

Organización de Datos (75.06) Segundo Cuatrimestre de 2014. Examen parcial, primera oportunidad. [2014_2c_Parcial_1]

Importante: Antes de empezar complete nombre y padrón en el recuadro. Lea bien todo el enunciado antes de empezar. Para aprobar se requiere un mínimo de 60 puntos (60 puntos = 4) con al menos 15 puntos entre los ejercicios 1 y 2. Este enunciado debe ser entregado junto con el parcial si quiere una copia del mismo puede bajarla del grupo de la materia. En el ejercicio 3 elija 2 de los 4 ejercicios y resuelva única y exclusivamente 2 ejercicios. **Si tiene dudas o consultas levante la mano**, está prohibido hablar desde el lugar, fumar o cualquier actividad que pueda molestar a los demás. El criterio de corrección de este examen estará disponible en forma pública en el grupo de la materia.

"C is quirky, flawed, and an enormous success.." (D. Ritchie)

#	1	2	3.1	3.2	4	5	6	7	Entrega Hojas:	Nombre:
Corr									Total:	Padrón:
Puntos	/15	/15	/10	/10	/15	/15	/10	/10	/100	Corregido por:

1) Se tiene un grafo dirigido representado mediante una lista de adyacencias. Dado un nodo inicial y un nodo final queremos que programe usando Map-Reduce un algoritmo que indique cuál es el camino mínimo entre ambos nodos. En clase vimos como calcular la longitud, ahora queremos saber cuál es el camino. (****) (15 pts)		2) Se tiene un archivo de texto de tipo (URL, user_id). Programar un script PIG que indique cuál es la URL visitada por mayor cantidad de usuarios. (**) (15pts)	
3) Resolver 2 (dos) y solo 2 de los siguientes ejercicios a elección (si resuelve mas de 2 el ejercicio vale 0 puntos, sin excepciones). En cada caso indicar V o F justificando adecuadamente sus respuestas. Si no justifica vale 0 puntos sin excepciones.			
a) Una función de hashing criptográfica produce muy pocas colisiones. (*) (10 pts)	b) La construcción de Davis-Meyer es necesaria para que la función de hashing produzca muy pocas colisiones. (***) (10 pts)	c) El efecto avalancha se produce cuando muchas claves hashlean a un mismo valor generando muchas colisiones. (**) (10 pts)	d) Una función de hashing para strings debe poder generar el resultado muy rapidamente. (**) (10 pts)
4) Suponiendo que solo son posibles los caracteres "A,B,C,EOF" comprimir el siguiente archivo usando PPMC de orden máximo 2 indicando el resultado final en binario del archivo comprimido. (****) (15 pts) "BCCCCC"		5) Considerando a cada línea como un documento y cada palabra como un término construya la matriz de términos x documentos usando TF-IDF. Finalmente indique cuantas dimensiones elegiría para usar LSI justificando adecuadamente su respuesta. (****) (15 pts) Salsa Tomate Pizza Muzzarella Tomate Fideos Tomate Salsa Fideos Fideos Pizza Tomate Tomate Fideos Salsa Salsa Salsa	
6) Si se tiene la siguiente familia de funciones LSH (0.15,0.85,0.85,0.15) indique de que forma quedaría amplificada usando r=3 y b=4. Finalmente interprete el resultado de la familia amplificada indicando que cantidad de falsos positivos o falsos negativos se producirían. (**) (10 pts)		7) Indique en que casos el código unario es mas conveniente que el código gamma o delta para almacenar los punteros de un índice invertido. Justifique adecuadamente. (****) (10 pts)	