

Predicting Prices, Volume and Trade for Meat and Dairy Products: An Agri Food Data Model

Suzanne McCarthy*, Andrew McCarren, and Mark Roantree

Insight Centre for Data Analytics, School of Computing, Dublin City University,
Dublin 9, Ireland.

suzanne.mccarthy@insight-centre.org, andrew.mccarren@dcu.ie,
mark.roantree@dcu.ie

Insight Technical Report Ref. McR-2017-01 (April 2017)

* Research funded by Science Foundation Ireland under grant number SFI/12/RC/2289

1 Introduction

The Agri food industry in Ireland accounts for approximately 5.7% of gross value added, 9.8% of Ireland's trade exports and 8.5% of national employment [2] and Agri companies are increasingly making use of data analytics. Start-ups and research groups are emerging for the purpose of addressing the need for data analytics unique to the Agri sector [1]. Extract-Transform-Load (ETL) as a framework for data integration gained popularity in the 1970's and quickly became the standard process for data integration. Traditional ETL grabs a 'snapshot' of the source data at given intervals and updates the data warehouse or data marts [5]. As part of this process, the source data is extracted, transformed according to relevant rules and loaded into a warehouse where it remains until it is needed for user queries.

Heterogeneity of data is often cited as a major challenge for data integration, especially for large quantities of data and diversity of structures. Heterogeneity of data leads to heterogeneity of data schemas and non-integrated data management systems needing multiple query languages. A metadata organisation system is required to prevent it from becoming an unstructured 'data swamp' [?]. At the core of this metadata system should be a canonical data model which captures all aspects of the data requirements of the overall system.

In this report, we present our Agri data Model which effectively provides the structure for an Agri Data Warehouse. This warehouse stores data from web and enterprise sources in a unified, homogeneous view created from disparate, structured and semi-structured datasets. Web data relevant to the Agri sector is usually freely available from official sources but there is little unity between sources, leading to semantic conflicts and data quality issues. A common model adds value to Agri businesses by making their Business Intelligence reports easier.

The scale and diversity of the Agri sector makes it very difficult to capture in a single data model and, for this reason, we narrow our focus to a specific subset of the Agri sector. The data model reflects the data model required for the food processing industry, and is designed to allow one to extract data marts that can be used to predict volume, prices and trade of meat and dairy products. In building this model, we focused on a couple of key products and processes. Our data sources are primarily in the areas of:

- International trade of Agri products such as beef cuts, animal carcasses etc, measured in both weight and value [27, 15, 12, 23, 24]
- Production levels and price of animal feed, i.e. grains etc [9, 16, 18, 22]
- Production of dairy products [29, 27, 11, 13, 14]
- Slaughter numbers of animals, measured in both weight and head [15, 27, 10, 9]
- Prices of animals, both live and whole carcasses [9, 10]
- Economic information such as GDP, population etc on an international level [20]
- Prices of supermarket products [26, 25, 19, 8, 28]

1.1 Report Structure

The remainder of this report is structure as follows: in §2, we present a high level overview of the facts and dimensions captured in our data model; in §3, the name and description of each attribute for every fact and dimension is given; and finally in §4, we present some conclusions.

2 Agri Data Model: An Overview

Our working assumption is that data will arrive from different sources and be dimensional in nature as dimensional data provides for the most powerful forms of analysis [3]. A dimensional schema is built in a starlike fashion with measurement data (quantity sold, price) at its core and the analytical attributes (region, product, date) at the tips of the star. The data model was built, dimensions and measures chosen, and data marts selected based on a number of business requirements, such as a need to view the weekly price of live pigs in Ireland for the previous year. Examples of some of our sources and a demonstration of the sort of use cases of our model can be found in [4]. The current sets of facts, dimensions and data marts are presented below.

2.1 Facts

The fact tables hold one or more measures of interest to the Agri sector along with links to each of the relevant dimensions, which provide context for the data. There are 27 fact tables which hold measures such as price, slaughters (measured in heads), production (measured in litres or kilograms), trade weight and trade value. Tables 1,2 and 3 show the fact table names, separated by granularities Yearly, Monthly and more fine-grained respectively.

Table 1. Facts: Yearly Granularity

fact_consumption_annual
fact_demographic_population
fact_economic_forecast
fact_economic_stats
fact_population_annual
fact_production_annual
fact_slaughter_annual
fact_supply_demand_annual
fact_vital_statistics

Table 2. Facts: Monthly Granularity

fact_cold_storage_monthly
fact_economic_forecast
fact_index_monthly
fact_normalised_gdp
fact_population_monthly
fact_price_monthly
fact_production_monthly
fact_sales_monthly
fact_slaughter_monthly
fact_stocks_monthly
fact_supply_demand_monthly
fact_trade_monthly

Table 3. Facts: biweekly, weekly and daily

fact_price_biweekly
fact_price_supermarket_weekly
fact_price_weekly
fact_production_daily
fact_production_weekly
fact_slaughter_weekly
fact_usda_all_info_weekly

Table 4. Dimensions

Static dimensions	Dynamic dimensions
dim_age_group	dim_currency_annual
dim_date_annual	dim_currency_daily
dim_date_daily	dim_currency_monthly
dim_date_monthly	dim_currency_weekly
dim_date_weekly	dim_geo
dim_gender	dim_index_type
dim_grade	dim_measurement_feature
dim_outlier	dim_population_economics_type
dim_price_type	dim_product
dim_source	dim_supply_demand_type
dim_stat_regime	dim_trade_product
dim_status	dim_unit
dim_trade_flow	

2.2 Dimensions

Dimension tables provide the real power in dimensional schemas as they facilitate the creation of datasets across multiple dimensions using the OLAP (OnLine Analytical Processing) methodology. In a purely structural sense, they are linked to the fact tables by way of Foreign Keys. Our data model contains 25 dimensions and some will apply to all fact tables, such as `dim_source` (links the data to the original source) and `dim_outlier` (the result of an anomaly detection algorithm as seen in [4]), while other dimensions that are frequently used would include `dim_geo`, `dim_product` and each of the date dimensions.

The dimensions are categorised into two sets: static dimensions, which are populated once and should rarely or never need to change, and dynamic dimensions, which are populated regularly from the source raw data.

2.3 Data Marts

The data model is large and it is unlikely that any analysis will require the entire data warehouse. Thus, a single business requirement will generally require the construction of a data mart which is a subset (similar to a database view) of the overall schema. Those data marts which are most popular, requiring constant reuse are now described. Each Data Mart is shown with the name of the relevant fact table, followed by the level of granularity and a list of each of the dimensions to which the fact table is linked. Many of the dimensions are shared between multiple fact tables and, as stated previously, there are some dimensions that will be found in all data marts. Fig 1 illustrates an example of a frequently used Data Mart where `price` is the measure and the rest of the attributes in the fact table are links to each of the dimensions.

Table 5: Data Marts

fact_cold_storage_monthly	fact_consumption_annual
Grain: Monthly	Grain: Yearly
<code>dim_date_monthly</code>	<code>dim_date_annual</code>
<code>dim_geo</code>	<code>dim_geo</code>
<code>dim_product</code>	<code>dim_product</code>
<code>dim_unit</code>	<code>dim_unit</code>
<code>dim_source</code>	<code>dim_source</code>
<code>dim_outlier</code>	<code>dim_outlier</code>

Continued on next page

Table 5 – *Continued from previous page*

fact_demographic_population	fact_economic_forecast
Grain: Yearly	Grain: Yearly
dim_date_annual	dim_date_annual
dim_geo	dim_geo
dim_gender	dim_product
dim_age_group	dim_measurement_feature
dim_status	dim_unit
dim_population_economics_type	dim_source
dim_unit	dim_outlier
dim_source	
dim_outlier	
fact_economic_stats	fact_index_monthly
Grain: Yearly	Grain: Monthly
dim_date_annual	dim_date_monthly
dim_geo	dim_geo
dim_population_economics_type	dim_product
dim_unit	dim_index_type
dim_currency_annual	dim_unit
dim_source	dim_source
dim_outlier	dim_outlier
fact_normalised_gdp	fact_population_annual
Grain: Monthly	Grain: Yearly
dim_date_monthly	dim_date_annual
dim_geo	dim_geo
dim_unit	dim_product
dim_source	dim_unit
dim_outlier	dim_source
	dim_outlier
fact_population_monthly	fact_price_biweekly
Grain: Monthly	Grain: Biweekly
dim_date_monthly	dim_date_daily
dim_geo	dim_geo
dim_product	dim_product
dim_unit	dim_price_type
dim_source	dim_grade
dim_outlier	dim_unit
	dim_currency_weekly
	dim_status
	dim_source
	dim_outlier

Continued on next page

Table 5 – *Continued from previous page*

fact_price_monthly	fact_price_supermarket_weekly
Grain: Monthly	Grain: Weekly
dim_date_monthly	dim_date_daily
dim_geo	dim_geo
dim_product	dim_product
dim_price_type	dim_unit
dim_grade	dim_currency_weekly
dim_unit	dim_source
dim_currency_monthly	dim_outlier
dim_status	
dim_source	
dim_outlier	
fact_price_weekly	fact_production_annual
Grain: Weekly	Grain: Yearly
dim_date_daily	dim_date_annual
dim_geo	dim_geo
dim_product	dim_product
dim_price_type	dim_unit
dim_unit	dim_source
dim_currency_weekly	dim_outlier
dim_source	
dim_outlier	
fact_production_daily	fact_production_monthly
Grain: Daily	Grain: Monthly
dim_date_daily	dim_date_monthly
dim_geo	dim_geo
dim_product	dim_product
dim_unit	dim_unit
dim_source	dim_source
dim_outlier	dim_outlier
fact_production_weekly	fact_sales_monthly
Grain: Weekly	Grain: Monthly
dim_date_daily	dim_date_monthly
dim_geo	dim_geo
dim_product	dim_product
dim_unit	dim_unit
dim_source	dim_source
dim_outlier	dim_outlier

Continued on next page

Table 5 – *Continued from previous page*

fact_slaughter_annual	fact_slaughter_monthly
Grain: Yearly	Grain: Monthly
dim_date_annual	dim_date_monthly
dim_geo	dim_geo
dim_product	dim_product
dim_unit	dim_unit
dim_source	dim_source
dim_outlier	dim_outlier
fact_slaughter_weekly	fact_stocks_monthly
Grain: Weekly	Grain: Monthly
dim_date_daily	dim_date_monthly
dim_geo	dim_geo
dim_product	dim_product
dim_unit	dim_unit
dim_source	dim_source
dim_outlier	dim_outlier
fact_supply_demand_annual	fact_supply_demand_monthly
Grain: Yearly	Grain: Monthly
dim_date_annual	dim_date_monthly
dim_geo	dim_geo
dim_product	dim_product
dim_supply_demand_type	dim_supply_demand_type
dim_unit	dim_unit
dim_source	dim_source
dim_outlier	dim_outlier
fact_trade_monthly	fact_usda_all_info_weekly
Grain: Monthly	Grain: Weekly
dim_date_monthly	dim_date_daily
dim_geo	dim_geo
dim_trade_product	dim_product
dim_trade_flow	dim_unit
dim_stat_regime	dim_source
dim_unit	dim_outlier
dim_currency_monthly	
dim_source	
dim_outlier	

Continued on next page

Table 5 – *Continued from previous page*

fact_vital_statistics

Grain: Yearly

dim_date_annual

dim_geo

dim_population_economics_type

dim_unit

dim_source

dim_outlier

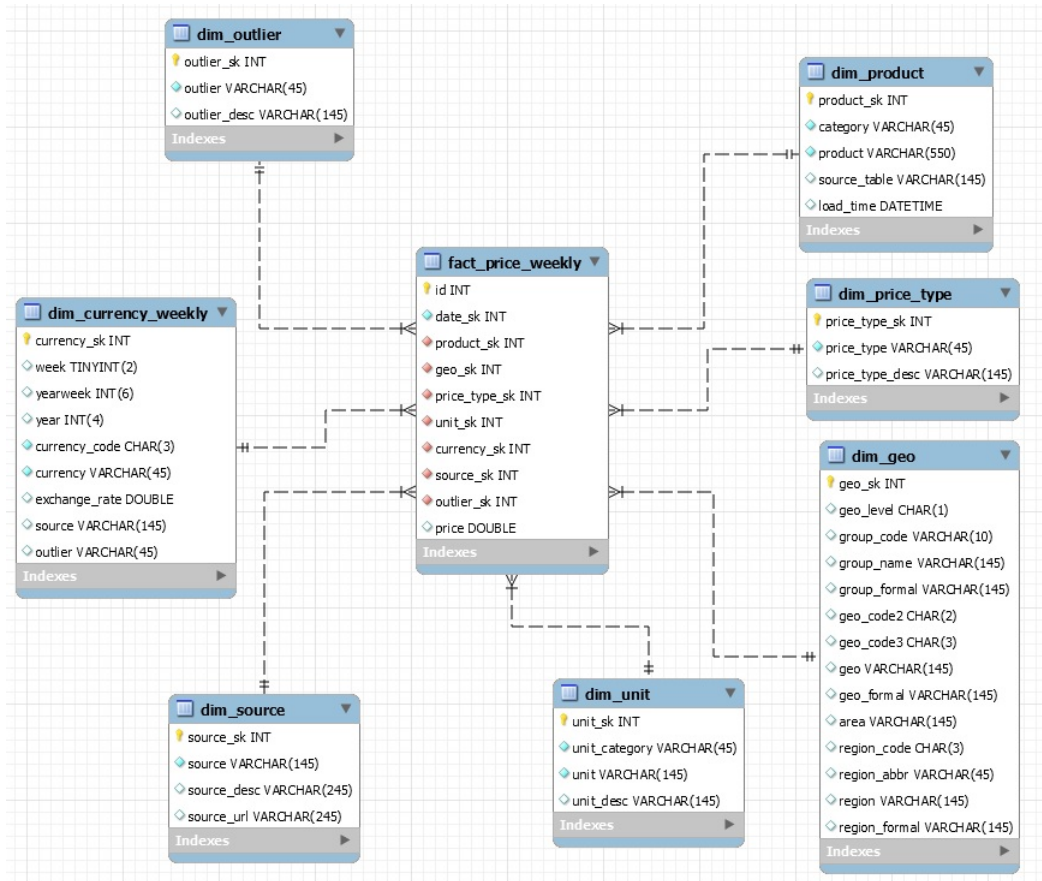


Fig. 1. Fact_price_weekly Data Mart

3 Agri Data Model: A Detailed View

For each of the dimension and fact tables in the Warehouse and listed in §2, a list of table attributes (columns) is provided along with a brief explanation of the semantics or purpose of the attribute. It can be seen that each dimension has a primary key which provides a surrogate key for the fact tables. The fact tables are then composed of an ID, one or more measures, and one or more foreign key links to each of the relevant dimensions, where the number of foreign keys represents the dimensionality for that measure.

Table 6. dim_age_group

age_group_sk	Primary key, auto-increment
age_group	Short description of age group e.g. 40-44
age_group_desc	Longer description of age group e.g. forty to forty-four

Table 7. dim_currency_annual

currency_sk	Primary key, auto-increment
year	year in format YYYY
currency_code	IMF currency code
currency	Name of currency unit
exchange_rate	Rate of exchange to SDR (Special Drawing Right)
source	Source website
outlier	Verdict of outlier detection algorithm

Table 8. dim_currency_daily

currency_sk	Primary key, auto-increment
year	year in format YYYY
quarter	quarter
month	month
yearmonth	year and month concatenated
week	week
yearweek	year and week concatenated
date	date
day	day
currency_code	IMF code of currency unit
currency	Name of currency unit
exchange_rate	Rate of exchange to SDR (Special Drawing Right)
source	Source website
outlier	Verdict of outlier detection algorithm

Table 9. dim_currency_monthly

currency_sk	Primary key, auto-increment
month	month
yearmonth	year and month concatenated
year	year in format YYYY
currency_code	IMF currency code
currency	Name of currency unit
exchange_rate	Rate of exchange to SDR (Special Drawing Right)
source	Source website
outlier	Verdict of outlier detection algorithm

Table 10. dim_currency_weekly

currency_sk	Primary key, auto-increment
year	year in format YYYY
week	week
yearweek	year and week concatenated
currency_code	IMF currency code
currency	Name of currency unit
exchange_rate	Rate of exchange to SDR (Special Drawing Right)
source	Source website
outlier	Verdict of outlier detection algorithm

Table 11. dim_date_annual

date_sk	Primary key, auto-incremented
year	year in format YYYY

Table 12. dim_date_daily

date_sk	Primary key, auto-incremented
date	date in format YYYY-MM-DD
date_iso	date in format YY-M-DD
week	week number
month	month in format MM
month_name	Month of the year e.g. January
month_abbr	Abbreviated month of the year e.g. Jan
quarter	Quarter number of the year
quarter_name	Name of quarter
year	year in format YYYY
day_of_month	day in format DD
day_of_week	day in format D
day_of_year	day in format DDD
day_name	Day of the week e.g. Monday
day_abbr	Abbreviated day of the week e.g. Mon
is_weekend	1 if date is weekend, otherwise 0
is_weekday	1 if date is weekday, otherwise 0
yearweek	year and week concatenated
yearmonth	year and month concatenated

Table 13. dim_date_monthly

date_sk	Primary key, auto-incremented
month	month in format MM
year	year in format YYYY
yearmonth	year and month concatenated

Table 14. dim_date_weekly

date_sk	Primary key, auto-incremented
week	week number
month	month number
year	year in format YYYY
yearweek	year and week concatenated
yearmonth	year and month concatenated

Table 15. dim_gender

gender_sk	Primary key, auto-incremented
gender	Short description of gender e.g. M
gender_desc	Longer description of gender e.g. Male

Table 16. dim_geo

geo_sk	Primary key, auto-incremented
geo_level	A numerical value to indicate the level of the geo: from 0 for a continent or other large group of countries to 5 for a subdivision of a country.
group_code	Abbreviated code for name of a large group of countries
group_name	Name of a large group of countries
group_formal	Formal name of a large group of countries
geo_code2	2-letter ISO code
geo_code3	3-letter ISO code
geo	Common name of individual country
geo_formal	Formal name of individual country
area	Common name of high-level subdivision of country
region_code	Code name of low-level subdivision of country
region_abbr	Abbreviated name of low-level subdivision of country
region	Common name of low-level subdivision of country
region_formal	Formal name of low-level subdivision of country

Table 17. dim_grade

grade_sk	Primary key, auto-incremented
grade	Short description of grade e.g. Standard
grade_desc	Longer description of grade e.g. standard based on the ministerial ordinance

Table 18. dim_index_type

index_type_sk	Primary key, auto-incremented
index_type	Index type
index_type_desc	Description of index type

Table 19. dim_measurement_feature

meas_feature_sk	Primary key, auto-incremented
meas_feature	Measurement feature
meas_feature_desc	Description of measurement feature

Table 20. dim_outlier

outlier_sk	Primary key, auto-incremented
outlier	Short description of verdict of outlier detection algorithm e.g. standard
outlier_desc	Longer description of verdict of outlier detection algorithm e.g. normal value

Table 21. dim_population_economics_type

pop_eco_type_sk	Primary key, auto-incremented
pop_eco_type	Population or economics type
pop_eco_type_desc	Description of population or economics type

Table 22. dim_price_type

price_type_sk	Primary key, auto-incremented
price_type	Short description of price type
price_type_desc	Longer description of price type

Table 23. dim_product

product_sk	Primary key, auto-incremented
category	Category of product
category_desc	Description of category
product_code	Product code if relevant
product	Name of product
product_source_table	Raw table from which product was extracted
load_time	Date and time when product was extracted from product_source_table

Table 24. dim_source

source_sk	Primary key, auto-incremented
source	Name of source website
source_desc	Description of source
source_url	URL to access source website

Table 25. dim_stat_regime

stat_regime_sk	Primary key, auto-incremented
stat_regime	Description of trade statistic regime

Table 26. dim_status

status_sk	Primary key, auto-incremented
status	Single character indicating status of data e.g. P
status_desc	Description of status e.g. Provisional

Table 27. dim_supply_demand_type

sup_dem_type_sk	Primary key, auto-incremented
sup_dem_type	Short description of aspect of supply or demand
sup_dem_type_desc	Short description of aspect of supply or demand
source_table	Raw table from which product was extracted
load_time	Date and time when product was extracted from source_table

Table 28. dim_trade_flow

flow_sk	Primary key, auto-incremented
flow	Single character indicating flow of trade goods e.g. E
flow_desc	Description of flow of trade goods e.g. export

Table 29. dim_trade_product

trade_product_sk	Primary key, auto-incremented
product_code	10-digit HS (Harmonised System) code
product_desc	Description of trade product

Table 30. dim_unit

unit_sk	Primary key, auto-incremented
unit_category	Value that unit measures e.g. price, weight
unit	Name of unit
unit_desc	Description of unit
source_table	Raw table from which unit was extracted
load_time	Date and time when unit was extracted from source_table

Table 31. fact_cold_storage_monthly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_monthly
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
storage	measure

Table 32. fact_consumption_annual

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_annual
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
consumption	measure

Table 33. fact_demographic_population

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_annual
geo_sk	Foreign Key of dim_geo
pop_eco_type_sk	Foreign Key of dim_product
gender_sk	Foreign Key of dim_gender
age_group_sk	Foreign Key of dim_age_group
unit_sk	Foreign Key of dim_unit
status_sk	Foreign Key of dim_status
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
population	measure

Table 34. fact_economic_forecast

economic_forecast_id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_annual
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
feature_sk	Foreign Key of dim_measurement_feature
unit_sk	Foreign Key of dim_unit
status_sk	Foreign Key of dim_status
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
forecast_stat	measure

Table 35. fact_economic_stats

economic_stats_id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_annual
geo_sk	Foreign Key of dim_geo
pop_eco_type_sk	Foreign Key of dim_population_economics_type
unit_sk	Foreign Key of dim_unit
currency_sk	Foreign Key of dim_currency_annual
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
economic_stat	measure

Table 36. fact_index_monthly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_monthly
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
index_type_sk	Foreign Key of dim_index_type
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
index_value	measure

Table 37. fact_normalised_gdp

gdp_id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_monthly
geo_sk	Foreign Key of dim_geo
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
gdp_index	measure

Table 38. fact_population_annual

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_annual
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
population	measure

Table 39. fact_population_monthly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_monthly
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
population	measure

Table 40. fact_price_biweekly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_daily
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
price_type_sk	Foreign Key of dim_price_type
grade_sk	Foreign Key of dim_grade
unit_sk	Foreign Key of dim_unit
currency_sk	Foreign Key of dim_currency_daily
status_sk	Foreign Key of dim_status
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
price	measure

Table 41. fact_price_monthly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_daily
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
price_type_sk	Foreign Key of dim_price_type
grade_sk	Foreign Key of dim_grade
unit_sk	Foreign Key of dim_unit
currency_sk	Foreign Key of dim_currency_daily
status_sk	Foreign Key of dim_status
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
price	measure

Table 42. fact_price_supermarket_weekly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_daily
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
currency_sk	Foreign Key of dim_currency_daily
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
price	measure

Table 43. fact_price_weekly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_daily
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
price_type_sk	Foreign Key of dim_price_type
grade_sk	Foreign Key of dim_grade
unit_sk	Foreign Key of dim_unit
currency_sk	Foreign Key of dim_currency_daily
status_sk	Foreign Key of dim_status
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
price	measure

Table 44. fact_production_annual

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_annual
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
production	measure

Table 45. fact_production_daily

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_annual
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
production	measure

Table 46. fact_production_monthly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_annual
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
production	measure

Table 47. fact_production_weekly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_annual
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
production	measure

Table 48. fact_sales_monthly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_daily
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
sales_value	measure

Table 49. fact_slaughter_annual

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_annual
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
slaughter_head	measure

Table 50. fact_slaughter_monthly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_monthly
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_weight_sk	Foreign Key of dim_unit for weight unit
unit_head_sk	Foreign Key of dim_unit for head unit
source_sk	Foreign Key of dim_source
outlier_weight_sk	Foreign Key of dim_outlier for weight measure
outlier_head_sk	Foreign Key of dim_outlier for head measure
slaughter_weight	measure
slaughter_head	measure

Table 51. fact_slaughter_weekly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_daily
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
slaughter_head	measure

Table 52. fact_stocks_monthly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_monthly
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
stocks	measure

Table 53. fact_supply_demand_annual

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_annual
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
sup_dem_type_sk	Foreign Key of dim_supply_demand_type
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
quantity	measure

Table 54. fact_supply_demand_monthly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_annual
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
sup_dem_type_sk	Foreign Key of dim_supply_demand_type
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
quantity	measure

Table 55. fact_trade_monthly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_monthly
reporter_sk	Foreign Key of dim_geo for reporter
partner_sk	Foreign Key of dim_geo for partner
trade_product_sk	Foreign Key of dim_trade_product
unit_weight_sk	Foreign Key of dim_unit for the weight unit
unit_value_sk	Foreign Key of dim_unit for the value unit
flow_sk	Foreign Key of dim_trade_flow
stat_regime_sk	Foreign Key of dim_stat_regime
currency_sk	Foreign Key of dim_currency_monthly
source_sk	Foreign Key of dim_source
outlier_weight_sk	Foreign Key of dim_outlier for trade_weight
outlier_value_sk	Foreign Key of dim_outlier for trade_value
trade_value	measure
trade_weight	measure

Table 56. fact_usda_all_info_weekly

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_daily
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
value	measure

Table 57. fact_vital_statistics

id	Primary key, auto-incremented
date_sk	Foreign Key of dim_date_daily
geo_sk	Foreign Key of dim_geo
product_sk	Foreign Key of dim_product
pop_eco_type_sk	Foreign Key of dim_populations_economics_type
status_sk	Foreign Key of dim_status
unit_sk	Foreign Key of dim_unit
source_sk	Foreign Key of dim_source
outlier_sk	Foreign Key of dim_outlier
vital_stat	measure

4 Conclusions

In this report, we presented the detailed structure of the facts and dimensions that comprise our multi-dimensional Agri data model. We also illustrated a number of the most popular sample data marts that provide the user interface to our data warehouse. Our aim is that this can provide a common model to which data marts produced from this warehouse will conform, while the model also remains extensible for future data sources. This report describes v1.0 of our data model but it is likely to extend over time, as new sources are added, which provide new measures (facts) or opportunities for further analyses (dimensions).

References

1. DIT Agriculture Analytics Research Group. <http://www.agrianalytics.org/>. 2017.
2. Central Statistics Office. <http://www.cso.ie/en/index.html>. 2018.
3. Kimball R. and Ross M. *The Data Warehousing Toolkit: The Complete Guide to Dimensional Modelling*, Wiley, 2002.
4. McCarren A, McCarthy S, O Sullivan C, and Roantree M. Anomaly Detection in Agri Warehouse Construction. Proceedings of the ACSW, ACM press, pages 1-17, 2017. DOI: 10.1145/3014812.3014829
5. Zhu Y, An L, and Liu S. Data updating and query in real-time data warehouse system. In International Conference on Computer Science and Software Engineering, CSSE 2008, 2008, Wuhan, China, pp 1295-1297, 2008.
6. AimisSimia. <http://aimis-simia.agr.gc.ca/>. 2017.
7. Alic. <http://www.alic.go.jp/english/index.html>. 2017.
8. Asda. <http://www.asda.com/>. 2017.
9. Bordbia. <http://www.bordbia.ie/Pages/Default.aspx>. 2017.
10. AHDB. <http://pork.ahdb.org.uk/>. 2017.
11. CLAL. <http://www.clal.it/en/>. 2017.
12. Comtrade. <http://comtrade.un.org/>. 2017.
13. DairyAustralia. <http://www.dairyaustralia.com.au/>. 2017.
14. Dairyco. <http://www.dairyco.org.uk/>. 2017.
15. Eurostat. <http://ec.europa.eu/eurostat>. 2017.
16. Futures. <http://future.aae.wisc.edu/>. 2017.
17. Global Dairy Trade. <https://www.globaldairytrade.info/>. 2017.
18. HGCA. <http://www.hgca.com/>. 2017.
19. Morrisons. <http://www.morrisons.com/>. 2017.
20. OECD. <http://www.oecd.org/>. 2017.
21. Pig333. <https://www.pig333.com/>. 2017.
22. Quandl. <https://www.quandl.com/>. 2017.
23. Statistics Canada. <http://www.statcan.gc.ca/start-debut-eng.html>. 2017.
24. Stats New Zealand. <http://stats.govt.nz/>. 2017.
25. Supervalu. <https://supervalu.ie/>. 2017.
26. Tesco. <http://tesco.ie/>. 2017.
27. United States Department of Agriculture. <http://www.usda.gov/wps/portal/usda/usdahome>. 2017.
28. Walmart. <http://www.walmart.com/>. 2017.
29. ZuivelNL. <http://www.zuivelnl.org/>. 2017.