**Insights using MySQL to explore Global Trade Item Number (GTIN) data structures**

Randy Lisbona, Marvin Scott, Vinh Le

**Abstract**

MySQL is a popular open source database format. GTIN (Global Trade Item Number) is a family of product ID codes, used worldwide between trading partners, to identify products via linear or 2D barcodes on the product or packaging. This project will provide insight and first impressions on the ease of use of mySQL using GTIN example data.

An open source GTIN subset > 100K records (.csv, .sql, or JSON format) will be downloaded and installed on a local mySQL database instance. Open source documentation will be reviewed to better understand the GTIN data. The schema will be checked for normalization and modified if necessary, Primary and foreign key indexes will be created/modified as needed. A graphical schema of the database will be created with example data from each table to visualize the relationship between tables. Exploratory queries will be created to summarize selected fields such as brand, and packaging level. A summary list of mySQL commands used in the analysis will be provided along with impressions on ease of use, intuitiveness, and effectiveness.

**Proposed Outline**

Introduction, why the GTIN dataset, (open source, 1M + records, a variety of tables. Where we got our tables, steps to load it. Steps to import BSIN code in JSON <http://brand-okfn.herokuapp.com/brand/>

Investigate feasibility of doing this project on webhosted mySQL like <https://www.000webhost.com/>

Background on GTIN from [www.gs1.org](http://www.gs1.org)

, who uses it, can we find/estimate worldwide usage. Describe various formats for GTIN =(companycode+Itemcode) How many unique companies and itemID’s can it encode. We have GTIN-13, there are a few others. Appears that there is not an open source version of the complete database, trading partners publish their data to companies selling the products? Can we explain why? Explain why price is not in this database.

EER diagram– Randy working on this. Steps in MySQL to create EER, Reverse engineering the schema in MySQL didn’t bring in the tables. Data Export to SQL script, import mysql reverse engineer script does work, steps to explore and define relationships. Need explanation of nearly empty skeleton tables found with just a few records. Base our queries on the

EER Graphic

Queries: Basic stats on each table, records, keys, do we need to add indexes? File size of zipped, unzipped, and MySQL database. Create several example summary queries, a few charts:

1. ItemCount by Company
2. Brand Count by company
3. Average Package size ?

Use Excel PowerQuery for charts, need to install SQL connector <https://support.office.com/en-gb/article/Connect-to-a-MySQL-database-Power-Query-8760c647-88b9-409d-b312-6ea8f84a269b?ui=en-US&rs=en-GB&ad=GB>

Insights:

1. Average package size
2. Min max avg number of items per company.
3. Brands per company
4. Ease of using MySQL for relationships, queries, data exploration

Summary, what we learned with each of us working on our own local copy, vs a web version if we can get that working.

Appendix

Github Repository <https://github.com/rlisbona/MSDS-7330-Term-Paper-1>

Data Sources

|  |  |
| --- | --- |
| Product Open Data – Subset of GTIN | <http://www.product-open-data.com/en/1-home.html> |
| POD database SQL Create and Load | <http://www.product-open-data.com/docs/pod_web_2014.01.01_01.sql.gz> |
| POD database Specification | <http://www.product-open-data.com/docs/POD-SPECS-2013.11.13_01.xlsx> |
| BSIN – Brand to Product (JSON) | <http://brand-okfn.herokuapp.com/brand/api/v1/brand/?format=json> |
| GS1.org – Standards Organization for GTIN | <http://www.gs1.org/> |
|  |  |

Related works

|  |  |
| --- | --- |
| None found so far |  |
|  |  |

POD Database

