



Title: Entorno de Aprendizaje de Procesos y Herramientas de Big Data para Alumnos de Licenciaturas en Ingeniería en Computación

Authors: HERNÁNDEZ-CABRERA, Jesús, MENDOZA-GONZÁLEZ, Omar and SÁNCHEZ-HERNÁNDEZ, Miguel Ángel

Editorial label ECORFAN: 607-8695

BCIERMMI Control Number: 2021-01

BCIERMMI Classification (2021): 271021-0001

Pages: 16

RNA: 03-2010-032610115700-14

ECORFAN-México, S.C.

143 – 50 Itzopan Street

La Florida, Ecatepec Municipality

Mexico State, 55120 Zipcode

Phone: +52 1 55 6159 2296

Skype: ecorfan-mexico.s.c.

E-mail: contacto@ecorfan.org

Facebook: ECORFAN-México S. C.

Twitter: @EcorfanC

www.ecorfan.org

Holdings

Mexico	Colombia	Guatemala
Bolivia	Cameroon	Democratic
Spain	El Salvador	Republic
Ecuador	Taiwan	of Congo
Peru	Paraguay	Nicaragua

1. Introducción
2. Objetivos
3. Contribución
4. Desarrollo
5. Descripción de la metodología
6. Gráficos
7. Resultados
8. Conclusiones
9. Referencias

Introducción

- Creciente demanda de profesionistas con conocimientos de en el área de Big Data.
- Necesidad de una capacitación temprana y continua respaldada por un entorno de enseñanza-aprendizaje teórico y práctico.
- Apoyar el aprendizaje de los alumnos con:
 - Herramientas
 - Ejemplos teóricos y prácticos
 - Material didáctico

Objetivos

Generar un entorno de enseñanza-aprendizaje de los procesos, herramientas y tecnologías Big Data, soportados por entregables concretos como son talleres, cursos, material didáctico, recursos en línea y con la implementación de un entorno distribuido con las herramientas Hadoop, disponible para alumnos y académicos de nivel licenciatura de la UNAM.

Contribución

Se ha capacitado a 70 alumnos de forma profesional con acercamiento real a la arquitectura, herramientas y procesos de manejo masivo de datos por medio de:

- Talleres prácticos
- Entorno preconfigurado con el ecosistema Hadoop
- Cursos intersemestrales

Desarrollo

1. Integración de Material didáctico.
2. Construcción de imagen de máquina virtual con el entorno de Hadoop.
3. Impartición de Taller Práctico.
4. Impartición de cursos intersemestrales
5. Construcción de un recurso MOOC.

Descripción de la metodología

- a) Investigación documental del marco teórico de los conceptos de Big Data.
- b) Desarrollo del contenido de los cursos, talleres, material didáctico y recursos en línea .
- c) Creación y configuración del entorno de trabajo para Big Data.
- d) Pruebas de operación del entorno para Big Data.
- e) Generación del material didáctico, que se incluirá en el repositorio de recursos digitales
- f) Impartir los cursos y talleres presenciales y a distancia.
- h) Publicación por medio de un MOOC de los materiales didácticos del curso y taller de Big Data.
- i) Generación de un reporte final.

Gráficos

Componentes del ecosistema en la máquina virtual.



Hadoop User Experience (HUE)



APACHE SQOOP



Flume



Apache
ZooKeeper



Apache Hive



YARN / Map Reduce



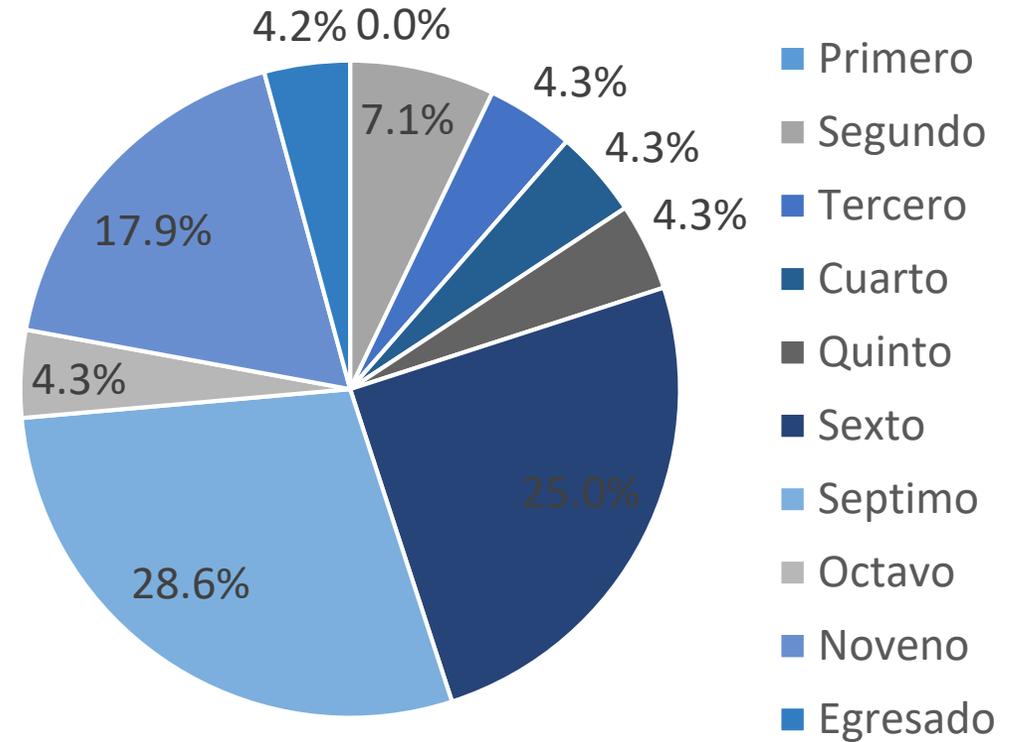
Hadoop Distributed File System



Linux Ubuntu

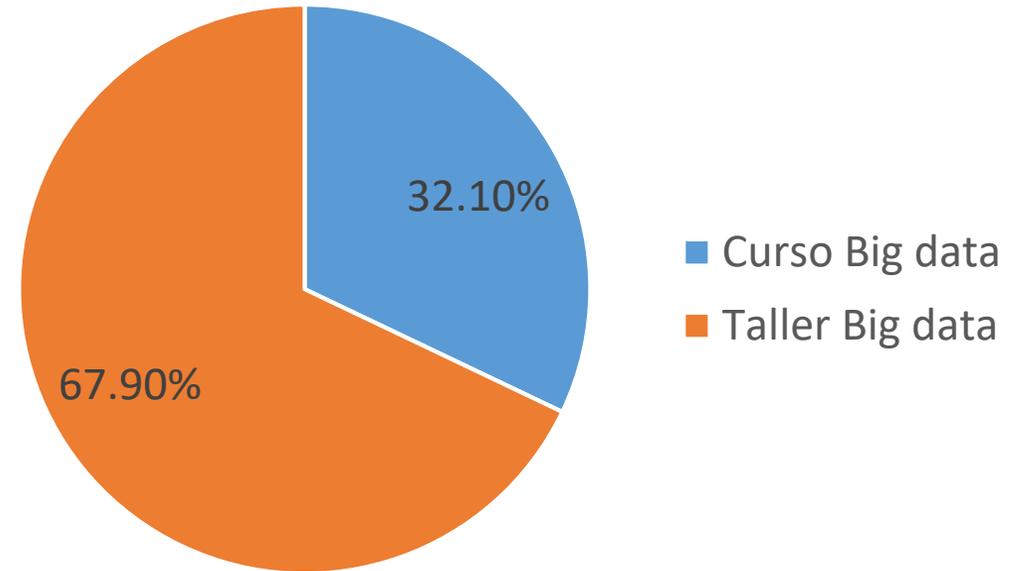
Gráficos

Porcentaje de respuestas a la pregunta “¿Qué semestre cursabas cuando tomaste el curso?”



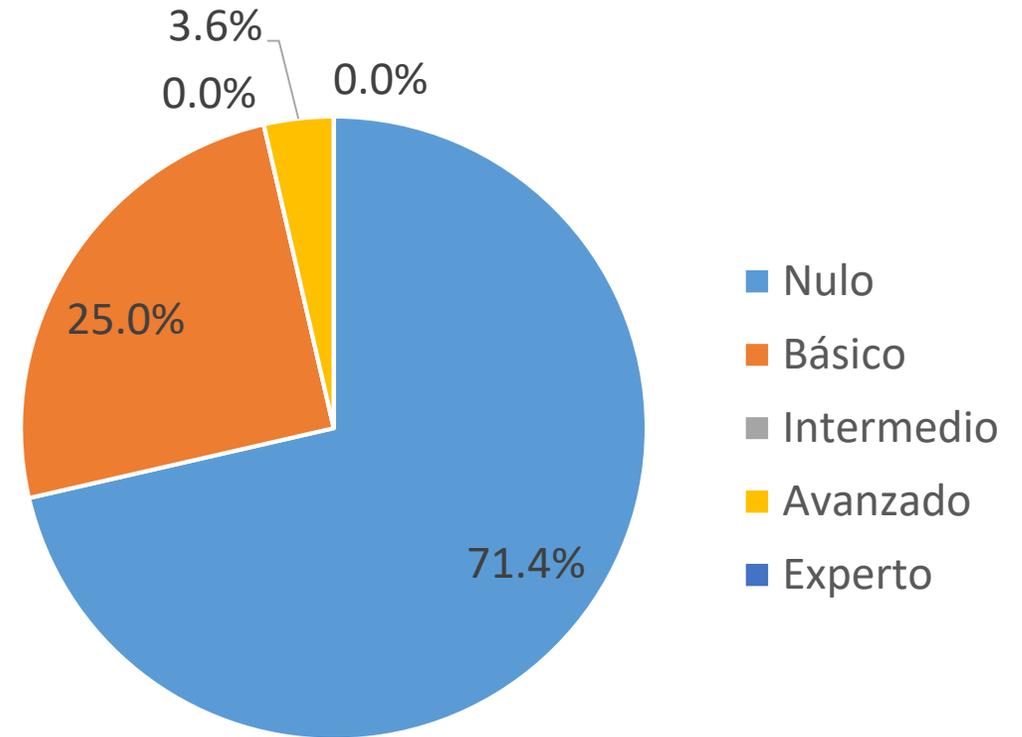
Gráficos

Porcentaje de respuestas a la pregunta “¿Qué curso tomaste?”



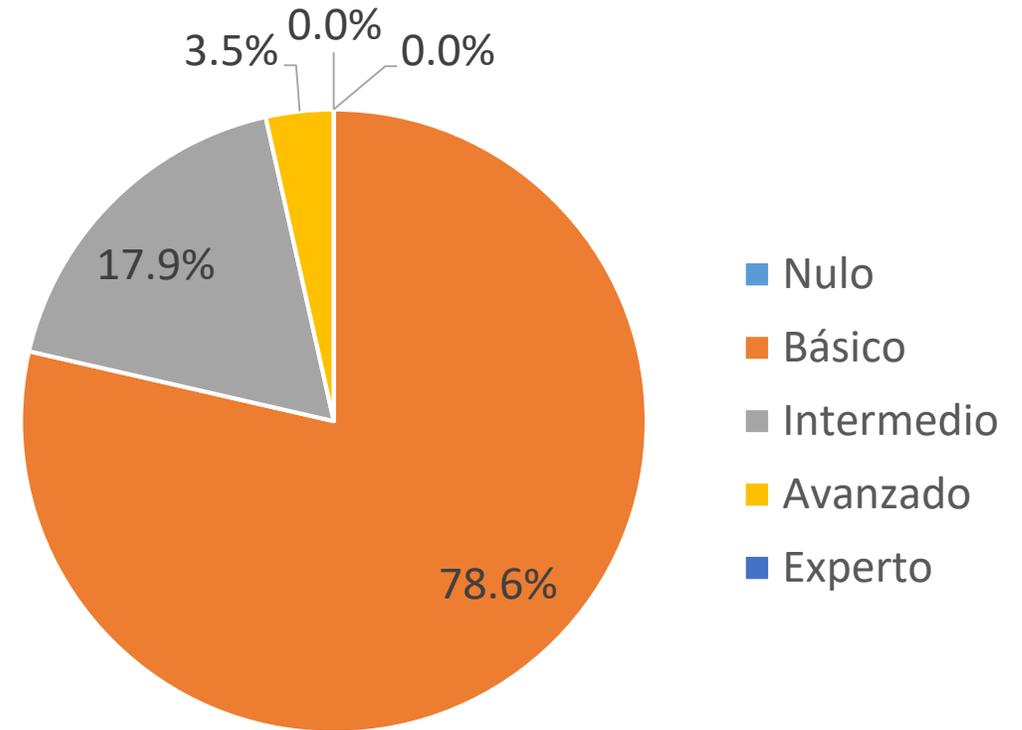
Gráficos

Porcentaje de respuestas a la pregunta “¿Con qué nivel previo de conocimientos en Big data contabas?”



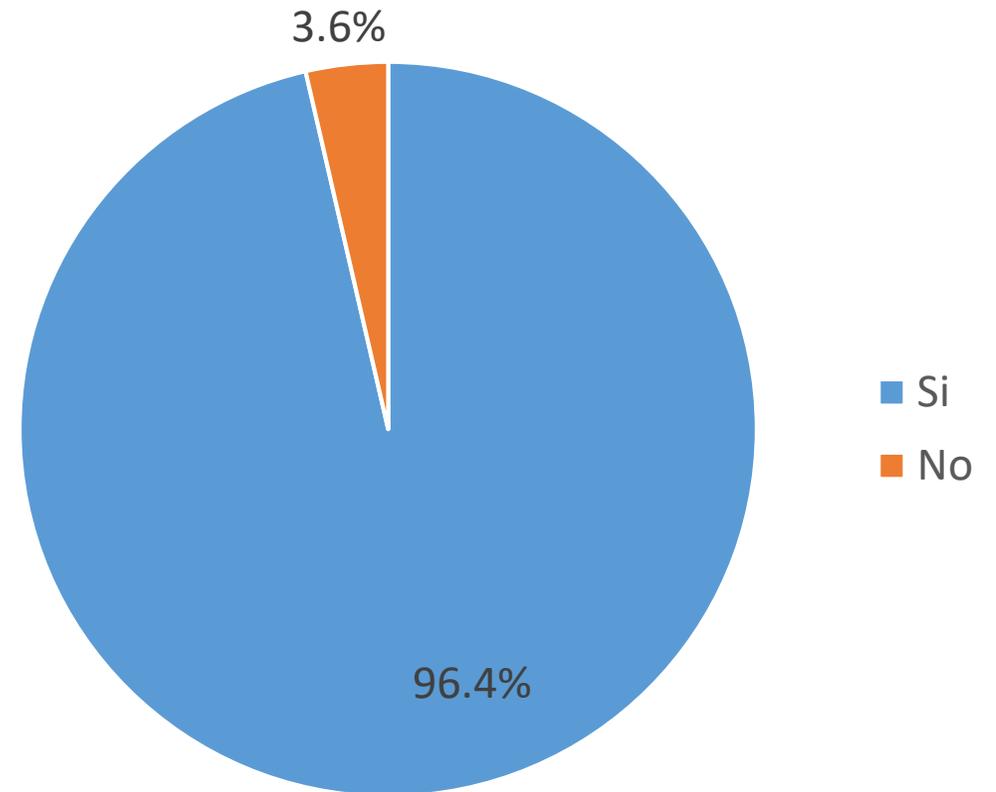
Gráficos

Porcentaje de respuestas a la pregunta “Después del curso/taller ¿Cómo consideras tu nivel de conocimiento del ecosistema Hadoop?”



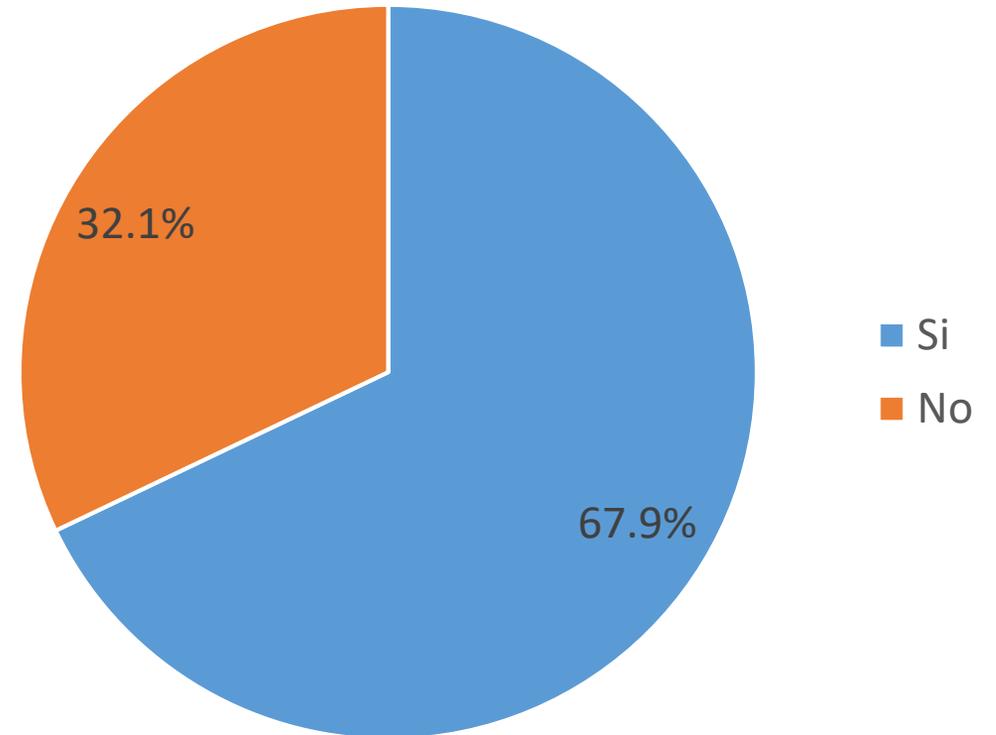
Gráficos

Porcentaje de respuestas a la pregunta “¿Consideras seguir aprendiendo temas de Big Data?”



Gráficos

Porcentaje de respuestas a la pregunta “¿El curso te sirvió de base para implementar los conocimientos de Big Data en otros proyectos?”



Resultados

- Los participantes indican haber obtenido conocimientos básicos del ecosistema de Hadoop, procesamiento e ingesta de datos
- Se alcanza una respuesta superior al 95% de los alumnos que desean seguir aprendiendo Big Data
- Un 70% de ellos lograron implementar su conocimiento en otros proyectos
- Seis de los alumnos que tomaron el taller práctico participaron también como apoyo a los instructores del proyecto, y tuvieron la responsabilidad de guiar y asesorar a los alumnos de los cursos

Conclusiones

- Inicio de una comunidad de aprendizaje en Big Data en la Facultad de Estudios Superiores Aragón
- Desarrollo de un entorno de práctica y material didactico para los cursos.
- El reforzamiento práctico que ayuda a que el alumno aumente su nivel de aprendizaje y se interese en seguir profundizando en el tema.
- Inicio de una segunda etapa que consiste en la creación de un clúster utilizando equipo de placa simple, que soporte el procesamiento masivo de datos
- Posibilidad de ejecutar proyectos de Big Data relacionados con problemáticas como impacto urbano, problemas de baches, tránsito municipal, líneas de producción industrial, seguimiento de redes sociales, entre otros.

Referencias

- Saroha, M., & Sharma, A. (2019). Big Data and Hadoop Ecosystem: A Review. *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*.
- Davoudian, A., & Liu, M. (2020). Big Data Systems. *ACM Computing Surveys*, 53(5),1–39.
- Sanagustin, M. M. (2016). Mooc Marker. En línea, disponible en: http://www.moocmaker.org/wp-content/files/D1.1-InformeMOOCLatam-vFINALDEFINITIVO_Spanish.pdf
- J. Moreno, E. B. Fernandez, M. A. Serrano and E. Fernández-Medina, "Secure Development of Big Data Ecosystems," in *IEEE Access*, vol. 7, pp. 96604-96619, 2019, doi: 10.1109/ACCESS.2019.2929330.
- Ting, K., & Cecho, J. (2013). *Apache Sqoop Cookbook*. O'Reilly Media.
- White, T. (2015). *Hadoop: The Definitive Guide*. O'Reilly & Associates.
- S. Ghemawat, H. Gombioff, S. Leung. "The Google file system, " In *Proc. of ACM Symposium on Operating Systems Principles*, Lake George NY, Oct 2003, pp 29-43.
- Sam R. (2017) *Expert Hadoop Administration, Managing Tuning and Securing SPARK,YARN and HDFS*. United States of America:Addison-Wesley
- Hadoop Documentación. [Online]. Available: <https://hadoop.apache.org/docs/r3.1.4/>
- Hive Documentación. [Online]. Available: <https://cwiki.apache.org/confluence/display/Hive/LanguageManual>
- Dean, Jeffrey & Ghemawat, Sanjay. (2004). MapReduce: Simplified Data Processing on Large Clusters. *Communications of the ACM*. 51. 137-150. 10.1145/1327452.1327492.
- Chang, Fay & Dean, Jeffrey & Ghemawat, Sanjay & Hsieh, Wilson & Wallach, Deborah & Burrows, Michael & Chandra, Tushar & Fikes, Andrew & Gruber, Robert. (2008). Bigtable: A Distributed Storage System for Structured Data. *ACM Trans. Comput. Syst.*. 26. 10.1145/1365815.1365816.



ECORFAN®

© ECORFAN-Mexico, S.C.

No part of this document covered by the Federal Copyright Law may be reproduced, transmitted or used in any form or medium, whether graphic, electronic or mechanical, including but not limited to the following: Citations in articles and comments Bibliographical, compilation of radio or electronic journalistic data. For the effects of articles 13, 162,163 fraction I, 164 fraction I, 168, 169,209 fraction III and other relative of the Federal Law of Copyright. Violations: Be forced to prosecute under Mexican copyright law. The use of general descriptive names, registered names, trademarks, in this publication do not imply, uniformly in the absence of a specific statement, that such names are exempt from the relevant protector in laws and regulations of Mexico and therefore free for General use of the international scientific community. BCIERMMI is part of the media of ECORFAN-Mexico, S.C., E: 94-443.F: 008- (www.ecorfan.org/booklets)