



Universidad Nacional de San Luis
Rectorado

Posgrado
"2014 - Año de Homenaje al Almirante Guillermo Brown,
en el Bicentenario del Combate Naval de Montevideo"

ES COPIA
OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL

SAN LUIS, 21 MAR 2014

VISTO:

El Expediente EXP-USL: 352/2014 mediante el cual se solicita la protocolización del Curso de Posgrado: **PROCESAMIENTO PARALELO Y DISTRIBUIDO SOBRE DATOS MASIVOS**; y

CONSIDERANDO:

Que el mencionado Curso se dictará en el ámbito de la Facultad de Ciencias Físico Matemáticas y Naturales desde abril hasta mayo de 2014 bajo la responsabilidad y la coordinación de la Dra. Graciela Verónica **GIL COSTA**, Corresponsable: Dra. Alicia Marcela **PRINTISTA**, Colaborador: Mag. Jair Edinson **LOBOS VILLEGAS**, todos de la Universidad Nacional de San Luis, con un crédito horario de 60 horas presenciales.

Que la Comisión Asesora de Posgrado de la Facultad de Ciencias Físico Matemáticas y Naturales recomienda aprobar el curso de referencia.

Que el Consejo de Posgrado de la Universidad Nacional de San Luis en su reunión del 18 de marzo de 2014, luego de su análisis acordó aprobar la propuesta como Curso de Posgrado.

Que corresponde su protocolización.

Por ello y en uso de sus atribuciones

EL RECTOR DE LA UNIVERSIDAD NACIONAL DE SAN LUIS

RESUELVE:

ARTÍCULO 1º.- Protocolizar el dictado del Curso de Posgrado: **PROCESAMIENTO PARALELO Y DISTRIBUIDO SOBRE DATOS MASIVOS**, en el ámbito de la Facultad de Ciencias Físico Matemáticas y Naturales desde abril hasta mayo de 2014 bajo la responsabilidad y la coordinación de la Dra. Graciela Verónica **GIL COSTA** (DNI N° 27.135.820), Corresponsable: Dra. Alicia Marcela **PRINTISTA** (DNI N° 16.836.867), Colaborador: Mag. Jair Edinson **LOBOS VILLEGAS** (Documento N° 94.999.140), todos de la Universidad Nacional de San Luis, con un crédito horario de 60 horas presenciales.

Felix D. Nieto Quintas
Felix D. Nieto Quintas
Rector
U.N.S.L.

M. Mercedes Campesinos
Dra. Mercedes Campesinos
Secretaria de Ciencia y Tecnología
UNSL

Cpde RESOLUCIÓN R N°

197



"2014 - Año de Homenaje al Almirante Guillermo Brown,
en el Bicentenario del Combate Naval de Montevideo"

Universidad Nacional de San Luis
Rectorado

ES COPIA
OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL

ARTÍCULO 2º.- Aprobar el programa del Curso de referencia, de acuerdo al ANEXO de la presente disposición.-

ARTÍCULO 3º.- Comuníquese, insértese en el Libro de Resoluciones y archívese.-

RESOLUCIÓN R Nº 197
mav

Mercedes Campderrós
Directora de Ciencia y Tecnología
UNSL

Dr. Felix D. Nieto Quintas
Rector
U.N.S.L



Universidad Nacional de San Luis
Rectorado

"2014 - Año de Homenaje al Almirante Guillermo Brown,
en el Bicentenario del Combate Naval de Montevideo"

ES COPIA
OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL

ANEXO

DENOMINACIÓN DEL CURSO: PROCESAMIENTO PARALELO Y
DISTRIBUIDO SOBRE DATOS MASIVOS

UNIDAD ACADÉMICA RESPONSABLE: Facultad de Ciencias Físico
Matemáticas y Naturales

CATEGORIZACIÓN: Perfeccionamiento

RESPONSABLE: Dra. Graciela Verónica GIL COSTA

CORRESPONSABLE: Dra. Alicia Marcela PRINTISTA

COLABORADOR: Mag. Jair Edinson LOBOS VILLEGAS

COORDINADOR: Dra. Graciela Verónica GIL COSTA

CRÉDITO HORARIO: 60 horas

MODALIDAD DE DICTADO: Presencial

FECHA DE DICTADO DEL CURSO: desde abril hasta mayo de 2014

FECHA PREVISTA PARA ELEVAR LA NÓMINA DE ALUMNOS

APROBADOS: noviembre de 2014

DESTINATARIOS: Egresados con título de grado universitario en Ciencias Informáticas
y en disciplinas afines a la temática del curso.

LUGAR DE DICTADO: Departamento de Informática – UNSL

CUPO: 15 personas.

FUNDAMENTACIÓN: La habilidad de manejar grandes volúmenes de datos requiere también la habilidad de desplegar y coordinar grandes volúmenes de hardware. Cada día nuevas técnicas y/o aplicaciones se proponen para el análisis y diseño eficiente de sistemas que deben procesar grandes cantidades de datos. Los sistemas más comúnmente utilizados son los de recuperación de la información como son los motores de búsqueda Web. Las técnicas utilizadas en estos sistemas complejos incluyen estrategias de procesamiento paralelo sobre diferentes arquitecturas de memoria compartida y distribuida, técnicas de planificación, estrategias de enrutamiento, heurísticas, índices, estrategias de ahorro de energía, estrategias de balance de carga, etc.

OBJETIVOS: Caracterizar los problemas de procesamiento sobre datos masivos en paralelo desde el punto de vista de una arquitectura de cluster de procesadores multi-core y de los paradigmas de programación aptos para ella.

CONTENIDOS MINIMOS:

- 1- Tipos de arquitectura de computadores
- 2- Modelos de computación para memoria compartida
- 3- Programación sobre Memoria distribuida/Compartida

Cpde RESOLUCIÓN R N°

197

Dr. Felicio Mendocinos
Rector
U.N.S.L.

Dra. Mercedes Campderros
Secretaria de Ciencia y Tecnología
UNSL




Universidad Nacional de San Luis
Rectorado


ES COPIA
OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL

- 4- Plataformas Distribuidas y dinámicas: CDN, Cloud y P2P
- 5- Procesamiento de datos masivos: Caso de estudio la Web
- 6- Stream Processing

PROGRAMA:

- 1- Arquitectura Multicore - Jerarquía Cache-Políticas de reemplazo
 - a. Memoria distribuida, compartida.
 - b. Límites de los procesadores de memoria compartida
 - c. Protocolos de coherencia de caché
 - d. Técnicas para optimizar el uso de memoria
 - e. Algoritmos de scheduling a nivel multicore (work stealing, basadas en sincronización y colas de trabajo)
- 2- Modelos de computación para memoria compartida
 - a. Motivación para un modelo estándar
 - b. Descripción del modelo
 - c. Caso de Estudio: Modelo Map-Reduce
 - d. El modelo de computación Multi-BSP
 - i. Definición del modelo y costo
 - ii. Algoritmos fundamentales utilizando Multi-BSP
- 3- Programación sobre Memoria distribuida/Compartida
 - a. Paralelismo a nivel de thread
 - b. Problemas de False-Sharing
 - c. Librerías de programación: OpenMP y MPICH
- 4- Desarrollo en Plataformas Distribuidas y Dinámicas
 - a. Programación en cluster de computadoras
 - b. DHT: CAN y Chord, SimPeer, Pastry
 - c. Plataforma Cloud Computing
- 7- Procesamiento de datos masivos: Caso de estudio la Web
 - a. Clasificación de información en la Web
 - b. Recuperación de la información en la Web
 - c. Motores de Búsqueda
 - d. Real Time Big Data
 - e. El nuevo desafío: Capacity Planning.
- 8- Stream Processing
 - a. Sistemas para procesamiento de flujos de eventos
 - b. Sistemas para procesamiento de eventos complejos
 - c. Colaboración entre sistemas de procesamiento de eventos
 - d. Aplicaciones para móviles
 - e. Caso Práctico: Estudio y análisis de herramientas de procesamiento de datos en streaming


Dr. Felix D. Nieto Quimbas
Rector
UNSL


Dra. Mercedes Campderrós
Secretaría de Ciencia y Tecnología
UNSL



Universidad Nacional de San Luis
Rectorado

"2014 - Año de Homenaje al Almirante Guillermo Brown,
en el Bicentenario del Combate Naval de Montevideo"

ES COPIA
OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL

SISTEMA DE EVALUACIÓN: Para aprobar el curso el alumno deberá presentar un proyecto integrador de las técnicas estudiadas en el curso.

BIBLIOGRAFÍA:

Afrati, F. N., Sarma, A. D., Salihoglu, S., and Ullman, J. D. (2012). Vision paper: Towards an understanding of the limits of map-reduce computation. arXiv preprint arXiv:1204.1754

Mike Barlow. Real Time Big Data. O'Reilly Media Inc. 2013. First Edition

Beaumont, O., Bonichon, N., and Eyraud-Dubois, L. (2008). Scheduling divisible workloads on heterogeneous platforms under bounded multi-port model. In Parallel and Distributed Processing, 2008. IPDPS 2008. IEEE International Symposium on, pages 1-7. IEEE.

Grama, A. (2003). Introduction to parallel computing. Pearson Education.

Jacobs, A. (2009). The pathologies of big data. Communications of the ACM, 52(8):36-44.

Neumeyer, L. Robbins, B. ; Nair, A. ; Kesari, A. S4: Distributed Stream Computing Platform. Pages 170 – 177. International Conference on Data Mining Workshops (ICDMW), 2010

Alberto Montresor and Mark Jelasity. PeerSim: A Scalable P2P Simulator. Proc. of the 9th Int. Conference on Peer-to-Peer. Pp. 99-100 , 2009.

Ewing Lusk and Anthony Chan. Early experiments with the OpenMP/MPI hybrid programming model. IWOMP'08 Proceedings of the 4th international conference on OpenMP in a new era of parallelism. 2008

Qingqing Gan, Torsten Suel: Improved techniques for result caching in web search engines. WWW 2009: 431-440.

L. G. Valiant. A Bridging Model for Multi-core Computing. In ESA, 2008.

L. G. Valiant. A Bridging Model for Parallel Computation Comm. ACM, 33(8): 103-111, 1990


E. Koskinen, M. Parkinson and M. Herlihy. Coarse-Grained Transactions; In POPL, 2010.

Georg Hager-Gerhard Wellein. Introduction to High Performance Computing for Scientists and Engineers. Chapman & Hall/CRC. Computational Science Series. 2011

The OpenMP API specification for parallel programming. <http://openmp.org>

K. Agrawal and Y. He and E. Leiserson. Adaptive work stealing with parallelism feedback. In Principles and Practice of Parallel Computing, pages 112-120. 2007.

Cpde RESOLUCIÓN R N° 197


Dr. Felix B. Nieto Quinones
Rector
U.N.S.L.


Dra. Mercedes Campderros
Secretaría de Ciencia y Tecnología
UNSL



"2014 - Año de Homenaje al Almirante Guillermo Brown,
en el Bicentenario del Combate Naval de Montevideo"

ES COPIA
USCA GUILLERMO SEGURA
Director de Demarchas

Universidad Nacional de San Luis
Rectorado

R. Hoffmann and M. Korch and T. Rauber. Performance Evaluation of Task Pools
Based on Hardware Synchronization. In Supercomputing Conference. 2004.

ARANCEL: Sin costo.

COSTOS Y FUENTE DE FINANCIAMIENTO: Departamento de Informática –
UNSL.

Cpde RESOLUCIÓN R N° **197**
may

Dra. Mercedes Campderrós
Secretaria de Ciencia y Tecnología
UNSL

Dr. Felix D. Nieto Quintas
Rector
U.N.S.L