

PROJECT SPECIFICATION

Architect Enterprise Data Lake for Medical Data Processing Company

Network Diagram

CRITERIA	MEETS SPECIFICATIONS
Use best practices when creating diagrams.	Diagram includes <ul style="list-style-type: none"><li>Label for all components</li><li>The diagram must include visual components representing the 4 layers: Ingestion, Processing, Storage and Serving</li><li>Layout should be readable and easy to follow</li></ul>
Create a detailed end-to-end architecture diagram	Diagram shows <ul style="list-style-type: none"><li>4 layers are represented in a way which communicates design principles</li><li>Where to keep the metadata information</li><li>Logos or tool name used in Storage and Serving layer</li><li>Multiple tool logos or names used in Ingestion &amp; Processing layers</li></ul>

Design Document

CRITERIA	MEETS SPECIFICATIONS
Question #1 <ul style="list-style-type: none"><li>Summarize the purpose of the document and identify the business scope of the project</li></ul>	Design document <ul style="list-style-type: none"><li>Defines “what” and “why” in less than 10 sentences.</li><li>Defines the target audience</li><li>Identifies at least 3 in-scope elements</li><li>Identifies at least 3 out of scope elements</li></ul>
Questions #2, #3 <ul style="list-style-type: none"><li>Document requirements and define architectural design principles</li></ul>	Design document includes <ul style="list-style-type: none"><li>Summary is provided of the problem statement and business requirements</li><li>Three design principles are identified and justification is provided to show how each principle is aligned with the company's technical and business requirements</li></ul>
Question #4 <ul style="list-style-type: none"><li>Define assumptions and risks</li></ul>	Design document includes <ul style="list-style-type: none"><li>3 relevant assumptions are explained, as well as how those assumptions will impact the design</li><li>Potential current and future risks are described</li></ul>
Question #5 (already completed in section 1 above)	Question #5 (already completed in section 1 above). No additional work is needed for this step.
Question # 6 (part 1) <ul style="list-style-type: none"><li>Design an effective Ingestion layer in a design architecture</li></ul>	Design document includes <ul style="list-style-type: none"><li>Plan for ingesting different types of data and data sources</li><li>Required tools are listed and justified</li><li>Plan for scaling</li><li>List of at least 3 tools that were considered but not selected, including an explanation of why each tool was not selected</li></ul>
Question # 6 (part 2) <ul style="list-style-type: none"><li>Design an effective Storage layer in a design architecture</li></ul>	Design document includes <ul style="list-style-type: none"><li>Plan to store a vast amount of data</li><li>Plan for handling 20% YoY Data Growth rate</li><li>Plan &amp; strategies to handle back-up and recovery</li><li>Plan to store custom metadata information, including what type of information is held by metadata</li><li>Explanation for selection of data format</li><li>Plan to secure data, including at least 2 possible techniques, tools, and/or considerations</li><li>List of any tools that were considered but not selected, including explanation of why each tool was not selected. Identify why any 3rd party tools could solve problems</li></ul>
Question # 6 (part 3) <ul style="list-style-type: none"><li>Design an effective &amp; scalable Processing layer in a design architecture</li></ul>	Design document includes <ul style="list-style-type: none"><li>Plan to process the data</li><li>Plan to enable ad-hoc querying capabilities</li><li>Plan to satisfy different processing needs</li><li>Identification of different tools involved</li><li>List of any tools that were considered but not selected, including an explanation of why each tool was not selected. Identify why any 3rd party tools could solve problems</li><li>Plan for scaling</li></ul>
Question # 6 (part 4) <ul style="list-style-type: none"><li>Design an effective Serving layer in a design architecture</li></ul>	Design document includes <ul style="list-style-type: none"><li>What is meant by serving layer?</li><li>Description of types(s) of stored data</li><li>Description of how the data is used</li></ul>
Questions #7 <ul style="list-style-type: none"><li>Effectively evaluating different storage and processing frameworks</li></ul>	Design document includes: <ul style="list-style-type: none"><li>Concluding &amp; relevant synthesized thoughts and intuition that can be used in the next steps of the project.</li></ul>
Question #8	Design document includes <ul style="list-style-type: none"><li>Links that students used, if any</li></ul>

Slide Show

CRITERIA	MEETS SPECIFICATIONS
Present & summarize about data lake	Slide show includes: <ul style="list-style-type: none"><li>Definition of data lake</li><li>Definition what a data lake is used for</li></ul>
Present components of the data lake	Slide show : <ul style="list-style-type: none"><li>Defines at-least 4 components (or modules or layers) which make up a data lake</li><li>Describes briefly each layer and what they do</li></ul>
Present different between Data Warehouse and Data Lake	Slide Show includes at least 4 unique differentiators between Data Lake and Data Warehouse
Present business value of the data lake	Slide show provides at least 4 business values of the data lake proposed design solution. The values must directly relate and solve the technical and business requirements of the company
Present data lake architecture design diagram to reinforce proposal	Slide show contains the same architecture diagram which was submitted as deliverable item #1.

Recorded Presentation

CRITERIA	MEETS SPECIFICATIONS
Articulate the Data Lake components, business value, and rationale behind the architecture in order to persuade a non-technical audience. ( The executive leadership of the company)	<ul style="list-style-type: none"><li>Video maximum length is 10 mins, minimum length is 6 mins.</li><li>Verbal component includes appropriate elaborations as appropriate on all slides. (expanding on key points)</li><li>Relevant data and examples are used to describe how Data Lake will solve the technical and business requirements for the company</li></ul>

Suggestions to Make Your Project Stand Out!

- Provide details on alternative tools for each layer along with the advantages and disadvantages of each approach.
- Provide recommendation on data archival strategy, describe how hot data vs cold data can be handled
- Research and provide other Data Lake public references from other companies who have successfully embraced a Data Lake strategy