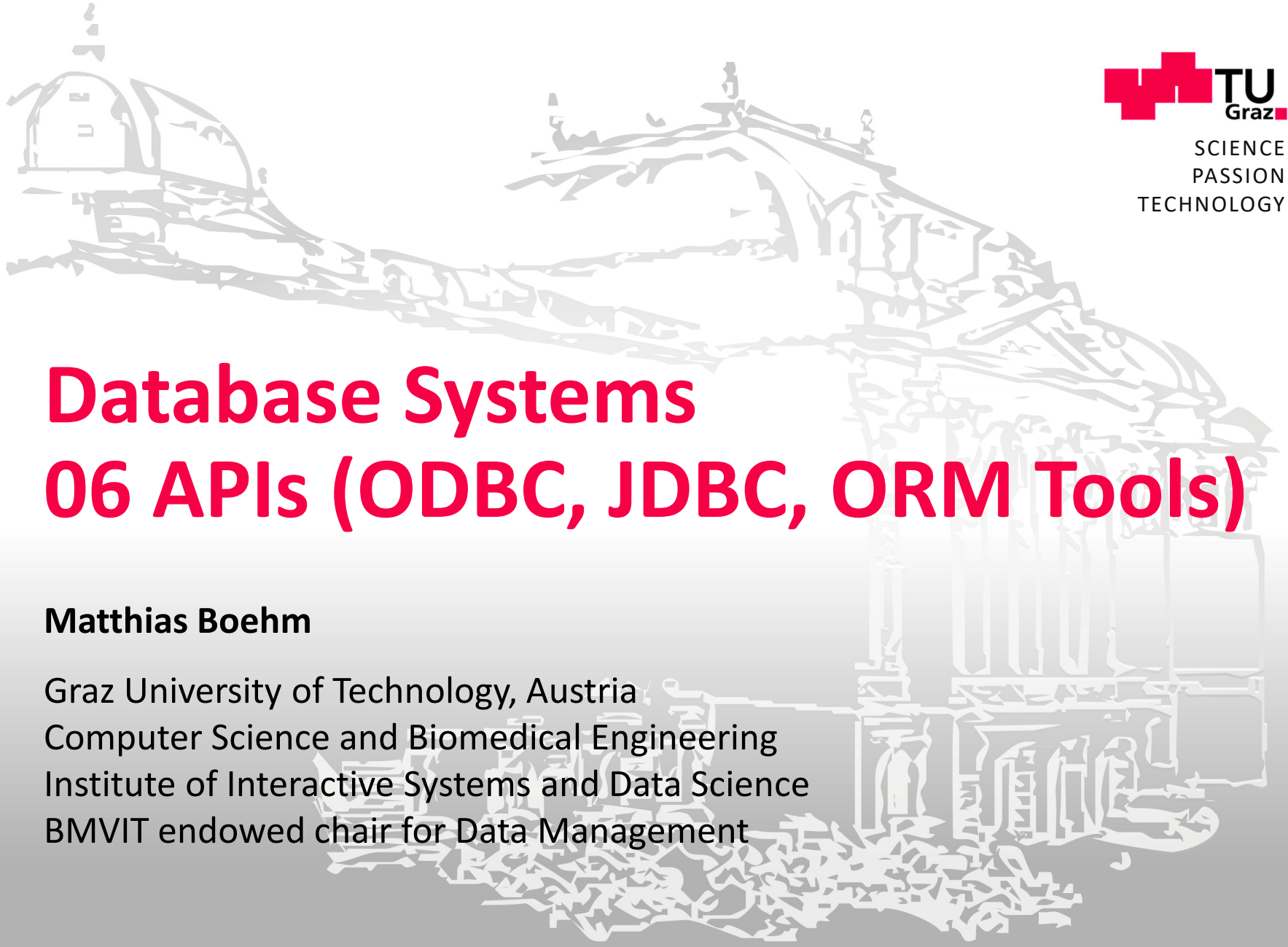


SCIENCE
PASSION
TECHNOLOGY

Database Systems
06 APIs (ODBC, JDBC, ORM Tools)

Matthias Boehm

Graz University of Technology, Austria
Computer Science and Biomedical Engineering
Institute of Interactive Systems and Data Science
BMVIT endowed chair for Data Management

Last update: Apr 08, 2019

ISDS

Announcements/Org

- **#1 Video Recording**
  - Since lecture 03, video/audio recording
  - Link in TeachCenter & TUbe (but not public yet)
  - **Issues lectures 04 and 05**

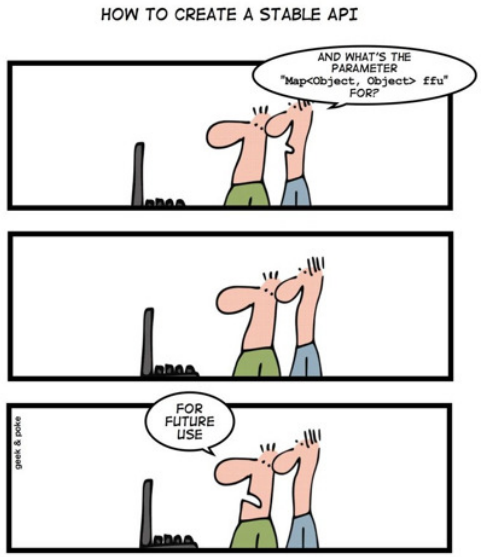
- **#2 Statistics Exercise 1**
  - Submissions so far: **261+19/400** (DB), **63+6/96** (DB1)
    ➔ **double check draft status! at least, try!**
  - Additional email submissions: **12**
  - **Last chance tomorrow EOD** (including 7 late days)

65.3%

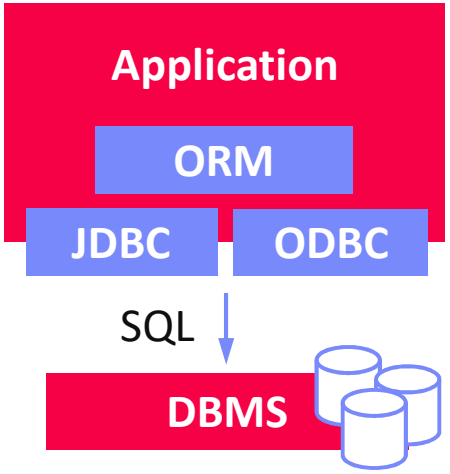
- **#3 Exercise 2**
  - Task details introduced today
  - Submission opens Apr 22, Deadline: **Apr 30 11.59pm**

What's an API and again, why should I care?

- **Application Programming Interface (API)**
  - Defined **set of functions or protocols** for system or component communication
  - Interface independent of concrete implementation ➔ **decoupling of applications** from underlying libraries / systems
  - API stability of utmost importance



- **Examples**
  - **Linux:** kernel-user space API → system calls, POSIX (Portable Operating System Interface)
  - **Cloud Services:** often dedicated REST (Representational State Transfer) APIs
  - **DB Access: ODBC/JDBC and ORM frameworks**



Agenda

- **Exercise 2: Query Languages and APIs**
- **Call-level Interfaces (ODBC/JDBC) and Embedded SQL**
- **Object-Relational Mapping Frameworks**

Exercise 2:
Query Languages and APIs

Recap: Soccer World Cup 1954-2014

- **Dataset**
  - Public-domain, derived (parsed, cleaned) from **Openfootball Worldcup Dataset**
  - Clone or download your copy from https://github.com/tugraz-isds/datasets.git
  - ➔ **Updated Apr 07 (improved data quality)**

- **Exercises**
  - 01 Data modeling (relational schema)
  - **02 Data ingestion and SQL query processing**
  - 03 Tuning, query processing, and transaction processing
  - 04 Large-scale data analysis (distributed data ingestions and query processing)

1954_2014_Squads.csv: The Squads file contains th
structure and examples look as follows.

```
#Year, Host_Country, Country, Jersey_Number,
1998,France,Austria,14,FW,Hannes Reinmayr,Stu
2014,Brazil,Germany,1,GK,Manuel Neuer,Bayern
2014,Brazil,Germany,11,FW,Miroslav Klose,Lazi
```

1954_2014_Matches.csv: The Matches file contains
and examples look as follows.

```
#Year, Host_Country, Match_ID, Type, Date, Lo
2006,Germany,572,Group A,Wed Jun/14,Signal Id
2010,South Africa,684,Round of 16,Sun Jun/27
2014,Brazil,761,Final,Sun Jul/13 16:00,Estádi
```

1954_2014_Goals.csv: The Goals file contains the g
time of the game. It's detailed structure and exam

```
#Year, Host_Country, Match_ID, Team, Player,
2014,Brazil,760,Netherlands,Daley Blind,17
2014,Brazil,760,Netherlands,Georginio Wijnald
2014,Brazil,761,Germany,Mario Götze,113
```

Task 2.1: Schema Creation via SQL (3/25 points)

- **Schema creation via SQL**
  - Relies on lectures **04 Relational Algebra** and **05 Query Languages (SQL)**
  - #1 Setup DBMS PostgreSQL
  - #2 Create database `db<studentID>` and **setup relational schema**

- **Recommended Schema**
  - **TODO (after Apr 9, 11.59pm)**

Task 2.2 Data Ingestion via CLI (9/25 points)

- **Data Ingestion via ODBC/JDBC**
  - Relies on lectures **05 Query Languages (SQL)** and **06 APIs (ODBC, JDBC)**
  - #3 Write a program that performs **stateful data ingestion**

- **Two-Phase Ingestion**
  - Find data at https://github.com/tugraz-isds/datasets/tree/master/soccerworldcup19542014/subsets

```
IngestData ./1954_2010_Squads.csv \
  ./1954_2010_Matches.csv \
  ./1954_2010_Goals.csv <host>:<port>/<database>

IngestData ./2014_Squads.csv ./2014_Matches.csv \
  ./2014_Goals.csv <host>:<port>/<database>
```

Task 2.3: SQL Query Processing (10/25 points)

- **SQL Query Processing**
  - Relies on lecture **05 Query Languages (SQL)**
  - #4 Write SQL script for the **10 given queries**

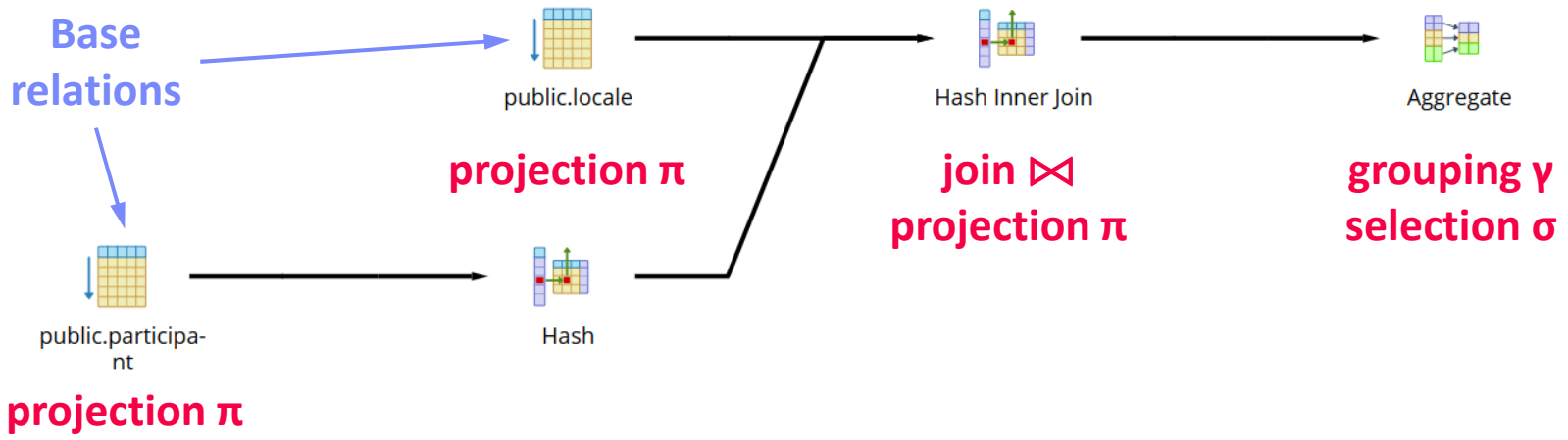
- **Example Queries**
  - **Q01:** In which tournaments did Austria participate between 1954 and 2014 (inclusive)?
  - **Q05:** With how many distinct clubs where players of Germany affiliated between 2002 and 2014?
  - **Q06:** How many players from Sturm Graz ever participated in a world cup tournament.
  - **Q07:** Which player(s) shot more than two goals in a single game in 2014?
  - **Q10:** Construct the final group table for Group G of the 2014 tournament (matches, wins, draws, losses, goal difference, points), ranked by points and goal differences.

Task 2.4: Query Plans (3/25 points)

- **Explain Query Plans**
  - Relies on lecture **04 Relational Algebra**
  - #5 Obtain and **analyze execution plans** of at least two queries

- **Example**
  - Recap: Participants and Locations of Database Research Self Assessment 2018
  - Text Explain

```
EXPLAIN VERBOSE SELECT L.location, count(*)
  FROM Participant P, Locale L WHERE P.lid = L.lid
  GROUP BY L.location HAVING count(*)>1
```



Call-level Interfaces (ODBC/JDBC) and Embedded SQL

Call-level Interfaces vs Embedded SQL

- **Embedded SQL**
  - Standardized in ISO/IEC SQL – Part 2: Foundation / Part 10 OLB
  - **Embedded (typically static) SQL in host language**
  - **Preprocessor** to compile CLI protocol handling
    ➔ **SQL syntax and type checking**, **but static** (SQL queries, DBMS)
  - **Examples:** ESQL for C/C++, SQLJ

- **Call-level Interfaces**
  - Standardized in ISO/IEC SQL – Part 3: CLI
  - **API of defined functions for dynamic SQL**
  - **Examples:** ODBC, JDBC, Python DB-API

Embedded SQL

- **Overview**
  - **Mix host language constructs and SQL** in data access program ➔ **simplicity?**
  - **Precompiler translates program** into valid host language program
  - Primitives for creating cursors, queries and updates, etc ➔ **In practice, limited relevance**

- **Example SQLJ**
  - Cursors with and without explicit variable binding

```
#sql iterator StudIter
   (int sid, String name);
StudIter iter;
#sql iter = {SELECT * FROM Students};

while( iter.next() )
   print(iter.sid, iter.name);

iter.close();
```

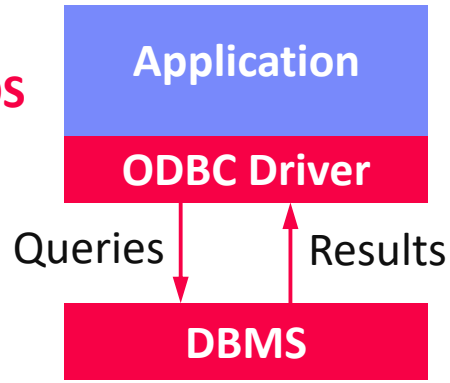
```
int id = 7;
String name;

#sql {SELECT LName INTO :name
   FROM Students WHERE SID=:id};

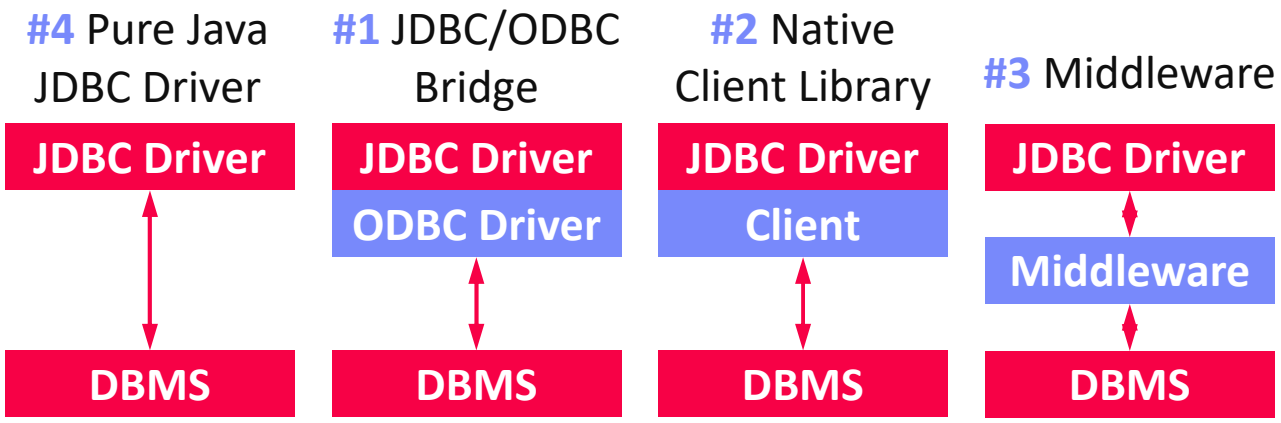
print(id, name);
```

CLI: ODBC and JDBC Overview

- **Open Database Connectivity (ODBC)**
  - **API for accessing databases independent of DBMS and OS**
  - Developed in the **early 1990s → 1992** by Microsoft (superset of ISO/IEC SQL/CLI and Open Group CLI)
  - **All relational DBMS have ODBC implementations**, good programming language support

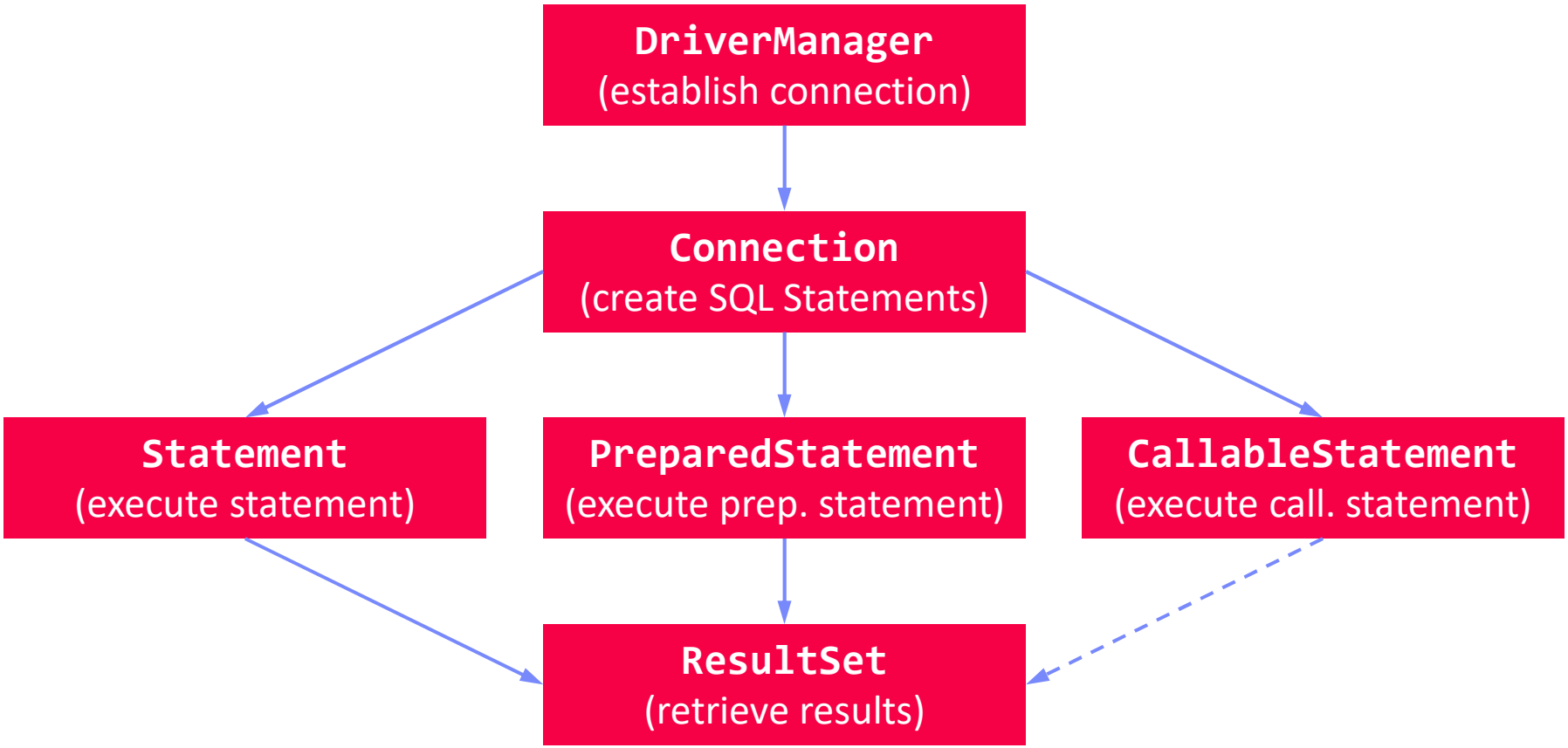


- **Java Database Connectivity (JDBC)**
  - **API for accessing databases independent of DBMS from Java**
  - Developed and released by Sun in **1997**, JDBC 4.0 (2006), JDBC 4.3 in Java 9
  - Most relational DBMS have JDBC implementations
  - Types of Drivers



Note: Reuse of drivers from open source DBMS

JDBC Components and Flow



JDBC Connection Handling

- **Establishing a Connection**
  - **DBMS-specific URL strings** including host, port, and database name
  - Stateful handles representing user-specific DB sessions
  - JDBC driver is usually a jar on the class path
  - **Connection and statement pooling** for performance

```
Connection conn = DriverManager
   .getConnection("jdbc:postgresql:"+
   "//localhost:5432/db1234567",
   username, password);
```

```
META-INF/services/
java.sql.Driver
```

- **JDBC 4.0**
  - **Explicit driver class loading and registration no longer required**
  - Improved connection management (e.g., status of DB connections)
  - Other: XML, Java classes, row ID, better exception handling

```
Class.forName(
   "org.postgresql.Driver");
```

JDBC Statements

- **Execute Statement**
  - Use for simple SQL statements w/o parameters
  - **Beware of SQL injection**
  - API allows fine-grained control over fetch size, fetch direction, batching, and multiple result sets

```
Statement stmt = conn.createStatement();
ResultSet rs = stmt.executeQuery(sql1);
...
int rows = stmt.executeUpdate(sql2);
stmt.close();
```

Note: PostgreSQL does not support fetch size but sends entire result

- **Process ResultSet**
  - Iterator-like cursor (app-level) w/ on-demand fetching
  - Scrollable / updatable result sets possible
  - Attribute access via column names or positions

```
ResultSet rs = stmt.executeQuery(
   "SELECT SID, LName FROM Students");

List<Student> ret = new ArrayList<>();
while( rs.next() ) {
   int id = rs.getInt("SID");
   String name = rs.getString("LName");
   ret.add(new Student(id, name));
}
```

JDBC Prepared Statements

- **Execute `PreparedStatement`**
  - Use for precompiling SQL statements w/ input params
  - Inherited from Statement
  - **Precompile SQL once**, and execute many times w/ different parameters

  ➔ **Performance**

  ➔ **No danger of SQL injection**

```
PreparedStatement pstmt =
   conn.prepareStatement(
     "INSERT INTO Students VALUES(?,?)");

for( Student s : students ) {
   pstmt.setInt(1, s.getID());
   pstmt.setString(2, s.getName());
   pstmt.executeUpdate();
}

pstmt.close();
```

- **Queries and Updates**
  - Queries → `executeQuery()`
  - Insert, delete, update → `executeUpdate()`

JDBC Callable Statements

- **Recap: (Stored Procedures, see 05 Query Languages (SQL))**
  - Can be **called standalone via CALL** <proc_name>(<args>);
  - Procedures return no outputs, but might have **output parameters**

- **Execute `CallableStatement`**
  - Create prepared statement for call of a procedure
  - Explicit registration of output parameters
  - Example

```
CallableStatement cstmt = conn.prepareCall(
   "{CALL prepStudents(?, ?)}");

cstmt.setInt(1, 2019);
cstmt.registerOutParameter(2, Types.INTEGER);
cstmt.executeQuery();

int rows = cstmt.getInt(2);
```

Preview Transactions

- **Database Transaction**
  - A transaction (TX) is a **series of steps** that brings a database from a **consistent state** into another (not necessarily different) **consistent state**
  - **ACID properties** (atomicity, consistency, isolation, durability)
  - See lecture **08 Transaction Processing and Concurrency**
- **Example**
  - Transfer 100 Euros from Account 107 to 999

```
START TRANSACTION ISOLATION LEVEL SERIALIZABLE;
   UPDATE Account SET Balance=Balance-100
      WHERE AID = 107;
   UPDATE Account SET Balance=Balance+100
      WHERE AID = 999;
COMMIT TRANSACTION;
```

- **Transaction Isolation Level**
  - **Tradeoff:** isolation (and related guarantees) vs performance
  - READ UNCOMMITTED (~~lost update~~, dirty read, unrepeatable read, phantom R)
  - READ COMMITTED (~~lost update~~, ~~dirty read~~, unrepeatable read, phantom R)
  - REPEATABLE READ (~~lost update~~, ~~dirty read~~, ~~unrepeatable read~~, phantom R)
  - SERIALIZABLE (~~lost update~~, ~~dirty read~~, ~~unrepeatable read~~, ~~phantom R~~)

JDBC Transaction Handling

- **JDBC Transaction Handling**
  - Isolation levels (incl NONE) and (auto) commit option
  - Savepoint and rollback (undo till begin or savepoint)
  - **Note:** TX handling on connection not statements

- **Beware of Defaults**
  - DBMS-specific default isolation levels

(SQL Standard: **SERIALIZABLE**, PostgreSQL: **READ COMMITTED**)

```
conn.setTransactionIsolation(
   TRANSACTION_SERIALIZABLE);
conn.setAutoCommit(false);

PreparedStatement pstmt = conn
   .prepareStatement("UPDATE Account
   SET Balance=Balance+? WHERE AID = ?");

Savepoint save1 = conn.setSavepoint();

pstmt.setInt(1,-100); pstmt.setInt(107);
pstmt.executeUpdate();

if( rand()<0.1 )
   conn.rollback(save1);

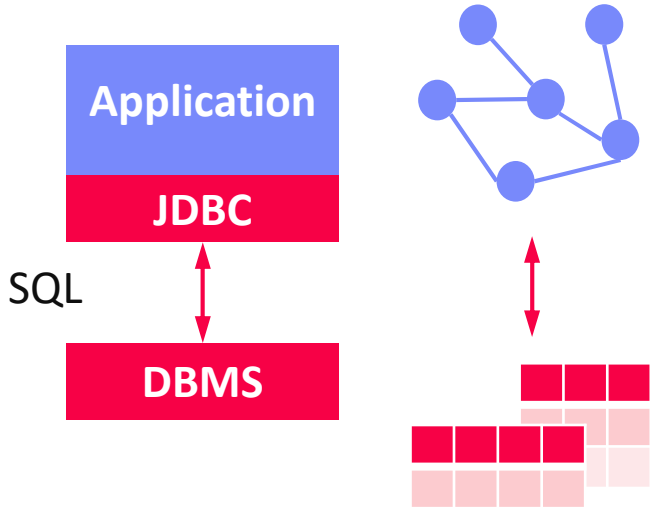
pstmt.setInt(1,100); pstmt.setInt(999);
pstmt.executeUpdate();

conn.commit();
```

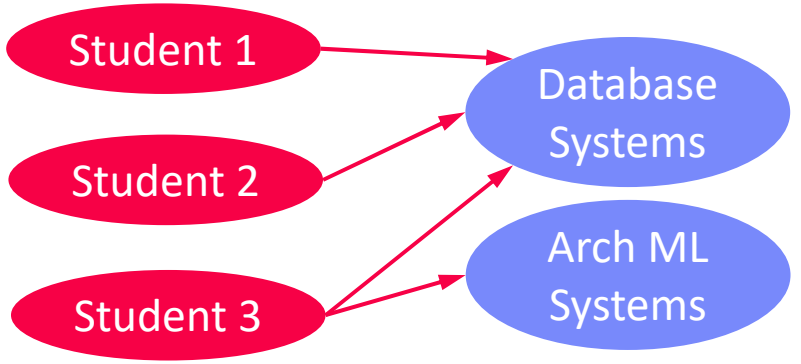
Object-Relational Mapping Frameworks

The "Impedance Mismatch" Argument

- **Problem Description**
  - Applications rely on **object-oriented programming languages** with hierarchies or graphs of objects
  - Data resides in **normalized "flat" tables** (note: ~~OODBMS~~, object-relational)
  - Application is responsible for **bridging this structural/behavioral gap**



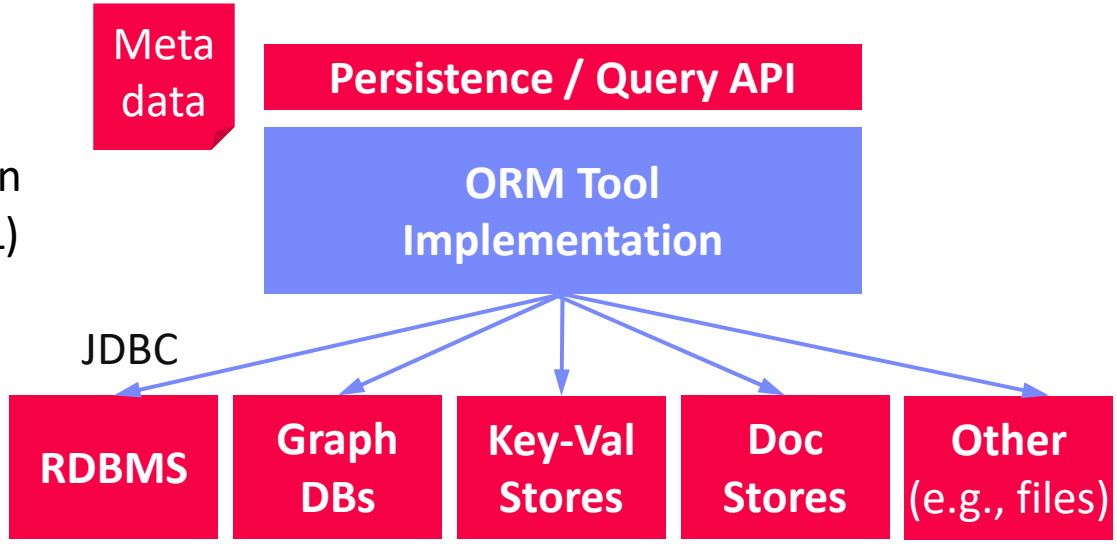
- **Example**
  - `SELECT * FROM Students`
  - ```
    SELECT C.Name, C.ECTS FROM
      Courses C, Attendance A
      WHERE C.CID = A.CID
        AND A.SID = 7;
    ```
  - `… A.SID = 8;`


    ```

Overview Object-Relational Mapping

- **Goals of ORM Tools**
  - Automatic **handling of object persistence lifecycle** and querying of the underlying data stores (e.g., RDBMS)
  - Reduced development effort ➔ **developer productivity**
  - Improved testing and independence of DBMS

- **Common High-Level Architecture**
  - **#1** Persistence definition (meta data → e.g., XML)
  - **#2** Persistence API
  - **#3** Query language / query API



History and Landscape

- **History of ORM Tools** (aka persistence frameworks)
  - Since 2000 J2EE EJB **Entity Beans** (automatic persistence and TX handling)
  - Since 2001 **Hibernate** framework (close to ODMG specification)
  - Since 2002 **JDO** (Java Data Objects) via class enhancement
  - 2006 **JPA** (Java Persistence API), reference implementation **TopLink**
  - 2013 JPA 2, reference implementation **EclipseLink**
  - Late 2000s/early 2010s: **explosion of ORM alternatives, but criticism**
  - **2012 - today:** ORM tools just part of a much more diverse eco system

- **Example Frameworks**
  - http://java-source.net/open-source/persistence
  - Similar lists for .NET, Python, etc

JPA – Class Definition and Meta Data

- **Entity Classes**
  - **Define persistent classes** via annotations
  - Add details for IDs, relationship types, and specific behavior on updates
  - Some JPA implementations require enhancement process as post compilation step

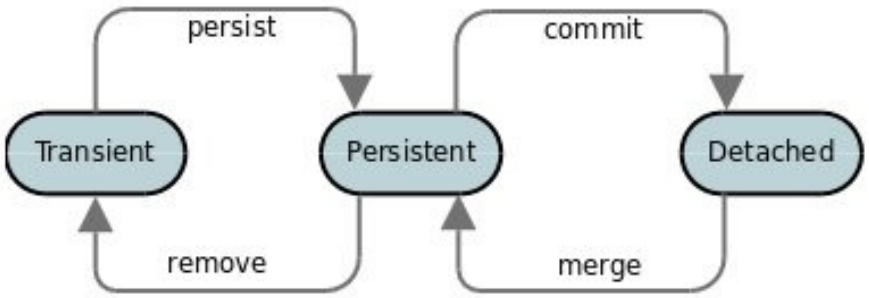
```
@Entity
public class Student {
   @Id
   private int SID = -1;
   private String Fname;
   private String Lname;
   @ManyToMany
   private List<Course> ...
}
```

- **Persistence Definition**
  - **Separate XML meta data** `META-INF/persistence.xml`
  - Includes connection details

```
<?xml version="1.0" encoding="UTF-8" ?>
<persistence
      xmlns="http://xmlns.jcp.org/xml/ns/persistence"
      xmlns:xsi=... xsi:schemaLocation=...>
   <persistence-unit name="UniversityDB">
      <class>org.tugraz.Student</class>
      <class>org.tugraz.Course</class>
      <exclude-unlisted-classes/>
      <properties> ... </properties>
   </persistence-unit>
</persistence>
```

JPA – Object Modification

- **CRUD Operations**
  - Insert by making objects persistent
  - Update and delete objects according to object lifecycle states
- **Lifecycle States**
  - Lifecycle state transitions via specific persistence contexts
  - Explicit and implicit transitions



[**Credit:** Data Nucleus, JPA Persistence Guide (v5.2), http://www.datanucleus.org/products/accessplatform/jpa/persistence.html#lifecycle]

```
EntityManager em = factory
   .createEntityManager();

tx.begin();

Student s = new
   Student(7,"Jane","Smith");
s.addCourse(new Course(...));
s.addCourse(new Course(...));

em.persist(s);

tx.commit();
em.close
```

JPA – Query Languages

- **JPQL: Java Persistence Query Language**
  - SQL-like object-oriented query language
  - Parameter binding similar to embedded SQL

```
EntityManager em = factory
   .createEntityManager();
Query q = pm.createQuery(
   "SELECT s FROM Student s
      WHERE s.age > :age");
q.setParameter("age", 35);

Iterator iter = q
   .getResultList().iterator();
while( iter.hasNext() )
   print((Student)iter.next());
```

- **JPQL Criteria API**
  - JPQL syntax and semantics with a programmatic API
  -
```
CriteriaQuery<Student> q = bld.createQuery(Student.class);
Root<Student> c = q.from(Student.class);
q.select(c).where(bld.gt(c.get("age"), bld.parameter(...)));
```

- **Native SQL Queries**
  - Run native SQL queries if necessary

```
em.createNativeQuery("SELECT *
   FROM Students WHERE Age > ?1");
```

A Critical View on ORM

- **Advantages**
  - **Simple CRUD operations** (insert/delete/update) and simple queries
  - **Application-centric development** (see boundary crossing)
- **Disadvantages**
  - **Unnecessary indirections** and complexity (meta data, mapping)
  - **Performance problems** (hard problem and missing context knowledge)
  - **Application-centric development** (schema ownership, existing data)
  - **Dependence** on evolving framework APIs
- **Sentiments** (additional perspectives)
  - Omar Rayward: Breaking Free From the ORM: Why Move On?, 2018 medium.com/building-the-system/**dont-be-a-sucker-and-stop-using-orms**-190add65add4
  - Vedra Bilopavlović: Can we talk about ORM Crisis?, 2018 linkedin.com/pulse/**can-we-talk-orm-crisis**-vedran-bilopavlovi%C4%87
  - Martin Fowler: ORM Hate, 2012 martinfowler.com/bliki/**OrmHate**.html

➔ **Awareness of strength and weaknesses / hybrid designs**

Conclusions and Q&A

- **Summary**
  - Call-level Interfaces (ODBC/JDBC) as fundamental access technology
  - Object-Relational Mapping (ORM) frameworks existing (**pros and cons**)

- **Exercise Reminder**
  - Exercise 1: **Last chance tomorrow EOD** (including 7 late days)
  - Exercise 2: Submission opens Apr 22, deadline: **Apr 30 11.59pm**

- **Next Lectures**
  - Apr 29: **07 Physical Design and Tuning**
  - May 6: **08 Query Processing**
  - May 13: **09 Transaction Processing and Concurrency**