# SEMINAR ON BIG DATA GENERATION TECHNIQUES

# DBSCALE: A SYNTHETICALLY DATASET SCALING TOOL

Sponsored by HK ITF | Developed by Database Research Group, HKPU
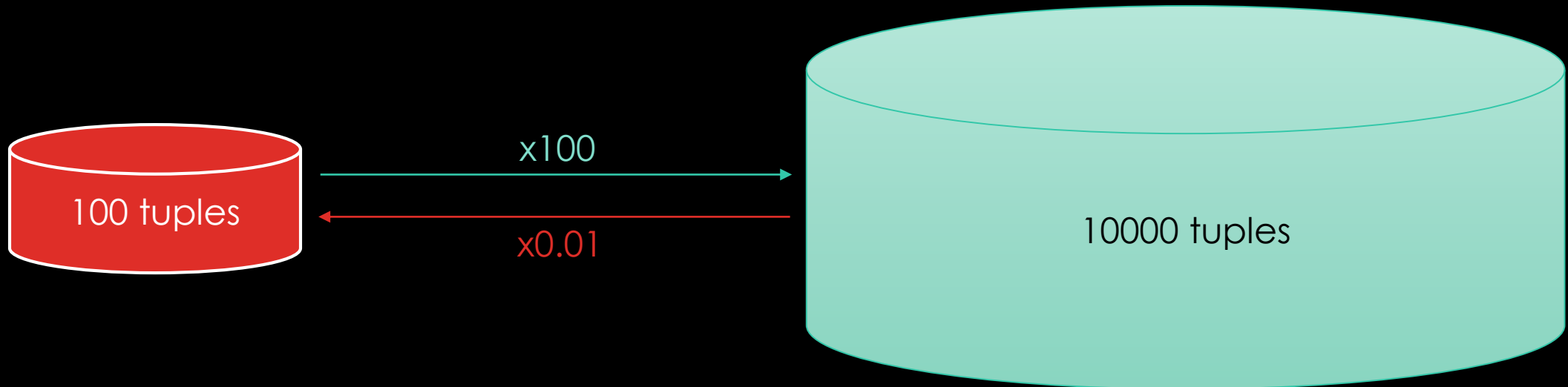
Cliz Sun

# PROMOTION OF DBSCALE

- Why database scaling?
- Why DBScale?
- Demo
- Q&A

# WHY DATABASE SCALING

- What is scaling?

x100

x0.01

100 tuples

10000 tuples

- Scaling helps:
  - Testing & Development
  - Planning
  - Prepare special dataset which is not available by normal generator

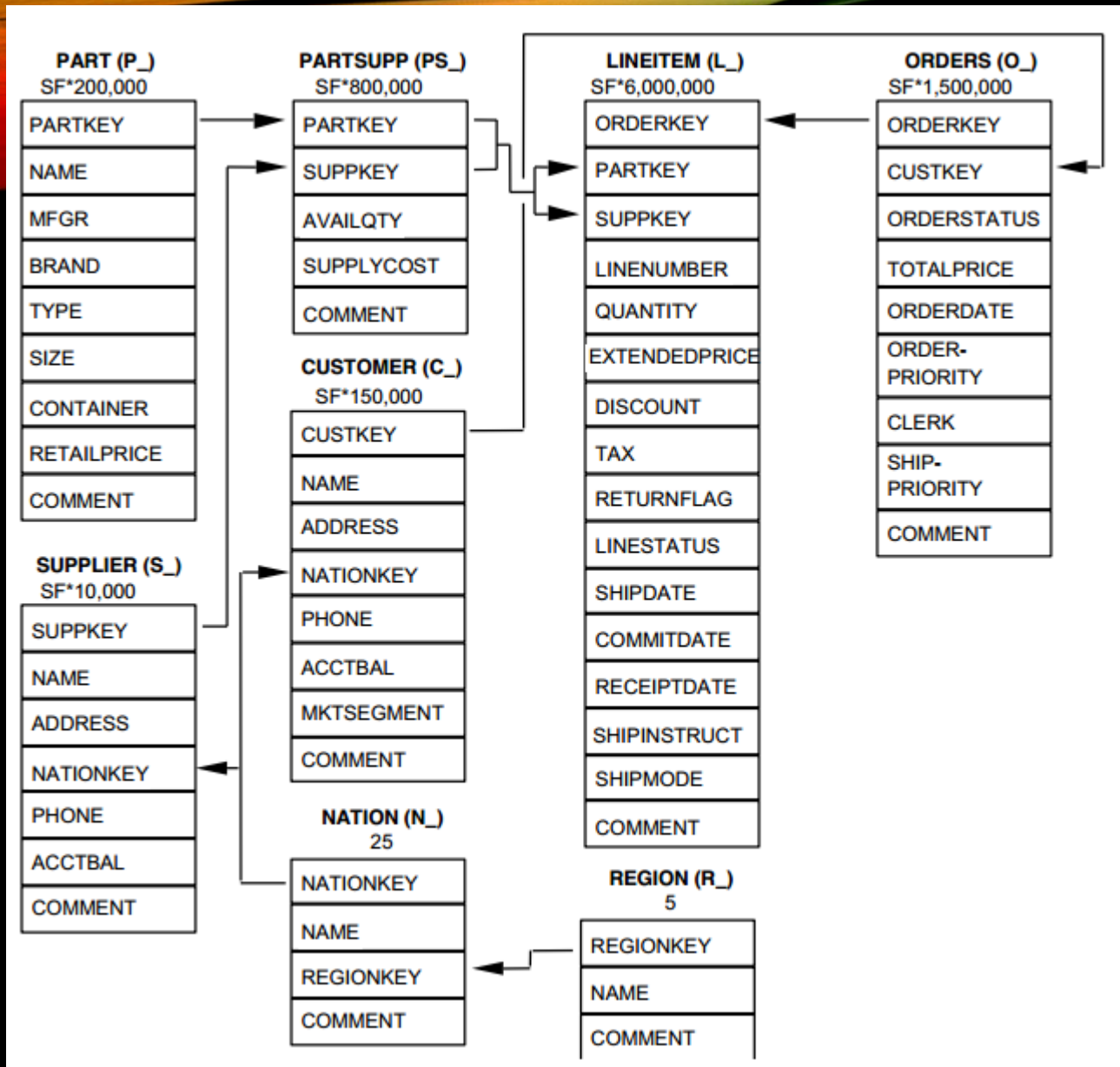# DBSCALE: POWERFUL AND FRIENDLY SCALING TOOL

- First tool to do this
- Generate from existing real data
- Keep characteristic (data composition & query results)
- For testing and planning

# SCALE UP WITH DBSCALE

- Test database: a sales warehouse database

# SCHEMA



**PART (P_)**
SF*200,000

- PARTKEY
- NAME
- MFGR
- BRAND
- TYPE
- SIZE
- CONTAINER
- RETAILPRICE
- COMMENT

**SUPPLIER (S_)**
SF*10,000

- SUPPKEY
- NAME
- ADDRESS
- NATIONKEY
- PHONE
- ACCTBAL
- COMMENT

**PARTSUPP (PS_)**
SF*800,000

- PARTKEY
- SUPPKEY
- AVAILQTY
- SUPPLYCOST
- COMMENT

**CUSTOMER (C_)**
SF*150,000

- CUSTKEY
- NAME
- ADDRESS
- NATIONKEY
- PHONE
- ACCTBAL
- MKTSEGMENT
- COMMENT

**NATION (N_)**
25

- NATIONKEY
- NAME
- REGIONKEY
- COMMENT

**LINEITEM (L_)**
SF*6,000,000

- ORDERKEY
- PARTKEY
- SUPPKEY
- LINENUMBER
- QUANTITY
- EXTENDEDPRICE
- DISCOUNT
- TAX
- RETURNFLAG
- LINESTATUS
- SHIPDATE
- COMMITDATE
- RECEIPTDATE
- SHIPINSTRUCT
- SHIPMODE
- COMMENT

**ORDERS (O_)**
SF*1,500,000

- ORDERKEY
- CUSTKEY
- ORDERSTATUS
- TOTALPRICE
- ORDERDATE
- ORDER-PRIORITY
- CLERK
- SHIP-PRIORITY
- COMMENT

**REGION (R_)**
5

- REGIONKEY
- NAME
- COMMENT

# SCALE UP WITH DBSCALE

- Demo flow:
  - Show original data
  - Scale up 2x
  - Compare database size
  - Compare query results (of 2 queries)

# INTEGRITY CONSTRAINT

- Set up special rule for generation of column data
- Use formulas:

  e.g. DELIVER DATE should be shipping date + 20day

  Normal distribution -> not keeping this constraint
  Use formula: DATEADD(SHIPDATE, 20) -> guarantee constraint

# Q&A SESSION

# MORE INFORMATION & TRAIL DOWNLOAD AT

- [Http://dbgroup.comp.polyu.edu.hk/?page_id=101](Http://dbgroup.comp.polyu.edu.hk/?page_id=101)