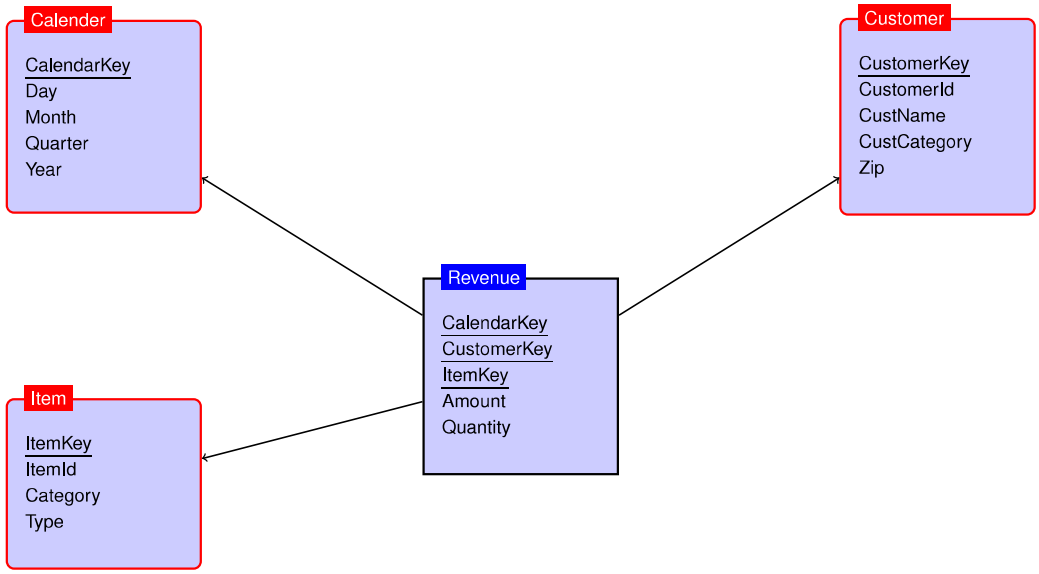
Q 1

SOLUTION: STAR SCHEMA



POPULATED TABLES

CALENDAR

CalendarKey	*Day*	*Month*	*Quarter*	*Year*
1	1	Jan	1	2004
2	2	Jan	1	2004
3	3	Jan	1	2004
4	4	Jan	1	2004
5	5	Jan	1	2004

ITEM

ItemKey	*ItemId*	*Category*	*Type*
1	M1	Memship	Platinum
2	M2	Memship	Gold
3	M3	Memship	Value
4	PSA	OneDayP.	Adult
5	PSS	OneDayP.	Senior
6	PSK	OneDayP.	Kid
7	AP1	Mrch.	T-Shirt
8	AP2	Mrch.	Hat
9	EQ1	Mrch.	Jump Rope
10	L-A	Spec. Evnt	All Day
11	L-H	Spec. Evnt	Half Day

REVENUE

CalendarKey	*CustKey*	*ItemKey*	*Amount*	*Quantity*
1	1	1	$1000	1
1	2	3	$300	1
1	3	3	$300	1
1	1	4	$20	1
1	3	4	$20	1
2	3	6	$3	1
1	3	7	$11	1
1	3	8	$9	1
2	2	8	$9	1
3	1	9	$36	3
4	4	10	$3500	1
5	5	11	$2200	1

CUSTOMER

CustKey	*CustId*	*CustName*	*CustCategory*	*Zip*
1	111	Joe	Ind	60611
2	222	Mary	Ind	60640
3	333	Sue	Ind	60611
4	CC1	Sears	Corp	60640
5	CC2	Boeing	Corp	60611

① See above. Note that this is not the unique answer.

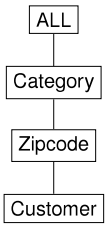
② There are several tasks involved when importing the data into the data warehouse. E.g., we need to *extract* zipcode information from $CorpCustNameLoc$; we need to perform aggregation $(price \cdot Quantity)$ for tuples in the $merchandise$ table; we might also need to deal with (near) duplicate object detection (e.g., the same "member" that appear in two data sources).

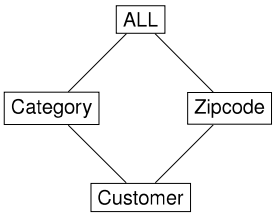
③ "Find the percentage of revenue generated by members in the last year" can be easily answered on the star schema by two aggregate queries on the fact table. Specifically, if the complete data cube has been built, the queries can be efficiently answered by the cuboid $(Year)$, and the cuboid $(Year, Category)$.

④ Since $CustName$ is not likely to be a good "level" for analysis (rather, it is a descriptive attribute), there are 4 levels on $Calendar$ dimension, 3 on $Item$, and 3 on $Customer$. Therefore, there are $(4+1)*(3+1)*(3+1) = 80$ in total.

Note that we could have different hierarchies on a dimension. E.g., we

could consider the hierarchy on the Customer dimension one of the following. They have different semantics, but do not affect the number of cuboids.



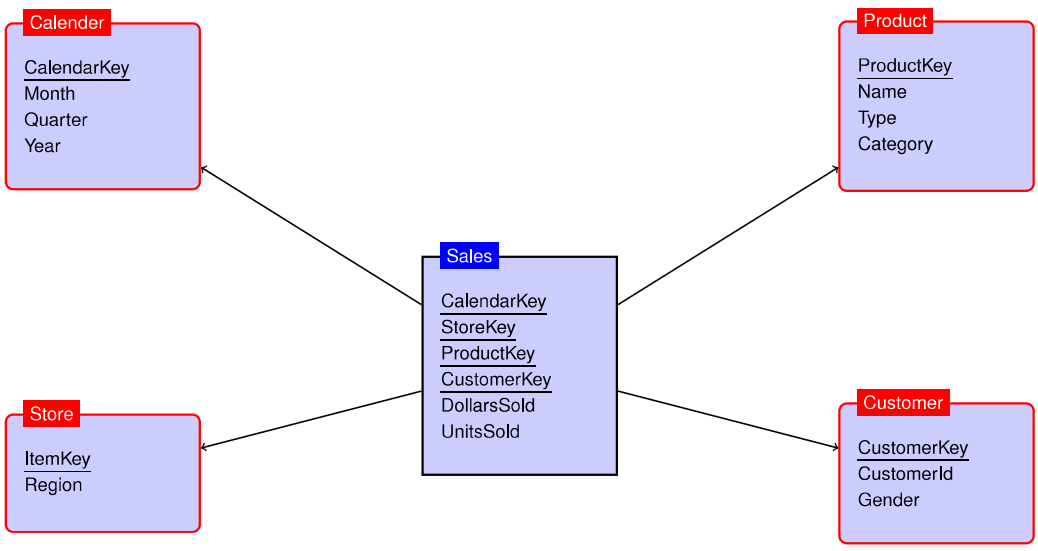


Solution: MDX Query

```
SELECT [Product].[Category].MEMBERS ON COLUMNS
       [Store].[USA].[CA].CHILDREN  ON ROWS
FROM   [Sales]
WHERE  ([Measures].[DollasSold])
```

```
SELECT [Product].[Category].MEMBERS ON COLUMNS
       [Store].[USA].[CA].CHILDREN  ON ROWS
FROM   [Sales]
WHERE  ([Time].[2007], [Measures].[DollasSold])
```


Solution: Star Schema



Populated Tables

CALENDAR

CalendarKey	*Month*	*Quarter*	*Year*
1	Jan	1	2003
2	Feb	1	2003

PRODUCT

ProKey	*ProName*	*ProType*	*Category*
1	Luvs 50	Diapers	Infant Care
2	Huggies 24	Diapers	Infant Care
3	High C	Vitamin	Dietary Supp

STORE

StoreKey	*Region*
1	East
2	Midwest

CUSTOMER

CustKey	*CustID*	*Gender*
1	12	Male
2	23	Male
3	34	Female

SALES (← Intermediate)

CalendarKey	*ProKey*	*StoreKey*	*CustKey*	*$Sold*	*UnitsSold*
1	1	1	1	$15	1
1	2	1	2	$20	1
1	2	1	1	$40	2
1	2	1	1	$20	1
1	2	1	1	$19	1
1	2	1	1	$19	1
2	3	2	1	$9	2
2	3	2	2	$9	1
2	3	2	3	$9	1

Populated Tables

SALES

CalendarKey	*ProKey*	*StoreKey*	*CustKey*	*$Sold*	*UnitsSold*
1	1	1	1	$15	1
1	2	1	2	$20	1
1	2	1	1	$98	5
2	3	2	1	$9	2
2	3	2	2	$9	1
2	3	2	3	$9	1