

Student name: Armando forgione

Student ID number: 2218650

Programme: BSc Applied Computing

Module: Data Warehousing & Data Mining

Module code: ACCA6026

Recovery Radio Ltd – Data Warehouse Report

Task 1

1. Difference Between OLTP Database and Data Warehouse

OLTP (Online Transaction Processing) systems are used for everyday operations. At Recovery Radio, these include the playlist database, advert sales, and request logs. They are designed to handle fast and frequent updates like inserting track plays or advert bookings. OLTP databases are usually normalized (split into many related tables) to avoid repetition and keep data consistent.

On the other hand, a Data Warehouse is used for analysis, not daily operations. It gathers data from different sources, transforms it, and stores it in a way that helps management make better decisions. For example, Recovery Radio can use a warehouse to see advertising trends, music popularity, or donations over time.

Data warehouses are usually denormalized and use dimensional models like a star schema. This structure is better for fast reporting and dashboards, even if it repeats some data.

While OLTP is great for recording what's happening now, a data warehouse is better for seeing the big picture and planning.

According to Inmon [1], OLTP systems are transaction-focused and support short, frequent updates, while a Data Warehouse is subject-oriented, integrated, time-variant, and non-volatile.

Kimball [2] also emphasises that a Data Warehouse should be denormalised to improve analytical performance.

For Recovery Radio, this means the DW must store historical trends such as advert revenue over months, popular tracks over time, and donation patterns.

. Data Warehouse Architecture and Source Systems

A data warehouse has multiple layers:

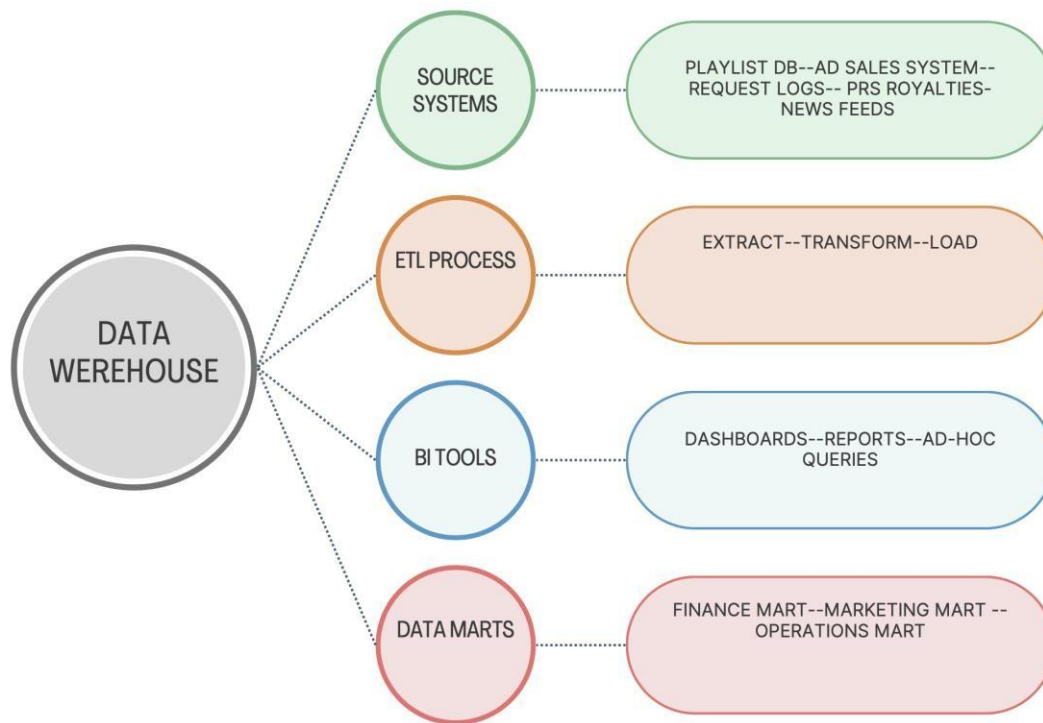
- **Source Systems:** OLTP playlist database, advert booking system, request logs, donation system, PRS royalty data, and news bulletins.
- **ETL Layer (Extract, Transform, Load):** Pulls data from sources, cleans and formats it, then loads it into the warehouse.
- **Data Warehouse Storage:** Main storage using dimensional models (star schema).
- **Data Marts:** Smaller parts of the warehouse for different teams.
- **Front-End Tools:** Dashboards, reports, and charts used by directors.

This setup lets Recovery Radio collect and analyze all data in one place, without affecting daily broadcasting.

A staging area is required to clean, validate, deduplicate, and standardise all incoming data before loading it into the DW. This follows IBM's recommended architecture for enterprise data warehousing [3].

The overall flow becomes:

Source Systems → Staging → Data Warehouse → Data Marts → Reporting Layer



This setup lets Recovery Radio collect and analyze all data in one place, without affecting daily broadcasting.

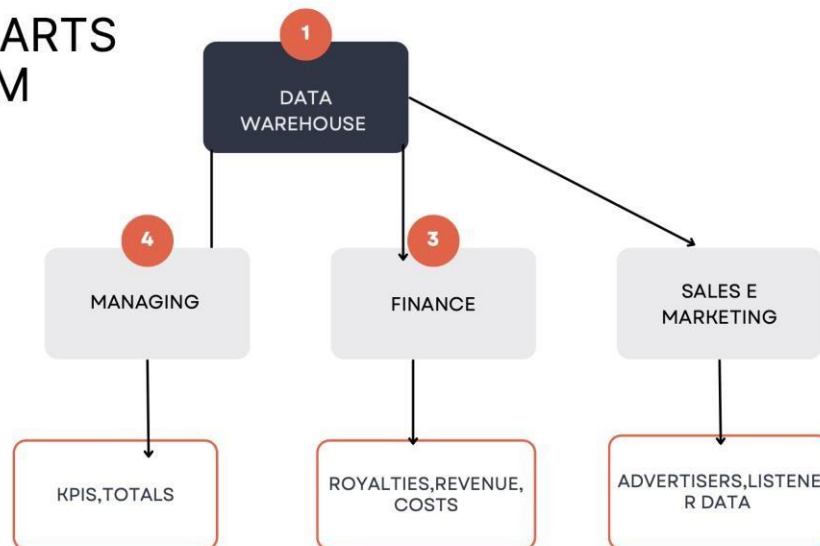
3. Use of Data Marts for Management Reporting

Data Marts are parts of the data warehouse designed for specific managers:

- **Managing Director Mart:** Revenue, costs, donations, and royalties. Useful for monitoring overall business performance.
- **Finance Director Mart:** Royalty payments, advert income, expenses. Helps with budgeting and financial planning.
- **Sales & Marketing Director Mart:** Listener requests, advert timing, campaign response. Helps improve marketing and adverts.

DATA MARTS DIAGRAM

Flowchart



HIRE

Each director sees only the data they need, making decision-making faster and easier.

Each Data Mart uses role-based access control (RBAC), ensuring that each director only views metrics relevant to their responsibilities. This improves security and performance [4].

4. Normalized vs Dimensional Data Structures

OLTP systems use a **normalized structure**, splitting data into many small tables. This is great for fast updates and keeping everything accurate. For example, Recovery Radio's playlist system might have separate tables for tracks, artists, and programmes.

A **data warehouse** uses a **star schema**. This means there's one central **Fact Table** (with numbers like play counts or advert revenue) and multiple **Dimension Tables** (with details like artist, programme, time).

🔗 Example:

- **Fact Table:** PlaylistFact
- **Dimensions:** DimTrack, DimArtist, DimProgramme, DimHospital, DimTime

Task 2

1. Internal and External Data Sources

Internal Sources:

- Playlist database (tracks, timestamps)
- Advert sales system (bookings, clients)
- Request logs (SMS, email, web)
- Programme schedules (types, presenters)
- Finance system (donations, costs)

External Sources:

- PRS (royalty rates)
- News bulletin providers
- Third-party music data (artist info, genres)

These all go through ETL to enter the data warehouse.

2. Star Schema Design

Fact Tables:

- **PlaylistFact:** Track plays, timestamp, ISRC, programme ID, hospital ID, duration, royalty cost
- **AdvertisingFact:** Advert ID, time slot, programme ID, client ID, duration, revenue

Dimension Tables:

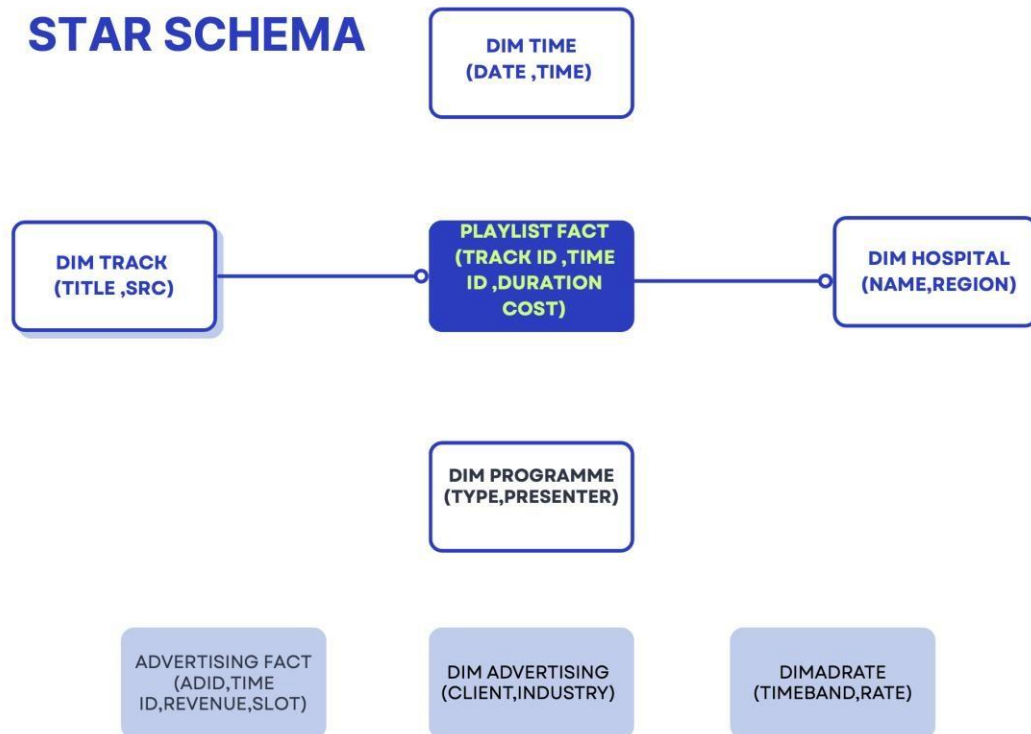
- DimTrack: TrackID, ISRC, Name, ArtistID, Genre, Year
- DimArtist: ArtistID, Name, Country, ActiveYears
- DimProgramme: ProgrammeID, Type, Presenter, Duration
- DimHospital: HospitalID, Name, Region, Frequency
- DimTime: Date, Month, Year, Day, Time of Day
- DimAdvertiser: Name, Industry, Direct/Agency
- DimPRSRate: TimeBand, Rate
- DimAdRate: TimeBand, Rate

All dimension tables use surrogate keys instead of natural keys. This allows the warehouse to manage Slowly Changing Dimensions (SCD).

Recovery Radio requires **SCD Type 2** for Artist, Track, Programme, and Advertiser dimension changes to preserve historical reports [2].

-

STAR SCHEMA



. Data Marts and Schema Extraction

Each department gets its own data mart:

- *Finance Mart: Playlist + advert facts, royalty rates*
- *Marketing Mart: Adverts + request data*
- *Operations Mart: Programmes + hospitals + playlists*

Each mart includes only the needed tables, making them fast and secure.

4. Simplified ETL Plan

- *Extract: Nightly pulls from playlists, advert sales, PRS, and requests*
- *Transform: Fix timestamps, format codes, calculate royalties and ad revenue*
- *Load: Load cleaned data into the warehouse, update tables, keep performance optimized*

Runs during off-hours to avoid problems with the live radio systems.

Detailed ETL Steps:

- Extract using timestamp watermarking to avoid full loads.
- Transform data by cleaning, deduplicating, standardising formats, and generating surrogate keys.
- Load dimensions first, then facts, ensuring referential integrity.
- Implement monitoring checks (row counts, invalid formats, broken links, duplicates).
- Store ETL logs for auditing and error recovery.

5. Data Quality and Governance *Key*

challenges:

- *Missing or duplicate request logs*
- *Wrong timestamps or IDs*
- *Royalty rate updates*

Solutions:

- *Validation checks*
- *Regular audits*
- *Documentation of data flows*
- *User access controls*

References

- *Inmon, W. H. (2005). Building the Data Warehouse*
- *Kimball, R., & Ross, M. (2013). The Data Warehouse Toolkit* • *Golfarelli, M., & Rizzi, S. (2009). Data Warehouse Design*
- *IBM. (n.d.). What is ETL?*
- *Microsoft. (n.d.). Data Warehousing Concepts*
- *Oracle. (n.d.). Dimensional Modeling*