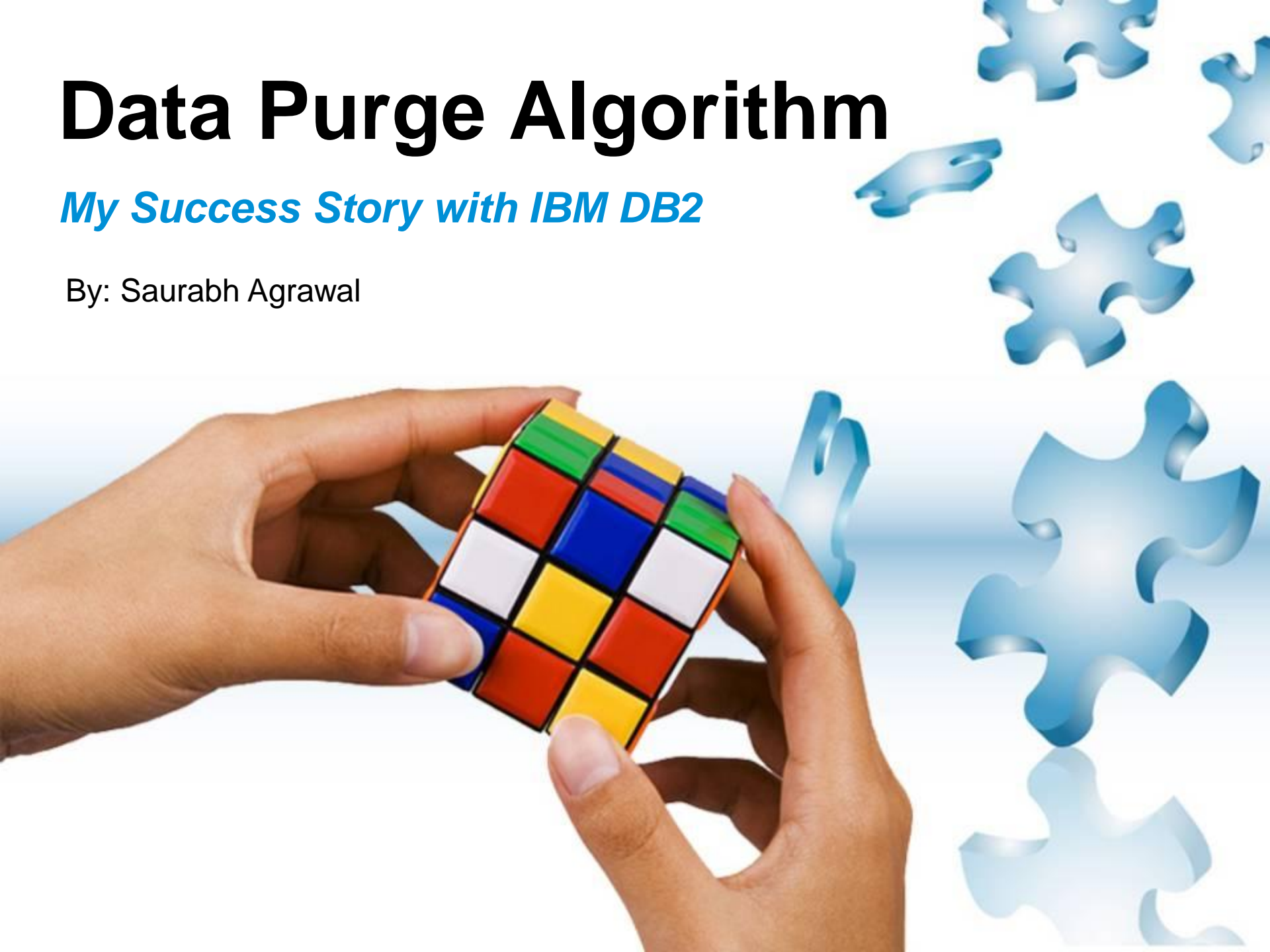


# Data Purge Algorithm

*My Success Story with IBM DB2*

By: Saurabh Agrawal



# Quick Recap

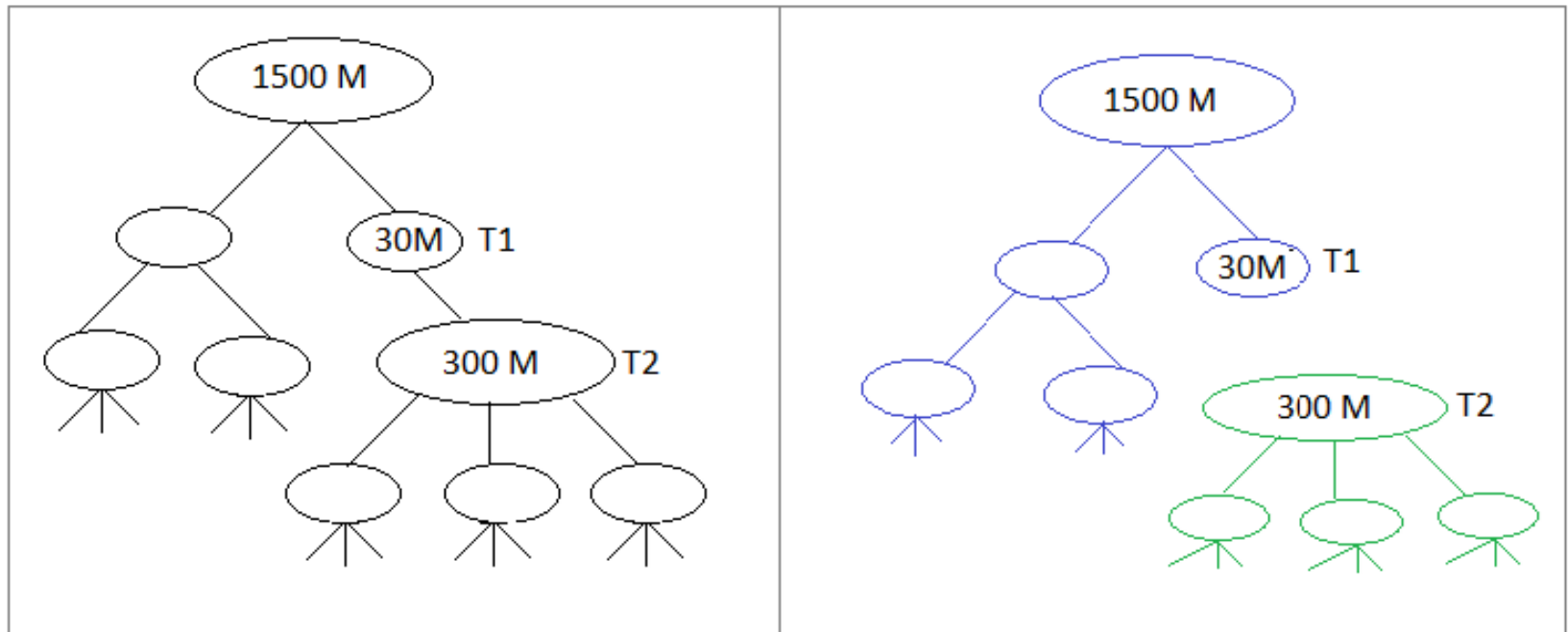
Refer: DB2Night Show: #124

- **Problem:** Purge data to reduce DB from 1.8TB to 0.8TB in size.
- Approach which failed?
  - Cascaded Delete.
  - PL/SQL block with commit interval.
- **Troubles?**



# Solution

- **HELP:** Thanks DB2 Forums !!!
  - Reload root table with **valid** data.
  - Move invalid data from children into exception tables.



# Configuring Parameters

DB2 DB CFG Parameters	DB2 Profile Registry Parameters
<i>STMT_CONC =LITERALS</i>	<i>DB2_SKIPINSERTED =ON</i>
<i>SECTION_ACTUALS =BASE</i>	<i>DB2_PARALLEL_IO =*</i>
<i>DFT_DEGREE =ANY</i>	<i>DB2_EVALUNCOMMITTED =ON</i>
<i>PCKCACHESZ =AUTOMATIC</i>	<i>DB2_USE_ALTERNATE_PAGE_CLEANING =YES</i>
<i>CATALOGCACHE_SZ =2000</i>	<i>DB2_SKIPDELETED =ON</i>
<i>LOGBUFSZ =2048</i>	
<i>CUR_COMMIT =ON</i>	
<i>UTIL_HEAP_SZ =524288</i>	
<i>BUFFPAGE =10000</i>	
<i>LOGFILSIZ =16384</i>	

*\*Choose numbers which best suits your environment.*

# Algorithm

```
db2 -x "select 'runstats on ' || trim(tabschema) || '.' || trim(tabname) || ';' from
syscat.tables where tabschema='SCHEMA' and type='T'" > runstats.sql
```

```
db2 "select substr(tabname,1,30) tabname,card from syscat.tables where
tabschema='SCHEMA' and type='T'" > OriginalCount.out
```

```
db2 "call get_dbsize_info(?,?,?,-1)" > OriginalDBSize.out
```

```
db2look -d <dbname> -a -e -l -x -c -o OriginalDB.ddl
```

```
db2 "select substr(tabname,1,30)tabname,card from syscat.tables where
tabschema='SCHEMA' and type='T' and card > 300000000" > BigChdTab.out
```

```
db2 -x "select 'alter table ' || trim(a.tabschema) || '.' || trim(a.tabname) || ' drop
constraint ' || a.CONSTNAME || ';' from syscat.references a,syscat.tables b where
a.tabschema=b.tabschema and a.tabschema='SCHEMA' and b.type='T' and
a.tabname=b.tabname and b.card > 300000000" > brkParent.sql
```

```
db2 export to tabname.ixf of ixf select * from tabname where <...>
```

```
db2look -d <dbname> -t <tabname> -e -o tabname.ddl
```

# Algorithm (contd...)

```
db2 "select 'alter table ' || trim(st.tabschema) || '.' || trim(st.tabname) || ' drop
constraint ' || st.constname || ';' from SYSCAT.KEYCOLUSE sk inner join
SYSCAT.TABCONST st on sk.TABNAME = st.TABNAME and
sk.TABSCHEMA=st.TABSCHEMA and st.tabschema='SCHEMA' where st.type in ('P','U')
and st.CONSTNAME =sk.CONSTNAME and exists(select 1 from syscat.tables a where
a.tabname=st.TABNAME and st.TABSchema=a.tabschema and a.card > 300000000)" >
dropPK.sql
```

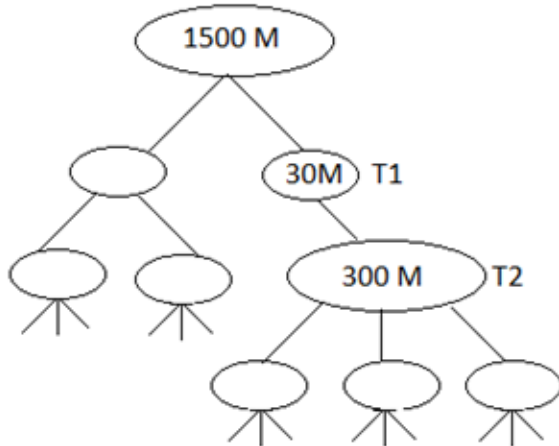
```
db2 "select 'drop index ' || trim(INDSCHEMA) || '.' || trim(INDNAME) || ';' from
syscat.indexes where TABNAME in (select tabname from syscat.tables where
tabschema='SCHEMA' and type='T' and card > 300000000) and tabschema='SCHEMA'"
> dropIndexes.sql
```

```
db2 "load from <export file>.ixf of ix replace into <tablename> DATA BUFFER 131072
SORT BUFFER 131072 CPU_PARALLELISM 16 DISK_PARALLELISM 16"
(use "nmon" command to check number of CPUs on your machine)
```

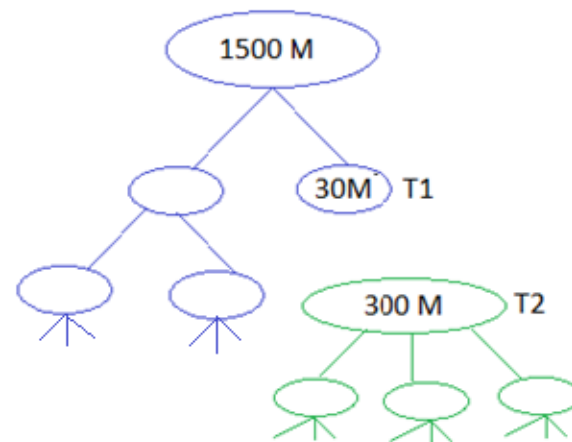
```
db2 "set integrity for <tabschema>.<tablename> immediate checked"
```

# How My DB Looks?

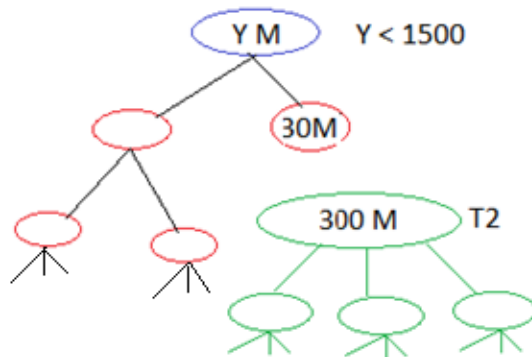
1. Original Structure



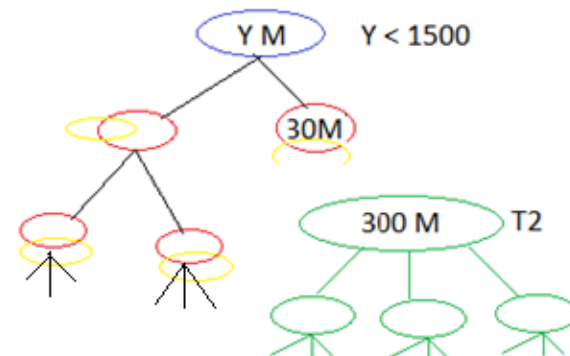
2. Before Algorithm



3. After Purge from Root



4. After creating EXP Tables



# Algorithm (contd...)

```
db2 -x "select 'CREATE TABLE ' || trim(TABNAME) || '_exp' || ' like ' || tabname || ';'
from SYSCAT.TABLES where STATUS='C' and type='T' and TABSCHEMA='SCHEMA'" >
createExceptionTab.sql
```

```
db2 -x "select 'Alter table ' || trim(TABNAME) || '_exp' || ' add column c1 TIMESTAMP
add column c2 CLOB; ' from SYSCAT.TABLES where STATUS='C' and type='T' and
TABSCHEMA='SCHEMA'" >> createExceptionTab.sql
```

Recreate the dropped indexes and PK constraints.

- Setting integrity of tables? Is it easy?

**SQL3608N** Cannot check a dependent table dependent-table-name using the SET INTEGRITY statement while the parent table or underlying table parent-table-name is in the Set Integrity Pending state or if it will be put into the Set Integrity Pending state by the SET INTEGRITY statement.



# Set Integrity: The Easy Way

- Create and execute the shell script.

```
db2 -x "select 'SET INTEGRITY FOR ' || TABSCHEMA || '.' || TABNAME || ' IMMEDIATE  
CHECKED FOR EXCEPTION IN ' || TABNAME || ' USE ' || TABNAME || '_exp ;' from  
SYSCAT.TABLES where STATUS='C' and type='T' and TABSCHEMA='$SCHEMA' ">  
chkset_integrity.sql
```

```
tabcnt=$(wc -l < chkset_integrity.sql)  
while [[ ${tabcnt} -gt 0 ]];  
do  
    db2 connect to $DBName >> output.out  
    db2 set schema $SCHEMA >> output.out  
    db2 -tvf chkset_integrity.sql >> output.out  
    db2 -x "select 'SET INTEGRITY FOR ' || TABSCHEMA || '.' || TABNAME || ' IMMEDIATE  
CHECKED FOR EXCEPTION IN ' || TABNAME || ' USE ' || TABNAME || '_exp ;' from  
SYSCAT.TABLES where STATUS='C' and type='T' and TABSCHEMA='$SCHEMA'  
order by card " > chkset_integrity.sql  
    tabcnt=$(wc -l < chkset_integrity.sql)  
done
```

# Algorithm (contd...)

Recursively use the same algorithm for the table hierarchy of isolated tables.

Turn OFF integrity for the tables which we had isolated.

```
db2 "set integrity for <tab1,tab2,...,tabn> off"
```

Recreate the dropped FK relationships with parents.

Turn ON integrity for these tables.

```
db2 "set integrity for <tab1 all,tab2 all,...,tabn all> immediate unchecked"
```

```
db2look -d <dbname> -a -e -l -x -f -o FinalDB.ddl
```

```
db2 -x "select 'runstats on ' || trim(tabschema) || '.' || trim(tabname) || ';' from  
syscat.tables where tabschema='SCHEMA' and type='T'" > runstats.sql
```

```
db2 "select substr(tabname,1,30) tabname,card from syscat.tables where  
tabschema='SCHEMA' and type='T'" > FinalCount.out
```

```
db2 "call get_dbsize_info(?,?,?,-1)" > FinalDBSize.out
```

Compare final DB size, counts and DB2LOOK outputs with the initial.

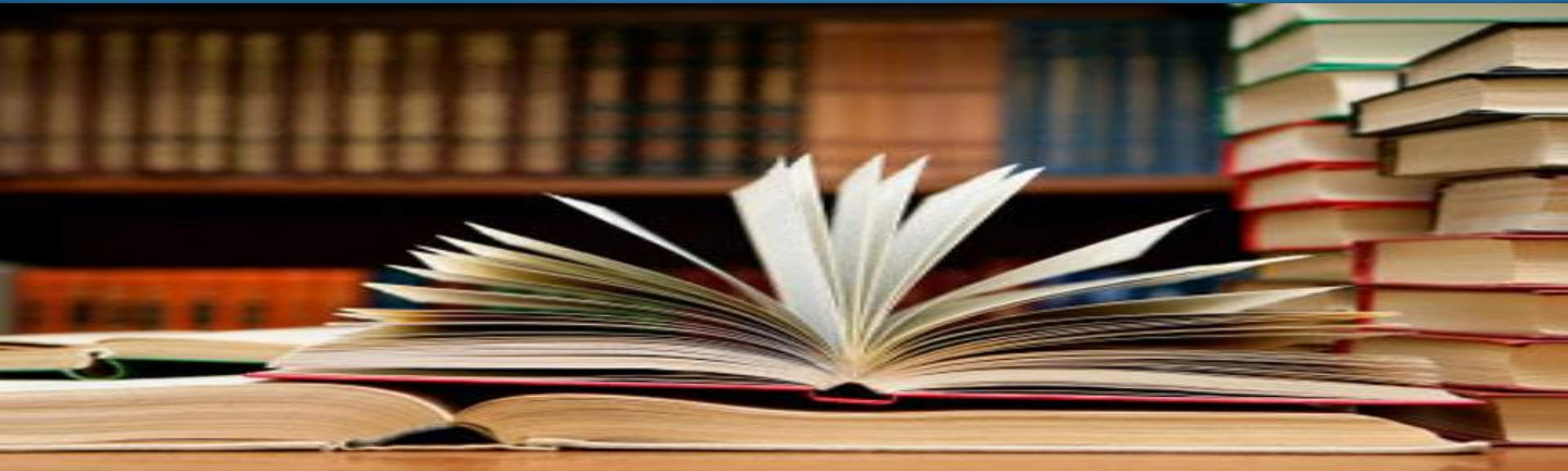
Drop the exception tables.

# That's My Success Story...

*“There is always a better solution for the problem,  
challenge is to figure it out !!!”*



# Quick Reference



- My Blog:  
<http://saurabhska.wordpress.com/2013/11/14/database-shrinking-purge-a-large-amount-of-data-from-database/>

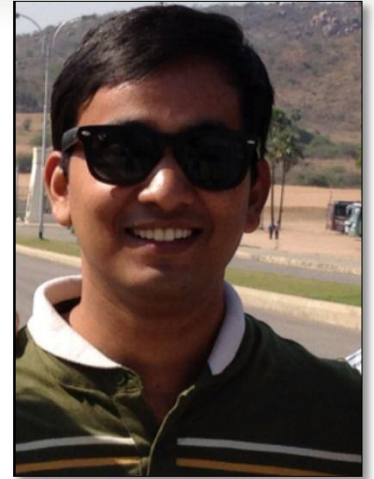




THANK YOU

# Contact Me

- LinkedIn : [in.linkedin.com/pub/saurabh-agrawal/23/845/a8a/](https://in.linkedin.com/pub/saurabh-agrawal/23/845/a8a/)
- Email: [saurabh.ska@gmail.com](mailto:saurabh.ska@gmail.com)
- Twitter: Follow me @saurabhska
- Personal Blog:  
<http://saurabhska.wordpress.com/>



**“PLEASE VOTE FOR ME !!!”**