



Herramientas de Big Data - Presentación

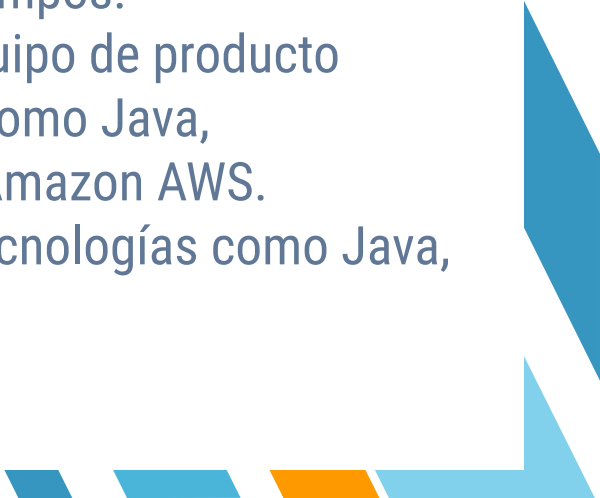
HBD - 01/2019





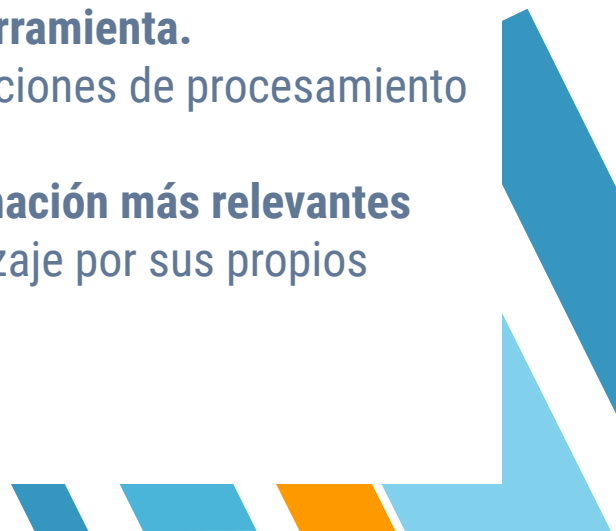
Presentación

Marcelo Turrin -mturrin@itba.edu.ar

- » **Ingeniero en Informática** del ITBA
 - » **Docente de grado** Software Engineer desde el 2006, actualmente dictando Introducción a la informática y Programación de Objetos Distribuidos.
 - » **2006-2007:** Desarrollador Java en la consultora Snoop..
 - » **2007-2015:** Desarrollador/Lider PM en la consultora Zaub en diversos proyectos para clientes en múltiples tecnologías y campos.
 - » **2015-2017:** Back End Developer en Flowics en el equipo de producto específicamente en backend utilizando tecnologías como Java, Elasticsearch, MongoDB, RabbitMQ, Redis, Hadoop, Amazon AWS.
 - » **2018:** Software Engineer en Elementum, como con tecnologías como Java, Weld, Cassandra, ElasticSearch, Kafka.
- 




Objetivos de la Materia

- » Proveer una **mirada comprensiva** de las herramientas que componen el ecosistema de Big Data actualmente.
 - » **Resaltar** aquellos **conceptos, tecnologías y proyectos** que poseen mayor tracción o proyección que apunta a que se conviertan en el standard.
 - » Proveer a los alumnos con las herramientas básicas para **entender el ecosistema y en qué casos es preciso utilizar cada herramienta**.
 - » Proveer **experiencia práctica** en el desarrollo de aplicaciones de procesamiento y guardado de datos altamente escalables.
 - » Proveer **acceso a las herramientas y fuentes de información más relevantes** para que los alumnos puedan continuar con el aprendizaje por sus propios medios una vez finalizada la cursada.
- 




Qué no vamos a ver

- » Herramientas propietarias o no Open Source.
 - » Detalles muy técnicos o de implementación.
 - » Implementaciones específicas para retail, medicina, agro, finanzas, etc.
 - » Armado de visualizaciones y/o dashboards.
 - » Detalles sobre cloud computing.
 - » Algoritmos de Machine Learning y Data Mining.
- 




Programa de la Materia

1. Introduction to Big Data.
 2. Hadoop and processing frameworks.
 3. Hadoop data ingestion and practice.
 4. From RDBMs to NoSQL and beyond.
 5. Introduction to Apache Spark.
 6. Apache Spark practice.
 7. Developing real-life Big Data projects by Andres Moratti, guest lecturer from Flowics.
 8. **Final Exam** and Real-time data processing and the future.
- 




MUY IMPORTANTE

- » El curso abarca muchas herramientas y conceptos que van a ser novedosos.
 - » Fundamental repasar los temas ya vistos antes de cada clase.
 - » La bibliografía sugerida les permite ahondar en cada uno de los temas que vamos a ver y más.
- 



Bibliografía: Libros

- » Tom White, **Hadoop: The Definitive Guide**, 4th Edition, O'Reilly Media, March 2015
 - » Alex Holmes, **Hadoop in Practice**, Second Edition, Manning Publications, September 2014
 - » Mark Grover, Ted Malaska, Jonathan Seidman, Gwen Shapira, **Hadoop Application Architectures**, O'Reilly Media, July 2015
 - » Nathan Marz and James Warren, **Big Data**, Manning Publications, April 2015
 - » Martin Kleppmann, **Designing Data Intensive Applications**, O'Reilly Media, April 2017
 - » Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia, **Learning Spark**, O'Reilly Media, January 2015
 - » Sandy Ryza, Uri Laserson, Sean Owen, Josh Wills, **Advanced Analytics with Spark**, O'Reilly Media, Inc., March 2015
 - » Neha Narkhede, Gwen Shapira, Todd Palino, **Kafka: The Definitive Guide**, To Be released.
- 



Bibliografía: Online

- » The Hortonworks Blog - <http://hortonworks.com/blog/>
 - » Cloudera Engineering Blog - <http://blog.cloudera.com/blog/>
 - » Databricks Developer Blog - <http://databricks.com/blog>
 - » Confluent Blog - <http://blog.confluent.io/>
 - » Weekly news about Big Data ecosystem - <http://www.hadoopweekly.com/>
 - » Martin Kleppmann Blog - <https://martin.kleppmann.com/>
 - » NoSQL Weekly - <http://www.nosqlweekly.com/>
 - » Netflix Tech Blog - <https://medium.com/netflix-techblog>
 - » Uber Tech Blog - <https://eng.uber.com>
- 