**Section 3: Architectural Design Challenge:**

**Collect event data:**

1. SDK for events monitoring (Segment or Google Tag Manager) – using JavaScript code implemented in the Wix website to collect events.
2. Kafka as event stream – collects and stores the captured events and enables fast parallel processing.
3. Apache Spark - for real-time processing of data from Kafka and integrating it with additional data.

**Process and Enrich Data:**

1. Google Cloud Storage – for raw storage of events before transformation.
2. DWH based on BigQuery – where the processed data is stored for efficient analysis and reporting.
3. An ETL system like Airflow – for running and scheduling batch data processing workflows to enrich data from additional sources.

**Make Data Accessible:**

1. Using Looker / Tableau – BI platforms for data analysis and creating interactive dashboards.
2. DWH BigQuery – stores all the information and transformed data and allows to query the data in real time or to create ad hoc reports using SQL.
3. Role-Based Access Control – defining permissions based on roles to ensure secure access to data.

**Technical Specifications:**

**Technologies & Providers:**

1. Segment - Offers direct streaming to various sources and advanced privacy management capabilities.
2. Kafka - High resilience to large data volumes with low latency.
3. Google Cloud Storage- A cost-effective and scalable cloud storage solution for raw data storage.
4. Spark streaming - Live event processing with high scalability.
5. BigQuery - A cloud-based solution with high performance for data analytics.
6. Looker - Advanced visualization tools for report management and data analysis.

**Handling High Volume & Data Quality:**

1. Apache Kafka is used for real-time event streaming, enabling the processing of events as they occur.
2. Spark is a framework used for processing the streaming data in real-time, allowing for scalability and high throughput.

**Monitoring & Alerting Strategy:**

1. Using monitoring tools such as Monte Carlo and Grafana to monitor query latency and data loads, a high percentage of events rejected due to errors, and Resource usage higher than usual .