**Teste teórico de analista de Business Intelligence**

**Candidato(a):** Thays Dami **Data:** 08/08/2020

**Questão 4 – Uma tabela de clientes possui uma coluna sexo com dois valores possíveis (M – Masculino e F – Feminino). Grande parte das consultas considera o sexo como critério de pesquisa na cláusula WHERE juntamente com outros campos. Que tipo de índice que deve ser utilizado nessa coluna?**

( ) Clustered Index.

( ) Nonclustered Index.

( ) Bitmap Index.

(x) Não deve ser utilizado um índice nessa coluna por sua alta densidade.

( ) Não deve ser utilizado um índice nessa coluna por sua alta seletividade.

**Questão 5 – De acordo com o T-SQL, quais são as cláusulas obrigatórias em uma query de SELECT?**

( ) As cláusulas FROM E SELECT.

( ) As cláusulas SELECT E WHERE.

(x) A cláusula SELECT.

( ) As cláusulas SELECT, FROM E WHERE

**Questão 7 - O que acontece após a execução do comando: SELECT TRY\_CAST(‘abc’ AS INT).**

( ) Um erro é gerado Um valor.

(x) null é retornado.

( ) Um valor inteiro é retornado.

( ) Uma string é retornada.

**Questão 8 - Em relação à clausula Where e Having podemos afirmar?**

( ) Ambas tem a mesma função.

( ) São funções diferentes.

(x) Ambas tem a mesma função mas o filtro da clausula where linha por linha e o Having após o agrupamento.

( ) Ambas acontecem durante o agrupamento.

**Questão 9 – Você está criando um pacote SSIS na sua máquina que aponta para uma base SQL Server com uma conta SQL e é executado via Job agendado. Após concluir o pacote remete para produção e no outro dia quando verifica o JobHistory tem o seguinte erro**

***DTS\_E\_OLEDBERROR. An OLE DB error has occurred. Error code: 0x80040E4D. An OLE DB record is available. Source: "Microsoft SQL Native Client" Hresult: 0x80040E4D Description: "Login failed for user ''."***

**O que você deve fazer para que o pacote execute corretamente a noite?**

( ) Mude todas as conexões para usar SQL Authentication.

(x) Mude todas as conexões para usar Windows Authentication

(x) Encriptar o pacote com "EncryptSensitiveWithPassword" ou "EncryptAllWithPassword" e forneça a senha cada vez que o usuário precisar executar.

(x) Crie um DTSConfig para fornecer informações de conexão para o pacote em tempo de execução.

**Questão 11 – Quais componentes são do MS-SQL Server Integration Services:**

( ) Designer SSIS, Cubos OLAP, Tarefas e Elementos de Fluxo de dados.

(x) Designer SSIS, Contêineres, Tarefas e Elementos de Fluxo de dados.

( ) Data Mart, Designer SSIS, Contêineres e Elementos de Fluxo de Dados.

( ) Data Mart, Designer SSIS, Tarefas e Elementos de Fluxo de Dados.

( ) Data Mart, Cubos OLAP, Contêineres e Tarefas.

**Questão 12 - Em um comando SQL, o operador LIKE é usado em uma cláusula WHERE para buscar um determinado padrão em uma coluna.**

(x) Certo.

( ) Errado.

**Questão 14 - Muitos autores consideram a tecnologia de Data Warehousing (o processo de fazer Data Warehouse) como sendo uma evolução natural do ambiente de apoio à decisão. As empresas utilizam Data Warehouse com mais frequência, pois há a necessidade de domínios de informações estratégicas que podem garantir respostas rápidas, assegurando, dessa forma, a competitividade no mercado concorrente e em constantes mudanças. O DW possui diversas características. “A arquitetura do Data Warehouse inclui, além de estrutura de dados, mecanismos de comunicação, processamento da informação para o usuário. ” Assinale, a seguir, a característica correspondente.**

( ) Não volátil.

( ) Integração.

( ) Variação de tempo.

( ) Orientado por assunto.

(x) Arquitetura do ambiente.

**Questão 15 - O objetivo dessa área é criar um ambiente intermediário de armazenamento e processamento dos dados oriundos de aplicações OLTP (Online Transaction Processing) e outras fontes, para o processo ETL (Extract Tranform Load), possibilitando seu tratamento, e permitindo sua posterior integração em formato e no tempo, evitando problemas após a criação do Data Warehouse e a concorrência com o ambiente transacional no consumo de recursos. A área citada é conhecida como:**

( ) Transaction area.

( ) Warehouse.

( ) Backup area.

(x) Staging area.

( ) Cube area.

**Questão 19 - VIEW é uma tabela virtual cujo conteúdo está definido por uma instrução SELECT**

(x) Certo.

( ) Errado.

**Questão 20 - No MS SQL Server, as tabelas criadas por meio do comando CREATE TABLE são temporárias se:**

( ) A opção TEMP é especificada logo após o termo CREATE.

( ) O comando é executado dentro de uma stored procedure.

( ) O usuário não possui privilégio para criação de tabelas.

(x) O nome da tabela é iniciado por #.

( ) A opção ON refere-se ao filegroup TEMP.

**Questão 21 – Descreva os modelos Start Schema (Ralph Kimball) e SnowFlake (Bill Inmon).**

O Star Schema (Esquema Estrela) é um modelo muito utilizado no Data Warehouse onde as dimensões ficam ligadas à uma tabela fato, que fica no centro, facilitando o apoio à decisão. Já o SnowFlake é outro modelo muito utilizado no Data Warehouse, e é bem semelhante ao Star Schema, mas possui mais dimensões. Ambos possuem pontos positivos e negativos para um DW e, o melhor modelo deve ser escolhido com base nas necessidades do projeto.

**Questão 23 –** O **que podemos entender por “Granularidade do dado”?**

A granularidade de dados se refere basicamente à quantidade de informações nos quais os campos são divididos. Um mesmo campo com uma grande quantidade de informações é pouco granular, porém, se estas informações forem segregadas em outros campos, os dados se tornam mais granulares e torna-se mais fácil de obter alguma informações em específico.

**Teste Big Data**

**2) Explain the difference between Amazon Athena and Redshift Spectrum as well as the main use cases for each of them.**

The both of services are used to query data using SQL, without managing infrastructure. The main difference is in cost, performance and functionality and the analyst needs to choose the best option for the project. Amazon Redshift is your best option when you need to gather data from several different sources - such as inventory systems, financial systems and retail sales systems - into a common format and store it for long periods in order to create business reports sophisticated based on historical data. Amazon Athena is great if you only need to run a quick query on some web logs to solve a performance problem at your location.

**3) You work for a start-up of photos processing and you need to swap the colors to black and white after loading them into Amazon S3. How can you do this on AWS??**

Creating an elastic transcoder preset

**5) Which technologies below are related to Big Data on Cloud?**

a. Kubernetes, Jenkins, Terraform

b. Azure SQL Server, AWS Lambda, AWS EC2

c. Google BigQuery, Apache Spark, Amazon Redshift

d. Digital Ocean, Packet, Javascript e. AWS, Google, Facebook

**6) Which file type is the best to read/write tabular data on big scales?**

a. CSV

b. Protobuf

c. Gzip

d. Parquet

e. JSON

f. Avro

**7) Choose all correct answers To real-time data processing which technology is best for the streaming layer?**

a. Apache Kafka

b. MySQL

c. MongoDB

d. Python

e. Apache Spark

**8) Explain the main points that define the concepts of ELT and ETL.**

ETL is “Extract, Transform and Load” process for data. ELT is “Extract, Load and Transform” process for data. Basically, in ETL the Load happens after the Transform and in ELT the Load happens before.

**9) Define in some lines the characteristics, 2 examples, and 2 use cases each for the following types of Databases:**

• Relational: A relational database is a type of database that stores and provides access to data points that are related to one another. Relational databases are based on the relational model, an intuitive, straightforward way of representing data in tables. In a relational database, each row in the table is a record with a unique ID called the key. The columns of the table hold attributes of the data, and each record usually has a value for each attribute, making it easy to establish the relationships among data points.. Two of the most famous is SQL Server and PostgreSQL and this type is frequently used to libraries and video store.

• Key Value: This kind of database work by storing arbitrary data accessible through a specific key. To store data, you provide a key and the blob of data you wish to save, for example a JSON object, an image, or plain text. To retrieve data, you provide the key and will then be given the blob of data back. In most basic implementations, the database does not evaluate the data it is storing and allows limited ways of interacting with it. Like example we hve Redis and Etcd and this type is frequently used to store configuration values and application variables and flags for websites and web applications.

• Documents: This type shares the basic access and retrieval semantics of key-value stores. Document databases also use a key to uniquely identify data within the database. As example we have MongoDB and RethinkDB and they’re good choice for rapid development because you can change the properties of the data you want to save at any point without altering existing structures or data.

• Graphs: Graph databases are a type of NoSQL database that takes a different approach to establishing relationships between data. Rather than mapping relationships with tables and foreign keys, graph databases establish connections using the concepts of nodes, edges, and properties. As example we have JanusGraph and Dgraph and they’re frequently used in social media.

• Timeseries: This type are data stores that focus on collecting and managing values that change over time. Although sometimes considered a subset of other database types, like key-value stores, time series databases are prevalent and unique enough to warrant their own consideration. As example we hava OpenTSDB and Prometheus databases and this type is frequently used tore monitoring or system performance information.in

• In-Memory: An in-memory is a database management system that primarily relies on main memory for computer data storage. As example we have SQL Server and MySQL and they’re frenquently used in applications where response time is critical.

**Teste Python**

**1. What is the avarage distance traveled by trips with a maximun of 2 passengers;**

Average distance = 2,981

**1. What is the avarage trip time on Saturdays and Sundays**

Average trip time = 0,326 hours