Based on the requirements outlined in the case study, here's a proposed solution for addressing the challenges faced by AdvertiseX:

**Data Ingestion:**

1. Ingestion System:

* Utilize Apache Kafka for real-time data streaming and Apache NiFi for batch processing.
* Configure Kafka topics for ad impressions (JSON), clicks/conversions (CSV), and bid requests (Avro).
* Implement NiFi dataflows for batch ingestion from various sources into appropriate Kafka topics.

2. Scalability:

* Use Kafka Connect for seamless integration with different data sources and scalability.
* Deploy Kafka clusters with appropriate partitioning to handle high data volumes.

**Data Processing:**

1. Data Transformation:

* Develop Apache Spark jobs to standardize and enrich data.
* Use Spark Structured Streaming for real-time processing of ad impressions, clicks, and conversions.
* Implement validation, filtering, and deduplication logic within Spark jobs.

2. Correlation:

* Utilize Spark SQL to join ad impressions with clicks/conversions based on user IDs and timestamps.
* Implement logic to handle discrepancies and missing data gracefully.

**Data Storage and Query Performance:**

1. Storage Solution:

* Choose Apache Hadoop Distributed File System (HDFS) or cloud-based storage like Amazon S3 for storing processed data.
* Utilize Apache Hive or Apache HBase for structured storage and efficient querying.

2. Optimization:

* Partition data in storage based on time or campaign ID for faster retrieval.
* Implement columnar storage formats like Parquet for better compression and query performance.

**Error Handling and Monitoring:**

1. Error Detection:

* Implement schema validation at ingestion points to detect anomalies in data formats.
* Use Apache Flink for complex event processing to identify discrepancies in real-time.

2. Alerting Mechanisms:

* Integrate monitoring tools like Prometheus and Grafana for real-time monitoring of data pipelines.
* Configure alerts for data quality issues using tools like Apache Airflow or custom scripts.

**Additional Considerations:**

* **Security**: Implement encryption mechanisms for data at rest and in transit.
* **Compliance**: Ensure compliance with data privacy regulations like GDPR and CCPA.
* **Resource Management**: Utilize container orchestration platforms like Kubernetes for resource management and scalability.
* **Documentation**: Maintain detailed documentation of data pipelines, schemas, and processing logic for future reference.

By implementing this solution, AdvertiseX can effectively ingest, process, store, and analyze vast amounts of data generated by ad impressions, clicks, and bid requests, while ensuring data quality, scalability, and performance.