

Es condición para aprobar sumar 3 puntos en la parte teórica y 3 puntos en la parte práctica.

1 – TEORIA

En cada punto marcar la opción correcta, pudiendo existir una o más de una opciones posibles.

- a) La principal diferencia entre DataLake y Datawarehouse (0,5 Pto.)
 a. Es que el primero permite realizar insights aplicando business analytics
☒ b. Es que el primero es un repositorio de datos sin una organización explícita al que cualquier usuario final puede acceder para extraer siempre información valiosa.
 c. Es que el primero es un ecosistema de información donde conviven distintas tecnologías de datos, organizado de forma tal que permita el análisis de información tanto estructurada como no estructurada.
 d. Es que el primero es un repositorio de datos donde se vuelca información de todo tipo para su posterior uso.
- b) Las siguientes técnicas de Tunning no se pueden aplicar en conjunto: (0,5 Pto.)
 a. Particiones + Índices
☒ b. Índice BitMap + Índice B-Tree
 c. Agregaciones + Estadísticas
☒ d. Ninguna de las anteriores
- c) Streaming Analytics se refiere a: (0,5 Pto.)
☒ a. Al proceso de buscar patrones en la información, haciendo uso de técnicas estadísticas y de minería de datos.
☒ b. A la capacidad de comunicar información en real time.
☒ c. La capacidad de realizar análisis de información mientras los eventos suceden y poder así tomar decisiones en tiempo real.
☒ d. A las tecnologías utilizadas para extraer información desde cualquier origen a partir de cualquier evento que en él se produzca.
- d) En general, una tabla de hechos (0,5 Pto.)
☒ a. Incluirá atributos de sus dimensiones que serán accedidas frecuentemente.
☒ b. Incluirá timestamps.
☒ c. Incluirá relaciones jerárquicas.
☒ d. Estará limitada a claves foráneas e indicadores.
- e)Cuál de los siguientes no es un componente típico de la arquitectura de tecnología de un DW (0,5 Pto.)
☒ a. Bases de datos, herramientas de ETL y herramientas de análisis
☒ b. Redes, sistemas de gestión de base de datos, y servidores
☒ c. Consultas, análisis, y herramientas de reportes.
☒ d. Plataformas de los sistemas fuentes.
- f) Cuáles de estas afirmaciones acerca de Datamining es incorrecta: (0,5 Pto.)
 a. La diferencia entre Datamining y estadística es la dirección hacia donde apunta la búsqueda.
 b. Uno de los objetivos de Datamining es encontrar patrones ocultos y repetitivos que expliquen el comportamiento de los datos en un contexto definido.
☒ c. A través de Datamining pueden encontrarse datos profundos en grandes volúmenes de información.
☒ d. Solo Datamining puede ayudar a predecir resultados futuros.
- g) Si en una carga de ETL se encuentran registros que no cumplen con las reglas de negocio, ¿qué solución NO debe seguirse? (0,5 Pto.)
☒ a. Descartar los casos, sin logueo.
☒ b. Rechazar los casos.
☒ c. Corregir los casos.
☒ d. Ninguna de las Anteriores.

1 - Teoría										2 - Práctica		Puntos	NOTA
1.a	1.b	1.c	1.d	1.e	1.f	1.g	1.h	1.i	1.j	2.a	2.b		
0,5 Pto	0,5 Pto	0,5 Pto	0,5 Pto	0,5 Pto	0,5 Pto	0,5 Pto	0,5 Pto	0,5 Pto	0,5 Pto	1 Ptos.	4 Ptos.		
0	0,5	0,5	0	0,5	0	0,5	0,5	0,5	0,5	0,75	2,5	6,5	6 (56%)

Apellido

Legajo:

- h) ¿Cómo se mide el Throughput Total (cantidad de I/O secuenciales) de un Datawarehouse? (0,5 Pto.)
- ☒ a. En MB/s.
 - ☐ b. En GB/s.
 - ☐ c. En KB/s.
 - ☐ d. En TB/s.
- i) ¿Que define a un científico de datos (0,5 Pto.)
- ☒ a. La habilidad de implementar insights valiosos para una compañía aplicando reglas de negocio.
 - ☐ b. El gran conocimiento de estadística y de diferentes tecnologías tanto de programación como manejo de datos.
 - ☒ c. La capacidad de superar el gap entre realizar análisis avanzado de datos y usar esos hallazgos para producir resultados alineados con los objetivos de una organización.
 - ☐ d. Ninguna de las anteriores
- j) OLAP y Data Mining son: (0,5 Pto.)
- ☒ a. Soluciones que compiten para la resolución de un mismo problema.
 - ☐ b. Sinónimos
 - ☒ c. Formas complementarias para realizar análisis de datos.
 - ☐ d. No tienen ninguna relación

2 – PRACTICA

Un Empleado de Recursos Humanos de una importante empresa ha comenzado a incursionar en Business Intelligence. Aprovechando lo nuevo del mercado sobre visualizaciones, ha trabajado en la construcción de un tablero a partir de información proveniente de planillas Excel que él solía elaborar.

El tablero ya ha sido aprobado por su Manager, quien solicito que el mismo sea publicado en el ambiente corporativo de Business Intelligence.

A raíz de esto, acudimos a usted para que a partir de la Imagen 1, diseñe el modelo para poder abastecer el tablero de acuerdo a las siguientes consignas:

- b) Diseñar el modelo Físico (Tablas y Columnas) a partir del modelo dimensional graficado en el punto anterior según las siguientes definiciones (4 Ptos.).
- Al menos una dimensión en modelo estrella (Con más de tres niveles)
 - Al menos una dimensión en copo de nieve, completamente desnormalizado (Con más de tres niveles)
 - Al menos una dimensión SCD tipo 2.
 - Al menos una dimensión degenerada.

Justificaciones atras

1 - Teoría										2 - Práctica		Puntos	NOTA
1.a	1.b	1.c	1.d	1.e	1.f	1.g	1.h	1.i	1.j	2.a	2.b		
0.5 Pto.	0.5 Pto.	0.5 Pto.	0.5 Pto.	0.5 Pto.	0.5 Pto.	0.5 Pto.	0.5 Pto.	0.5 Pto.	0.5 Pto.	1 Pos.	4 Pos.		

MSBPM EMPLOYMENT REPORT - FALL 2016

Imagen 1

Total number of graduates: 164

80%

% Placed within 3 months after graduation

% students seeking employment: 90%

Average sign-on bonus \$9,000*

Average Salary \$84,000

Work Experience

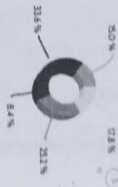
Work Experience	<2 yrs.	2-4 yrs.	>4 yrs.	Overall
Business Degree	\$71,000	\$72,000	\$74,000	\$72,000
Non-Business Degree	\$70,000	\$71,000	\$72,000	\$71,000
Master's Degree	\$75,000	\$76,000	\$77,000	\$76,000
PhD	\$80,000	\$81,000	\$82,000	\$81,000

Average Salary by Region

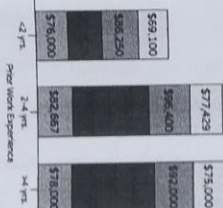
Presence of BAPM students



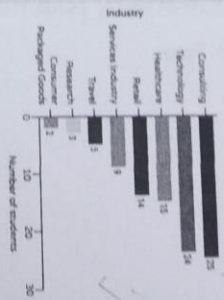
Job Titles



Average Salary by job titles



BAPM students by Industry



Average number of companies interviewed with

7

Average number of offers: 1.8

NO TONER
EN QUE NTA

La dimensión sc2 es companion, ya que puede comenzar el nombre de la empresa y no quiere perder la cantidad de entrevistados que se tuvo con ella y su nombre anterior.

CompSK	nombre	compasu	CompSK	nombre	version
1	mac	nombre a LOL	1	MAC	1
			2	LOL	2

cada vez que cambia agrega un nuevo SK, como lo nombre ^{entrevistado} y ^{segunda} versión. ^{ya que} Para la dimensión degenerada use "años de experiencia" (años CompSK) ya que el número solo de por sí deja en claro a que nos referimos y no tiene más ^{columnas} que agregar y que justifiquen una LK para este caso.

Para modelo estrella use la dimensión Geo1, la nombre Geo1.LK y puse todos los sububios en 1 misma tabla.

Agregue para llegar a 4 "Zona" ^{de} "peligrosidad" que se refiere a los lugares donde hay incidencia de huracanes y se afecta al estudiante y, "costa" para hacer una agrupación sea mas del lado del atlántico o del pacífico.

Decide separar años de experiencia de tiempo, porque puede ser que pase el tiempo y le persona este desempleada.

Para saber el máximo, mínimo y promedio ^{funciones} usa ~~datos~~ en base a salario y su BT.

Al igual que los promedios los averigua con los hechos de las BT y funciones.

Para saber si están ^{Buscando empleo} ~~desempleados~~ use otra degenerada dimensión en la BT de alumnos.

Necesito encontrar 4 dimensiones para hacer estrella y el copo de nieve por ese tiempo fue hasta 2010.

FISICO

siglo_LK
sigloSK
decadaSK

(ii) ✓

decada_UK

decadaSK
sigloSK
decadaSK

siglo_UK
sigloSK
decadaSK
decadaSK

Año_LK

AñoSK
decadaSK
sigloSK
decadaSK
decadaSK

Titulo_LK

TituloSK
descripcion

Alumno_RR

AñoSK
cicloSK
industriaSK
anilloSK
bitioSK
tempSK
BosondosSK

carA Alumnos

(iv) ✓

Geo1_LK

zonaSK
descripcion
comunaSK
descripcion
RegiónSK
descripcion
EstadoSK
descripcion

→ 51 B1 P1 M1

(i), P1 B1 B1

TORON 1

SOLA SK

Companies_LK

compSK
nomere

Industria_LK

industriaSK
descripcion

SALARIO_RT

bitioSK
AñoSK
RangoSK
AnilloSK
Borus
Salaries

Geografia_R_RT

mapSK
dibujo

(iii) SCD TIPS 2!

