Database Answers

Ruxley Towers,UK

Data Management by Example

Barry Williams
March 25th 2017
barryw@databaseanswers.org

Second Edition, March 25th. 2017

1. Management Summary

1.1 Why did I write this book ?

My purpose is to summarise my experience over 20 years in a format that anybody can use, refer to and contribute to.
Therefore, over a period of time it can be used as a reference that will evolve to always reflect our experience.
If you could like to add comments based on your experience and your opinion I would be very pleased to hear from you.
You can reach me on my email address given above.

In the meantime, I hope you enjoy this book and find it helpful.

1.2 Our Approach

Our Objective is to provide an Answer to any Question related to Enterprise Data Management.
Our Approach is to maintain a Manual of Best Practice for Practitioners to use :-

- Build a Library of Techniques that Work for Practitioners
- Define our Reference Data Architecture with Components (such as a Data Warehouse)
- For each Component :-
  - Create Data Models for a range of Industries (eg Banking, Logistics and Retail)
  - Generate SQL Scripts
- We have created starting-points for over 20 Industries on this page :-
  - http://www.databaseanswers.org/data_models/POC_Cloud_Services.htm

Our current Reference Data Architecture is on this page

- http://www.databaseanswers.org/reference_data_architecture.htm

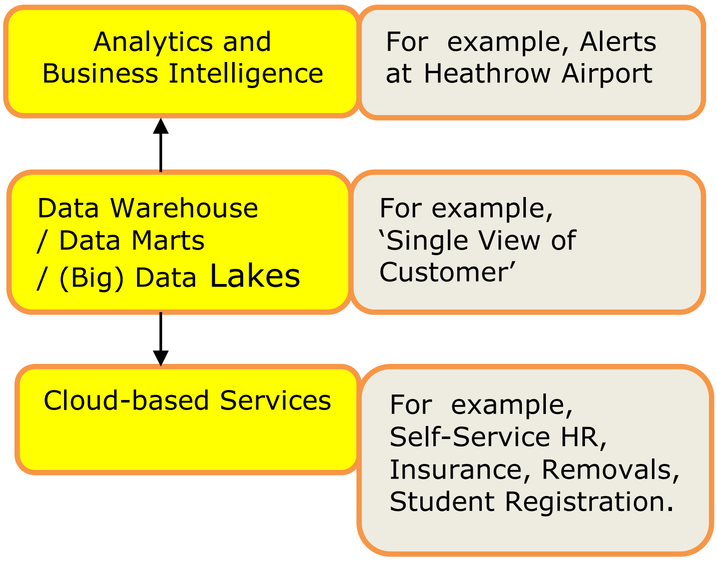
2. Reference Data Architecture

2.1 High-Level Overview

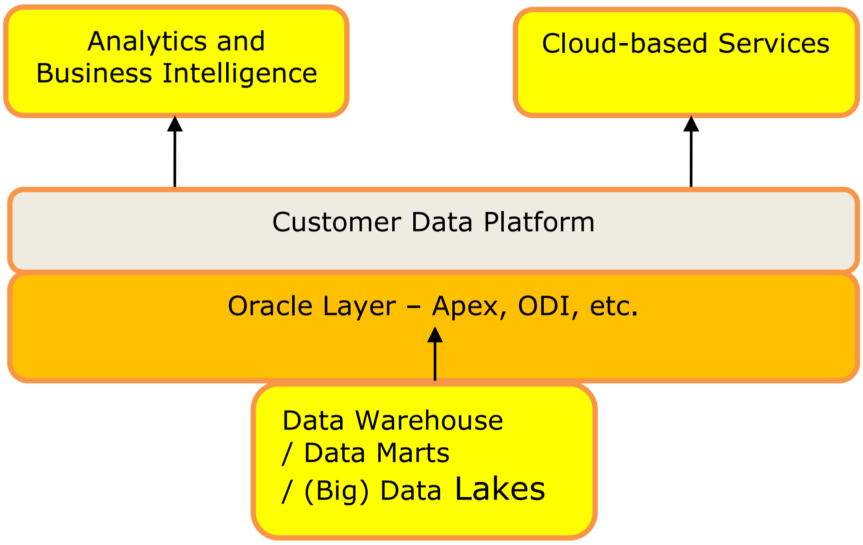
The Stages in our Reference Data Architecture provide the framework for our study of Best Practice in Data Management.

These are the three top-level view of the building-blocks.
We show that the Data Warehouse complex provides source data fort both the Analytics layer and the Cloud-based Services.

This shows our Layered Arvhitecture :-


Analytics and Business Intelligence
For example, Alerts at Heathrow Airport
Data Warehouse / Data Marts / (Big) Data Lakes
For example, 'Single View of Customer'
Cloud-based Services
For example, Self-Service HR, Insurance, Removals, Student Registration.

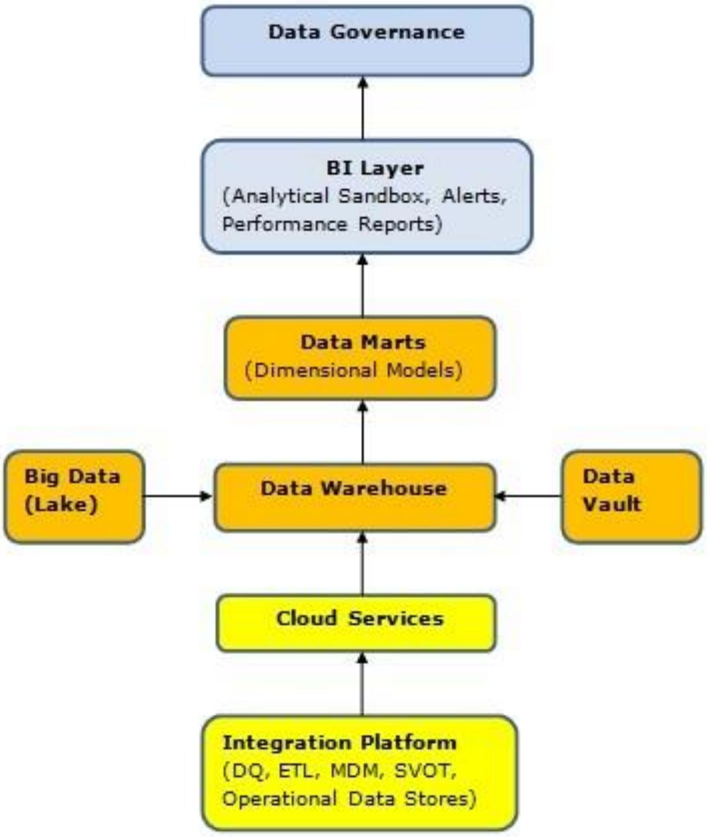
This shows how the Data Layer provides data to the Analytics and the Cloud Services layers :-


Analytics and Business Intelligence
Cloud-based Services
Customer Data Platform
Oracle Layer – Apex, ODI, etc.
Data Warehouse / Data Marts / (Big) Data Lakes

2.2 Details

A more details description of the Stages is shown here and we discuss the What and Why of each Stage.

We would be pleased to receive your comments and our email address is on the first page.



3. Data Governance

3.1 What is it ?

Data Governance is the control of change within an IT and Data environment.
This requires some activities that are general, such as clearly-defined Roles and Responsibilities, and some that are specific to Data Management, such as Data Lineage.
The Sarbanes-Oxley Act came into law in the USA in 2002 and introduced much higher standards of financial compliance for any publicly-quoted company.
This means that Chief Executives have to state that the figures in their Annual Reports are 'the truth, the whole truth and nothing but the truth'.
This in turn has focussed a great deal of attention on Data Governance.
It means, for example, that Data Lineage is now 'front and centre' because it must e possible to trace every item of data in the Annual Reports down to the source of the data in an Operational System within the organisation.

3.2 Why is it important ?

It is important because there is an increasing need for accountability, for a 'Single View of the Truth' and for an understanding of how data is controlled within complex organisations.
All of these factors require a better awareness and better control of changes to the way is sourced, retrieved and processed between Data Sources (qv) and Performance Reports (qv).

3.3 What will I learn ?

You will learn how Data Governance can be implemented and how to determine if an organisation has good Governance in place.
You will also learn how to prepare for Certification and where to go for more detailed information.

3.4 Best Practice

Many people talk about Data Governance and many organisations talk about adopting it.
The reality is that progress is rather slow in implementing Governance.
This is primarily because it requires a commitment from the top to the bottom of management.

3.4.1 Useful Links

The Data Governance Institute –

- http://www.datagovernance.com/

Excellent review of Sarbanes-Oxley in Wikipedia –

- http://en.wikipedia.org/wiki/Sarbanes-Oxley_Act

LinkedIn has a number of relevant Groups, and here are just two of them :-

- Data Stewardship & Governance
- The Data Quality Association

3.5 Templates

This Section shows two examples of Templates.

3.5.1 Data Lineage

This table shows the example of Profitability.

Data Item	Source	Description	Lineage
Profitability	Annual Report	The difference between Revenue and Costs for a specified time-period.	Revenue is obtained from the Billing System. Costs are derived from the Purchasing System.

3.5.2 Roles and Responsibilities

This table shows the example of Profitability.

Role name	Incumbent	Responsibilities
Data Governance Mgr	Jane Doe	Responsible for changes to Roles and Responsibilities
Data Steward	John Doe	Responsible for changes to content of the Information Catalogue

3.6 FAQs

FAQ.1 What is Data Governance ?

Data Governance can be defined simply as 'Doing things right' in Enterprise Data Management by complying with the appropriate rules, policies and procedures.

These will all be designed to make sure that data used throughout the Enterprise is good-quality data, certainly when it appears in Performance reports.

FAQ.2 Why should my organisation have a Data Governance function ?

The existence of a **Data Governance function is a measure of the maturity of Data Management within an organization**

The first steps should be to establish a thin slice of Data Governance from top to bottom

- Wikipedia on Data Governance - http://en.wikipedia.org/wiki/Data_governance
- Alignment of Enterprise Architecture with Business Goals –

http://www.information-management.com/infodirect/2009_115/enterprise_architecture_togaf-10015189-1.html?ET=informationmgmt:e886:2099687a:&st=email

If you are active in this area, you should consider joining a professional organizational.
This helps you to network with your peer group and will encourage you to keep up-to-date in knowledge and professional practice.

Here are two organisations that are planning active roles in Data Governance :-
i) The Data Governance Institute (Membership starts at $150 for individuals) :-
http://www.datagovernance.com/

ii) The Data Governance and Stewardship Community of Practice ($150/year) :-
- http://www.datastewardship.com/
It includes coverage of some very useful Case Studies :-
http://www.datastewardship.com/content.aspx?page_id=22&club_id=885168&module_id=37956

It also maintains a Data Governance Software Web Site :-
http://www.datagovernancesoftware.com/
and Sarbanes-Oxley Web Site - http://www.sox-online.com/

FAQ.3 How do I get a top-down view of Data Management in my organisation ?

Answers to this question are at different levels :-

- Data Governance at the top-level
- Master Data Management at the mid-level
- Data Integration at the mid-level
- Data Owners and Sources at the lowest level
- Information Catalogue mandated as the central repository of all this information
- Appropriate procedures in place to control all of these factors.

4. BI Layer

4.1 What is it ?

The BI Layer sits between the Data Marts or Data Warehouse and the Presentation Layer.
It has become more important with the growth of big data and now incorporates aspects of

4.2 Why is it important ?

It is important because the need for more functionality has increased with Big Data analytics.

4.3 What will I learn ?

You will learn how to identify the requirements related to the user interface and analytical functionality.

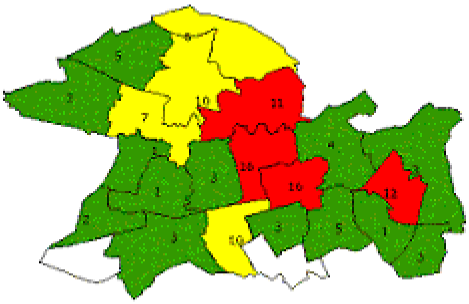
4.4 Best Practice

Best Practice involves articulating user requirements and interactions in terms of user data structures.

4.5 Templates

4.5.1 Map showing KPIs

This Map shows Key Performance Indicators (KPIs) for the Wards in a Local Authority
Each Ward is displayed in either Red, Amber or Green, depending in whether the KPIs Threshold values are reached or exceeded.
Red indicates a situation that requires urgent management attention, amber is a warning and green is within acceptable limits.



4.5.2 Reports at the Regional Level

This Report shows the total count of Customers gained and lost in an imaginary South-East Region

Rpt.1 Total Customers Gained and Lost by Week						
Date selected: End of May, 2012						
Week Ending		**Location**		**Total Gained**		**Total Lost**
March 6th		SE Region		10		10
March 13th		SE Region		20		20
March 20th.		SE Region		30		30
March 27th.		SE Region		40		40
April 3rd/		SE Region		50		50
April 10th.		SE Region		30		30
April 17th.		SE Region		20		20
April 24th.		SE Region		10		10

.

4.5.3 Reports at the City Level

This Report shows the total count of Customers gained and lost for London in the South-East Region.

RPt.1 Total Customers Gained and Lost by Week						
Date selected: End May, 2012						
Week Ending		**Location**		**Total Gained**		**Total Lost**
March 6th 09		London		1		1
March 13th 09		London		2		2
March 20th. 09		London		3		3
March 27th. 09		London		4		4
April 3rd/ 09		London		5		5
April 10th. 09		London		3		3
April 17th. 09		London		2		2
April 24th. 09		London		1		1

4.5.4 Reports for Parking Tickets

This table shows a sample Template of unrealistic data for Parking Ticket Reports.
The Template is available on this page of the Database Answers Web Site :-
http://www.databaseanswers.org/Parking_Rpts/PK06_TotalPaidPCNs_withPaymentMethod_demo_rpt.xls

PK.6 - Report on Total PCNs Paid with Payment Methods

Date selected: Month of January, 2011

PCN Type		**Source**		**Payment Method**		**PCNs Paid**		**Amount Paid**
PCN - BLE		H		Credit Card		5		£300.00
PCN - BLE		O		Cheque		186		£11,160.00
PCN - BLE		O		Credit Card		1		£60.00
PCN - BLE		O		Postal Order		4		£240.00
PCN - BLE		U		Auto Phone Payment		594		£35,700.00
PCN - CCTV		H		Credit Card		3		£150.00
PCN - CCTV		H		Debit Card		5		£250.00
PCN - CCTV		O		Cheque		171		£8,700.00
PCN - CCTV		O		Postal Order		2		£100.00
PCN - CCTV		U		Cash		50		£2,500.00
PCN - CCTV		U		Cheque		5		£250.00
PCN - DTE		H		Credit Card		28		£1,680.00
TOTAL						**10,000**		**£500,000**

4.6 FAQs

FAQ.1 Does your Chief Exec have Report requirements that you cannot meet ?

In order to respond to this situation appropriately, it is necessary to have an Information Catalogue, a Data Architecture and Data Lineage.
The solution then involves the following Steps :-
Step 1) Produce a draft Report for the Chief Execs approval
Step 2) Trace the lineage and perform a 'gap analysis' for all new data items.
Step 3) Talk to the Data Owners and establish when and how the data can be made available.
Step 4) Produce a Plan and timescale
Step 5) Review your Plan with the Chief Exec and obtain this agreement and formal sign-off.
Step 6) Deliver !!!

Performance Reports take data from Data Marts and many of the same considerations apply when it comes to determining **Best Practice**.
One difference is that is necessary to have a clearer understanding of the business operations and how the right kind of Performance Reports can provide insight to the business users.

This leads to the need for a management education process to be in place so that the evolution of Performance Reports can be planned in a logical manner, from basic summaries, to KPIs, Dashboards and so on.

FAQ.2 How do I produce Integrated Performance Reports for senior management ?

The key action here is to establish a unified Reporting Data Platform.
This will involve aspects previously discussed, including MDM, CMI and will certainly involve Data Lineage.
Senior Management will want to take a view of the integrated data and not focus on details of derivation.

Therefore, we have to follow the MDM approach with Data Lineage for each item in the Integrated Performance Reports.

FAQ.3 What are Key Performance Indicators ('KPIs')

Key Performance Indicators ('KPIs') are in common use and represent one aspect of Best Practice.

A variation of this approach are Key Quality Indicators,('KQIs') which are used to monitor and manage Data Quality.

Dashboards and Scorecards are often used in association with KPIs.

FAQ.4 Where can I find a Tutorial on Reporting ?

Here's a Tutorial from Database Answers on Integrated Performance Reporting –
- http://www.databaseanswers.org/tutorial4_integrated_performance_reporting/index.htm

In broad terms, there are three areas involved :-

i) Determine the Data Sources from the Data Marts
ii) Choose the commercial Report-Writer
iii) Create Data Validation and Transformation procedures

FAQ.5 How do I get certified as a Microsoft BI Specialist ?

Certification can be described as 'Necessary but not sufficient'. In other words, some employers consider it as evidence that you have the necessary technical knowledge and skills to be a Database Administrator, but without any experience, it will not guarantee you a job.

If you take your profession seriously and are committed to self-improvement, then you should certainly consider getting certified in the DBMS of your choice.
Here is a Web Link discussing the role of Microsoft Certified Technology Specialist in SQL Server Business Intelligence :-

http://www.microsoft.com/learning/mcp/mcts/bi/default.mspx

FAQ.6 How do I manage request for changes to Reports ?

When you are planning to produce Reports, it is vital to plan for changes to avoid disappointment.
The most common response when Users get their much-anticipated Reports for the first time, is for them to say – "Oh dear, that isn't really what I wanted'.
Even when the Reports meet their Requirements, which will have been well-documented, and probably signed-off by the Users, they still want changes made.

There are some technical things you can do, including specifications for Report Templates which capture the features in families of similar Reports.

From a procedural point of view, you can discuss with the Users, how they see the patterns of future changes, and try to understand the operational environment. This will help you see how the Reports fit into their management style and

You can identify a progression from KPIs (Key Performance Indicators), Traffic Light Reports (using Red, Amber and Green to indicate the seriousness of situations being reported on), Dashboards, Scorecards
This will help you to arrange for the appropriate management education so that you and your Users are always in step, with your planning for what is just around the corner.

FAQ.7 What are the Qualities for Success in Performance Reporting
To be successful in this area of Performance Reporting it is useful to be able to see things from the User's perspective and formulate the layout and content of the Reports accordingly
People who are successful working in this area are happy to work with End-Users and formulate Report requirements in a style that can be easily understood and implemented by the developers who might be the Report specialist.

They are subsequently able to implement the inevitable changes requests by the End-user and manage the expectations of the End-user and developers.

5. Data Marts

5.1 What is it ?

Wikipedia provides a good reference on Data Marts - http://en.wikipedia.org/wiki/Data_mart
Data Marts are always built around Dimensions, such as Dates, Regions and Customers.
The other kinds of data are derived figures, such as totals.

5.2 Why is it important ?

Data Marts are important because they make it very easy to assemble data for Reports.

5.3 What will I learn ?

You will learn what a typical Data Mart looks like, how to design one and other useful facts.

5.4 Best Practice

Your Dimensions will always be Foreign Keys to Tables in your Data Warehouse.
Best Practice suggests that you position the Dimensions at the top of the Data Mart, and listed in alphabetical order.

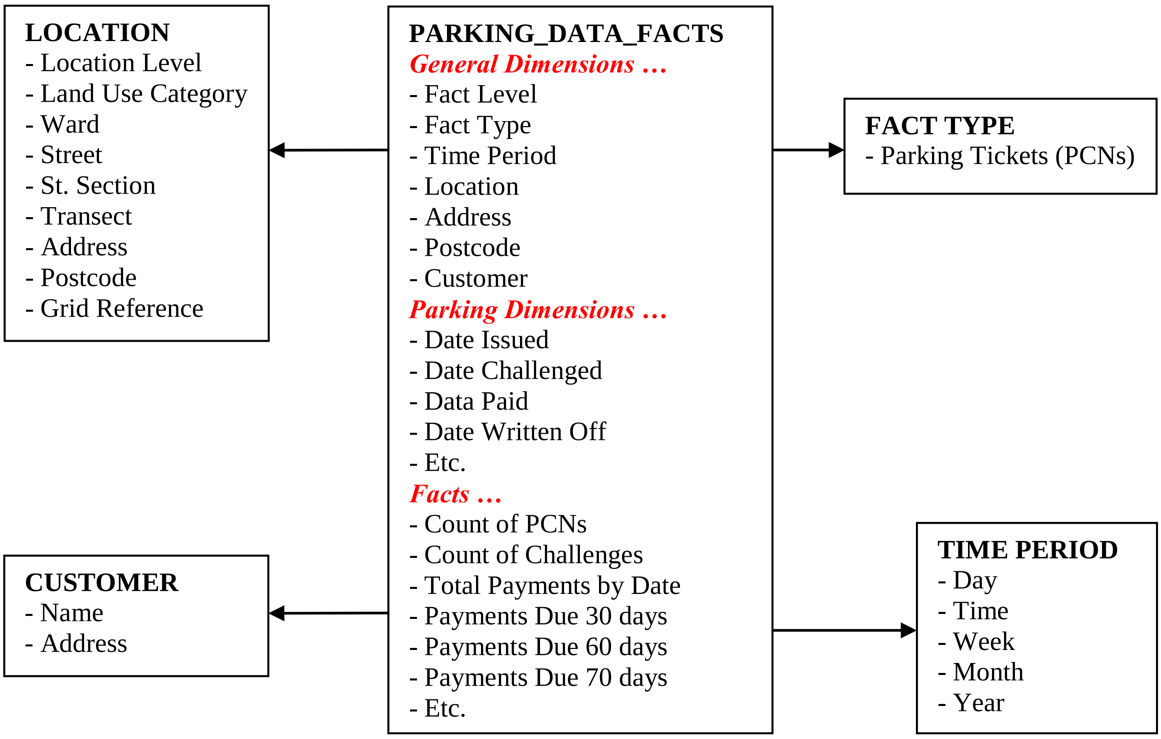
5.5 Templates

5.5.1 Data Mart for Parking Tickets

This diagram shows a Data Model for a Data Mart to hold data about Parking Tickets issued by a Local Authority in the UK.

It was produced in a Word document from early discussion with the End-User and was very helpful in establishing communication and a collaborative method of working.
End-users find to easier to understand and agree to this kind of Data Model than a formal ERD.
This approach is therefore recommended.

Each Fact is associated with a number of Dimensions.
The 'FACTS' Table contains the list of data items which is available.
The other Tables are called 'Dimensions' and define how the Facts can be analysed.



LOCATION
- Location Level
- Land Use Category
- Ward
- Street
- St. Section
- Transect
- Address
- Postcode
- Grid Reference

PARKING_DATA_FACTS

General Dimensions …
- Fact Level
- Fact Type
- Time Period
- Location
- Address
- Postcode
- Customer

Parking Dimensions …
- Date Issued
- Date Challenged
- Data Paid
- Date Written Off
- Etc.

Facts …
- Count of PCNs
- Count of Challenges
- Total Payments by Date
- Payments Due 30 days
- Payments Due 60 days
- Payments Due 70 days
- Etc.

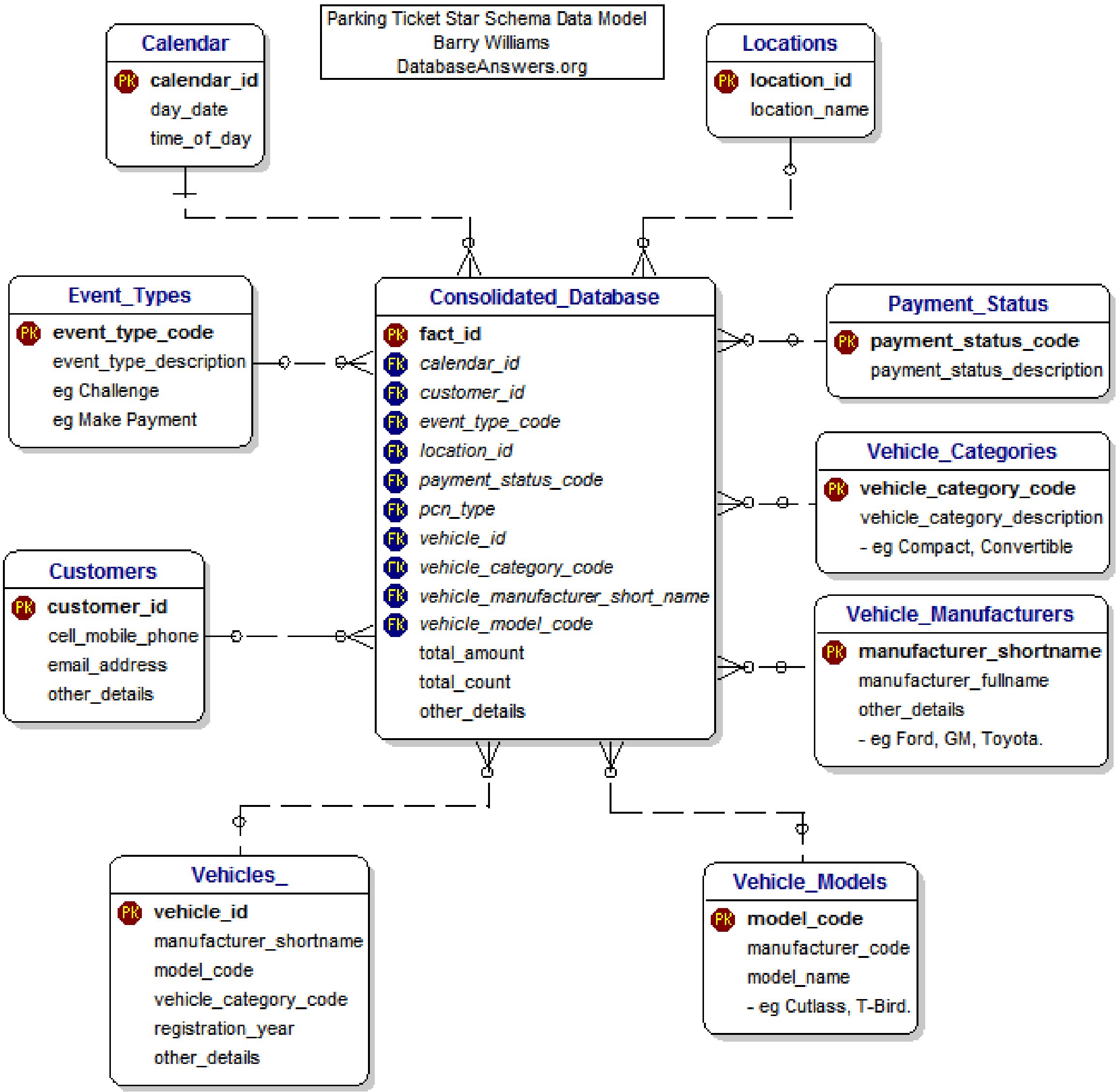
FACT TYPE
- Parking Tickets (PCNs)

CUSTOMER
- Name
- Address

TIME PERIOD
- Day
- Time
- Week
- Month
- Year

5.5.2 Data Mart as a Data Model

This diagram was produced by a Data Modelling Tool and is the complete analysis of all the data required.



5.6 FAQs

FAQ.1 How do I design a Data Mart ?

The first step is to think about the Data Mart as a place where you simply throw all available data and provide 'hooks' so that any combination of data can easily be retrieved.
Briefly, the Key fields in Tables involved become Dimensions in a Data Mart.
Facts include all the basic data plus any derived data, typically averages, percentages and totals under various headings.

FAQ.2 What are the Qualities for Success in designing Data Mart ?

To be successful in designing Data Marts it is important to have a talent for visualizing the User's Requirements and for translating this to a formal design of Dimensions and Facts, together with the most important aspect, which is the derivation of the data required from the underlying basic data.

FAQ.3 How do I improve the performance of my Data Mart ?

Every DBMS produces what is called an Execution Plan for every SELECT statement.
The steps to improving the performance involve checking this Execution Plan against the Indexes that exist, and making sure that the Query Optimizer has used the appropriate Indexes to obtain the best performance.
This is a specialized area where DBA's spend a lot of their time when they are looking after production databases where speed is a mission-critical factor.

Data Marts are always created to support Business Intelligence, which includes Performance Reports, Balanced Scorecards, Dashboards, Key Performance Indicators and so on.
Best practice always requires user involvement and a generic design to support a flexible approach to meeting changing requirements.
Users will always want changes to their first specifications of their requirements.
The insight that they obtain from the first Reports helps them identify more precisely what their long-term requirements will be.
Therefore flexibility is important.
A well-designed Data Mart will anticipate the areas where flexibility is required.
The design process should always follow two steps :-

- Production of generic design for the Data Mart
- Implementation of the design with a specific Data Mart software product.

6. Data Warehouse

6.1 What is it ?

I think of a Data Warehouse as a repository of centralised data from multiple source which has been transformed to be consistent with an Enterprise Data Model.
It is commonly used to provide a 'Single View of Corporate Data'

Wikipedia has a useful entry for a Data Warehouse at this page :-

- https://en.wikipedia.org/wiki/Data_Warehouse

It contains the following description :-

- A **data warehouse** is a used for reporting and data analysis, and is considered a core component of business intelligence which stores integrated data from one or more disparate sources. They store current and historical data and are used for creating analytical reports for knowledge workers throughout the enterprise. Examples of reports could range from annual and quarterly comparisons and trends to detailed daily sales analysis.
- The data stored in the warehouse is uploaded from the operational systems (such as marketing or sales). The data may pass through an operational data store for additional operations before it is used in the DW for reporting.

6.2 Master Data Management

Wikipedia has a useful entry for Master Data Management (MDM) at this page :-

- https://en.wikipedia.org/wiki/Master_data_management

which says :-

- "The data that is mastered may include:

reference data – the business objects for transactions, and the dimensions for reports and analysis

MDM is a part of the 'Single View of the Truth'.
It provides a solution to known problems and many major players, such as Informatica, offer commercially available solutions.

In a Blog by Erik Haahr posted on March 9th. 2017, he stated that the Gartner Group suggests four different styles of MDM Hub implementations :-

- (Gartner http://www.gartner.com/technology/home.jsp)

Here are the four Styles :-

1) Registry Style
The MDM Hub is the centre of Reference but stores only an Index used to retrieve master data from relevant backend systems
2) Centralised Style
The MDM Hub updates necessary in all backend system and is both a system of reference and a system of entry.
3) Coexistence Style

The MDM Hub stores all Master Data. The data entry procedure updates the Master data which then updates all the backend systems.

4) Consolidation Style
   The MDM Hub is the system of reference for reporting purposes and stores all master data so there is no need to get data from backend systems.

   My personal preference is the Registry Style because it is simple, neat and elegant and I have used it with great success.
   Here is a link to my Data Model for a Customer Master Registry :-
   - http://www.databaseanswers.org/data_models/customer_master_index/index.htm

7. Big Data (Lake)

7.1 What is it ?

Wikipedia has a useful entry for Data Lakes at this page :-

- https://en.wikipedia.org/wiki/Data_lakes

where it states :-
"A **data lake** is a method of storing data within a system or repository, in its natural format,[1] that facilitates the collocation of data in various schemata and structural forms, usually object blobs or files."
Two important features are

1. data stored 'in its natural form" – in other words, no data transformation
2. "collocation of data in various schemata" – in other words, data is simply stored in the format it arrives in.

I like the phrase 'Data Lake' because it is a simple and convenient way to describe something that would otherwise be difficult and complex to describe.

8. Data Mining

8.1 What is it ?

Wikipedia has a useful entry for Data Mining at this page :-

- https://simple.wikipedia.org/wiki/Data_mining

It states (in summary) :-
Data Mining is about finding new information in a lot of data with the aim of finding data that is both new and useful.

In many cases, data is stored so it can be used later. The data is saved with a goal. For example, a store wants to save what has been bought. They want to do this to know how much they should buy themselves, to have enough to sell later.

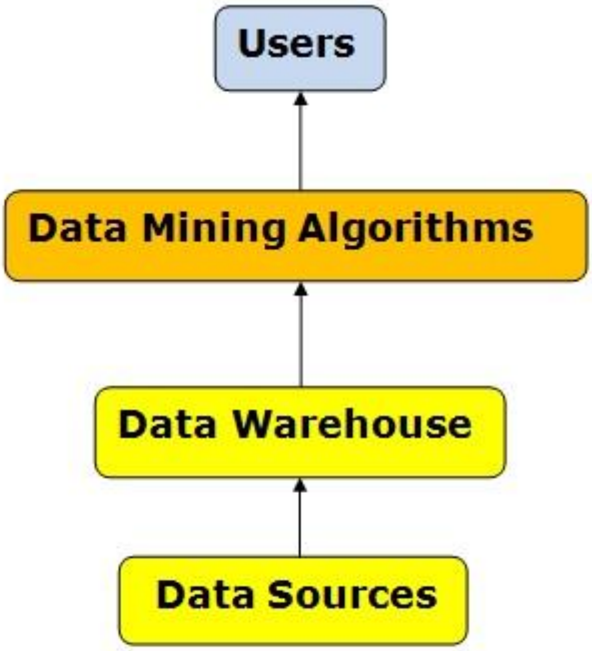
Saving this information, makes a lot of data.

The data is usually saved in a database. The reason why data is saved is called the first use.

We have Data Models on this page :-

- http://www.databaseanswers.org/data_models/data_mining/index.htm

and the Conceptual Model looks like this :-



9. Data Modelling Theory

9.1 Inheritance

Wikipedia has a useful entry for Inheritance at this page :-

- https://en.wikipedia.org/wiki/Inheritance_(object-oriented_programming)

In our Database Answers Web Site we have several Inheritance-related Data Models which provide valuable insight into the theory and practice of Inheritance, including these :-

- Aircraft –
  - http://www.databaseanswers.org/data_models/aircraft_and_inheritance/index.htm
- City Tourist Guide -
  - http://www.databaseanswers.org/data_models/city_tourist_guide/index.htm
- Insurance and eClaims –
  - http://www.databaseanswers.org/data_models/insurance_and_eclaims/index.htm
- Retail Customers –
  - http://www.databaseanswers.org/data_models/retail_customers/retail_customers_and_inheritance.htm
- Roles, Inheritance and Sub-types :-
  - http://www.databaseanswers.org/data_models/roles_inheritance_and_subtypes/index.htm
- Union Grievances –
  - http://www.databaseanswers.org/data_models/union_grievances/union_grievances_inheritance_model.htm
- Vehicle Maintenance -
  - http://www.databaseanswers.org/data_models/vehicle_maintenance_with_inheritance/index.htm

On this page, we have a detailed discussion :-

- Design Notes –
  - http://www.databaseanswers.org/inheritance_design_notes.htm

10. Data Vault

10.1 What is it ?

Wikipedia has a useful entry for Data Vault Modelling at this page :-

- https://simple.wikipedia.org/wiki/Data_vault_modeling

It states (in summary) :-

“**Data vault modeling** is a database modeling method to preserve different sets of historical data from different sources. It is also a method of looking at historical data that deals with issues such as auditing

Data Vault Modeling focuses on several things. First, it emphasizes the need to trace where all the data in the database came from. Each row has extra attributes that describe where the data came from, and at what time it was loaded. This feature lets auditors find the source of the values.
It is an approach developed by Dan Linstedt.

11. Commercial Data Integration Products

11.1 What is it ?

The process of organising and managing data from different sources.
It typically involves cleaning-up and transforming data to a standard format for subsequent processing.
This can typically be part of a Data Integration activity.

This is an introduction to various Commercial Platforms that provide some enterprise Data Integration facilities.

11.2 Why is it important ?

It is important because a common requirement is to identify multiple data sources.

11.3 What will I learn ?

You will learn how to identify multiples sources and formats in order to identify how to convert

11.4 Some Commercial Products

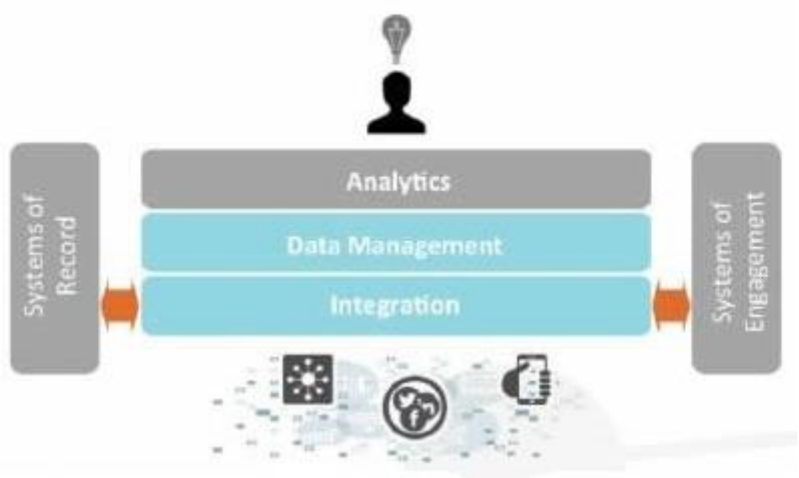
11.4.1 Liaison Alloy Platform

Here are some quotes from the Liaison Web Site :-

“Conceived from the ground up to address today’s technology disruptors, ALLOY is a next generation cloud platform for solving today’s integration and data management challenges.

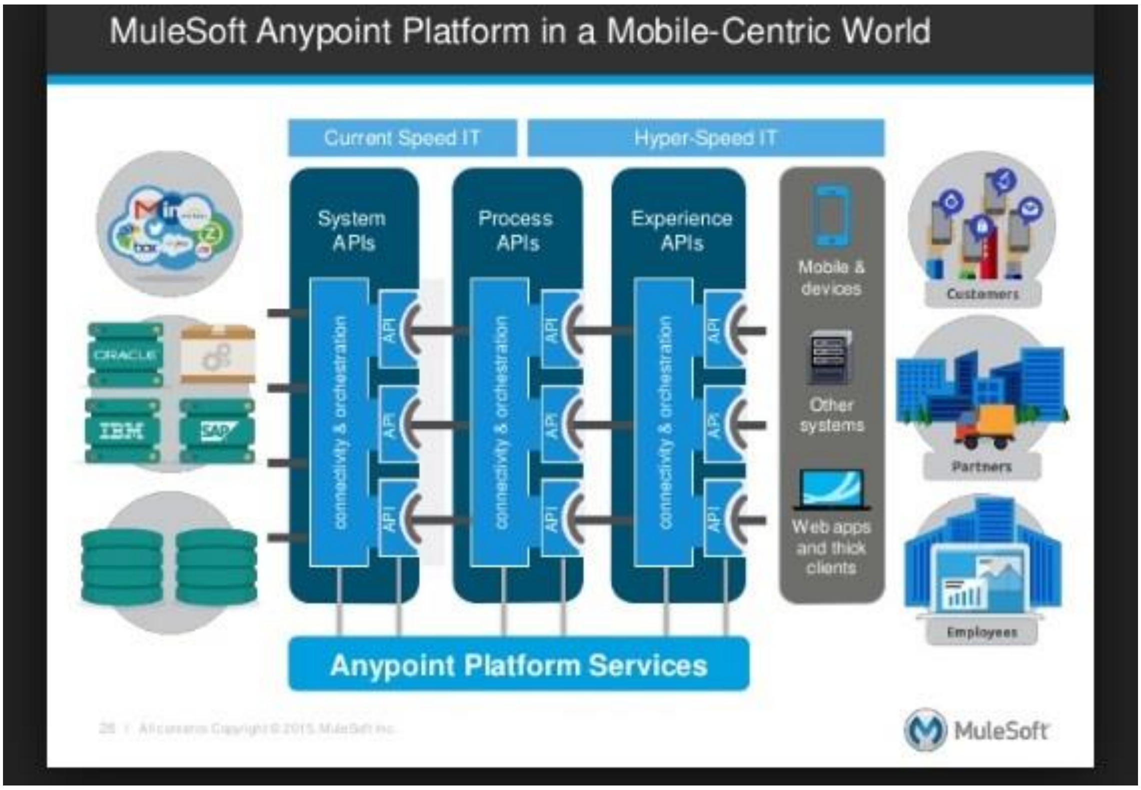
- **ALLOY provides** unified integration and data management capabilities as managed services, buffering the complexities of increasing data volume and variety
- **ALLOY connects** any two application end points: cloud, mobile, device, on-premises, etc.
- **ALLOY persists** data in a big data repository, providing on-demand, self-service access to clean, quality data
- **ALLOY provides** built-in security and compliance
- **ALLOY is an efficient alternative** to DIY integration models such as ESB or iPaaS at a time when connections are growing exponentially

- https://www.liaison.com/liaison-alloy-platform
- http://www.idevnews.com/stories/6515/Liaison-Alloy-Platform-Redefines-Integration-and-Data-Management



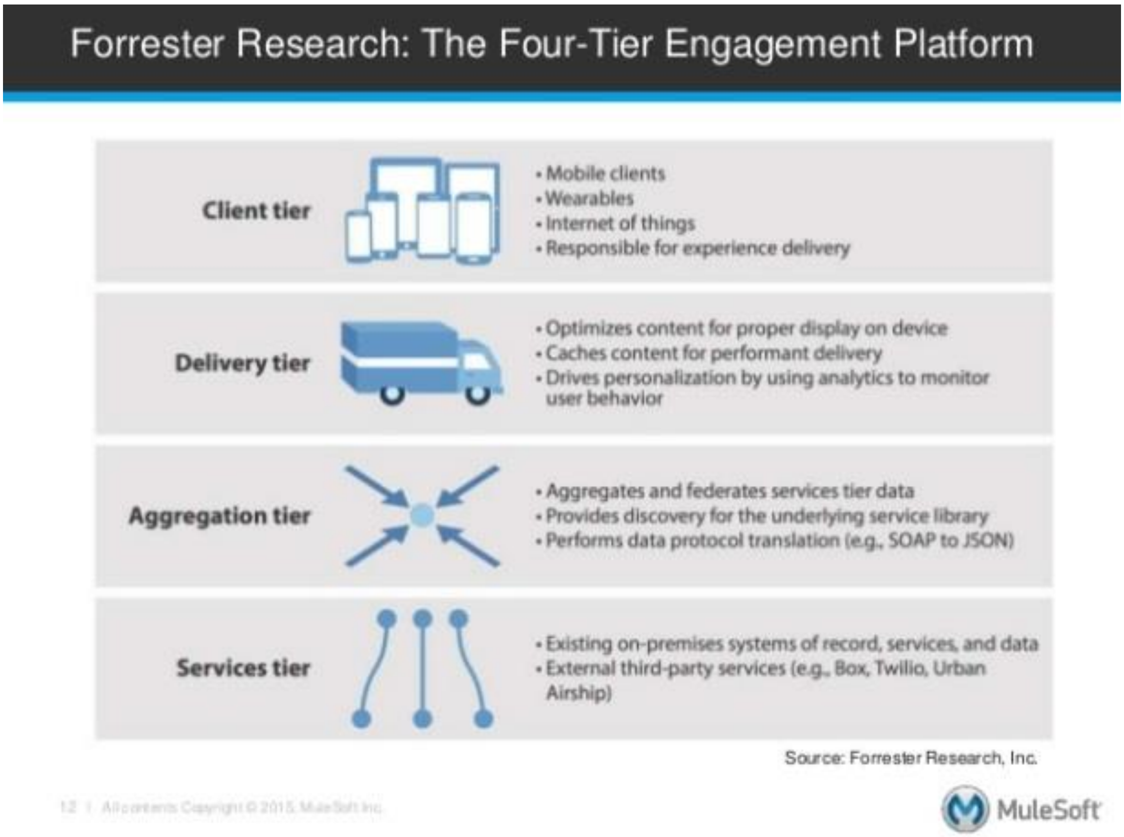
11.4.2 Mulesoft

11.4.2.1 Mulesoft's Anypoint Platform


MuleSoft Anypoint Platform in a Mobile-Centric World

26 | All contents Copyright © 2015, MuleSoft Inc.

11.4.2.2 Mulesoft and Forrester



Forrester Research: The Four-Tier Engagement Platform

Client tier

- Mobile clients
- Wearables
- Internet of things
- Responsible for experience delivery

Delivery tier

- Optimizes content for proper display on device
- Caches content for performant delivery
- Drives personalization by using analytics to monitor user behavior

Aggregation tier

- Aggregates and federates services tier data
- Provides discovery for the underlying service library
- Performs data protocol translation (e.g., SOAP to JSON)

Services tier

- Existing on-premises systems of record, services, and data
- External third-party services (e.g., Box, Twilio, Urban Airship)

Source: Forrester Research, Inc.

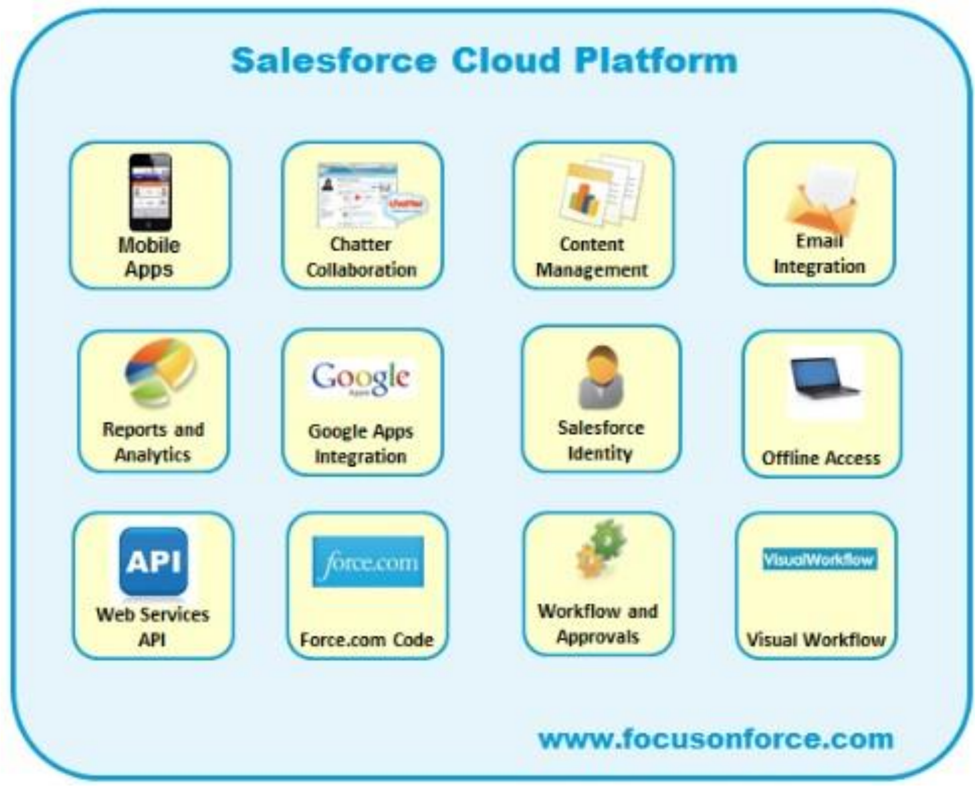
12 | All contents Copyright © 2015 MuleSoft Inc.

MuleSoft

11.4.3 Salesforce

11.4.3.1 Salesforce Cloud Platform

- http://focusonforce.com/platform/salesforce-platform-overview/

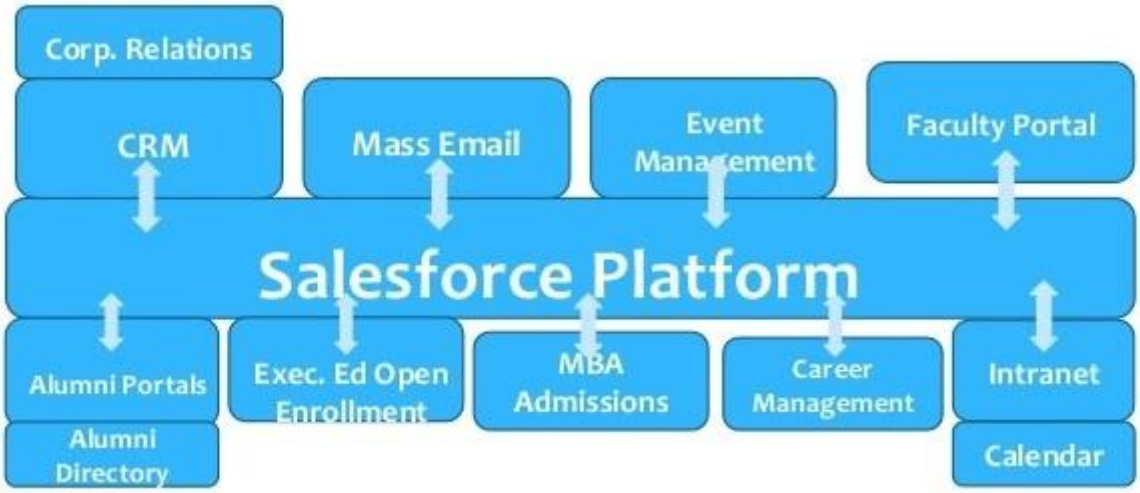


11.4.3.2 Salesforce and Events

On this page :-

- https://www.slideshare.net/salesforcefoundation/georgetown-university-and-st-norbert-college-improving-recruiting-efficiency-webinar

We like this slide because it combines the words Platform and event.



11.4.4 SAP and Google (Kronva)

They have a Netweaver Platform
On this page :-

- http://www.kronva.com/

11.4.5 Software AG

Digital Business Platform for SAP :-

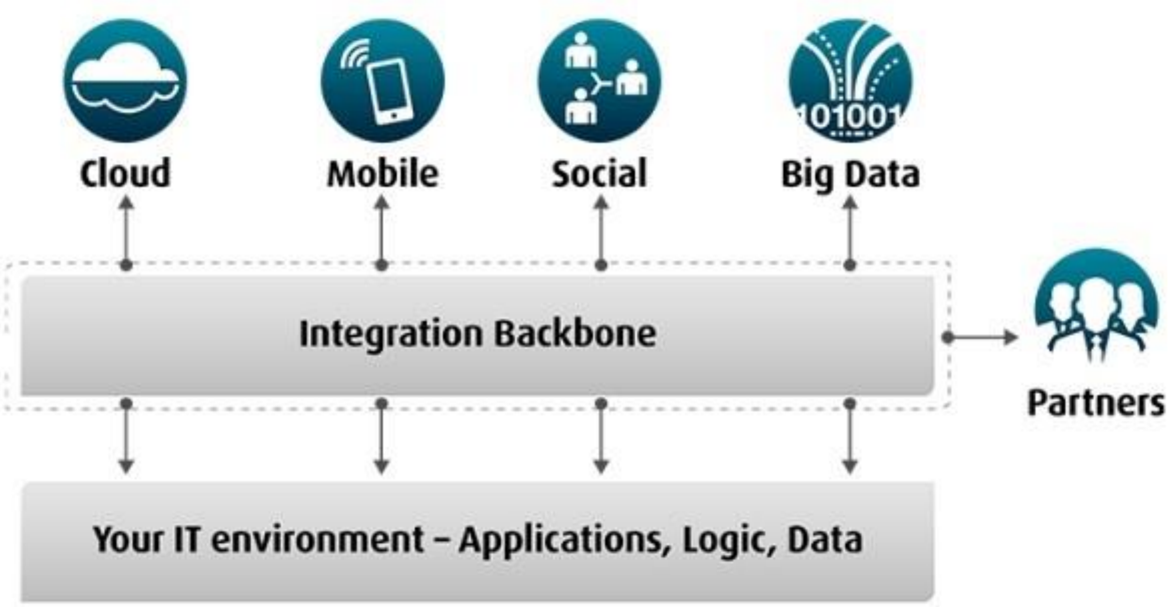
- https://marketplace.softwareag.com/apps/48105#!features/SAP_process_design

Claims and Policy Management for Insurance :-

- https://marketplace.softwareag.com/apps/37895#!overview

Here is their DBP Integration Platform or Webmethods Integration Platform
On this page :-

- http://www2.softwareag.com/corporate/products/webmethods_integration/integration/default.aspx



12. Data Sources

12.1 What is it ?

The process of organising and managing data from different sources.
It typically involves cleaning-up and transforming data to a standard format for subsequent processing.
This can typically be part of a Data Integration activity.

12.2 Why is it important ?

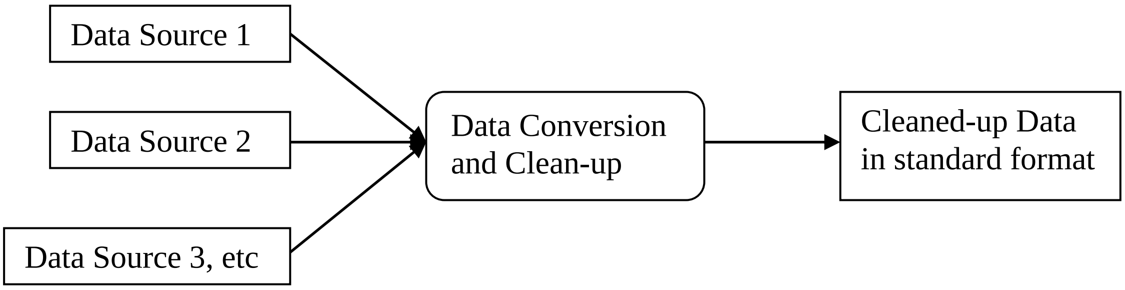
It is important because a common requirement is to identify multiple data sources.

12.3 What will I learn ?

You will learn how to identify multiples sources and formats in order to identify how to convert them to a common format for subsequent processing.

12.4 Best Practice

Best Practice looks like this :-


Data Source 1
Data Source 2
Data Source 3, etc
Data Conversion and Clean-up
Cleaned-up Data in standard format

12.5 Templates

The Templates make it possible to record the details of all the data formats, sources and data stewards.

12.6 FAQs

This Wikipedia entry is a useful introduction :-

- https://en.wikipedia.org/wiki/Operational_data_store

13. Information Catalogue

13.1 What is it ?

An Information Catalogue is a Repository of Information related to Information systems.

13.2 Why is it important ?

It is important because it provides a single point of reference and consistent definitions for data items, such as 'What is a Customer' – is it somebody who has actually purchased or simply made an enquiry ?

13.3 What will I learn ?

You will learn how to design and build an Information Catalogue.

13.4 Best Practice

You can get started on this page of our Database Answers Web Site that list some commercially available Data Dictionaries :-

- http://www.databaseanswers.org/data_dictionaries.htm

Here is the page listing our Data Models for Data Dictionaries :-

- http://www.databaseanswers.org/data_models/data_dictionary/index.htm

This page lists some very interesting Questions and Answers about Data Dictionary :-

- http://www.databaseanswers.org/data_models/data_dictionary/facts.htm

13.5 Templates

A suitable Template for getting started is a simple table.
An example for a Template for Entries in a Glossary is show on this page :-

- http://www.databaseanswers.org/template_for_new_style_pages.htm

13.6 FAQs

FAQ.1 How do I publish an Information Catalogue ?

It is good to start with a Spreadsheet and then move to a stand-alone Access Database before finally migrating to an Internet-based Database which is published on a Web Site.

FAQ.1 How can I be sure of Success with an Information Catalogue ?

To be successful in maintaining and publishing an Information Catalogue it is beneficial to enjoy detail and to have an interest in ensuring that all interested parties are on the same Page.

It is also useful to have an eye for detail and to have an appreciation for the way in which the separate Components within the Information Catalogue are interrelated.

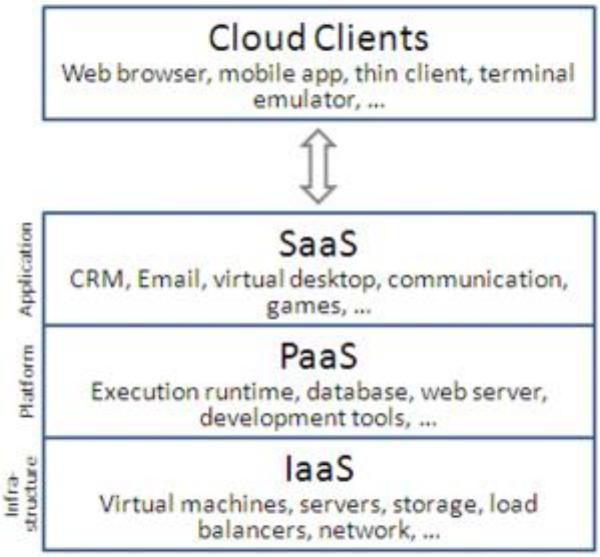
14. Cloud Services

14.1 What is it ?

Cloud Services can be defined as a conceptual IT Delivery system with no concern of hardware, software or operating system.
Wikipedia does not offer an entry for Cloud Services.
The nearest it suggests is Cloud Computing Service Models :-

- https://en.wikipedia.org/wiki/Cloud_computing#Service_models

It includes this diagram, which is describes as :-
“Cloud computing service models arranged as layers in a stack” :-


Cloud Clients
Web browser, mobile app, thin client, terminal emulator, ...
Application
SaaS
CRM, Email, virtual desktop, communication, games, ...
Platform
PaaS
Execution runtime, database, web server, development tools, ...
Infra-structure
IaaS
Virtual machines, servers, storage, load balancers, network, ...

We define Cloud Services as a “Conceptual IT Delivery system with no concern of hardware, software or operating system”.

In other words, a ‘User-eyes Business View such as Banking, Insurance, Retail and Travel.

- Event-Driven Platform http://www.databaseanswers.org/data_models/event_driven_platform/index.htm

14.2 Why is it important ?

It is important because it represents a convenient way to think about the user and their interaction with IT Services.
In addition, we can combine this with a Model-View-Controller which is a well-established Application Architecture.
This provides us with a very powerful approach to discussing Cloud Services.

14.3 What will I learn ?

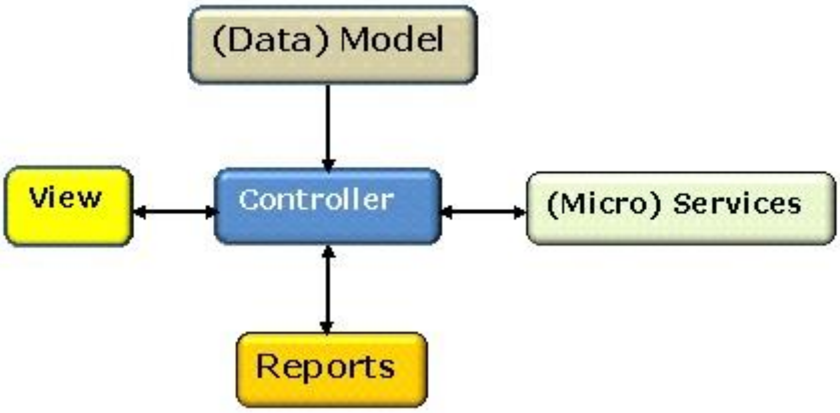
You will learn about a Model-View-Controller and how it can be the foundation for a Cloud Services Architecture.

14.4 Best Practice

This provides us with a very powerful that we show on our Database Answers Web site :-

- http://www.databaseanswers.org/data_models/mvc_model_view_controller/index.htm

which looks like this :-



15. Integration Platform

15.1 What is it ?

Data Integration is the process of producing one stream of related data from a number of streams of data from different sources.
For example, Customer data can be obtained from retail purchases and telephone bills.
The key activity involves matching data for the same Customer from different streams.
In the case of Customers, this can be achieved by matching facts about a Customer such as names, addresses, gender and data of birth.

15.2 Why is it important ?

It is it important because it provides a 'Single View of the Truth'

15.3 What will I learn ?

You will learn how to match data using appropriate external standards.
For external, in the UK, the Government maintains a so-called 'Post Office Address Format' or PAF-File and commercial software is available to simplify the matching process.

15.4 Best Practice

This page on our Database Answers Web Site is a good starting point :-

- http://www.databaseanswers.org/enterprise_data_integration.htm

15.5 Templates

This page on the IBM Web Site is a useful introduction to the more general topic of Master Data Management :-

- http://www.ibm.com/analytics/us/en/technology/master-data-management/

15.6 FAQs

This page has some very useful links to help you get started on Customer Data Integration:-

- http://www.databaseanswers.org/customer_data_integration.htm

Here is a useful set of links for commercial De-Duping software :-

- http://www.databaseanswers.org/deduping.htm

15.7 ETL – Extract, Transform and Load

ETL is a very important component of the Integration Platform.
This Wikipedia entry is a very useful introduction:-

- https://en.wikipedia.org/wiki/Extract,_transform,_load

The function of ETL is to take data from multiple sources, clean it up and transform it so that it can be loaded into a Data Warehouse.
It might include providing a 'Single View of the Truth', so that, for example, a Customer called John could be recognised as Johnny, Jon or Jonno.
It is common to find Libraries of Transformation Utilities being used that reflect corporate standards, such as closing Dates for Sales Orders.

Appendix A. Validation of our Referance Data Architecture

My Data Architecture is very important and provides the foundation for my work.
I decided to validate it using some of my Library of more than 1.500 Data Models.
I chose 20 of my favourites and was very pleased with the results which I have documented on this Page of my Database Answers Web Site :-

- http://www.databaseanswers.org/data_models/POC_Cloud_Services.htm

The 20 Models I chose are as follows :-

1 Air Transport
2 Banking – Investment
3 Banking - Self-Service (Retail)
4 Doctors and Patients
5 Gym Training Diary
6 HR Self-Service
7 Insurance Self-Service
8 Law Enforcement
9 Local Government
10 Logistics
11 Olympic Games
12 Pharmaceutical Companies
13 Phone Bills
14 Restaurant Guides
15 Retail Sales
16 Student Self-Service Registration
17 Telecommunications
18 Utilities
19 Waste Management
20 Wine Stores

Appendix B. Templates

I would be pleased to have your suggestions for improving this book, based on your experience and thoughts.

Please use this Template and email it to me at

- barryw@databaseanswers.org .

This will qualify you for automatic inclusion in my club of potential Partners.

Appendix C. Industry Platforms

C.1 What is it ?

I define a Data Platform as a series of Data Layers that establish an Architectural component for a common theme such as Banking, Insurance or Retail.
We have included a Wikipedia reference here because they are reliable and helpful :-

- https://en.wikipedia.org/wiki/Personalized_marketing#DMPBig

Under the heading of 'Personalised marketing' Wikipedia makes this statement :-'A data management platform (DMP) is a centralized computing system for collecting, integrating and managing large sets of structured and unstructured data from disparate sources."
Wikipedia gives a number of examples, including Oracle's BlueKai which Wikipedia defines as third-party data collecting company offering a cloud-based data platform to personalize online, offline, and mobile marketing campaigns.

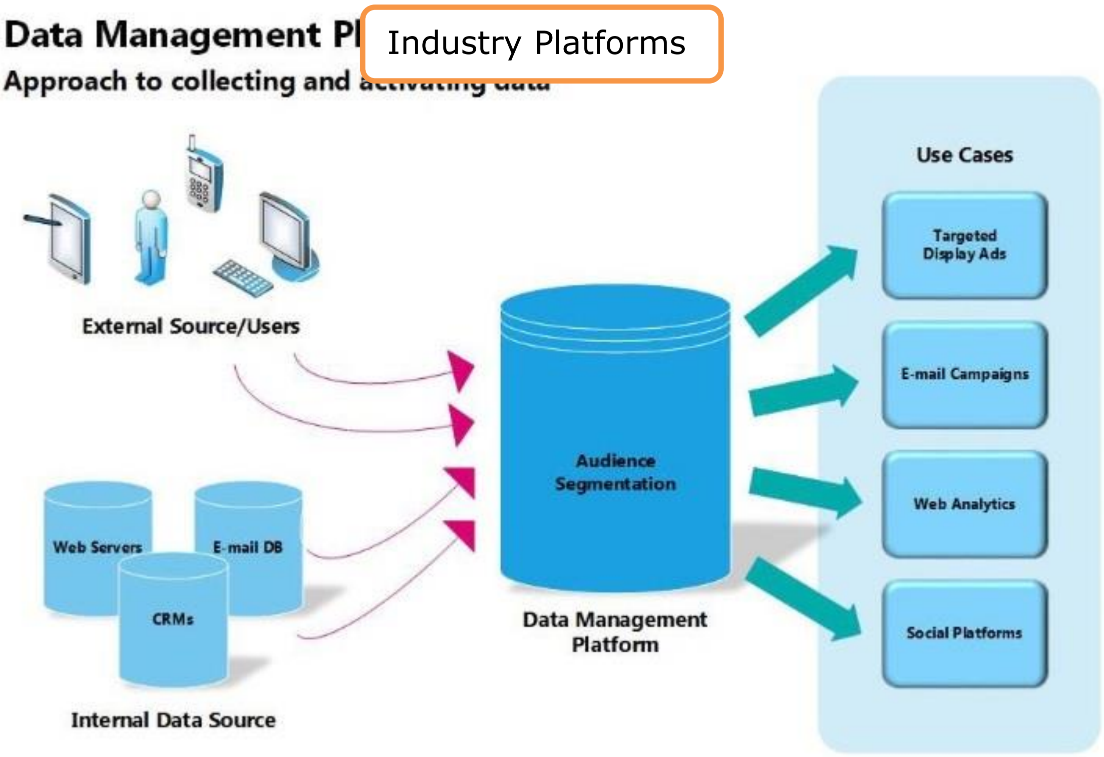
C.2 Why is it important ?

A Data Platform is important because it provides a coherent artefact based around data that has a common theme (eg Banking) and structure (ie Layers)

C.3 What will I learn ?

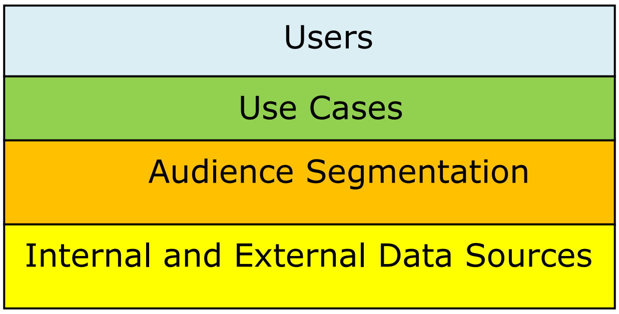
You will learn how to design a Data Management Platform.
We start with this Wikipedia diagram and then redefine it so it show a series of horizontal Layers.

C.3.1 Wikipedia diagram


Industry Platforms

C.3.2 Our version of the diagram

Here we show our layered version.



C.4 Best Practice

You can get started on this page of our Database Answers Web Site that list some commercially available Data Dictionaries :-

- http://www.databaseanswers.org/data_dictionaries.htm

Here is the page listing our Data Models for Data Dictionaries :-

- http://www.databaseanswers.org/data_models/data_dictionary/index.htm

This page lists some very interesting Questions and Answers about Data Dictionary :-

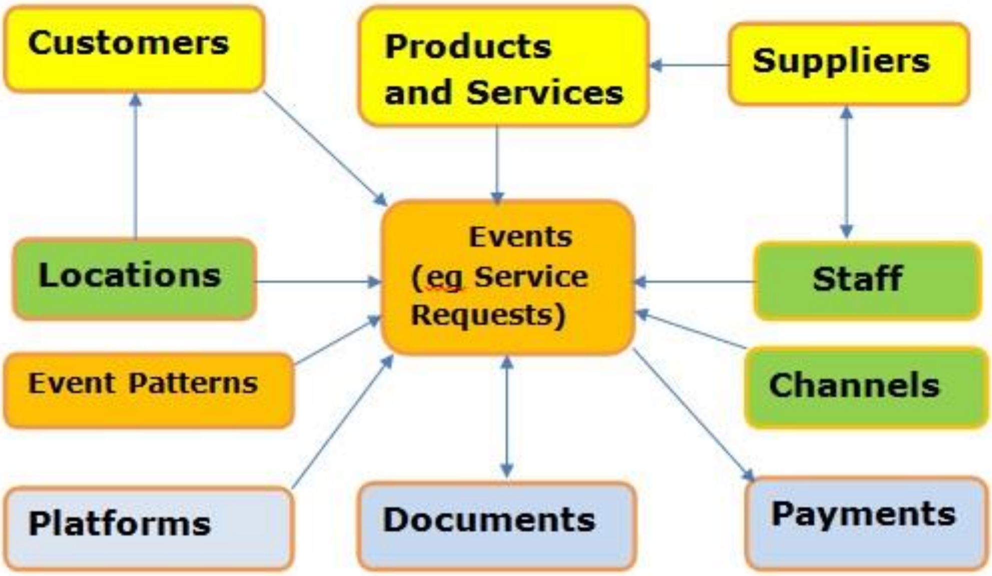
- http://www.databaseanswers.org/data_models/data_dictionary/facts.htm

C.5 Canonical Data Model

A very important element of an Industry Data Platform is a Canonical Data Model.
It is discussed in detail on this page :-

- http://www.databaseanswers.org/data_models/canonical_data_models/index.htm

and the Conceptual Data Model looks like this :-



As you can see, it is a very Event-oriented Model and we use it as the basis for all of our Industry-oriented Platforms.