

Teste teórico de analista de Business Intelligence

Nome do Candidato (a): _____

Data ___ / ___ / ___

Crie um repositório no github e ponha os resultados e os código lá. Envie o link de acesso ao seu repositório criado.

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.
- ☐ () 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.
- ☐ () 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.
- ☐ 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.
- ☐ 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 '<User_Name>'."

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

- ☐ Mude todas as conexões para usar SQL Authentication.
- ☐ Mude todas as conexões para usar Windows Authentication
- ☐ Encriptar o pacote com "EncryptSensitiveWithPassword" ou "EncryptAllWithPassword" e forneça a senha cada vez que o usuário precisar executar.
- ☐ Crie um DTSCConfig 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.
- ☐ 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.

- ☐ 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.
- ☐ 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 Transform 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.
- ☐ Staging area.
- ☐ Cube area.

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

- ☐ 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.
- ☐ O nome da tabela é iniciado por #.
- ☐ A opção ON refere-se ao filegroup TEMP.

Questão 21 – Descreva os modelos Star Schema (Ralph Kimball) e Snowflake (Bill Inmon).

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

Teste Big Data (Daqui para baixo está em inglês)

1) You work on a start-up that developed a bracelet to track down data about the health of inpatients. Each bracelet sends the data in JSON every 6 seconds to be analyzed and stored. These data will be used to generate a daily report on the Health Portal and

you need to come up with a real-time solution for analytics that is durable, scalable and parallel to support the whole operation.

Describe and justify the possible choices for the following architecture components:

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

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??

4) An organization implemented a streaming solution, on which a data goes through a Kinesis Data Stream and a Kinesis Data Stream until it is stored on Redshift and is made available to analysis. A new product requirement specifies some events which should be processed with a minimum delay and could trigger some actions afterward.

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.

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

- Relational:
- Key Value:
- Documents:
- Graphs:
- Timeseries:
- In-Memory:

Teste Python

Baixe o arquivo e resposta as perguntas abaixo: (use pandas e numpy para lhe ajudar)

1. What is the average distance traveled by trips with a maximum of 2 passengers;
- 2 - What is the average trip time on Saturdays and Sundays;
- 3- To be able to provision your entire environment in a public cloud, preferably AWS.