**MLOps Assignment:**

There are many possibilities for architecting (even more online/continual learning is desired).

Here is a high level architecture diagram.

A diagram of a software development process

Description automatically generated

1. **Data Ingestion Layer**:
   * Use **Azure Event Hubs** to ingest real-time data from the 1000 devices.
   * Event Hubs can handle the high throughput of data from the 100 sensors per device.
2. **Data Storage**:
   * Store raw data in **Azure Blob Storage** or **Azure Data Lake Storage** for large-scale, unstructured data storage.
3. **Data Processing and Transformation**:
   * Employ **Azure Databricks** for data processing and transformation.
   * Use Spark clusters in Databricks to read the .parquet files, preprocess the data, and do exploratory analysis, and some feature engineering.
4. **Processed Data Storage:**
   * Processed and cleaned data can be stored in a more structured format in **Azure SQL Database** or **Azure Cosmos DB** for efficient querying.
   * SQL can also be used to store predictions over a small window to feed into the dashboard.
5. **Model Training and Evaluation**:
   * Leverage **Azure Machine Learning Service** for training and evaluating machine learning models.
   * Use Azure ML pipelines to automate the model training and evaluation process, which can be scheduled or triggered as needed.
6. **Model Deployment**:
   * Deploy the trained models to **Azure Kubernetes Service (AKS)** for model serving.
   * AKS can host your machine learning models as containerized services.
7. **Web-Hosted Dashboard**:
   * Develop a dashboard using **Azure Web Apps** or **Azure App Service**.
   * The dashboard can visualize data, display model predictions, and alert users about anomalies.
8. **Alerting and Notification System**:
   * Implement an alerting system using **Azure Functions** and **Logic Apps** for real-time anomaly detection notifications.
   * Integrate with email services or other notification systems for sending alerts to users.
9. **Monitoring and Logging**:
   * Use **Azure Monitor** and **Application Insights** for monitoring the performance of your web applications and AKS deployments.
   * Store and analyze logs using **Azure Log Analytics** for insights into the system's operation and performance.

**Optional Part:**

* **Online and Batch Model Updates**: Integrate a system for online learning using Azure Stream Analytics for real-time data processing and model updating.  
  Online learning is an interesting topic. Features for online models also need to be incremental. Neural Networks and SGD are inherently stateful learners, where incremental updates are possible. Stream analytics has isolation forests possibility as well.  
   For slow batch updates, use Azure ML pipelines to retrain models with the full dataset on a regular schedule.
* **Multi-Team Collaboration**: Utilize Azure DevOps for code management, version control, and CI/CD pipelines. Different teams have different responsibilities, so stakeholder and expectation management is really important. Regular updates and client meetings are also critical.
* **External Data Integration**: If some data is external, consider using **Azure Data Factory** for data integration. It can automate the movement and transformation of data from various external sources into your system.