# Assignment 2

1. You should identify dimensions, map dimensions to data sources, and specify dimension hierarchies. For each dimension, you should identify its data sources and attributes in each data source. For hierarchical dimensions, you should indicate the levels from broad to narrow.
2. The dimensions in the problem are Franchise, MemberType, ServiceCategory, Merchandise, Member and Calendar. Franchise, MemberType, ServiceCategory, Merchandise, Member come from the ERD. The calendar dimension is a standard warehouse dimension. Calendar is a hierarchical dimension.

* Franchise
* FranchId: ERD only
* FranchRegion: ERD only
* FranchPostalCode: ERD only
* FranchModelType: ERD only
* MemberType
* MemTypeId: ERD only
* MemTypeName: ERD only
* MemTypePrice: ERD only
* ServiceCategory
* ServCatId: ERD only
* ServCatName: ERD only
* ServCatPrice: ERD only
* Merchandise
* MerchId: ERD only
* MerchName: ERD only
* MerchPrice: ERD only
* MerchType: ERD only
* Member
* MmbrId: ERD only
* MmbrName: ERD only
* MmbrZip: ERD only
* MemTypeId: ERD only
* FranchId: ERD only
* MmbrEmail: ERD only
* Calendar
* Date columns in ERD (MmbrDate, ServPurchDate and SaleDate) and spreadsheet. Hierarchical (year → month → day)

1. You should specify measures, related data sources, and measure aggregation properties.
2. The measure mostly come from the MemberType, ServiceCategory, Merchandise and Contains. Measures from the related tables are important to associate with the MemberType, ServiceCategory, Merchandise and Contains table.

* MemTypePrice (MemberType): Semi Additive measure
* ServCatPrice (ServiceCategory): Semi Additive measure
* MerchPrice (Merchandise): Semi Additive measure
* Qty (Contains): Additive measure
* Amount (Worksheet): Additive measure

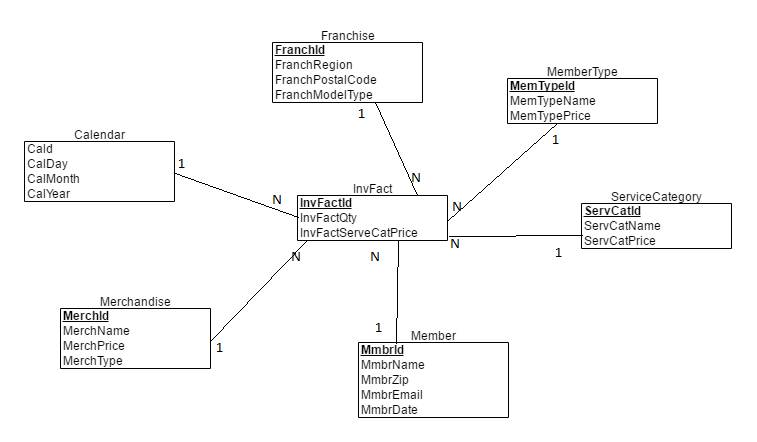
1. Identify the grain in your dimensional design using the business needs as a guideline. You should then indicate relative storage requirements for the grain using the statistics for the data sources. Using the cardinality estimates provided, you should determine either the fact table size or sparsity and then compute the unknown grain size variable. For example, you should compute sparsity if the fact table size is given.
2. The most detailed grain is the combination of individual Franchise, individual MemberType, individual ServiceCategory, individual Merchandise, individual Member and date.

* 650 Franchise: Sum of Franchise rows and special event worksheet.
* 20 ServiceCategory
* 500 Merchandise
* 50,150 Member: Sum of member row and unique customers per special event
* Days per year: 365
* 700,000 fact table: Sum of Sale row, Contains row and Service Purchase Row
* Sparsity estimate:

1- (fact table size / product of dimensions)

1-( 700,000/ (650\*20\*500\*50,150))= 0.999

1. Extend your analysis to design a star schema (or variation) to support inventory analysis. For each table, you should define the table name, primary key, and columns. You do not need to write complete CREATE TABLE statements.

4)

1. Identify summarizability potential problems in your star schema and indicate preferred resolutions of the summarizability problems. For incomplete dimension-fact relationships, you should also indicate if columns in a dimension table allow null values.
2. MmbrDate is an incomplete fact dimension relationship as the membership date is missing in the franchise spreadsheet. It is probably not possible to add to existing data but second data source possibly can be changed in future so membership date is collected on the spreadsheet. If membership data is same as franchise date, the same date can be used as a default value.

There are also missing values for MerchPrice and MerchType for the merchandise from the spreadsheet. Although the merchandise information is mandatory, these missing values make the relationship missing for the MerchPrice and MerchType columns. Additional data collection this incompleteness as no reliable default value exists.

1. You should populate your data warehouse tables based on the data in the sample tables and spreadsheet. You do not need to write SQL INSERT statements or insert data into database tables. You can just show table listings in your solution document. You should indicate mappings from data sources into tables. For example, a mapping may involve generating new primary key values for a data warehouse table or using a default value for a missing value.

|  |  |  |  |
| --- | --- | --- | --- |
| **Franchise** | | | |
| *FranchId* | *FranchRegion* | *FranchPostalCode* | *FranchModelType* |
| F1 | Northwest | 98011 | Full |
| F2 | Mountain | 80111 | Medium |
| F3 | Central | 45236 | Limited |

Sample data for the Franchise Table

|  |  |  |
| --- | --- | --- |
| **MemberType** | | |
| *MemTypeId* | *MemTypeName* | *MemTypePrice* |
| M1 | Platinum | $1,000 |
| M2 | Gold | $800 |
| M3 | Value | $300 |
|  |  |  |

Sample data for the MemberType Table

|  |  |  |
| --- | --- | --- |
| **ServiceCategory** | | |
| *ServCatId* | *ServCatName* | *ServCatPrice* |
| SC1 | Ball machine | $15 |
| SC2 | Private lesson | $75 |
| SC3 | Adult class | $150 |
| SC4 | Child class | $125 |

Sample data for the ServiceCategory

|  |  |  |  |
| --- | --- | --- | --- |
| **Merchandise** | | | |
| *MerchId* | *MerchName* | *MerchPrice* | *MerchType* |
| MC1 | Wilson balls | $3 | Balls |
| MC2 | Wilson racket | $200 | Racket |
| MC3 | Adidas shoes | $100 | Shoes |
| MC4 | Racket stringing | $40 | Racket |

Sample data for the Merchandise Table

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Member** | | | | | | |
| *MmbrId* | *MmbrName* | *MmbrZip* | *MemTypeId* | *MmbrDate* | *FranchId* | *MmbrEmail* |
| 1111 | Joe | 98011 | M1 | 1-Feb-2009 | F1 | joe@serv1.com |
| 2222 | Mary | 80112 | M2 | 1-Jan-2010 | F2 | mary@serv2.com |
| 3333 | Sue | 45327 | M3 | 3-Mar-2011 | F3 | sue@serv3.com |
| 4444 | George | 45236 |  |  | F3 | george@serv4.com |

Sample data for the Member Table

|  |  |  |
| --- | --- | --- |
| ***InvFact*** | | |
| *InvfactId* | *InvFactQty* | *InvFactServeCalPrice* |
| F1 | 2 | $15 |
| F2 | 1 | $75 |
| F3 | 1 | $150 |
| F1 | 1 | $125 |

Sample data for InvFact