Data Analysis Fundamentals
Understanding of Data

Satyam Govila

customer_details.xlsx	25-Apr-2021 at 4:50 PM	37 KB	Micros...cument
customer_reviews.csv	15-Dec-2020 at 11:16 PM	1.5 MB	CSV Document
location_details.xlsx	25-Apr-2021 at 4:50 PM	42 KB	Micros...cument
order_details.xlsx	04-May-2021 at 8:56 PM	371 KB	Micros...cument
order_payments_new.xlsx	25-Apr-2021 at 4:50 PM	202 KB	Micros...cument
Orders_2015.xlsx	12-Jan-2021 at 7:48 PM	289 KB	Micros...cument
Orders_Cleaned.xlsx	18-Jan-2021 at 8:48 PM	1.7 MB	Micros...cument
Orders.xlsx	25-Apr-2021 at 4:50 PM	884 KB	Micros...cument
Returns.csv	25-Apr-2021 at 4:50 PM	6 KB	CSV Document
transactions_updated.xlsx	22-Apr-2021 at 11:10 PM	507 KB	Micros...cument
Vendors.csv	25-Apr-2021 at 4:50 PM	143 bytes	CSV Document

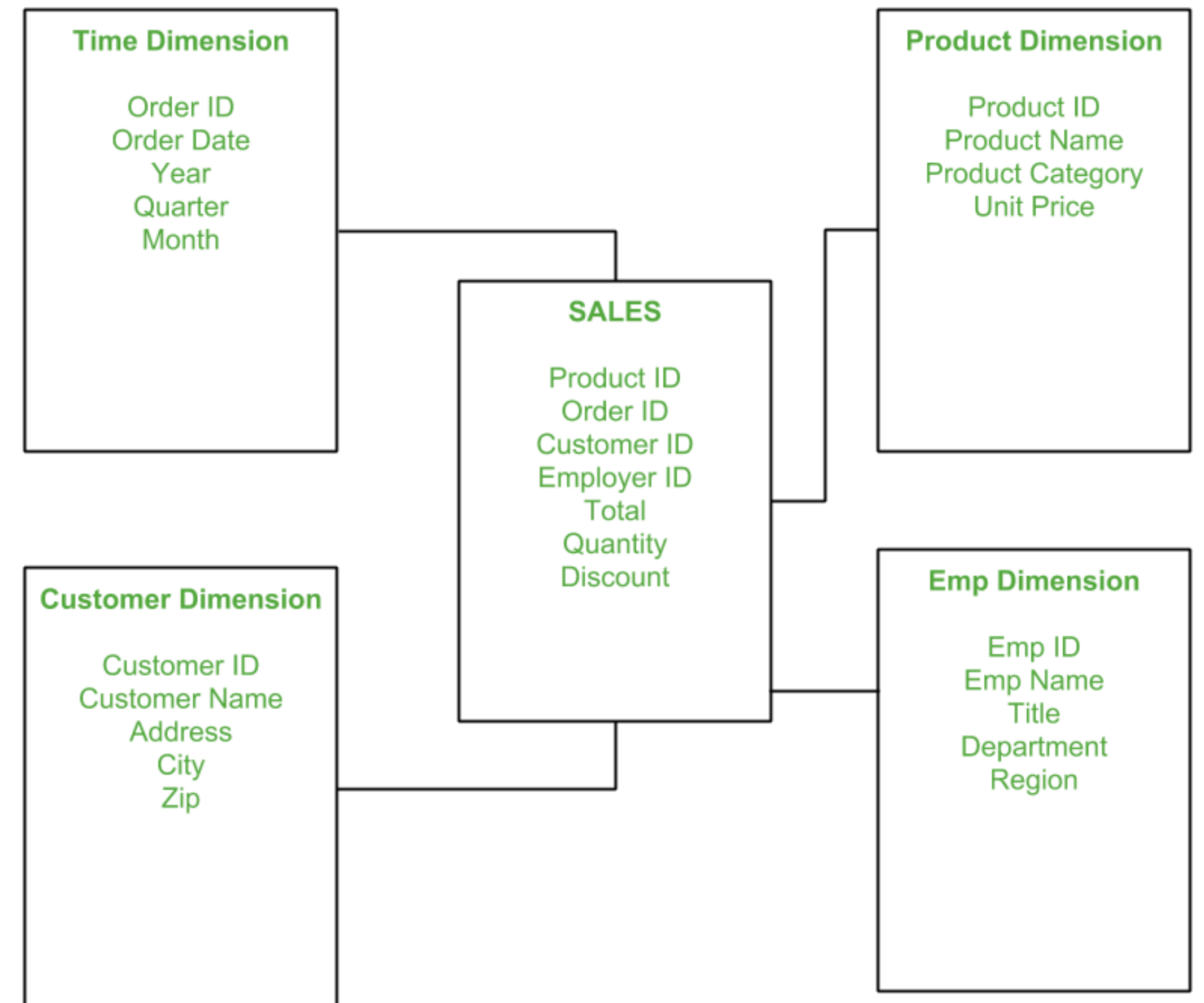| ID	order_id	order_status	order_purchase_date	order_approved_at	order_delivered_carrier_date	order_delivered_customer_date	order_estimated_delivery_date
1	CA-2016-152156	delivered	1/12/18 7:27 PM	1/12/18 7:37 PM	1/15/18 11:19 PM	1/17/18 10:08 PM	43131
2	CA-2016-138688	delivered	7/24/17 5:19 PM	7/24/17 5:30 PM	7/26/17 12:58 AM	8/8/17 3:07 PM	42958
3	US-2015-108966	delivered	3/7/18 11:46 PM	3/8/18 12:49 AM	3/8/18 8:12 PM	3/16/18 1:58 PM	43187
4	CA-2014-115812	delivered	6/5/18 11:36 AM	6/5/18 7:15 PM	6/6/18 2:59 PM	6/11/18 9:38 PM	43297
5	CA-2017-114412	delivered	2/15/18 11:27 PM	2/17/18 3:35 PM	2/19/18 11:53 PM	3/14/18 4:52 PM	43179
6	CA-2016-161389	delivered	6/29/17 12:10 AM	6/29/17 3:41 PM	6/29/17 3:41 PM	7/10/17 5:36 PM	42935
7	US-2015-118983	delivered	1/19/18 9:07 PM	1/20/18 9:09 AM	1/29/18 8:17 PM	2/20/18 7:27 PM	43153
8	CA-2014-105893	delivered	3/29/18 12:00 PM	3/29/18 12:48 PM	4/2/18 4:46 PM	4/9/18 7:26 PM	43207
9	CA-2014-167164	shipped	6/28/18 12:52 PM	6/28/18 1:11 PM	7/4/18 3:20 PM		43315
10	CA-2014-143336	delivered	3/15/18 8:23 PM	3/16/18 8:31 PM	3/20/18 11:42 PM	3/29/18 7:22 PM	43196
11	CA-2016-137330	delivered	3/27/18 10:14 PM	3/27/18 10:30 PM	3/29/18 1:04 AM	3/29/18 6:28 PM	43199
12	US-2017-156909	delivered	8/3/18 1:32 PM	8/4/18 5:05 AM	8/6/18 1:43 PM	8/14/18 12:28 AM	43335
13	CA-2015-106320	delivered	1/21/17 5:08 PM	1/21/17 5:22 PM	1/24/17 7:47 AM	1/27/17 6:12 PM	42796
14	CA-2016-121755	delivered	9/26/17 11:08 AM	9/26/17 11:24 AM	9/28/17 7:33 PM	10/3/17 6:35 PM	43031
15	US-2015-150630	shipped	2/28/18 8:57 AM	2/28/18 10:40 AM	3/5/18 4:10 PM		43182
16	CA-2017-107727	delivered	9/13/17 3:07 PM	9/13/17 3:25 PM	9/14/17 7:27 PM	9/22/17 4:48 PM	43012
17	CA-2016-117590	delivered	8/9/18 8:36 PM	8/9/18 8:50 PM	8/15/18 6:23 AM	8/25/18 12:46 AM	43348
18	CA-2015-117415	delivered	12/7/17 9:40 AM	12/7/17 10:31 AM	12/7/17 8:21 PM	12/11/17 6:52 PM	43102
19	CA-2017-120999	delivered	5/30/18 2:41 AM	5/30/18 3:24 AM	5/30/18 1:39 PM	6/22/18 10:41 PM	43299
20	CA-2016-101343	delivered	6/27/17 10:45 AM	6/27/17 11:04 AM	6/27/17 2:56 PM	6/29/17 4:45 PM	42926
21	CA-2017-139619	delivered	12/8/17 4:45 AM	12/12/17 3:50 AM	12/12/17 5:43 PM	1/9/18 6:04 PM	43103
22	CA-2017-118255	delivered	3/9/18 3:09 PM	3/9/18 5:00 PM	3/12/18 6:42 PM	3/26/18 9:07 PM	43193
23	CA-2014-146703	delivered	9/11/17 11:50 AM	9/11/17 12:05 PM	9/14/17 8:57 PM	9/19/17 9:47 PM	43012
24	CA-2016-169194	delivered	6/12/18 9:52 AM	6/12/18 10:21 AM	6/12/18 3:31 PM	6/19/18 5:57 PM	43293
25	CA-2015-115742	delivered	7/3/18 1:17 AM	7/3/18 1:30 AM	7/3/18 1:58 PM	7/11/18 10:52 PM	43320
26	CA-2016-105816	delivered	5/23/17 3:28 PM	5/24/17 3:35 PM	6/1/17 4:27 PM	6/12/17 4:48 PM	42909
27	CA-2016-111682	delivered	11/29/17 6:27 PM	11/29/17 7:31 PM	11/30/17 10:32 PM	12/11/17 9:44 PM	43091
28	CA-2015-135545	delivered	1/14/18 6:49 PM	1/14/18 6:59 PM	1/16/18 8:15 PM	1/20/18 2:28 PM	43130
29	US-2015-164175	delivered	8/23/17 3:40 PM	8/23/17 5:33 PM	8/24/17 4:25 PM	8/29/17 9:36 PM	42991
30	CA-2014-106376	delivered	9/10/17 10:32 PM	9/12/17 4:50 AM	9/13/17 6:26 PM	9/18/17 6:48 PM	43006
31	CA-2016-119823	delivered	8/21/18 4:45 PM	8/21/18 5:08 PM	8/22/18 3:18 PM	8/23/18 9:03 PM	43339
32	CA-2016-106075	delivered	5/27/18 5:59 PM	5/27/18 6:14 PM	6/6/18 4:10 PM	6/12/18 8:48 PM	43284
33	CA-2017-114440	delivered	5/20/17 2:48 AM	5/23/17 3:50 AM	6/5/17 1:58 PM	6/16/17 4:52 PM	42895
34	US-2015-134026	processing	10/5/16 10:44 PM	10/6/16 3:51 PM			42713
35	US-2017-118038	delivered	7/28/17 11:25 AM	7/28/17 11:35 AM	7/28/17 7:58 PM	8/7/17 8:41 PM	42964
36	US-2014-147606	delivered	12/5/17 10:34 AM	12/6/17 2:50 AM	12/8/17 9:41 PM	1/4/18 4:57 PM	43105
37	CA-2016-127208	delivered	3/7/18 3:37 PM	3/7/18 3:55 PM	3/8/18 11:25 PM	3/10/18 12:05 AM	43179
38	CA-2014-139451	delivered	8/20/18 5:38 PM	8/20/18 5:50 PM	8/21/18 1:09 PM	8/22/18 7:29 PM	43336
39	CA-2015-149734	delivered	2/9/18 1:48 PM	2/9/18 2:15 PM	2/9/18 10:53 PM	2/22/18 5:14 PM	43165
40	US-2017-119662	delivered	8/2/18 2:41 PM	8/3/18 2:35 PM	8/6/18 8:31 AM	8/10/18 10:12 PM	43328
41	CA-2017-140088	delivered	1/2/18 7:20 PM	1/2/18 7:32 PM	1/3/18 5:34 PM	1/27/18 2:27 PM	43136
42	CA-2017-155558	delivered	4/18/18 3:40 PM	4/19/18 8:15 AM	4/25/18 1:34 PM	5/3/18 1:03 PM	43238
43	CA-2016-159695	shipped	3/17/18 3:32 PM	3/17/18 3:48 PM	3/20/18 9:08 PM		43203

Customer ID	customer_email	Customer Name	Segment	Country	City	State	Postal Code	Region
SC-20770	dinther@outlook.com	Stewart Carmichael	Corporate	United States	Decatur	Alabama	35601	South
RL-19615	msroth@me.com	Rob Lucas	Consumer	United States	Montgomery	Alabama	36116	South
BS-11590	dbindel@gmail.com	Brendan Sweed	Corporate	United States	Gilbert	Arizona	85234	West
TS-21610	geekgrl@comcast.net	Troy Staebel	Consumer	United States	Phoenix	Arizona	85023	West
TW-21025	cremonini@me.com	Tamara Willingham	Home Office	United States	Scottsdale	Arizona	85254	West
AG-10900	benits@msn.com	Arthur Gainer	Consumer	United States	Tucson	Arizona	85705	West
TN-21040	unreal@live.com	Tanja Norvell	Home Office	United States	Phoenix	Arizona	85023	West
AA-10375	stern@verizon.net	Allen Armold	Consumer	United States	Mesa	Arizona	85204	West
ES-14080	dawnsong@hotmail.com	Erin Smith	Corporate	United States	Tucson	Arizona	85705	West
BB-10990	dmouse@outlook.com	Barry Blumstein	Corporate	United States	Sierra Vista	Arizona	85635	West
MD-17350	dcoppit@gmail.com	Maribeth Dona	Consumer	United States	Fayetteville	Arkansas	72701	South
DV-13045	mastinfo@msn.com	Darrin Van Huff	Corporate	United States	Los Angeles	California	90036	West
BH-11710	mastinfo@msn.com	Brosina Hoffman	Consumer	United States	Los Angeles	California	90032	West
ZD-21925	fmerges@hotmail.com	Zuschuss Donatelli	Consumer	United States	San Francisco	California	94109	West
EH-13945	fmerges@hotmail.com	Eric Hoffmann	Consumer	United States	Los Angeles	California	90049	West
RA-19885	mhouston@optonline.net	Ruben Ausman	Corporate	United States	Los Angeles	California	90049	West
KM-16720	dmouse@outlook.com	Kunst Miller	Consumer	United States	Los Angeles	California	90004	West
DN-13690		Duane Noonan	Consumer	United States	San Francisco	California	94122	West
JS-15685		Jim Sink	Corporate	United States	Los Angeles	California	90036	West
KD-16345	benits@msn.com	Katherine Ducich	Consumer	United States	San Francisco	California	94122	West
LS-16975	mstrout@me.com	Lindsay Shagiari	Home Office	United States	Los Angeles	California	90004	West
LC-16885	ozawa@me.com	Lena Creighton	Consumer	United States	Roseville	California	95661	West
SH-19975	mhouston@optonline.net	Sally Hughsby	Corporate	United States	San Francisco	California	94122	West
HA-14920	specprog@icloud.com	Helen Andreada	Consumer	United States	Pasadena	California	91104	West
SP-20650	wildfire@yahoo.com	Stephanie Phelps	Corporate	United States	San Jose	California	95123	West
TT-21070	dmouse@outlook.com	Ted Trevino	Consumer	United States	Los Angeles	California	90045	West
CS-12130	world@att.net	Chad Sievert	Consumer	United States	Los Angeles	California	90004	West
FM-14290	dbindel@gmail.com	Frank Merwin	Home Office	United States	Los Angeles	California	90032	West
MZ-17515	unreal@live.com	Mary Zewe	Corporate	United States	Redlands	California	92374	West
LE-16810	unreal@live.com	Laurel Elliston	Consumer	United States	Whittier	California	90604	West
JH-15910	fmerges@hotmail.com	Jonathan Howell	Consumer	United States	Los Angeles	California	90032	West
DB-13120	horrocks@verizon.net	David Bremer	Corporate	United States	Santa Clara	California	95051	West
LH-17155	geekgrl@comcast.net	Logan Haushalter	Consumer	United States	San Francisco	California	94109	West
KC-16540	unreal@live.com	Kelly Collister	Consumer	United States	San Diego	California	92037	West
AH-10210	shaffei@sbcglobal.net	Alan Hwang	Consumer	United States	Brentwood	California	94513	West
KH-16690	bancboy@icloud.com	Kristen Hastings	Corporate	United States	San Francisco	California	94110	West
HA-14920	mastinfo@msn.com	Helen Andreada	Consumer	United States	San Francisco	California	94109	West
BB-10990	horrocks@verizon.net	Barry Blumstein	Corporate	United States	Inglewood	California	90301	West
JO-15280	lukka@yahoo.ca	Jas O'Carroll	Consumer	United States	Los Angeles	California	90004	West
FA-14230	comdig@yahoo.com	Frank Atkinson	Corporate	United States	Long Beach	California	90805	West
GK-14620	cremonini@me.com	Grace Kelly	Corporate	United States	Hesperia	California	92345	West
DB-13615	carcus@me.com	Doug Bickford	Consumer	United States	Los Angeles	California	90045	West
AC-10420	cvrcek@me.com	Alyssa Crouse	Corporate	United States	San Francisco	California	94122	West
CC-12550	dawnsong@hotmail.com	Clay Cheatham	Consumer	United States	San Francisco	California	94122	West
JD-16060	bancboy@icloud.com	Julia Dunbar	Consumer	United States	San Francisco	California	94109	West
AS-10285	wildfire@yahoo.com	Alejandro Savely	Corporate	United States	San Francisco	California	94109	West
KH-16510	benits@msn.com	Keith Herrera	Consumer	United States	San Francisco	California	94110	West
CJ-12010	pgolle@mac.com	Caroline Jumper	Consumer	United States	Huntington Beach	California	92646	West
PB-19150	horrocks@verizon.net	Philip Brown	Consumer	United States	Los Angeles	California	90004	West
SP-20860	geekgrl@comcast.net	Sung Pak	Corporate	United States	Los Angeles	California	90045	West

Star Schema : Facts and Dimensions

- A fact table is the central table in a star schema of a data warehouse.

- A fact table stores quantitative information for analysis and is often denormalised.

- A fact table works with dimension tables.

- A fact table holds the data to be analysed, and a dimension table stores data about the ways in which the data in the fact table can be analysed.

- Thus, the fact table consists of two types of columns. The foreign keys column allows joins with dimension tables, and the measures columns contain the data that is being analysed.

Time Dimension

Order ID
Order Date
Year
Quarter
Month

Product Dimension

Product ID
Product Name
Product Category
Unit Price

SALES

Product ID
Order ID
Customer ID
Employer ID
Total
Quantity
Discount

Customer Dimension

Customer ID
Customer Name
Address
City
Zip

Emp Dimension

Emp ID
Emp Name
Title
Department
Region

Star Schema : Facts and Dimensions

Time ID	Product ID	Customer ID	Unit Sold
4	17	2	1
8	21	3	2
8	4	1	1

Now we can add a dimension table about customers:

Customer ID	Name	Gender	Income	Education	Region
1	Brian Edge	M	2	3	4
2	Fred Smith	M	3	5	1
3	Sally Jones	F	1	7	3

Hierarchical data
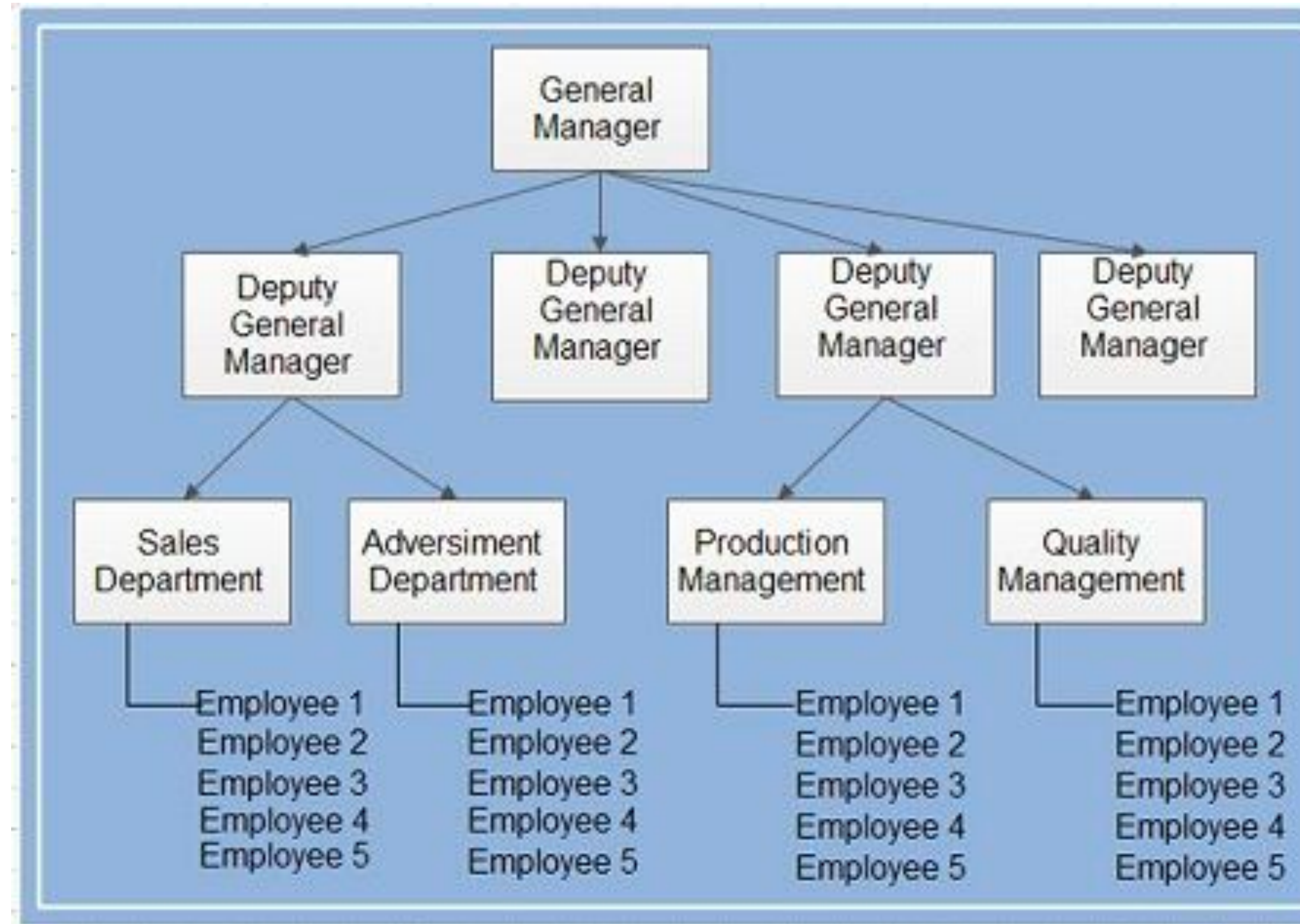
Data Granularity

- Data granularity is a measure of the level of detail in a data structure.

- In time-series data, for example, the granularity of measurement might be based on intervals of years, months, weeks, days, or hours.

- For ordering transactions, granularity might be at the purchase order level, or line item level, or detailed configuration level for customized parts.

- The name field could represent the full name or have separate entries for first name, middle name, and last name.

Any Questions?