Use-Case Description:

MediSecure HealthTech, a hypothetical healthcare analytics startup based in San Francisco, aims to use sensitive medical data collected from multiple hospitals across the U.S. and Europe. The company's goal is to build a real-time predictive analytics platform to help diagnose and treat cardiovascular diseases effectively. Their dataset includes about 100,000 unique patient records, covering highly sensitive details such as demographics, diagnostics, genetic markers, and lab results. The platform must provide highly accurate predictions quickly—ideally in real-time—while strongly protecting patient privacy and minimizing risks of re-identification.

Decision-Making Tool Walkthrough:

MediSecure used our PET Advisor to help select appropriate privacy technologies. The process included answering a total of 27 key questions to help PET Advisor have a clear understanding of the situation and then guided them to the best privacy-preserving solutions with implementation techniques. For example, when asked, "How much risk of re-identification is acceptable?" the team chose "Very low (<1%)" because of the highly sensitive nature of medical data. This decision highlighted Differential Privacy (DP) as especially suitable, giving it higher priority due to its strong mathematical guarantees against re-identification.

Next, when asked about their preferred privacy protection method considering their risk tolerance, MediSecure selected "I want rigorous mathematical privacy guarantees even if some randomness is introduced." This further confirmed that Differential Privacy was ideal, reinforcing its selection. Meanwhile, given their need for fast, accurate results, the tool decided to exclude public-key cryptography because it would introduce latency issues. Instead, based on the scoring system integrated inside the tool, the final decision opted to combine Differential Privacy with Trusted Execution Environments (TEEs), which align well with their real-time analytics requirements. Furthermore, MediSecure team will also be asked questions regarding to law compliance. For example, questions like, does your dataset include medical or health-related information? Do you collect or process information about people in the EU/EEA? If the answers for both are yes, the tool will include HIPAA and GDPR in the law section to tell the team these are the potential laws that the data should obey.

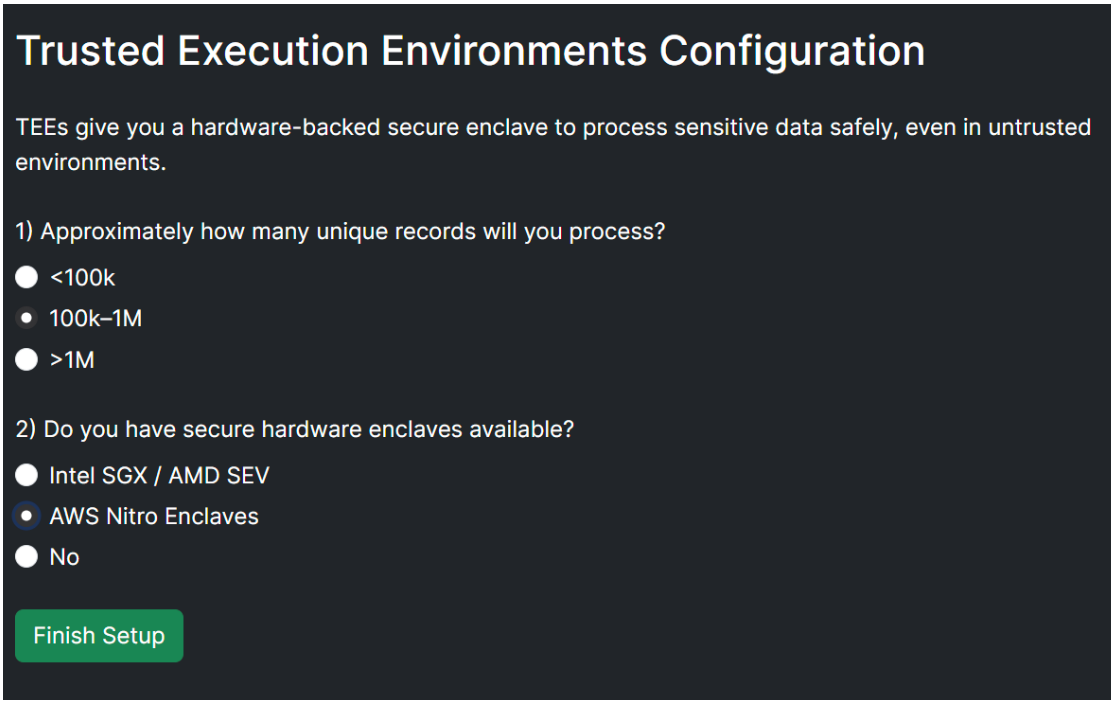
Concrete Recommendations:

Based on the decision-making tool, MediSecure received the following privacy-preserving recommendations with implementation guidance:

- **Differential Privacy (DP)**:

    - Privacy budget: ε ≈ 20 per query, total ε ≈ 1000 per day (based on approximately 50 queries daily).
    - Implementation tip: Calibrate DP methods (Laplace or Gaussian mechanisms) to meet these ε-values, managing the privacy budget carefully each day.
- **Trusted Execution Environments (TEEs)**:
    - Recommended use of AWS Nitro Enclaves, leveraging AWS Key Management Service (KMS) for secure enclave attestation.
    - Follow AWS Nitro CLI documentation closely to ensure proper setup and management.
- **Legal and Regulatory Compliance**:
    - Compliance with HIPAA, GDPR, and CCPA/CPRA is mandatory due to the sensitive medical data collected from multiple jurisdictions.

The implementation guidance was based on some further questions ask as shown in the following picture for TEE and DP.



Trusted Execution Environments Configuration

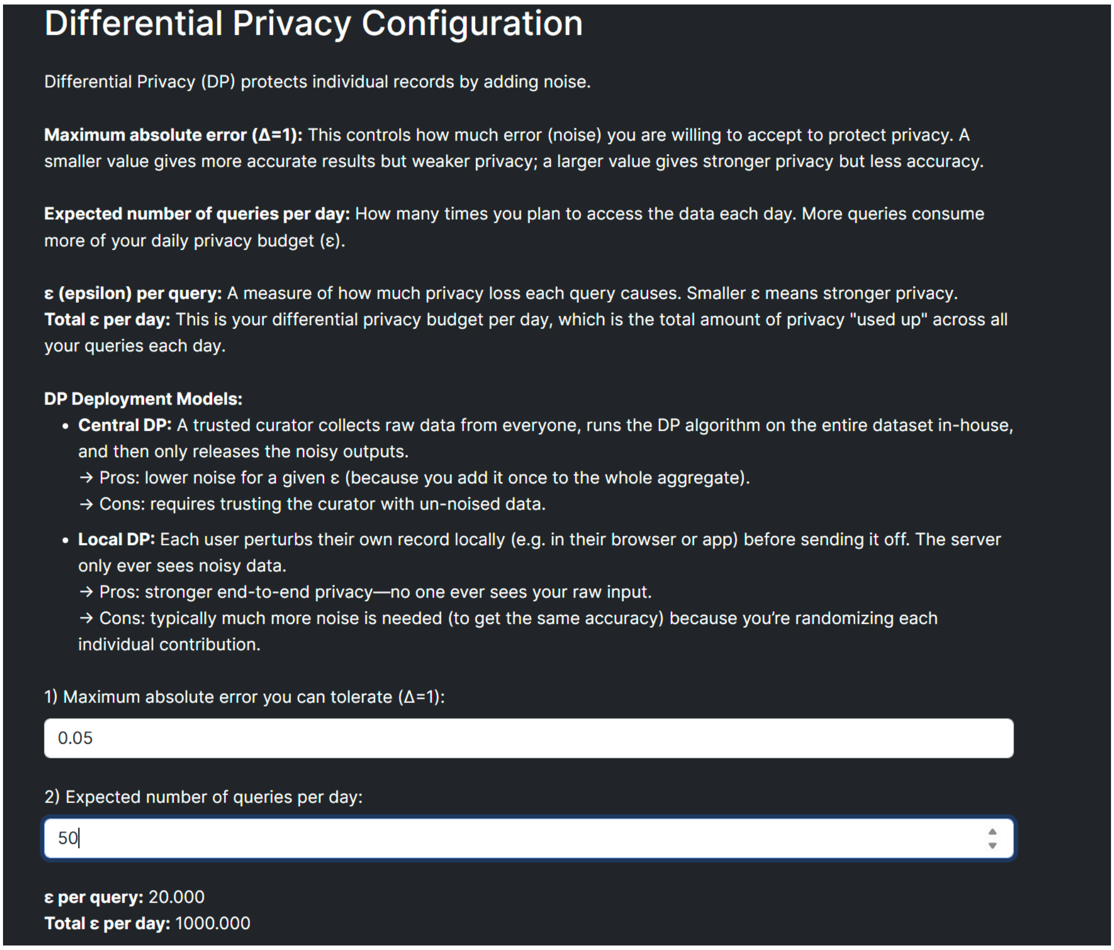
TEEs give you a hardware-backed secure enclave to process sensitive data safely, even in untrusted environments.

1) Approximately how many unique records will you process?

○ <100k
● 100k–1M
○ >1M

2) Do you have secure hardware enclaves available?

○ Intel SGX / AMD SEV
● AWS Nitro Enclaves
○ No



Differential Privacy Configuration

Differential Privacy (DP) protects individual records by adding noise.

Maximum absolute error (Δ=1): This controls how much error (noise) you are willing to accept to protect privacy. A smaller value gives more accurate results but weaker privacy; a larger value gives stronger privacy but less accuracy.

Expected number of queries per day: How many times you plan to access the data each day. More queries consume more of your daily privacy budget (ε).

ε (epsilon) per query: A measure of how much privacy loss each query causes. Smaller ε means stronger privacy.
Total ε per day: This is your differential privacy budget per day, which is the total amount of privacy "used up" across all your queries each day.

DP Deployment Models:

- **Central DP:** A trusted curator collects raw data from everyone, runs the DP algorithm on the entire dataset in-house, and then only releases the noisy outputs.
  → Pros: lower noise for a given ε (because you add it once to the whole aggregate).
  → Cons: requires trusting the curator with un-noised data.
- **Local DP:** Each user perturbs their own record locally (e.g. in their browser or app) before sending it off. The server only ever sees noisy data.
  → Pros: stronger end-to-end privacy—no one ever sees your raw input.
  → Cons: typically much more noise is needed (to get the same accuracy) because you're randomizing each individual contribution.

1) Maximum absolute error you can tolerate (Δ=1):

0.05

2) Expected number of queries per day:

50

ε per query: 20.000
Total ε per day: 1000.000

Justification of Recommendations:

Differential Privacy was recommended because it provides strong, mathematically backed privacy guarantees, significantly reducing the chance of individual re-identification. With an ε-value carefully chosen at around 20 per query, it balances stringent privacy needs with the accuracy required for medical analytics. TEEs, particularly AWS Nitro Enclaves, provide secure, isolated environments for real-time analytics, crucial for maintaining data security and rapid response times. These enclaves integrate smoothly with MediSecure's existing cloud infrastructure.

Technical, Policy, and Legal Challenges:

- **Technical**:
  - The guidance for DP assumes the delta value is 1, which might not be the same in the real time situation, so choosing 20 might not be enough. The guidance is only acting as a reference range.
  - Carefully managing the privacy budget (ε) daily without sacrificing accuracy or privacy.

    - Seamlessly integrating DP into real-time analytics processes.
    - Correctly configuring and optimizing performance of secure hardware enclaves.
- **Policy**:
    - Establishing clear, consistent policies to manage daily privacy budgets.
    - Ensuring comprehensive team training to handle complex privacy technologies effectively.
- **Legal**:
    - Ensuring strict compliance with HIPAA, GDPR, and CCPA regulations through diligent governance and regular audits.
    - Addressing challenges around international data transfers and ensuring proper legal safeguards are in place.

Practical Privacy Guarantees:

The recommended combination of Differential Privacy and TEEs provides robust privacy protections. Differential Privacy ensures mathematically proven limits on data exposure and re-identification risks. TEEs add another security layer, safeguarding data during computation from unauthorized access and external threats. Still, careful daily management of the privacy budget is essential to maintain strong privacy guarantees consistently.

In conclusion, MediSecure’s combination of Differential Privacy and TEEs, supported by comprehensive legal compliance, offers strong and practical privacy protections while fulfilling stringent accuracy and real-time operational demands.