REPORT

Upon a comprehensive analysis of the provided use cases—Computer Vision and Clinical Natural Language Technology for Health Care, Machine Learning and Deep Learning in Health Care, and Interoperability Technology for Health Care—I have concluded that consolidating all requirements within a unified cloud-native environment, supported by multiple microservices, offers a cost-effective solution. This integrated approach not only streamlines operations but also presents an opportunity to enhance workforce productivity while mitigating overhead usage costs.

Computer Vision & Clinical Natural Language Technology for Health Care:
The selected use case involves Computer Vision and Clinical Natural Language Technology for reviewing unstructured and semi-structured clinical records. This includes reducing medical errors and improving diagnosis, prescriptions, care plans, and predictions.

Azure Databricks for Warehousing:
Advantage: Databricks provides a unified analytics platform, seamlessly integrating structured and unstructured data processing. Its scalability, flexibility, and support for advanced analytics make it an ideal choice for diverse warehousing tasks.
Why Not Synapse: While Synapse Analytics excels in structured data warehousing, Databricks' versatility and advanced analytics capabilities make it more suitable for this scenario.

Azure Form Recognizer for Document Understanding:
Advantage: Azure Form Recognizer enhances document understanding, automating the extraction of structured information from clinical records. This specialized service ensures accurate data extraction for further analysis.
Integration with Databricks: By integrating Form Recognizer with Databricks, the extracted data becomes an integral part of the analytical pipeline, enriching the dataset and enabling more insightful analytics.

Azure Data Lake Storage (ADLS):
Advantage: ADLS is designed to handle both structured and unstructured data at scale. In healthcare analytics, clinical records often come in various formats, including images, text, and structured data. ADLS accommodates this diversity efficiently.

Automated Data Extraction:
Opportunity: Utilize Azure Form Recognizer to automate the extraction of structured information from clinical records.
Strategic Action: Integrate Form Recognizer into the Databricks workflow to automate the data extraction process, reducing manual efforts and minimizing errors.

Machine Learning & Deep Learning in Health Care:

Azure Databricks for Analytics and Machine Learning:
Advantage: Azure Databricks offers a unified analytics platform seamlessly integrated with Apache Spark, demonstrating excellence in managing diverse data types, including unstructured data. This makes it the ideal solution for healthcare analytics. Moreover, it provides pre-attached Python Jupyter notebooks, enabling the execution of various algorithms such as Random Forest, and enhancing the platform's versatility for advanced analytics tasks.

Integration with Azure Machine Learning: Databricks can be integrated with Azure Machine Learning to enhance machine learning workflows, combining collaborative analytics with advanced model training and deployment.

Azure Time Series Insights for Anomaly Detection:
Advantage: For time-series anomaly detection in healthcare operations, Azure Time Series Insights can be employed. It complements Databricks by providing insights into time-related patterns.

Why not HDInsight?
While HDInsight is a robust big-data analytics platform provided by Azure, I've carefully considered the limitations evident in the data provided, specifically regarding the volume or size of data. After a thorough analysis of available options, I've opted for Databricks as the preferred choice.

Interoperability Technology for Health Care :

Azure API Management for Scalable API Architectures:
Advantage: Azure API Management facilitates the creation, publishing, and management of APIs, offering a scalable and cloud-native solution for interoperability. It ensures the seamless integration of disparate healthcare systems.
Integration with Azure Streaming and Batch Services: Azure API Management can be integrated with Azure Streaming and Batch services to enable both streaming and batch ETL/ELT processes.

Azure Event Hubs for Streaming Data Ingestion:

Advantage: Azure Event Hubs provides a scalable and real-time event streaming platform. It is well-suited for ingesting streaming data from diverse healthcare sources, ensuring timely and efficient interoperability.
Integration with Azure API Management: Event Hubs can be integrated with Azure API Management to handle streaming data as part of the interoperability architecture.

Azure Data Factory as Orchestrator(For building ETL/ELT pipelines in conjunction with Azure DataBricks:

Advantage: Azure Data Factory serves as the orchestrator for unified data integration. It efficiently manages both batch ETL and ELT processes, ensuring seamless handling of unstructured clinical PDF documents and structured or semi-structured JSON medical claims and records.
Integration with Azure API Management: Azure Data Factory orchestrates the interoperability workflow, integrating seamlessly with Azure API Management and other services.

**Now all this can be also implemented in the AWS environment, below are the microservices that can be used instead:
Azure From Recognizer: ---> Amazon Textract
Azure Databricks for Warehousing: ---> Amazon Redshift
Azure Form Recognizer for Document Understanding: ---> Amazon Textract
Azure Data Lake Storage (ADLS): ---> Amazon S3 (Simple Storage Service)
Machine Learning & Deep Learning in Health Care: ---> Amazon SageMaker
Azure Time Series Insights for Anomaly Detection: ---> Amazon CloudWatch Anomaly Detection
Azure API Management for Scalable API Architectures: ---> Amazon API Gateway
Azure Event Hubs for Streaming Data Ingestion: ---> Amazon Kinesis
Azure Data Factory as Orchestrator: ---> AWS Step Functions

** Opting for an all-open-source tool approach, including building our Optical Character Recognition (OCR), may initially require a slightly higher capital investment. However, it comes with potential challenges, such as reduced support in case of issues compared to leveraging cloud-native services. Another consideration is platform limitations.
My insights are drawn from extensive experience collaborating with various multinational corporations. In practice, the predominant tools are those adopted by organizations lacking their internal software running on specific infrastructure.
In essence, if tasked with designing a solution based on the provided data, an ideal scenario would involve having access to critical details such as transaction data volume and the size of existing databases or archived files. This comprehensive understanding would further refine my approach to providing better solutions.

Bibliography:

<Proctored & Paid certificate Examinations taken>
Exam DP-203: Data Engineering on Microsoft Azure
Exam AI-102: Designing and implementing a Microsoft Azure AI Solution
Exam DP-100: Designing and implementing a Data Science solution in Azure.