CHINOOK DATABASE DOCUMENTATION

==========

OVERVIEW

The Chinook database is a sample database that represents a digital media store. It was created as an alternative to the Northwind database and contains real data from an iTunes library combined with fictional customer and employee data.

PURPOSE & USE CASES

- Learning database concepts and design patterns
- Practicing data analysis and visualization
- Building RAG (Retrieval-Augmented Generation) systems
- Testing e-commerce and inventory applications
- Developing APIs with realistic data
- Creating business intelligence reports

DATABASE STATISTICS

- 11 Core Tables
- Data spanning 2009-2013
- Multiple relationships and foreign keys
- Covers music, video, and digital media

==========

DATABASE TABLES

1. ARTISTS

Description: Stores information about music artists and bands.
Key Fields: Artist ID, Artist Name
Purpose: Central table for organizing all artists in the media store.

2. ALBUMS

Description: Contains album information linked to artists.
Key Fields: Album ID, Title, Artist ID
Purpose: Links multiple tracks to a single album and artist.

3. TRACKS

Description: Individual songs or media files with detailed metadata.
Key Fields: Track ID, Name, Album ID, Media Type ID, Genre ID, Composer, Duration, Size, Unit Price
Purpose: Core table containing all purchasable media items.

4. GENRES

Description: Music and media genre classifications.
Key Fields: Genre ID, Genre Name
Purpose: Categorizes tracks by musical style (Rock, Jazz, Classical, etc.).

5. MEDIATYPES

Description: Different formats of media files.
Key Fields: Media Type ID, Name
Purpose: Defines file formats like MP3, AAC, MPEG video, etc.

6. PLAYLISTS

Description: User-created or curated music playlists.
Key Fields: Playlist ID, Playlist Name
Purpose: Allows grouping of tracks into custom collections.

7. PLAYLISTTRACK

Description: Junction table linking playlists to tracks.
Key Fields: Playlist ID, Track ID
Purpose: Enables many-to-many relationship between playlists and tracks.

8. CUSTOMERS

Description: Customer information and contact details.
Key Fields: Customer ID, First Name, Last Name, Company, Address, City, State, Country, Postal Code, Phone, Email, Support Rep ID
Purpose: Stores all customer data for the media store.

9. EMPLOYEES

Description: Employee records including organizational hierarchy.
Key Fields: Employee ID, Last Name, First Name, Title, Reports To, Birth Date, Hire Date, Address, City, State, Country, Phone, Email
Purpose: Manages staff information and organizational structure.

10. INVOICES

Description: Customer purchase records and billing information.
Key Fields: Invoice ID, Customer ID, Invoice Date, Billing Address, Total
Purpose: Tracks all sales transactions with complete billing details.

11. INVOICELINES

Description: Individual line items on each invoice.
Key Fields: Invoice Line ID, Invoice ID, Track ID, Unit Price, Quantity
Purpose: Details each item purchased in a transaction.

==========

TABLE RELATIONSHIPS
---

1. Artists → Albums (One-to-Many)
   Each artist can have multiple albums.

2. Albums → Tracks (One-to-Many)
   Each album contains multiple tracks.

3. Genres → Tracks (One-to-Many)
   Each genre categorizes many tracks.

4. MediaTypes → Tracks (One-to-Many)
   Each media format applies to many tracks.

5. Playlists ↔ Tracks (Many-to-Many)
   Through PlaylistTrack junction table.

6. Customers → Invoices (One-to-Many)
   Each customer can have multiple purchases.

7. Invoices → InvoiceLines (One-to-Many)
   Each invoice contains multiple line items.

8. Tracks → InvoiceLines (One-to-Many)
   Tracks can be sold multiple times.

9. Employees → Customers (One-to-Many)
   Support reps manage multiple customers.

10. Employees → Employees (Self-referencing)
    Implements organizational hierarchy.

==========

COMMON USE CASES

EDUCATIONAL TRAINING
Teaching database concepts, normalization, and query optimization.

RAG SYSTEMS DEVELOPMENT
Building question-answering systems over structured data.

BUSINESS ANALYTICS
- Sales analysis and revenue trends
- Customer segmentation
- Product performance tracking
- Geographical market analysis

APPLICATION DEVELOPMENT
Testing e-commerce systems and inventory management.

DATA SCIENCE & VISUALIZATION
Practicing data analysis and creating visualizations.

API DEVELOPMENT
Building REST APIs with realistic relational data.

==========

BUSINESS INSIGHTS

SALES & REVENUE ANALYSIS
Complete sales history enables revenue trend analysis, peak period identification, and customer lifetime value calculations.

CUSTOMER MANAGEMENT
Comprehensive data supports customer segmentation, targeted marketing, and support workload distribution.

INVENTORY ORGANIZATION
Media items organized by artist, album, genre, and format for effective catalog management.

ORGANIZATIONAL STRUCTURE
Employee hierarchy supports team structure evaluation and workload assessment.

PRODUCT PERFORMANCE
Track sales data reveals bestsellers, genre preferences, and pricing optimization opportunities.

==========

DATA CHARACTERISTICS

TIME PERIOD: 2009-2013 transaction data

GEOGRAPHIC COVERAGE: International customer and employee data from multiple countries

CONTENT DIVERSITY:
- Diverse musical genres
- Various media formats
- Different pricing tiers

DATA QUALITY:
- Real iTunes library data combined with realistic fictional data
- Maintains referential integrity
- Properly normalized structure
- Realistic data volumes

==========

WHY CHOOSE CHINOOK?

Chinook provides a realistic, well-structured dataset that mirrors real-world business scenarios.

KEY ADVANTAGES:
• Proper database normalization
• Diverse relationship types
• Realistic data volumes
• Well-documented structure
• Industry-standard schema design
• Suitable for education and testing

==========

END OF DOCUMENTATION