DATA WAREHOUSE SCHEMA DESIGN JUSTIFICATION

University Library Analytics Project

1. ARCHITECTURAL APPROACH SELECTION

Chosen Approach: Kimball (Bottom-Up) Methodology

Reasons:

1. **Project Timeline Constraint:** 2.5-week deadline favors rapid deployment
2. **Focused Business Needs:** Specific library analytics requirements (not enterprise-wide)
3. **Departmental Scope:** Single subject area (library usage) aligns with Kimball's data mart focus
4. **Implementation Speed:** Quicker ROI with immediate departmental reporting

Alternative Considered: Inmon (Top-Down)

- Rejected due to longer implementation time
- Overkill for single-department analytics needs
- Higher complexity without proportional benefit

2. DIMENSIONAL MODEL SELECTION

Chosen Model: Star Schema

Reasons:

1. **Query Performance:** Fewer joins than Snowflake schema
2. **Power BI Compatibility:** Star schema is optimal for Power BI relationships
3. **Simplicity:** Easier for end-users to understand
4. **Maintenance:** Simpler ETL processes
5. **Project Requirements:** Supports all required OLAP operations (drill-down, roll-up, slicing, dicing)

Alternative Considered: Snowflake Schema

- Rejected due to unnecessary normalization
- Would split dimensions (e.g., date hierarchy) without performance benefit
- Adds complexity for minimal storage savings

3. FACT TABLE DESIGN

Chosen Design: Consolidated Fact Table

Table: fact_library_usage

Reasons:

1. **Unified Metrics:** Single source for cross-service analysis
2. **Shared Dimensions:** All library services (books, digital, rooms) use same dimensions
3. **Simplified Reporting:** One query can analyze all service types
4. **Conformed Dimensions:** Ensures consistency across metrics

Measures Included:

- Count metrics: loan_count, download_count, booking_count
- Duration metrics: loan_duration_days, download_duration_minutes, booking_duration_hours

Alternative Considered: Galaxy Schema (Multiple Fact Tables)

- Rejected as overly complex for current requirements
- Could be implemented later if reporting needs diverge significantly

4. DATE DIMENSION STRATEGY

Chosen: Single Date Dimension Table

Reasons:

1. **Centralized Time Intelligence:** One source for all time-based calculations
2. **Hierarchy Support:** Built-in day→month→quarter→year hierarchy
3. **Holiday/Weekend Flagging:** Pre-calculated business logic
4. **ISO Standardization:** Resolves source system date format inconsistencies

5. SURROGATE KEY STRATEGY

Chosen: Auto-increment Integers for Dimensions

Reasons:

1. **Performance:** Smaller than natural keys (VARCHAR)
2. **Stability:** Immune to source system ID changes
3. **Integration:** Facilitates slowly changing dimensions (Type 2 if needed)
4. **Consistency:** Uniform key structure across all dimensions

6. STAGING AREA DESIGN

Chosen: Three Separate Staging Tables

Reasons:

1. **Source Isolation:** Each source system maintains its original structure
2. **Error Containment:** Data quality issues contained in staging
3. **Audit Trail:** load_timestamp tracks data ingestion
4. **Incremental Loading:** Supports delta processing

7. SCALABILITY CONSIDERATIONS

Future-Proofing Decisions:

1. **Index Strategy:** Appropriate indexes on foreign keys and frequently filtered columns
2. **Partitioning Ready:** Date-based partitioning possible on fact table
3. **Extension Points:** Schema can accommodate new library services
4. **Performance Optimization:** Indexed views can be added for common queries

8. ALIGNMENT WITH BUSINESS REQUIREMENTS

Business Requirement	Schema Feature	How It's Addressed
Monthly trends comparison	Date dimension hierarchy	Easy aggregation by month
Department analysis	Standardized department in dim_student	Consistent grouping
Peak usage times	Time slot dimension	Time-based analysis
OLAP operations	Star schema design	Native support for drill-down/roll-up
Real-time dashboards	Optimized star schema	Fast query performance

CONCLUSION

The chosen Kimball star schema with consolidated fact table provides:

- **Rapid implementation** within 2.5-week timeline
- **Optimal performance** for Power BI dashboards

- **Flexibility** for future enhancements
- **Alignment** with all specified business requirements
- **Maintainability** through clear, documented design