**STEDI Human Balance Analytics – Big Data Architecture Proposal (Azure-Based)**

**A. Executive Summary**

This proposal outlines how STEDI can build a complete big data analytics system on **Microsoft Azure** to handle human balance data from its wearable devices. The goal is to create a setup that collects sensor data in real time, processes it efficiently, and turns it into insights the company can use to improve device calibration, product development, and research accuracy.

The outcome of this project is a scalable and secure analytics environment where STEDI can store, process, and visualize large volumes of sensor data. Once this is in place, the company will be able to see live data streams, run machine learning models that predict balance stability, and make faster, more informed decisions. The end benefit is a smarter, data-driven workflow that supports better product quality and customer outcomes.

**B. Business Problem Recap**

STEDI’s technology relies on collecting data from thousands of IoT-enabled balance devices. Each device constantly sends readings that measure motion, angle, and pressure. The company’s main challenge is that it collects too much data for its current systems to manage effectively.

The data arrives fast, in huge quantities, and in many different formats. Right now, this creates delays in analysis and prevents the team from making quick adjustments or training accurate predictive models. Without a scalable analytics framework, important information about customer performance and product calibration gets buried in raw data. Solving this problem means STEDI can turn those readings into real-time insights and long-term product improvements.

**C. Needs Assessment**

1. **How Big Data Analytics Helps**

Big data analytics is exactly what STEDI needs to make sense of all that sensor data. Azure’s cloud services can automatically collect information from devices, clean and organize it, and then make it available for analysis and visualization. By setting up automated pipelines, STEDI can track balance performance across all users, identify trends, and predict stability issues before they happen.

This approach also supports machine learning, so the company can train models that help fine-tune the balance devices or detect patterns that might indicate fall risk. In short, big data analytics lets STEDI go from reactive analysis to proactive insight.

1. **Data Analysis Methods**

The solution will combine several types of analysis:

* **Streaming analysis** using Azure Stream Analytics for real-time sensor data.
* **Batch analysis** through Azure Databricks to study long-term trends.
* **Predictive modeling** with Azure Machine Learning to forecast balance loss.
* **Visualization and reporting** in Power BI to make the results easy for stakeholders to understand.

These tools work together so STEDI can see both the moment-by-moment data and the bigger patterns over time.

**D. Solution Design**

1. **Databases and Processing Systems**

The core of this solution will include:

* **Azure IoT Hub** to connect and manage all balance devices securely.
* **Azure Event Hubs** to handle real-time streaming data.
* **Azure Data Lake Storage Gen2** to store both raw and processed data.
* **Azure Databricks** for running ETL (extract, transform, load) processes and building machine learning models.
* **Azure Synapse Analytics** for querying, analyzing, and preparing structured data.

1. **Large-Scale Data Models**

The system will use two main models:

* A **time-series model** for all the raw sensor readings, organized by device and timestamp.
* A **dimensional (star-schema) model** in Synapse for reporting and analytics, with fact tables for balance scores and dimension tables for customers, devices, and training sessions.

1. **Integrated Analytics Solution**

Here’s how the data will move:  
IoT Device → IoT Hub → Event Hub → Data Lake (raw) → Databricks (clean and transform) → Synapse (query and report) → Power BI (visualize).

Machine learning models trained in Azure Machine Learning will generate balance scores and predictive insights that appear directly in Power BI dashboards. Data Factory pipelines will automate this entire workflow.

1. **Supporting Services**

* **Orchestration:** Azure Data Factory will schedule and monitor the data flows.
* **Security:** Azure Active Directory will manage user access; all data will be encrypted in transit and at rest.
* **Monitoring:** Azure Monitor and Log Analytics will track performance, errors, and usage.

**E. Justification of Choices**

1. **Reasoning Behind Each Choice**

Every tool in this design was chosen to balance scalability, cost, and ease of integration. Azure IoT Hub and Event Hubs make it simple to handle thousands of devices sending data at once. Databricks is built on Spark, so it can process massive datasets quickly and supports advanced analytics in Python or SQL. Synapse Analytics serves as a powerful data warehouse that connects smoothly with Power BI for dashboards.

These tools fit naturally within STEDI’s Microsoft environment, meaning there’s less setup friction and better long-term support.

1. **How the Design Meets the Business Need**

This architecture directly addresses STEDI’s data challenges. It can handle both real-time and historical data, allowing the company to run immediate analyses while still maintaining full historical records. The system supports predictive analytics, improves security, and scales easily as the number of devices grows. Most importantly, it turns STEDI’s existing raw sensor data into clear, visual insights that can guide engineering, marketing, and product teams alike.

**F. Future Enhancements**

Once the system is running smoothly, STEDI could expand it in several ways:

* Use **Azure Digital Twins** to simulate human-device interaction for testing and training.
* Add **edge processing** with Azure IoT Edge so devices can process simple tasks locally before sending data to the cloud.
* Train more advanced machine learning models using deep learning methods.
* Integrate data from other products, such as smart shoes or posture trackers, to create a full picture of human balance.
* Automate retraining of machine learning models whenever new data arrives.

These updates would continue improving accuracy and efficiency over time.

**G. Implementation Plan**

Here’s a simple outline of how to build this system:

1. **Set up core infrastructure** (IoT Hub, Event Hub, Data Lake).
2. **Connect and configure devices** to start sending live data.
3. **Create data pipelines** with Data Factory and process data using Databricks.
4. **Load curated data** into Synapse for analytics.
5. **Train and deploy machine learning models** in Azure ML.
6. **Design Power BI dashboards** to share insights with key teams.
7. **Test, secure, and deploy** the solution for production use.

A realistic timeline for full rollout is around **four to six months**, with incremental deliverables after each stage.

**H. References**

* Microsoft Azure Architecture Center. (2025). *Big Data and Analytics Solutions.* https://learn.microsoft.com/azure/architecture
* Udacity. (n.d.). *STEDI Human Balance Analytics Case Study.* Data Engineering Nanodegree.
* GitHub repositories used for conceptual reference:
  + <https://github.com/deepSalva/project-stedi-human-balance-analytics>
  + <https://github.com/pkdaisu/STEDI-Human-Balance-Analytics>
  + <https://github.com/faqeel/udacity-dend-stedi-human-balance-analytics>
* Microsoft Azure Databricks Documentation. (2025). https://learn.microsoft.com/azure/databricks
* Microsoft Azure Synapse Analytics Documentation. (2025). https://learn.microsoft.com/azure/synapse-analytics