Database Administration

José Orlando Pereira

HASLab / Departamento de Informática
Universidade do Minho

2020/2021

Scope of DBA

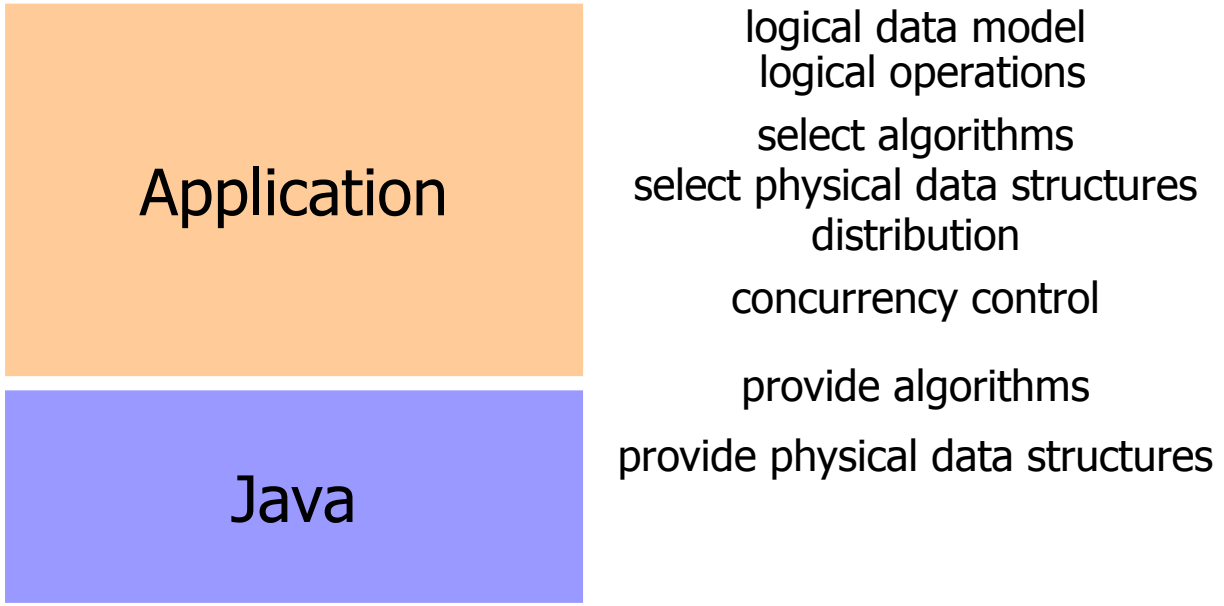
- What role? What tasks are performed?
- What skills are necessary?

Case study

- Manage invoices:
  - Add invoice
  - List invoices
- More operations:
  - List invoices for product
  - Top 10 products
- Assumption:
  - Large number of users

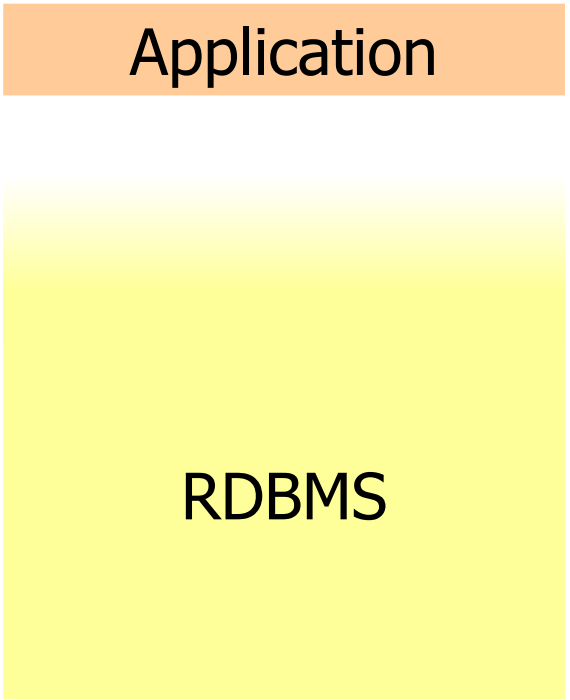
Customer X	# 123
prod. x	54
prod. y	21
prod. z	63
Total:	138

Java version

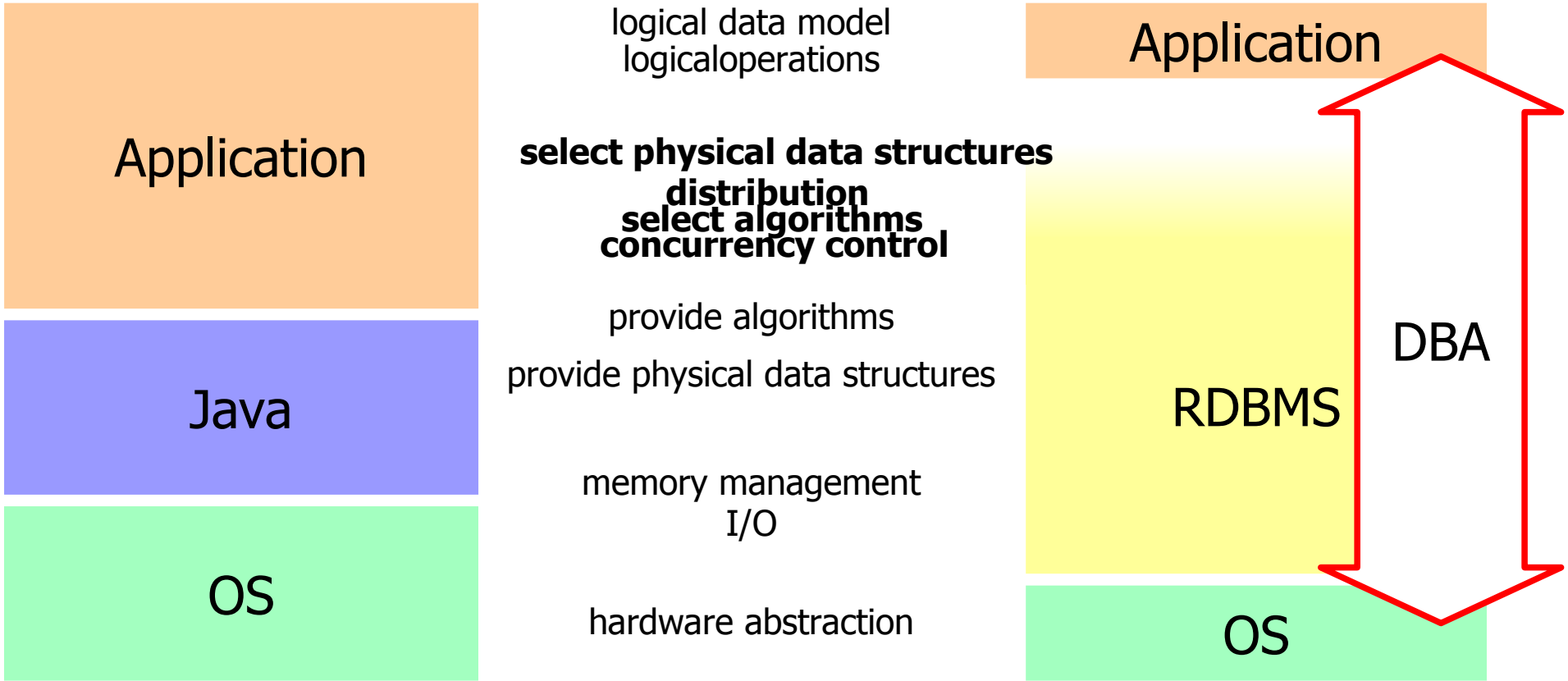


RDBMS version

logical data model
logical operations



Role and tasks



Role and tasks

- Application development:
  - Bridge the gap between the (high level) application and DBMSs
- Database operation:
  - Interaction with the OS
  - Backups, security, etc

Skills

- In-depth knowledge of DBMSs (plural!)
- Experimentation and deduction
- Ingenuity in problem solving

Roadmap

- Physical structures
- Query processing and optimization
- Isolation
- Recovery
- Replication and "sharding"
- NoSQL and NewSQL

Relational

ACID

"big data"

Warning!

The slides are not enough!

Main references

(1) H. Garcia-Molina, J. Ullman and J. Widom. **Database Systems: The Complete Book**. Prentice-Hall, 2006 (**2nd Edition**).

(2) J. Gray and A. Reuter. **Transaction Processing: Concepts and Techniques**. Morgan-Kaufmann, 1993.

(3) M. Tamer Özsu, P. Valduriez. **Principles of Distributed Database System**. Springer, 2011. (**3rd Edition**).

(4) I.Varley. **No Relation: The Mixed Blessings of Non-Relational Databases**. http://ianvarley.com/UT/MR/Varley_MastersReport_Full_2009-08-07.pdf

(5) PostgreSQL Documentation https://www.postgresql.org/docs/11/static/index.html

Main references

	(1)	(2)	(3)	(4)
Physical Structures	**13**	14		
Query Processing	15			
Indexes and Views	8, 14	15		
Query Optimization	16			
Isolation	18	7		
Recovery	17	9, 10, 11		
Repl. & Sharding			13	
NoSQL				*
Parallel Query	20.1-20.4		14	

Information

- Grading:
  - 60% written exam
  - 40% project
  - Both 8/20 minimum
- Contacts:
  - Blackboard
  - jop@di.uminho.pt
  - DI 2.16 / 253 604 449

Case study

- Tables:
  - Client: Id, Name, Address, Data(*)
  - Product: Id, Description, Data(*)
  - Invoice: Id, ProductId, ClientId, Data(*)
- Pre-populate Client and Product with 2^n items

(*) Strings with arbitrary data...

Case study

- Sell:
  - Add invoice record
- Account of a specific client:
  - names of items sold to that client
- Top 10 products:
  - 10 most sold products
- Generate client and product ids with:
  `rand.nextInt(MAX)|rand.nextInt(MAX)`

Benchmarking

- Repeat workload for a variable number of client threads
- Discard initial and final periods
- Measure:
  - Response time (duration of transactions)
  - Throughput (rate of execution)