

**75.06 Organización de Datos.**  
**Calendario 1er Cuatrimestre 2019**  
[G##] = Gradiance /// [Num] = capítulo apunte

Día	Temas	Día	Temas
11/3	Administrivia, Introducción a Data Science, Introducción a Big Data, Jerarquía de Almacenamiento, Tipos de Datos, Arquitecturas Lambda. [G0][G0.5][1]	14/3	DataFrames. Análisis Exploratorio de Datos, Introducción a Pandas y Visualización de datos. [2,3]
18/3	DataFrames. Análisis Exploratorio de Datos, Pandas Avanzado. El paradigma split-apply-combine. Mas visualización de datos. [2,3]	21/3	DataFrames. Análisis Exploratorio de Datos, Pandas muy avanzado. <b>Vencimiento para la formación de grupos.</b> [G1][2,3]
25/3	Spark I. Hadoop. HDFS. Teoría de Map Reduce. Introducción a Spark. RDDs, modelo de ejecución, transformaciones y acciones en pyspark. [4]	28/3	Spark II. Transformaciones y Acciones II. Ejemplos con pyspark. Joins. Broadcast join.[4]
1 /4	Spark III. Estructuras de alto nivel. Dataframes, SparkSQL. [G2][4]	4/4	ITI I. Complejidad de Kolmogorov. NCD. Teoría de Compresión de Datos. Huffman estático y dinámico. Compresión Aritmética. [5]
8/4	ITI II. La Familia LZ de compresores. Block Sorting. PAQ. [G3][5]	11/4	Hashing. Construcción de Funciones de Hashing, Hashing Universal, Hashing Perfecto, Cuckoo Hashing, Teorema de Johnson-Lindenstrauss. Feature Hashing. [6].
15/4	LSH. Minhashes, amplificación de minhashes. LSH para la distancia de Jaccard, LSH para la distancia angular, LSH para la distancia euclidiana. [G4][7].	18/4	<b>FERIADO</b>
22/4	Reducción de Dimensiones. SVD y PCA. Multidimensional Scaling, ISOMAP, Laplacian Eigenmaps, TSNE. [G5][8] <b>Entrega de TP1.</b>	25/4	Information Retrieval I. Introducción al problema de Search. Índices Invertidos, construcción de índices, almacenamiento del léxico y punteros. Resolución de consultas. N-gramas.[9]
29/4	Information Retrieval II. Resolución de consultas ranqueadas. Método del coseno. BM25. LSI. Evaluación de consultas. [G6][9]	2/5	<b>CLASE DE EJERCICIOS</b>
6/5	<b>PARCIAL</b>	9/5	Machine Learning I. Teoría de Machine Learning. Teorema NFL. Underfitting y Overfitting. Naive Bayes. KNN [G7][10,11]
13/5	Machine Learning II. Clasificación. Modelos Lineales: Logistic Regression, Perceptron, Redes Neuronales. Deep Learning. Tensorflow. Keras. [G8][12]	16/5	Machine Learning III. Clasificación, Modelos basados en Árboles: Árboles de decisión, Random Forest, Boosting, XGBoost. Ensamblados.[12].
20/5	Machine Learning IV. Feature Engineering y casos de estudio. <b>Entrega de notas del Parcial.</b>	23/5	Clustering. Clustering jerárquico. K-Means, K-Means online, K-Means++, Clustering espectral. DBScan y HDBScan. [G9][13].
27/5	<b>RECUPERATORIO</b>	30/5	Recomendaciones. Sistemas basados en contenido. Collaborative Filtering. Matrix Factorization. Factorization Machines. [G10][14]
3/6	Streaming. Algoritmos de Streaming. [G11][15].	6/6	Page Rank y derivados. Pagerank, topicrank, trustrank, simrank, textrank, visualrank. [G12][16]
10/6	Redes Sociales. Análisis de Redes Sociales, características y algoritmos. [G13][17] <b>Entrega de Notas del Recuperatorio.</b>	13/6	<b>Clase especial</b>
17/6	<b>FERIADO</b>	20/6	<b>FERIADO</b>
24/6	<b>CORRECCION FINAL DE TPS</b>	27/6	<b>CORRECCION FINAL TPS</b>
1/7		4/7	<b>Segundo Recuperatorio, Coloquio y Examen por Promoción.</b>