



# JOURNEY TO VERTICA+ROLAP SOLUTION

Ramūnas Balukonis, Senior Developer, adform



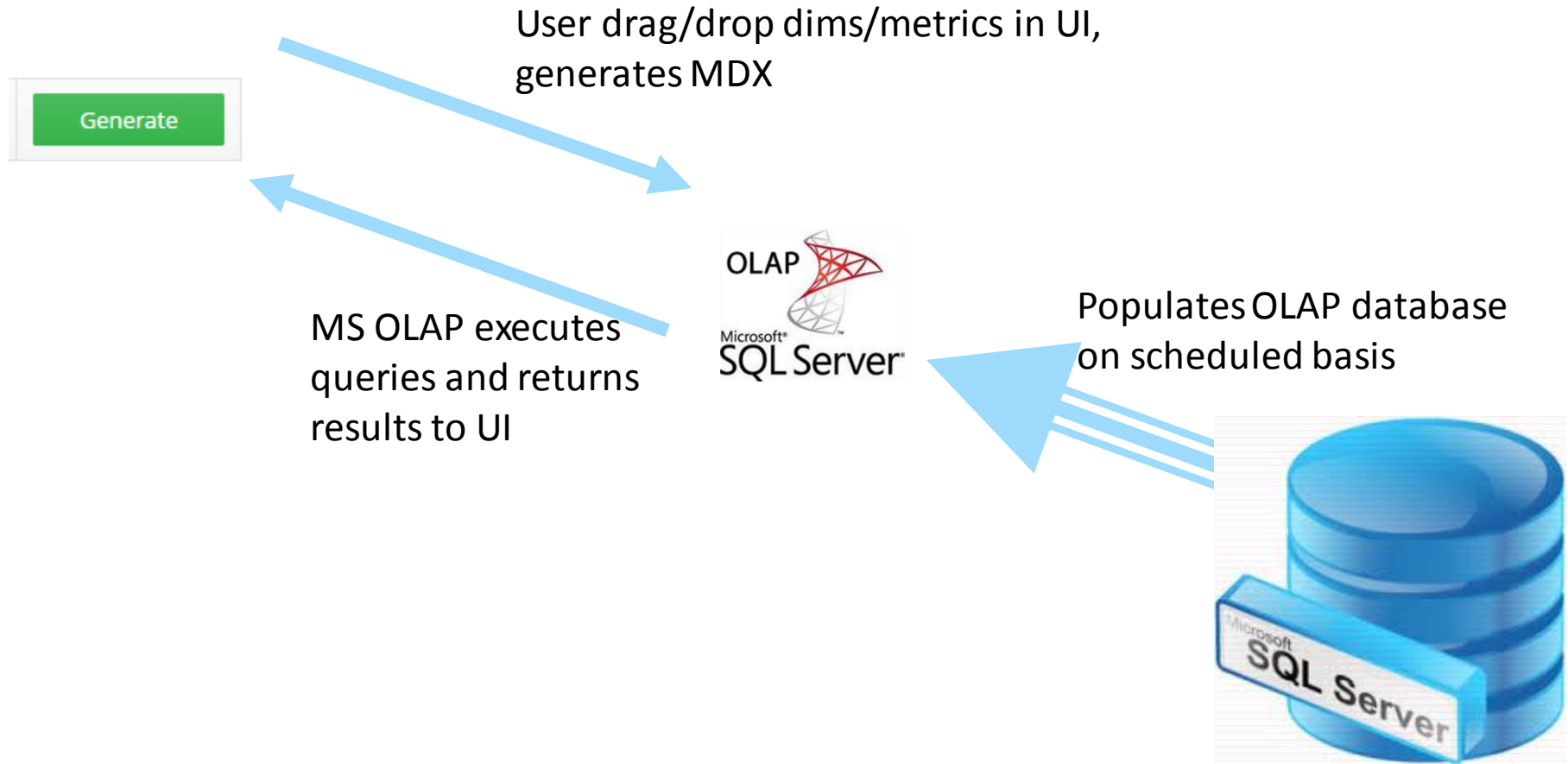
# ABOUT

- Years in SW development: 10+
- Years in database world: 10+
- Years in BI/analytics world: 10+
- Number of accidentally dropped tables on production: 0
- Best time in 10km run: 45:30
- Number of children made: 3

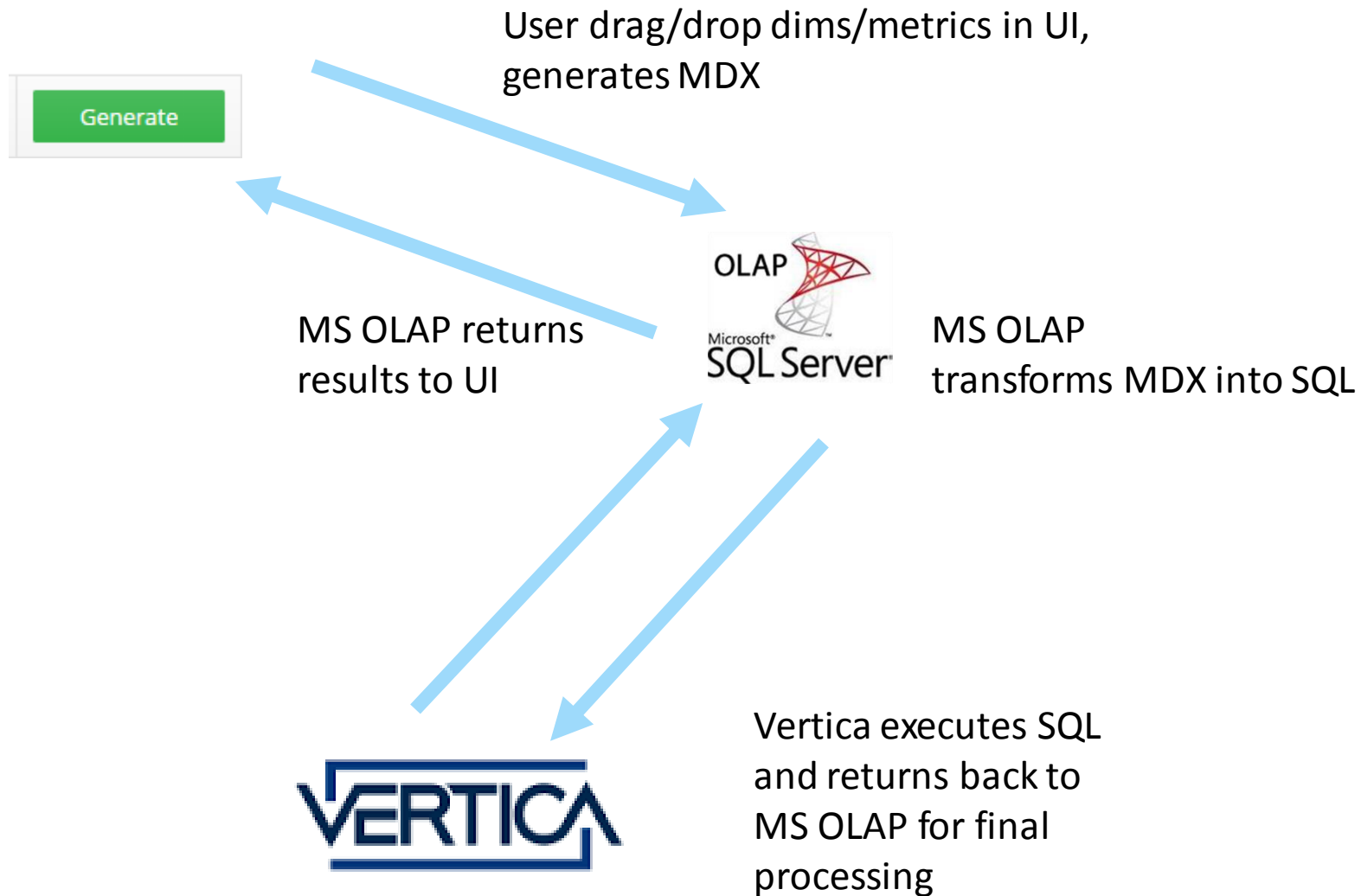
# WHAT IS VERTICA



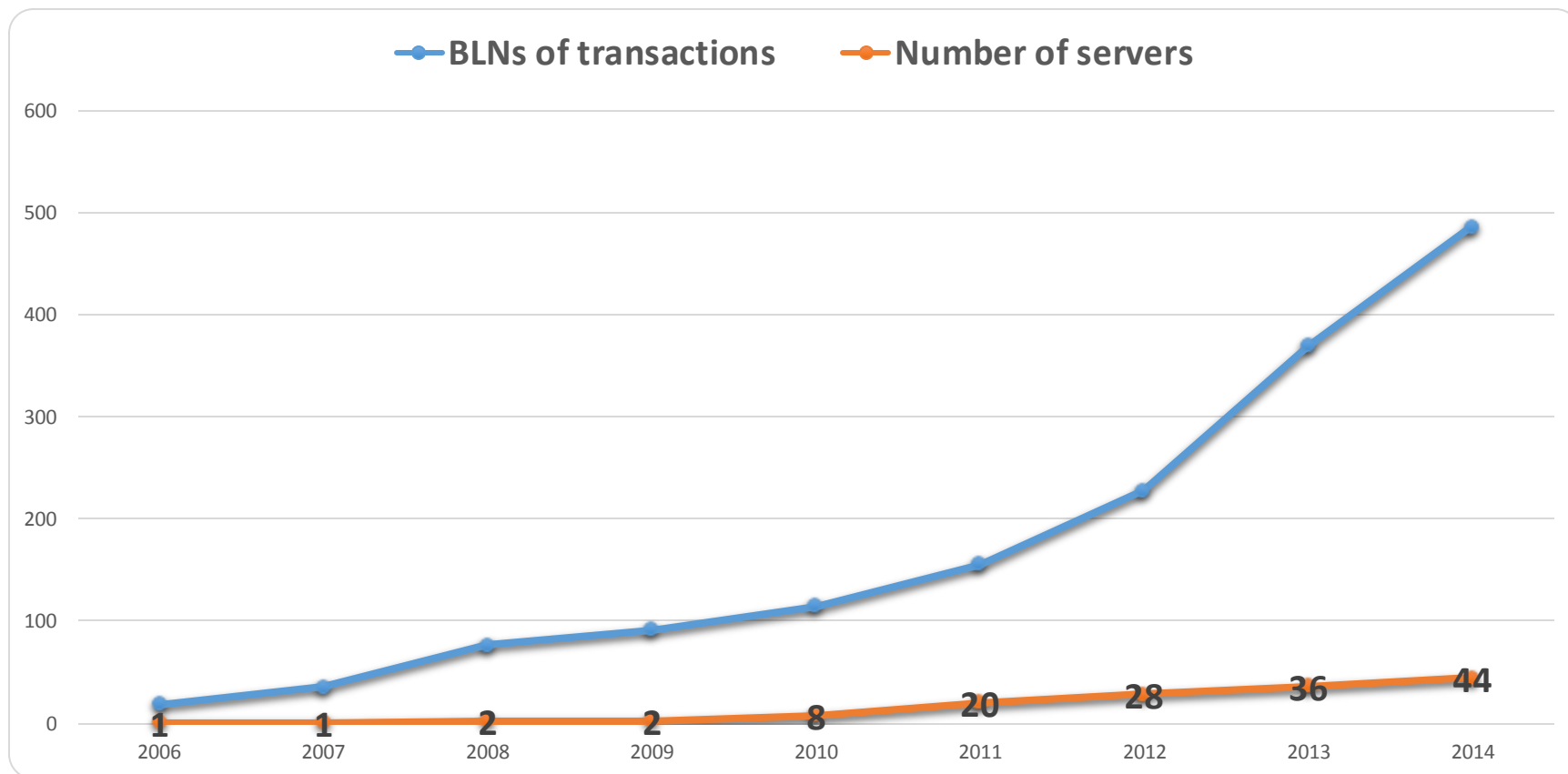
# WHAT IS [MICROSOFT] MOLAP



# WHAT IS [MICROSOFT] ROLAP

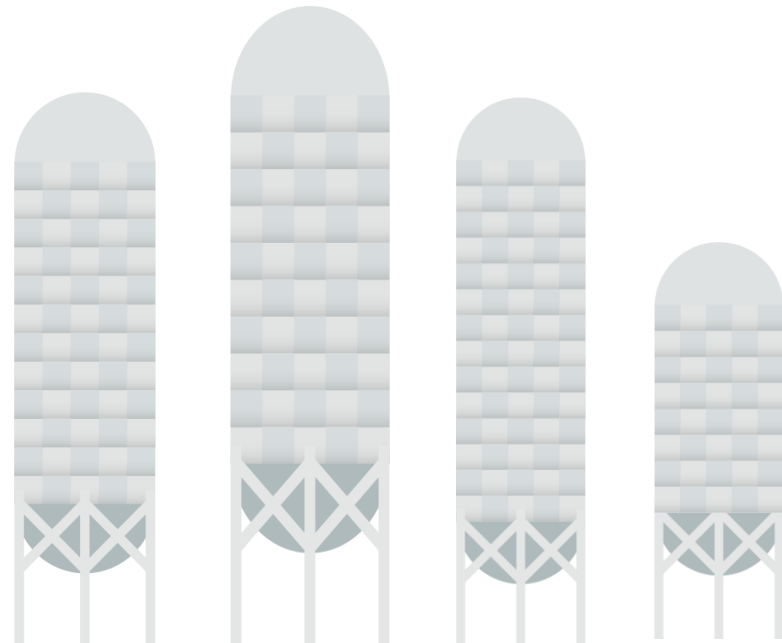


# ADFORM: BLNS OF TRANSACTIONS PER YEAR

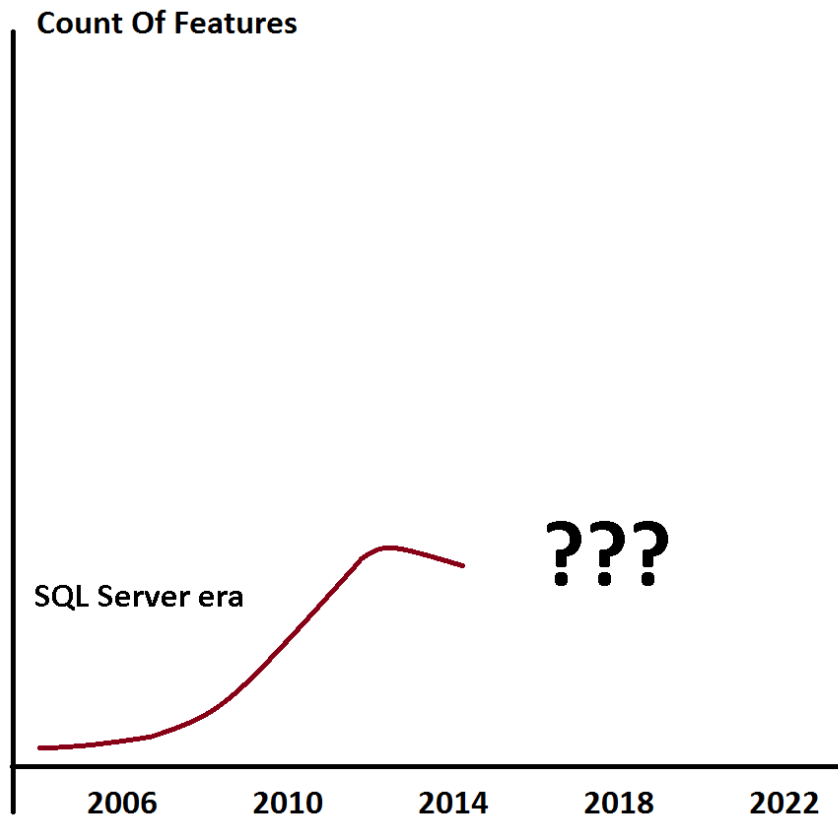


# SILO BASED ARCHITECTURE

- Scaling
- Development cost
- Licenses
- Support
- Cross SILO selection



# WHERE WE WERE 2 YEARS AGO





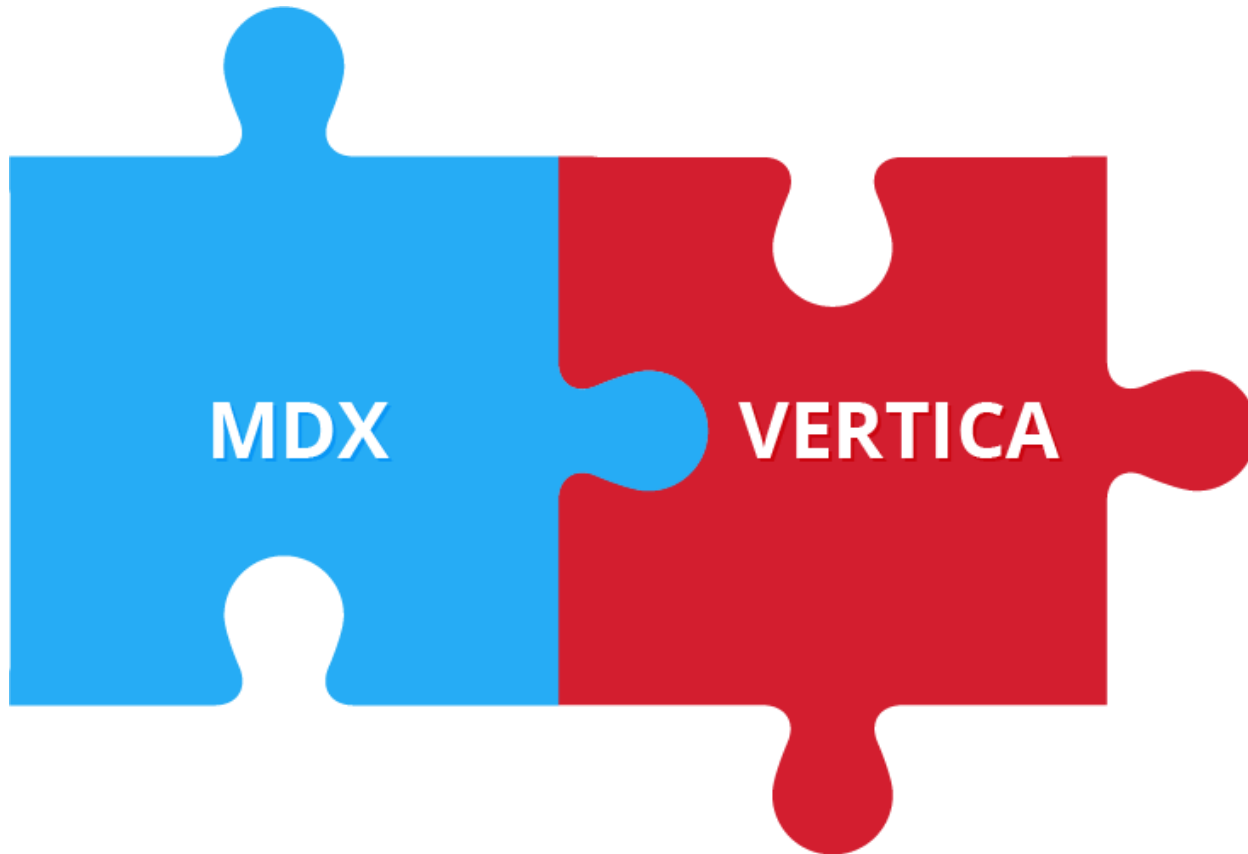
# OUR NEEDS FOR BIG DATA ANALYTICS

- Easy scaling(plug and play)
- Query performance up to moments
- Accept and execute MDX queries
- No downtime window
- Near real time latency
- Inessential data loss and data discrepancies
- Shorter time to market

# OUR NEEDS FOR BIG DATA ANALYTICS



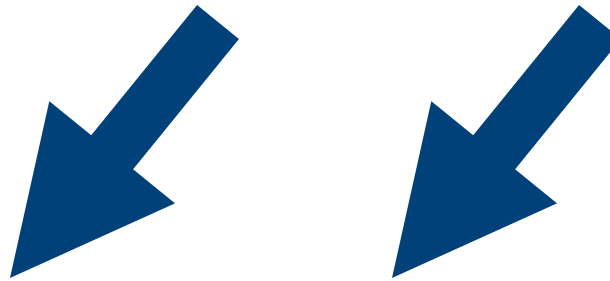
# SSAS MDX + VERTICA = ROLAP



# MDX SAMPLE

“Days” table

“Unloads” table



MDX

```
SELECT{[Measures].[impressions - days], [Measures].[Impressions UL - Unloads]} ON 0 ,  
{[Time].[Date].&[2015-05-01]:[Time].[Date].&[2015-05-31]}  
ON 1  
FROM main
```

SQL

```
SELECT  
SUM("OLAPCubes_v_rolap_TransactionDays"."Impressions") AS "OLAPCubes_v_rolap_TransactionDaysImpressions0_0"  
...  
,"OLAPCubes_v_rolap_TransactionDays"."LogTime" AS "OLAPCubes_v_rolap_TransactionDaysLogTime0_12"  
FROM "OLAPCubes"."v_rolap_TransactionDays" AS "OLAPCubes_v_rolap_TransactionDays"  
WHERE "OLAPCubes_v_rolap_TransactionDays"."LogTime">=  
'2015-05-01 00:00:00'  
AND "OLAPCubes_v_rolap_TransactionDays"."LogTime" <=  
'2015-05-31 00:00:00'  
GROUP BY "OLAPCubes_v_rolap_TransactionDays"."LogTime";
```

---

```
SELECT  
SUM("OLAPCubes_v_rolap_TransactionUnloads"."ImpressionsUL") AS "OLAPCubes_v_rolap_TransactionUnloadsImpressionsUL_0"  
...  
,"OLAPCubes_v_rolap_TransactionUnloads"."LogTime" AS "OLAPCubes_v_rolap_TransactionUnloadsLogTime0_12"  
FROM "OLAPCubes"."v_rolap_TransactionUnloads" AS "OLAPCubes_v_rolap_TransactionUnloads"  
WHERE "OLAPCubes_v_rolap_TransactionUnloads"."LogTime" >=  
'2015-05-01 00:00:00'  
AND "OLAPCubes_v_rolap_TransactionUnloads"."LogTime" <=  
'2015-05-31 00:00:00'  
GROUP BY "OLAPCubes_v_rolap_TransactionUnloads"."LogTime";
```

# MYTH #1



Insufficient ROLAP implementations

# GOOD NEWS ABOUT TRANSITION

ROLAP



MOLAP



SELECT

```
{[impressions - days], [Impressions UL - Unloads]} ON 0  
, {[Time].[Date].&[2015-05-01]:[Time].[Date].&[2015-05-31]} ON 1  
FROM main
```

# MYTH #2



SQL queries generated by ROLAP engine  
are inefficient

# THINGS TO CONSIDER FOR OPTIMIZATION

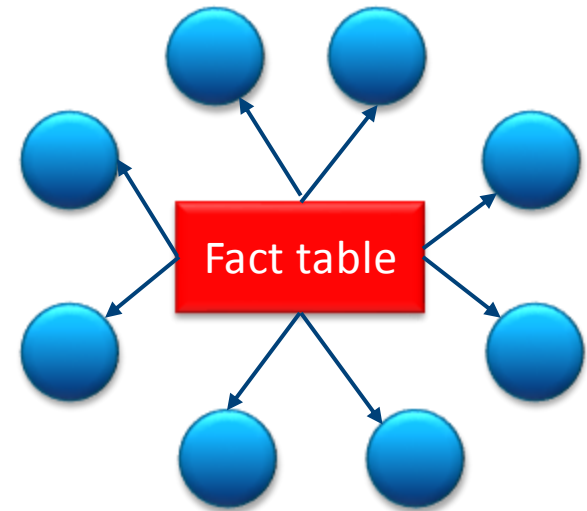
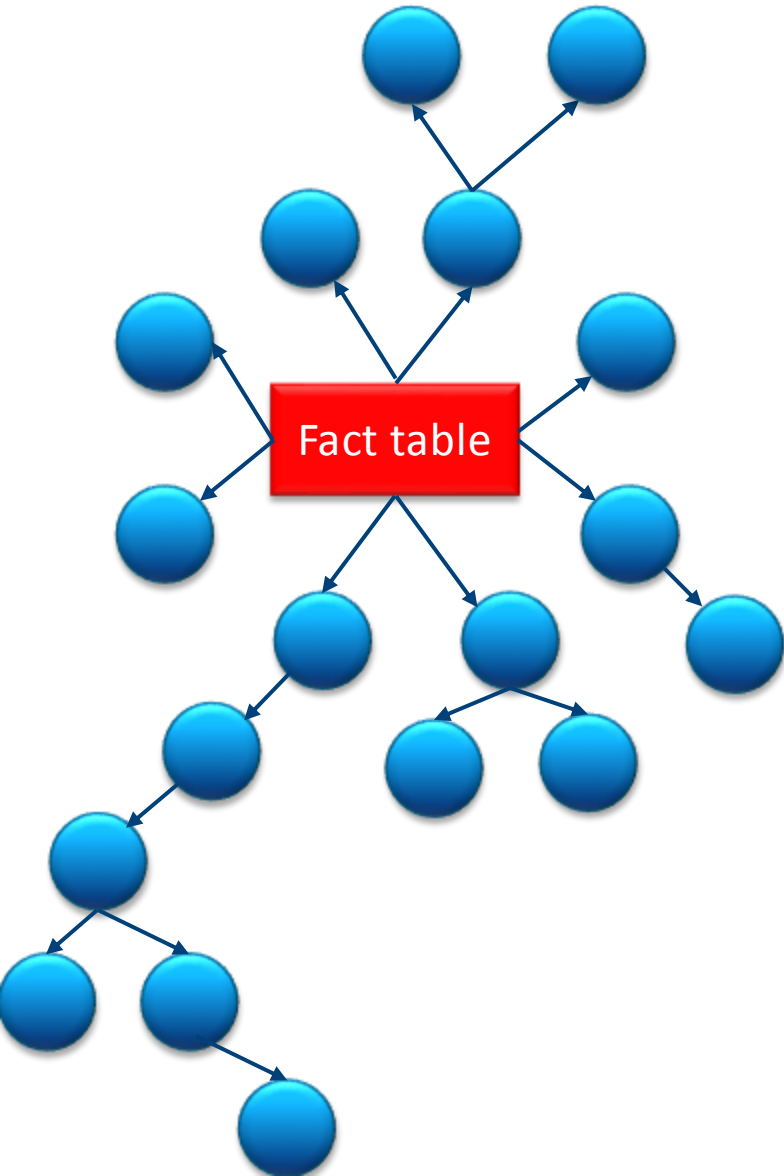
- OLAP Data Model
- Cartridge file
- Use correct version of MS OLAP
- Dimensions in MOLAP



# THINGS TO CONSIDER FOR OPTIMIZATION

- Unsegmented tables in Vertica for dimensions to avoid broadcast
- Resource Pools
- Avoid using expensive updates/deletes, use partition switching instead

# MATERIALIZATION



# MYTH #3

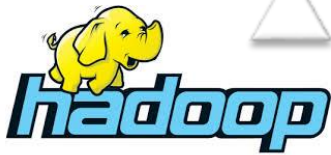


ROLAP is always slower than MOLAP

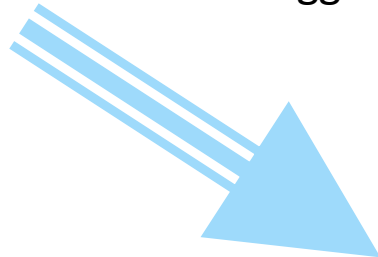
# CURRENT SITUATION

- Avg Performance 2 sec
- Throughput – up to 15 K queries per hour
- Licensed Volume – 100 TB
- Load latency – 10 minutes
- MDX and MS OLAP to generate and run queries
- 32 fact tables; 250 dimensions ; 190 metrics
- Biggest table 100 bl of rows

# DATA LOADING



Hadoop produces deltas;  
Deltas are stored in Hadoop for backups;  
Aggregates deltas for past periods



Vertica imports deltas every 10 mins;  
Aggregates deltas for past periods



OLTP databases  
replicates dimension  
changes to Vertica



OLAP imports  
dimensions every 3  
hours;

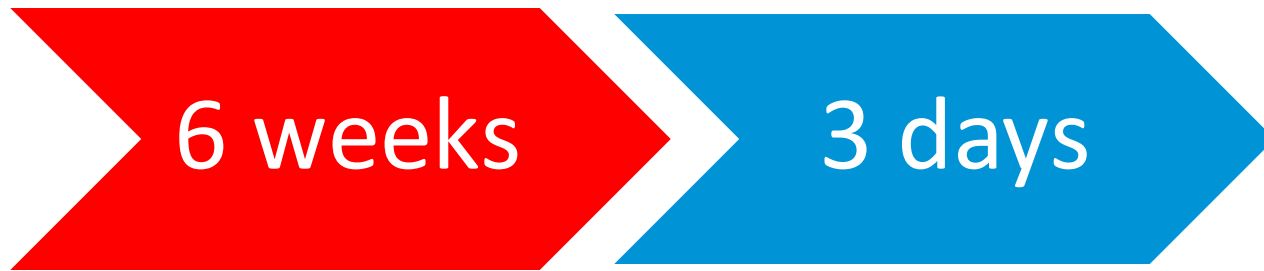


---

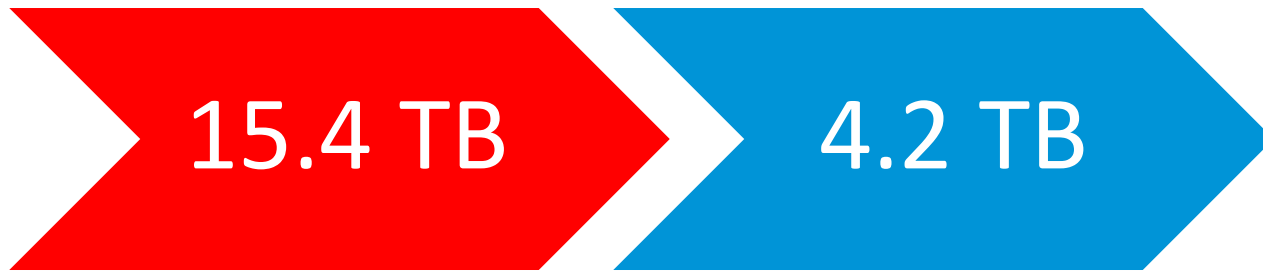
**HAPPY END**

---

# MOLAP VS ROLAP: ADDING 4 NODES TO CLUSTER

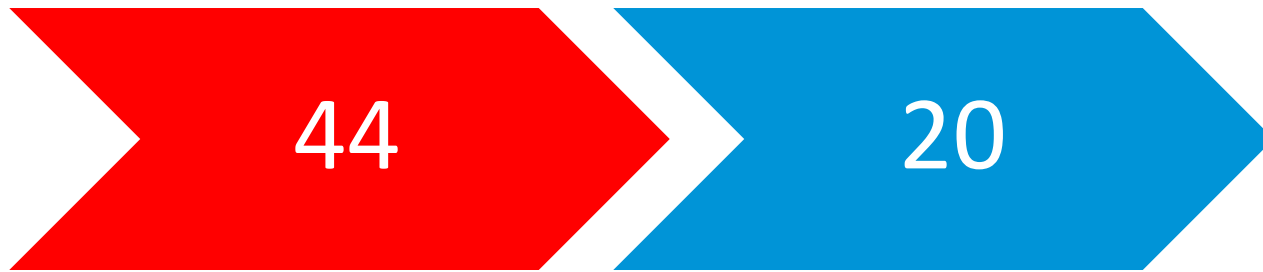


# MOLAP VS ROLAP: TABLE 35 BLNS OF ROWS

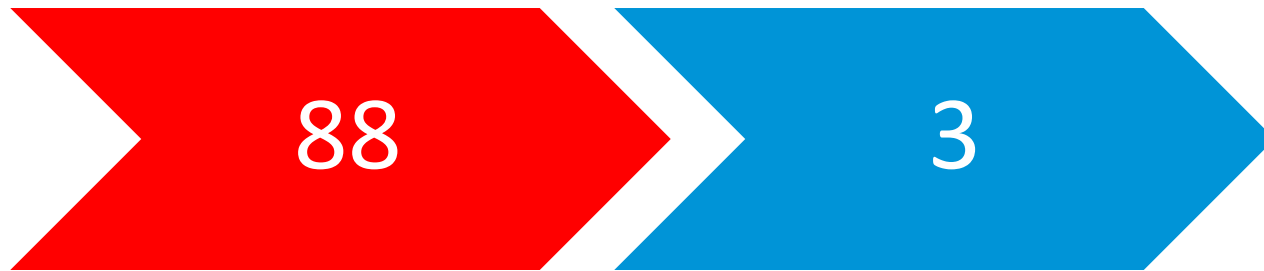




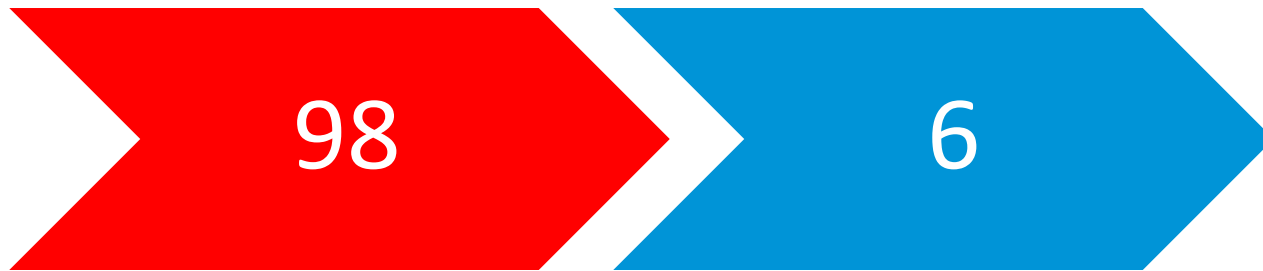
# MOLAP VS ROLAP: NUMBER OF MACHINES



# MOLAP VS ROLAP: LINES OF CODE FOR TYPICAL RELEASE



# MOLAP VS ROLAP: WORKING HOURS ON NON PEAK HOURS (WEEKENDS+NIGHTS) PER YEAR





# LESSONS LEARNED

- Rolap solution is only for transition period
- Data latency
- Additional layer: development costs
- Additional layer: licenses
- Additional layer: adds some time transferring queries and data
- ROLAP MDX always slower than SQL



# NEXT STEPS

Fully eliminate MDX and rely on SQL

---

**NOW IS REALLY HAPPY END**

---

adform

---

Questions?

---

@AdformInsider