ODS 系統開發規格書 (SRS)

1. 資料整合與 ETL 規格 (核心功能)

- **多源異質整合**：須支援從多種來源（如 Informix, SQL Server, 檔案如：EXCEL 檔或 CSV 檔,或 API 串接）擷取資料。
- **同步頻率**：須支援**近即時 (Near Real-time)** 同步或高頻率批次處理（例如每 5-15 分鐘一次），且不得影響來源系統效能。
- **異動資料擷取：系統應支援以增量方式同步資料，僅針對有變動的資料進行更新，而非全量覆蓋。**
- **清洗規則 (Data Cleaning)**：定義明確的轉換邏輯，包含異常值剔除、日期格式統一（如 YYYY-MM-DD）及空值處理。

2. 資料模型與儲存規格

- **原始性與細粒度**：ODS 須保留與來源系統一致的**原子資料 (Atomic Data)**，不進行高度彙總，以供後續溯源。
- **生命週期管理**：
  - **快照儲存**：應支援儲存特定時間點的營運快照。
  - **清理機制**：須具備過期資料自動清理或封存邏輯（例如僅保留 1~3 個月的營運明細）。
- **擴展性**：採用分散式架構或高擴展性儲存，確保未來資料量增長時可平滑擴容。
- **向量資料庫支援：**須支援向量資料庫之儲存與查詢能力，以利進行相似度檢索等應用需求，並可與既有資料模型整合。

3. 系統效能與穩定性

- **高併發處理**：須支援多個下游客戶端（如 Tableau、Dashboard）同時查詢，且查詢平均響應時間需在 X 秒內（依實際需求設定）。
- **斷點續傳**：當網路或 ETL 過程中斷時，系統必須能從最後成功的紀錄點自動恢復。
- **監控與告警**：須提供圖形化監控介面，顯示資料同步狀態、延遲時間，並在作業失敗時透過 Email 發送告警。

4. 安全與權限規範

- **資料遮罩 (Masking)**：針對個資欄位（如身分證號、手機號碼）須具備自動遮罩或去識別化功能。
- **稽核紀錄 (Audit Log)**：完整記錄所有資料擷取、存取及異動的操作人員與時間。

- **存取控制**：支援基於角色 (RBAC) 的欄位級 (Column-level) 存取權限控管。

5. 驗收交付物清單

1. **資料字典 (Data Dictionary)**：包含 ODS 所有資料表名稱、欄位屬性、來源系統對應關係 (Mapping)。
2. **ER 圖 (ER Diagram)**：完整的實體關係圖。
3. **ETL 邏輯說明書**：載明資料轉換、清洗與異常處理的演算法。
4. **效能測試報告**：包含壓力測試與資料同步延遲測試結果。