Data Mining UV TechPlat

- Intro to the University of Valencia
- UV's Technological Platform
- The problem to solve
  - Catalog & Document all APPs, services and their underlying infrastructure
- The solution
- Practical use scenarios

Josep Vidal
System Programmer

University of Valencia

Founded in 1499

IT services since 1955

University of Valencia

- Located in an autonomous region with his own language and culture

Ciutat de les Arts
I les Ciències

Pilota
Valenciana

Falles

University of Valencia

Some numbers

- Students Enrolled:
  - 46,000 in diplomas, degrees and engineering
  - 4,400 in masters and doctorates
  - 4,400 in postgraduate studies.
- Staff:
  - 3,400 teachers and researchers
  - 1,800 technicians and administrative and services staff
- Internationalization
  - Has 2,000 students in undergraduate studies with 26% of master's students coming from institutions from around the world.

University of Valencia

- Rankings

- Shanghai Jiao TongUniversity Ranking:
  - The UNIVERSITAT is ranked as the 4th highest Spanish university and places between the top 201 and 300 best universities in the world.
- Erasmus Mobility
  - The UNIVERSITAT is the **2nd university in Europe in receiving** students and the 8th university in Europe in sending students abroad.
- Spanish foundation for science and technology
  - The UNIVERSITAT is the 4th Spanish university in research.
- It is one of the top ten universities in the world to **learn Chinese** language and culture, according to the Chinese government.
- The UV provides more work placements/internships in businesses and institutions in Valencia: 6,500 students participate through agreements with over 2,400 businesses and institutions.
- The UV has made a big commitment to solar energy: the **first urban solar park in Spain**.

VNIVERSITAT DE VALÈNCIA

- Wide range of studies
  - Basic and experimental sciences
  - Social, economic and legal sciences
  - Educational Sciences
  - Health sciences
  - Engineering
  - Humanities

UV Technological Platform

UV demands for a lot of IT resources, services and applications that have been built in the last 35 years. We like to call it UV Technological Platform

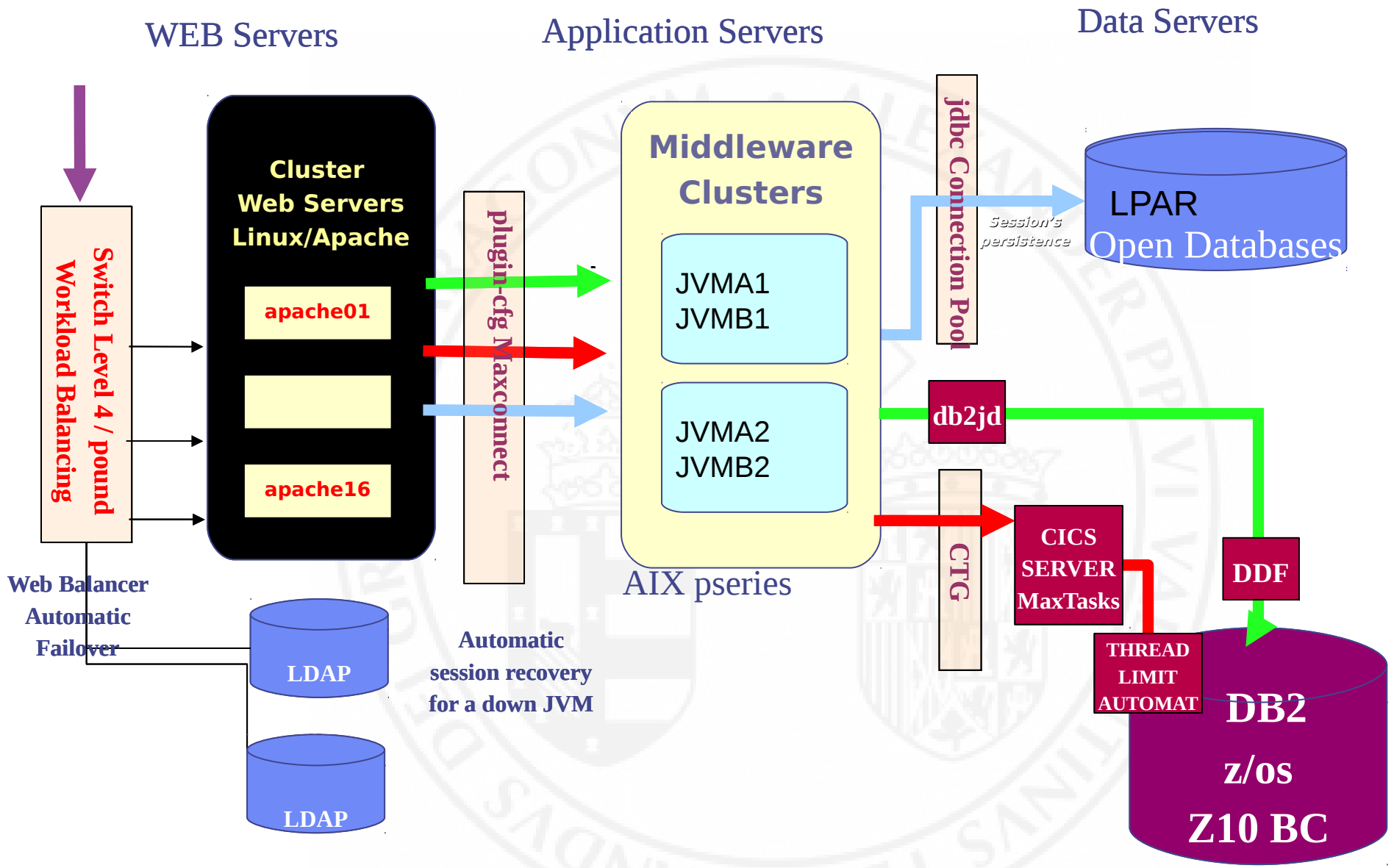
UV Technological Platform

- 4 Millions URLs served with a 0.3 seconds response time
- 120 Java Enterprise Applications with 200K web components used to run academic business (students records, payroll, research, accounting, etc ..)
- 4K online COBOL/CICS/DB2 business logic transactions
- 500 batch programs
- 50 millions of COBOL lines generated by EGL (business logic)
- 15 millions of documents generated by APPs

UV Technological Platform

- 40 TB of disc spread between 2 storage arrays
- 5 application servers but 95% of APPs running in WebSphere
- Main databases: DB2 z/os (65 % Apps) + Oracle (35% Apps)
  - 12K + database tables and thousand of millions of records.
  - 40 different datasources (User/Server/Instance/Port)
  - **First student records dates back 1980**
- 40 IT workers daily accessing systems, apps & data (DBAs, operators, system & app programmers, analyst, etc ..) + outsourced app providers employees
- 15M of documents, temporal files, configurations … stored in a shared filesystem
- 3 main environments: DEV, TEST & PROD running in the same HW (x86-64,pseries, zseries) but isolated by virtualization, WLMs, security (RACF)

General Architecture



UV Technological Platform

Problem to solve

- Need for documentation to organize IT service
  - Develop the Catalog of IT services
  - Protocolize apps promotion from test to production
  - Evaluate the migration of the technological platform to other architectures
- BUT …
  - UV IT pioneers started in the late 70's to develop information systems when computing engineering wasn't mature enough
  - Back then knowledge of our TechPlat fitted in the mind of our IT workers
- WAIT: All is up & running. So why not use data mining to find apps, data, infrastructure & dependences?

Problem to solve

- Challenge: To document APPs to know
  - Which (distributed) components conform an enterprise application?
  - Which infrastructure uses?
    - Necessary for management to authorize intervention over production infrastructure (servers, storage, middleware, DB objects, etc ...)
  - What is the purpose of the APP & each of its components?
  - Which datasources an APP accesses?
  - Implement control of version of different APP's components

DataMining UV TechPlat

- computational process of discovering patterns in large data sets **(data, files, apps classes & binaries, configurations, logs)**
- The overall goal of the data mining process is to extract information from TechPlat data sets and transform it into an understandable structure for further use
  - Interesting information: Application dependences, business transactions, datasources, infrastructure, runtime, level of service ...

DataMining UV TechPlat

- A series of distributed batch mining tools developed in bash scripting, REXX, perl and jython (java + python) to report:

  + Middleware Topology (nodes, clusters, application servers)

  + Application state in each cluster

  + JDBC data sources details (username, host, DB name) and applications using them

  + CICS programs and JEE applications calling them

  + DB2 z/os plans used by CICS transactions

  + Shared file system directories used by apps

  + Services used by APPs in other systems

DataMining UV TechPlat

- Batch takes 80 hours to complete
  - WLM to avoid affecting level of service
- Difficult to find some patterns:
  - Some datasources can be found in config files, while others are embedded in source code
  - brute-force search over millions of files, lines of code, databases and configurations
- Example:
  - STEP 1: Extract all datasources from middleware APIs
  - STEP 2: Search all the code & configurations, to see which applications are using which datasources

DataMining UV TechPlat

- COBOL/CICS/DB2 z/os Report:
  - STEP 1: Extract all names of transactions, plans, programs from CICS transaction monitor (mainframe)
  - STEP 2: Search all the code, to see which applications are calling which transactions and for what (source code comments, XMLs, etc ..)
  - STEP 3: Cross all info. **4K online transactions auto-documented:** Description (business utility), name of transaction, program called & plan used to access DB2 z/os

DataMining UV TechPlat

Example of APP report :

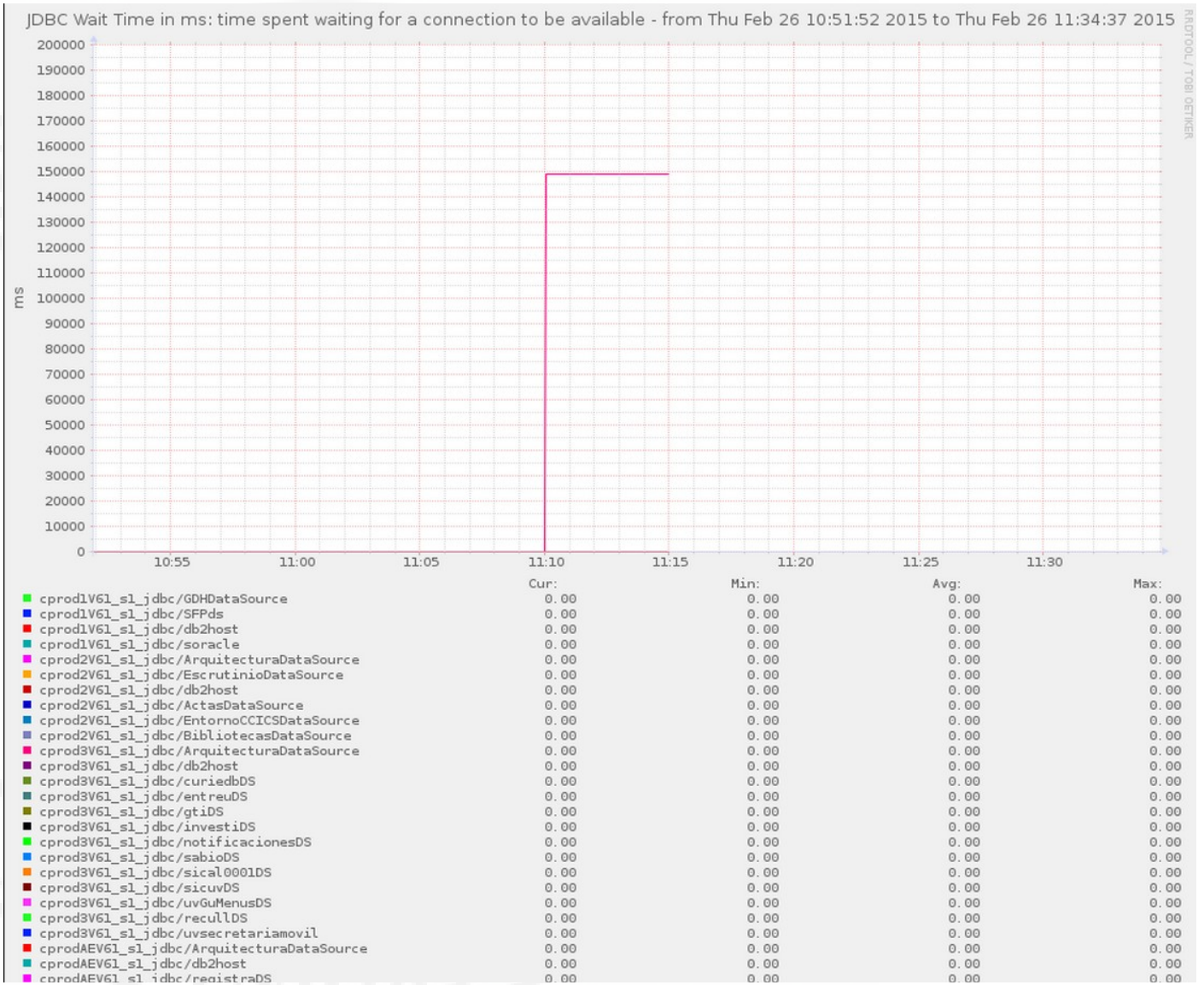
```
######################### REPORT DATE ######################################
Mon Jan 19 08:59:36 CET 2015
######################### EAR NAME #########################################
uvGdi.ear
######################### APP NAME #########################################
GDI:Gestión Docencia Impartida
######################### RUNTIME ##########################################
cprod1V61_s2
######################### APP ID/DESC ######################################
    68 Gestión Docencia Impartida

######################### LAST UPDATED #####################################
Jan 16 14:15
######################### NAS SUBDIRECTORIES ###############################
 UvGdi
######################### REFERENCES  ######################################
 gdi.uv.es mmedia.uv.es www.uv.es
######################### DATASOURCES ######################################
---- DataSource Name db2host ----:
databaseName "UVDST" ServerName "xxx.uv.es" portNumber "5025" USERNAME: "APLXXXX"
---- DataSource Name replica ----:
URL "jdbc:oracle:thin:@xxxx.uv.es:1533:repli11" USERNAME: "xxxxxx"
######################### DB2 Z/OS PLANS ###################################
USUARIOS
######################### CICS PROGRAMS  ###################################
NAME: CON_ENCUESTAS_NPI
PROGRAM: PUENTEM
TRANSACTION: OCE1
DESCRIPTION: Consulta la lista de encuestas para un dia
CALLED FUNCTION: FSRENCUE
DB2 PLAN: DOCTOR OCA PROCQMF USUARIOS
**********
NAME: CON_INCIDEN_TUTORIAS
PROGRAM: PUENTEM
TRANSACTION: OCE1
DESCRIPTION: Consulta la lista de tutorias
CALLED FUNCTION: FSRINTUT
DB2 PLAN: DOCTOR OCA PROCQMF USUARIOS
…....
```

DataMining UV TechPlat

- Some Interesting Reports:
  - Global Status:
    - https://webges.uv.es/private/j2ee_report/ptstatus.html
  - Level of Service:
    - http://monitor.uv.es/estadisticasweb/0
  - Big Picture:
    - https://webges.uv.es/private/j2ee_report/report.txt
  - APP Based Reports:
    - International Affairs APP Report:
      - https://webges.uv.es/private/j2ee_report/apps/uvMovilidad.ear
  - DB Links Report:
    - https://webges.uv.es/private/j2ee_report/dblinks/uvEntreuWeb.ear

Practical Usage Scenarios

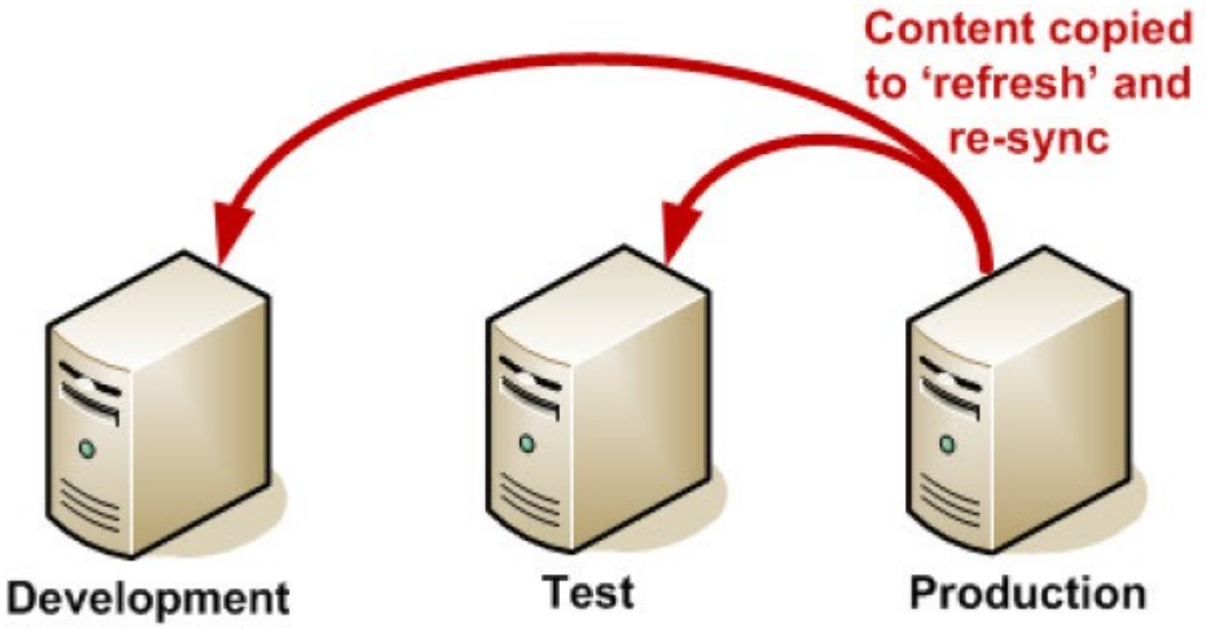


Problem determination:

E-goverment app waiting 150 seconds to obtain a connection from datasource.

Practical Usage Scenarios

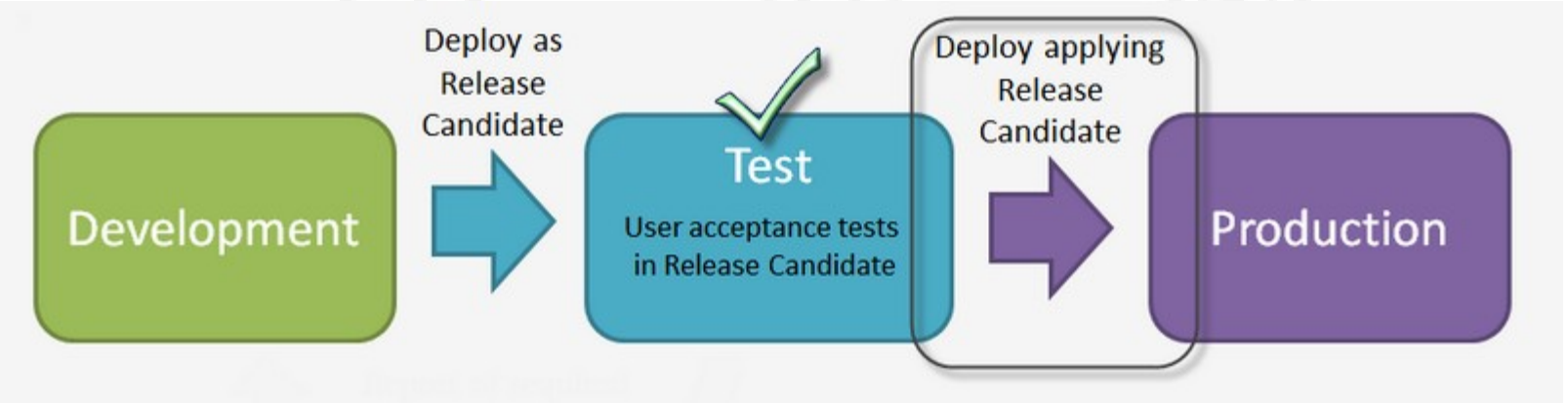
Application/DB cloning:



Quality Assurance: Need to refresh test & dev apps/data from production
Which databases/tables/applications/components need to be refreshed to test student enrollment application?

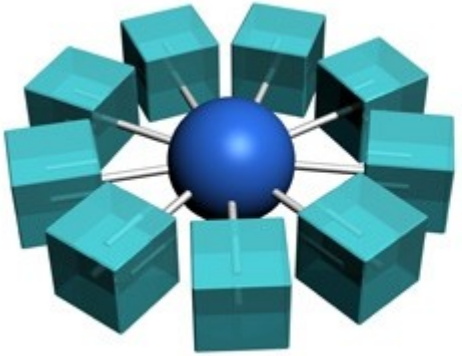
Practical Usage Scenarios

Application promotion:



Each time an app/service is promoted to production, an APP report is sent to operators, DBAs, programmers, etc ..
Anytime we know who promotes, what components and to which environment
An APP composed by different components (COBOL, JAVA) executed in different servers (mainframe, Unix, Linux) is copied from test to prod automatically, without manual intervention
Helps to implement change management

Practical Usage Scenarios



SU SOLUCIÓN EN
ESQUEMA NACIONAL DE SEGURIDAD
ENS
PLAN DE ADECUACIÓN

Some APPs/Services, like e-goverment are protected by law (SPANISH NATIONAL SECURITY FRAMEWORK)
SPANISH NATIONAL SECURITY FRAMEWORK: Protects citizens (APP users) rights (safety of systems, data, communications, availability of service, ...)
Need to catalog techplat assets used by APPS like e-goverment to implement special measures required by law
E-goverment is a complex APP with thousand of procedures implemented accessing several datasources
Risc management

DataMining UV TechPlat

- Any Questions?
- Thank you

  for coming!

Josep Vidal
System Programmer