Frequently Asked Questions from Project

1. What is the reason to Migrate/Develop the Project using Hadoop/Spark?
2. What is the Data Ingestion mechanism in your project?
3. In which format your project data is coming? (is it a delimited data, fixed length data, completely unstructured data etc)
4. What is size (volume) of data in your project?
5. What is the per day MAX load of data size to your cluster?
6. What type of hadoop project you involved? Is it on premises/cloud integrated?
7. What is your project cluster size? (Sometimes...explain your node/ram capacity).
8. Have you faced any Data transfer ERRORs while you are copying the data to HDFS? How did you resolve the same error? Can you explain any one such scenario?
9. Can you explain the architecture of your project?
10. . What are the tools used as part of your project? And for which purpose (WinScp/FileZilla🡪 for data transfer from one os to another, Putty for Remote Connect of Nodes,SVN/CVS/GitHub for Code Repository, JIRA for Issue Tracker, Citrix to connect to our VPN(Virtual Private Network) etc)
11. . Can you list out the activities you do on a daily basis as part of your project?
12. . What is your team structure/ team size/team distribution? How do your coordinate with your team members in case of dependencies?
13. . Explain any biggest technical challenge faced/solved out of your project deliverables?
14. . Which process methodology you are following in the project?
15. . Who is your architect/lead of your project?

.