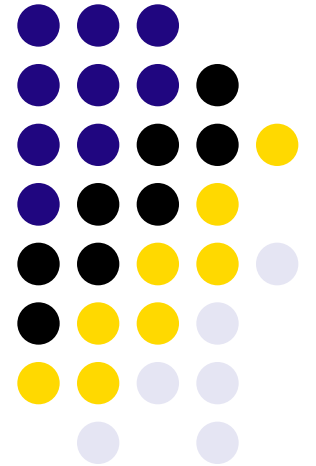
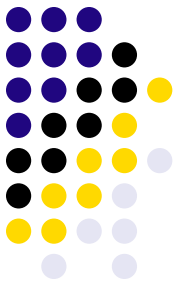


# Growing a Data Warehouse to 50 Terabytes and Beyond

Hüsnü Şensoy

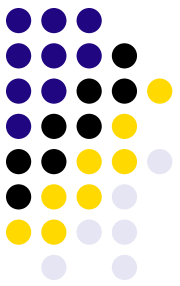


# About Turkcell



- Leader GSM operator of Turkey established in February 1994.
- Third GSM operator in Europe in terms of subscriber (~36 million).
- First and only Turkish company ever to be listed on **New York Stock Exchange**.
- Member of Board of Directors of **GSMA** since 2003.
- 25th company of **INFOTECH 100** list.



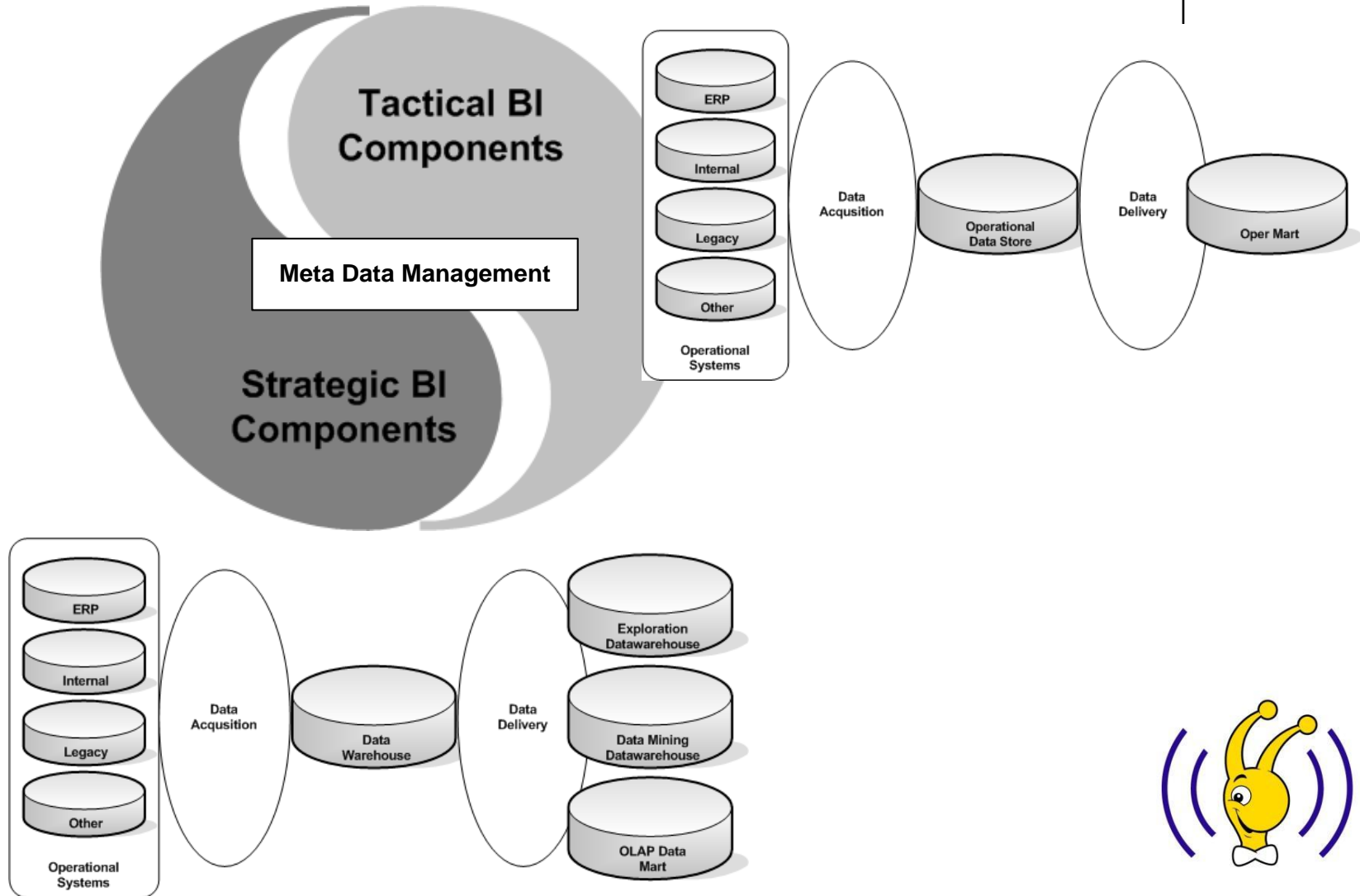
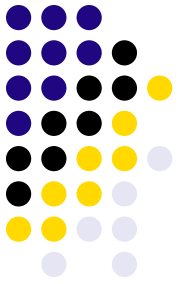


# About Me

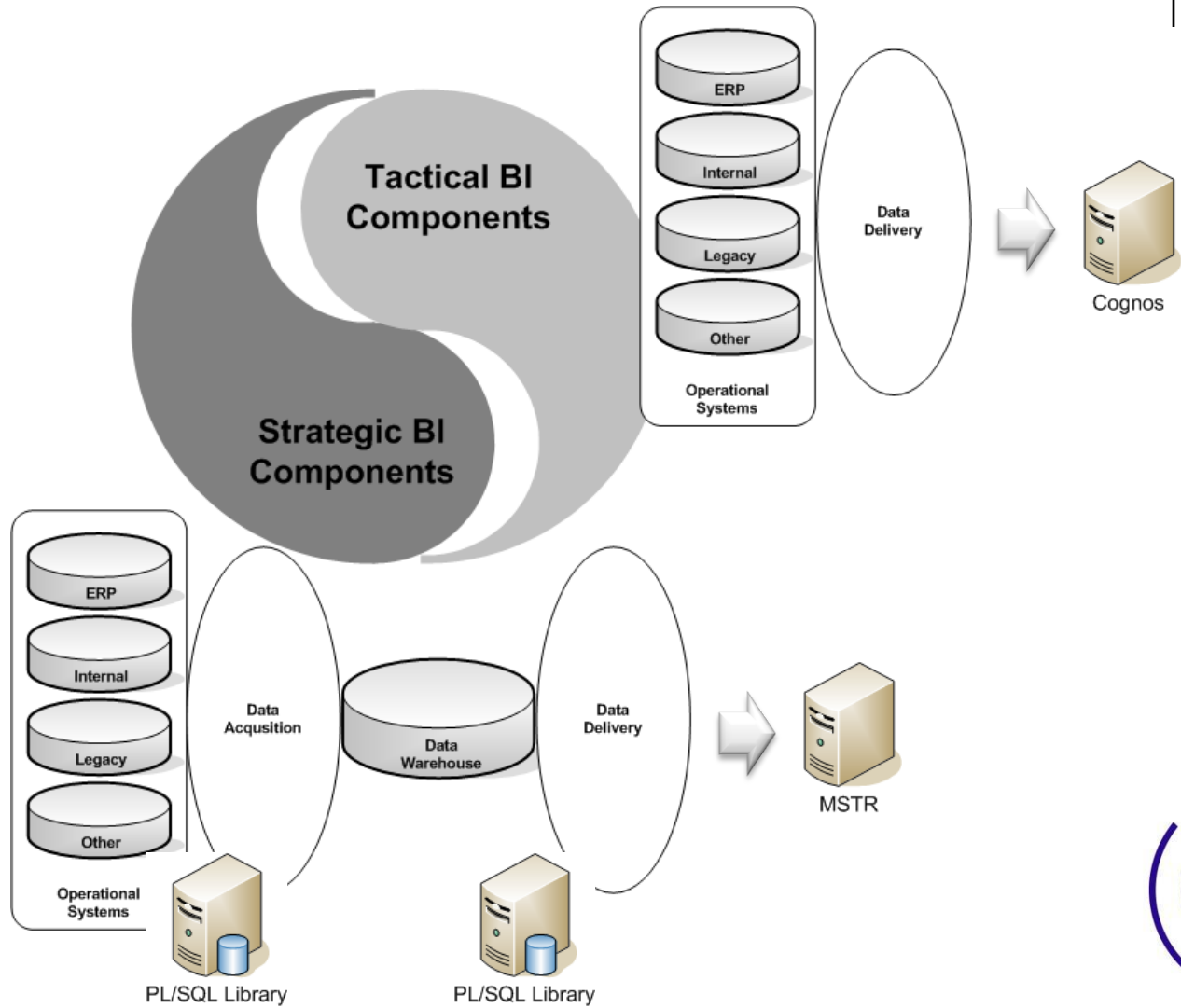
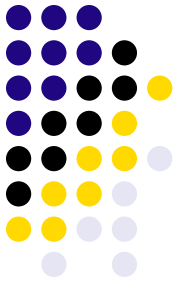
- Previously working as an Oracle developer and development team leader.
- Currently working as a VLDB DBA in Turkcell BIS Team.
- Previously studying **Machine Learning Theory** and currently studying **I/O Scheduling on Grid Architecture** in Bogazici University.
- Mainly interested in
  - ASM
  - Backup & Recovery of VLDB
  - Oracle Physical Design of Data warehouses



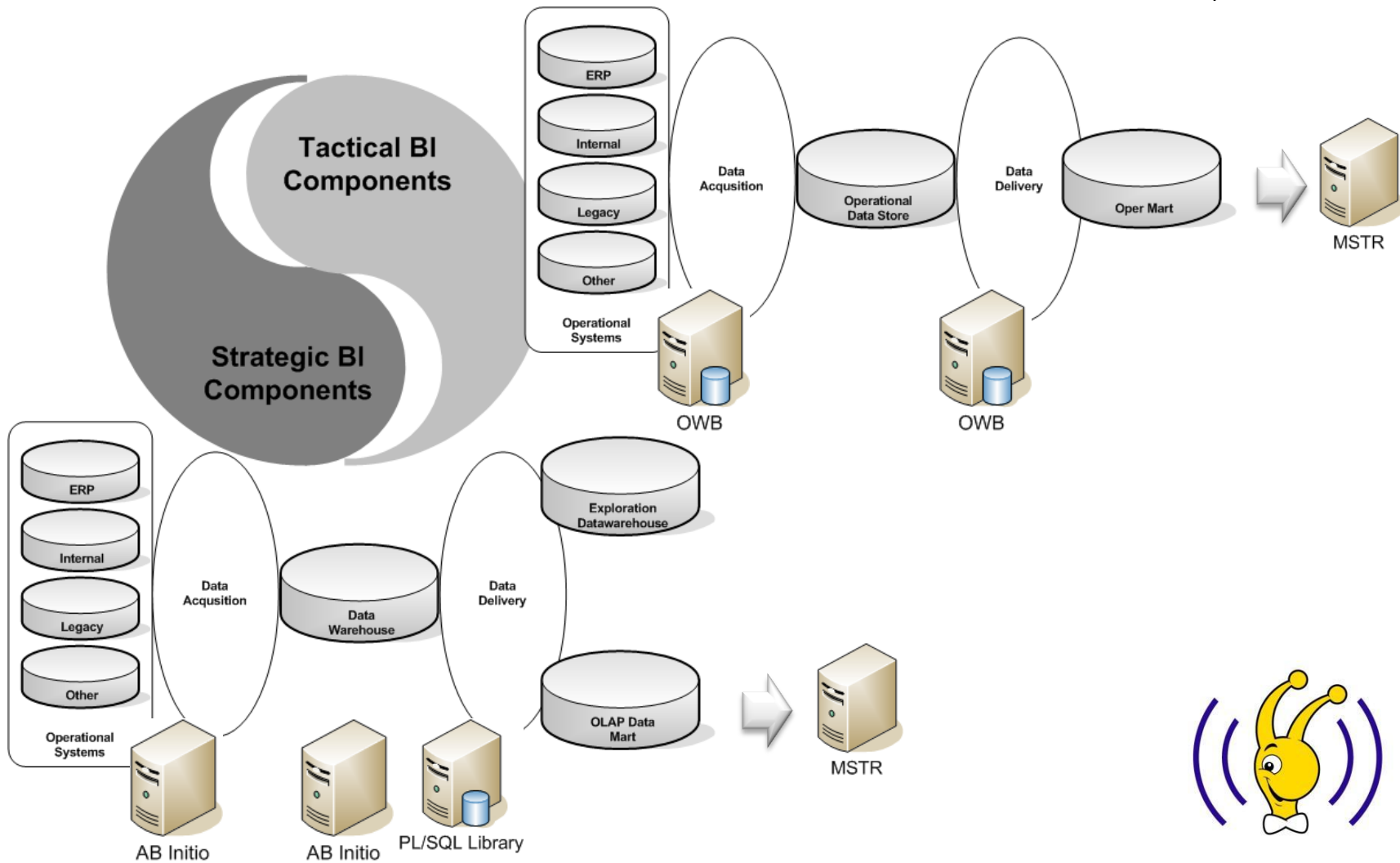
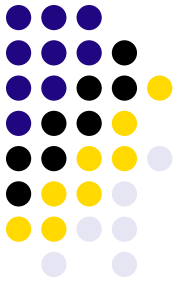
# An Efficient BI Architecture



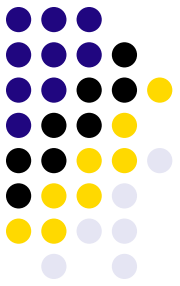
# Previous Picture In Turkcell



# New Picture In Turkcell



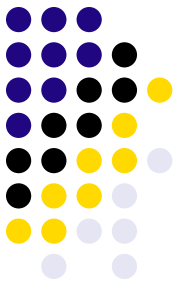
# Implementing Those in Oracle



- Presentations, ER diagrams, and sketches for explaining logical design are usually beautiful things in obtaining customer satisfaction.
- Mapping what they want in logical design to physical design is a totally different view point (Sometimes really a pain). Paradigm is
  - What customer thinks hard is usually easy
  - What customer thinks easy is usually hard
- What is included in this presentation is a subset of physical design best practices of Turkcell data warehouse reengineering project.



# I/O



## IN this presentation

- Storage structure we have used in Turkcell datawarehouse
- DBA Weaponary in Turkcell Datawarehouse
- Backup & Recovery in Turkcell Datawarehouse

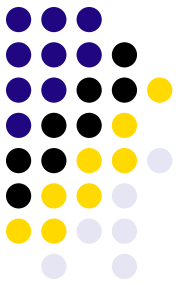
## OUT of this presentation

- Partitioning
- Compression
- Materialized Views
- Bitmap Indexes
- ...

But feel free to talk about those and others also after the presentation...







Growing a Data Warehouse to 50 Terabytes and Beyond

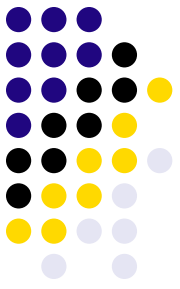
# STORAGE

**E-mail** : [husnu.sensoy@gmail.com](mailto:husnu.sensoy@gmail.com)

**Blog** : <http://husnusensoy.wordpress.com>



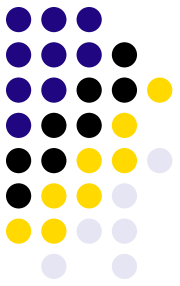
# Previous Volume and File System Layer



- Previous Turkcell BIS systems all built on Veritas VX & FS.
- Veritas VX & FS are really a good products in terms of
  - Availability (File System Check, Snap, Check Point, etc.)
  - Manageability (Dynamic file system sizing, soft striping, etc.)



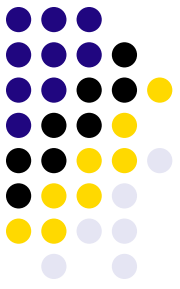
# But Veritas and others...



- They are not Oracle aware. They don't bother what is the I/O characteristics difference between a redolog and a data file. Even they don't bother the difference between a binary and a data file. So UNIX/Database administrator should do manual work to create those distinctions.
- **File System Boundary** is usually a pain in neck for Oracle VLDB DBAs.
- Every month we add more and more disks to our storage pool. But data has to be spread over disks by manual work to ensure homogenous I/O.



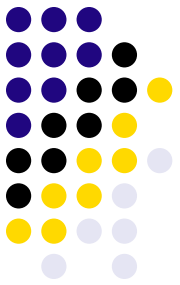
# We have decided to use ASM because...



- ASM is a totally Oracle aware product. Internally it is capable of making the distinction between different Oracle files.
- The storage boundary in ASM is the *Disk Group* itself, which means there is no boundary as long as you have more disks to add.
- Once you have added disks to ASM, it automatically rebalances the storage extents among disks and ensures I/O homogeneity over disks within the *Disk Group*.
- All those additional benefits are totally **FREE**...



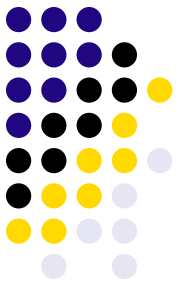
# Myths about ASM



- ASM is not as high available as other products.
  - For 10 months we haven't faced single crash related to ASM instance failure in our any production/test/development systems.
  - Currently the best option for RAC.
- Backup is a problematic issue with ASM.
  - You can use RMAN, or disk vendor snapping tools with ASM (BCV, SRDF, etc.).
- ASM has relatively moderate performance with compared to other products.
  - As far as I have tested before putting it to production, ASM is usually 10% better (sometimes equal) than a Veritas system (ODM enabled).
- ASM is not easy to manage.
  - It has been 10 months but I don't still know many ASM commands by hard.
- Once you start to use ASM, you can not use any other file system or volume manager
  - Oracle can work in peace with ASM and other file systems together.



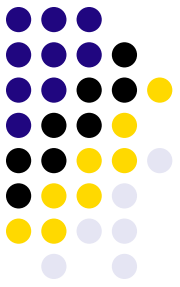
# Things to be Careful in ASM Implementation



- Are you sure that your hardware gives the desired throughput ?
  - Striping configuration (if any)
  - RAID configuration (if any and external).
  - Storage array cache
  - SAN/DAS/NAS configuration
- Do you have a multipathing product ? You **SHOULD** have one.



# Ensuring the Limits of Your Hardware. But how ?!?



## ORION

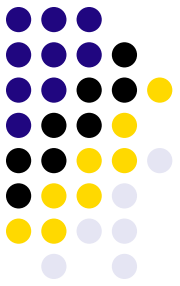
- An advanced tool developed by Oracle capable of emulating Oracle I/O by using native OS I/O libraries used by Oracle itself.
- Allows defining
  - I/O Characteristics
    - Random
    - Sequential
    - Mix
  - Array Cache Size
  - Read % + Write %
- Very easy to use

## 11g Calibration

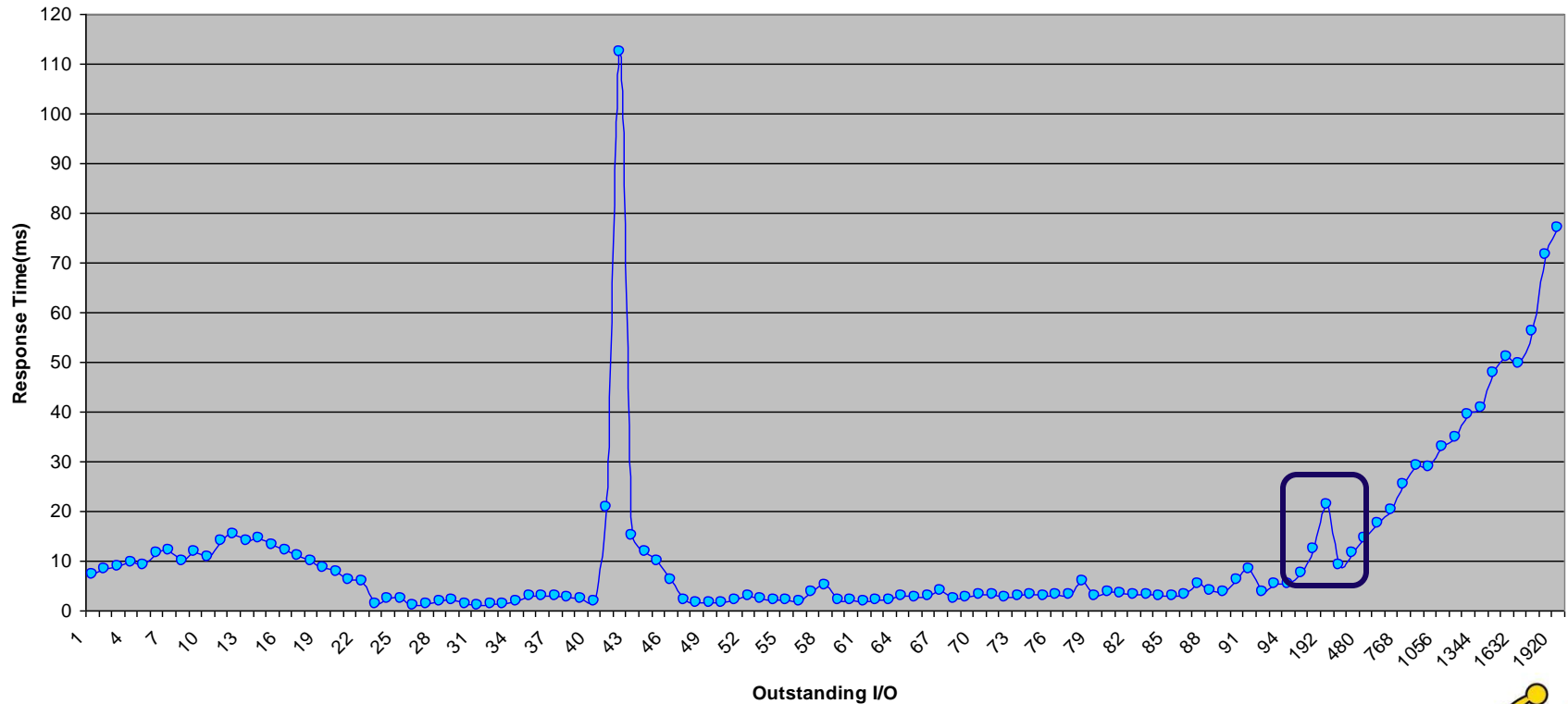
- Has less flexibility.
- Performs large sequential and small random read IO to compute
  - Max IOPS
  - Max Throughput
- Requires Oracle database installation
- Fully integrated within EM



# What will those tools will present ?

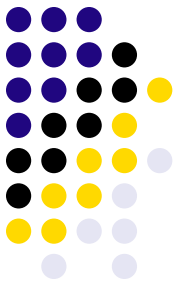


32K Random Read

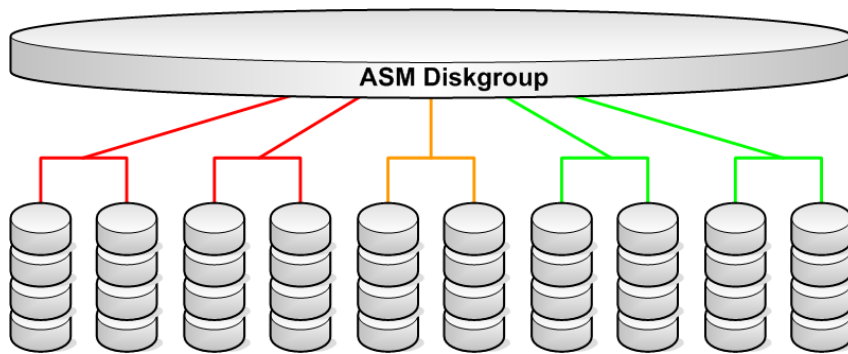




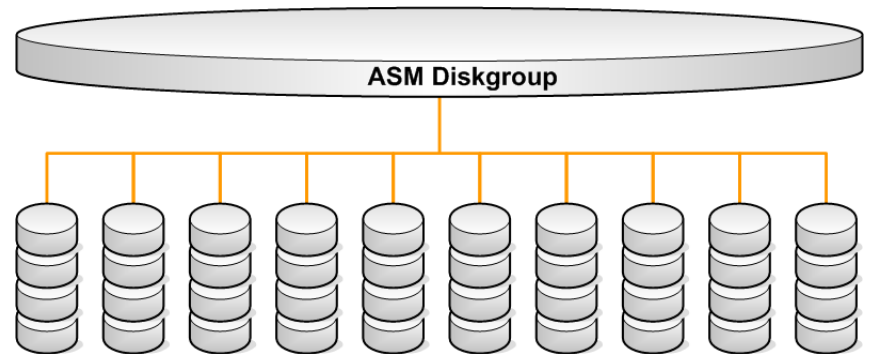
# Multipathing



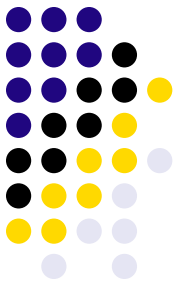
## Multipathing Disabled



## Multipathing Enabled



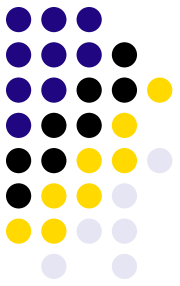
# Multipathing



- So far we have used
  - *EMC Powerpath* with *10g ASM* on *Solaris 5.10*
  - *Veritas DMP* with *11g ASM* on *HP-UX 11.31*
  - *HP Native Multipathing* (not in compatibility matrix yet) with *11g ASM* on *HP-UX 11.31*
- All are good. All are performing well. There are only some configuration nuances.

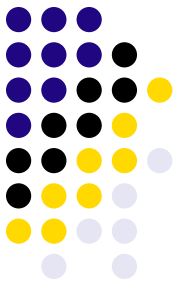


# Storage Summary



- If you are already a large database or planning to grow, start thinking about ASM.
- Test everything to prove to yourself and your managers that many negative gossips about ASM are indeed myths.
- Know about the limits of your hardware and never expect ASM to do better than that limit.
- Oracle currently doesn't provide a multipathing product for UNIX systems. So use storage vendors' products for
  - Availability
  - Performance





Growing a Data Warehouse to 50 Terabytes and Beyond

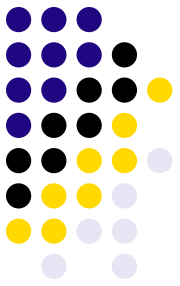
# ORACLE WEAPONARY OF VLDB DBA

**E-mail** : [husnu.sensoy@gmail.com](mailto:husnu.sensoy@gmail.com)

**Blog** : <http://husnusensoy.wordpress.com>



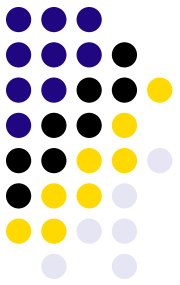
# DBA Weaponary



- There are many features in Oracle databases targeting VLDB administration.
- I will be pointing out some of them that can be used in your daily operations.



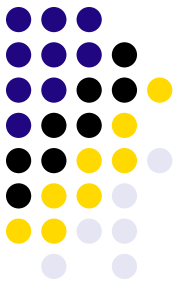
# BIGFILE Tablespaces 😊



- Once the size of your data warehouse grows, one of your hobbies become **“add one more data file”** game.
- That’s because the maximum size of a *SMALLFILE* tablespace data file is limited with your block size, but the same thing doesn’t hold for growing data amount.
- You can strictly decrease the number of data files in by *BIGFILE* tablespace usage. This will decrease your
  - Operational management cost.
  - Most likely increase your backup/restore speed for especially tape backups.



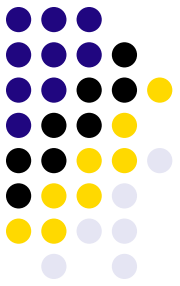
# Pitfalls about BIGFILE Tablespaces ☹️



- Remember that you can have one and only one data file under a *BIGFILE* tablespace. If you are using a file system, once a *BIGFILE* tablespace reaches the file system boundary, you are in trouble !!!
  - So use ASM 😊
- In 10g, your backup durations may be dominated by the size of even a single *BIGFILE* tablespace size independent of the number of channels.
  - In 11g, “*Parallel Backup & Restore for Very Large Files*” feature removes this problem.



# Resumable Operations 😊

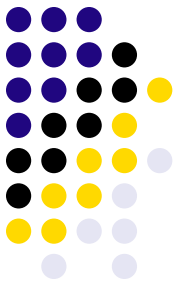


- At the final phase of ETL, load operations some times may not be completed due to
  - Quota restriction on users.
  - Tablespace sizes.
  - MAXEXTENT segment parameters.
- After a 3 hours long data load, would you like your load to be interrupted?!?
- Resumable operations allow Oracle to pause (not stop) operations for a defined amount of time in case that operation fails due to above reasons.
- By 10g this feature becomes a DBA feature rather than just a developer feature.
  - RESUMABLE\_TIMEOUT server parameter.





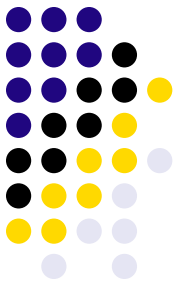
# Pitfalls about Resumable Operations ☹️



- Be sure that your ETL administrators or users are aware of the existence of this option. Otherwise you will receive some phone calls saying  
*“What’s the problem of production system ? It has been 5 hours since we’ve started to load the database and it is still not completed. Previously, it just took 3 hours”*
- They can just follow the changes in **user\_resumable** view about the current status of their loads.



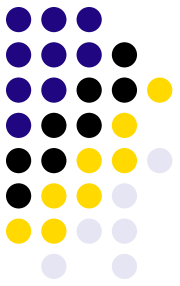
# Resource Manager



- Data warehouse  $\equiv$  ***Parallelism*** & **Aggrasive CPU Usage.**
- Data warehouse customer  $\equiv$  A bunch of people want  $\infty$  DOP and  $\infty$  CPU power.
- Everybody desire for all available CPU capacity but even in our large data warehouse we have only 104 CPUs.
- You need a customizable way of distributing available CPU power in order to avoid  $\infty$  CPU run queues.



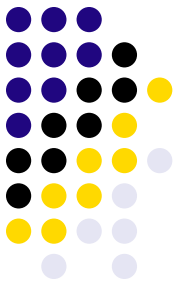
# What Resource Manager can “mainly” Do for You ?



- Distribute available CPU quantum in a hieracycal manner.
- Limit the number of parallel processes for a statement executed by some group of sessions.
- You can distinguish sessions according to their different session attributes. Such as,
  - Login user
  - Action
  - Module
- Change the quantum usage percentage if some operation takes longer than a certain time period.
- Automatically **snipe** idle sessions.



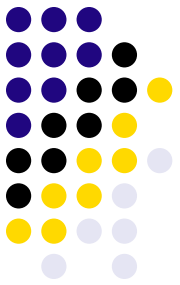
# Pitfalls about Resource Manager ☹️



- Ensure that you have configured “Idle Time” option carefully otherwise Oracle can snipe:
  - An RMAN session
  - A SQL\*Loader Routine
- You can never ensure the initial distribution of CPU quanta are perfectly good. Use “Monitor” option in Enterprise Manager to see who is using and waiting for CPU. So you can fine tune Resource Manager plans.



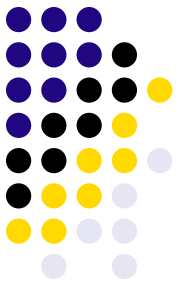
# Parallel Operations 😊



- For sure this is one of the most important weapons.
- In a multi-core environment (even in a single CPU) parallelization is something that will increase your throughput to a certain point.
- In Oracle you can parallel
  - Direct Path Load Data
  - Build indexes
  - Refresh Materialized Views
  - ...



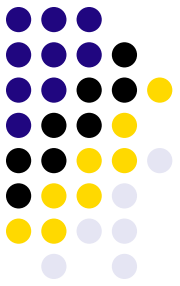
# Pitfalls about Parallel Operations ☹️



- Your strongest weapon can shoot you.
- Keep in mind that parallelism is not something scalable in a multi user environment. Most probably some customers are very unhappy when some are being satisfied.
- In order to ensure that your DOP level is calibrated, keep your eyes on CPU run queue lengths. A value less than 10 is good as a rule of thumb.

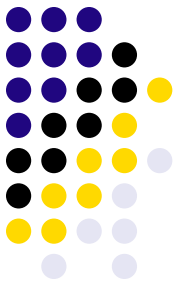


# Summary of Your Weaponary



- There are bunch of weapons you can use by 10g and 11g. Some others are:
  - Network Aware Active Database Duplicate
  - Incident Packaging Services (IPS)
  - Data Pump
  - Fine Grain Access Control
  - Grid Control
- Use them
  - Save time
  - Make your data warehouse environments more and more integrated.
- Know the **RESTRICTIONS** of your weapons, otherwise you will shoot yourself.





Growing a Data Warehouse to 50 Terabytes and Beyond

# BACKUP & RECOVERY

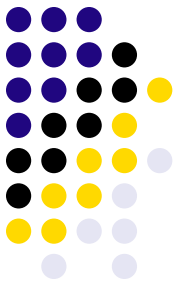
**E-mail :** [husnu.sensoy@gmail.com](mailto:husnu.sensoy@gmail.com)

**Blog :** <http://husnusensoy.wordpress.com>





# Data Warehouse Backup & Recovery



- With compared to other type of databases data warehouses have some different characteristics in terms of backup & recovery:
  - Size of data is huge so backup & recovery seems to be a problem for data warehouse databases.
  - In many data warehouses it is easy to decide for optimum backup start time.
  - A great percentage of data (above 90%) is static within 24 hours period.

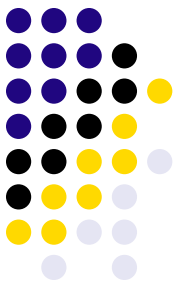




# Backup Cost Computation 1

- \$/TB cost of a first tier disk storage = \$6.000
- A high end tape driver costs \$20.000 .
- \$/TB cost of cartage = \$200
- Assume we have a 100 TB data warehouse and we have 6 tape drivers dedicate to our data warehouse
  - Cost of disk backup = \$600.000
  - Cost of tape backup = \$140.000





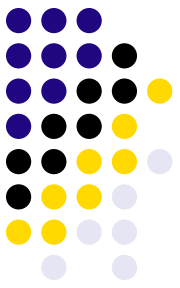
# Backup Cost Computation 2

- \$20,000 cost tape driver can reach maximum backup speed of 300 MB/s so you can reach up to 1800 MB/s in total with 6 drivers.
- 100 TB first tier storage configuration can handle write speed ~ 2-3 GB/s.

Use tape backup solution for your data warehouse it is almost as fast as disk backup and much cheaper.

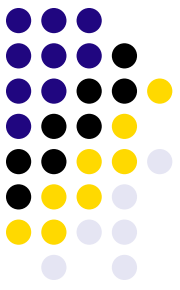


# And your production disk group has crashed one day...



- If you were using disk based backup solution you can recover your system within 1-3 minutes including the time to search recovery command syntax from google.com.
- But this is not an option, because you have a cheaper tape solution:
  - You most probably be able to restore your data warehouse in 1 day.
  - If any of  $8,79609E+14$  bits aren't corrupted (96,8 % - 100 %).



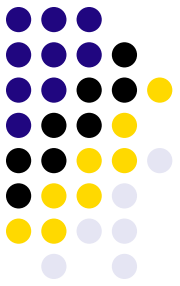


# Moral of the Story

- You don't backup to break world's fastest backup window record.
- You backup to bring your system back in case of a crash.
- So newer think from backup speed perspective first. Always ask "What will be my recovery duration ?"
- Even if your data warehouse doesn't require to be high-available, 1 day-long interrupt will not be accepted by your customers.



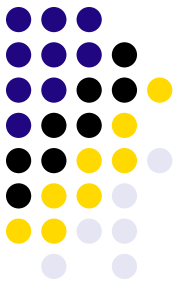
# Turkcell Way of Oracle Backup



- Daily RMAN Incrementally Updated Backup with check logical option and on a separate ASM Disk group. This allow us:
  - 15 -45 minutes backup windows/per day.
  - 1-2 minutes recovery time in case of disk group crash.
  - Blockrecovery in seconds.
- We don't backup everything
  - *STAGING\_BTS*, *INDEX\_BTS*, etc.
  - Use *RMAN* “**CONFIGURE EXCLUDE TABLESPACE**” command.



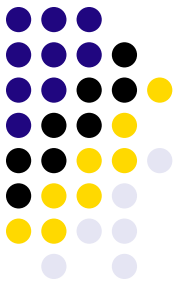
# Summary of Backup & Recovery



- Data warehouses are specialized databases in terms of backup & recovery
  - Large in size.
  - +90% static in 24h windows.
  - Mostly static after ETL processes
  - Overall they are not hard backup !!!
- New tape driver technologies are usually good in terms of backup speed.
- But your concern is not to backup, it is to recover.
- Try to push your managers towards tier-2, tier-3 even SATA technology for swift recovery.
- Use tape drivers for archiving purposes and long term backups.



# What is Next For Turkcell Data Warehouse

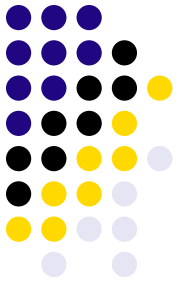


- Two upcoming big project for Turkey and Turkcell
  - Mobile Number Portability
  - 3G
- More reports, more data, more users.
- The reports desiring for faster response times.
- And we are ready with Oracle Database Technologies...





# Questions & Comments



E-mail : [husnu.sensoy@gmail.com](mailto:husnu.sensoy@gmail.com)

Blog : <http://husnusensoy.wordpress.com>

