New Zealand's Electoral Roll Database

A PostgreSQL case study

Finlay Thompson - finlay@catalyst.net.nz

PostgreSQL Conference 2007 - Ottawa

New Zealand's electoral system

- MMP electoral system
- 120 members of parliament
- Westminster system, with only one house
- Elections every 3 years
- 70 electorates

Provision for Māori voters

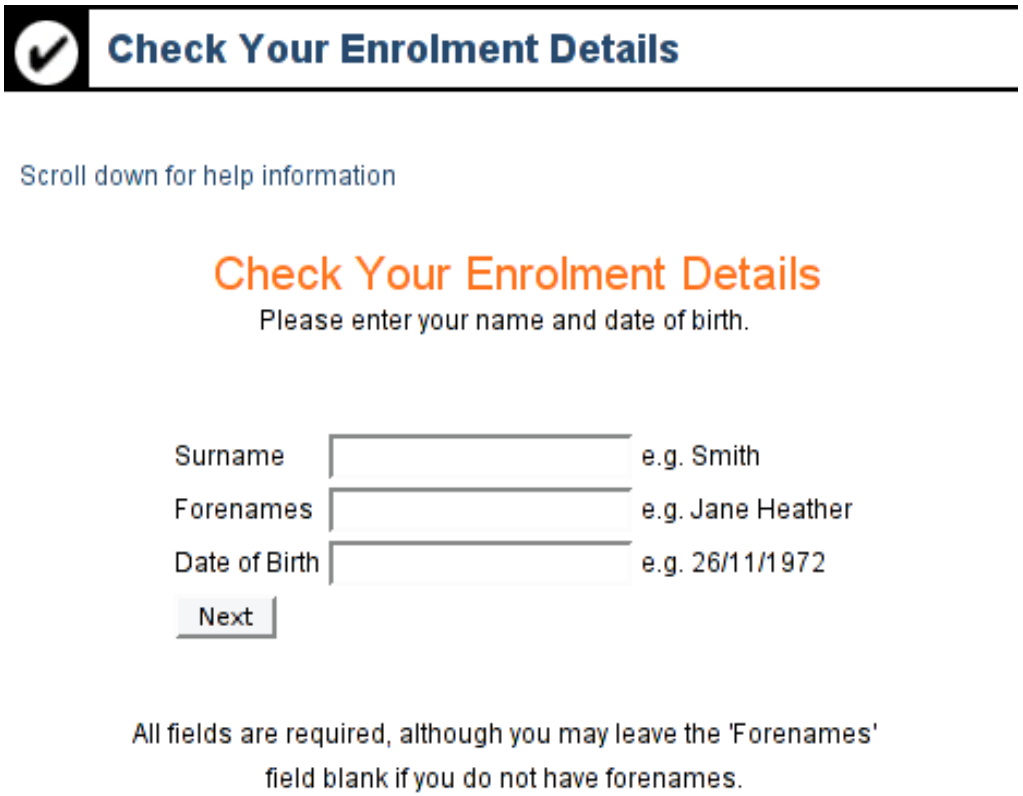
Māori Party leaders,
Tariana Turia & Pita Sharples

- Treaty of Waitangi in 1840
- Māori seats established permanently in 1876
- Two electoral rolls: General and Māori
- Māori option exercised every five years

Registrars of Electors

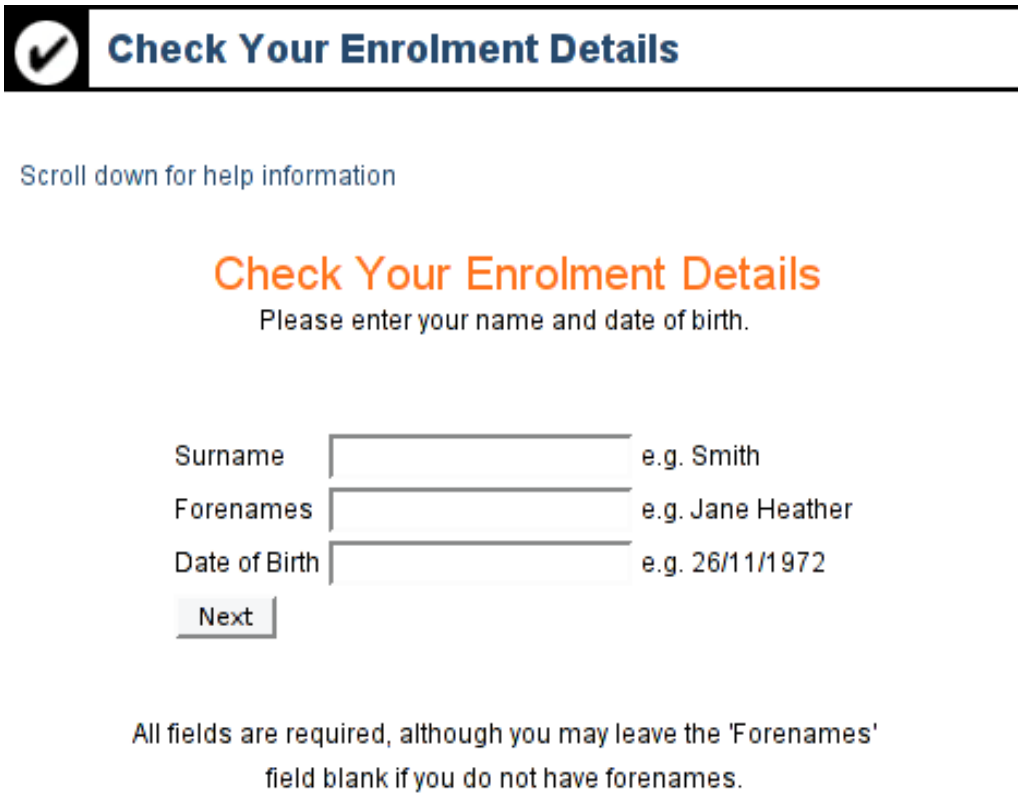
- 70 Registrars of electors maintain the electoral roll
- Legally the roll consists of signed forms stored in electoral offices!
- The Electoral Enrolment Centre supports the Registrars
- Offices located all over the country
- 2.9 million voters, 95.4% enrolled

Enrolling online



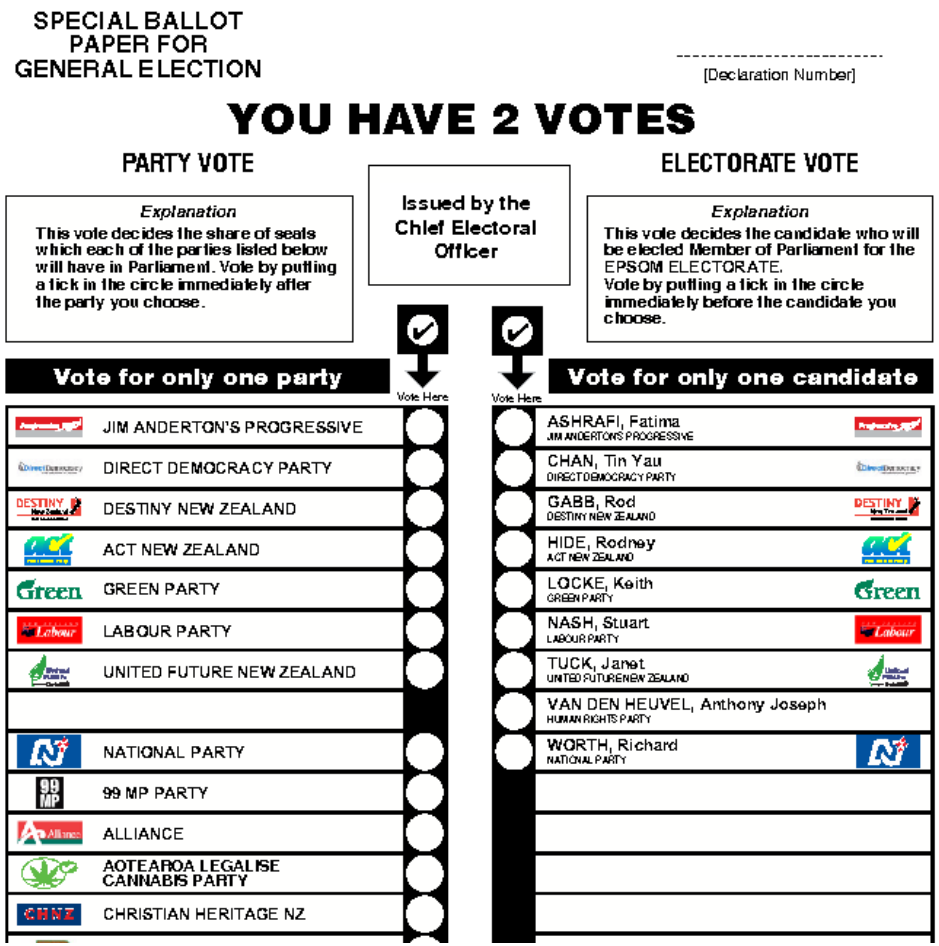
- Voters can check and update details on roll
- New enrolments made
- 81 thousand online transactions in 2005
- Abuse detection systems in place

Enrolling online



- Voters can check and update details on roll
- New enrolments made
- 81 thousand online transactions in 2005
- Abuse detection systems in place
- Voters not actually enrolled until signed form returned

Download Voting Papers



SPECIAL BALLOT PAPER FOR GENERAL ELECTION

[Declaration Number]

YOU HAVE 2 VOTES

PARTY VOTE

Explanation
This vote decides the share of seats which each of the parties listed below will have in Parliament. Vote by putting a tick in the circle immediately after the party you choose.

Issued by the Chief Electoral Officer

ELECTORATE VOTE

Explanation
This vote decides the candidate who will be elected Member of Parliament for the EPSOM ELECTORATE.
Vote by putting a tick in the circle immediately before the candidate you choose.

Vote for only one party

Vote Here

- JIM ANDERTON'S PROGRESSIVE
- DIRECT DEMOCRACY PARTY
- DESTINY NEW ZEALAND
- ACT NEW ZEALAND
- GREEN PARTY
- LABOUR PARTY
- UNITED FUTURE NEW ZEALAND
- NATIONAL PARTY
- 99 MP PARTY
- ALLIANCE
- AOTEAROA LEGALISE CANNABIS PARTY
- CHRISTIAN HERITAGE NZ

Vote Here

Vote for only one candidate

- ASHRAFI, Fatima — JIM ANDERTON'S PROGRESSIVE
- CHAN, Tin Yau — DIRECT DEMOCRACY PARTY
- GABB, Rod — DESTINY NEW ZEALAND
- HIDE, Rodney — ACT NEW ZEALAND
- LOCKE, Keith — GREEN PARTY
- NASH, Stuart — LABOUR PARTY
- TUCK, Janet — UNITED FUTURE NEW ZEALAND
- VAN DEN HEUVEL, Anthony Joseph — HUMAN RIGHTS PARTY
- WORTH, Richard — NATIONAL PARTY

- Forms for making special votes available online
- Eligibility checked
- Especially used by overseas voters

Maintaining the roll

Information direct from the public

- New enrolments come in the through the post
- Update forms when people change details
- Direct update campaigns
- Approximately 1.5 million new records / edits each year

Maintaining the roll

Information direct from the public

- New enrolments come in the through the post
- Update forms when people change details
- Direct update campaigns
- Approximately 1.5 million new records / edits each year

Data match campaigns

- Other government agencies provide data
- Potential inaccuracies are identified
- Correspondence generated to matched people
- 179 thousand data matches in March 2007

Primary uses for electoral data

Roll productions

- Rolls get printed on a regular schedule
- Provided to libraries, political parties, etc.
- Election day rolls (general and local elections)

Primary uses for electoral data

Roll productions

- Rolls get printed on a regular schedule
- Provided to libraries, political parties, etc.
- Election day rolls (general and local elections)

Special vote checking

- some people have a good reason not to be published
- In order to vote they need to make a special declaration
- Special votes are checked against the electoral roll data in the days after an election

Other uses for electoral data

Nationwide statistical analysis

- The country is divided into land parcels called mesh blocks
- There are approx. 33,000 mesh blocks covering the country
- Electoral roll data is shared with the Dept. of Statistics to allow them to maintain the mesh blocks

Other uses for electoral data

Nationwide statistical analysis

- The country is divided into land parcels called mesh blocks
- There are approx. 33,000 mesh blocks covering the country
- Electoral roll data is shared with the Dept. of Statistics to allow them to maintain the mesh blocks

Postal address sanitation

- NZ Post keeps a database of all addresses
- Data matching and sharing helps ensure accuracy in both data sets

Why we are using PostgreSQL

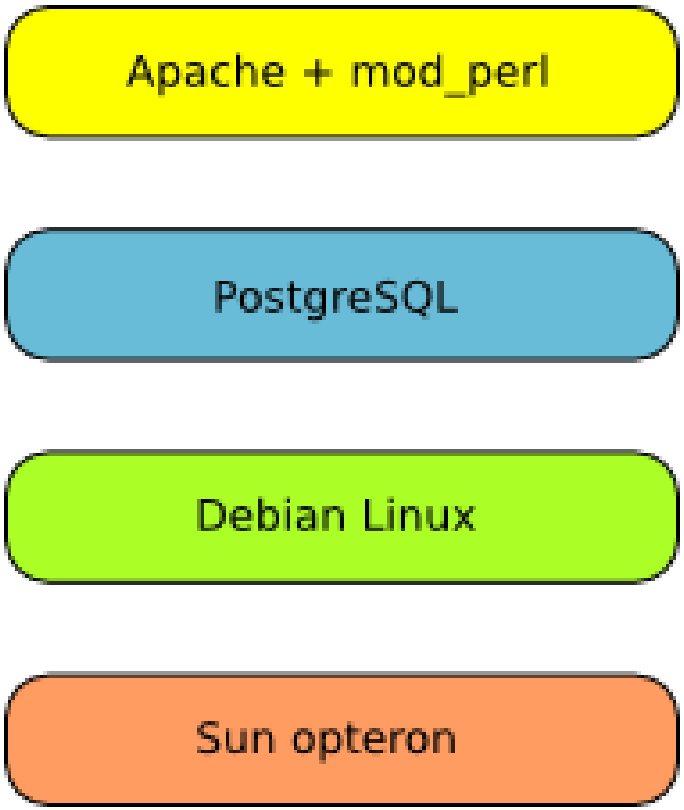
Life before PostgreSQL

- Distributed database with overnight synchronisation
- A client / server architecture
- Oracle databases, Visual Basic code base

Life before PostgreSQL

- Distributed database with overnight synchronisation
- A client / server architecture
- Oracle databases, Visual Basic code base
- System was unresponsive
- Needed to be replaced

Making a switch to open source



New architecture

- Centralised database with increased data integrity
- Internet to provide web based access over ADSL modems
- Linux, PostgreSQL, Apache and Perl

Choosing open source

Cons

- New direction for organisation
- No big international company backing
- Considered risky

Choosing open source

Cons

- New direction for organisation
- No big international company backing
- Considered risky and unusual

Choosing open source

Cons

- New direction for organisation
- No big international company backing
- Considered risky and unusual

Pros

- Established track record naging the SRS system
- No license costs or issues
- Flexibility of deployment and generic hardware

Choosing open source

Cons

- New direction for organisation
- No big international company backing
- Considered risky and unusual

Pros

- Established track record naging the SRS system
- No license costs or issues
- Flexibility of deployment and generic hardware
- Cost effective

Choosing PostgreSQL

- Completely open source license
- Strong support for data constraints
- Transaction and concurrency scaling
- Replication using RServ (soon Slony)
- Active and friendly development community
- Plays happily on debian linux

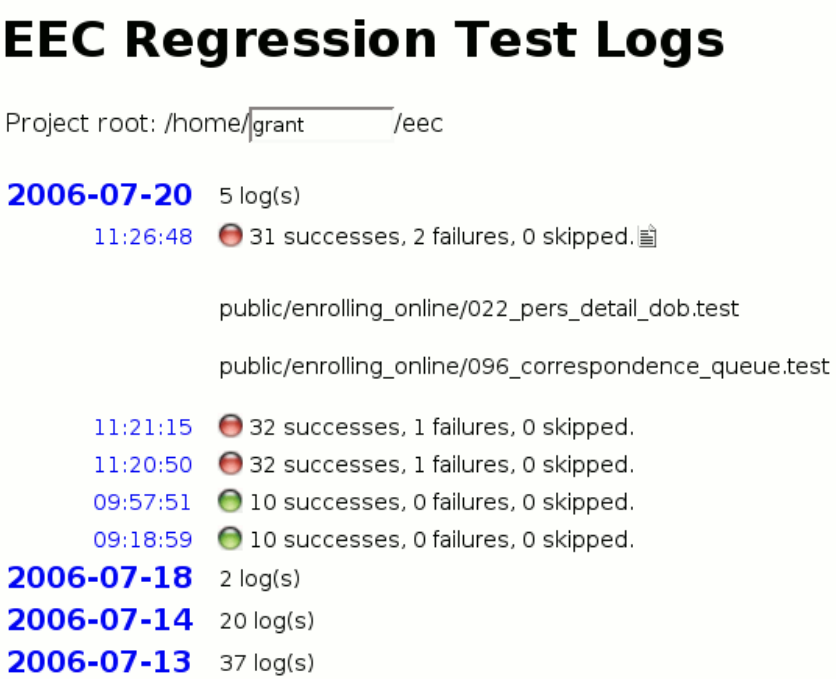
Living with constant development

- Changes to business rules
- New feature requirements
- Increases in data volumes

Living with constant development

- Changes to business rules
- New feature requirements
- Increases in data volumes
- Constant change!

Continuous integration



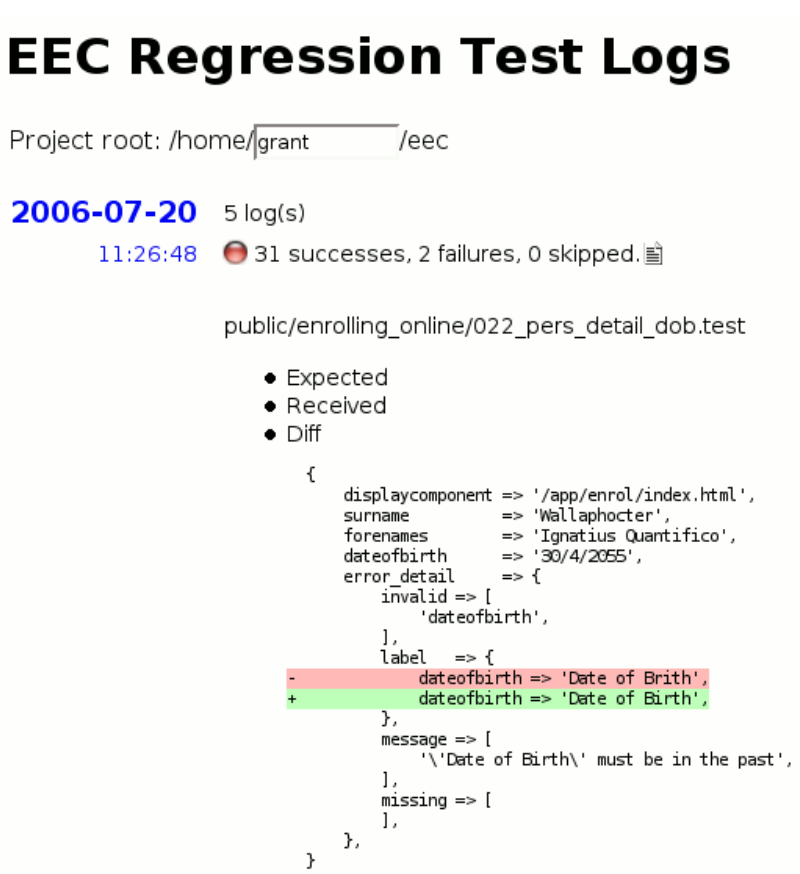
EEC Regression Test Logs

Project root: /home/grant /eec

2006-07-20 5 log(s)
11:26:48 31 successes, 2 failures, 0 skipped.
public/enrolling_online/022_pers_detail_dob.test
public/enrolling_online/096_correspondence_queue.test
11:21:15 32 successes, 1 failures, 0 skipped.
11:20:50 32 successes, 1 failures, 0 skipped.
09:57:51 10 successes, 0 failures, 0 skipped.
09:18:59 10 successes, 0 failures, 0 skipped.
2006-07-18 2 log(s)
2006-07-14 20 log(s)
2006-07-13 37 log(s)

- Regression tests get run automatically
- Code cannot get deployed until all tests are passing
- Tests are run continuously, all day, every day

Continuous integration



EEC Regression Test Logs

Project root: /home/grant /eec

2006-07-20 5 log(s)

11:26:48 31 successes, 2 failures, 0 skipped.

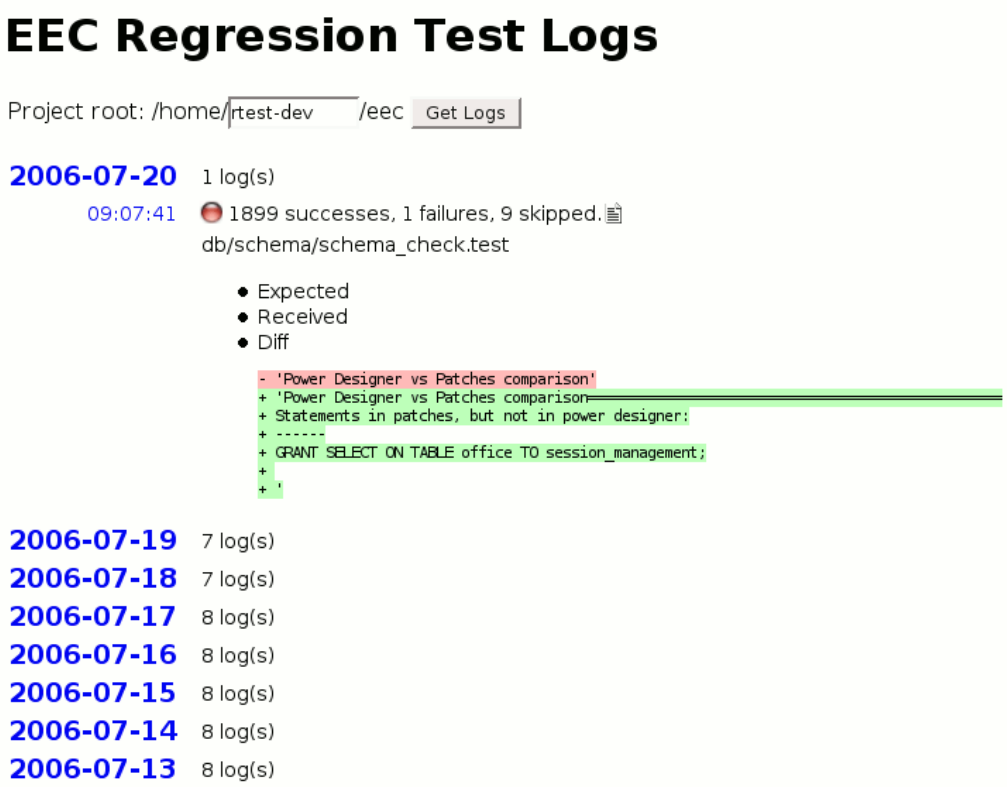
public/enrolling_online/022_pers_detail_dob.test

- Expected
- Received
- Diff

```
{
    displaycomponent => '/app/enrol/index.html',
    surname          => 'Wallaphocter',
    forenames        => 'Ignatius Quantifico',
    dateofbirth      => '30/4/2055',
    error_detail     => {
        invalid => [
            'dateofbirth',
        ],
        label   => {
-           dateofbirth => 'Date of Brith',
+           dateofbirth => 'Date of Birth',
        },
        message => [
            '\'Date of Birth\' must be in the past',
        ],
        missing => [
        ],
    },
}
```

- Regression tests get run automatically
- Code cannot get deployed until all tests are passing
- Tests are run continuously, all day, every day
- Developers are sent notification of test failures

Continuous integration



EEC Regression Test Logs

Project root: /home/rtest-dev /eec Get Logs

2006-07-20 1 log(s)

09:07:41 1899 successes, 1 failures, 9 skipped.

db/schema/schema_check.test

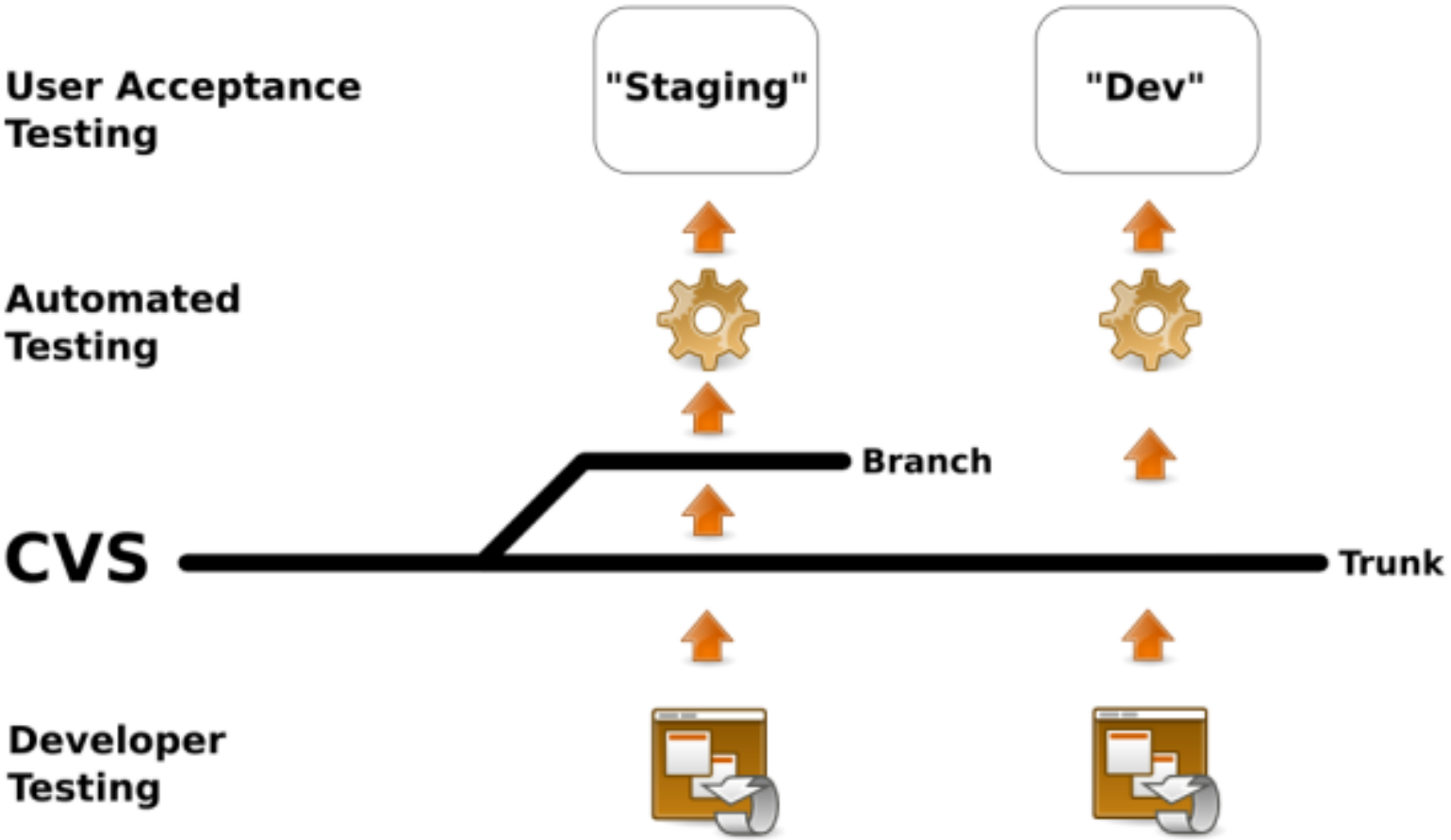
- Expected
- Received
- Diff

```
- 'Power Designer vs Patches comparison'
+ 'Power Designer vs Patches comparison
+ Statements in patches, but not in power designer:
+ ------
+ GRANT SELECT ON TABLE office TO session_management;
+
+ '
```

2006-07-19 7 log(s)
2006-07-18 7 log(s)
2006-07-17 8 log(s)
2006-07-16 8 log(s)
2006-07-15 8 log(s)
2006-07-14 8 log(s)
2006-07-13 8 log(s)

- Regression tests get run automatically
- Code cannot get deployed until all tests are passing
- Tests are run continuously, all day, every day
- Developers are sent notification of test failures
- Name and shame

Continuous integration



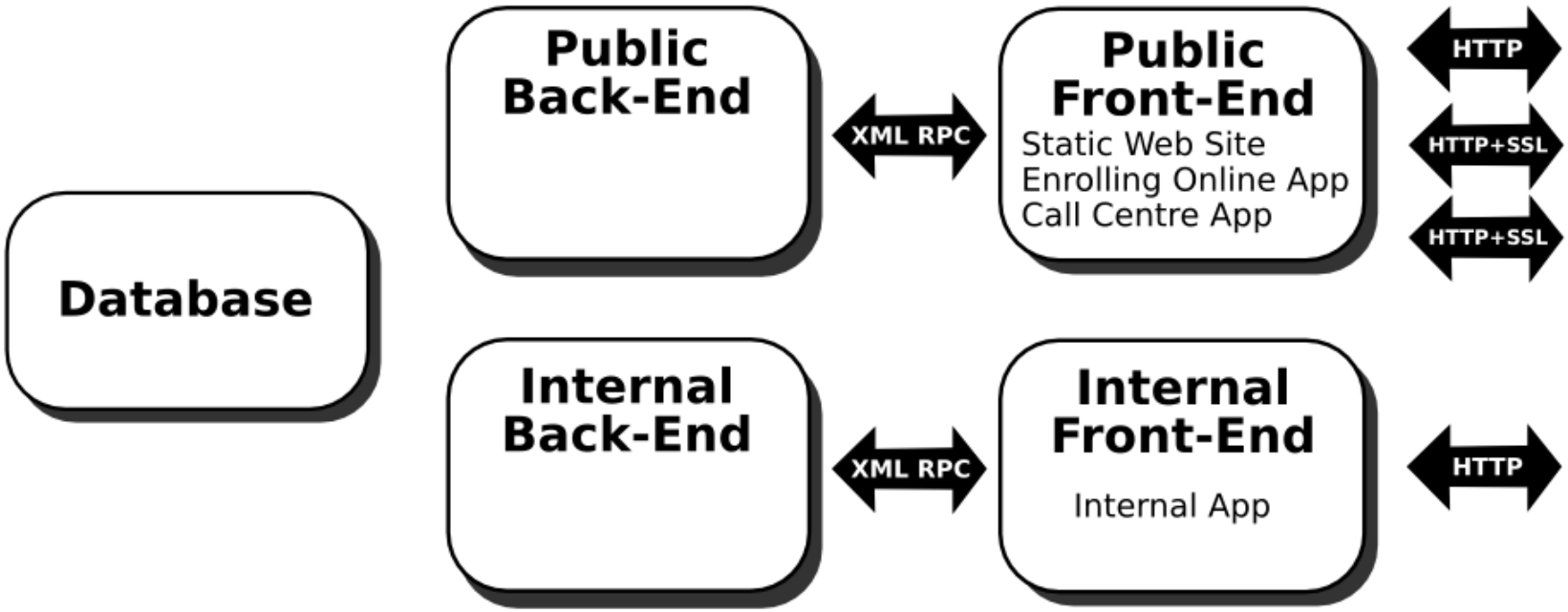
Continuous integration

- Makes development faster and easier
- Increased confidence to make big changes to code base
- Works against code rot

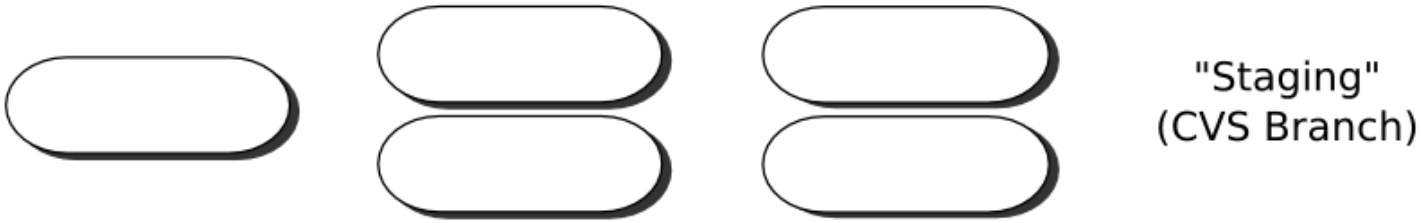
Continuous integration

- Makes development faster and easier
- Increased confidence to make big changes to code base
- Works against code rot
- Provides the required level of stability

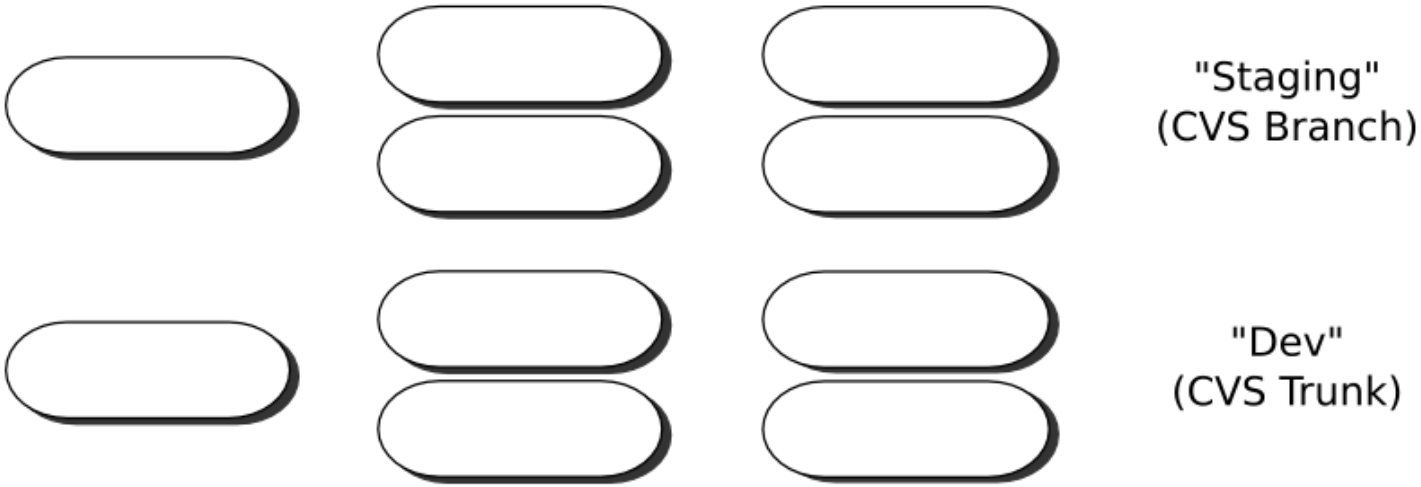
System architecture



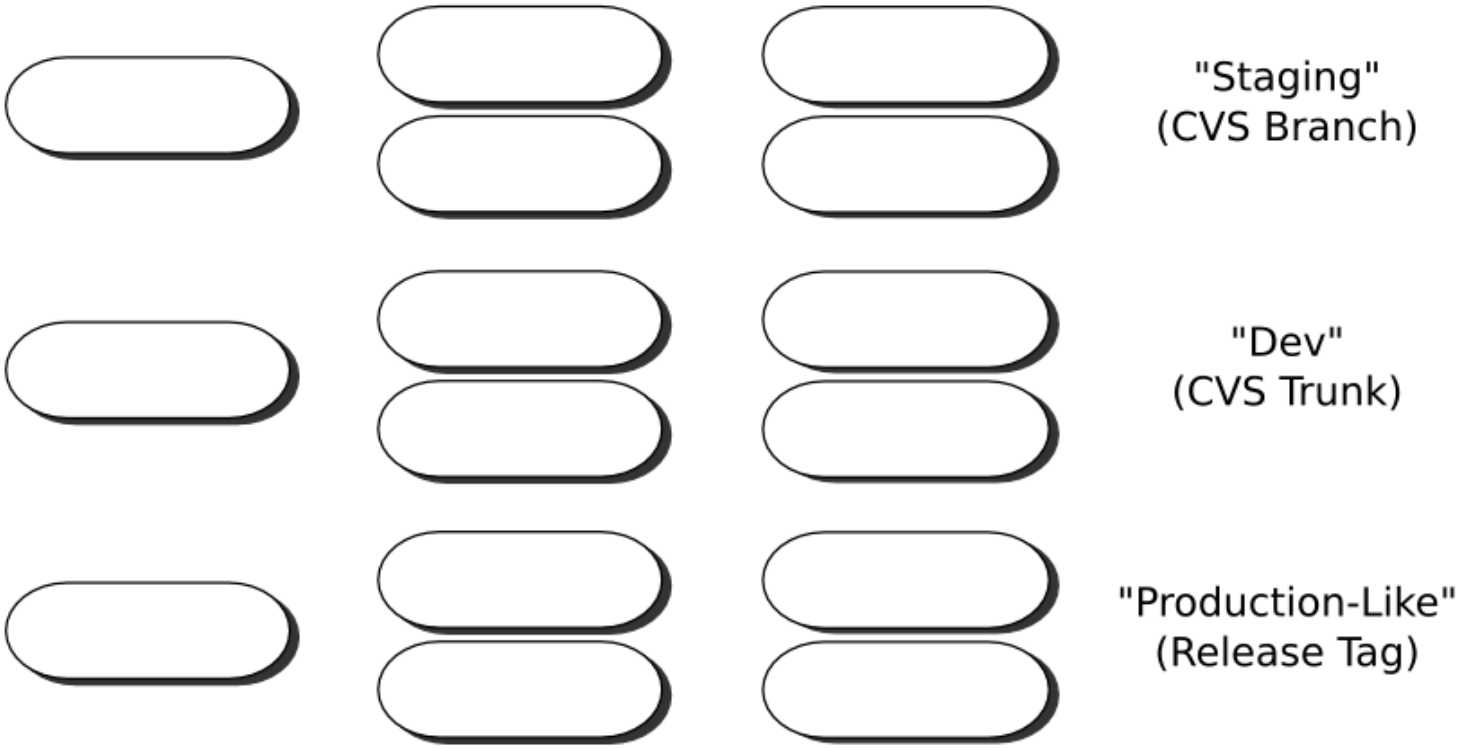
User acceptance testing



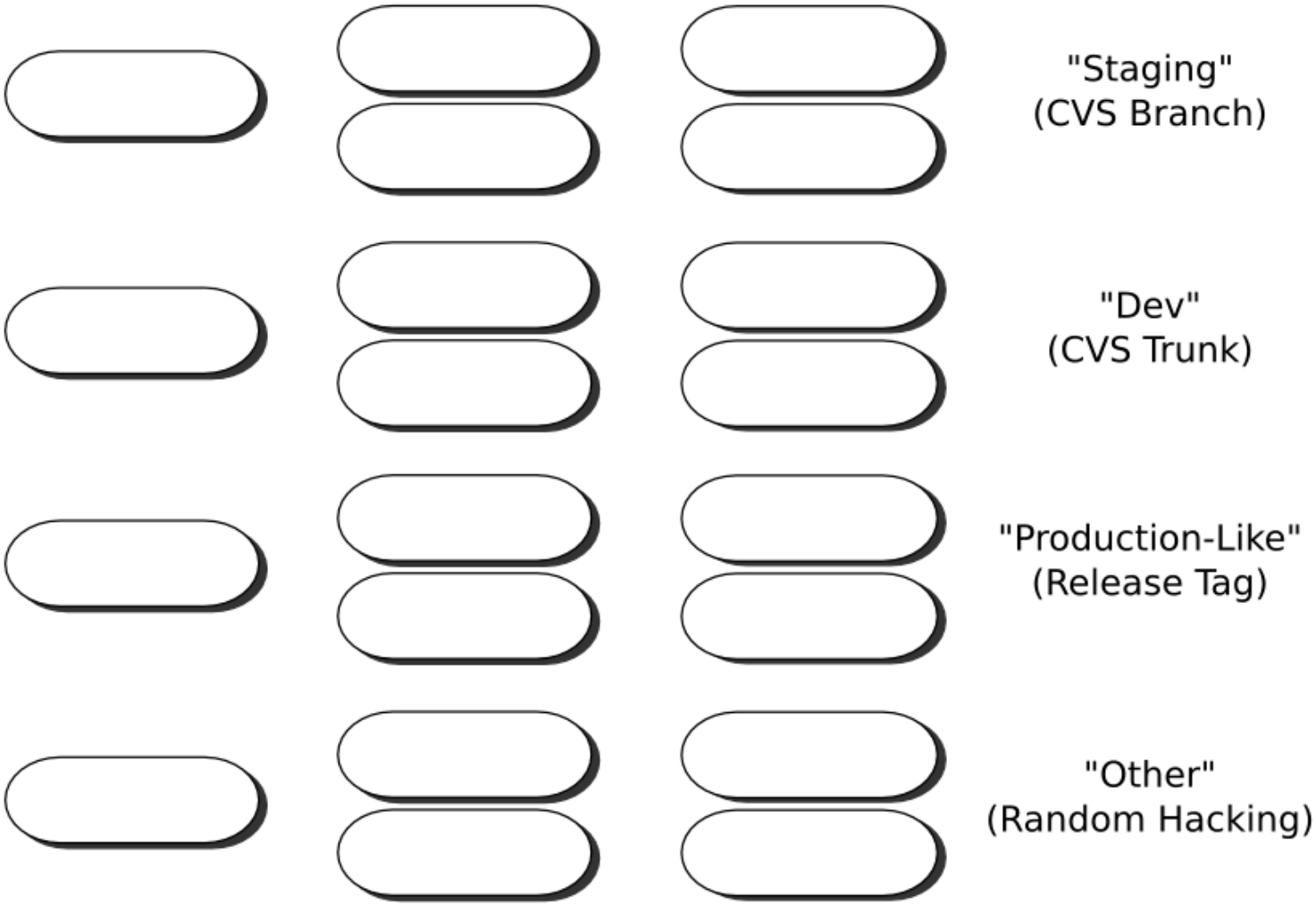
User acceptance testing



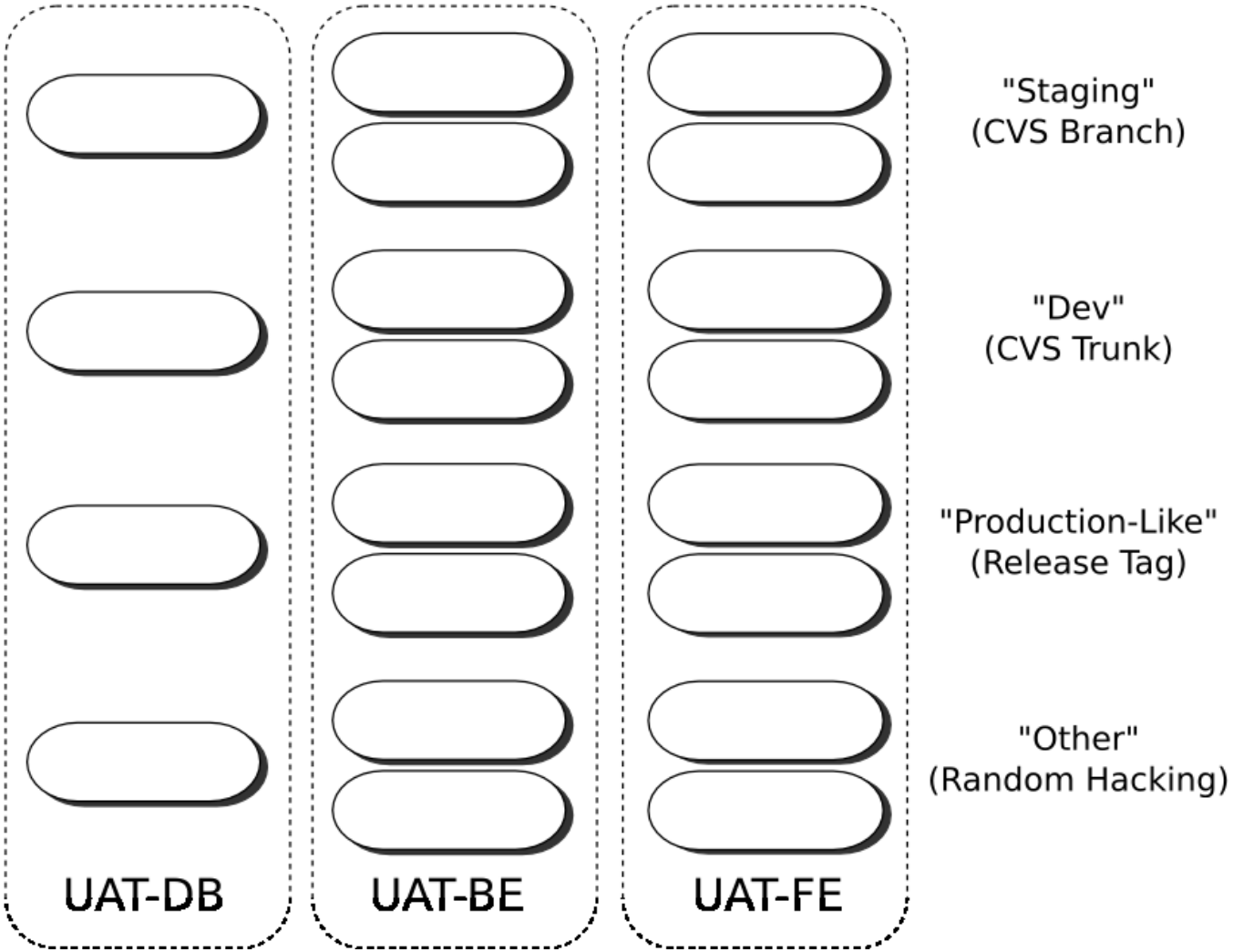
User acceptance testing



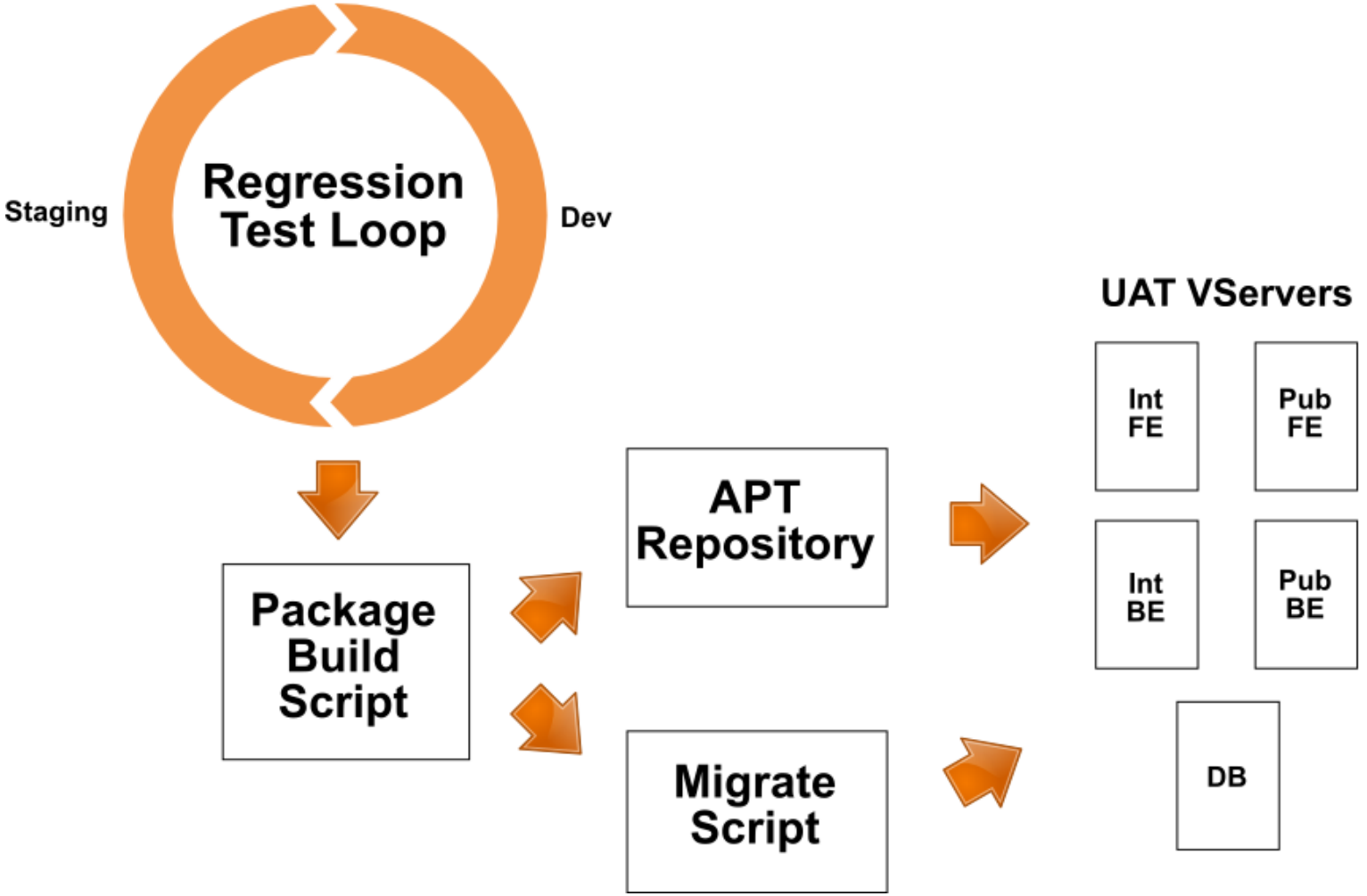
User acceptance testing



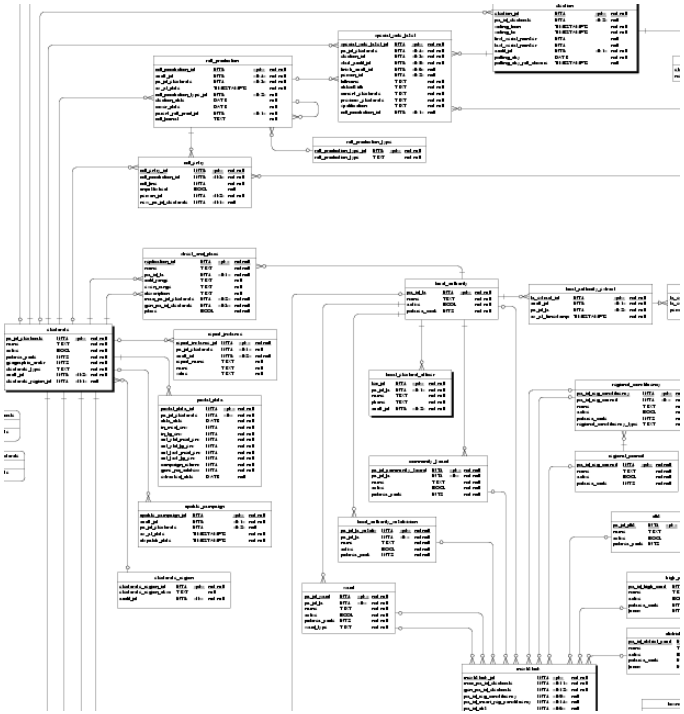
Linux VServers



User acceptance testing

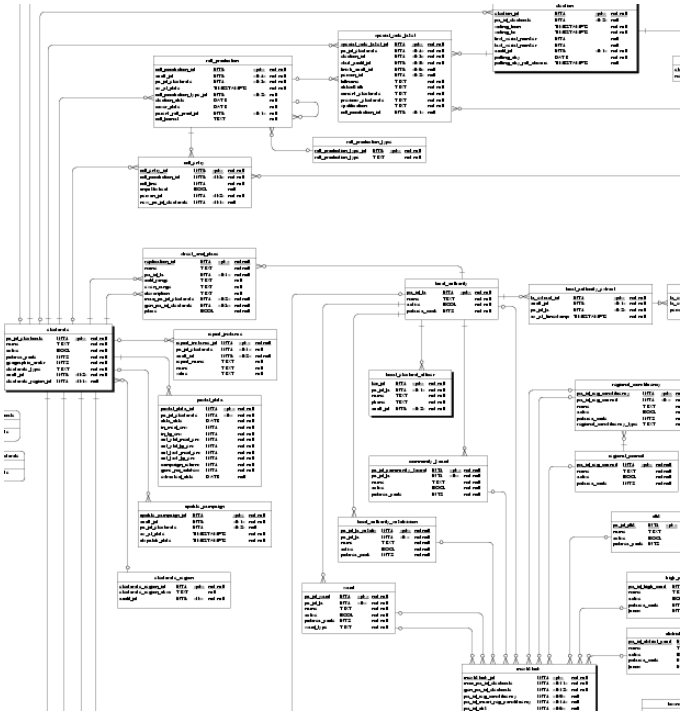


Database schema patching



- Approx. 300 database objects
- Rolling database patches
- Patches checked in with associated code
- Modeled in Power Designer

Database schema patching



- Approx. 300 database objects
- Rolling database patches
- Patches checked in with associated code
- Modeled in Power Designer
- Regression test the model

The configuration problem is hard

- Complexity is the enemy of stability
- 36 Computers
- 25 Virtual Linux servers

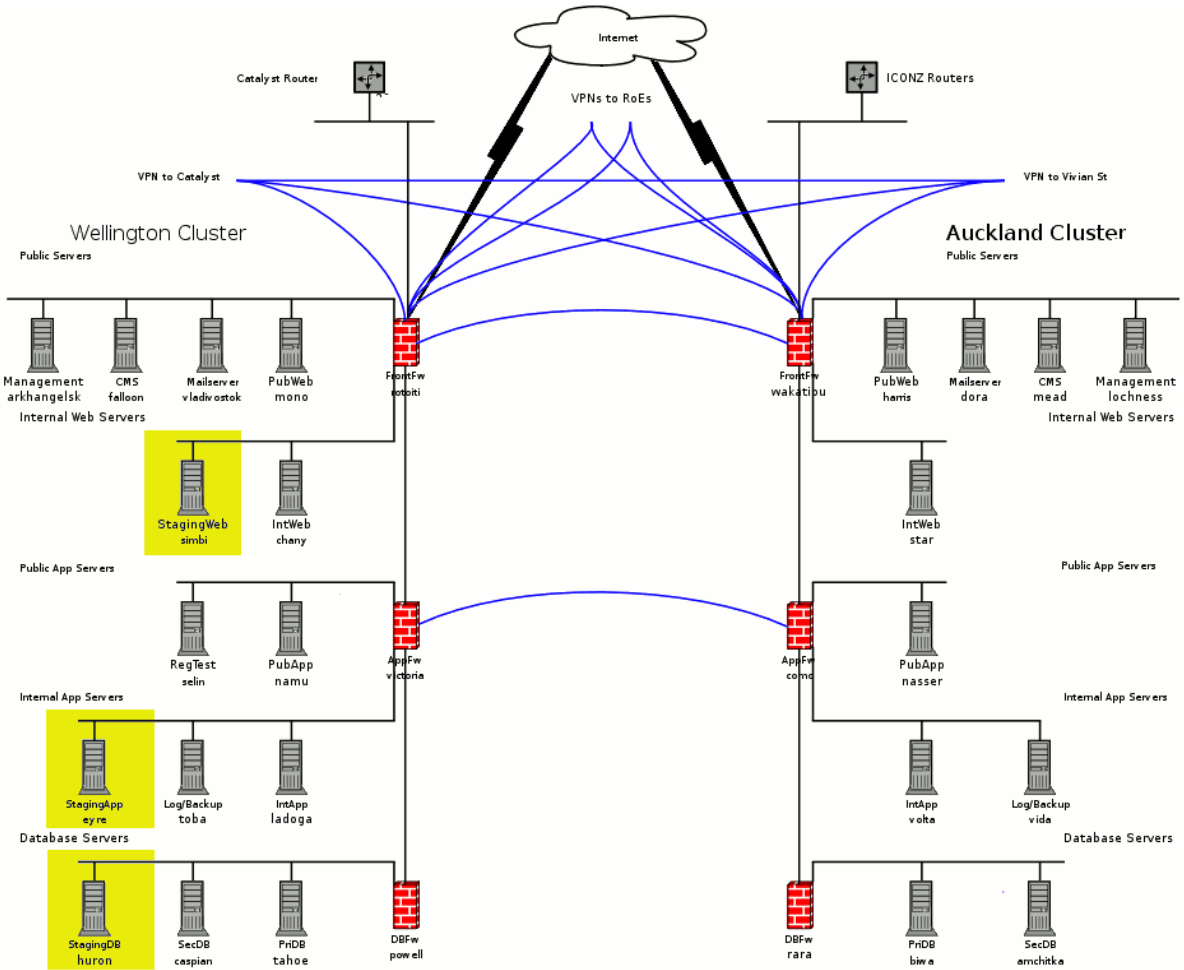
The configuration problem is hard

- Complexity is the enemy of stability
- 36 Computers
- 25 Virtual Linux servers
- Many different software packages
- Hundreds of configuration settings

The configuration problem is hard

- Complexity is the enemy of stability
- 36 Computers
- 25 Virtual Linux servers
- Many different software packages
- Hundreds of configuration settings
- Make it automatic!

The configuration problem is hard



The configuration problem is hard

Automatic installation and update

- All changes start with updated code
- Automatically applied in regression testing and user acceptance environments
- Identical code applied in production

The configuration problem is hard

Automatic installation and update

- All changes start with updated code
- Automatically applied in regression testing and user acceptance environments
- Identical code applied in production

Tools

- FAI - fully automatic installation (debian)
- APT - lots of custom debian packages
- CFengine - cluster aware configuration
- CVS - code repository

Conclusions

Planning for constant change

- Need to expect requirements to constantly change
- Computer systems are complex
- Automatic testing and deployment help manage risk

Conclusions

Planning for constant change

- Need to expect requirements to constantly change
- Computer systems are complex
- Automatic testing and deployment help manage risk

Using open source tools

- Flexibility to make decisions according to technical requirements, not license agreements
- Not locked into particular vendor
- Security in being able to fix problems ourselves

Conclusions

Electoral Enrolment Centre

- Remains a reliable and effective system through many changes
- On going costs directed at improvement not licenses
- Medium to long term security in the technology