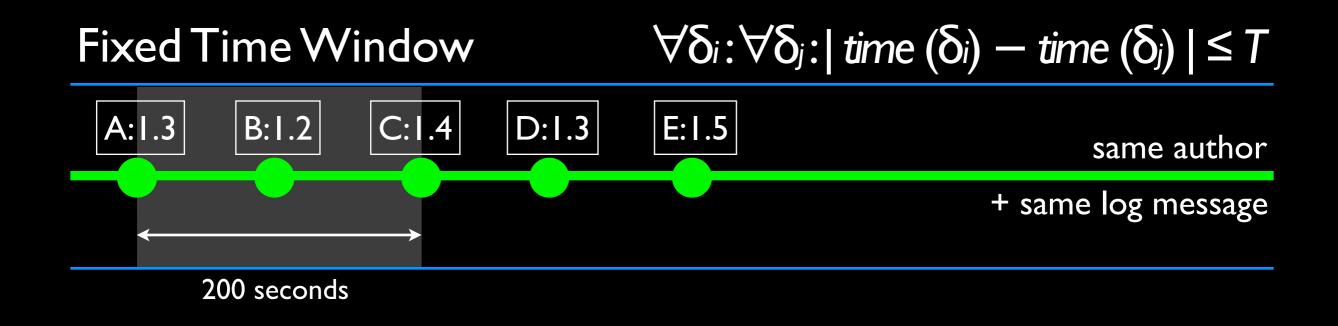
All changes by the same developer, with the same message, made at the same time belong to one transaction.

Fixed Time Window $\forall \delta_i : \forall \delta_j : | \text{ time } (\delta_i) - \text{ time } (\delta_j) | \leq T$ A:1.3 B:1.2 C:1.4 D:1.3 E:1.5 same author + same log message

All changes by the same developer, with the same message, made at the same time belong to one transaction.



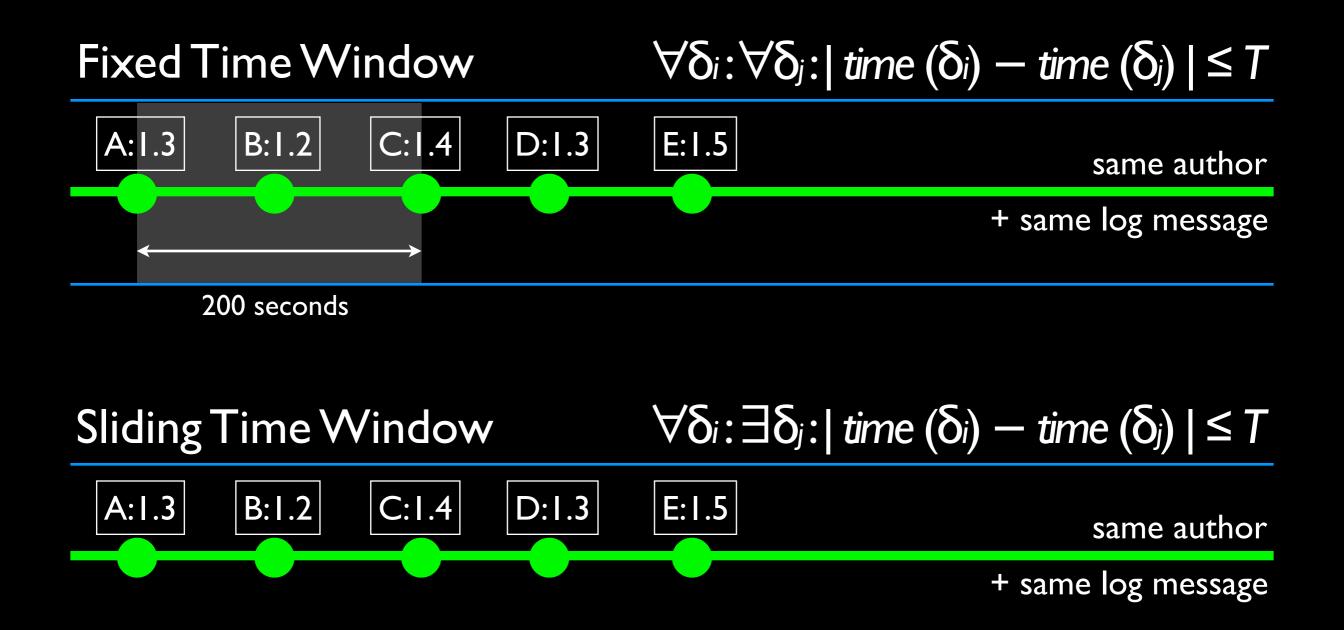
All changes by the same developer, with the same message, made at the same time belong to one transaction.



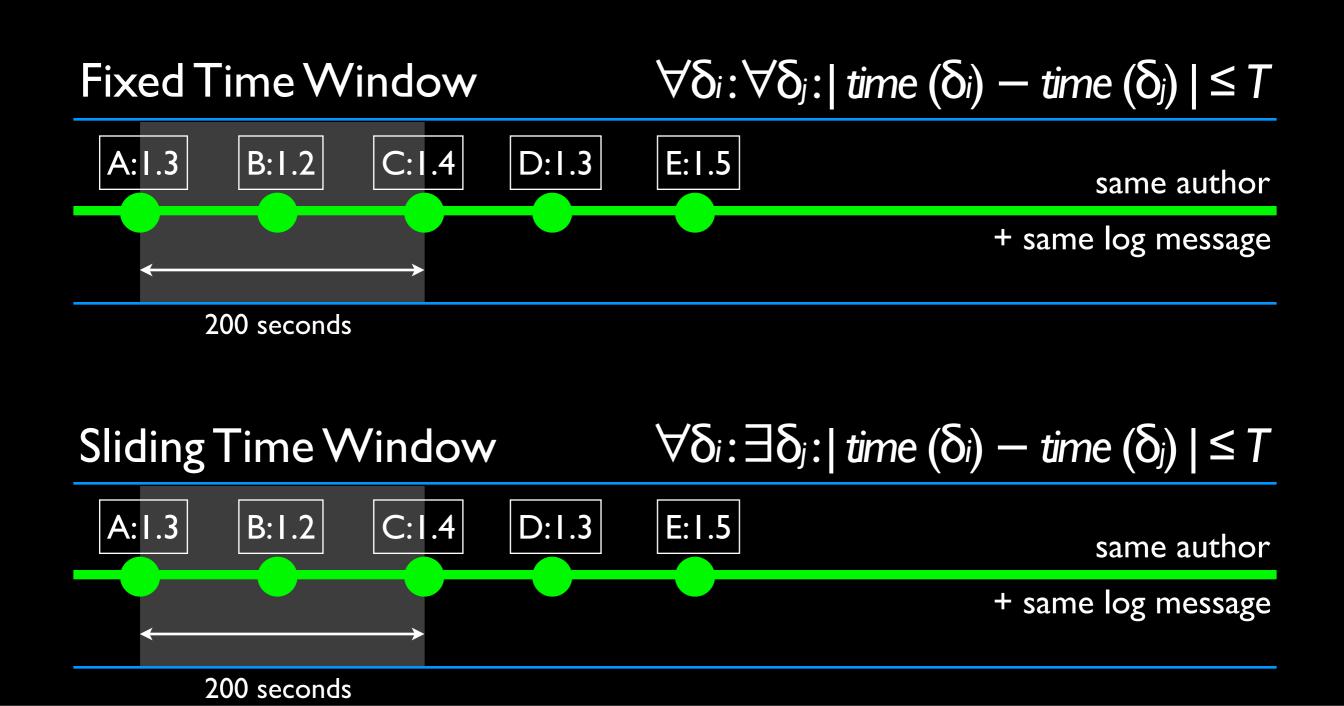
Sliding Time Window

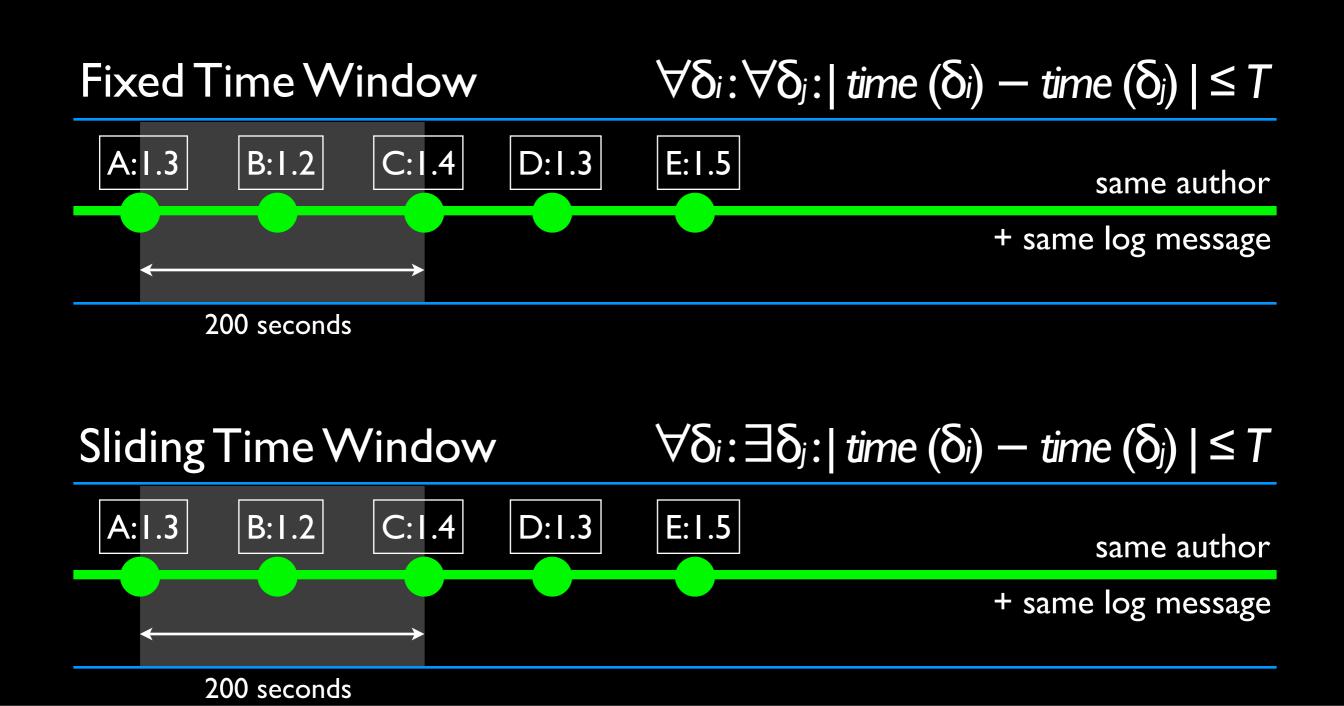
A:1.3	B:1.2	C:1.4	D:1.3	E:1.5

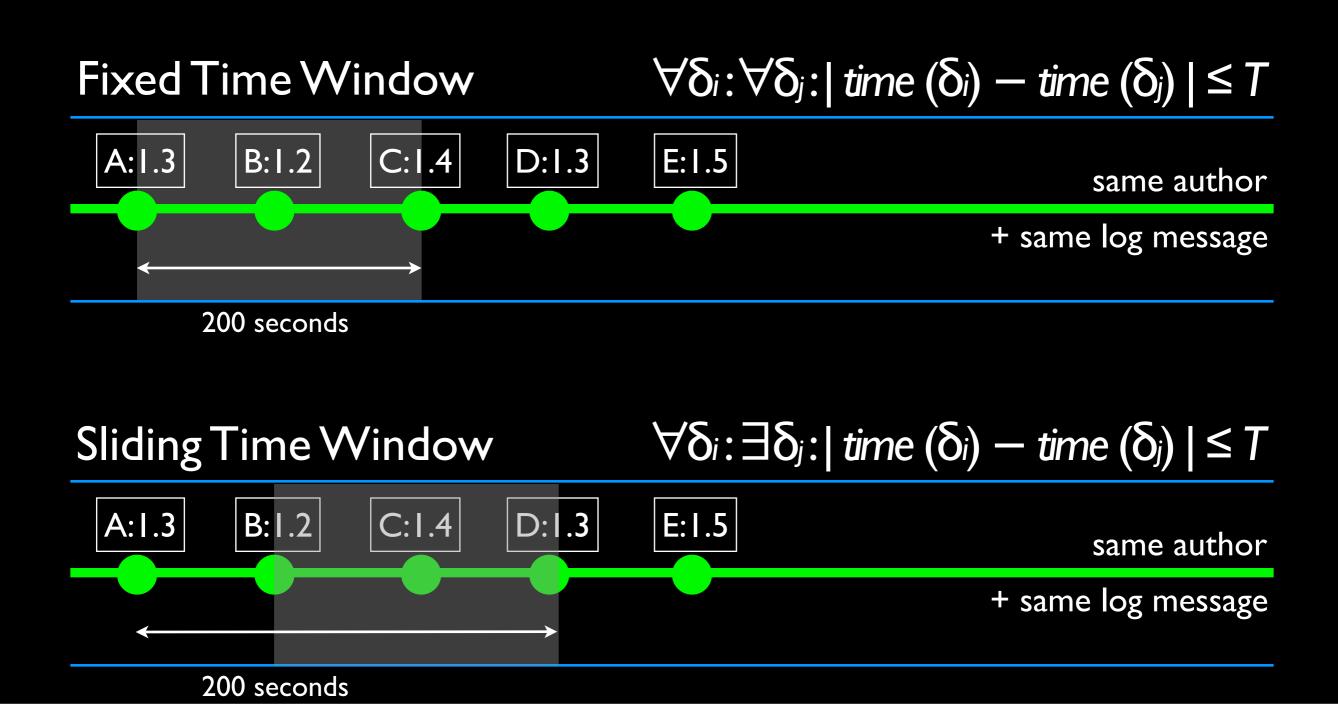
All changes by the same developer, with the same message, made at the same time belong to one transaction.

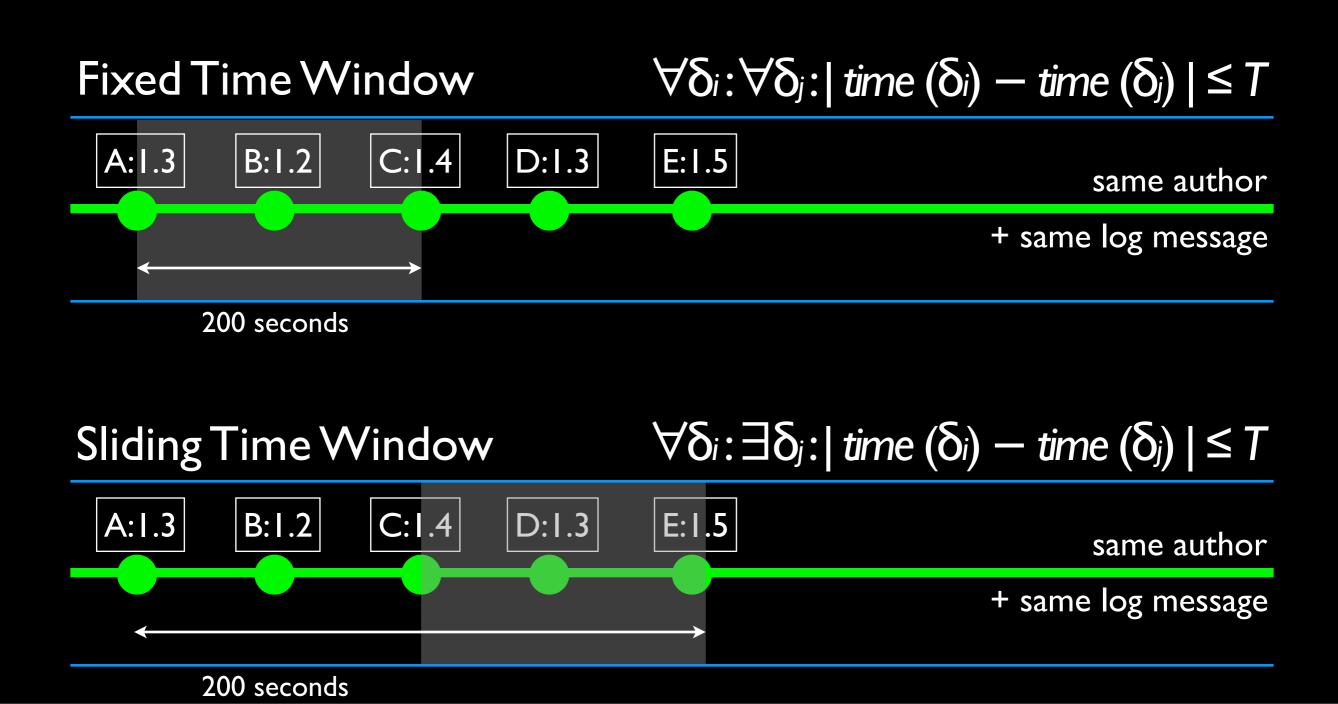


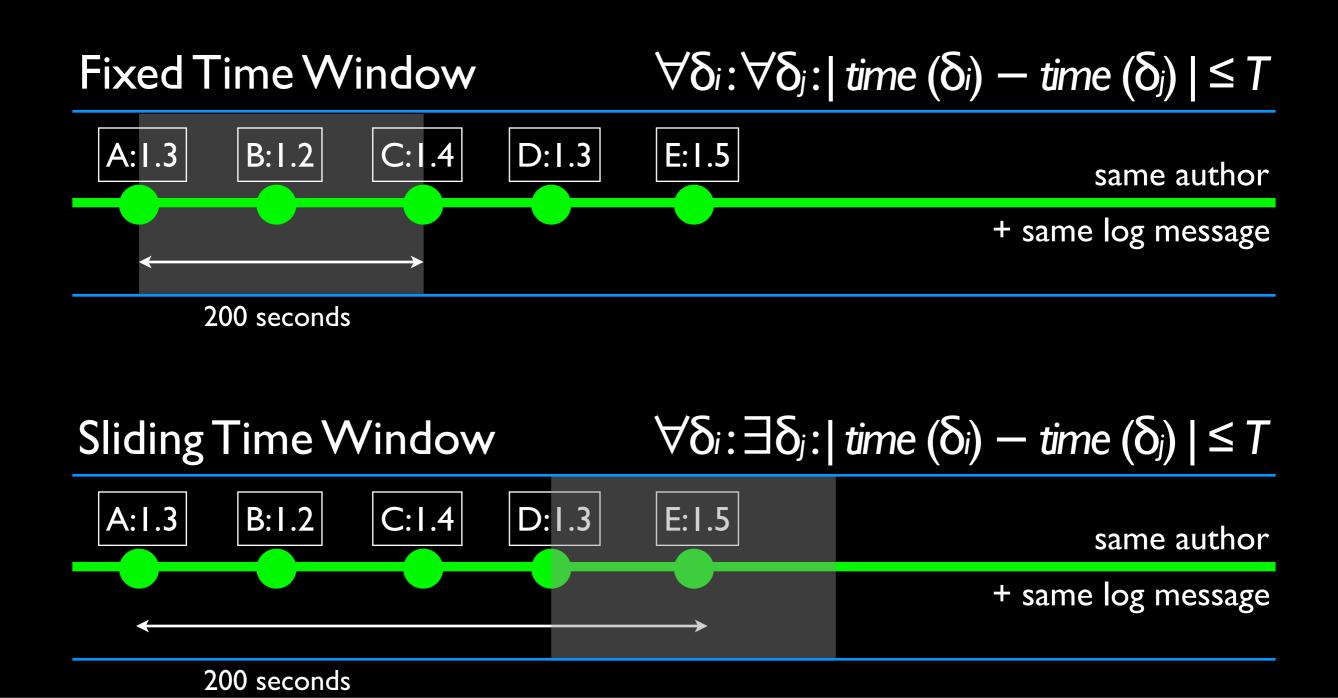
All changes by the same developer, with the same message, made at the same time belong to one transaction.

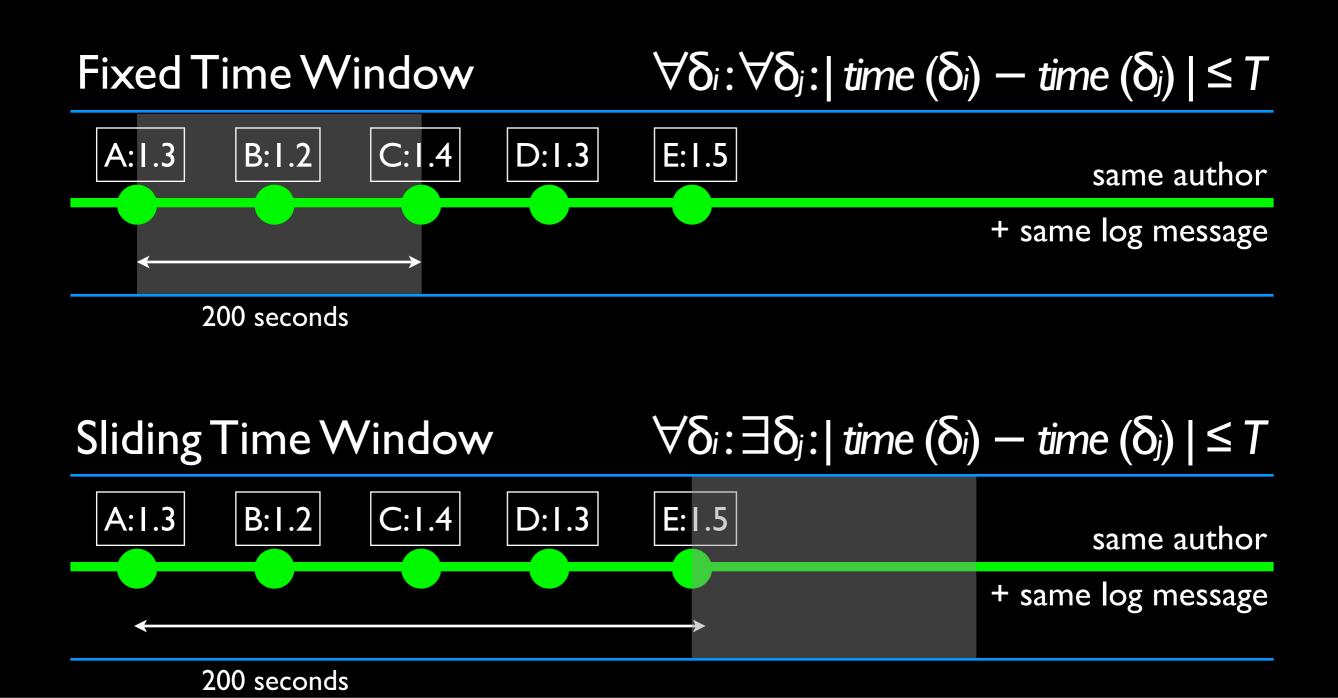


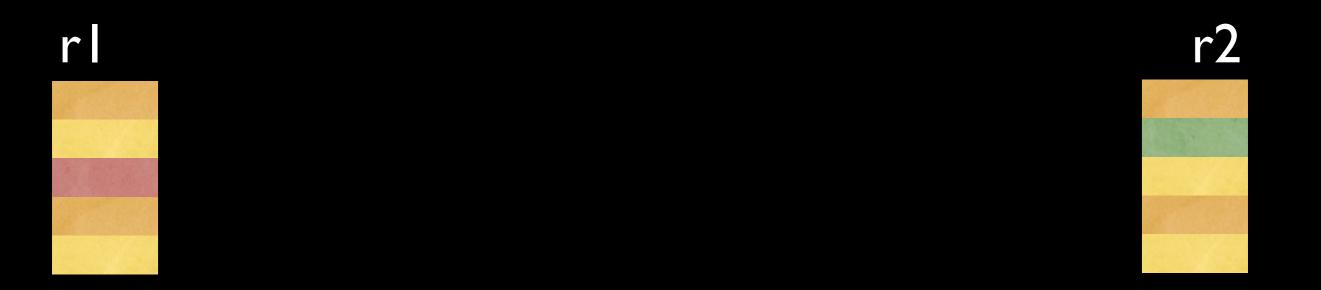






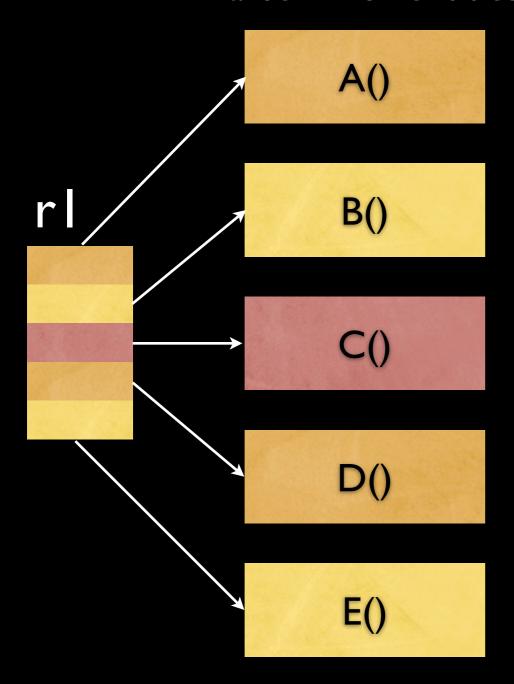




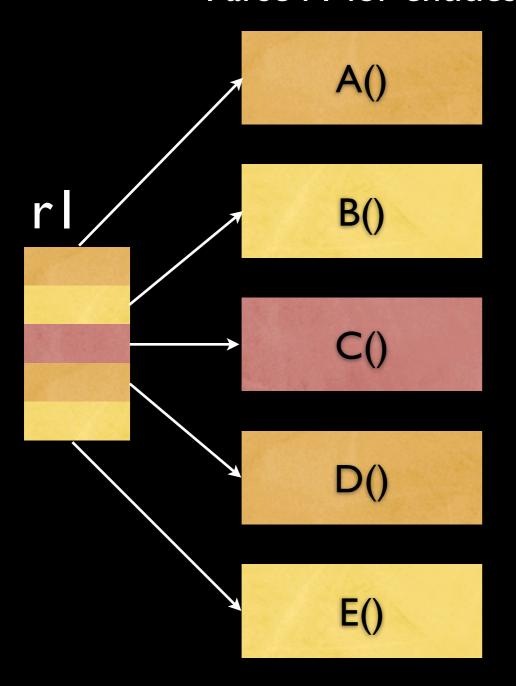


r2

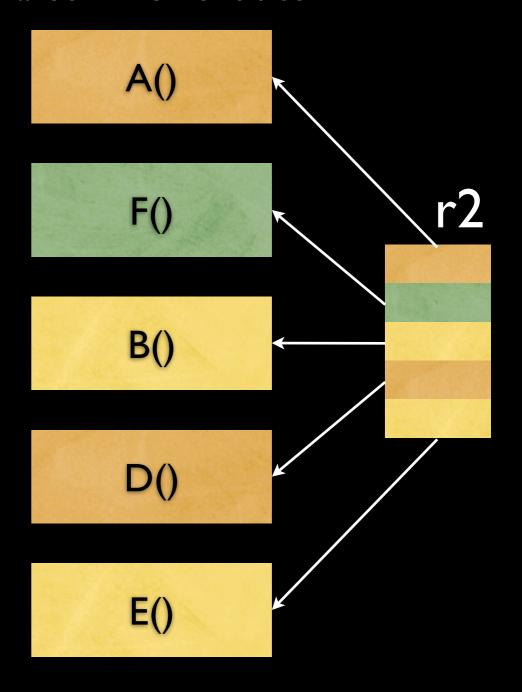
Parse rl for entities

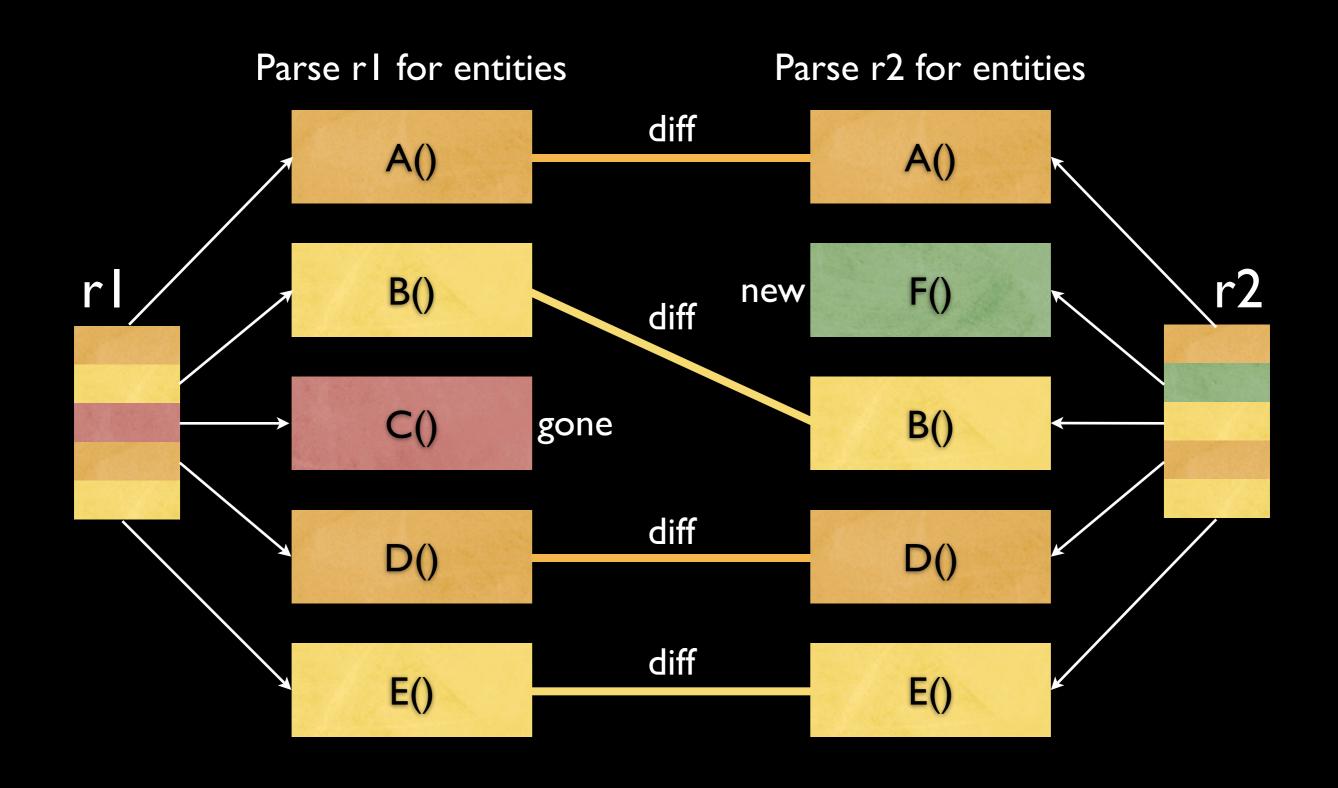


Parse rl for entities



Parse r2 for entities





Fine-grained changes

```
public static void initDefaults(IPreferenceStore store) {
   store.setDefault(OPEN_STRUCTURE_COMPARE, true);
   store.setDefault(SYNCHRONIZE_SCROLLING, true);
   store.setDefault(SHOW_PSEUDO_CONFLICTS, false);
   store.setDefault(INITIALLY_SHOW_ANCESTOR_PANE, false);
   store.setDefault(SHOW_MORE_INFO, false);
   store.setDefault(IGNORE_WHITESPACE, false);
   store.setDefault(PREF_SAVE_ALL_EDITORS, false);
   //store.setDefault(USE_SPLINES, false);
   store.setDefault(USE_SINGLE_LINE, true);
   //store.setDefault(USE_RESOLVE_UI, false);
   store.setDefault(PATH_FILTER, ""); //$NON-NLS-1$
}
```

Fine-grained changes

Fine-grained changes

Collecting Tokens

```
public static void main(String[] args) {
 System.out.println("Hello Portland.");
 System.out.println("Hello eTX.");
     P-public
                           T-String[]
     P-static
                           V-args
     T-System
                           T-System
     V-out
                           V-out
     M-println
                           M-println
     Y-"Hello Portland." Y-"Hello eTX."
```

Collecting Tokens

Modularization

Inheritance

Method calls

Variable usage

Exceptions

Comments

Various

packages, imports

extends, implements

inner vs final call

identifiers, types

throws, throw, catch

Javadoc, blocks, lines

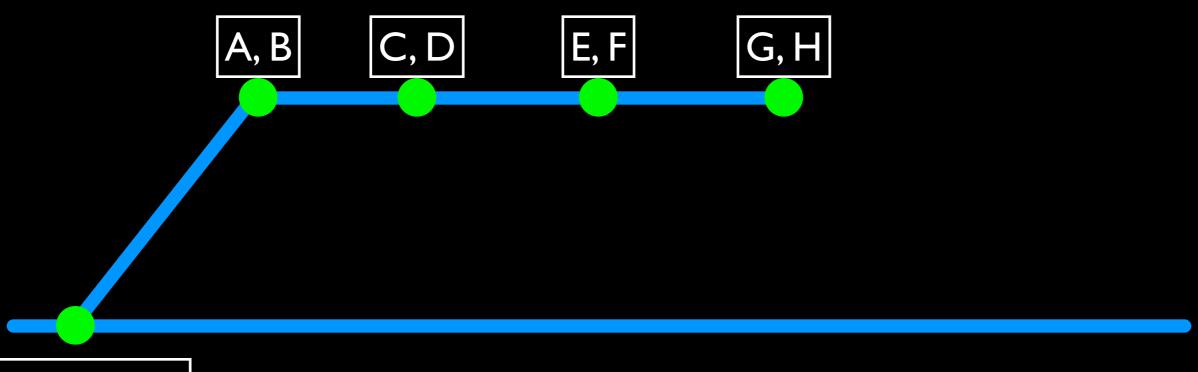
modifiers, literals, operator, keywords

Noise: Large Transactions

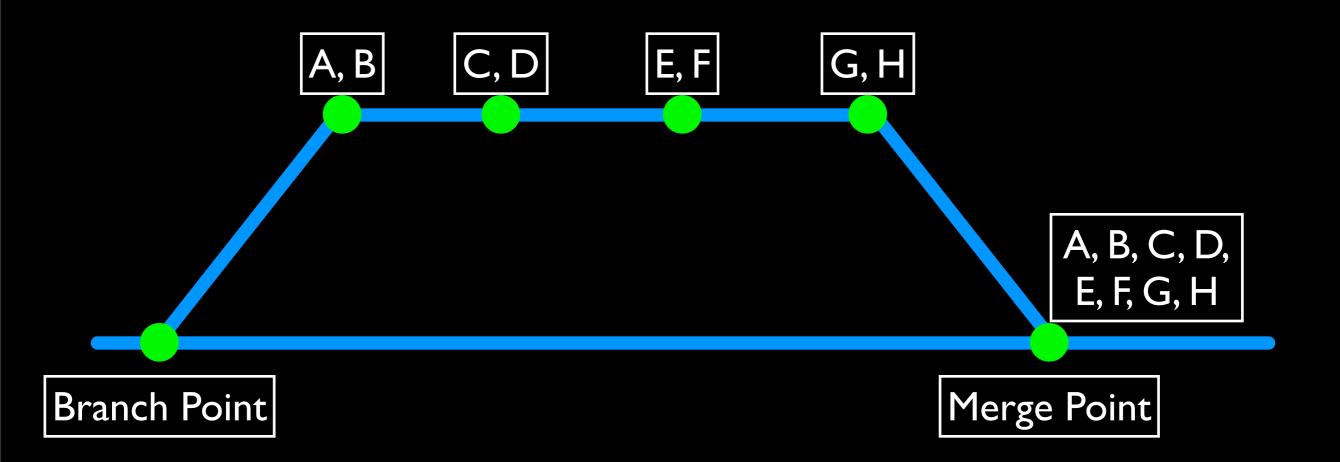
- "Change #include filenames from <foo.h> to <openssl.h>." (552 files)
- "Change functions to ANSI C." (491 files)

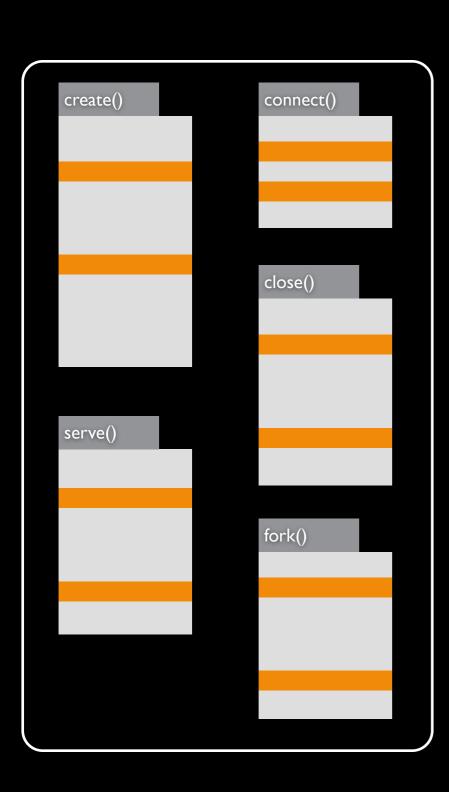
Select an upper bound N for transaction size.

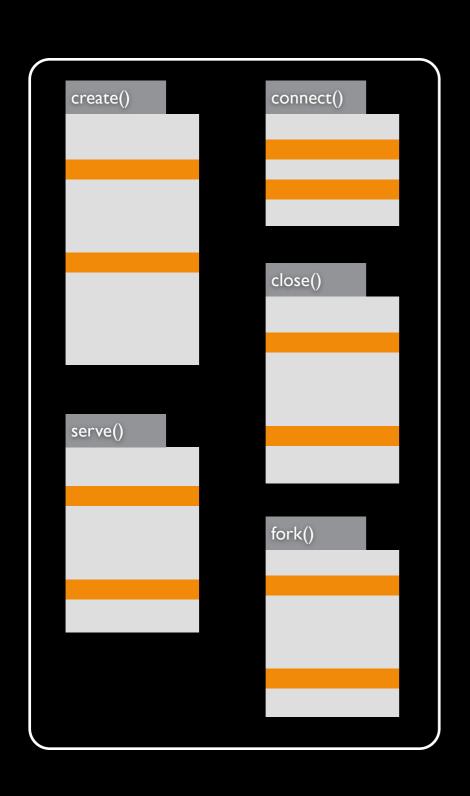




Branch Point

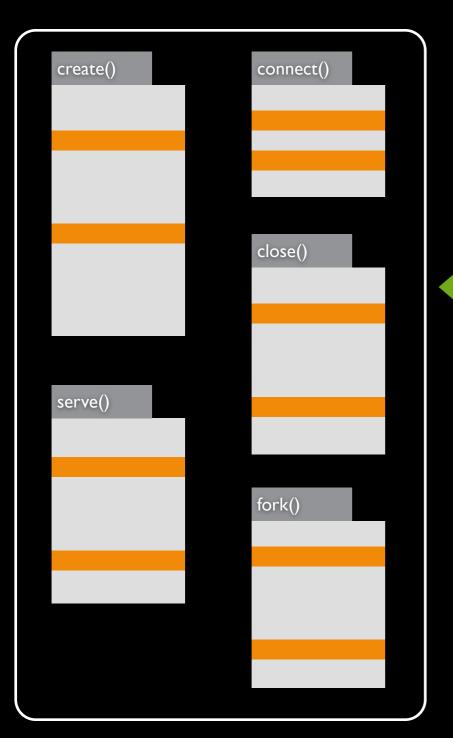






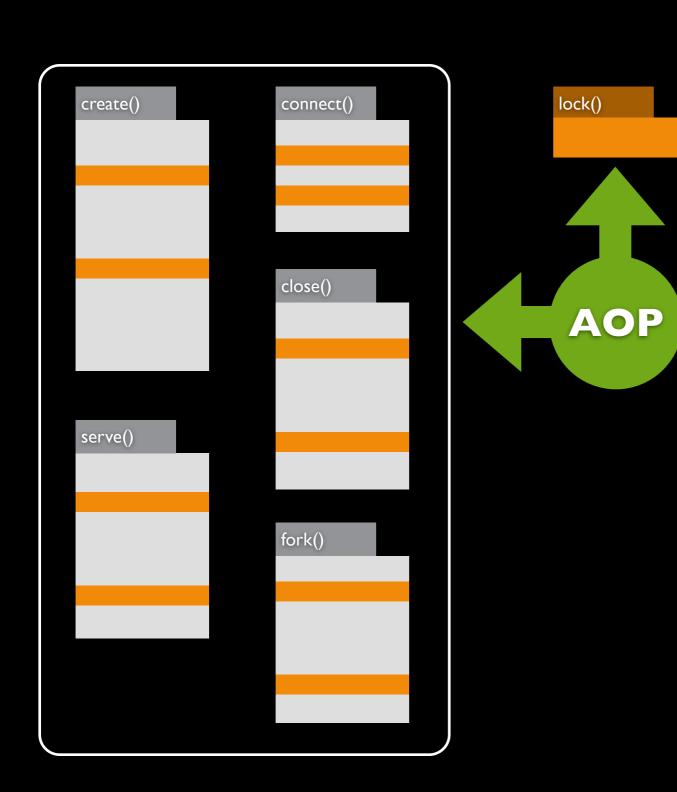
lock()

Typical Aspects
enter()/leave()
log()
lock()/unlock()





Typical Aspects
enter()/leave()
log()
lock()/unlock()



Typical Aspects enter()/leave() log() lock()/unlock()

Aspect Mining: Identifying Aspects in Legacy Code

```
SELECT token_name, COUNT(DISTINCT element_id)
FROM cvs_tokens_method NATURAL JOIN cvs_revisions
WHERE token_type='M' AND change_type='CHG'
GROUP BY transaction_id, token_name
ORDER BY COUNT(DISTINCT element_id) DESC;
```

```
SELECT token_name, COUNT(DISTINCT element_id)
FROM cvs_tokens_method NATURAL JOIN cvs_revisions
WHERE token_type='M' AND change_type='CHG'
GROUP BY transaction_id, token_name
ORDER BY COUNT(DISTINCT element_id) DESC;
```

token_name	count
getString	1462
lock	1284
unlock	1284
error	996
isValidWidget	988

```
SELECT token_name, COUNT(DISTINCT element_id)
FROM cvs_tokens_method NATURAL JOIN cvs_revisions
WHERE token_type='M' AND change_type='CHG'
GROUP BY transaction_id, token_name
ORDER BY COUNT(DISTINCT element_id) DESC;
```

token_name	count
getString	1462
lock	1284
unlock	1284
error	996
isValidWidget	988

Locking Mechanism

```
public static final native void XFree(int address);
public static final void XFree(int /*long*/ address) {
     lock.lock();
     try {
            XFree(address);
       finally {
           lock.unlock();
```

Locking Mechanism

```
public static final native void XFree(int address);
public static final void XFree!
                                        address) {
     locklast
     try
                 OCATIONS.
```

One line fixes

>301	Expression
150	Method call
120	Literal (boundaries, constants)
73	Keyword (true/false confusion)
60	Operator
39	Identifier
39	Comment (!!)

One line fixes

>301	Expression
150	Method call
120	Literal (boundaries, constants)
73	Keyword (true/false confusion)
60	Operator
39	Identifier
39	Comment (!!)

One line fixes

```
>301 Expression
```

150 Method call

Bug report: I 19638 - Typo in Toolltem.getControl 252c252

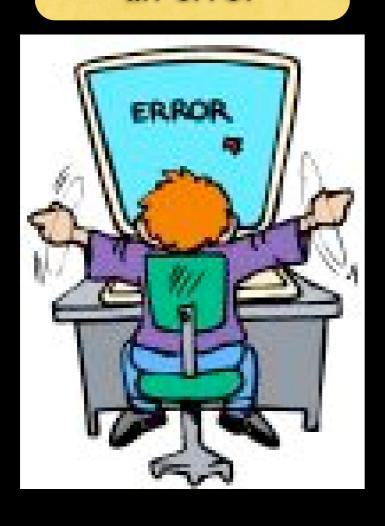
< * the item when the items is a <code>SEPARATOR</code>.

> * the item when the item is a <code>SEPARATOR</code>

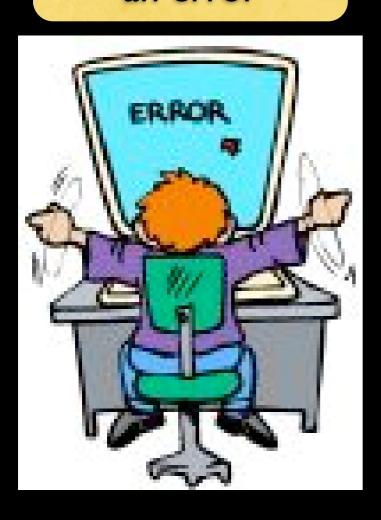
39 Comment (!!)



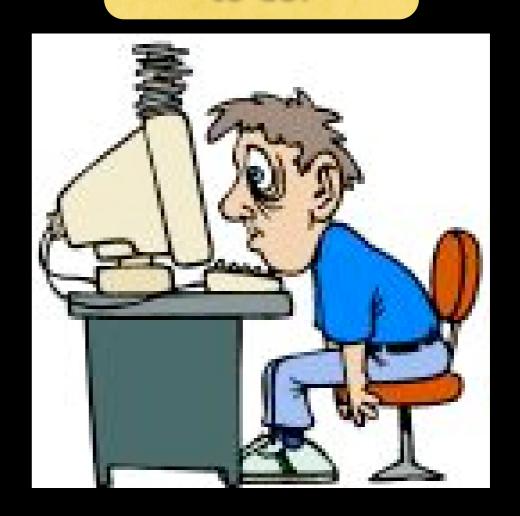
User encounters an error



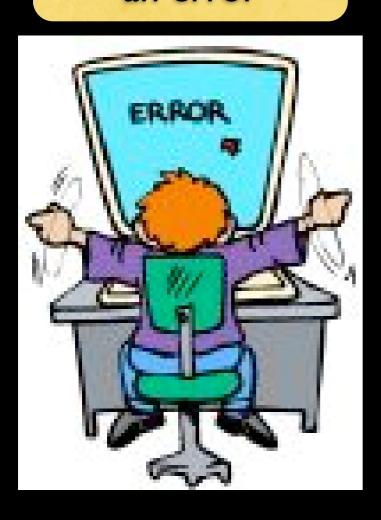
User encounters an error



No idea what to do?



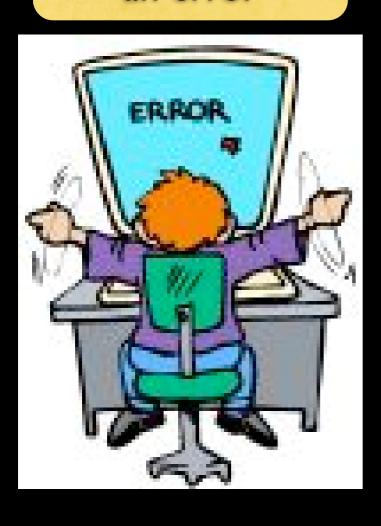
User encounters an error



Take slightly extreme measures



User encounters an error



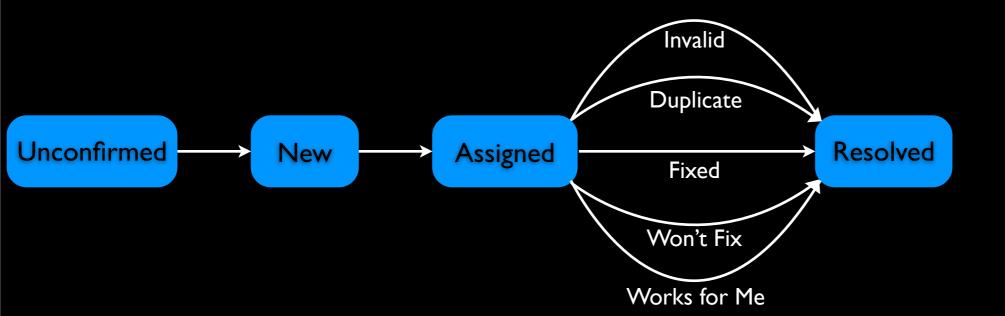
Report the bug!

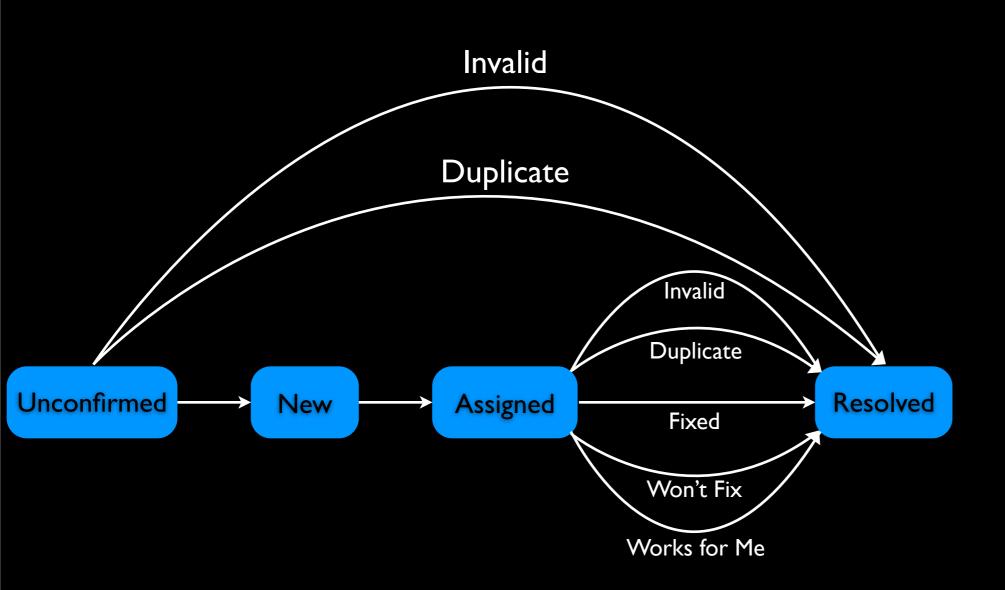


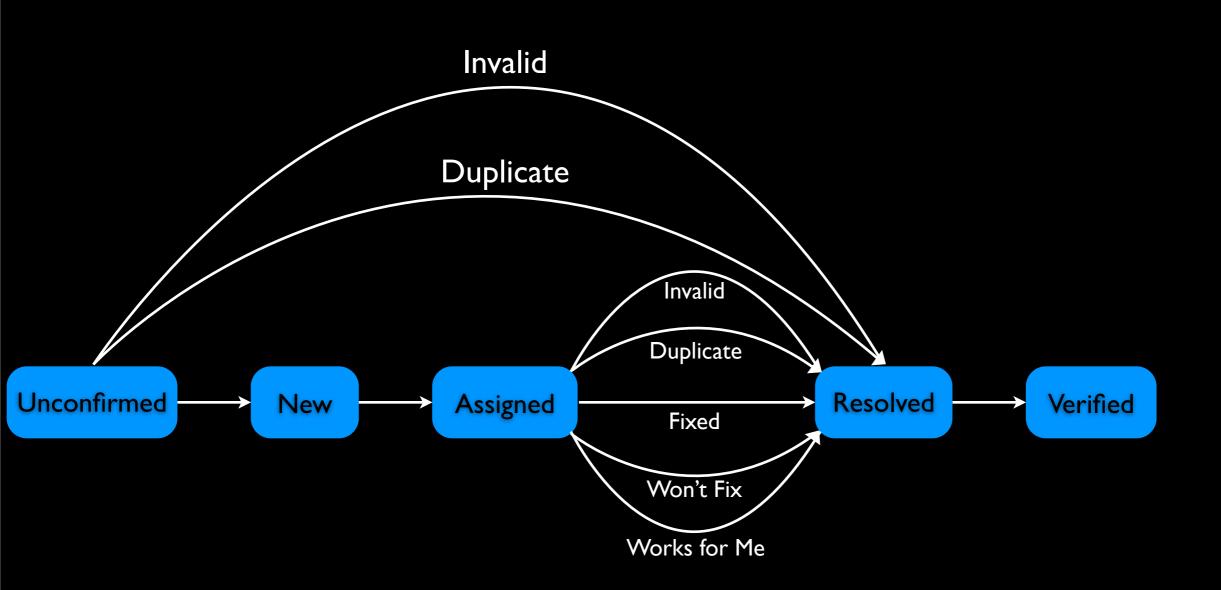
Unconfirmed

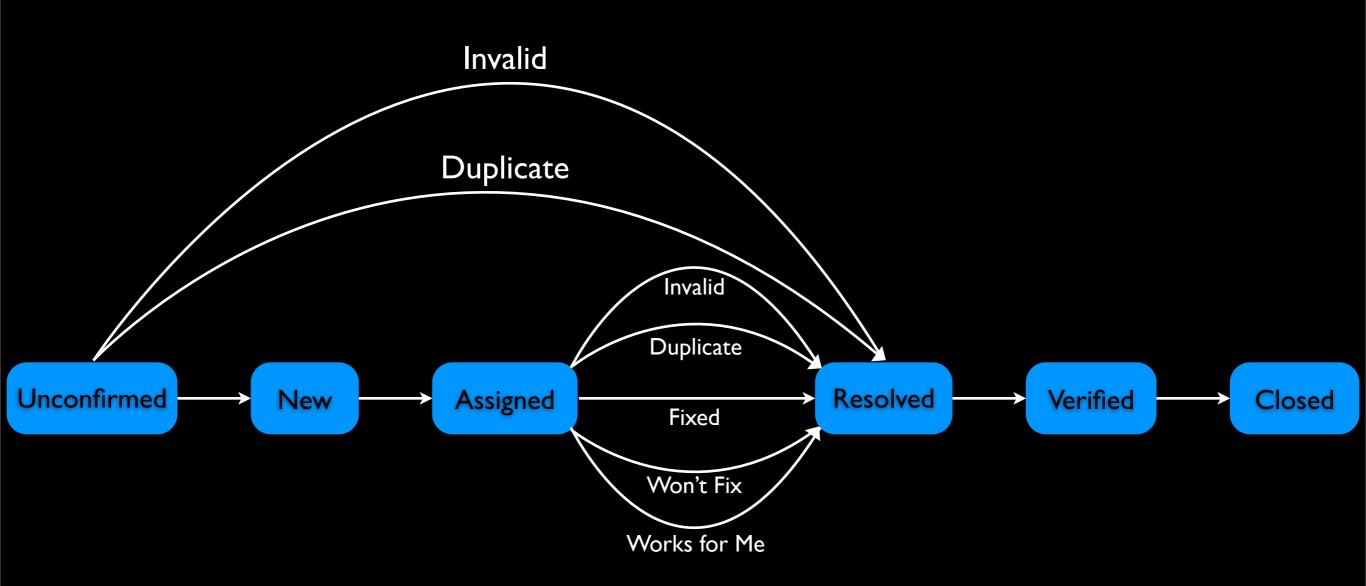


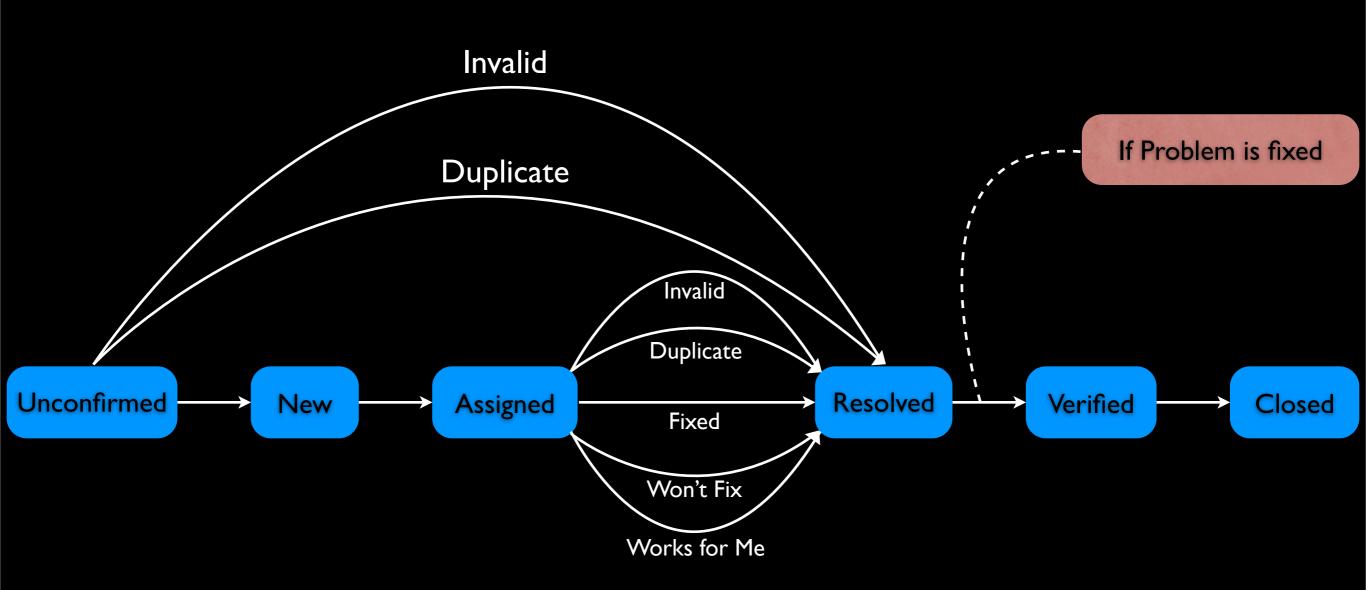


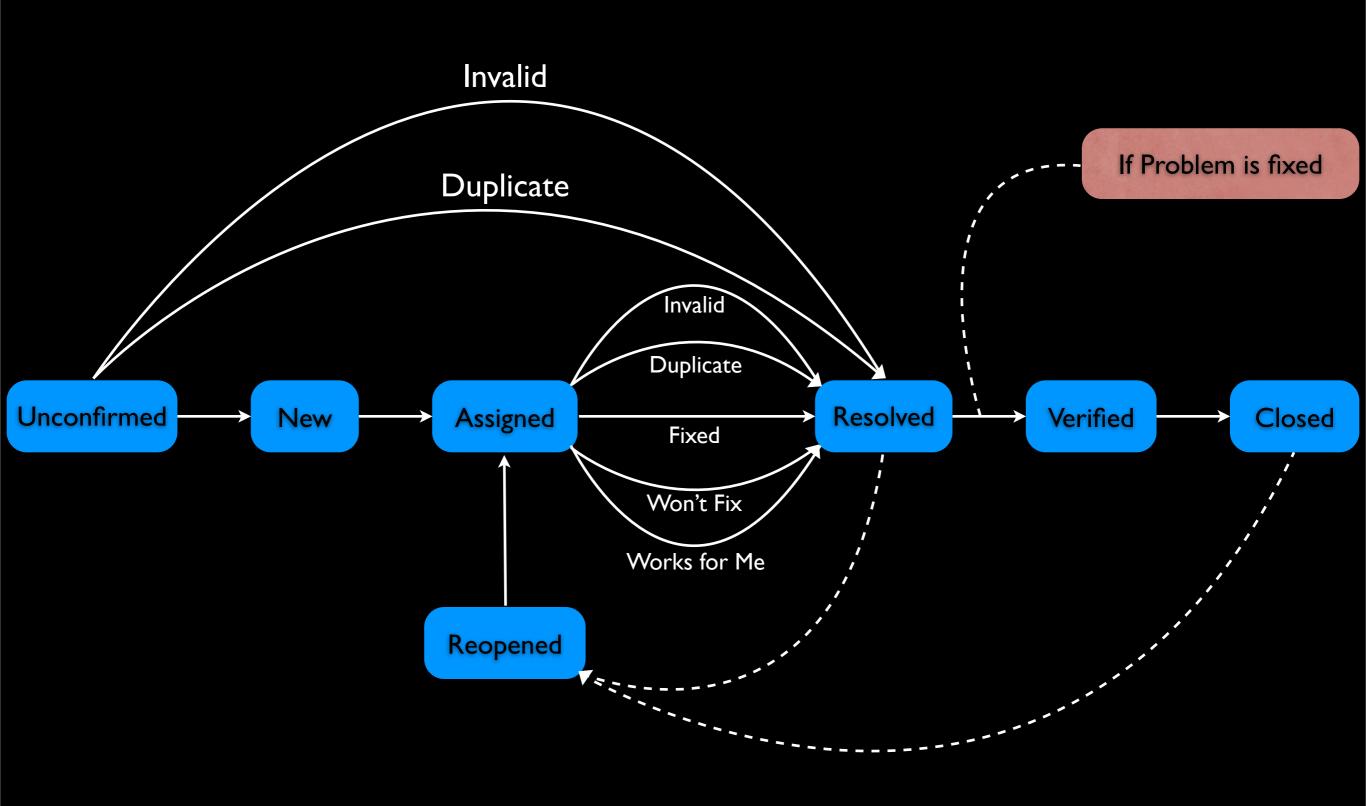














A commercial tool for

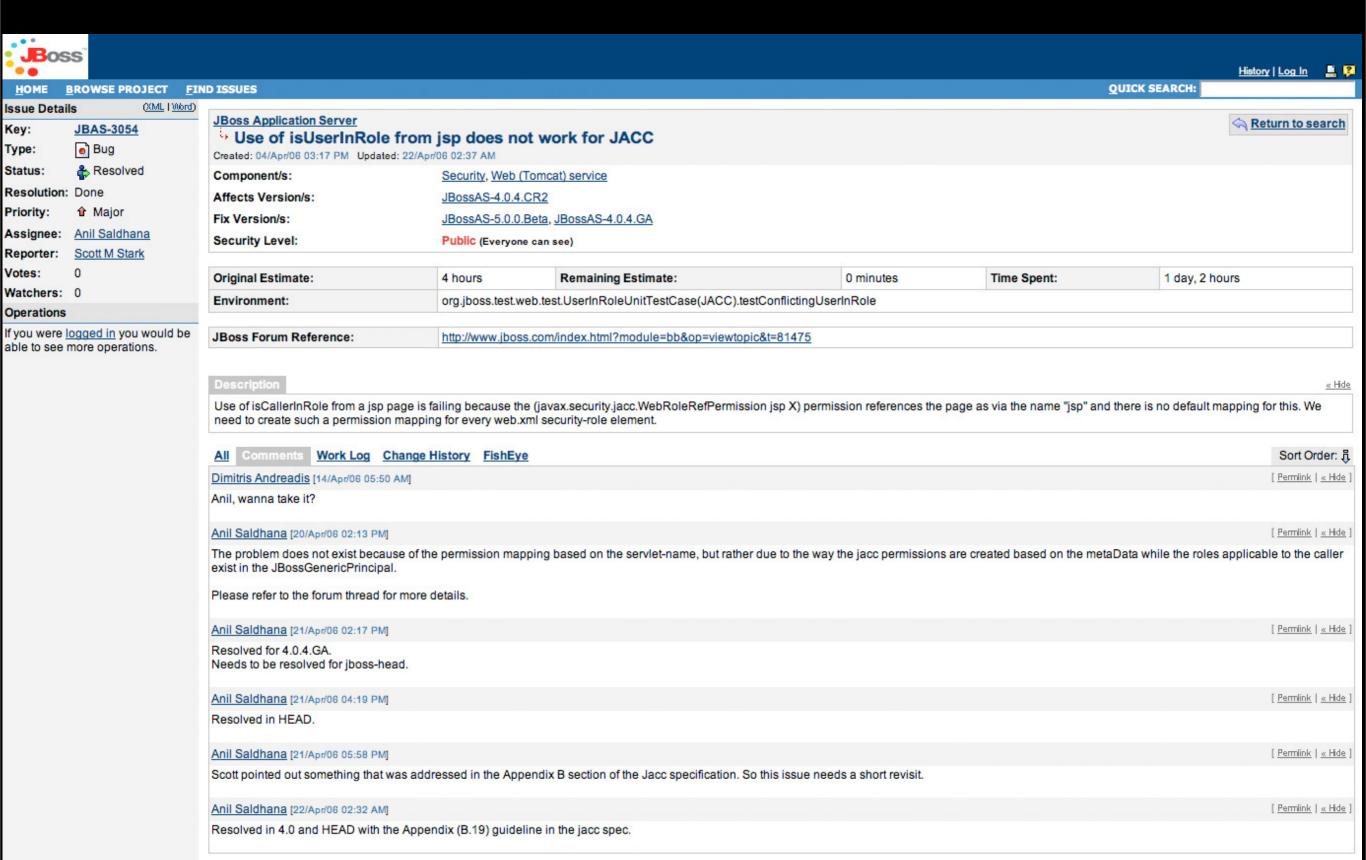


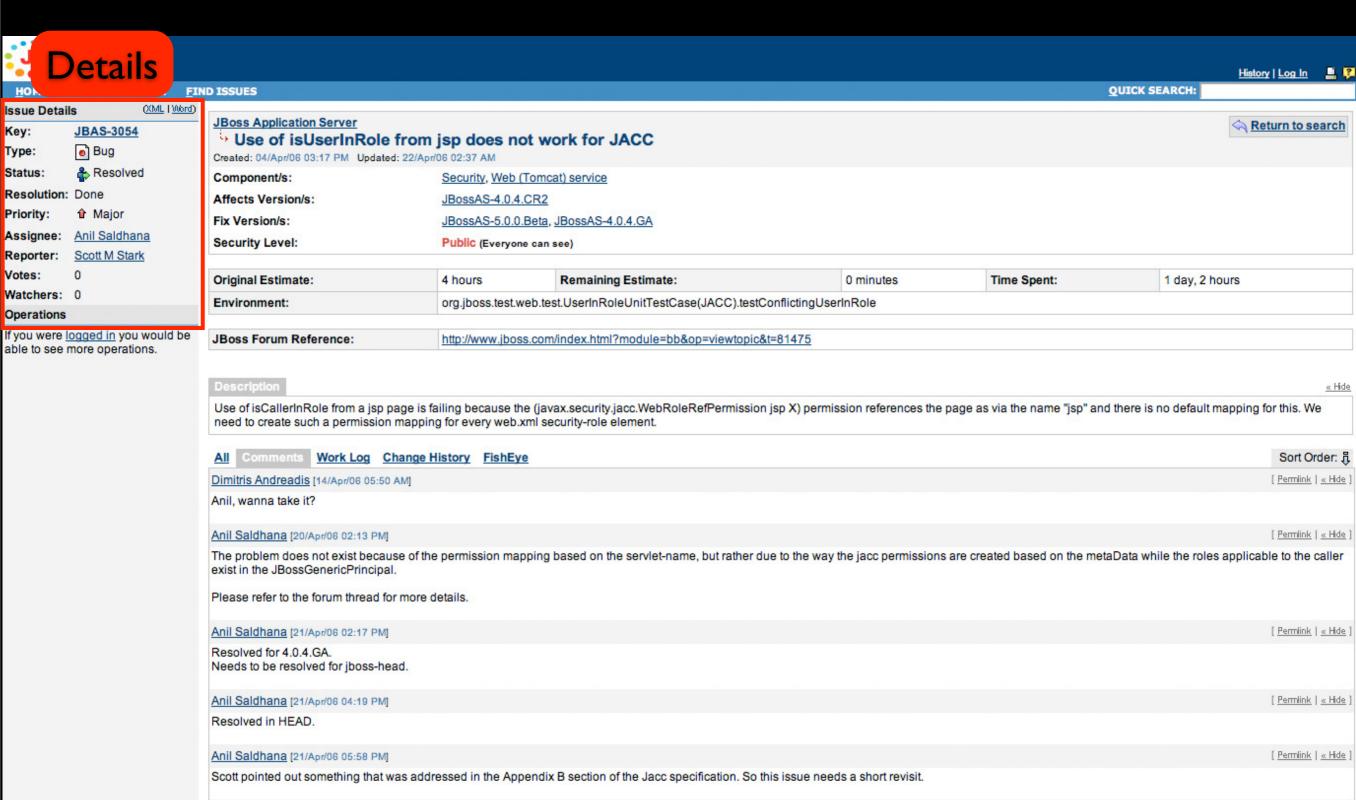
Bug Tracking, Issue Tracking and Project Management

Developed by Atlassian www.atlassian.com/software/jira



Developed by Atlassian www.atlassian.com/software/jira

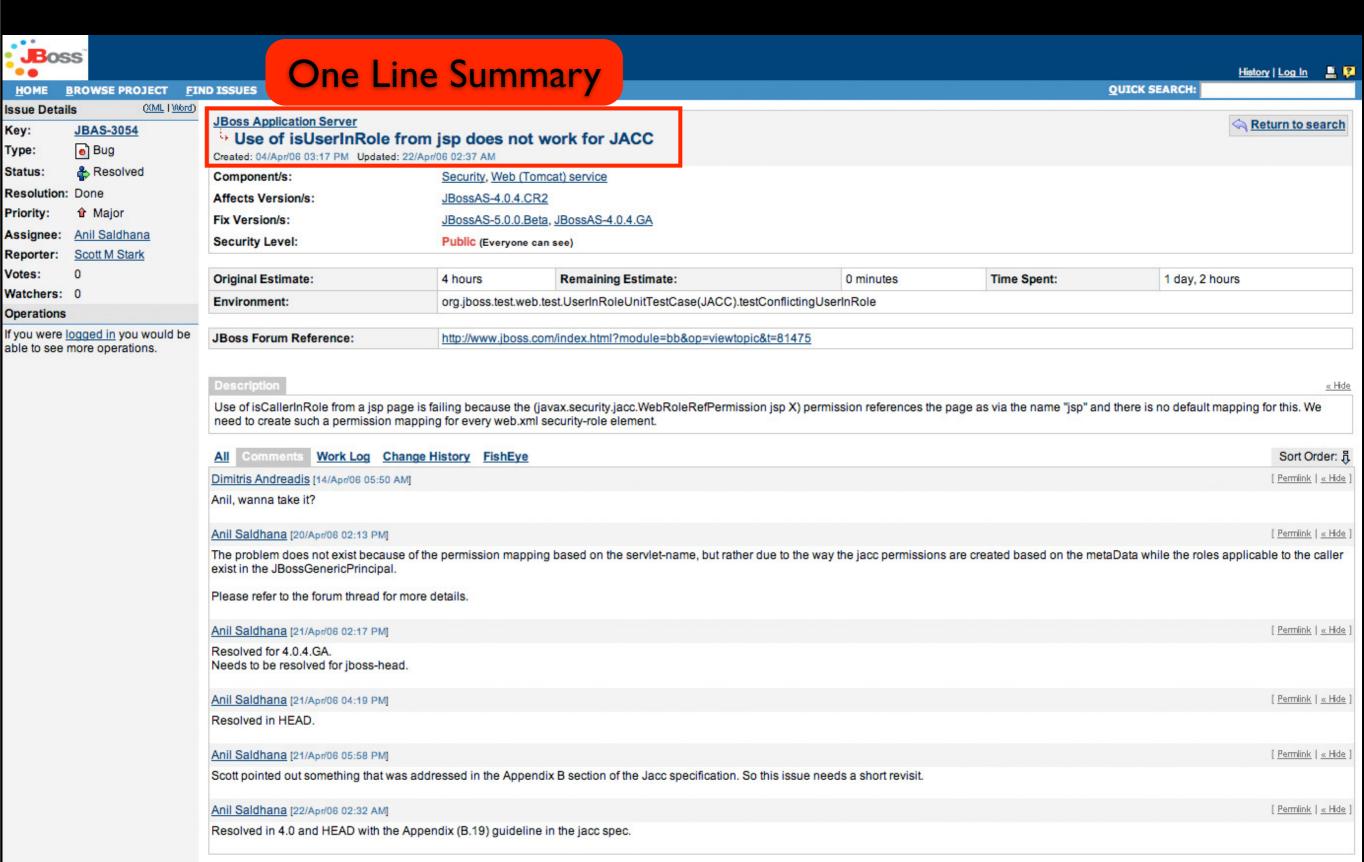


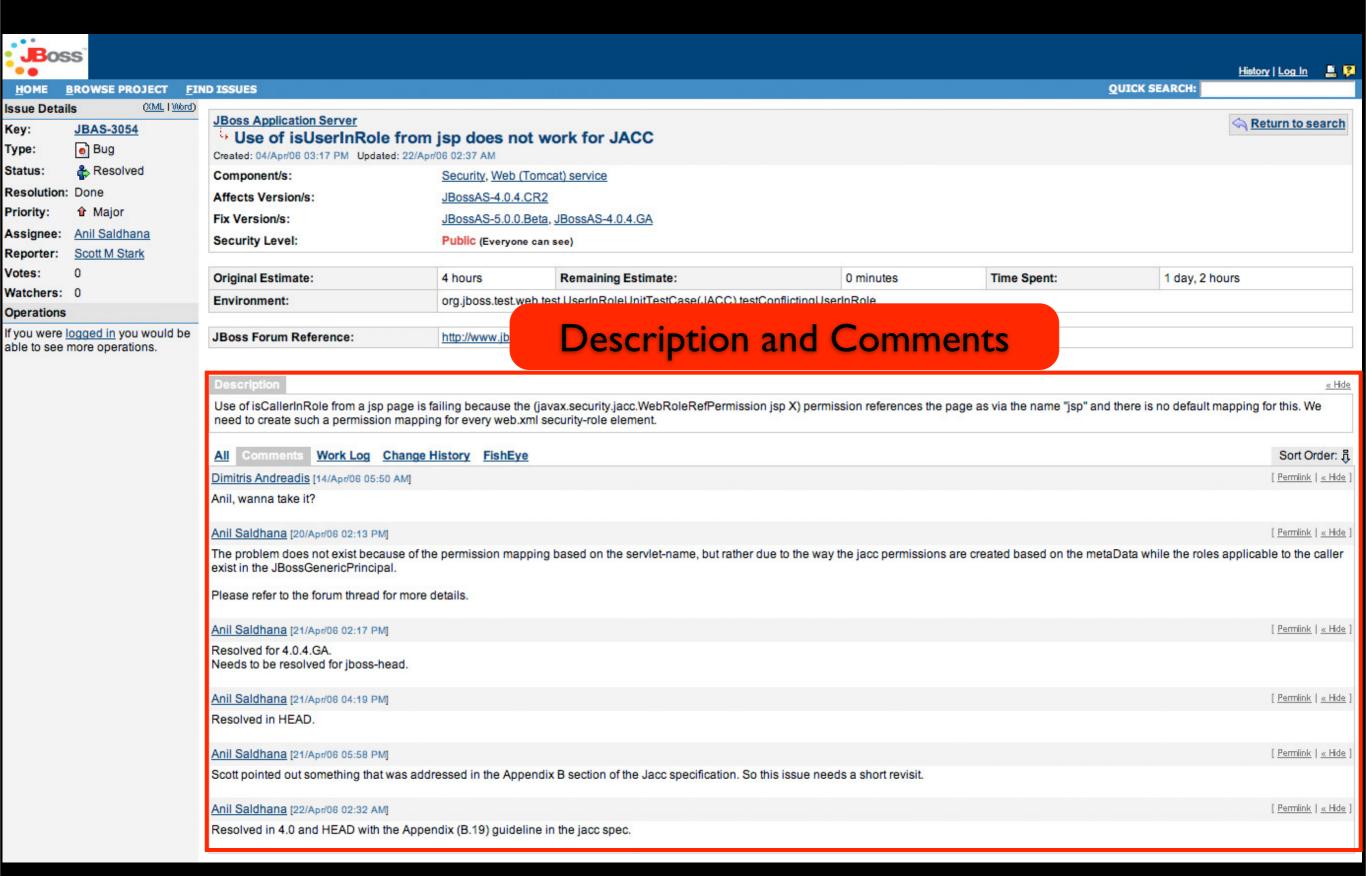


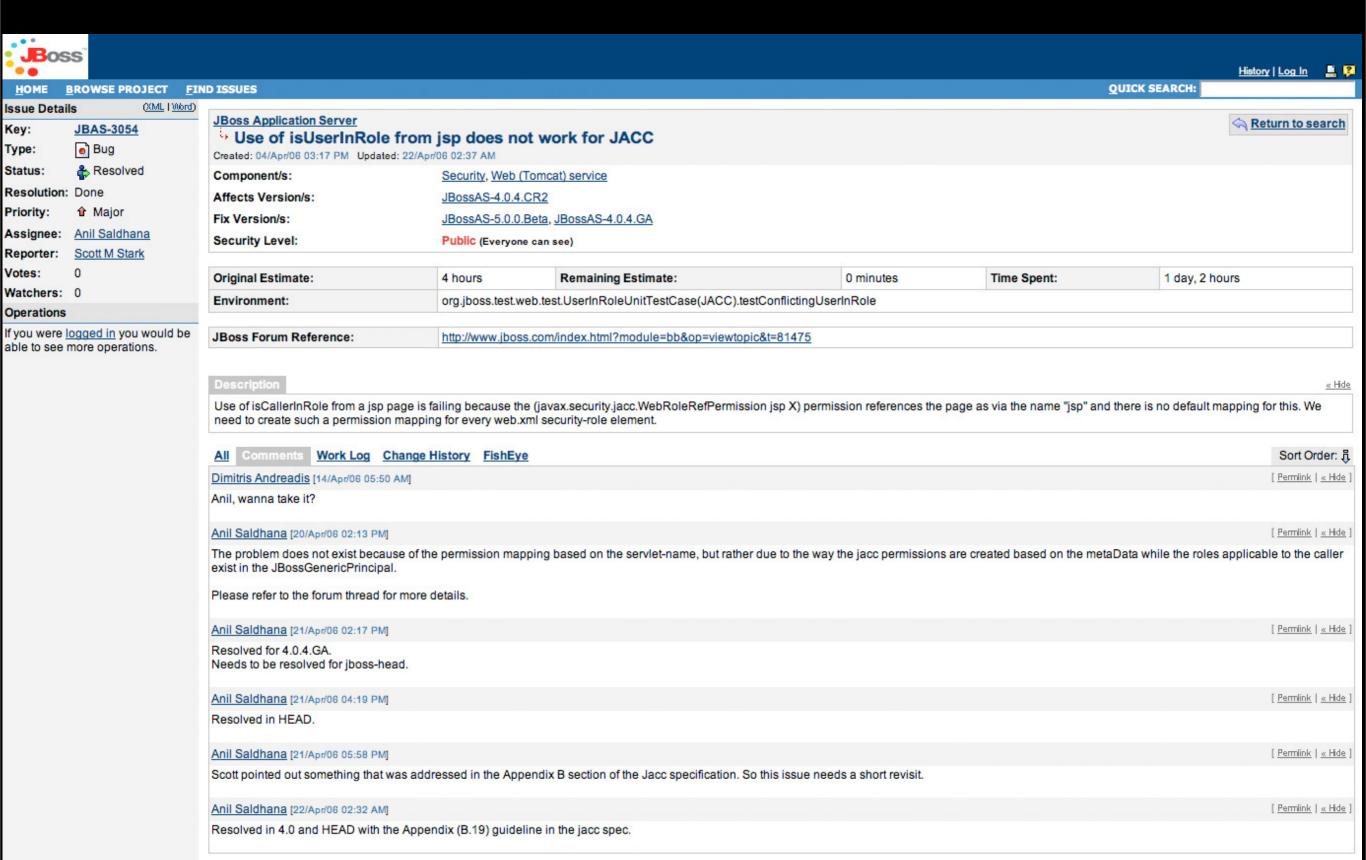
Anil Saldhana [22/Apr/06 02:32 AM]

Resolved in 4.0 and HEAD with the Appendix (B.19) guideline in the jacc spec.

Permlink | « Hide







Linking Bug Reports to Changes

teicher 2003-10-29 16:11:01

fixes issues mentioned in bug 45635: [hovering] rollover hovers

- mouse exit detection is safer and should not allow for loopholes any more, except for shell deactiviation
- hovers behave like normal ones:
 - tooltips pop up below the control
 - they move with subjectArea
 - once a popup is showing, they will show up instantly

TA-RE

An exchange language for mining software repositories

- Results are difficult to reproduce since different tools extract data using different methods.
- Benckmarking is virtually impossible.
- Other domains have repositories or corpora (plural of corpus) of data sets, e.g. PROMISE, UCI Repository, Reuters corpus.
- TA-RE presents software evolution data including
 CVS transactions and bug data in an XML format.

Summary

- Databases simplify the exploration of CVS.
- Sliding time windows are superior to fixed ones.
- Fine-grained analyses are feasible and worthwhile.
- Outlook: CVS mapped with bug data facilitates research in defect analysis and prediction, code clone investigation, cost estimation drivers, code decay and many more topics.