**Major Project Proposal**

**Team Id: 25\_CS\_AI\_4A\_09**

**Team Details:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S No** | **Full Name** | **Roll No** | **Branch &**  **Section** | **Mob No** |
| 1 | Ayush Verma | 2101641520040 | CS-AI-3A | 6393868175 |
| 2 | Aakarshit Srivastava | 2101641520001 | CS-AI-3A | 7275011914 |
| 3 | Ankita Singh | 2101641520021 | CS-AI-3A | 8115520277 |
| 4 | Bhaskar Banerjee | 2101641520042 | CS-AI-3A | 9696781896 |
| 5 | Diksha Tiwari | 2101641520049 | CS-AI-3A | 9555718386 |

**Project Title:**

|  |
| --- |
| ***DataFlowX: Cloud Migration Express*** |

**Domain: (Select all relevant Options)**

|  |  |
| --- | --- |
| **1.** Software-Web Application | 2. Software-Mobile Application |
| 3. Artificial Intelligence/Machine  Learning/Deep Learning | 4. Computer Vision/Image  Processing |
| 5. Blockchain | 6. Internet of Things |
| 7. Natural Language Processing | **8. \*Big Data / Cloud Computing** |

**Problem Statement:**

Design and implement a data analytics pipeline for a platform using Cloud Data Migration Services. The pipeline should seamlessly migrate transactional data from the existing database/resources to Amazon Redshift or Google Cloud for analytics purposes, perform necessary transformations and aggregations, and enable real-time monitoring and analysis of sales trends, customer behaviour, and inventory management. The solution should ensure scalability, reliability, and cost-effectiveness while adhering to security, compliance requirements and including factors like:

1. Massive Database delay
2. Categorization and Consolidation
3. Adaptive Scheduling
4. Process Allocation and Management
5. Workflow Execution and Observation

**Proposed Solution:**

Developing an automated data processing pipeline for categorizing and analyzing database generated by multiple services within an enterprise environment. The pipeline should include scheduling mechanisms for efficient data ingestion, transformation, and migration of services into a centralized storage solution, such as Amazon S3. Additionally, the pipeline should facilitate the seamless migration of services to new infrastructure while ensuring minimal downtime and data integrity. The solution should optimize resource utilization, enable real-time monitoring of pipeline activities, and support scalability to accommodate future growth and evolving business requirements.

**#.** Optimize and index a massive database, including historical records, to improve query performance, reduce latency, and enhance overall efficiency.

**#.** Creating a real-time data categorization system that prioritizes incoming data streams based on predefined criteria. The system dynamically adjusts categorization rules to reflect changing priorities, facilitating efficient resource allocation and enabling rapid response to critical data events for enhanced operational efficiency in real-time.

**#.** Implement an intelligent scheduling system using machine learning and optimization algorithms to automate and optimize service scheduling processes. By analyzing historical data and business objectives, the system generates optimal schedules to maximize efficiency and minimize costs. Continuous learning ensures adaptation to changing patterns, enhancing operational efficiency and customer satisfaction.

**#.** Seamlessly transition services from legacy or on-premises infrastructure to cloud-based environments, ensuring minimal disruption, data integrity, and security. Leveraging cloud benefits such as scalability and agility to optimize service delivery and operational efficiency.

**Unique/Distinctive feature of the solution:**

|  |
| --- |
| The distinctive feature of the solution lies in its comprehensive approach to migration, encompassing not only the technical aspects but also the strategic and operational considerations. Key aspects that make this solution stand out include:   * **Holistic Assessment**: Conducting a thorough assessment of the existing infrastructure, including applications, dependencies, and data, to ensure a complete understanding of the migration scope and potential challenges. * **Risk Mitigation Strategies**: Implementing robust risk mitigation strategies to address potential issues such as data loss, downtime, and compatibility issues during the migration process. * **Automated Migration Tools**: Utilizing automation tools and scripts to streamline migration tasks, reduce manual effort, and minimize the likelihood of errors. * **Scalable Architecture**: Designing a scalable architecture that can accommodate future growth and evolving business requirements, ensuring that the migrated services remain agile and adaptable. * **Continuous Monitoring and Optimization**: Implementing mechanisms for continuous monitoring of migrated services post-migration, allowing for timely detection of issues and optimization of performance and resource utilization. * **User Training and Support**: Providing comprehensive training and support to users and stakeholders to facilitate a smooth transition to the new environment and maximize the benefits of the migrated services. * **Documentation and Knowledge Transfer**: Documenting the migration process and knowledge transfer to internal teams, enabling them to maintain and support the migrated services effectively in the long run.   By incorporating these distinctive features, the solution ensures a successful and sustainable migration of services to the cloud, delivering tangible benefits in terms of cost savings, scalability, and operational efficiency. |

**Tools/Technology Uses:**

**Hardware requirements:**

1. **Local Development Machine**:
   * Processor: Multi-core processor for running development tools and IDEs efficiently.
   * Memory (RAM): At least 8GB RAM for smooth operation of development environments and virtualization.
   * Storage: Solid-state drive (SSD) for fast read/write speeds, at least 256GB for storing project files and development environments.
2. **Networking**:
   * High-speed internet connection for accessing cloud services and resources.
   * Ethernet or Wi-Fi adapter for network connectivity.
3. **Testing Environment**:
   * Depending on the project's requirements, additional hardware may be needed for setting up testing environments, such as virtualization servers or physical servers for testing deployments.
4. **Backup and Storage**:
   * External storage devices or network-attached storage (NAS) for backups and storing local project files.
5. **Security**:
   * Firewall/router for network security.
   * Antivirus software for endpoint protection.
6. **Miscellaneous**:
   * Monitor, keyboard, and mouse for local machine setup.
   * Power backup solutions (e.g., uninterruptible power supply or UPS) to prevent data loss during power outages.

**Software requirements:**

1. **Operating System**:
   * Depending on the preferences of the development and operations teams, the project may require an operating system such as Linux (e.g., Ubuntu, CentOS) or Windows Server.
2. **Containerization and Orchestration**:
   * Docker Engine: Software for creating, managing, and running containers.
   * Kubernetes: Container orchestration platform for automating deployment, scaling, and management of containerized applications.
3. **Services**:
   * AWS ,AZURE ,Google Cloud platform and Machine Learning Algorithms using Multiple interfaces.

(To be Filled by Faculty/Evaluator)

**Proposal Evaluation:**

1. Right Identification of the Problem (Appropriate selection of the problem)?
   1. Excellent b) Good c) Needs Improvement d) Unacceptable

1. Relevance of the Solution (Adequately addressing the problem/need)?
   1. Excellent b) Good c) Needs Improvement d) Unacceptable

1. Innovativeness in the Solution (Distinctive innovative components/features of the solution)?
   1. Excellent b) Good c) Needs Improvement d) Unacceptable

1. Uniqueness of the Solution (Intellectual Property Component)?
   1. Excellent b) Good c) Needs Improvement d) Unacceptable

**Improvements/ Suggestions by the Evaluator:**

1.

2.

3.

4.

**Name of Faculty:**

**Designation:**

**Signature with Date:**

Guidelines:

* + One Proposal per team will be submitted by the team leader only.
  + A Team can have maximum 5 Members.
  + Upload the document in .doc or .pdf format with font size 12, single spacing, Times New Roman font only.