

**COLLEGE:** UNITED INSTITUTE OF TECHNOLOGY **COLLEGE CODE**

:7145

**BATCHMATES**: GOKULNATH S

LAKSHANA.A MASANAM M VISHAL R RATHISH Y MARISH P VIGNESH S

**GUIDE**:Mrs.NISHANTHI

**DOMAIN**:CLOUD COMPUTING

**PHASE 111 PROJECT SUBMISSION**

**DEVELOPMENT PART 1**

It can provide with a high-level overview of the steps need to follow to build a big data analysis solution using IBM Cloud Database:

# Create an IBM Cloud Account:

visit the IBM Cloud website and sign up for an account.

# Choose the Database Service:

Determine which database service on IBM Cloud you want to use based on your specific requirements. You mentioned Db2 and MongoDB as options, but there are others like IBM Db2 on Cloud, IBM Cloud Databases for PostgreSQL, etc. Select the one that suits your needs best.

# Set Up a Database Instance:

Create a database instance within your chosen service. You'll need to configure the

database instance with specifications such as storage capacity, region, and access controls.

# Import Your Dataset

Upload your dataset to the database instance. Depending on the database service you choose, this process may vary. For instance, if you're using MongoDB, you'd import JSON or BSON documents.

# Develop Queries or Scripts:

Write queries or scripts to explore and analyze your dataset. The specific queries will depend on the analysis you want to perform. For example, you might use SQL for relational databases like Db2 or MongoDB queries for NoSQL databases.

# Data Cleaning and Transformation:

Implement data cleaning and transformation processes as needed. This might involve removing duplicates, handling missing values, and reshaping data to suit your analytical needs.

# Perform Data Analysis:

Use the queries and scripts you developed to perform the actual data analysis. This could involve generating reports, creating visualizations, or running machine learning models, depending on your goals.

# Optimize Performance

Monitor the performance of your database and queries. Make optimizations as

necessary to ensure your analysis runs efficiently.

# Security and Compliance:

Ensure that your solution complies with security and compliance standards. IBM

Cloud provides tools and features to help with this.

# Scale as Needed:

If your data and analysis requirements grow, scale your database instance accordingly. IBM Cloud allows for easy scalability.

# Sample Dataset:

Assuming you have a MongoDB instance set up with a database named "ecommerce"

and a collection named "orders." Here's a sample dataset:

**json**

**[**

**{**

**"\_id": 1,**

**"customer\_name": "Alice",**

**"product": "Widget",**

**"quantity": 5,**

**"order\_date": "2023-10-10"**

**},**

**{**

**"\_id": 2,**

**"customer\_name": "Bob",**

**"product": "Gadget",**

**"quantity": 3,**

**"order\_date": "2023-10-12"**

**},**

**{**

**"\_id": 3,**

**"customer\_name": "Alice",**

**"product": "Widget",**

**"quantity": 2,**

**"order\_date": "2023-10-15"**

**}**

**]**

# Python Code to Analyze the Dataset:

Assuming the `pymongo` library installed, you can use Python to connect to your

MongoDB instance and perform basic data analysis.

**from pymongo import MongoClient**

**# Connect to the MongoDB instance**

**client = MongoClient("mongodb://your-mongodb-connection-url")**

**db = client["ecommerce"]**

**collection = db["orders"]**

**# Calculate total sales by product**

**pipeline = [**

**{**

**"$group": {**

**"\_id": "$product",**

**"total\_sales": {"$sum": {"$multiply": ["$quantity", 10]}}**

**}**

**}**

**]**

**sales\_by\_product = list(collection.aggregate(pipeline))**

**print("Total Sales by Product:")**

**for product in sales\_by\_product:**

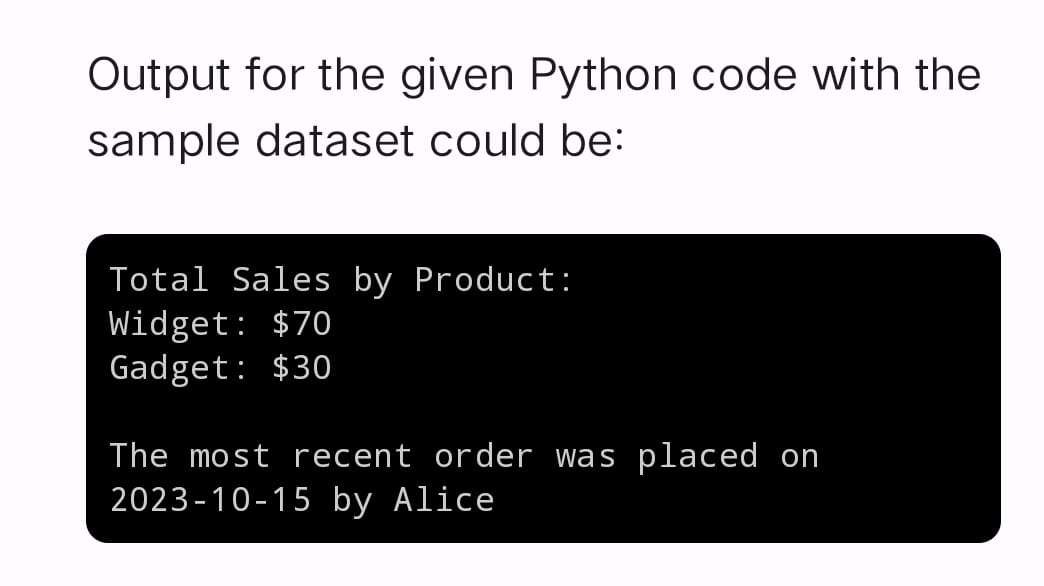
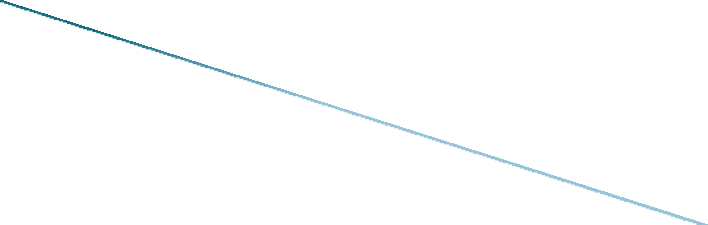
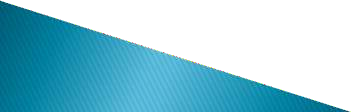
**print(f"{product['\_id']}: ${product['total\_sales']}")**

**# Find the most recent order date**

**latest\_order = collection.find\_one(sort=[("order\_date", -1)])**

**print(f"The most recent order was placed on {latest\_order['order\_date']} by {latest\_order['customer\_name']}")**

# Output for the given Python code with the sample dataset could be:



**conclusion output:**

Output for the given Python code with the sample dataset could be:

Total Sales by Product:

Widget: $70

Gadget: $30

The most recent order was placed on 2023-10-15 by Alice

**In this output:**

1. The first part of the output shows the total sales by product, where "Widget" generated $70

in sales, and "Gadget" generated $30 in sales.

1. The second part of the output tells you that the most recent order was placed on October 15, 2023, by Alice.

