MARINE ASSESSMENT SERVICES

To design a detailed Data Science approach for the customer's problem of augmenting surveyors in answering their queries and reducing the burden on Subject Matter Experts (SMEs) while considering multiple geographies and time zones, we can follow these steps:

1. Data Collection and Pre-processing:

a) Gather the internal documents in the form of PDFs and Word documents from the client.
b) Extract text from these documents and pre-process it to remove noise, such as special characters, stop words, and perform stemming/lemmatization.
c) Collect and pre-process the regulatory information specific to the geographies of interest from the provided HTML URLs.

2. Knowledge Base Creation:

a) Combine the pre-processed text data from all sources to create a comprehensive knowledge base.
b) b. Organize the knowledge base by categories, such as vessel types, safety regulations, maintenance procedures, etc.

3. Intent Recognition:

a) Implement an intent recognition system using Natural Language Processing (NLP) techniques to identify the purpose and context of the surveyor's query.
b) Train the intent recognition model on a labeled dataset containing example queries and their corresponding intents.

4. Query Understanding:

a. Develop a query understanding module that extracts relevant keywords and entities from the surveyor's query using Named Entity Recognition (NER) and keyword extraction techniques.
b. This step will help in identifying the key elements of the query to find relevant information in the knowledge base.

5. Response Generation:

a. Based on the recognized intent and extracted query elements, create an answer retrieval mechanism to fetch relevant information from the knowledge base.
b. Utilize NLP techniques like TF-IDF, word embeddings, or transformer-based models (such as BERT) to find the most relevant documents or paragraphs.
c. Rank the retrieved information based on relevance and credibility to generate a concise and accurate response.

6. Multilingual and Geographical Considerations:

a. If the surveyors operate in multiple geographies with different languages, incorporate multilingual support in the model.
b. Use language detection techniques to identify the language of the surveyor's query and provide responses in the appropriate language.

c. Ensure that the knowledge base includes regulatory information specific to different geographies.

7. Time Zone Adaptation:

a. Consider time zone differences to ensure timely responses to surveyor queries.
b. Implement a system that prioritizes queries based on their urgency and time sensitivity.
c. Notify surveyors about the expected response time, especially when dealing with critical issues.

8. Continuous Learning and Improvement:

a. Set up a feedback mechanism to gather user feedback on the responses provided by the QueryBoT.
b. Use this feedback to continuously improve the system's performance, update the knowledge base, and enhance the intent recognition and response generation modules.

9. Deployment and Integration:

a. Deploy the QueryBoT as a user-friendly web-based interface accessible to surveyors across geographies.
b. Integrate the system with the client's existing communication platform or ticketing system for seamless interaction between surveyors and the QueryBoT.

10. Security and Privacy:

a. Ensure that the knowledge base and user data are stored securely and comply with data privacy regulations.
b. Implement necessary measures to protect sensitive information.

By following the above Data Science approach, the customer can leverage NLP technology to empower surveyors with faster query clarifications, reduce the burden on SMEs, and enhance the efficiency and accuracy of their marine assessment services worldwide.