

Sistemas de Software Distribuidos y Bases de Datos Distribuidas

P. Pesado^(1,2), H. Ramón⁽¹⁾, P. Thomas⁽¹⁾, M. Boracchia⁽¹⁾, R. Champredonde⁽¹⁾, A. Pasini⁽¹⁾, F. Chichizola⁽¹⁾, M. Iglesias⁽¹⁾, L. Marrero⁽¹⁾, M. B. Albanessi⁽¹⁾, L. Delia⁽¹⁾, G. Ricci⁽¹⁾

⁽¹⁾Instituto de Investigación en Informática LIDI (III-LIDI)
Facultad de Informática – UNLP

⁽²⁾Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC)

{ppesado, hramon, pthomas, marcosb, rchampre, apasini, francoch, miglesias, lmarrero, balbanesi, ldelia,gricci}@lidi.info.unlp.edu.ar

1 CONTEXTO

Esta línea de Investigación forma parte del Subproyecto “Sistemas de Software Distribuido. Aplicaciones” dentro del Proyecto “Sistemas de Software Distribuidos. Aplicaciones en procesos industriales, E-government y E-learning” del Instituto de Investigación en Informática LIDI.

2 RESUMEN

El proyecto de investigación y desarrollo en Sistemas de Software Distribuido, abarca aspectos que van desde los fundamentos del desarrollo (Técnicas de Ingeniería de Requerimientos, Metodologías de Gestión y Desarrollo de Proyectos, Técnicas de Planificación, Métricas, Normas de Calidad, Web-Systems) hasta la concepción de aplicaciones específicas orientadas a los procesos industriales (Sistemas de Planeamiento de Producción, Control Industrial en tiempo real), procesos de E-government (Voto electrónico) y la aplicación de Tecnología Informática para Educación en ambientes distribuidos (Educación a Distancia, Software educativo específico, Educación basada en la WEB).

En particular el objetivo de este subproyecto es realizar investigación y desarrollo en temas relacionados con los aspectos de Ingeniería de Software que se orientan al desarrollo e implementación de proyectos concretos de Sistemas Distribuidos (incluyendo Web Based Systems), y el manejo de datos físicamente distribuidos.

Palabras claves: *Informática - Sistemas Distribuidos – Ingeniería de Requerimientos – Planificación – Metodologías de Desarrollo – Web-Systems, Bases de Datos Distribuidas*

3 INTRODUCCION

Un sistema distribuido consiste en un conjunto de computadoras autónomas conectadas por una red y con soporte de software distribuido. Permite que las computadoras coordinen sus actividades y compartan los recursos de hardware, software y datos, de manera tal que el usuario percibe una única facilidad de cómputo integrada aunque esta pueda estar implementada por varias máquinas en distintas ubicaciones [1].

El desarrollo de sistemas distribuidos es una necesidad a partir de la utilización de redes de computadoras y de computadores personales de alta performance.

Algunas ventajas del procesamiento distribuido son:

- Mejora de la disponibilidad: la operación es factible en una configuración reducida cuando algunos nodos están temporalmente no disponibles.
- Configuración más flexible: una aplicación puede configurarse de distintas maneras, seleccionando el número apropiado de nodos para una instancia dada.
- Control y administración más localizada: un subsistema distribuido, ejecutando en su propio nodo, puede diseñarse para ser autónomo, de modo que puede ejecutar en relativa independencia de otros subsistemas en otros nodos.
- Expansión incremental del sistema: si existe sobrecarga, el sistema puede expandirse agregando más nodos.
- Costo reducido: con frecuencia una solución distribuida es más barata que una centralizada.

- Balance de carga: en algunas aplicaciones la carga total del sistema puede ser compartida entre varios nodos.
- Manejo eficiente de datos distribuidos físicamente.
- Mejora en el tiempo de respuesta: los usuarios locales en nodos locales pueden obtener respuestas más rápidas a sus requerimientos.

Las aplicaciones del procesamiento distribuido abarcan prácticamente todas las áreas productivas de la sociedad actual. En este proyecto se enfocan metodologías y aplicaciones de Sistemas Distribuidos en áreas tales como Planeamiento de la Producción industrial, Sistemas centrados en la WEB, Manejo de datos distribuidos heterogéneos, E-learning e E-government.

Las características de los Sistemas Distribuidos conducen a la utilidad de desarrollar prácticas de Ingeniería de Software que apunten a los distintos aspectos del desarrollo de sistemas desde la captura de requerimientos y la planificación, pasando por las metodologías de desarrollo, hasta la verificación y simulación de procesos distribuidos, incluyendo el aseguramiento de calidad.

La Ingeniería de Software comprende la aplicación de principios científicos para realizar la transformación ordenada de un problema en una solución elaborada de software, y el mantenimiento subsecuente de ese software hasta el final de su vida útil [2].

Las personas y proyectos que adoptan un enfoque ingenieril para el desarrollo de software, generalmente atraviesan una serie de fases o estados conformando un ciclo de vida. Este ciclo de vida está guiado por una planificación que indica la secuencia de acciones a realizar, y los productos generados por la aplicación del plan (inclusive el mismo plan) están gestionados por diferentes Metodologías de Gestión y Desarrollo [3].

En el recorrido del ciclo de vida del desarrollo del software, la fase inicial comprende a la Ingeniería de Requerimientos.

La Ingeniería de Requerimientos permite comprender, documentar y acordar sobre el alcance del problema, teniendo esto impacto directo sobre la Planificación y la Gestión del Proyecto de acuerdo a la Metodología de desarrollo seleccionada para el mismo [4]. Este no es el único impacto que justifica el énfasis en la Ingeniería de Requerimientos, ya que esta disciplina permite establecer claramente *que* se debe hacer posponiendo el *como* a etapas posteriores en el ciclo de vida, constituyendo el fundamento de la construcción de un Sistema de Software [5].

El ciclo de vida del desarrollo de software puede ser considerado como una secuencia de procesos

que imponen claramente las actividades a realizar para la generación de productos [6].

La necesidad de supervisar estos procesos de desarrollo de software torna imprescindible la utilización de métricas. La medición se puede aplicar al proceso de software, con el intento de mejorarlo continuamente [7]. Se utiliza además, como ayuda a la estimación y gestión en proyectos futuros, el control de calidad y la evaluación de productividad. Finalmente el ingeniero de software puede utilizar los indicadores de medición obtenidos para la toma de decisiones tácticas a medida que el proyecto evoluciona de acuerdo a la planificación. En muchos casos es necesario adaptar o ampliar las métricas existentes para poder utilizarlas en sistemas distribuidos, en particular de tiempo real.

Asociado con la Ingeniería de Software de Sistemas Distribuidos, está el problema de utilizar un entorno WEB para los servicios que ofrece el Sistema. La tendencia creciente al desarrollo de arquitecturas centradas en un servidor (o un conjunto de servidores distribuidos) que ofrecen una interfaz WEB a los usuarios ha generado un importante desarrollo de la Investigación en metodologías y herramientas orientadas a Sistemas WEB, así como ha obligado a establecer nuevas métricas y parámetros de aseguramiento de la Calidad para tales Sistemas. [8] [9]

Por último, el modelo distribuido de datos hace posible la integración de BD heterogéneas proveyendo una independencia global del administrador de bases de datos respecto del esquema conceptual. En general interesa implementar una integración tal que reúna varios modelos de datos, representando cada uno de ellos características propias de organizaciones diferentes, asociadas para un trabajo conjunto.

Este modelo de distribución normalmente tiene un soporte multiprocesador tipo MIMD sobre diferentes esquemas de comunicaciones. En tal arquitectura se reproducen los problemas de distribución óptima y migración de datos y procesos y de tolerancia a fallas.

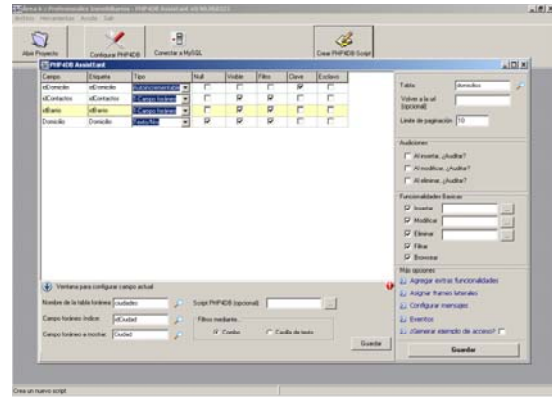
En particular interesa estudiar el comportamiento de grandes bases de datos sobre arquitecturas distribuidas de redes LAN y WAN. [10] [11]

4 LINEAS DE INVESTIGACION Y DESARROLLO

- Conceptos de procesamiento distribuido. Arquitectura, comunicaciones y software.
- Metodologías de especificación, validación y desarrollo de sistemas de software distribuido.
- Metodologías ágiles de desarrollo.
- Ingeniería de Requerimientos, en particular de sistemas distribuidos.
- Planificación de tareas para desarrollo de sistemas distribuidos.
- Bases de Datos Distribuidas
- Métricas para la estimación, seguimiento de proyectos y evaluación de software.
- Herramientas de integración y mantenimiento de proyectos distribuidos.
- Lenguajes y ambientes para procesamiento distribuido.
- Reingeniería de sistemas complejos que migran por down-sizing a esquemas cliente-servidor distribuidos.

5 RESULTADOS ESPERADOS/OBTENIDOS

- Avanzar en la capacitación continua de los miembros de la línea de investigación.
- Desarrollar soluciones a problemas concretos de software de sistemas distribuidos, poniendo énfasis en el desarrollo de metodologías y herramientas específicas para clases de aplicaciones. Evaluar la calidad, eficiencia, y relación de esfuerzo/costo obtenidos con las soluciones propuestas.
- Definir un proceso para proyectos de software de sistemas distribuidos.
- Generar una herramienta que soporte el proceso definido y que permita una evaluación de calidad, eficiencia, y relación de esfuerzo/costo obtenidos con las soluciones propuestas.
- Se ha desarrollado una herramienta que permite generación de código [12] [13] PHP interactuando con diferentes motores de BDD. Esta herramienta permite describir vistas (en base al modelo E/R), y a través del uso de eventos y descripción de datos (FormDataSet) automatizar tareas del proceso de desarrollo. Esto se ha aplicado en varios proyectos en ejecución del III-LIDI.



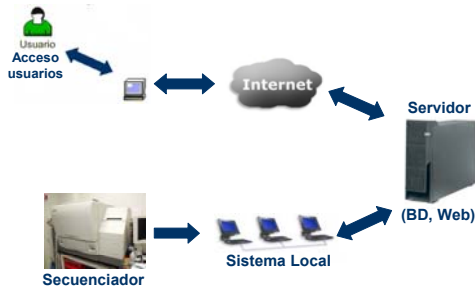
- Se ha utilizado una herramienta WEB para seguimiento de proyectos en forma colaborativa con el objetivo de lograr una planificación adecuada de un sistema distribuido.



Algunas transferencias de resultados logradas:

- Análisis y Diseño del Banco de Datos Genético de ganado vacuno y equino de la Pcia. de Bs. As. (centralizado/distribuido) y de los algoritmos para trazabilidad de ejemplares, identificación individual y determinación de paternidad para ganado mayor . El sistema desarrollado integra aplicaciones de dos escenarios distintos, por un lado un sistema WEB que permite toda la interacción con los potenciales usuarios del sistema, que solicitan a la unidad de procesamiento el análisis a realizar, envían las muestras a procesar, consultan sobre el estado de avance de la solicitud y reciben los resultados. Por el otro un sistema local que interactúa con un secuenciador que analiza las muestras de ADN y envía resultados. En particular el almacenamiento, tipificación y procesamiento de un gran número de muestras son la base para un programa de identificación genética que apunte a resolver situaciones tales como identificación de paternidad entre animales de una especie, estudio de casos de abigeato en ganado mayor y trazabilidad a partir de marcadores moleculares y secuenciamiento de

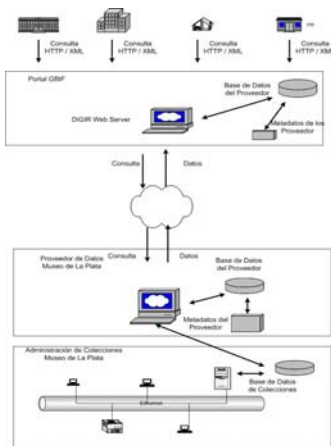
ADN. Este proyecto se desarrolla con la Facultad de Veterinaria de la UNLP en el marco de un Proyecto FONTAR de la Agencia Nacional de Promoción Científica y Tecnológica.



- Ampliación y escalabilidad del sistema desarrollado por el III-LIDI para el Museo de Ciencias Naturales de la UNLP y la Red Nacional de Colecciones Biológicas.

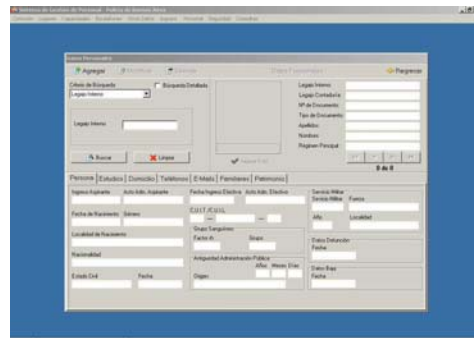
Este sistema WEB permite el tratamiento de Bases de Datos heterogéneas, esta necesidad surge de que los Museos de Ciencias Naturales albergan gran variedad de colecciones con características distintas y físicamente distribuidas, lo que hace que el almacenamiento, actualización y recuperación de estos datos sea un tema de estudio e investigación de mucho interés. La diversidad biológica de los elementos a almacenar, la localización física de los mismos y la distribución de los operadores/usuarios de la información, constituyen una clase de sistema distribuido complejo.

Actualmente se está trabajando con el Museo de Ciencias Naturales de La Plata (que integra la Red Nacional de Colecciones Biológicas), sobre 35 colecciones que involucran 2.400.000 registros.



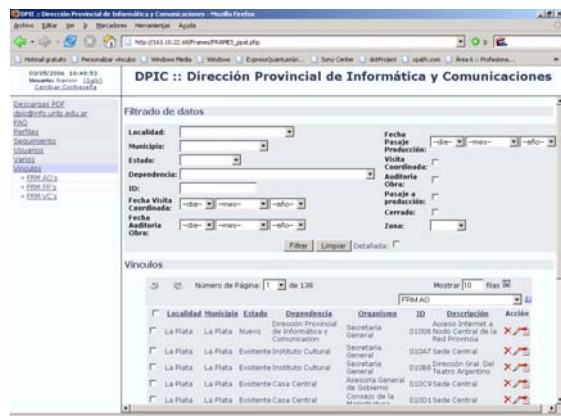
- Análisis, Diseño e Implementación de un Sistema de Información para la gestión estratégica de RRHH del Ministerio de Seguridad de la Provincia de Bs. As. Subsidiado por la CIC.

Se desarrolló un Sistema de Información contemplando los nuevos cuerpos policiales: Policía Buenos Aires 2, Comunal, de Distrito, Siniestral, de Custodia de Objetivos Fijos y Traslado de Detenidos, de Investigaciones del Tráfico de Drogas ilícitas, Central de Atención Telefónica de Emergencias y Cuerpo de Baqueanos. Este Sistema es apto para el tratamiento de datos de todos los regímenes estatutarios sirviendo de apoyo a la gestión de unidades organizativas centralizadas y descentralizadas del Sistema de Seguridad Pública.



- Auditoría de la Red Unica Provincial de Comunicación de Datos del Gobierno de la Provincia de Buenos Aires a través de la RedPIBA de la CICPBA.

Definición del manual de los procesos y productos necesarios para la aceptación de los nodos de acuerdo a la licitación del Gobierno de la Provincia de Buenos Aires para proveer "Servicio de Transmisión de Datos y Canales de Ordenes". Implementación del software de seguimiento del proyecto *Red Unica Provincial de Comunicación de Datos*, que vincula 135 Municipios (con alrededor de 1300 nodos) planificando la verificación de las instalaciones y la puesta en servicio (incluyendo el control de ancho de banda en tiempo real) de toda la red.



- Desarrollo de un sistema WEB de Gestión de Evaluaciones del sistema científico. El producto ha sido registrado y se ha utilizado para evaluación de artículos en Congresos y proyectos acreditados por Universidades Nacionales.



5. FORMACION DE RECURSOS HUMANOS

Los integrantes de esta línea de investigación dirigen Tesinas de Licenciatura en la Facultad de Informática, Tesis de Magister en temas relacionados con el proyecto y participan en el dictado de materias correspondientes al Doctorado de Ciencias Informáticas de la Facultad de Informática de la UNLP tales como Bases de Datos Distribuidas.

Colaboran en este proyecto los pasantes alumnos Germán Cáseres, Nicolás Luna y Danae Lopez.

6. BIBLIOGRAFIA

- [1] G. Coulouris. Distributed Systems – Concepts and Design. Addison-Wesley. 1994.
- [2] R. Pressman. Ingeniería de Software: Un Enfoque Práctico. McGraw-Hill. 2002
- [3] R. W. Wysocki. Effective Project Management: Traditional, Adaptive, Extreme, .Wiley .2003
- [4] Loucopoulos, P; Karakosas, V.. Systems Requirements Engineering. .McGraw Hill. Book Company. 1995
- [5] G. Kotonya and I. Sommerville,. Requirements Engineering: Processes and Techniques, Wiley. 1998
- [6] Pleegeer. Ingeniería de Software: Teoría y Práctica. Prentice-Hall. 2002
- [7] Stephen Kan. .Metrics and Models in Software Quality Engineering (2nd Edition). Addison Wesley. 2003
- [8] Offutt J., “Quality Attributes of Web Software Applications”. IEEE Software: Special, Issue on Software Engineering of Internet Software 19 (2):25-32, Marzo/Abril 2002.
- [9] Wu, Y. y Offutt, J. “Modeling and testing web-based Applications”. <https://citeseer.ist.psu.edu/551504.html>: 1-12, Julio 2004
- [10] Silberschatz A et al: “Fundamentos de Bases de Datos”, Tercera Edición Mc Graw Hill 1998
- [11] Ozsu M. Valduriez, P. : “Principles of Distributed Database Systems”, Segunda Edición. Prentice Hall 1999
- [12] Mello S. J., “Executable UML”, Addison-Wesley, 2002.
- [13] S. Ceri, P. Fraternali, A. Bongio, “Web Modeling Language (WebML): a modeling language for designing web sites”. Computer Networks, vol 33, 2000.