**Requirements Document for Database Backup and Archival Solution**

**1. Introduction**

**1.1 Purpose**

This document outlines the functional and non-functional requirements for a comprehensive database backup and archival solution. The system is intended to address challenges in managing, backing up, and archiving large datasets (minimum 200 GB) while ensuring data integrity, compliance with retention policies, and minimal impact on database performance.

**1.2 Scope**

The solution will include backup, archival, and restore functionalities with a web-based user interface for managing operations. It will support major relational and non-relational databases, on-premises, and cloud storage, ensuring scalability and platform independence.

**2. Functional Requirements**

**2.1 User Interface**

* Provide an intuitive web-based UI for initiating, scheduling, and managing backup and archival operations.
* Support dashboard views for:
  + Task progress monitoring.
  + Performance metrics.
  + Error notifications.
* Enable both one-time and recurring schedules with customizable frequency (e.g., daily, weekly, monthly).
* Allow on-demand backups.

**2.2 Backup Features**

* Support full, incremental, and differential backups.
* Provide scheduling options for backups at predefined intervals.
* Ensure data integrity checks post-backup.

**2.3 Archival Features**

* Support long-term data archival based on retention policies.
* Provide efficient compression to optimize storage usage.

**2.4 Restore Features**

* Enable partial or full data restoration.
* Validate the integrity of restored data.

**2.5 User Management**

* Role-based access control with the following roles:
  + **Admin:** Full access to system settings, scheduling, and monitoring.
  + **View-Only:** Restricted to viewing status, progress metrics, and logs.
* Notification and alert management:
  + **Admin:** Alerts for task completion, failure, and errors.
  + **View-Only:** Notifications for monitoring purposes.

**2.6 Monitoring and Logging**

* Provide detailed logs for all operations.
* Offer real-time progress monitoring and performance metrics.
* Support integration with external monitoring and alerting systems.

**3. Non-Functional Requirements**

**3.1 Scalability and Performance**

* Handle datasets of at least 200 GB efficiently, with scalability for larger datasets.
* Minimize resource usage (CPU, memory, and network) during operations.
* Ensure acceptable backup and restoration times to reduce downtime.

**3.2 Compatibility**

* Support major relational databases (e.g., PostgreSQL, SQL Server) and optional non-relational databases.
* Ensure compatibility with on-premises and cloud storage solutions (e.g., AWS S3, Azure Blob Storage).
* Operate seamlessly on Windows and Linux environments.

**3.3 Reliability and Testing**

* Include robust error handling and logging mechanisms.
* Simulate scenarios for:
  + Dataset handling exceeding 200 GB.
  + Backup failures and error recovery.
  + Restoration integrity verification.
* Automated testing scripts for validating functionality and performance.

**3.4 Security**

* Implement secure authentication and role-based access control.
* Ensure data encryption during storage and transit.

**3.5 Usability**

* Ensure the user interface is intuitive and accessible.
* Provide comprehensive documentation, including user guides and configuration manuals.

**4. Deliverables**

1. **Software Solution:** A reliable tool for database backup and archival, including customizable retention policies and efficient compression.
2. **User Interface:** A web-based dashboard for scheduling, monitoring, and managing tasks.
3. **Documentation:**
   * User guides and technical specifications.
   * Configuration and troubleshooting manuals.
4. **Automated Testing Scripts:** Scripts for validating backup, archival, and restoration processes.
5. **Performance Benchmarks:** Metrics for backup/restore times, resource usage, and scalability.
6. **Monitoring Integration:** Compatibility with existing monitoring systems.

**5. Constraints**

* Ensure efficient handling of large datasets (minimum 200 GB) with scalability.
* Balance performance with storage cost through optimized compression techniques.
* Maintain acceptable resource usage and minimize impact on database performance.
* Ensure cross-platform compatibility for Windows and Linux environments.
* Adhere to operational downtime thresholds during backup and restoration.

**6. Assumptions**

* The organization has sufficient infrastructure to support storage and processing requirements.
* Users are trained in accessing and using the system effectively.
* Compliance with local and organizational data retention and security policies is ensured.