**On-Platform Data Pipeline-**

**Content**: -

|  |  |
| --- | --- |
| **Serial No** | **Content Available** |
| 1. | Introduction |
| 2. | List Of On-Platform Data Sources Available |
| 3. | What is Data Ingestion Process |
| 4. | What is Data Processing Pipeline |
| 5. | Dataset with their Respective Table Description |

**Introduction: -**

An On-Platform data source providing social media data refers to an interface or system that enables the retrieval, collection, and analysis of data generated from social media platforms. This can include APIs, data streams, and third-party tools.

These data sources serve as critical components for building applications, conducting research, and deriving insights from social media interactions and trends.

Data is a two-part structure that ensures a clear separation of concerns between data ingestion and processing. Each part can be scaled and optimized independently, allowing for a robust data pipeline that can handle diverse data sources and processing needs. The two different Parts are DIP (Data Ingestion Process) and DPP (Data Processing Process.)

**List Of On-Platform Data Sources Available: -**

|  |  |
| --- | --- |
| **Serial No** | **On-Platform sources** |
| 1. | CMP/Cerebro |
| 2. | CAP |
| 3. | GAM |
| 4. | salesforce |
| 5. | Workfront |
| 6. | Criteo |
| 7. | Citrus |
| 8. | Campaign Aggregate Table |
| 9. | Rubix |
| 10. | Store Presentation/Signage |
| 11. | Instore/ServiceHub |

**What is Data Ingestion Process**: -

A data ingestion pipeline typically focuses on efficiently collecting, processing, and managing data, especially in the context of a cloud environment and Elasticsearch for search and analytics. Below is a structured outline of such a pipeline:

### **Data Ingestion Process Overview**

1. **Data Sources**: Identify the various data sources from which data will be ingested.
2. **Data Ingestion Tools**: Use tools to collect data and push it into a processing layer.
3. **Data Processing**: Transform and prepare the ingested data for storage and analysis.
4. **Data Storage**: Store the processed data in a suitable format, often in Elasticsearch.
5. **Monitoring and Management**: Utilize tools like Cerebro for managing Elasticsearch and monitoring the pipeline.

**What is a Data Processing Pipeline-**

Creating a data processing pipeline involves integrating various components to handle data ingestion, processing, storage, and analysis. Below is a high-level overview of how such a pipeline can be structured, assuming you’re referring to a typical setup for data processing.

### **Data Processing Pipeline Overview**

1. **Data Ingestion**: Collect data from various sources.
2. **Data Processing**: Transform and analyze the ingested data.
3. **Data Storage**: Store processed data in a suitable format.
4. **Data Visualization/Analysis**: Analyze or visualize the data for insights.