Github link to task 1 - https://github.com/nitinsoni1/AegisCovenant\_Assignment\_Internshala/blob/main/AegisCovenant\_Assignment\_1\_Internshala.ipynb

Here's a diagram demonstrating my proposed approach for managing and processing data from the three sources - MySQL database, JSON files stored in an S3 bucket, and a REST API:

+-------------+

| MySQL |

| Database |

+-------------+

|

|

+-----------------+

| Data Storage |

| and Syncing |

| (AWS S3 Bucket) |

+-----------------+

|

|

+-----------------------------------------------------+

| Data Processing |

| (Python + AWS Lambda + AWS Glue) |

| |

| |

+------------+------------+ +------------------+-------------+

| | | |

+-------+----------+ +--------------+---------+ +------+---------+ +--------+--------+

| JSON data files in | | REST API Endpoint | | AWS Lambda to | | MySQL Database|

| S3 bucket (Amazon | | (Amazon API Gateway)| | process and | | (on-premises |

| S3 service) | | for fetching data | | transform data| | or cloud) |

+--------------------+ +---------------------+ +--------------------+ + ------------------+

The proposed approach involves the following tools and services:

* **MySQL database**: This will be the source of data that will be used in the analysis dashboard. It can be hosted on-premises or in the cloud, depending on the organization's requirements.
* **AWS S3 Bucket**: This will be used to store JSON data files that are produced by various sources. The data can be ingested into the S3 bucket using AWS Lambda functions, AWS Glue, or other data integration services.
* **AWS Lambda**: This is a serverless computing service provided by AWS that allows for event-driven computing. Lambda functions can be used to perform data processing and transformation on data ingested from various sources, including JSON files and REST APIs.
* **AWS Glue**: This is a fully managed ETL (extract, transform, load) service provided by AWS that allows for automated data processing and transformation. It can be used to create ETL jobs that can read data from various sources, including JSON files and databases, and store the transformed data in the desired format in a data store of choice.
* **Python**: This will be used as the primary programming language for data processing and transformation. It is widely used in the data science community and has a rich set of libraries and tools that make it easy to work with various data sources and formats.
* **REST API Endpoint**: This is an endpoint exposed by the REST API that can be used to fetch data. It can be secured using authentication and authorization mechanisms provided by the API Gateway.
* **API Gateway**: This is a fully managed service provided by AWS that allows for the creation, management, and deployment of REST APIs. It can be used to expose the REST API endpoint securely and provide features such as authentication, rate limiting, and caching.

Overall, this approach provides a scalable, secure, and cost-effective way of managing and processing data from various sources for the analysis dashboard.