Here's a simplified breakdown of optimizing a backend for high performance and scalability, particularly for IoT applications:

**1. Cache Server Implementation:**
Utilize a cache server (similar to Redis) to store frequently accessed data, reducing database load and enhancing response times.

**2. Queue:**
A Laravel queue is a feature that allows users to defer time-consuming tasks to the background. This speeds up web requests to the application.

**3. Selective Data Retrieval:**
Opt for selective data retrieval methods (e.g., SELECT statements) instead of fetching entire datasets, minimizing network overhead and processing time.

**4. Data Compression Techniques:**
Apply compression techniques such as gzip for transmitting IoT data, reducing bandwidth usage and accelerating data transfer between devices and the backend.

**5. Middleware Optimization:**
Employ middleware to intercept requests and optimize functions for specific tasks like data validation, authentication, or encryption without compromising system speed.

**6. Custom Middleware for Caching:**
- Develop custom middleware to manage caching strategies, intercepting requests to serve responses from the cache if available, hence decreasing backend load.

These strategies collectively aim to enhance backend performance and scalability in IoT applications by minimizing resource utilization, optimizing data handling, and efficiently managing incoming data streams.