High-Level Architecture:

Data Sources:

Historical Health Data

Any external data (like regional health trends or generic health advice)

Data Ingestion Layer:

Tools and processes to import and initially process the data.

Data Storage Layer:

Raw Data Storage (Data Lake)

Processed Data Storage (Data Warehouse)

Processing & Analytics Layer:

Data Cleaning & Transformation

Machine Learning Modeling

Consumption Layer:

Dashboards & Reporting

Recommendation Delivery

Management & Monitoring:

Security & Compliance

Monitoring Tools

Low-Level Architecture:

Data Sources:

Historical Health Data:

Patient Profiles

Health Metrics from wearables

Historical health events or incidents

External Data Sources:

APIs for health trends

CSVs or databases with generic health advice or benchmarks

Data Ingestion Layer:

Batch Processing:

Tools: Apache Nifi, Talend, Sqoop

Process: Scheduled data ingestion jobs, error handling

Data Storage Layer:

Data Lake:

Tools: Hadoop HDFS, AWS S3

Structure: Raw files, possibly partitioned by ingestion date or data source

Data Warehouse:

Tools: Snowflake, Redshift, BigQuery

Structure: Organized tables optimized for query performance, Star or Snowflake schema

Processing & Analytics Layer:

ETL Processes:

Tools: Apache Spark, Google Cloud Dataflow

Process: Data cleaning, transformation, and loading into the data warehouse

Machine Learning:

Tools: Scikit-learn, TensorFlow, PyTorch for modeling; Apache Mahout for recommendations

Process: Anomaly detection, health outcome prediction, recommendation engine training

Consumption Layer:

BI Dashboards & Reporting:

Tools: Tableau, PowerBI

Purpose: Visualize data insights, historical trends, and anomalies

Recommendation Delivery:

Mechanism: Web interfaces, mobile apps, email systems

Management & Monitoring:

Security & Compliance:

Encryption at rest and in transit

Access controls and user authentication

Monitoring & Maintenance:

Tools: Grafana, Prometheus, AWS CloudWatch

Process: Monitor data ingestion, ETL, and model performance; alerting for anomalies or issues