

Short Questions:

Q.1 : Data Science: Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract or extrapolate knowledge and insights from noisy, structured and unstructured data. It extracts valuable information from data for business decision-making, strategic planning and other uses by combining math and statistics, computer science, specialized programming, advanced analytics, artificial intelligence (AI), and machine learning with specific subject matter expertise which can be used to guide decision making and strategic planning.

Q.2 : Artificial Intelligence: Artificial Intelligence (AI) refers to the simulation of human intelligence that is processed by computer systems. In other words, AI is the ability of a digital computer or computer-controlled system to perform tasks commonly associated with intelligent beings. Artificial intelligence, intelligence demonstrated by machines, is a field of data science that combines computer science and robust datasets to enable problem-solving.

Q.3 : Machine Learning: Machine Learning (ML) is a field of AI that teaches computer to learn from data and statistics, which can make prediction based on learning. It focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy. In its simplest form, ML is the process of using mathematical models of data to help a computer learn without direct instruction.

Q.4 : Future of data science: According to IBM, data science jobs will grow by 30% in this year. In future, data science would continue to evolve into a new phase depending on the need of the hour. The advancement in machine learning is the key contributor towards the future of data science. The increase of complex data science algorithms will be subsumed in packages in a magnitude making them quite easier to deploy, which is accomplished by data scientists. Data Scientists would exist as long as there exists data.

Real life applications of data science:

- Security (ex: Fraud detection, cyber risk analysis)
- Text, voice and advanced image recognition
- Transport (ex: Airline route planning)
- Targeted Advertising
- Healthcare industry (ex: Image analysis in medicine, genetics and genomics, medical diagnosis, drug development and virtual assistants such as health bots)
- Spam e-mails and comments
- Agriculture
- Weather forecasting
- Market basket analysis
- Banking and finance

Q.5 : Why python is so popular: Python is an immensely popular programming language for a long time because of the following characteristics:

- Simplified or English-like syntax
- Flexibility
- Supports OOP
- Versatility (ex: Python can be used in ML, AI, data science, web and game development, hardware programming)
- Having open resources (ex: Github, Youtube, Reddit, Quora, Medium, Stack Overflow, etc.) which makes it accessible to everyone
- Having a huge libraries for data science (ex: TensorFlow, Pandas, Matplotlib, NumPy, Keras, Scikit-learn, PyTorch, etc.)
- Having data frame libraries (ex: Django, Beautiful Soup)
- Having calculator libraries
- Platform independent

Future of python: Python has managed to reach a level of popularity that is higher than other popular programming languages like Java, C, C++ etc. in last 25 years. There is no other language that could compete with it as it is growing at a rapid pace. In future, it will surely remain one of the fastest growing programming languages in the world. Python will be the language of the future because it is planning to go to a new level with the involvement of Artificial Intelligence.

Q.6 : Why python for data analytics: For its immense capabilities of rapid development in data analytics, python is known worldwide. Python is hugely used in data analytics for the following reasons:

- Easy to learn (Simple and consistent)
- Built-in data analytics tools
- Stable and flexible
- Huge libraries collection
- Graphics and visualization

Q.7 : Importance of python in machine learning:

- Python is easy to understand allows for quick data validation.
- Python comes with a large number of libraries. Some of the libraries are scikit-learn, TensorFlow, pylearn2.
- Python allows easy and powerful implementation.
- Python has many modules that can be fastly used for machine learning.
- Friendly syntax and human-level readability.
- Python is having a huge community.

Q.8 : Importance of python in deep learning & NLP:

Natural Language Processing (NLP) is a popular machine learning technique used to analyze text content and deep learning is a part of machine learning. Python is now-a-days a must for deep learning and NLP for the following reasons:

- Extensive selection of libraries and frameworks
- Great community and popularity
- Platform independence
- Tokenization
- Stop word removal
- Stemming
- Lemmatization
- Modelling

Q.9 : Supervised Learning with examples: Supervised learning refers to the process of making an algorithm to learn to map an input to a particular output which is accomplished by using labeled training data to help us predict outcomes for unforeseen data. In other words, it is an approach to create artificial intelligence, where a computer algorithm is trained on input data that has been labeled for a particular output. Examples: Image and object recognition, predictive analysis (house price, spam detection), sentiment analysis, etc.

Q.10 : Unsupervised Learning with examples: Unsupervised learning refers to the process of making an algorithm to learn to analyze and cluster unlabeled datasets. It is a machine learning technique in which the users do not need to supervise the model. In other words, it is the use of AI algorithms to identify patterns in data sets containing data points that are neither classified nor labeled. Example: Dimension reduction and clustering, anomaly detection, neural networks, etc.

Jobs:

Q.11 : Role and responsibilities of a Database Engineer:

- Designing, architecting, building and continuously developing a dedication and highly available database infrastructure.
- Analyzing, planning, developing, deploying, and managing critical database systems.
- Using various partitioning schemes to optimize the database performance for scalability.
- Creating automation for repeating database tasks.
- Configuring, patching, upgrading and optimizing database software and related products.
- Assisting with testing, diagnosis and resolution of database issues.
- Recommending and implementing new features for Oracle database for high availability and scalability.
- Communicating, collaborating and working effectively in a global environment.

Q.12 : Role and responsibilities of a Data Analyst:

- Analyzing data models and structures to identify users, relationships, and dependencies.
- Formulating and fulfilling requests for data from users.
- Making data more accessible, understandable and usable and assisting with client deliverables.
- Developing test plans and cases for validating the integrity of data.
- Determining cause and effect of data anomalies and providing recommendations for procedural or programmatic resolutions.
- Designing and developing reports, data extracts and dashboard.
- Creating robust methods to detect missing or poor data.
- Having ability to effectively communicate in English, both verbally and written, with all levels of staff and management.

Q.13 : Role and responsibilities of a Data Engineer:

- Building data pipelines to bring together information from different source systems.
- Integrating, consolidating, cleansing data and structuring it for use in analytics applications following industry best practices.
- Troubleshooting software and processes for data consistency and integrity.
- Working with project manager, business users and other technical team members effectively.
- Documenting data design tasks or project requirements.
- Providing ongoing maintenance and process improvements for data initiatives.

Q.14 : Role and responsibilities of a Data Scientist:

- Applying a diverse set of tactics such as statistics, quantitative reasoning, and machine learning to analyze key performance indicators.
- Developing data driven models to forecast the performance and growth of products.
- Providing comprehensive analytics, develop tools and resources to empower data access and self-service to partner teams.
- Developing and maintaining high quality software for machine learning projects.
- Producing solutions driven by exploratory data analysis from complex and high-dimensional datasets.
- Contributing to the development, application and evaluation of predictive models and algorithms for extraction of knowledge from the data and exercise the ability to transfer skills across application domains.

Q.15 : Role and responsibilities of a Machine Learning Engineer:

- Applying machine learning technologies to build statistical models with large amounts of data.
- Analyzing large and complex data sets to derive valuable insights, inform feature and product development.
- Using machine learning, natural language processing, and graph analysis to solve modeling and ranking problems across discovery, ads and search.
- Writing clean, efficient, and sustainable code and executing it.
- Breaking down complex problems and implementing creative algorithm.
- Working with cross functional teams to architect on-sensor machine learning algorithms.
- Maintaining and improving the performance of existing machine learning solutions.
- Deploying ML models under the constraints of scalability, correctness, and maintainability.