

Thinking in Columns, Not Rows



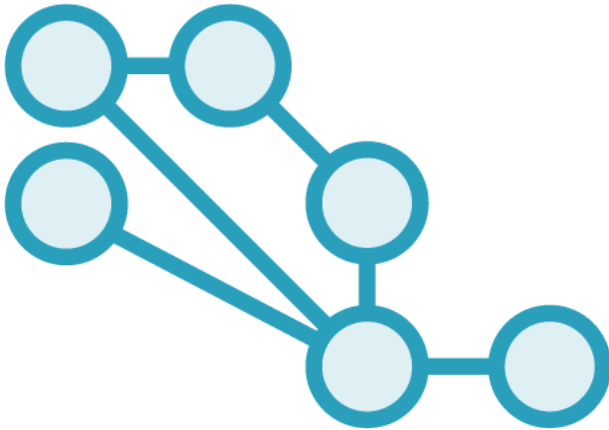
Eugene Meidinger

DATABASE DEVELOPER

@sqlgene www.sqlgene.com



Two Uses for Databases



OLTP

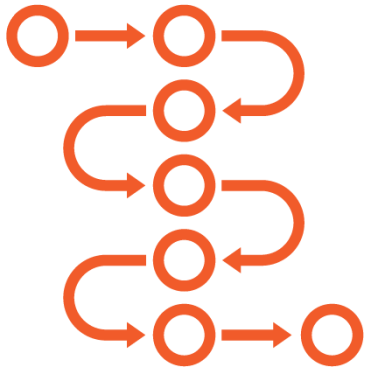
Applications designed for day to day use and frequent updates



OLAP

Applications designed for daily reporting and heavy reads

Optimizing for OLAP



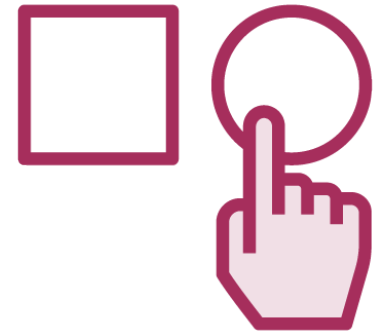
Single-column
aggregations



Large number of
rows



Repeated values



Need to quickly
apply filters

Columnar databases are
highly optimized for
analytical reporting.



Columnar Databases



Thinking in Columns

Order ID	Order Date	Customer ID	Amount	Type	State
10001	01/01/2018	1	\$2.00	Modest	PA
10002	01/01/2018	1	\$2.00	Sorcerous	WV
10003	01/01/2018	2	\$2.00	Stone	OH
10004	01/01/2018	2	\$2.00	Tropical	PA
10005	01/01/2018	3	\$2.00	Marshy	WV
10006	01/01/2018	3	\$2.00	Meadow	OH
10007	01/01/2018	4	\$2.00	Forest	PA



Row-store Databases

Order ID	Order Date	Customer ID	Amount	Type	State
10001	01/01/2018	1	\$2.00	Modest	PA
10002	01/01/2018	1	\$2.00	Sorcerous	WV
10003	01/01/2018	2	\$2.00	Stone	OH
10004	01/01/2018	2	\$2.00	Tropical	PA
10005	01/01/2018	3	\$2.00	Marshy	WV
10006	01/01/2018	3	\$2.00	Meadow	OH
10007	01/01/2018	4	\$2.00	Forest	PA



Column-store Databases

Order ID	Order Date	Customer ID	Amount	Type	State
10001	01/01/2018	1	\$2.00	Modest	PA
10002	01/01/2018	1	\$2.00	Sorcerous	WV
10003	01/01/2018	2	\$2.00	Stone	OH
10004	01/01/2018	2	\$2.00	Tropical	PA
10005	01/01/2018	3	\$2.00	Marshy	WV
10006	01/01/2018	3	\$2.00	Meadow	OH
10007	01/01/2018	4	\$2.00	Forest	PA



Column-store Databases

Order ID	Order Date	Customer ID	Amount	Type	State
10001	01/01/2018	1	\$2.00	Modest	PA
10002	01/01/2018	1	\$2.00	Sorcerous	WV
10003	01/01/2018	2	\$2.00	Stone	OH
10004	01/01/2018	2	\$2.00	Tropical	PA
10005	01/01/2018	3	\$2.00	Marshy	WV
10006	01/01/2018	3	\$2.00	Meadow	OH
10007	01/01/2018	4	\$2.00	Forest	PA



Vertipaq

The engine used to store data as columns, alternatively known as xVelocity.



DirectQuery

The engine that translates DAX formulas into relational SQL queries.





Leveraging Timely On-premises Data with Power BI

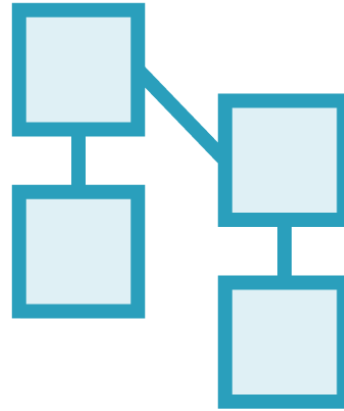
Eugene Meidinger



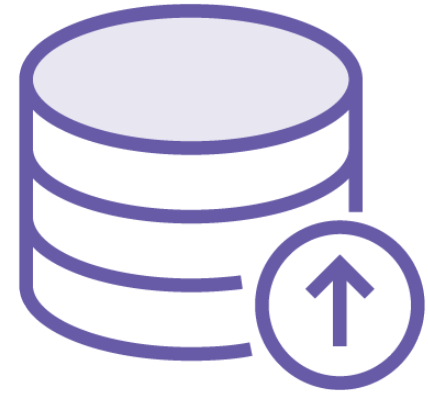
Quick Review



Star Schema



Vertipaq



DirectQuery

Compression and Encoding



Types of Encoding

Value encoding

Dictionary encoding

Run-length encoding



Value Encoding

Account
6001
6002
6003
6004
6005
6006
6007

85000000



Dictionary Encoding

Color
1
2
2
3
3
3
3

Key	Color
1	Blue
2	Green
3	Red



Run-length Encoding

Color
Blue
Green
Green
Red
Red
Red
Red



Color
Blue, 1
Green, 2
Red, 4

Columnar storage turns
repeated values from a
waste of space
to an asset.



Demo



Visualizing run-length encoding

Comparing compression rates



Related	
1	1
1	1
1	1
1	1
2	2
2	2
Four 1s	Four 1s
Four 2s	Four 2s



Unrelated			
1	1	1	1
1	1	1	1
1	1	2	2
1	1	2	2
2	2	1	Two 1s
2	2	1	Two 2s
Four 1s	2	2	Two 1s
Four 2s	2	2	Two 2s



Unique

a

b

c

d

e

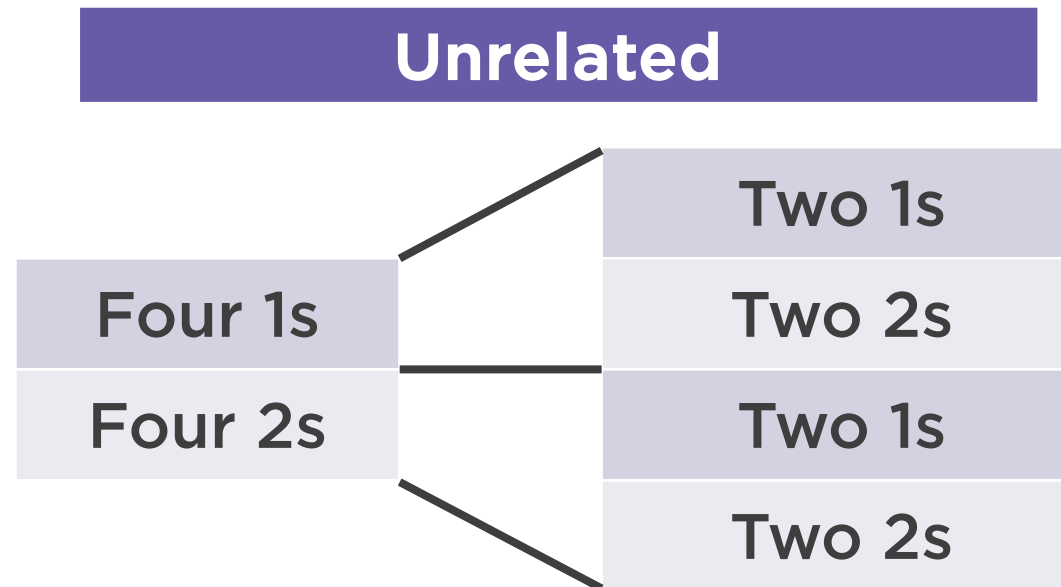
f

g

h



Related	
Four 1s	Four 1s
Four 2s	Four 2s



Unique
a
b
c
d
e
f
g
h



Summary



OLAP is designed for reporting

Columnar databases are optimized for OLAP

Columnar storage plus encoding allows for great compression

