**Assignment: Big Data Engineering Solutions Across Industries**

*Objective:* Leverage the capabilities of Hadoop, Hive, Hbase, Spark, Solr, Kafka, and NiFi to engineer a real-world data solution. This project will emphasize data ingestion, transformation, storage, and analytical querying.

*Duration:* 2 Weeks

**Assignment Breakdown:**

1. **Selection of Use Case:** Choose one of the listed use cases below. Each represents a data engineering challenge faced by a specific industry.
2. **Data Pipeline Design:** Design your data flow and pipeline, specifying the role of each component.
3. **Implementation:** Set up data ingestion, transformation, storage, and create SQL queries to derive analytical insights.
4. **Presentation Creation:** Compile a comprehensive PowerPoint detailing the problem, your solution’s architecture, the role of each component, and insights gained from your SQL analytics.
5. **Demo Recording:** Produce a video where you walk viewers through your data pipeline, explain design choices, and demonstrate your SQL analytics.
6. **Submission:** Upload both your PowerPoint presentation and your demo video.

**Use Cases:**

1. **E-Commerce - Product Inventory Management:**
   * **Problem:** Ingest sales data and manage product inventory in real-time, offering insights into stock levels, sales trends, and reorder points.
   * **Sample Data:** [E-Commerce Dataset](https://www.kaggle.com/carrie1/ecommerce-data)
2. **Finance - Transaction Log Analysis:**
   * **Problem:** Streamline large volumes of transaction logs, identifying transaction volumes, peak transaction times, and other important metrics.
   * **Sample Data:** [Bank Transactions Dataset](https://www.kaggle.com/jeanmidev/smart-meters-in-london)
3. **Social Media - User Interaction Analysis:**
   * **Problem:** Ingest and analyze user interactions on a social media platform, deriving insights about user engagement, most active times, and trending posts or topics.
   * **Sample Data:** [Social Media Interaction Dataset](https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings)
4. **Tourism - Hotel Booking Analysis:**
   * **Problem:** Process and analyze hotel booking data to identify trends, booking preferences, and cancellations.
   * **Sample Data:** [Hotel Booking Dataset](https://www.kaggle.com/jessemostipak/hotel-booking-demand)

**Submission Guidelines:**

1. **PowerPoint Presentation:**
   * Title Slide: Use Case Name, Your Name, Date
   * Problem Overview: Brief on the chosen industry challenge.
   * System Architecture: Flow diagram showcasing the pipeline and role of each component.
   * Data Ingestion & Transformation: Elaborate on data sources and transformation processes.
   * Storage Strategy: Justify storage choices, schemas, and structures.
   * Analytical Queries: Showcase SQL queries and their results.
   * Insights & Observations: Deductions from the analytical results.
   * Conclusion: Wrap-up and possible future improvements.
2. **Video Demo:**
   * Duration: 10-15 minutes.
   * Introduce the selected use case.
   * Delve into the designed data pipeline.
   * Walk through data ingestion, transformation, and storage.
   * Demonstrate analytical SQL queries and discuss results.
   * Conclude with insights and takeaways.
3. **Upload:** Ensure the files are correctly named and upload them to the designated platform.

**Evaluation Criteria:**

1. Depth of Understanding
2. Logical and Efficient Data Pipeline
3. Clarity in Role of Each Component
4. Analytical Capability with SQL
5. Presentation and Demonstration Quality

Best wishes! We’re eager to see your big data prowess in action!