



TERM PROJECT

**MINE SAFETY AND HEALTH
ADMINISTRATION(MSHA)**

PREPARED FOR:
PROFESSOR: JAMES MARKULIC



ENTERPRISE DATABASE MANAGEMENT

**IS 631 - 101
FALL 2016**

PREPARED BY:

ANCHITA LIKHAR – ARJUN BANUR ANANTH – SNIGDHA VULLI

VENELA SUBRAMANYAM

Table of Contents

I.	Introduction	3
II.	Data Size.....	3
III.	Establishing Primary Keys.....	4
IV.	Business Rules	6
V.	Entity Relationships – Business Rules	8
VI.	Challenges.....	9
VII.	Data Querying Performance	11
VIII.	Summary	13
IX.	References.....	14

I. Introduction

The Mine Safety and Health Administration (MSHA) government entity must conduct frequent mine inspections and investigations yearly. MSHA is required to inspect each underground mine four times a year and each surface mine twice a year for health and safety compliance. Certain mines with high levels of explosive or toxic gasses are inspected more often. Inspections are also conducted in response to complaints of hazardous conditions. MSHA inspectors work out of district offices for coal mines and metal and nonmetal mines located across the country. On finding any violation of a health or safety standard, an inspector will write a citation that specifies the standard violated and evaluates the gravity of the violation by several factors, including likelihood of injury. An inspector may also issue an order to withdraw due to imminent danger, failure to correct a violation within the appropriate time period, inadequate training of miners, or an unwarrantable failure to comply. The purposes of these investigations are to obtain and utilize information relating to health and safety conditions, root-cause analysis of accidents, causes of diseases, and physical/bodily impairments originating in such mines. The aggregation of the data is vital for reporting and analyzing for the overall safety and care of personnel and population.

II. Data Size

The originally imported data resided within one table for all Mine, Inspect, Contractor, Controller, Commodity, Employee, State, MineOffice, History, CurrentState and MineSample data. The data was divided into the following tables respectively: MSHA_Commodity consisting

of 120 distinct rows, MSHA_Contractor consisting of 10,022 distinct rows, MSHA_Controller consisting of 31,283 distinct rows, MSHA_CurrentState consisting of 51,534 distinct rows, MSHA_Employee consisting of 66,407 distinct rows, MSHA_History consisting of 31,237 distinct rows, MSHA_Inspect consisting of 2,405,362 distinct rows, MSHA_Mine consisting of 66,407 distinct rows, MSHA_MineOffice consisting of 100 distinct rows, MSHA_MineSample consisting of 51,534 distinct rows, MSHA_CommodityLookup consisting of 120 distinct rows and lastly MSHA_State with 66,407 rows. Each of the newly created tables were created to house all the relevant data specific to each of the 12 (Commodity, Contractor, Controller, CurrentState, Employee, History, Inspect, Mine, MineOffice, MineSample, CommodityLookup, State) data.

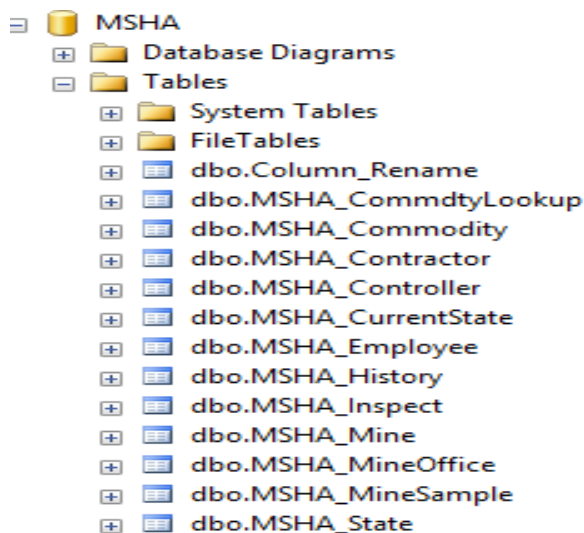
III. Establishing Primary Keys

Table	Primary Key	Reasoning for PK	PK Origin
MSHA_Commodity	msha_Mine_full_sic_cd	The commodity table is driven by the unique value of the commodity code	This is a code derived from the SIC codes to use as a primary key for the primary commodity extracted at a mine.
MSHA_Contractor	msha_mine_contractor_cntctr_id	The contractor table is unique to the contractors themselves such as an employee table. The contractor's individual ID makes each row unique such as an employee ID.	Identification number assigned by MSHA for a legal entity acting as a contractor at a mine.
MSHA_Controller	msha_Mine_controller_id	The controller table is unique to the controllers themselves such	Identification number assigned by MSHA Assessments Center for a

		as an employee table. The controller's individual ID makes each row unique such as an employee ID.	Legal Entity acting as a controller of an operator.
MSHA_CurrentState	msha_Mine_curr_mine_nm	The CurrentState Table is driven by unique Current Mine name.	Unique Mine name assigned to the mine by MSHA
MSHA_Employee	msha_Mine_mine_id	The Employee table is driven by unique Mine id at which they work	Identification number assigned to the mine by MSHA. It is a unique primary key to join to all other tables.
MSHA_History	msha_controller_history_oper_id	The history table is driven by unique operator id that worked at the Mine	Operator id is Identification number assigned to the operator working at Mine by MSHA.
MSHA_Inspect	msha_Inspect_event_no & msha_Mine_mine_id	The inspection event number is unique to the inspection but mines can have multiple inspections with event numbers associated. For this reason, a composite key of event number and mine ID were used.	msha_Inspect_event_no A unique number used to identify each inspection, investigation, audit, survey, etc. (event). msha_Mine_mine_id Identification number assigned to the mine by MSHA. It is a unique primary key to all other tables.
MSHA_Mine	msha_Mine_mine_id	The mine table is driven by the unique values of the mine IDs	Identification number assigned to the mine by MSHA. It is a unique primary key to join to all other tables.
MSHA_MineOffice	msha_Mine_office_cd	The MineOffice table is driven by unique Office Cd's	Identification number (Cd's) assigned to the mine by MSHA
MSHA_MineSample	msha_Mine_curr_mine_nm	The MineSample table is driven by	Unique Mine name assigned to the mine by MSHA

		unique Mine name	
MSHA_CommodityLookup	msha_Commodity_Lookup_cmm dty_cd	The commodity Lookup table is driven by unique commodity look up Cd's	Identification number (Cd's) assigned to the mine by MSHA
MSHA_State	msha_Mine_mine_id	The State Table is driven by a unique mine id	Identification number assigned to the mine by MSHA. It is a unique primary key to join tables

All the above tables were created in management studio using MSHA_Full table. Below is the screenshot of all the table that were created. The DML and DDL for the table creation and the table updation was submitted in the moodle. The relationships between each of these tables are shown in the entity relationship diagram (ER model).



IV. Business Rules

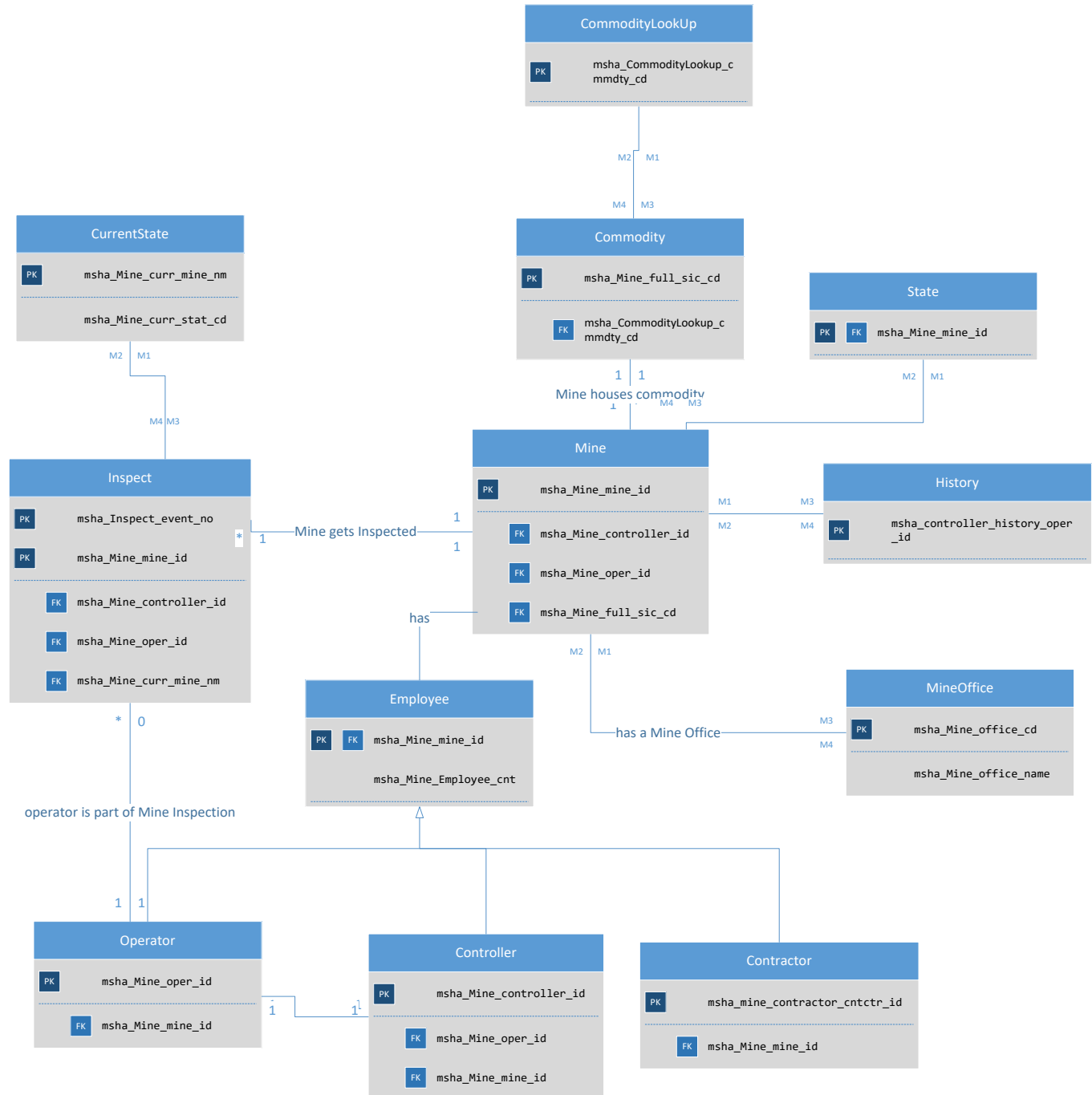
By distributing the compiled data across these 12 newly created tables the data is then made available for querying in a more organized and sound manner for multiple querying purposes. Each

of the 12 tables (Commodity, Contractor, Controller, CurrentState, Employee, History, Inspect, Mine, MineOffice, MineSample, CommodityLookup, State) house only the data that is pertinent to the entity and for its relationships to other tables. The tables were created mainly for the reason of clearly distinguishing the tasks that were performed by various entities that worked at the mine. By doing so it clearly gives us a picture of how each entities are related to each other. Further helping us to identify those operations which might directly or indirectly impact the environment. Clearly the dataset in its initial stage lacked any particular structure. All we are trying to do here in our project is to break down the initial unstructured data into multiple tables and add relationships among these tables. In doing which we hope to succeed in running multiple query to understand more about the places were the gold ore is being mined and many more.

Below table highlights the relationship between each of the table that was created.

Entity A	In Relation to	Entity B
Controller	Supervises	Operator
Operator	Works in	Mine
Mine	Gets	Inspected
Mine	Houses	Commodities
Contractor	Operates	Mine
Inspection	Includes	Operator
Employee	Works in	Mine
CurrentState	Status of	Mine
History	Events at	Mine
MineOffice	Includes	Controller
MineSample	Gets	Inspected
CommodityLookup	Supervised by	Operator
State	Location of	Mine

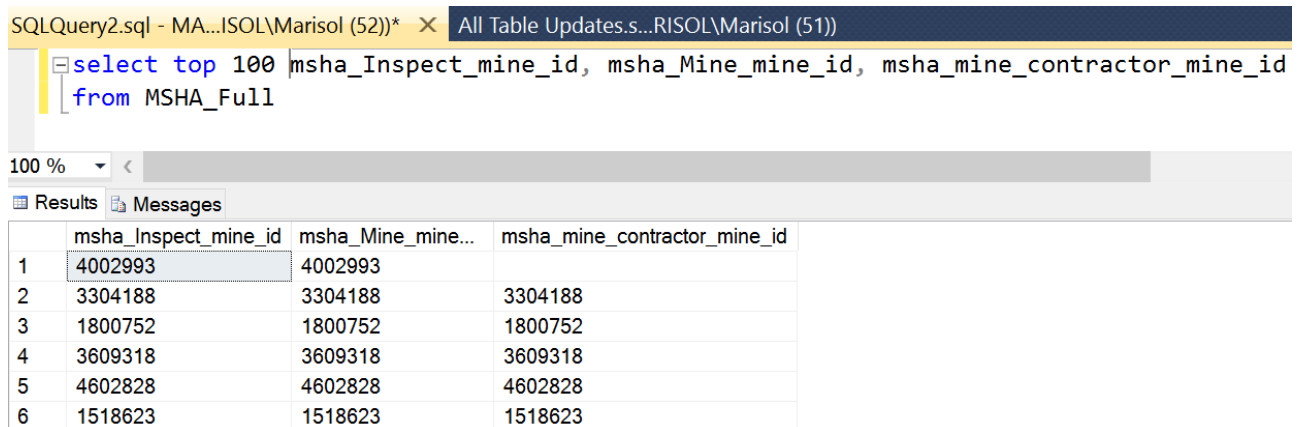
V. Entity Relationships – Business Rules



Please note: these screenshots of the tables are not representative of all the data in the actual database for each of the tables.

VI. Challenges

The actual MSHA data lacked structure with respect to table organization. To expand the table we would require tables for mines(MSHA_Mine),for the operators who work in the mines (MSHA_Operators),for the separate inspection events(MSHA_Inspect),for the contractors who operate the mines (MSHA_Contractor), for the controllers who oversee the operators during work and inspections (MSHA_Controller) and lastly, for the commodities that are obtained from the mines (MSHA_Commodities).One of the challenges was to segregate data in a 3rd normal form which were of varying name conventions for columns in the original import that housed duplicate data.



The screenshot shows a SQL query window with the following query:

```
select top 100 msha_Inspect_mine_id, msha_Mine_mine_id, msha_mine_contractor_mine_id
from MSHA_Full
```

The results pane shows the following data:

	msha_Inspect_mine_id	msha_Mine_mine...	msha_mine_contractor_mine_id
1	4002993	4002993	
2	3304188	3304188	3304188
3	1800752	1800752	1800752
4	3609318	3609318	3609318
5	4602828	4602828	4602828
6	1518623	1518623	1518623

To create redundancies, we had to be familiar with data within each of the columns and would have to use several queries. Other challenges were that we could not figure out the controller's relationship had to be in conjunction to the other tables. Initially we thought that only 5 tables were needed, later we got to know that the controller's existence was solely to oversee the mine operator and to serve as the main point of supervision during inspections. So while the controller has a direct correlation to the mine there is also a direct correlation to the operator which means that the

operator is needed to exist as its own entity. Once the operator table was created it was easier to see the operator had correlations to the inspections, mine and controller and all the rest of the tables. Analyzing the relationships and justifying the data existence we could understand the missing data. Later we further continued by creating various other tables such as State, these states provide mine rescue services to mines within their respective jurisdictions. According to each MSHA-State agreement, inspections of rescue stations are to be conducted every 6 months by a designated MSHA representative in the district where the station is located. Further we saw all those attributes which can be separated from the MSHA_full and created tables for MineOffice, CommodityLookup, MineSample, History for containing all the Mine related histories. It was difficult in linking all the tables by finding the common entities. But we eventually figured it out in the later point of the project as we developed more understanding towards the concepts.

On Some of the missing elements are explained in the 2 examples below:

Table	Data	Justification
MSHA_Mine	msha_Mine_oper_id	May not contain data (null values) if this record has a status of New Mine. If it is a New Mine, this information will be entered into the system at a future date.
MSHA_Controller	msha_controller_history_ctrlr_end_dt	Date the controller ceased to control the operator. May not contain data (null values) if controller work is still active.

VII. Data Querying Performance

Filtering the data to include only the observations you need can dramatically improve query speed. It was also important to drop indexes before loading data into existing tables. It is important to break down the larger tables to smaller ones to run query. If your query hits one or more tables with millions of rows or more, it could affect performance. In our project MSHA dataset was pretty huge for us to run any query. Since the processing time was more. we break down the MSHA_full table into multiple tables based on various facts and in that way we were able run the query much more efficiently. Microsoft SQL Server Management studio has given great performance and ease of use with the MSHA data. The querying capabilities have given great use to the available data.

Some examples of business use include: Compliance officer requests for listing of all mines that have had fatalities during operation. MSHA has the option to conduct a Part 50 reporting audit at a mine where a fatal accident occurs. Factors MSHA will consider in deciding whether to conduct a Part 50 audit after a fatality include the following:

MSHA received complaints by miners or a miners' representative that reportable accidents were not being reported by the mine operator;

MSHA otherwise has reason to believe there has been underreporting of accidents at a mine; or

In our project we ran a simple query to capture all mines with a fatality related inspection and data was drilled down further to capture all mines with fatality occurrences greater than 1 time, occurrences during a specified calendar year, or occurrences related to a specific contractor.

Example 1: From the query example 1.0 you can deduce that contractors such as Shapiro Brothers and Vander

```

SELECT MSHA_Inspect.msha_Inspect_event_no, MSHA_Inspect.msha_Mine_mine_id,
MSHA_Inspect.msha_Inspect_cal_yr, MSHA_Contractor.msha_mine_contractor_cntctr_id,
MSHA_Contractor.msha_mine_contractor_cntctr_nm
FROM MSHA_Inspect
    join MSHA_Contractor
    on MSHA_Inspect.msha_Mine_mine_id = MSHA_Contractor.msha_Mine_mine_id
WHERE msha_Inspect_acty_cd = 'AFA' --AFA is the inspection code for fatalities
ORDER BY MSHA_Inspect.msha_Mine_mine_id, MSHA_Contractor.msha_mine_contractor_cntctr_id

```

	msha_Inspect_event_no	msha_Mine_mine_id	msha_Inspect_cal_yr	msha_mine_contractor_cntctr_id	msha_mine_contractor_cntctr_nm
103	5018890	1100610	1982	ICZ	Deraway & Wells Scrap Proc Inc
104	6005888	1100610	1996	ICZ	Deraway & Wells Scrap Proc Inc
105	5193735	1100611	1983	BFN	Shapiro Brothers Of Ill Inc
106	5225404	1100611	1983	BFN	Shapiro Brothers Of Ill Inc
107	5018821	1100617	1982	IVV	Vander Parts Company
108	5062574	1100617	1982	IVV	Vander Parts Company
109	5957025	1100617	1991	IVV	Vander Parts Company
110	5957280	1100617	1991	IVV	Vander Parts Company
111	5440810	1100628	1986	P750	Oxford Crane Inc
112	5958354	1100628	1991	P750	Oxford Crane Inc

Example 2:

SQL server can provide a query for a contractor soliciting information on all mines that produce Lime.

```

SELECT msha_Mine_mine_id, msha_Mine_curr_mine_nm, msha_Mine_state_abbr, msha_Mine_sic_desc
FROM MSHA_Mine
WHERE msha_Mine_sic_desc = 'Lime'
ORDER BY msha_Mine_state_abbr

```

	msha_Mine_mine_id	msha_Mine_curr_mine_nm	msha_Mine_state_abbr	msha_Mine_sic_desc
7	0100643	Roberta Lime Plant (Kilns)	AL	Lime
8	0101264	Alabaster Plant	AL	Lime
9	0102053	Suggsville Pit And Plant	AL	Lime
10	0102146	Demopolis Quarry	AL	Lime
11	0102353	EPES LIME PIT	AL	Lime
12	0102960	Sharp Ag Mine	AL	Lime
13	0103227	Southern Lime	AL	Lime

VIII. Summary

To conclude, Microsoft SQL Server Management Studio proved to be easy to use, Import, manage and query huge datasets. Through this project we developed much more understanding towards the SQL language and it's use in the real world. The functionality of the SQL server management studio was easy to understand even being beginner. Recreation and normalizing the data was relatively easy. While running simple queries for business use such as two examples, in the data query and performance section helped us know the importance of normalization and how it can make life easier.

MSHA project helped us understand how the database knowledge can help an organization make much more informed decision. By having all the related information in place. It becomes easy to write a query and make sense out of the awful data.

In today's world Databases allow companies to store virtually any type of data. Their speed and relatively low cost makes them popular among businesses with large customer bases.

IX. References

Avi Silberschatz, Henry F. Korth, S. Sudarshan, Database System Concept, McGraw- Hill, ISBN 0-07-352332-1, 6th edition.

<https://www.msha.gov/compliance-enforcement/mine-inspections>

<http://arlweb.msha.gov/READROOM/HANDBOOK/HANDBOOK.HTM>

<http://arlweb.msha.gov/regs/complian/ppm/pmvol3d.htm>

Database Management Systems (3rd edition) by Ramakrishnan and Gehrke.

<https://www.reference.com/technology/companies-use-database-e0c5d2ba994c2360>

<https://learnatvivid.com/solutions/courses/msha>

<https://msdn.microsoft.com/en-us/library/ms188927.aspx>

<http://stackoverflow.com/questions/2221929/what-is-the-importance-of-schema-in-sql-server>