

Questões de Data Analytics

1 - Para apresentar os resultados da consultoria você deve contextualizar o momento atual da indústria, passando por todas as revoluções industriais. Quais são os fatores que levaram à quarta revolução industrial?

A primeira revolução industrial ocorreu em 1764 com o desenvolvimento de máquinas a vapor e teares mecânicos, a segunda ocorreu em 1870 onde houve o desenvolvimento das linhas de produção, do aço, da energia elétrica e os motores. A terceira em 1970 com o avanço da eletrônica e o primeiro PLC.

A indústria 4.0 é o novo conceito de indústria, está sendo considerada a 4^a revolução industrial, ela engloba as principais inovações tecnológicas dos campos de automação, controle e tecnologia da informação, aplicada aos processos de manufatura.

Características da indústria 4.0: Operação em tempo real, virtualização, descentralização, orientação a Serviço, modularidade, conectividade, redução de custos, ajustes muito rápidos na linha de produção, personalização dos produtos, novos modelos de negócio.

O que se espera dos trabalhadores da 4 revolução industrial: novas habilidades e qualificações, redução de funções repetitivas e braçais, entre outras novas funções que devem surgir com o tempo.

2 - Algumas áreas possuem muitos dados a serem processados e a equipe técnica sugeriu que fossem usados algoritmos paralelos. Quais tipos de problemas podem ser resolvidos com algoritmos paralelos?

A programação paralela é a prática de dividir uma determinada tarefa em tarefas menores que possam ser executadas de forma simultânea e independente.

Os benefícios dessa técnica: utilizar melhor o poder de processamento, apresentar um melhor desempenho, permitir compartilhar dados e recursos e atender um maior número de usuários.

Os problemas que podem ser resolvidos com esses algoritmos possuem as seguintes características:

- Podem ser particionados em subproblemas ou unidades de trabalho que podem ser resolvidos simultaneamente.

- Podem executar múltiplas instruções a qualquer momento no decorrer da resolução do problema.
- Podem ser resolvidos em menor unidade de tempo com múltiplos recursos computacionais do que com único recurso computacional

3 - Considere as três tipologias de Computação em Nuvem apresentadas na disciplina. Explique brevemente como cada uma delas funciona.

A computação em nuvem é um paradigma da infraestrutura de computação, com disponibilização através da internet de servidores que podem ser reconfigurados dinamicamente com relação aos seus recursos de memória, armazenamento e processamento, ou seja, com alta escalabilidade. A nuvem pode ser pública, híbrida ou privada.

Os tipos de oferta de computação em nuvem são:

- SaaS: a aplicação funciona diretamente na nuvem. Não há necessidade de instalação na máquina do cliente.
- IaaS: disponibiliza os recursos necessários para que o usuário faça a implantação, configuração e utilização das suas aplicações.
- PaaS: o provedor fornece uma plataforma que é usada para desenvolver e disponibilizar as aplicações.

O nível de controle do usuário é baixo no SaaS, intermediário no PaaS, e alto no IaaS.

4 - Você deverá criar um banco de dados para armazenar todas as informações coletadas durante o projeto. Porém, você ainda deve decidir qual estrutura será usada. Quais as principais diferenças entre Data Warehouse, Data Mart e Data Lake?

Data Warehouse: é um conjunto de dados baseado em assuntos, integrado, não volátil em relação ao tempo, de apoio às decisões gerenciais. Pode ser relacional ou multidimensional.

Data Mart: refere-se a cada uma das partes de um Data Warehouse corporativo. É um subconjunto do DW que contém os dados para um setor específico da empresa, ou seja, corresponde às necessidades de informação de uma determinada comunidade de usuários.

Data Lake: são os dados em grandes volumes e em seu estado natural, vindos de todos os tipos de fontes, onde os usuários podem "mergulhar" e tirar amostras. É um lago cheio de dados.

5 - Qual a diferença entre IoT e Edge Computing?

O conceito de **IoT** apresenta dispositivos que possuem a capacidade de se conectar à internet para enviar ou receber dados.

Para a **Edge Computing**, ou Computação em Borda em português, esses dispositivos são ainda mais qualificados. Os dispositivos são capazes de realizar análises e processamentos avançados. É um modelo distribuído que pode evitar a necessidade do envio de dados para uma nuvem remota ou um sistema centralizado para realizar o processamento. Dessa forma facilita também processamentos de dados em tempo real. É uma grande tendência para o futuro da análise de dados.