Facoltà di Ingegneria

Corso di Laurea Specialistica in Ingegneria Informatica

Tesi di Laurea

In-Memory Database: Competitive Landscape and Performance Analysis

Laureando

Valerio Barbagallo

Relatore

Prof. **Paolo Merialdo**

Università Roma Tre

Tutor Aziendale

Dr. **Michele Aiello**

Applications Director & Co. Founder

ERIS4 s.r.l.

Anno Accademico 2007-2008

Scio me nihil scire

Acknowledgements

I would like to thank all the people who have helped and inspired me during the writing of this thesis.

I especially want to thank my tutor Michele Aiello for his guidance during my research. With his enthusiasm and great efforts, he was always accessible and willing to help. I am also grateful to all the people working at ERIS4, who have given me hospitality during this thesis. In alphabetical order they are, apart from my tutor, Stefano Antonelli, Augusto Coppola, Patrizio Munzi, Mario Zullo. In the last 8 months I worked in a company's climate which was cordial and quite where everyone helps each other.

I want to express my gratitude also to Prof. Paolo Merialdo, who during the last years has given me many advices.

I wish to thank all the friends who in the last five years made the university a convivial place to work. They are too many to list them all, but a particular thanks to Giordano Da Lozzo, who is my friend since the high school.

A special thanks is for Chiara Gaetani, who during the last two years has given me the gift of serenity.

Lastly, and most importantly, I wish to thank my parents, Lucia Taccetti

and Antonio Barbagallo. To them I will not spend any word, i can only dedicate this thesis.

Abstract

This thesis introduces the reader to the argument of in-memory database systems, providing a competitive landscaper between several IMDB's vendors, and, subsequently, analyzing and comparing the performance of different systems to each other. The purpose of this thesis is to find the in-memory database which provides the best performance for a specific application usage: a real time prepaid system, such as a real time authorization, rating and billing system for a telco company.

This objective was achieved with the development of a proper test suite, simulating a real time prepaid system, and consequently of a portable benchmark application, which was able to execute the test suite defined. The need to create a specific test suite and the portable benchmark application comes from the fact that database's performance depends on a specific set of criteria used by the benchmark itself, and they are mainly: the test scenarios, the test implementation and the platform used to execute the benchmark. This means that there is not a better database systems, but there are only databases that are more suitable for a particular task or application usage. Of course this is a general statement and actually we will find out that in some cases certain database systems might be "overall" better than other,

but as a rule of thumb we have always to keep in mind that the best way to choose our database system, for applications that require cutting edges performances, is to create a specific test scenario which simulates the real life application's behavior.

The thesis is divided into two parts: the first is made of two chapters which introduce in-memory databases and the competitive landscape, while the second is dedicated to the benchmark application developed during this work.

The first chapter gives a brief introduction to the in-memory databases, explaining what they are, their use, their strength and weakness. Particularly interesting is the comparison against traditional DBMS, which shows the key differences.

In chapter two, we will make a competitive landscape between the different IMDB's vendors. Every in-memory database will be analyzed investigating their advantages and disadvantages, the stability and reliability of the project and how much development is going on. Eventually we will also provide some example code.

The chapter three is the first chapter of the second part. It therefore introduces the reader to the database performance analysis' problem, illustrating the difficulty behind the measurement and the interpretation of the performance. This in fact leads to a definition of an axiom on which is based all this work: there is not a slower or a faster database, because they are only slower or faster given a specific set of criteria in a given benchmark.

The fourth chapter deals with the creation of a proper test suite, based

on our needs, and therefore the creation of a test scenario based on a real time prepaid system. This test suite will be used as a test scenario for the benchmark application.

The fifth chapter analyzes more or less deeply different open source database benchmarks trying to find a benchmark for our needs, or, where none are suitable, to take some idea for the development of a new benchmark application.

The sixth chapter describes the new benchmark application developed for our needs, starting form the specification defined from the requirements and then describing the application from different points of view. Subsequently the discussion moves on the application usage, starting from the application's configuration and ending with the extension and implementation of new database adapters and test scenarios.

Chapter seven deals with the results gathered from the execution of the test suite (as described in chapter three) using our database benchmark application. The results are analyzed to clarify their meaning and to come up with a report on all the in-memory database analyzed in this work. We will see which one of them is better suitable for a real time prepaid system.

The final chapter lists some of the most important contributions brought by this thesis work and how it can be used as a starting point for future researches and developments on the benchmark application.

Contents

List of Figures

List of Tables

Part I

Competitive Landscape

Chapter 1

A Brief Introduction

"While familiar on desktops and servers, databases are a recent arrival to embedded systems. Like any organism dropped into a new environment, databases must evolve. A new type of DBMS, the in-memory database system (IMDS), represents the latest step in DBMSes' adaptation to embedded systems" [Graves 02].

In this chapter we are going to make a brief introduction to the in-memory databases, explaining what they are, their use, their strength and weakness.

1.1 Definition

An in-memory database (IMDB), also called main memory database (MMDB), in-memory database system (IMDS) or real-time database (RTDB), is a database management system that relies on main memory for data storage. While the bandwidth of hard disks is just 1 order of magnitude slower than the main memory's bandwidth, the disk access time is about 3 order of

magnitude slower than the RAM access time, and thus in-memory databases can be much more faster than traditional database management systems (DBMS).

1.2 History

Initially embedded systems developers produced their own data management solutions. But with the market competition requiring smarter devices, applications with expanding feature set will have to manage increasingly complex data structures. As a consequence, these data management solutions were outgrowing, and became difficult to maintain and extend.

Therefore embedded systems developers turned to commercial databases. But the first embedded databases were not the ideal solution. They were traditional DBMS with complex caching logic to increase performance, and with a lot of unnecessary features for the device that make use of embedded databases. Furthermore these features cause the application to exceed available memory and CPU resources.

In-memory databases have emerged specifically to meet the performance needs and resource availability in embedded systems.

1.3 Application Scenario

Often in-memory databases run as embedded database, but it's not their only use. Thanks to their high performance, these databases are particularly useful for all that kind of applications that need fast access to the data. Some

examples:

- real time applications which don't need to be persisted either because it doesn't change, or the data can be reconstructed: imagine a routing table of a router with millions of record and data access in less than few milliseconds; the routing table can be rebuilt [Fowler 05].

- real time applications with durability needs which capture, analyze and respond intelligently to important events, requiring high performance in terms of throughput and mainly latency (traditional DBMS can be clustered to increase the througput, but with no great benefits in terms of latency). Infact almost all IMDBs can be persistent on disk while still keeping higher performance compared to traditional DBMSs.

- in-memory databases are also very useful for developers of traditional database systems for testing purpose: in a enterprise application running a test suite can take long; switching to an IMDB can reduce the whole build time of the application.

1.4 Comparison against Traditional DBMS

In-memory databases eliminate disk I/O and exist only in RAM, but they are not simply a traditional database loaded into main memory. Linux systems already have the capability to create a RAM disk, a file system in main memory. But a traditional database deployed in a such virtual hard disk doesn't provide the same benefits of a pure IMDB. In-memory databases are

less complex than a traditional DBMS fully deployed in RAM, and lead to a minor usage of CPU and RAM.

Comparing IMDBs with traditional databases we can find at least 3 key differences:

- Caching.
- Data-transfer overhead.
- Transaction processing.

1.4.1 Caching

All traditional DBMS software incorporates caching mechanisms to keep the most used records in main memory to reduce the performance issue introduced by the disk latency. The removal of caching brings to the elimination of the following tasks:

- cache synchronization, used to keep the portion of the database loaded in main memory consistent with the physical database image.
- cache lookup, a task that handle the cached page, determining if the data requested is in cache

Therefore, removing the cache, IMDBs eliminate a great source of complexity and performance overhead, reducing the work for the CPU and, comparing to a traditional DBMS fully deployed in main memory, also the RAM.

1.4.2 Data-Transfer Overhead

Traditional DBMS adds a remarkable data transfer overhead due not only to the DB cache, but to the file system and its cache too. In contrast an IMDBs, which eliminate these steps, have little or no data transfer.

There is also no need for the application to make a copy of the data in local memory, because IMDBs give to the application a pointer to the data that reside in the database. In this way the data is accessed through the database API that protect the data itself.

1.4.3 Transaction Processing

In an hard-disk database the recovery process, from a disaster or a transaction abort, is based on a log file, updated every time a transaction is executed. To provide transactional integrity, IMDBs mantain a before image of the objects updated and in case of a transaction abort, the before image are restored in a very efficient way. Therefore another complex, memory-intensive task is eliminated from the IMDB, unless the need to add durability to the system, because there is no reason to keep transaction log files.

1.5 ACID Support: Adding Durability

When we choose a database we expect from it to provide ACID support: atomicity, consistency, isolation and durability. While the first three features are usually supported by in-memory databases, pure IMDBs, in their simplest form, don't provide durability: main memory is a volatile memory and loses

all stored data during a reset.

With their high performance, IMDBs are a good solution for time-critical applications, but when the durability need arises they may not seem a proper solution anymore. To achieve durability [Gorine 04], in-memory database systems can use several solutions:

- On-Line Backup.
- Transaction logging.
- High availability implementations.
- Non volatile RAM (NVRAM).

1.5.1 On-Line Backup

On-line backup is a backup performed while the database is on-line and available for read/write. This is the simplest solution, but offer a minimum degree of durability.

1.5.2 Transaction Logging

A transaction log is a history of actions executed by a database management system. To guarantee ACID properties over crashes or hardware failures the log must be written in a non-volatile media, usually an hard disk. If the system fails and is restarted, the database image can be restored from this log file.

The recovery process acts similarly to traditional DBMS with a roll-back or a roll-forward recovery. Checkpoint (or snapshot) can be used to speed up

this process. However this technique implies the usage of persistence memory such as an hard disk, that is a bottleneck, especially during the resume of the database.

Although this aspect, IMDBs are still faster than traditional DBMS:

1. transaction logging requires exactly one write to the file system, while disk-based databases not only need to write on the log, but also the data and the indexes (and even more writes with larger transactions).

2. transaction logging may be usually set to different level of transaction durability. A trade-off between performance and durability is allowed by setting the log mechanism to be synchronous or asynchronous.

1.5.3 High Availability Implementation

High availability is a system design protocol and associated implementation: in this case it is a database replication, with automatic fail over to an identical standby database. A replicated database consists of failure-independent nodes, making sure data is not lost even in case of a node failure. This is an effective way to achieve database transaction durability.

In a simile way to transaction logging, a trade-off between performance and durability is achieved by setting the replication as eager (synchronous) or lazy (asynchronous).

1.5.4 NVRAM

To achieve durability a IMDB can support non volatile ram (NVRAM): usually a static RAM backed up with battery power, or some sort of EEPROM.

In this way the DBMS can recover the data also after a reboot.

This is a very attractive durability option for IMDBs. NVRAM in contrast to transaction logging and database replication does not involve disk I/O latency and neither the communication overhead.

Despite this, vendors rarely provide support for this technology. One of the major problems to this approach is the limited write-cycles of this kind of memory, such as a flash memory. On the other hand there is some new memory device that has been proposed to address this problem, but the cost of such devices is rather expensive, in particular considering the huge amount of memory needed by IMDBs.

Chapter 2

In-Memory Database Overview

There is a variety of in-memory database systems which can be used to maintain a database in main memory, both commercial and open source. Although all of them share the capability to maintain the database in main memory, they offer different sets of feature.

In this chapter we will make a competitive landscape of the most famous in-memory databases.

2.1 The Analysis' Structure

Every in-memory database will be analyzed investigating their advantages and disadvantages, the stability and reliability of the project and how much development is going on. Eventually we will go to the development stage and, in some case, we will also have a look "under the hood" to see how the DB works.

Although not all in-memory database systems share the same features,

in terms of advantages and disadvantages, it is possible to identify a subset of common advantages and disadvantages they all have. We will therefore enunciate these common features before starting the specific analysis for each IMDB.

2.1.1 Common Advantages

The common advantages in-memory database systems share are:

Lightweigth is one of the common advantages IMDBs share. This kind of databases is really simple: they don't implement any of the complex mechanism used by traditional database to speed up the disk I/O. Thus, usually, an in-memory database consist of a simple jar, whose size is less than 1 MB.

Robustness is a direct consequence of the previous point: (citing Henry Ford) it is impossible to break something that doesn't exist.

High Performance is another common feature for every IMDB, because they all store the whole database in main memory, avoiding disk access.

2.1.2 Common Disadvantages

On the other side, all IMDBs require **high quantity of main memory**. It is larger as the database image increases, but this doesn't mean that every IMDB uses the same RAM quantity for the same database image. A common question is about what may happen when the RAM is not enough. Altought

this question is really interesting, generally an IMDB makes the assumption there is always enough main memory to hold the database image.

Furthermore, since in-memory databases rely on main memory, which is a volatile memory, they may **lose all the stored data** with a power off, unless they don't use proper mechanisms to achieve durability, as already described in paragraph 1.5.

But these mechanisms, used to add durability to an in-memory database system, make the database's **startup terribly slow**, because the database image is not stored on the hard disk, or on another persistent memory, and it needs to be reconstruct at every reboot, for example by reading the transaction log file and repeat all the operations executed before the power off.

This issue is always mitigated by the use of database's **snapshots**, or commits, which store a fixed database image on a persistent memory, avoiding, in the case of a reboot, the reconstruction of the database image from the start of the transaction log file, and therefore saving a lot of time. Although this mechanism if properly used dramatically decreases the database startup time, it is still very **slow**, and in addition also the snapshot of the database is very slow and furthermore it slows down, if not freeze, the whole database performance during the execution of the snapshot.

2.1.3 Common references

The source of informations described in the following sections comes from the web, mainly from sourceforge or ohloh.net. Both these web site offer a variety of news such as the size of developer team, the frequency of commits,

the date of the last realese, etc.

Another source is, naturally, the IMDBs' official web site. Although this is a huge source, it cannont be considered impartial. Therefore the following analysis may not be consistent with facts.

2.2 Prevayler

Prevayler is an object persistence library for Java, written in Java. It keeps data in main memory and any change is written to a journal file for system recovery. This is an implementation of an architectural style that is called System Prevalence by the Prevayler developer team.

Therefore Prevayler, being an object prevalence layer, provides transparent persistence for Plain Old Java Objects. In the prevalent model, the object data is kept in memory in native, language-specific object format, and therefore avoid any marshalling to an RDBMS or other data storage system. To avoid losing data and to provide durability a snapshot of data is regularly saved to disk and all changes are serialized to a log of transactions which is also stored on disk [Wuestefeld 01].

All is based on the Prevalent Hypothesis: that there is enough RAM to hold all business objects in your system.

2.2.1 Advantages

Prevayler is a lightweight java library, just 350 KB, that is extremely simple to use. There is no separate database server to run. With Prevayler you can program with real objects, there is no use of SQL, there is no impedance mis-

match such as when programming with RDBMS. Moreover Prevayler doesn't require the domain objects to implement or extend any class in order to be persistent (except `java.io.Serializable`, but this is only a marker interface).

It is very fast. Simply keeping objects in memory in their language-specific format is both orders of magnitude faster and more programmer-friendly than the multiple conversions that are needed when the objects are stored and retrieved from an RDBMS. The only disk access is the streaming of the transactions to the journal file that should be about one thousand[1] transactions per second on an average desktop computer.

The thread safety is guaranteed. Actually, in the default Prevayler implementation all writes to the system are synchronized. One write at a time. So there's no threading issues at all. Therefore there is no more multithreading issue such as locking, atomicity, consistency and isolation.

Finally Prevayler supports the execution only in RAM, like a pure in-memory database should work. But it can also provide persistence through a journal file, as we described above. Moreover Prevayler supports snapshots of the database's image, which serves to speed up the database's boot, and server replication, enabling query load-balancing and system fault-tolerance [Prevayler 08]. This last feature is really promising, because in-memory databases suffer the start up process, allowing it to be used in an enterprise application. Althought this feature is not ready yet: see paragraph 2.2.4 for further details at page 19.

[1] This is what Prevayler team says on their official web site on Mar 26, 2008.

2.2.2 Disadvantages

On the other hand of Prevayler's simplicity, there is some restriction due to the fact that only changes are written in the journal file: transaction must be completely deterministic and repeteable in order to recover the state of the object (for instance it's not possible to use `System.getCurrentMillis()`).

In addition, when Prevayler is embedded in your application, the only client of your application's data is the application itself [Hobbs 03]. While deploying Prevayler as a server, the access to the data is still limited to whom who speaks the host language (you can't use another programming language unless you make your own serialization mechanism) and, at the same time, knows the model objects. For all these reasons there are no administration and migration tools.

Another problem is related to the Prevalent Hypothesis. If, for any reason, the RAM is not enough and an object model is swapped out of main memory, Prevayler will become very slow, more than traditional RDBMS's [Miller 03]. In fact Prevayler doesn't use any mechanism to optimize the disk I/O, such as traditional DBMS's mechanisms (eg: indexing, caching etc.) Anyway this issue belongs to all pure in-memory databases.

2.2.3 Project Info

Klaus Wuestefeld is the founder and main developer of Prevayler, an open source project. This project started in 2001, from what sourceforge reports, and had a great development until 2005. The community is still active, but the last update is dated at 25/05/2007 when version 2.3 (the latest)

was released. The development team is composed of 8 developers, including Klaus Wuestefeld.. The current development status is declared to be production/stable.

2.2.4 Usage

In this example, and in all the followings, we are going to use a simple Plain Old Java Object as business object that need to be persisted: `Number`. It has one single private field, which is the value of the number itself, and the relative getter and setter methods.

```
public class Number{
  private int value;
  public Number(){}
  public Number(int value) {
    this.value = value;
  }
  public int getValue() {
    return value;
  }
  public void setValue(int number) {
    this.value = number;
  }
}
```

Listing 2.1: Implementation of the object Number

Prevayler usage is quite different from a traditional RDBMS with JDBC driver. Prevayler is accessed through a native driver, and it acts similarly to a framework: your business objects must implement the interface `java.io.Serializable` in order to be persisted (which is only a marker interface); and all modifications to these objects must be encapsulated in transaction objects. Therefore our business object will appear as follow:

```
public class Number implements Serializable { ...
```

Basic: main memory and transaction logging

This example is about the default Prevayler's usage: how to create an in-memory database without losing durability. This property is achieved with a transaction log file, whose usage is totally transparent. To initialize Prevayler Database you need to create a Prevayler, providing it with the directory where you want to save your database, and the object which is going to be saved in this Prevayler database. This step can be compared to a `CREATE TABLE` in SQL, but only one object will be saved in your Prevayler database. Therefore, from a RDBMS perspective where the table is a class and each row is one instance, you may want to initialize Prevayler with a `List` or a `Map` of `Number`. Here is an example:

```
public class Main{
  private static final String DIRECTORY_DB = "numberDB";
  public static void main(String[] args) throws Exception {
    Prevayler prevayler = PrevaylerFactory.createPrevayler(
        new HashMap<Integer,Number>(),DIRECTORY_DB);
    new Main().fillPrevaylerDb(prevayler);
    new Main().readPrevaylerDb(prevayler);
  }
  ...
```

It's important to understand that the state of the object you use to create this database will not be saved in the database itself. To insert any `Number` in the database a transaction must be executed:

```
private void fillPrevaylerDb(Prevayler prevayler) throws
    InterruptedException {
  for (int i = 0; i<100; i++){
    prevayler.execute(new InsertNumberTransaction(new Number
        (i)));
    System.out.println("The value of the number inserted is
        = "+i);
  }
}
```

The parameter of the method `execute` must be a class that extends `org.prevayler.Transaction`. Every insert, update or delete (in other words: any write operation) requires to be executed inside a transaction. In this particular case this is the the code:

```
public class InsertNumberTransaction implements Transaction{
  private Number number;
  public InsertNumberTransaction(Number number) {
    this.number = number;
  }
  public void executeOn(Object prevalentSystem, Date ignored) {
    Map<Integer,Number> map = (Map<Integer, Number>) prevalentSystem;
    map.put(number.getValue(), number);
  }
}
```

It's important to note that when you stop the database, and then you restart it, Prevayler will execute all the transactions exactly the same number of times they were executed before shutting down the process [2]. While a snapshot should avoid this behavior.

Finally, reading from Prevayler database is really simple and doesn't require any transaction:

```
private void readPrevaylerDb(Prevayler prevayler) {
  Map<Integer,Number> map= (Map<Integer,Number>) prevayler.prevalentSystem();
  Set<Integer> keys = map.keySet();
  for (Integer key : keys) {
    Number number = map.get(key);
    System.out.println("Reading the number " + number.getValue());
  }
}
```

[2]This is the reason why the transactions must be deterministic

Only RAM

This example show how to make Pervayler run only in main memory, without the writes to the journal file, in the case you want a database even faster and you don't care for durability or you don't have write permission. There is only one line of code which need to be modified to be able to run Prevayler in such a way:

```
Prevayler prevayler = PrevaylerFactory.↘
    createTransientPrevayler(new HashMap<Integer,Number>());
```

Instead of creating a persistent prevayler, you just need to `createTransientPrevayler`↘. You can also decide to create the transient prevayler from a snapshot, and then work only in main memory: really useful for the execution of your test case. But you can't take a snapshot while using transient prevayler, therefore you need to disable the snapshot in your code when switching from the default to the transient prevayler, otherwise a `IOException` will raise.

With this method you have no durability, but it is even faster than the first example, about 10 times faster. And moreover it is more scalable, because there is no more bottleneck caused by the hard disk.

Server Replication

Prevayler can support also a server replication modality. Also in this case, only a small change to the database initialization is needed. Quite obviously you need two different kind of initialization: server side and clients side.

As regards the server, you need to specify the port number and, then, simply create the object `Prevayler`:

```
public class MainServer {
```

```
  private static final String DB_DIRECTORY_PATH = "↘
     numberReplicaDB";
  private static final int PORT_NUMBER = 37127;
  public static void main(String[] args) throws Exception {
    PrevaylerFactory factory = new PrevaylerFactory();
    factory.configurePrevalentSystem(new HashMap<Integer,↘
       Number>());
    factory.configurePrevalenceDirectory(DB_DIRECTORY_PATH);
    factory.configureReplicationServer(PORT_NUMBER);
    Prevayler prevayler = factory.create();
    ...
    // execute your transactions
    ...
    // The server will continue to listen/run for incoming ↘
       connections
    ...
  }
}
```

As for the clients, you have to tell Prevayler not only the port number, but the ip address too. Only one line of code changes from the server:

```
public class MainReplicant {
  private static final String DB_DIRECTORY_NAME = "numberDB"↘
     ;
  private static final int PORT_NUMBER = 37127;
  private static final String SERVER_IP = "10.0.2.2";

  public static void main(String[] args) throws Exception {
    PrevaylerFactory factory = new PrevaylerFactory();
    factory.configurePrevalentSystem(new HashMap<Integer,↘
       Number>());
    factory.configurePrevalenceDirectory(DB_DIRECTORY_NAME);
    factory.configureReplicationClient(SERVER_IP, ↘
       PORT_NUMBER);
    Prevayler prevayler = factory.create();
    ...
    //execute your transactions
  }
}
```

With this setup, you can stop, restart and add clients without losing your data, and keeping it synchronized with the server and the other clients, apparently without any concurrent exception. But when you try to kill the server, or for any other reason the server stops, while any client is still active this is the message you get:

```
The replication logic is still under development.
java.io.EOFException
```

Reading the javadoc comes out that this feature is *reserved for future implementation*. Contacting Prevayler's author, Klaus Wuestefeld, by email, he said this feature will be probably done within the 2009, because he needs it for other projects.

2.3 HSQLDB

HSQLDB is a relational database written in Java. It is based on Thomas Mueller's Hypersonic SQL discontinued project and therefore its name which stands for Hypersonic SQL DataBase. Thomas Mueller then developed a new database system which he called H2 and he closed the Hypersonic SQL project in 2001. Therefore a number of developers, who where using the closed project for their work, got togheter through the internet and formed the HSQLDB development group.

Each hsql database consists of between 2 to 5 files, all named the same but with different extensions, located in the same directory. For example, the database named "test" consists of the following files:

- test.properties contains general settings about the database;
- test.script contains the definition of tables and other database objects, plus the data for non-cached tables;
- test.log contains recent changes to the database;
- test.data contains the data for cached tables;

- test.backup is a zipped backup of the last known consistent state of the data file.

Hsql 1.8.0 is also the database engine in OpenOffice 2.0.

2.3.1 Advantages

HSQLDB is an open source relational Java database and it can run on Java runtime from version 1.1. As for most IMDBs, it is a lightweight library of about 100 kilobytes in the smaller version and 600 kilobytes in another. It has a JDBC driver and supports a large subset of SQL standards and therefore it is quiet easy switch to HSQLDB if you are using another relational DBMS through SQL.

It offers both in-memory and disk based table, therefore it can act as an in-memory database system, which writes on the disk only the change to the database, or as a traditional database system, recording all the data to the hard disk. Hence HSQLDB let you choose the way you prefer to use it and switch easily between them. Obviously the first will be faster, but the second will save a huge amount of time when the database system is restarted.

In addition HSQLDB can work in server and embedded mode, allowing this time too an easy switch between these two modalities. Usually the server mode is used while developing a new application, so saving the time when it is restarted, and then switch to embedded mode for deployment for better performance.

HSQLDB also includes some tools such as a minimal web server, in-memory query and management tools.

2.3.2 Disadvantages

The most important disadvantages is of course that HSQLDB supports only a low transaction isolation level when queries are made from a Java program, the level 0 which means read uncommitted transaction. Therefore the concurrence must be managed inside the Java program. For example HSQLDB can support the SERIALIZABLE isolation level if you write the database access code as public static Java methods and call them as stored procedures.

Another feature to take into consideration if you want to use HSQLDB is that it supports data to a maximum of 8GB, and although for common applications it is a huge value, if we want to use it in enterprise applications it is very often not enough.

Finally it is important to notice how HSQLDB doesn't support a strict level of durability. In fact if the database system is not shutted down nicely, ie by sending the commands `CHECKPOINT` or `SHUTDOWN` via JDBC connection before killing the process, then occasionally HSQLDB will lose some data changes.

2.3.3 Project Info

As already explained in the introduction, HSQLDB is based on the hypersonic SQL project and started in 2001. Actually the latest version is 1.8.0.8, released at 30/08/2007. The project is currently in production/stable development status, and the development team which is composed by 34 programmers is still working. In fact we are close to the next release which will be

1.9.0, whose features are already known: an extensive range of new SQL and JDBC capabilities, increased scalability, and better query optimization have been achieved by a rewrite of several internal components and the addition of some major new ones.

2.3.4 Usage

HSQLDB usage is very simple because it is almost the same of all relational database systems with JDBC drivers. Also in this example, as for Prevayler, we will use the Number object, showed in the listing 2.1, as an object to be stored and retrieved. While for Prevayler we need a different implementation for the object Number, altougth it is very small, HSQLDB uses a JDBC driver and therefore uses SQL language, and there is no need for a different implementation.

The default HSQLDB behavior is to take data in main memory and only changes are written to the disk. In the following listing we show how to execute HSQLDB embedded, or in other word in in-process mode:

```
public class Main{
  private static final String databaseURL="jdbc:hsqldb:file:↘
     db/test";
  private static final String username="sa";
  private static final String password="";
  private static final String driver="org.hsqldb.jdbcDriver"↘
     ;
  private static final String dropTableQuery = "DROP TABLE ↘
     NUMBER";
  private static final String createTableQuery = "CREATE ↘
     TABLE NUMBER (number integer)";
  private static final String saveNumberQuery = "INSERT INTO↘
      NUMBER VALUES (?)";
  private static final String findNamberQuery = "SELECT * ↘
     FROM NUMBER WHERE NUMBER=?";

  public static void main(String[] args) throws Exception {
    Class.forName(driver);
```

```
    Connection con = DriverManager.getConnection(databaseURL↘
        ,username,password);
    createTable(con);
    int numberValue = new Random().nextInt();
    Number number = new Number();
    number.setNumber(numberValue);
    saveNumber(con,number);
    Number numberRetrieved = findNumber(con,numberValue);
    dropTable(con);
    if (numberRetrieved.getNumber()==number.getNumber())
      System.out.println("The numbers have the same value");
  }
  private static Number findNumber(Connection con, int ↘
      numberValue) {
    Number result = null;
    try {
      PreparedStatement st = con.prepareStatement(↘
          findNamberQuery);
      st.setInt(1, numberValue);
      ResultSet resultSet = st.executeQuery();
      if (resultSet.next()){
        result = new Number();
        result.setNumber(resultSet.getInt(1));
      }
      System.out.println("Number found correctly");
    } catch (SQLException e) {
      e.printStackTrace();
    }
    return result;
  }
  private static void saveNumber(Connection con, Number ↘
      number) {
    try {
      PreparedStatement st = con.prepareStatement(↘
          saveNumberQuery);
      st.setInt(1, number.getNumber());
      st.executeUpdate();
      System.out.println("Number saved corerctly");
    } catch (SQLException e) {
      e.printStackTrace();
    }
  }
  private static void createTable(Connection con) {
    try {
      PreparedStatement st = con.prepareStatement(↘
          createTableQuery);
      st.executeUpdate();
      System.out.println("Table created correctly");
    } catch (SQLException e) {
      e.printStackTrace();
    }
```

```
  }
  private static void dropTable(Connection con) {
    try {
      PreparedStatement st = con.prepareStatement(↘
          dropTableQuery);
      st.executeUpdate();
      System.out.println("Table deleted correctly");
    } catch (SQLException e) {
      e.printStackTrace();
    }
  }

}
```

Listing 2.2: HSQLDB implementation

The database systems is started directly by the JDBC driver at the first connection, and therefore it runs as an embedded database. This modality provides very high performance and full portability. It's important to point out that after the word "file" in the `databaseURL`, there is the path where the files are stored on the disk, which in this case is only a relative path which tells HSQLDB to store the "test" database in the subfolder "db".

It is also possible to execute HSQLDB all in main memory without writing any file to the disk. This is done just by using a different `databaseURL`: instead of the word "file" we need to use the word "mem", and then write only the database name:

```
private static final String databaseURL="jdbc:hsqldb:mem:↘
    test";
```

2.4 Hxsql

Hxsql is a quite recent database system written in Java and it is built on HSQLDB, therefore it shares most of the features of HSQLDB and it is fully compatible with the open source version. Also the version numbering is based

on the one used by HSQLDB. The idea behind HyperCtremeSQL consists in the capability to replace the HSQLDB jar in an existing application with no change to the Java code or SQL queries.

The main differences with HSQLDB are a new indexing engine and a new persistence core. These changes provides higher performance and make Hxsql to go over some of the limitations of HSQLDB. For example Hxsql is not limited anymore to 8GB of data storage, but by default this limit is now 16GB, which can moreover be extended to 128GB. Furthermore it has also a lower footprint, which now is about 350-400KB.

Although it seems a better than HSQLDB, which is already a very good database system, its code is closed source and we can't test if the real performance are the same that we read on its home page. Anyway this is a product that need to be mentioned, especially for those who are interested in HSQLDB and in a possible upgrade.

2.5 H2

H2 is a relational database management system written in Java by Thomas Mueller, who is the same author of HypersonicSQL, the ancestor of HSQLDB, and also, originally, main developer of HSQLDB itself. In fact the name H2 stands for Hypersonic 2, however H2 does not share any code with HypersonicSQL or HSQLDB.

2.5.1 Advantages

This is an open source database written in Java, and it is built from scratch, without sharing any code with Hsql. Therefore this database system benefits from all the experience cumulated by Mueller with his previous work.

H2 is able to create both in-memory tables, as well as disk-based tables. Therefore tables can be temporary or persistent on a durable device.

As regards the administration tools, H2 includes an embedded web server and a browser based console, in order to support the portability of the database system.

In addition, H2 implements a simple form of high availability: when used in client-server mode, the database engine supports hot failover (commonly known as clustering). However, the clustering mode must be re-enabled manually after a system failure.

2.5.2 Disadvantages And Project Info

Although this database systems is reach of interesting features, it is quite recent, in fact the project started in May 2004, but it was first published in December 2005. In fact the latest stable release is the 1.0.79, while there is also a beta release, whose version number is 1.1.104.

Another disadvantage is that there is only one developer: Thomas Mueller. Moreover H2 is only a Mueller's hobby, therefore the project still don't has a good support.

2.5.3 Usage

H2 is able to run both in embedded mode and in server mode, allowing, in this second modality, the clustering of more servers, and therefore the high availability. We will show both of this modality in the following example. Also in this example, as for Prevayler, we will use the Number object, showed in the listing 2.1, as an object to be stored and retrieved.

Embedded Mode

The implementation for H2 to run in embedded mode is almost equal to the implementation of HSQLDB, showed in the listing 2.2. It's not by chance that the author of H2 was also main developed at HSQLDB project. There are only two differences: the database URL and, obviously, the JDBC driver used to connect to the database. The following example shows these two differences:

```
public class Main{

    private static final String databaseURL="jdbc:h2:file:db↘
       /test";
    private static final String username="sa";
    private static final String password="";
    private static final String driver="org.h2.Driver";

    ...
}
```

Clustering / High Availability Mode

H2 supports a simple clustering / high availability mechanism. The architecture of a clustering mode is composed by two database servers, which run on two different computers, and on both computers there is a copy of the

same database. If both servers run, each database operation is executed on both computers. If one server fails (power, hardware or network failure), the other server can still continue to work. From this point, the operations will be executed only by one server until the other server is restored. Clustering can only be used in server mode (the embedded mode does not support clustering). It is possible to restore the cluster without stopping the server, however it is critical that no other application is changing the data in the first database while the second database is restored, so restoring the cluster is currently only a manual process.

In this example the two databases reside on the same computer, but we will use them as they were on two different machines, and therefore with remote access (we will not use localhost as an address to connect to the database). The following steps describe how to make it runs in clustering mode:

1. We need to create 2 different directories, server1 and server2, that will store our databases.

2. To start the first server we need to use the following command:

```
java -cp h2.jar org.h2.tools.Server -tcp -tcpAllowOthers↘
    -tcpPort 9101 -baseDir server1
```

   The option "-cp" specifies the classpath to h2.jar, if it's not in the CLASSPATH, and the option -baseDir specifies the path to the directory where your server will store the database.

3. Subsequently we need to start the second server with the following command:

```
java -cp h2.jar org.h2.tools.Server -tcp -tcpAllowOthers↘
    -tcpPort 9102 -baseDir server2
```

4. After the two server are running, we need to use the "CreateCluster" tool to initialize the cluster. This will automatically create a new, empty database if it does not exist. The command is the following:

```
java -cp h2.jar org.h2.tools.CreateCluster -urlSource ↘
   jdbc:h2:tcp://192.168.150.151:9101/test -urlTarget ↘
   jdbc:h2:tcp://192.168.150.151:9102/test -user sa -↘
   serverlist 192.168.150.151:9101,192.168.150.151:9102
```

5. It is now possible to connect to the databases using the following JDBC URL:

```
jdbc:h2:tcp://192.168.150.151:9101,192.168.150.151:9102/↘
   test
```

This clustering mode can now be tested simply by using the example code of the embedded modality and only changing the database URL. Now H2 is providing high availability and you can also stop one server (by killing the process), without noticing any issue, because the other server is still running, and therefore the database is still accessible. Although this works, when a server crash, to restore the cluster, you first need to delete the database that failed, then restart the server that was stopped, and re-run the CreateCluster tool.

2.6 Db4o

Db4o, database for objects, is an high-performance, embeddedable open source object database for Java and .NET developers. It is developed, com-

mercially licensed and supported by db4objects, Inc., which is a privately-held company based in San Mateo, California. The company was incorporated in 2004 under CEO Christof Wittig, with the financial backing of top-tier Silicon Valley investors including Mark Leslie, founding CEO of Veritas, Vinod Khosla, founding CEO of SUN Microsystems, and Jerry Fiddler, founding CEO of Wind River.

2.6.1 Advantages

Db4o is an open source database system with a small footprint and it is much more faster than traditional relational database management systems. Object persistence is easy with db4o, it can be done with only 1 line of code.

It can run both in embedded and server mode and it can store the data on the disk or keep them all in main memory only. And since Db4o supports both Java and .Net, it is possible to access the same database from these 2 different platforms.

In addition db4o supports automatic object schema evolution for the basic class model changes (field name deletion/addition). More complex class model modifications, like field name change, field type change, hierarchy move are not automated out-of-the box, but can be automated by writing a utility update program.

Rather than using string-based APIs, such as SQL, native queries allow developers to simply use the programming language itself (e.g., Java, C#, or VB.NET) to access the database and thus avoid a constant, productivity-reducing context switch between programming language and data access API.

2.6.2 Disadvantages

Db4o can run both as in-memory database system or as a disk based database, but it doesn't have an in-memory mode with some form of durability (e.g. with a transaction log file). Moreover the in-memory mode is not so much more performant than the disk based mode.

For some native queries db4o will have to instantiate some of the persistent objects to run the native query code, thus slowing down the database's performance.

There's not much "out-of-the-box-support" for clustering several db4o-databases at the moment. There's a simple query-a-cluster-implementation in com.db4o.cluster which could guide you only as a starting point.

2.6.3 Project Info

This project started in 2000 and it is continuously updated: there is also a continuous build available for download every day. The latest version is the 7.6 and the number of developers working at this project are about 20.

According to db4o community website the amount of registered members has grown to over 45000 providing various sources of documentation: tutorial, reference documentation, API documentation, online paircasts and blogs.

2.6.4 Usage

Also in this example, as for Prevayler, we will use the Number object, showed in the listing 2.1, as an object to be stored and retrieved. In the following code we can see how it is easy to use Db4o:

```
public class Main {

  private static final String databaseFileName = "test";

  public static void main(String[] args) throws Exception {
    ObjectContainer db = createEmptyDB();
    fillDB(db);
    simpleQuery(db);
    nativeQuery(db);
    db.close();
  }

  private static void nativeQuery(ObjectContainer db) {
//    List<Number> numberList = db.query(new NumberPredicate
    ());
    List<Number> numberList = db.query(new Predicate<Number
        >() {
      public boolean match(Number candidate) {
        return candidate.getNumber() > 2;
      }
    });
    for (Number number : numberList)
      System.out.println(number.getNumber());
  }

  private static void simpleQuery(ObjectContainer db) {
    ObjectSet<Number> numbers = db.queryByExample(new Number
        (2));
    while (numbers.hasNext())
      System.out.println(numbers.next().getNumber());
  }

  private static void fillDB(ObjectContainer db) {
    db.store(new Number(1));
    db.store(new Number(2));
    db.store(new Number(2));
    db.store(new Number(3));
    db.store(new Number(4));
  }

  private static ObjectContainer createEmptyDB() {
    new File(databaseFileName).delete();
    ObjectContainer db = Db4o.openFile(databaseFileName);
    return db;
  }

}
```

Note that the Number class here does not require any interface implementations, annotations or attributes added. It can be just any application

class including third-party classes contained in referenced libraries. All field objects (including collections) are saved automatically. This feature speed up the whole process of developing an application, because the persistence layer become very easy both to write and change.

The example previously showed described how to run Db4o in a disk based mode, but if we want to execute it all in main memory, we need to change just few lines of code during the creation of the database, as reported in the following code:

```
Configuration configuration = Db4o.newConfiguration();
MemoryIoAdapter memoryIoAdapter = new MemoryIoAdapter();
configuration.io(memoryIoAdapter);
ObjectContainer db = Db4o.openFile(configuration,↘
    databaseFileName);
```

Note that Db4o by default has not high performance in read operations. If you want to speed them up you can use an index, like a relational database system. The following code shows how it is possible to create an index based on the field called "number":

```
configuration.objectClass(Number.class).objectField("number"↘
    ).indexed(true);
```

2.7 Derby

Apache Derby is a Java relational database management system that can be embedded in Java programs and used for online transaction processing.

2.7.1 Advantages

Apache Derby is an open source project and it is currently distribuited as Sun Java DB. It has a 2MB disk space footprint and it can be easily embedded in

Java programs. The embedded database engine is a full functioning relational database engine and therefore allow a full support to transactions.

In addition Derby can run as a server database system. A network server increases the reach of the Derby database engine by providing traditional client server functionality.

Derby offers the following administration utilities:

- ij: a tool that allows SQL scripts to be executed against any JDBC database;
- dblook: schema extraction tool for a Derby database;
- sysinfo: utility to display version numbers and class path.

2.7.2 Disadvantages

Derby can't be run as a totally in-memory database, or in other words, it can't run only in main memory. Therefore if you don't have any durability need, you can't benefit of the higher performance granted by the execution of a database only in main memory.

2.7.3 Project Info

This project was previously distributed as IBM Cloudscape until IBM open sourced it as Derby, with the hope to accelerate development of Java-based applications and drive more innovation around Linux and Java. At the moment Apache Derby is distributed by Sun as Java DB.

Derby project started in 2005, but its progenitor was first released in 1997. The latest version is the 10.4.1.3 released at 24/04/2008. The development team is composed by 32 developers, standing at the information on ohloh.net.

2.7.4 Usage

Using Derby as an embedded db is really simple. You need only to set the appropriate database URL and the driver, without specifing any username and password.

Also in this example, as for Prevayler, we will use the Number object, showed in the listing 2.1, as an object to be stored and retrieved. Moreover the code for this example is the same used for HSQLDB and H2, and it is showed in the listing 2.2, involving creates/drops table and inserts/retrieves. In the following lines of code we will just show the differences between the implementation of HSQLDB and Derby:

```
public class Main{

  private static final String databaseURL="jdbc:derby:↘
     derbyDB;create=true";
  private static final String driver="org.apache.derby.jdbc.↘
     EmbeddedDriver";
  ...

}
```

2.8 SQLite

asdasd

2.8.1 Advantages

asdasd

2.8.2 Disadvantages

asdasd

2.8.3 Project Info

asdasd

2.9 Firebird

asdasd

2.9.1 Advantages

asdasd

2.9.2 Disadvantages

asdasd

2.9.3 Project Info

asdasd

2.10 MySql

asdasd

2.10.1 Advantages

asdasd

2.10.2 Disadvantages

asdasd

2.10.3 Project Info

asdasd

2.11 ExtremeDB

EXtremeDB is an ACID-compliant embedded database management system (DBMS) designed for embedded systems and applications with real-time performance requirements. It is a proprietary software of McObject company and therefore it is not open source. As most IMDBs, also eXtremeDB has a small footprint of about 50K. In addition it provides both native API for C/C++ language, and SQL API, called eXtremeSQL which implements much of the ANSI SQL-89 specification.

For software applications that require high availability, eXtremeDB High Availability Edition is designed to ensure that changes to a master database

and identical standby databases succeed or fail together, and enables deployment of multiple fully synchronized eXtremeDB-HA databases within the same hardware device or across multiple, widely distributed systems. A predictable response time is created via the HA protocol's time cognizance if a master or replica database fails to respond within a pre-set limit, it is automatically decommissioned and replaced in its role by a responding eXtremeDB-HA database instance.

EXtremeDB-64 increases maximum database size by adding support for 64-bit micro-processors and operating systems, and therefore it increases the amount of memory a system can address from approximately 3GB to more than one terabyte. This database systems is also highly portable, in fact there are many platforms supported by McObject, such as linux and real time linux distribution, sun solaris, RTX real time extension for windows, and many others.

2.12 Polyhedra

* Closed source * Polyhedra is a family of relational database management systems * The original version of Polyhedra was an in-memory database management system which could be used in high availability configurations * in 2006 Polyhedra FlashLite was introduced to allow databases to be stored in Flash memory * All versions are fully client-server to ensure the data is protected form misbehaving application software, and use the same SQL, JDBC interfaces * Cross-platform * Available in C/C++ and Java * its fault-tolerance model is based on the hot-standby approach (to minimise

hardware costs) rather than clustering (which is better for load-sharing) * The first release is in 1993 * The last release is 7.1 in 28/09/2007

2.13 TimesTen

* TimesTen provides a family of real-time infrastructure software products designed for low latency, high-volume data, event and transaction management * TimesTen's core DataServer product uses in-memory database technology to speed access to data * Offers a Cache product that interoperates with an Oracle database backend * Supports jdbc interface * The project started in 1996 * Acquired by Oracle in 2005

2.14 Csql

Supports only a limited set of features Csql should offer Atomicity, Consinstency and Isolation, but no Durability Supports jdbc interface It should support SQL DML statements which contains only single table It is used mostly as a cache for a disk-based database Last update date at 02/06/2008

2.15 SolidDB

* It started in 1992 * It is a relational dbms offered by Solid, an IBM Company * Includes both in-memory and on-disk engines * Can be configured using a hot-standby architecture to achieve six 9's data availability (99.9999%) with no single point of failure * Two-node architecture can also be used for

load-balancing * SolidDB 6 has shown linear scalability from 2 to 16 cores * Allows full ACID transactions

2.16 MonetDB

* Open spource * Designed to provide high performance on complex queries against large databases * Focus its query optimization effort on exploiting CPU caches * Uses a sql and a jdbc interface * The development of MonetDb? started in 1979 * And became a open source project in 2003 * The last release is 4.16 and date at 26/02/2007

2.17 RDM Embedded

* Supports both disk based and in-memory tables * Fully ACID transaction compliant * RDM Embeddeds legacy has 20+ years of history * The Java API is based on Java Native Interface technology * RDM Embedded has implemented a SQL API set to support applications that manage the database through SQL commands * RDM Embedded Mirroring System assists developers in creating Fault Tolerant applications; however the design does not attempt to make fault tolerance an engine featur * With RDM Embedded advanced replication engine, application databases can be replicated for fault tolerance and high availability

2.18 FastDB

* Open source * It is a object-relational main memory embedded database system * It is written in C++ * It is possible to use a Java Native Interface to access to the database * There is only 1 developer * The project started in 2007

2.19 QuiLogic

* Available for many platform, including .NET, but not for Java * The maximum size of the database is about 2GB

2.20 Pico4v2

2.20.1 Advantages

2.20.2 Disadvantages

2.20.3 Project Info

2.20.4 Usage

2.21 Conclusion

Tabular result here!

Part II

Performance Analysis

Chapter 3

Performance Analisys Introduction

"Every database vendor cites benchmark data that proves its database is the fastest, and each vendor chooses the benchmark test that presents its product in the most favorable light" [Burleson 02].

In this chapter we will introduce the database performance analysis' problem, and so we will discuss about how to measure the performance, how to understand which is the fastest and how to choose the best database depending on your needs. This preface is not specific to in-memory databases, but can be generalized for every DBMS.

3.1 Impartiality Problem

The most important problem related to performance analysis, and benchmarking, is the validity, understood as impartiality, of the results. It's im-

possible to develop a benchmark which produces the same results obtained by real applications, and it's very hard to have the results similar to real performance too. Moreover every benchmark may produce valid results only for a small set of applications, and therefore there is the need to use different benchmarks. All these difficulties were emphasized by vendors who, with benchmarks wars and benchmarketing, complained for fake results, and consequently brought to a lack of trust in benchmarks. In this section we discuss about these difficulties, starting from a benchmark's definition until an analysis of the main differences in benchmarks which produce different results.

3.1.1 Benchmark Introduction

Before starting all this discussion, it's necessary to explain the benchmark's meaning in computer science. The term benchmark refers to the act of running a program, or a set of programs, against an object in order to evaluate its performance, but it is also mostly used to represent the program or programs themselves. In this work we refer to software benchmarks run against database management systems. The main purpose of a benchmark is to compare the performance of different systems across different platforms. With the advance of computer architecture, it is become more difficult to evaluate system's performance only by reading its specifications. Therefore a suite of tests is developed to compare different architecture. This suite of tests, intended as validation suite, is also used to verify the correctness, the properties of a software.

In particular benchmarks mimics specific workload giving a good measure of real world performance. But ideally a benchmark should substitute the application simulated only if it is unavailable, because only the real application can show real performance. Moreover when performance is critical, the only benchmark that matters is the inteded workload.

The benchmark's world is full of challenges starting from a benchmark development, to the interpretetion of the results. Famous benchmark tend to become useless after some time, because the vendors tune their products specifically for that benchmark. In addition most benchmarks focus only on the speed which is expressed in term of throughput, and forget many other important features. Moreover many server architectures degrade drastically near 100% level of usage, and this is a problem which is not always taken into consideration. Some vendor may publish benchmarks at continuous at about 80% usage, and, on the other hand, some other benchmark doesn't take this problem into account, and execute only tests at 100% usage.

3.1.2 Benchmark History

As already highlighted in the previous paragraph, with the advance of computer architecture the performance analisys became more difficult: it's not possible to evaluate a system only by reading its specifications. A computer program, a benchmark, simulating a specific workload, became a solution to this problem.

The first benchmarks were developed internally by each vendor to compare their database's performance against the competitors. But when these

results were published, they weren't considered reliable, because there was an evident conflict of interest between the vendor and its database [Gray 93]. This problem was not solved by a benchmark pubblication by a third party, which usually brought to a benchmark war.

Benchmark Wars

Also when benchmarks were published by third parties, or even by other competitors, the results were always discredited by losing vendors, who complained for the numbers, starting often a benchmark war. A benchmark war started when a loser of an important and visible benchmark reran it using specific enhancements for that particular test, making them get winning numbers. Then the opponent vendor again reran the test using better enhancements made by a "one-star" guru. And so on to "five-star" gurus. Every result so obtained by a benchmark could not be considered a valid result, firstly because the vendors themselves don't give any credit to the result, secondly because the result, with its relative numbers, changes too much frequently.

Benchmarketing

Benchmarketing is a variation of benchmark wars. Due to domain-specific benchmarks there was always a benchmark which rated a particular system the best. Such benchmarks may perform only a specific operation promoting one database instead than others. For example a test may execute only one type of transaction, or more reads than writes and therefore promote an in-memory database over a traditional DBMS. To summarize, a benchmark

may simulate different scenarios favouring a particular database. Therefore each vendor promotes only the benchmarks which highlight the strengths of their product, trying to impose them as a standard. This leaded to a ploriferation of confuse benchmarks and didn't bring any benefit to the benchmark's reputation.

Although these phenomenons were drastically reduced by the foundation of the Transaction Processing Performance Council (TPC) in 1988, they are still alive nowday.

3.1.3 Benchmark Differences

It is now evident how hard is to understand which database is the fastest. Benchmarks cannot be properly used to analyze the performance of several databases and choose simply the best. Every benchmark has a bias, testing some particular aspect of database systems, such as writes, reads, transactions and so on. Moreover every benchmark operation can be implemented in different ways by the same database: creating a connection for every operation or using a connection pool; use different transaction isolation level; load the whole database in RAM or split it in different hard disk partition; etc. Furthermore the same benchmark, with the same implemention for every database, can show dissimilar results when it runs on different platforms (hardware and software).

All these differences are grouped by three categories:

1. Test scenario: reads, writes, etc.

2. Test implementation: there are different way to implement the same transaction.

3. Execution platform.

3.1.4 The Axiom

By this reasoning comes out the axiom which will guide the following chapters, and the work behind this thesis:

"There is not a slower or a faster database: databases are only slower or faster given a specific set of criteria in a given benchmark".

This sentence comes from an article by Bernt Johnsen [Bernt 05], who, in response to a benchmark comparing HSQL and Derby, stated that it's easy to make a benchmark, but it is always hard to communicate the meaning of the results. The interpretation depends on too many criteria, so that it's not easy to say which database is the fastest, unless specifying the set of criteria used in the benchmark.

3.2 Measures

Even if we cannot state which database is the best by analyzing their performances on standard benchmarks, we can still measure other parameters that can be later interpreted to choose the database system which best fit our needs.

When choosing a database, the most important features to evaluate and compare are:

Throughput : is the number of transactions per second a database can sustain. This is the most important feature to consider when evaluating a database system. The most representative application scenario to understand the meaning of the throughput is an on-line transaction processing system. This kind of application requires the database to sustain a certain number of transactions per second, based on the number of users, and not every database can suite the needs of the application itself. Another example is real time applications: they require even higher throughput as well as very low latencies. Therefore it is crucial to understand if a certain database is suitable for an application in terms of transactions per second.

Latency/responsiveness : is the time the database takes to execute a single transaction. This is the most important parameter in real application where the response time is a vital feature. In some case, where transactions are executed synchronizedly by a single user, latency can be measured as the inverse of throughput.

CPU load : is the average percentage of the CPU processing time used by the database. It becomes an even more important parameter when the database is not the only service running on a server. CPU is a precious and limited resource, shared by all processes in a particular machine, and although database systems are usually deployed on dedicated servers, embedded databases live collated with the application that uses them. Many in-memory database don't even have a "stand aone" version and are mainly used as embedded db.

Disk I/O : is the measure of the amount of data transferred to and from the disk. It is often a bottleneck for every traditional databases. Altought pure in-memory databases never access to the disk, when adding durability through a transaction log file, disk I/O is a bottleneck for IMDBs too. This parameter could be considered less important since all databases will incure in the same performance bottleneck, but it is possible that different DB coudl use the disk in different ways and suffer more or less impacts from it.

File size : is the size of the database image on the hard disk. While traditional DBMS' store objects (tables), indexes and a transaction log file on the file system, in-memory databases usually store only a journal file containing all the transaction executed on the database. Altought this may seem a small file, it can become very huge, even more than the database image. This measure is interesting whereas each database use different data structures.

RAM usage : is the quantity of RAM a process uses while running. Talking about in-memory database, this can be another way to measure the size of the database image. In addition, this is a critical value to take in mind: IMDBs only works correctly and efficiently under the hypothesis that the RAM is enough to contain the whole database. Therefore this can be an important parameter to take into consideration when deciding if a DBMS fits your requirements.

Startup time is the time the database needs to become operational. This is a very important parameters for applications which need high avail-

ability, and in this case the startup time plays a crucial role to respect the maximum time the application can be off-line or not completely operative.

Shutdown time is the time the database takes to shut down and kill the process.

3.3 Choosing a Database

From the previous sections, it's now clear how difficult it is to use benchmarks to prove which database is the fastest. Even with a fair benchmark, which can be very useful to understand the performance of databases, it is still difficult to choose a database: performance is only one factor to consider when evaluating a database [Burleson 02].

Other factors to consider are:

- The cost of ownership.
- The availability of trained DBAs.
- The vendor's technical support.
- The hardware on which the database will be deployed.
- The operating system which will support the database.

In other words, it is very difficult to choose the right database for our needs, and, of course, while evaluating databases and benchmarks there is absolutely *no winner*.

For this reason, what we will introduce in the next chapter is a suite of tests which has been developed with these concepts in mind. What we tries to achieve is a way for people to judge, analyze, verify and evaluate any DBMS (especially in-memory databases) with a chance of developing their own set of test and run the test suite on any hardware and under (virtaully) any operating system. The application is Java-based and so it ansures the widest coverage of both hardware and operating systems.

Chapter 4

Test Suite

In this chapter we are going to discuss the first of the three big differences, described in paragraph 3.1.3, which make a database faster or slower: the test scenarios used to run a benchmark and to analyze database's performance. These test scenarios will be used to build test cases, a collection of different scenarios, one or more, executed concurrently. All these test cases will compose the test suite we are going to use for database benchmarking. Therefore in this chapter besides a tests' description, a suite of test will be created, allowing us to analyze the databases' features of our interest.

The tests are divided into three categories: base test case, load test case and acid test case. The first category contains simple tests configured as a race between databases, and it is used to obtain an approximate idea in an early stage. Instead, the second category simulate real application scenario, and it can give us a detailed analysis of the real database's performance. Finally the last category is built to understand if the database is ACID compliant.

4.1 Why Define Test Scenarios

The axiom, previously described in paragraph 3.1.4 at page 50, expresses clearly the difficulty to analyze databases' performance and how every result obtained in a given benchmark depends on the specific set of criteria used in the benchmark itself. In paragraph 3.1.3 there is also a description of the three major categories in which the criteria fall. The first of these categories is test scenarios: different test scenarios may show completely different performance results. It's not possible to avoid this behavior, but we can define clearly every tests so that we will be aware of the differences between them.

Moreover we can model these tests in a way they simulate real application scenarios, so that real performances will be very close to the results obtained by the tests.

4.2 Base Test Case

Base test case is a collection of very simple tests, which are configured mostly as a race. This is exactly what the majority of benchmarks does, particularly Poleposition [Grehan 05].

Every test can execute different read/write operations on the database, such operations are grouped into *tasks*. The word *task*, used extensively in the following paragraphs, refers to a collection of operations. Each task is then repeated, and since all the operations in a task are synchronized and executed consecutively, the number of *task per second* is the same of the *transaction per second* of each operation involved by the same task. Therefore *task, or*

task per second and *transaction, or transaction per second* can be used in this contest without any distinction, and they all refer to the operations' *iteration*.

The key features of these kind of test are:

- A fixed number of tasks' iteration before the test stops: so tests are configured as a race where every database must execute a certain number of transaction.
- A fixed amount of time before the test stop: as an alternative to a fixed number of transaction, every test run for a specific amount of time, executing the maximum number of iteration.
- Different kinds of objects: tests must be able to create/retrieve/update/delete simple flat objects with few fields, or complex flat objects with many fields, or still hierarchical objects, and so on. This feature let us test effectively the performance on the objects used in the domain of our interest.
- Single task: to keep these test simple it's better to avoid concurrent tests, but it is still possible to implement concurrently the different operation executed on the database.

4.2.1 Race Test: Timed

We already said base tests are configured inherently as a race. These tests can be used to show the maximum throughput the database can reach doing a particular operation on a specific object. Metaphorically it's like a rocket

car in the desert trying to reach its speed limit. In a real application this is rarely useful, but can give us an idea of the database's limits.

In order to create a test scenario we have to define the domain object/s involved in the test, the operations on which the test loop and when the test should stop. For example:

- The object represent a person, with only two fields: an id number and a name. This is a simple flat object.
- The only operation executed by the test is a write operation: every time an object will be added to the database.
- 60 seconds is the stop condition of the test: after 60 seconds of iteration of writes (objects person inserted in the database) the test stops.

Clearly the time is not a significant measure, every test's execution run for the same amount of time: 60 seconds. Instead of the time, a more interesting measure to take is the throughput. This test shows the maximum theoretical value for the throughput, which in a real usage scenario will never be outperformed.

This test, and its results, are not useful to understand the real performance and potentiality of the database and so they are not useful to choose the database for our needs, but it can be used in an early stage to reduce the databases' set which will be analyzed extensively with further tests. In other words we can throw away every databases whose maximum throughput is not enough for our needs.

4.2.2 Race Test: Transactions

This test is really similar to the previous test. The only difference is in the stop condition:

- The object represent a person, with only two fields: an id number and a name. This is a simple flat object (same as before).
- The only operation executed by the test is a write operation: every time an object will be added to the database (same as before).
- 1.000.000 of iteration is the stop condition of the test: when 1.000.000 of writes are executed (objects person inserted in the database) the test stops.

While in the previous test the duration was meaningless, this time, like in a race, it shows which database is the fastest (the winner of the race). Despite this, the duration is still of little use. Also for the throughput, the same considerations made before are still valid.

But this new test is useful also to take other interesting measures, such as the file size, which, as explained in par. 3.2, can become very huge, even more than the database image, and therefore it is a critical measure. We can evaluate how the file size increase with the number of objects (person) inserted in the database. This is because, differently to the previous one, every test's execution perform a fixed amount of iteration.

4.3 Load Test Case

Restrictions introduced by base tests are very simple to read and they can be useful for a first look and for fast and approximate results. These restrictions are due to the impossibility to simulate real complex application scenarios. Substantially the main restrictions are:

1. Real application scenario are rarely single thread/task: to stick to the axiom at paragraph 3.1.4, the best way to understand which database fits our needs is to make test scenarios as much realistic as possible.

2. The second restriction is a direct result of the first: trying to make test scenarios more realistic, it is necessary to introduce some sort of control for the throughput. When a test stress the database engine to its maximum level, some other mechanisms may not work properly, such as the garbage collector, caching, indexing, etc. In addition, when a test concurrently accesses the database, this restriction is even more evident: if a thread is not limited in its transactions per second, it will impact the performances of other threads, not to mention the locking that will occur on the DB.

Base test case offer very important and useful results, but when it comes to test the average load of a real application of our interest, they are not enough. These restrictions are solved by load test case, which allow a better simulation of the application scenario. The key features of these tests are:

- Multi-tasking: different task, and therefore different sets of operations, can be executed concurrently against the database.

- A bond on the throughput: in order to simulate real application usage, for every task it is possible to specify the maximum throughput in terms of iteration per second, if the database can reach it.

From these features comes the name "load test case", which means the simulation of an average, or specific, load on a certain database. This idea, and the need of this kind of tests, can simply be exaplined by a Bernt's metaphor, who has been already cited for the axiom at paragraph 3.1.4. The metaphor is:

"I don't buy the fastest car in the market. I don't even buy the fastest car I can afford. I buy a car I can afford that fits my needs... e.g. drive to the northern Norway with 2 grown-ups, 3 kids, a kayak, sleeping bags, a large tent, glacier climbing equipment, food, clothes and so on. Can't do that with a Lamborghini" [Bernt 05].

This metaphor explain exactly the need not for the fastest database, but for a database which can sustain the load of the application without any complexity during its normal functioning, such as a database snapshot freezing all writes operation, or a bug/memory leak making the database crash.

4.3.1 Real Time Prepaid System

Keeping these features in mind, we want to create a load test case to analyze exactly which performance a database system offers for a practical use case: a real time prepaid system, such as a telephone company. Basically the test is the concurrent execution of 3 different task, each simulating complex sce-

narios: services authorization and management, balance checks and accounts management.

We start describing the domain objects used by this test, and then we move on to the analysis of the different tasks involved.

Domain Objects

Domain objects involved by this test are typical for a telephone company who handle telephone calls for every person, which is rappresented by an account, and which can access to many service through a web application, offered by the company for its customers. There are four domain objects used byt this test:

Account : this object represent a customer in the real time prepaid system. It contains all the typical information needed by a telephone company, such as the balance, the type of subscription, etc. In other words, this is a complex flat object, that has many private fields but no hierarchies. In a real application there are millions of accounts instantiated in the database, one for every customer. This dimension is also very important to make the simulation as real as possible, in order to get realistic results.

MSISDN : it is the unique number which identify a subscriber in a GSM or UMTS mobile network, it is the telephone number. Each msisdn object is linked to an account. Therefore there are also millions of MSISDN objects in the database, referring to a real application.

Webuser : this represents the customer logged in the company web site. It

contains the username and the password to access to web services. In common with MSISDN, it is linked to an account too, and both are simple flat objects: a very simple object with few fields.

Session : when a customer start a call, an object session is created, and it keep all the information about the call which is going on, until the call ends. After the call this object is deleted. A session contains the start time of the call, the time of the last unit and all the relevant information for the authorization process to take place. Each session references to an account, the one who started the call. But there are not as many sessions as accounts, not everyone will start a call at the same time, except New Year's Day! A reasonable size for the session objects is the number of hundreds of thousands concurrently going on.

Balance Check Task

This task simulate a customer checking his residual balance for the prepaid card, a SIM in the case of the telephone company. First, the customer logins in the website or calls the dedicated number, and then receives all the details on his balance.

Each task, as already explained, corresponds to a transaction composed of different operations. This leads to illustrate how the task work in terms of operations executed:

1. *Read* randomly the MSISDN if the customer makes a call or the webuser if he access to the website.

2. *Read* the account referenced by the MSISDN or webuser, and then

check the residual balance, a simple account's field.

So this task executes a total of two read for every iteration.

Another important parameter to make this task a part of a load test is the amount of transactions per second this task should sustain. The average throughput for check balance task, considering there are millions of accounts in the database, is about ten transactions per second.

Accounts Management Task

The account management task simulates the subscription of new customers and consists pf the creation of an account object and the relative MSISDN and webuser.

The operations involved by this task are:

1. *Write* of a new account: all informations of a customer are inserted in the system by creating a new account object.

2. *Write* of a new MSISDN, containing the telephone number of the new subscriber.

3. *Write* of a new webuser with username and password for the customer.

To sum up, this task executes three writes on the database for every iteration.

To simulate a real scenario, the average throughput is about ten transactions per second.

Services Authorization and Management Task

This task simulates a call started by a customer. After checking the balance, and therefore if the account is allowed to start a call or use a service, a

new session is created and updated every 30 seconds, the unit time. On the average a call lasts about two minutes. When a call ends the session is deleted.

In terms of operations executed on the database, it can be described as follow:

1. *Read* the MSISDN of the customer starting the call.
2. *Read* the account referenced by the MSISDN, and its residual balance, and the other parameters, to check the user's permissions for the service requested.
3. *Write* a new object session and *update* account's balance.
4. *Update* the session every 30 seconds and the account's balance.
5. *Delete* the session after the call is ended.

This is the most complex and important task, because it represents a very frequent task. No wonder if the average throughput is about 2000 transactions per second.

Real Time Prepaid System Summary

To sum up this whole load test we have to define the objects used by the test, the task concurrently executed, and when the test stop.

There are four domain objects involved by this test. Three of these are simple flat objects (objects with few private fields): MSISDN, webuser and session. The forth is a complex flat object (objects with many private fields): account. To test the system in its normal operational scenario we will have

the system preloaded by some millions of objects before starting the actual test.

Three task are executed concurrently. One is the major task, simulating a customer's call, and it runs 2000 times per second. This task is the bottleneck of a real application, and it could be also tested alone in a base test to understand the database's limits in running this task, and therefore to have an idea of the database potentiality. The other two are secondary tasks, in fact the throughput is limited to ten transactions per second.

This test is a load test and therefore we are not interested in stopping it after a certain amount of transactions executed. But what we want is to let the test run for many minutes until to many hours to understand if the database can sustain the load generated by the test.

4.4 Acid Test Case

Base and load test case are used to analyze databases' performance, but, as already said, a database is not made only by performance: there are many other parameters to take into consideration before judging a database, such as ACID properties. To make these tests an all-round test suite, we could add some tests for verifying databases' ACID properties. The above corresponds to the work done by the Transaction Processing Performance Council with their TPC-C benchmark [Council 07]: *performance is not the only benchmark's goal, but ACID properties are tested too.*

Especially the durability property is most interesting to analyze when talking about in-memory databases. Pure IMDB have no durability at all,

while it can be added in different degree, as already described in paragraph 1.5. Therefore would be very useful to know how strong is the durability solution implemented by the in-memory database, although listening to vendors' words their database management systems should offer always strong durability.

Nevertheless, testing ACID properties is not a simple goal to achieve. Particularly the durability is the hardest property to be tested. Also the tests developed by TPC are not intended to be an *exhaustive quality assurance test.*

Chapter 5

Database Benchmark Softwares Overview

The difficulty to analyze objectively databases' performance has already been explained and understood in chapter 3. The major problems were found in the differences of test scenarios, implementation and execution platform. In the previous chapter we tried to minimize the first issue by modeling different test scenarios like real application scenarios. This is not a panacea, but in this way results produced by these tests are more valid for the applications they simulate than generic tests.

The focus is now moved on the second and third issues: the tests' implementation and the execution platform, in other words an application benchmark used to run test scenarios. In this chapter we will analyze, with different levels of detail, various open source benchmarks, trying to find a benchmark for our needs, or, in the case it's not suitable, to take some ideas for a future development of a new benchmark application.

5.1 Benchmark Requirements

Before starting the analysis, it's necessary to make clear the requirements a benchmark should have to run our tests. These requirements will give us a set of parameters which can be used to evaluate properly a benchmark. But before stating all the requirements, it's important to point out some general characteristics for a benchmark:

- More than one single metric: a benchmark should show the results with different metrics in order to provide a better awareness of the results themselves.
- *The more general the benchmark, then less useful it is for anything in particular*: it's impossible to make a benchmark for everything, it's better to put some limitations and produce proper results for a specific task.
- Based on a defined workload: only the essential attributes of a workload should be used to create a benchmark.

These general characteristics and some of the following requirements are taken by the TPC presentation at the sigmoid conference in 1997 [DeWitt 97], a bit old work, but this council is still active and of course they know how to make a benchmark.

The requirements we are going to analyze are: portability, understandability, flexibility, detailed report, visual report, both for relational and object databases and easy to use. In addition we will illustrate some benefits a benchmark should bring.

5.1.1 Portability

A benchmark is a set of tests executed against a system, a database in this case, to analyze the performance with the main purpose to compare them. The comparison is usually between different systems when you are choosing the best in term of performance for your needs. But when you already have chosen the system it is also very useful to compare the benchmark's results on the same database between different hardware platform. In order to make this possible, both the system to be tested and the benchmark must be able to run on several platforms. While databases usually run on different operating systems, it's not the same for benchmarks. When the benchmark runs on a client's network, and therefore it is for client-server databases, it can be considered portable, because the execution platform doesn't matter. But when we are using embedded databases, the benchmark itself must live togheter with the database system, and therefore able to run on different OS, which means it is portable.

Thus, when testing embedded databases too, a portable benchmark can help us not only in the decision of the best database system given a certain platform, but also in choosing a suitable platform for our database.

5.1.2 Understandability

This feature may seem obvious, but it is not when we are reading thousands of incomprehensible words or graphs with many lines and no legends or captions on the axis. The capability of being understood is essential to make the benchmark useful. Understandability is intended in terms of clear re-

sults and their easy interpretation: as already explained in paragraph 3.1.2 benchmarks' results are easy to be tricked and cheated, and therefore they must be clear to avoid any religion's war and consequently the distortion of the results.

Moreover, it's important to take in mind there are different stakeholders interested in benchmarks' results: engineers working at the database engine, managers selling the system and customers interested in buying a database. Thus a good benchmark must be able to be easily understood by many people and provide a clear view or views of the results.

5.1.3 Flexibility

This is not a requirement for every benchmark, but it is for our purpose. We want a benchmark which can run the test suite described in chapter 4. That is the collection of tests we gathered and we want to run against a database system to understand its performance. Therefore we need a benchmark which with few programming can execute our tests.

Furthermore, recalling the axiom in paragraph 3.1.4, benchmarks' results may depend on too many criteria, and therefore a good benchmark must be flexible in order to simulate the real application scenarios we are interested in, scenarios which we may add in the future. Although we said in this paragraph's introduction a benchmark should not be general because it becomes less useful, it is also true that a good benchmark must be based on a workload, the workload of the application which will use the database system we are going to choose. Therefore flexibility is a key feature for a benchmark

when comes the need to run a own test suite based on the workload of our application.

5.1.4 Detailed Report

Recalling the first general feature for a benchmark in the introduction of this paragraph ("not only a single metric"), results must provide different measures allowing a better understanding of databases' performance. It's not by chance we are talking about understandability. These two characteristics are strictly connected. Of course many measures for every test will make the result harder to understand but will provide more informations to interpret the databases' behavior. On the other hand only one measure is easier to understand but may hide the database behavior and the reasons behind that. Therefore a compromise is needed: in each test we are not interested in all the measures that it's possible to take, but only a subset of those can be chosen. The reference measurements have been already introduced in paragraph 3.2 and they represent a subset of all possible measures for a database.

In addition more measures the benchmark's application takes, more it might alter the database performance, and therefore a compromise in the number of measures is always a good choice, but keeping in mind that only one measure will not tell much about databases performance.

5.1.5 Visual Report

In order to make the results even more understable, and to make the detailed measures more readable, visual reports will play an essential role. Graphs are

able to express thousands of numbers in one or few colored lines. Graphs are perfect not only to show the final result of a test, but also the whole trend, from the beginning to the end. Therefore any oscillation in the measures can become evident and easy to read. It will not be hidden by an average value reported at the test's end. Of course graph can also be almost useless if they report only the final value. It depends by the implementation. For example graphs such as those from Poleposition benchmark, which will be analyzed in paragraph 5.5, may be very hard to understand and they simply describe an average value.

Furthermore visual reports are usually used with the main purpose to compare system between them, and this is another task they play perfectly. Of course there are many parameters to understand before being able to read properly a graph, but then in a simple graph every line can represent a different database system, allowing an easy comparison between them. This is really appreciated by non technical people, such as managers or customers. For this purpose graphs are a must, and therefore this is a feature we want from a benchmark application.

5.1.6 Both Relational and Object Database

The benchmark we are looking for must be able to work with both relational and object databases. Relational databases are usually able to run as stand-alone server and accessed via SQL. Instead object databases are used mainly as embedded databases and accessed via a native interface for a specific programming language. Anyway in-memory databases are used mainly

as embedded databases and therefore we need a benchmark which can use embedded databases accessed by both SQL (eg: JDBC) and native interface. Although in-memory databases are not a new technology, they became widely used in the last few years, therefore many benchmarks do not provide any support for embedded databases.

When testing both relational and object databases another problem comes out: the benchmark application can test the database systems in different ways, using a relational point of view or an object oriented point of view, or both. This may produce different results in the performance analysis. The knowledge of this feature is important during the results' analysis.

5.1.7 Easy to Use

This is a commonplace for every application, nobody wants something impossible to use. As for other requirements, a compromise is always needed, especially for this feature. The ease of use, in this case, means the possibility to modify several test's parameters, such as the parameters already explained in chapter 4, and also to be able to bring some little variation to the workload simulated, in order to improve the test when the needs change. This ease of use, understood as a variation in parameters, should be granted without any modification in the application's lines of code, but for example in an xml file, or other configuration files, or with some input parameters.

Of course this may seem similar to the flexibility requirement, that we discussed in paragraph 5.1.3, but while flexibility is intended as the capability for engineers to implement completely new test scenarios, ease of use is the

capability to execute the workloads implemented with few variations, such as the initial database's state, the exact operations' order, etc. This requirement is needed for all the non technical people who want to analyze the database performance.

5.1.8 Benefits

Discussing about requirements, it's important to note also the benefits a benchmark should bring, the aims it is built around. Referring again to the TPC presentation at the sigmoid conference in 1997 [DeWitt 97], the benefits a good benchmark should bring are:

- Definition of the playing field: a benchmark should clearly point out the database performance and the hypothesis behind these performance.
- Progress's acceleration: a benchmark allow to measure the database's performance and therefore engineers are able to do a great job once an objective is measurable and repeatable.
- Schedule of the performance agenda: this is something also managers and customers can understand. Every release can have clearly declared its goals, such as X increment in the throughput, and it is easy to measure realese to realese progress.

The TPC work illustrate also how good benchmarks have a lifetime, because they firstly drive industry and technology forward, but when all reasonable advances have been made, benchmarks can become counter productive, encouraging artificial optimizations. But this reasoning falls down when the

workload, on which the benchmark is based, changes. And the modification of the workload, from little modification to completely new implementation, is a requirement for the benchmark we are looking for.

5.2 The Open Source Database Benchmark

We start now the analysis of several database benchmarks trying to understand if they are suited for our needs, or to take some idea for a future development. Discussing about these benchmarks we will illustrate advantages and disadvantages. Surfing the web and looking for databases benchmark "The Open Source Database Benchmark" is one of the first results.

5.2.1 Advantages

First of all this database benchmark is open source and this mean that everyone can contribute to the benchmark development.

A more interesting feature to point out is the test suite used by this benchmark application: it is based on AS3AP, the ANSI SQL Standard Scalable and Portable benchmark. This is a relational query benchmark divided in two section: single-user tests and multi-user tests. Therefore this is a collection of SQL script which can be used against a database with a minimum effort in benchmark's programming.

5.2.2 Disadvantages

On the other hand there are several disadvantages. This project started in early 2001 and its last release is 0.21 dated at January 2006. Therefore both

the latest release number and date are bad numbers: this software is outdated and it is still in alpha release. Also the AS3AP benchmark is outdated and very few information are available on the web.

From a relational point of view, this benchmark is portable, because it runs on a network client, but it is written in C language, and it is not portable from an embedded database point of view. When porting a program written in C from a platform to another, there is almost always some code's lines which need to be changed, especially when using sockets or other system calls.

In addition this benchmark is able to run only SQL script against relational database server, thus it needs a lot of programming for the procedures' implementation to test embedded/object databases. What's more we want primarily test embedded/in-memory databases and not relational database servers.

At last, another disadvantage is the lack of detailed visual reports. This benchmark works only by the command line interface and no graphs are produced. Moreover there is no direct comparison between different database systems: in order to compare two databases, manually the benchmark must be runned against them and then compare the throughput reported. There is also no comparison in the throughput's trend during the execution of the test.

5.2.3 Conclusion

In conclusion it is clear how this database benchmark doesn't fit to our needs. Not even one requirement is completely satisfied and some of them are taken in no consideration. The reports are neither detailed or visual, and embedded in-memory databases are not supported. With a lot of programming this benchmark can also become a good one, but it's more convenient develop a completely new benchmark than extend the open source database benchmark to fit our requirements.

What we have learned from this application is very few: only the AS3AP is a nice discovery. It is an old standard SQL test suite, which can be used to understand and to learn some new kind of test, new interesting workload. But AS3AP is very old and outdated and therefore useless.

5.3 Transaction Processing Performance Council

In the past years, without a standards body to supervise the testing and publishing, vendors begin to publish extraordinary marketing claims running specific benchmarks. The Transaction Processing Performance Council, TPC, is a non-profit corporation founded in 1988 by eight leading software and hardware companies in response to benchmarketing and benchmark wars. The TPC currently has 20 full members: AMD, BEA, Bull, Dell, EnterpriseDB, Fujitsu, Fujitsu Siemens, HP, Hitachi, IBM, INGRES, Intel, Microsoft, NEC, Netezza, Oracle, Sun Microsystems, Sybase, Teradata and Unisys.

The main aim of TPC was the provision of uniform benchmark tests. Therefore they, recnognizing that different types of applications have different workload, created several benchmarks. Besides the outdated benchmark which are TPC-A, B, D, R and W, there are the followings:

- TPC-App: an application server and web services benchmark. It simulates the activities of a business-to-business transactional application server operating in a 24x7 environment.
- TPC-C: an on-line transaction processing (OLTP) benchmark created in 1992; a complete computing environment where a population of users executes transactions against a database.
- TPC-E: a new on-line transaction processing workload which simulates a brokerage firm.
- TPC-H: a decision support benchmark. It consists of a suite of business oriented ad-hoc queries and concurrent data modifications.

5.3.1 Advantages

Of course there are many advantages when discussing about TPC benchmarks, and above all there is the experience this consortium has accumulated in the last twenty years. Since 1988 TPC is working on benchmarks expressing how benchmark results depend on the application's workload, on the system design and implementation, including the execution platform, and on the application requirements, such as high availability or strict durability.

This experience, and the big names of the members, contributed to impose TPC benchmarks as a standard, especially for OLTP performance.

Talking about standards, the most famous TPC benchmark is TPC-C, the on-line transaction processing benchmark. The scenario simulated by TPC-C is very common real application scenario, and therefore all its success. This is a moderately complex OLTP modeling a wholesale supplier managing orders, whose workload consists of five transaction types:

1. New order: enter a new order from a customer.
2. Payment: update customer balance to reflect a payment.
3. Delivery: deliver orders (done as a batch transaction).
4. Order status: retrieve status of customer's most recent order.
5. Strock level: monitor warehouse inventory.

The benchmark take a measure of the throughput, that in TPC terms is a measure of maximum sustained system performance. It is based on the number of "new order" transactions per minute while the system is executing all the other transactions types. In addition, while running the benchmark, 90% of each type of transaction must have a user response time less than 5 seconds, except "stock level" which is 20 seconds.

But throughput is not the only metric used by TPC, there is also a price/performance metric. The price is simply divided by the throughput and tell you how much is the cost for a transaction. It's important to note that this price is not the cost of the hardware, or another component. It

include all cost dimension of the system environment: terminals, backup storage, servers, software, and three year maintenance.

TPC-C requires also transactions to be ACID (atomicity, consistency, isolation and durability) and therefore TPC included different tests to demonstrate that the ACID properties are supported and enabled during the execution of the benchmark. These tests are not intended to be exhaustive quality assurance tests. Therefore these test are a necessary condition, but not a sufficient, to demonstrate the ACID properties. In the TPC-C specification document [Council 07] there is the description of different kind of scenarios which can be used to test the acidity of the database system.

Lastly TPC doesn't work only for creating good benchmarks, but also a good process for reviewing and monitoring those benchmarks. A nice metaphor written by Kim Shanley in 1998, chief operating officer at TPC, says good benchmarks are like good laws: as Aristotle said, if the laws are not obeyed, do not constitute good government. And this is the meaning for the process reviewing and monitoring. Although this last concept about process monitoring is really important, this is not an interesting feature for the benchmark application we are looking for, simply because it doesn't regards the application itself.

5.3.2 Disadvantages

The transaction processing performance council offer many advantages, but its benchmarks are not exempt from disadvantages. First of all TPC benchmarks are only specification and there is no implementation available: each

vendor must implement its own benchmark and then the council will review and monitor the process to check if the specification are kept. Although it is possible to find some free TPC-C implementation in internet, they are all unofficial benchmarks.

In addition all the available implementations, as for the official private implementations, are for relation database only and run SQL script. Not by chance all the TPC-C specification are expressed in terms of relational data model with conventional locking scheme. But in the introduction of TPC-C specification [Council 07] it is clearly stated that it's possible to use any commercial database management system, server or file system that provides a functionally equivalent implementation. The relation terms, such as "table", "row" or "column", are used only as an example of a logical data strucuteres.

Moreover all these benchmarks have been used for high end system and for vendors who produce both hardware and database management system. Although it's possible to use a proper implementation of these benchmarks for every kind of DBMS the comparison between the official results will be almost always of few interest: because the official results regards only high end systems and because the results obtained by the proper implementation will not get any credit by the TPC consortium.

5.3.3 Conclusion

To sum up the whole discussion about TPC benchmarks, especially TPC-C, it's evident how this is not suitable for our needs, but instead all this work

can be used as an inspiration. There are different reasons why this doesn't fit the requirements previously exaplined. The council itself wrote in the TPC-C specification [Council 07] a sentence that clearly state why this is not what we are looking for:

"Benchmark results are highly dependent upon workload, specific application requirements, and systems design and implementation. Relative system performance will vary as a result of these and other factors. Therefore, TPC-C should not be used as a substitute for a specific customer application benchmarking when critical capacity planning and/or product evaluation decisions are contemplated".

This is an emblematic sentence and it can be considered as an extension of the axiom enunciated in paragraph 3.1.4 where we said that performance depend on a huge number of criteria. The meaning of this sentence is that TPC-C is a good benchmark used to compare different systems for a generic purpose OLTP, without critical performance requirements. In this case, you need a benchmark which simulates the workload of your application, or the application itself. While TPC benchmarks are not flexible, they can't be used to run our test scenarios. In fact TPC-C benchmark is simply a complex application scenario, a benchmark specification, and not an application benchmark, which we could eventually extend to run our test suite.

But this benchmark and this analysis is not useless. This is a source of inspiration for an extension of our test suite: from the implementation of a new load test case similar to TPC-C (it is a standard de facto for generic OLTP); to the implementation of some, not strict, ACID test. Moreover the TPC work faced many problems in the last year in benchmark's definition,

accumulating a lot of experience, and this is a part of our work.

5.4 Apache JMeter

Apache JMeter is an Apache Jakarta project, a set of open source Java solutions and part of the Apache Software Foundation, which encourages a collaborative, consensus-based development process under an open software license. Apache Jmeter is designed to load test functional behavior and measure performance. Originally developed with the main purpose to test web applications, it has expanded to other resources, both static or dynamic: files, servlets, perl script, java objects, databases, FTP servers and more.

5.4.1 Advantages

Apache JMeter is a 100% pure Java desktop application, and therefore allowing complete portability. In addition JMeter doesn't need any installation, it's possible to download the binaries and simply execute them. This is perfectly suited for our need to test in-memory database on different platforms. Also most of the database systems we will test are written in Java e mainly for Java. In addition JMeter offers a powerful GUI which let you set your test in an easy way, and therefore ease of use is accomplished.

A key feature of JMeter is the capability to load test: there are not only the classic stress tests but load test too, which can simulate a group of user and different load types. Tests in JMeter are represented by a test plan that is a collection of user's behavior. Each user usually executes different actions, such as a SQL query or a HTTP request, simulating therefore a

certain behavior. For each user is also possible to specify several parameters in order to simulate a particular load. Figure 5.1 is an example of a JMeter configuration test for a database via JDBC. It shows some of the features previously described such as the thread group, which is a set of users with identical behavior.

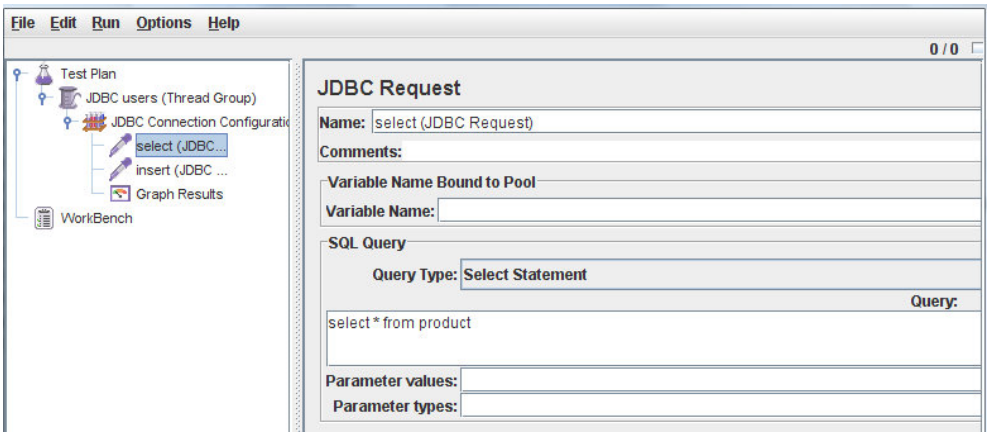

Figure 5.1: JMeter GUI

Moreover these load tests can also be executed concurrently by JMeter and in two different ways:

- Each thread group can simulate a collection of users with the same behavior. Each user is a thread, and the amount of thread is a parameter of the thread group. Therefore it's possible to simulate different users with the same behavior.

- It is also possible to implement different thread groups and so different behavior.

The ability to execute load test and concurrent test give us all the flexibility

we need to run our test suite, particularly load test case such as the real time prepaid system explained at paragraph 4.3.

Another important feature in favor of JMeter is the capability to load test also databases via JDBC. As already said the figure 5.1 is an example of a JMeter configuration test for a database via JDBC. It shows how to simulate a group of user with the same behavior: they firstly execute a select and then an insert. The whole database configuration is setted up by the JDBC Connection Configuration where we need simply to specify the driver to be used and the usual connection parameters, such as the database URL, username and password.

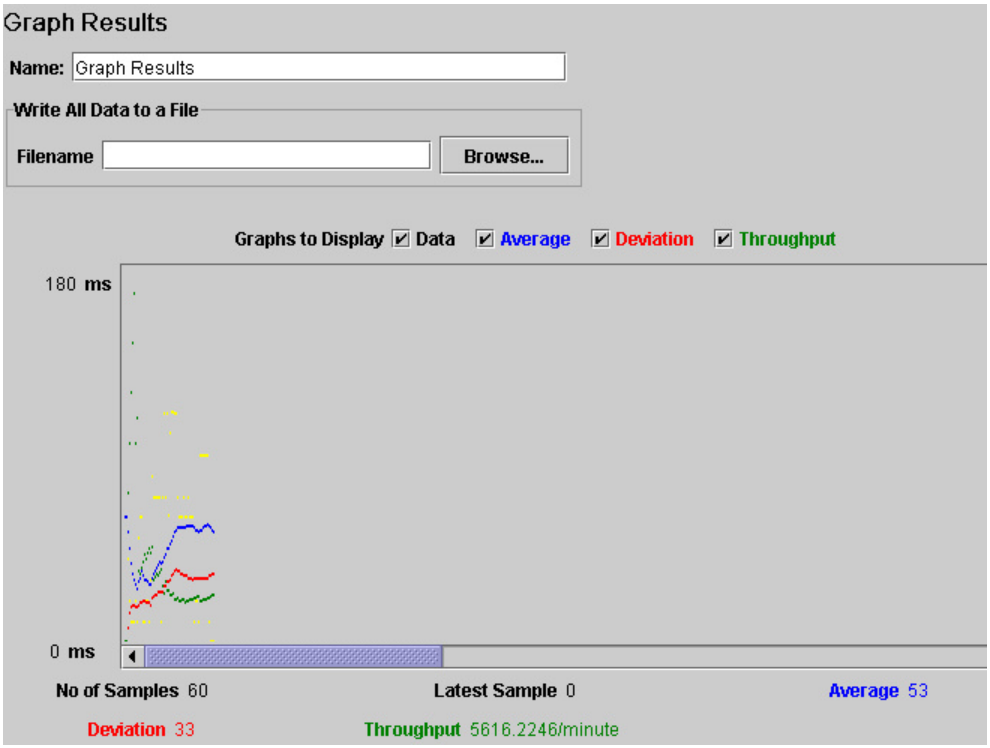

Figure 5.2: JMeter Graph

In figure 5.1 there is another element which we have not still explained:

the Graph Result. JMeter is also capable to represent the results in a graphical form. In figure 5.2 there is an example of a very simple graph obtained during a database load test (this image is taken from the Apache JMeter tutorial). This Graph Result is only a specific Listener which is possible to configure in the test plan, but there are many other listeners such as Table Result or Monitor Result and so on. It is also possible to execute JMeter from the command line interface, saving precious resources while running the test, logging all the results in a file, and then use a specific application to represent them in better ways.

Last but not least, JMeter is based on a plugin architecture and therefore it is highly extensible. Every element can be extended. For example a new Listener with a particular graph can be added, and also new Timers. This feature makes JMeter really flexible, and so another requirement is accomplished.

5.4.2 Disadvantages

To make this analysis complete, we have also to evaluate the disadvantages of JMeter. And, as first thing, it's evident how JMeter doens't allow load testing of in-memory databases or other kind of databases which don't have a JDBC interface. It only acts with databases through SQL, and it is quite obvious: at the moment there is no a standard for database native interface. Therefore, although JMeter works with relational databases, it will not work with all the in-memory databases without a proper extension, an implementation of a new plugin, because there is no plugin available for our needs.

Moreover JMeter doesn't compare different systems between them, but instead it measures the performance of one system at time. Therefore there is no direct comparison, while the comparison is our main purpose. In fact JMeter was originally designed to test web applications in order to find the best tuning for the web server and the relative application. This means that JMeter is not used to find the best web server with the same application, which in other words it can be said that it is not used to find the best database for our application scenario. Nonetheless it can be also used to compare different systems, but not with much ease of use.

Also the visual representation of the result, the Graph Result, is not that great, and of course JMeter is not famous for graph results. It is not very fluid, and it has some bugs, such as when the test time is too much the graph goes in overflow.

5.4.3 Conclusion

Now, after the analysis of advantages and disadvantages, we can conclude this investigation about Apache JMeter. It seems a very interesting application benchmark with a lot of features which meet most of our requirements such as:

- portability, because it is written in Java and it is a desktop application;
- flexibility, which comes from the plugin architecture;
- ease of use, tanks to the GUI;
- visual and detailed report.

Although all these advantages, there is still something which is not perfect. Firstly JMeter can't test both object and relational databases, and this is an important limitation considering our main purpose is to test in-memory databases. Nonetheless JMeter can test java objects, and therefore it can work with java native interface, and however it is possible to write a new plugin to add the functionalities we need. But this solution will also add a considerable amount of work in programming the new plugin, losing partially or completely the usefulness of an already existing application benchmark.

Secondly also the graph listeners already implemented in the application are not very useful. They can only be used for a fast and approximate analysis. In order to create a good graph with the resulting data stored in a file, there are two possibilities: implement a new graph listener or use another application to represent the result. In both cases there is again the necessity to write new code's lines.

In addition JMeter doesn't allow an easy comparison between different results, and therefore, with the purpose to compare several database performance, there is no more ease of use.

Finally, JMeter is a great application and the extension with new plugins and other elements is very attractive, especially considering this is an open source application widely used. But not only the amount of code to write may be equal or more than the code needed to write a specific application for our needs; in this way we are also forced to move in a more complex application, although JMeter is easily extensible. However, taking in mind JMeter is not the best application to compare different performance, this is not a proper way to work. Nevertheless the extension of JMeter is still very

interesting.

5.5 Poleposition

Poleposition is an open source database benchmark, developed to compare database engines and object-relational mapping technology. Poleposition is built to be a framework to help developers in evaluating databases performance, through a simple implementation with only few lines of code. In fact the impetus behind Poleposition came from the observation that developers evaluating candidate databases for future applications often resorted to constructing ad hoc benchmarks rather than using "canned" benchmark tests (or relying on vendor-provided data). This is entirely understandable: to properly evaluate a database for a specific project, you would want to exercise that database in ways that correspond to the application's use of it [Grehan 05]. This is the same concept we have already explained in paragraph 3.1.3 and paragraph 3.1.4.

Poleposition use the metaphor of a race, how it is clear by its name, to help the developers in understanding the framework structure.

5.5.1 Advantages

This is another really interesting database benchmark, full of advantages. Above all, again, it is a Java desktop application, without the need of any installation, except for eventual database servers. Therefore the portability specified as first of our requirements is granted, allowing the test of database systems on different platforms.

In addition also the flexibility is provided by Poleposition: in fact it is a framework, for database benchmarking, allowing the implementation of new tests and of new database systems. The framework is also simplified by the metaphor used to describe the whole application,and which is evident by the name itself. This application benchmark is configured like a championship car racing, where:

- A *circuit* is a set of timed test cases that work against the same data.
- A *lap* is a single (timed) test.
- A *team* is a specific database category or engine that requires specific source code.
- A *car* is a specialized implementation of a team, and therefore every database system can use several implementations.
- A *driver* is an implementation of a circuit for a team .

In favor of Poleposition there is also the reporting tool: results are available both in HTML format and in a PDF file, providing a fast and clear idea of the results obtained. The PDF file and the HTML result are a collection of circuits, which actually are:

- Melbourne: writes, reads and deletes unstructured flat objects of one kind in bulk mode.
- Sepang: writes, reads and then deletes an object tree.
- Bahrain: write, query, update and delete simple flat objects individually

- Imola: retrieves objects by native id.
- Barcelona: writes, reads, queries and deletes objects with a 5 level inheritance structure.

These circuits are a collection of laps: a write, a read, a delete etc. And each lap show the numeric results with both a table (table 5.1) and a graph (figure 5.3).

t [time in ms]	objects:3000	objects:10000	objects:30000	objects:100000
db4o/6.4.14.8058	379	7365	1398	4820
Hibernate/hsqldb	716	799	2617	14916
Hibernate/mysql	1442	2948	9437	35853
JDBC/MySQL	2881	1872	5470	18422
JDBC/Mckoi	1202	2842	9371	33530
JDBC/JavaDB	970	907	5454	8676
JDBC/HSQLDB	276	57	199	931
JDBC/SQLite	20381	46545	138990	483594

Table 5.1: Poleposition time table: melborune circuit and write lap

The table 5.1 shows the time the "write" lap takes during the melbourne circuit. For each car/team is shown the time in milliseconds they take to execute a certain number of writes: 3000, 10000, 30000 and 100000. Therefore it's easy to compare them and understand who is the fastest, the winner of the race. The same result is also represented by a graph, figure 5.3, which clearly shows which database is above the others. To note how on the y axis the time is in a logarithmic scale and therefore higher is the value and faster is the database system. On the HTML result there is also another kind of chart, which measures the memory consumption, but on the PDF file this is not present.

Poleposition can also be used to test both object and relational database systems. This is possible because it was developed with the aim to com-

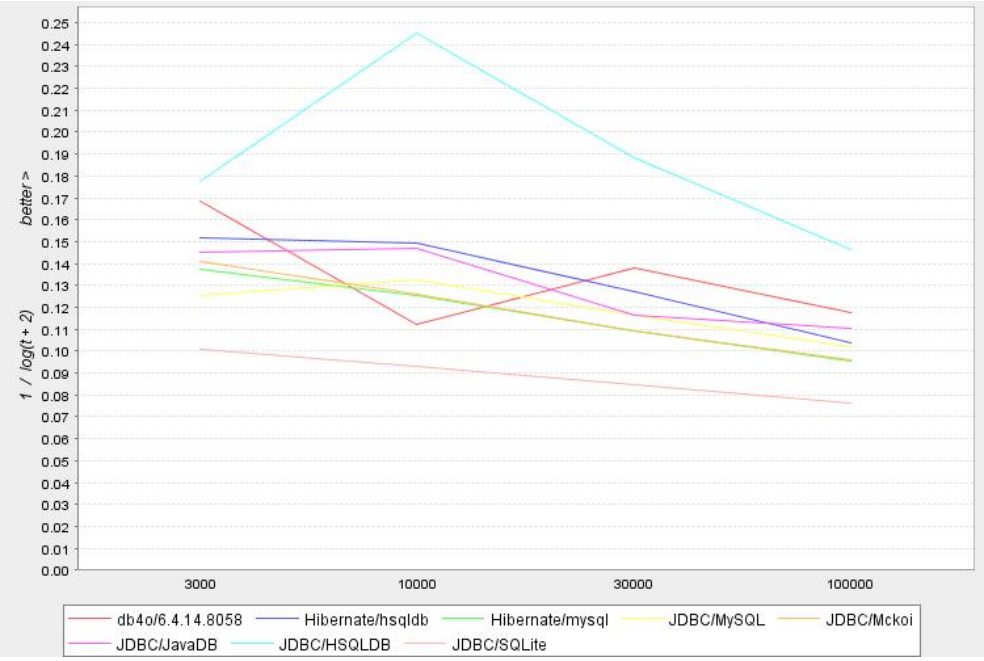

Figure 5.3: Poleposition time graph: melborune circuit and write lap

pare database engines and object-relational mapping technology. Hence this makes poleposition a really interesting database benchmark for our needs.

Finally, as last advantage, there is something that is not properly a feature of the application, instead it is an idea the programmers took in mind during the development of the application, and therefore it reflected on the quality of the application. The idea we are talking about is the following: *"Objectivity must be the most important goal of the benchmark tests. The results show very clearly that there can not be one single best product. Depending on your application, some circuit results may be very important for you and some may be completely irrelevant."* This statement perfectly accords to the axiom described in paragraph 3.1.4, and on which is based our work.

5.5.2 Disadvantages

The metaphor may help to understand the initial problem of the benchmark, but on the other hand, it becomes a trap, a limit for the benchmark itself: if they continue with the car race they will soon run out of gas [Grehan 05]. Any further extension will be forced to follow the rail imposed by the metaphor. Moreover this metaphor makes the benchmark a simple race, and there is no load testing.

Although poleposition's output is very catchy, it is not easily readable and it lacks of detail: time and memory are the only measures taken in the HTML result and in the PDF file only the time is represented. In addition every test and graph/table lack of an explanation of what it is testing and representing, making the comprehension both of the test and the result very opaque.

Moreover the results are always too much summarizing. For example in figure 5.3, and thus also in table 5.1, only the total time of the test is represented, and so we can calculate the throughput dividing the number of objects inserted with the total time. But it's not possible to know if this throughput was constant or growing or oscillating too, there is no representation of the throughput's trend during the execution of the test. This contributes to make the test even less clear, hiding the real database systems' behavior. For example, looking at figure 5.3, we can't understand why HSQLDB and Db4o are so much oscillating. Of course it can be the real database performance, but it could also be a mistake in the test implementation or in the database configuration. Certainly a more detailed report of the result could be helpful.

Furthermore this framework only allow single-user testing, because, as the authors said in [Grehan 05], *"some sort of multi-threaded test to simulate multiple user would be exceedingly difficult without extensive modifications to the framework. But if that could be done, PolePosition could provide enterprise-level testing"*. Nevertheless they hope in the future to introduce this new functionality, but at the moment it is still not working.

Finally this database benchmark is quite recent: it started in 2005 and actually there is still only one release available for download, posted in June 2005. Looking through the SVN repository on sourceforge, it's possible to notice some new changes and activity in the last years, but there is still no new release available for download. In addition the code of poleposition is not that clear and object oriented as they said in [Grehan 05].

5.5.3 Conclusion

As first impact Poleposition seems to be exactly what we are looking for:

- the portability is granted by the Java language;
- new test can be implemented allowing flexibility;
- there are very catchy graph report;
- and it can works both with relational and object databases.

But a deep analysis of this feature showed that it is not suited for our needs. Particularly the results are very poor, and the graphs don't show any database behavior during the execution of the test: in fact these graphs

shouldn't be line charts, but it would be better if they were bar charts, because they don't show any trend.

But a very important disadvantage, which exclude Poleposition from our choice, is the lack of load testing, also due to the incapability to execute concurrent tests. This is a major disadvantages, especially when we want to simulate real application scenarios, which is a consequence of the axiom on which is based our work.

In addition this benchmark is not easy to use, although it may seem easy. It's impossible to understand the meaning of the results without looking at the code: *"if you consider basing your choice of database on this benchmark, bring along lots of time to look at what every test does. You will not be able to judge the results without looking at the source code that produces them."* And therefore this benchmark is not for non technical people who don't know Java language, but also for engineers is not so easy, because they need a lot of time to understand properly the benchmark results, looking through the code's lines.

It could be possible to extend this framework with the features we need and the authors themselves exhort this possibility. On top of that there is a substantial amount of code to write and the framework code is not simple and easily extensible: for example there are methods with hundreds of line of code and with many many nested conditional instructions, which is not object oriented programming, losing all the maintainability such a programming paradigm offers.

However this application is very catchy and it is a great source of inspiration for a future development. Very interesting features, although they are

not perfect, are the report in a PDF file, the graphs, and the idea to make a database benchmark framework.

5.6 Benchmark Overview Summary

Coming to a conclusion we sum up the pros and cons of the database benchmark applications analyzed in the table 5.2. It shows only the most interesting negative and positive features we found in these benchmarks, in order to understand what we need and take inspiration for a future development of a new benchmark application.

Benchmark	Pro	Con
The open source database benchmark	Open source AS3AP	SQL only C language No comparison No visual reports Latest release: 0.21 (2006)
TPC-C benchmark	Standard Simulate OLTP systems Price/performance metric Acidity testing For every DBMS	For high end vendor No implementation available
Apache JMeter	Java desktop application Load testing Plug-in architecture JDBC plug-in	No plug-in for IMDB No comparison Bad graphs
Poleposition	Java desktop application Database benchmark framework Both relational and object DB Results comparison Catchy report	The metaphor is a trap Lack of detail in results Only single-user testing Quite recent

Table 5.2: Benchmark summarizing table

The starting idea was to find an application with the requirements explained in paragraph 5.1, so that we could be able to run our test suite. But the final result is that there is no benchmark which perfectly fits our needs.

This means there are only two possibilities to execute our test suite:

- choose one of the benchmarks previously analyzed and then extend it, in order to add the functionalities we are looking for;
- develop a new benchmark starting from our requirements.

The first idea was already taken into consideration during the benchmark analysis and rejected in the conclusion paragraphs. In summary:

- The first benchmark, the open source database benchmark, was completely unsuitable. Not even one requirement was satisfied. The only positive aspect was the use of AS3AP, the ansi SQL standard scalable and portable benchmark (portable only for relational database server), which can be helpful to the creation of new test case.
- The transaction processing performance council, with their TPC-C, is very interesting because it is a standard and it can simulate a generic OLTP system. But there is no implementation available, and anyhow this will not help us to execute are test suite. Instead this can be a good source for inspiration, both for a new load test case and for some ACID test case. So it is a nonsense talking about extension.
- Apache JMeter is a very good general purpose benchmark application and offers a plug-in for testing relational database server through JDBC. Therefore to use this tool for our aim we need to write a new plug-in, which may not worth the effort, especially considering that JMeter allows an efficient measurement of the performance, but not

an easy comparison between different system, and moreover the graphs are quite poor, although extensible.

- Poleposition, on the other hand, is suited for both relational and object database systems, but its testing capability are quite poor. The lack of concurrent testing, and therefore of load testing, prevent us to execute our test suite. A possible extension of this application is not the best solution, especially looking at the source code which is not well organized.

Therefore the best solution is to develop a new benchmark application. Although this decision, the analysis done is not useless, in fact we understood which features can be used from other benchmarks, and how to develop them, and what should be avoided. But this will be explored in the next chapter.

Chapter 6

The New Database Benchmark Application

The previous chapter is dedicated to the research of a suitable benchmark application for our needs which is able to execute our test suite. The analysis didn't bring to any available benchmark neither to a benchmark easily extensible in order to implement the features we want. Therefore the conclusion was to develop a new benchmark application, starting from our requirements and treasuring the knowledge acquired during the analysis.

In this chapter we describe the new benchmark application developed for our needs, starting form the specification defined from the requirements and then describing the application from different points of view. Subsequently we will discuss also about the application usage, from the application's configuration to the extension and the implementation of new databases and tests.

6.1 Analysis: Requirements Specification

This section is dedicated to the analysis of the benchmark application developed during this thesis. We start from the specifications' definition, which come from the requirements described in paragraph 5.1 and from the functional requirements resulting from the capability to execute the test suite described in chapter 4.

In order to define the specifications from the requirements on which the benchmark application is based, it's primarily important to understand what is a specification in relation to requirements. A requirement is a singular documented need of what a particular product or service should be or do. On the other hand a specification is an explicit set of requirements, or in other words it is a well-formed requirement, a formal requirement, an implementable requirement.

Now it is clear how the discussion done in chapter 5 is not useless. Through that analysis we obtained a better understanding of the requirements, and a suggestion to implement them. This will help us in the process of requirements specification.

The table 6.1 summarizes the specifications we are going to analyze in the following paragraphs.

6.1.1 Portability

We already recognized the value of portability as a requirement when testing embedded databases, because it allows us also to chose the best platform for the database we are testing and to study the difference in the performance

Requirements	**Specifications**
Portability	Java
Understandability	clear PDF file with graphs
Flexibility	extensible framework
Detailed Report	several measures
Visual Report	graphs with trend
Both Relational and Object Databases	databases interface written in Java
Ease of Use	XML both for test and report configuration

Table 6.1: Benchmark specifications

with a platform's change. From the experience gained with the benchmarks' analysis we know that both JMeter and Poleposition grant the portability we need. And this is granted thanks to the Java language. In fact they are both Java desktop application who are able to be executed on several platform because the Java Virtual Machine allows Java to run everywhere (*"write once, run everywhere"*).

Therefore this requirement brings to a simple specification: when you want to grant portability to an application, the Java language is a good solution.

6.1.2 Understandability

When a benchmark is not easily understandable, it is useless. Benchmarks are too easily tricked and cheated, and thus they are meant for different stakeholders. Looking at JMeter and Poleposition we understand how a graph can help to explain the numeric results and to compare several systems. Particularly Poleposition offers a very catchy PDF result, containing all the

tests with all the results expressed both in tabular and graphic results. But in this PDF what the test does is not very clear. Then the need of clear results which explain also the test's operations. For this purpose JMeter is a bit better, because with a tree structure it shows exactly the task and all the operation involved by the test.

Therefore what we need to grant this requirement is a PDF result, containing all the information about the tests:

- what the test does, in terms of operations involved on the database;
- graph result which catch the attention of the reader and quickly express the meaning of the numeric result.

This is the specification implicated by understandability.

6.1.3 Flexibility

Benchmark always suffers the limitation introduced by the workload simulated. The performances measured by the benchmark are highly dependent on the workload, and therefore for an application, whose workload is not similar to the benchmark's one, the results are nearly useless. To overcome this limitation, flexibility is a key feature. In this contest flexibility is meant as the capability of the benchmark application to run complete new tests, and so to simulate new workloads. Furthermore it is also intended as the capability to run the benchmark against new database systems. Also this time JMeter and Poleposition show a possible solution to this problem, but they are not equals.

As regards Poleposition, it works like a framework, in fact the main idea behind this application is to make a framework for databases benchmarking. Therefore Poleposition allow an easy implementation for both new tests and database systems, although it is not so easy and flexible as a framework should be, especially for new tests. In both cases the solution involves the programming of new code's lines.

JMeter, instead, is able to create and execute new tests thanks to a graphic editor which can be used to create very complex scenarios. In regard to the implementation of new database systems, at the moment JMeter can test only SQL databases, but a possible solution may be the development of a new plug-in. In fact JMeter is based on a plug-in architecture, which allow a good evolution of the software.

In conclusion the resultant specification is: create a benchmark application which works as a benchmark, with tests that can be implemented by programming, and with pluggable databases.

6.1.4 Detailed Report

A well detailed report with a lot of interesting measures will tell you much about the database behavior, while only one single metric will simply make your trust go away when you read it.

The first who stated this idea was the transaction processing performance council, who, in fact, adopt two different measures: the first is the throughput and the second is a price/throughput measure. This because the TPC-C benchmark doen't compare directly different systems, but every vendor run

it on his own, therefore without a price/performance measure, it is very hard to understand if the result depends from more expensive computers. This is not our case, but we agree that one measure doesn't explain the real database performance.

JMeter, instead, has several listeners which are able to take different measures, and they can be composed in many ways inside a test plan. So JMeter allows the user to take exactly the measures he wants. And this is applicable also for graph listeners.

In this case case JMeter shows a very good solution, and this is exactly what we like: the possibility for the user to chose the measures to display as final result.

6.1.5 Visual Report

When you face complex results with a lot of numbers, a visual representation will make them clearer. Thus when using graphs also a comparison between different systems will be easier, because it becomes a comparison between different colored lines. Furthermore graphs are very helpful if you want to analyze not only the final result of a test, but the whole trend.

Although Poleposition makes use of graphs, these don't represent any test's trend, they simply represent the final value of the test. Nevertheless Poleposition with its graphs allow a very easy comparison between different database systems and the graphs are very catchy.

On the other hand JMeter provides very stark graphs, without any comparison between different systems. But it is possible to represent each mea-

sure on the graphs and in addition they show the performance's trend during the test's execution. Using appropriate listeners, or writing new ones, it is possible to represent almost everything.

To sum up we want graphs who let us compare different database systems and draw different measures, with the capability to analyze the trend during the test's execution, but in a catchy way. Something like Poleposition graphs with in addition the trend representation. Furthermore the user should chose which graphs create, specifying the measure to drawn on the graph's axis.

6.1.6 Both for Relational and Object Databases

Relational and object database system usually doesn't share the same interface. From the Java point of view, relational databases are usually accessed through JDBC, while object databases usually use native interfaces. It's not easy to make a database benchmark work with two different technologies, there is an impedance mismatch.

Poleposition is able to test both relational and object databases, but at the cost of a completely new implementation of the whole test for every database which is to be tested. JMeter instead works only with relational databases, but it provides also the possibility to test Java objects, which is very useful thinking at Java native interfaces to access the databases. Nevertheless this means that an object oriented point of view will be used for object databases and at the same time a relational point of view will be used for relational databases, therefore the final result will have less meaning. With the purpose of benchmarking and comparing it's better to use a uniform

point of view.

Since the benchmark application we want to develop must work with both relational and object database systems and since most of our database are object databases, we will use an object oriented point of view when benchmarking databases and therefore every database must implement a Java native interface for each test. Every database must implement a own interface, a driver, because there isn't yet any standard for object oriented querying.

6.1.7 Ease of use

Easy of use is intended as a variation in the tests'parameters, such as the initial database state and others already explained in chapter 4, allowing little variations to the workload simulated, in order to improve the test when the needs change.

This ease of use should be granted without any modification in the application's lines of code. For this purpose Poleposition uses several properties files while with JMeter the user can modify the whole test, without any programming, and therefore he can also modify every parameters in the test without touching any code's lines.

Our idea consists in an xml file which contains all the tests' configuration. The xml file should be organized with a tree structure such as the JMeter's representation of the test plan. This xml must contain all these variable parameters. Furthermore we would like to configure also the final report of the benchmark, specifying which graphs draw and which measures represents

for each test.

6.1.8 Test Suite's Specification

More specifications come from the capability for the benchmark application to execute the test suite described in chapter 4. These are functional specifications because they describe the requested application's behavior. In this particular case, therefore, they come from the test suite and we are going to analyze them.

In order to run the test suite we need the benchmark to be able to:

- execute test concurrently and therefore allocate a thread for each task composing a test;
- run each task of the test at most for a specific amount of time, which may be different for every task;
- iterate each task for a certain maximum number of iteration, different for each task;
- run each task without consuming all the resources and therefore limiting the transactions per second, the throughput of the task.

All these informations must be taken by the benchmark application as input without any programming, and therefore they must be specified in the XML file.

6.2 Design And Development

Software design is a process of problem-solving and planning for a software solution. After the purpose and specifications of software are determined, software developers will design or employ designers to develop a plan for a solution. It includes low-level component and algorithm implementation issues as well as the architectural view.

In this section we will design and develop the application using different point of view, or it's better to say we will show how the system has been realized by describing it through different point of view. Although a common temptation is to try to analyze and describe a system with a single, heavily overloaded model *"it's not possible to capture the functional features and quality properties of a complex system in a single comprehensible model that is understandable by and of value to all stakeholders"* [Nick Rozanski 06].

Therefore different views will be used in the following paragraphs to describe separately each the major aspect of the benchmark application. The views used are the functional view, the information view, the concurrency view and the development view. Collectively they describe the whole system.

6.2.1 Functional View

A functional view defines the system's functionalities, demonstrating how the system performs the functions required. This is probably the easiest view for stakeholders to understand the system and it is the starting point for the definition of the other architectural views. The figure 6.1 represents the benchmark application from a functional point of view which takes into

account the specifications summarized in table 6.1. It represents the system by an external point of view, showing functional capabilities, external interfaces, internal structure and design philosophy, such as the low coupling between the database and test implementations.

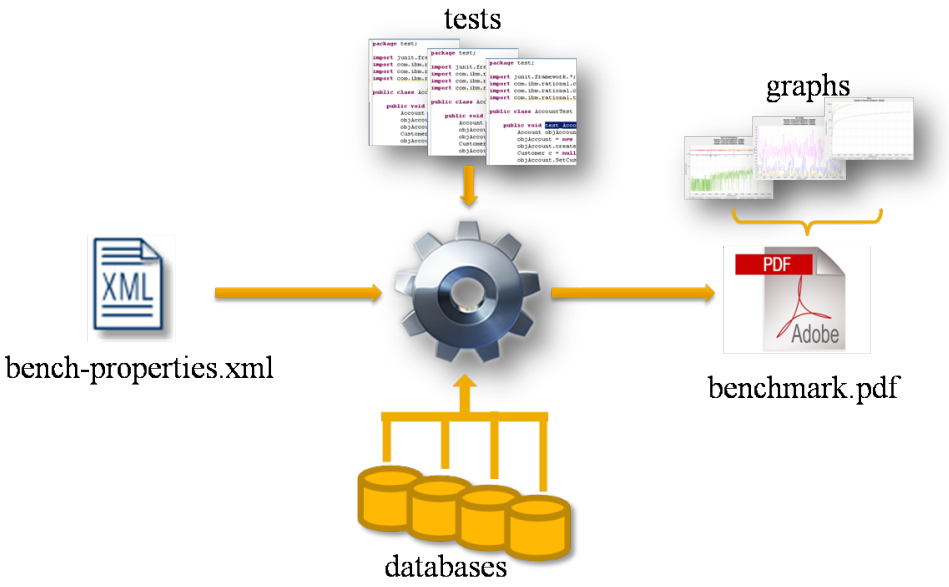

Figure 6.1: Functional view: the system from an external point of view

The benchmark application, represented in the figure 6.1, is composed with the following parts:

- The application's input is an XML file, the bench-properties.xml. It contains a substantial part of the tests' configuration, that is a collection of variable parameters for each test, such as the task, the operations' order, the time, the maximum throughput, the initial database's state etc.

- The group of databases shows how they can be plugged in the bench-

mark, in order to execute benchmark's tests on several and new databases.

- New tests can be implemented in the benchmark application allowing the execution of specific workloads.
- The output is a PDF file containing all the results obtained by the performance analysis.
- The results in the output file are represented mostly by graphs.
- Finally a gear represents the application as a black box. The figure just shows that the application works as a framework with extensible databases and tests.

In other words, the benchmark application takes as input an XML file, and through pluggable databases and tests, it produces a PDF file containing all the result, represented also with graphs.

6.2.2 Information View

The information view focus on a small number of entities that your stakeholders view as significant or meaningful, primarily considering the user stakeholders, but also taking into account the concerns of other stakeholders types such as maintainers. The idea behind the model is to keep it simple. Although our application doesn't store any information, it manipulates, manages and creates information.

The figure 6.2 can be considered a kind of information view because it describes the entities involved by the benchmark application. Anyway this model is very useful because it provides a description of the entities, helping

the comprehension of how the benchmark works. This view is not only for developers who can eventually implement new features and extend the current application, but it is dedicated to every stakeholders. In fact it explain very well how a test works and how it is composed, and this comprehension is essential for a right use of the XML file, but is is even essential for understanding the benchmark's results.

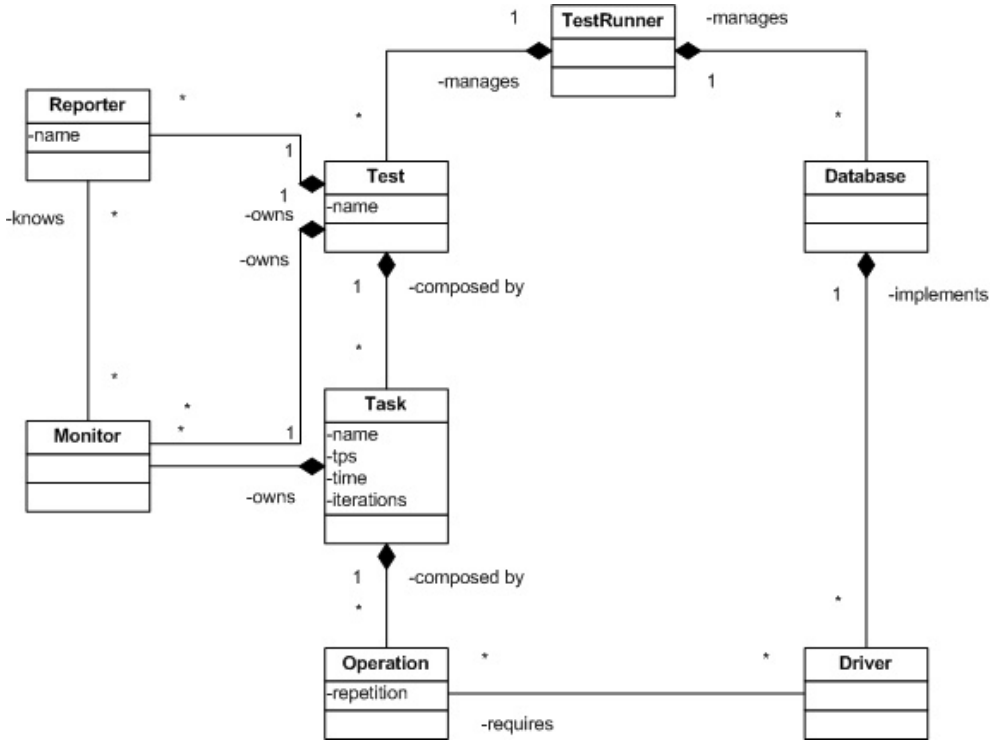

Figure 6.2: Information view: class diagram

Analyzing more in detail the figure 6.2, it is so composed:

Test : this is the most important entity in the benchmark application. It represent the test which can be started on a specific database. Each test has a name which is configurable by the XML and helps the readability

of the output. Furthermore each test is composed by tasks and has a collection of monitors, which measure the test's performances, and reporters.

Task : this is substantially a set of operations which are executed synchronized. Each task has also a collection of monitors which measure the task's performances. As for the test, also the task got a name to help the readability of the PDF file, but, very interesting, are some special attributes which are used to define the test scenario to simulate. These attributes, setted in the XML file, are the transactions per second (tps), the time and the number of iterations. They have been already explained as specifications in paragraph 6.1.8.

Operation : this is the class to write to extend the database benchmark framework. Every operation represent one or more method invocation to the database.

Monitor : this entity is dedicated to the performances' measurement. It monitors some specific measure of the test/task to which it is linked, such as the throughput or the cpu usage.

Reporter : this is a monitor of monitors. In fact they register and put togheter several measures of different monitors. This entity is dedicated to the production of graphs in the output file, and therefore it also has a name, helping the readability. A reporter can be declared directly in the XML file and can represent exactly what the user wants.

Database : this entity represents a specific database, such as Prevayler,

Pico4 or HSQLDB. It is a collection of drivers.

Driver : this is a driver required for the execution of a specific operation, which may require also more than one driver. In other words this can be seen as the implementation of a specific database for an operation.

TestRunner : this entity runs the benchmark, executing every tests on every databases. In fact it manages a collection of test and another collection of databases.

6.2.3 Concurrency View

The concurrency view describes the concurrency structure of the system, mapping functional elements to concurrency units to clearly identify the parts of the system that can execute concurrently, and shows how this is coordinated and controlled.

The concurrency of our benchmark application is very simple, and it is based only on threads, but an explanation of it will help in the comprehension of how tests work and, therefore from the definition of new tests and databases to a "consapevole" results' analysis.

The figure 6.3 shows how the concurrency is organized, mapping the active entities, already seen in figure 6.2 in the previous paragraph, to threads. The benchmark runs in only one process of the operating system, in the Java virtual machine, except the processes of the in-memory databases. All the other concurrent elements are divided into different threads:

- The is only one test running per time, and after it starts the reporters, the monitors and the tasks it waits the end of every tasks.

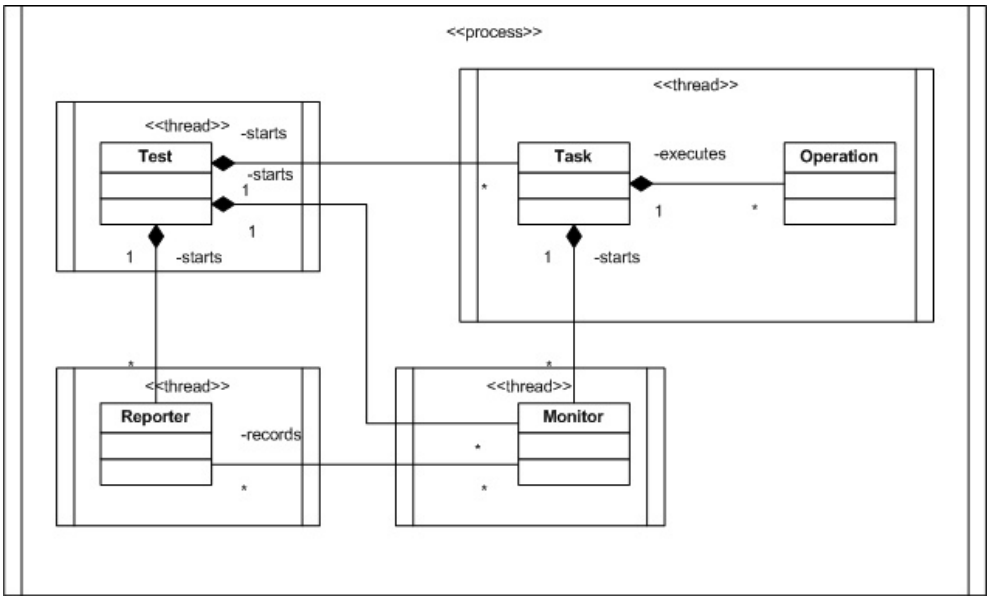

Figure 6.3: Concurrency view: threads

- Each monitor is associated with a thread so that they can measure the performance of the test and of the different task concurrently.

- Each reporter runs in a own thread, and, after it is started by the test, it records the values measured by the monitors.

- Each task is executed concurrently with the other tasks and so there is a dedicated thread for each task. After it starts the monitors associated, it executes in a repetitive and synchronized way all the operations which compose the task itself. Therefore all the operations belonging to a specific task are executed in the same thread.

6.2.4 Development View

A considerable amount of planning and design of the development environment is often required to support the design and build of software for complex systems. Things to think about include code structure and dependencies, build and configuration management of deliverables, system-wide design constraints, and system-wide standards to ensure technical integrity. It is the role of the development view to address these aspects of the system development process. The concerns of the development view we will analyze are the modules organization, the codeline organization and some of the technologies used by the application. Furthermore we will also analyze an UML sequence diagram for a test's execution.

Module Organization

As regards the module organization, the figure 6.4, which is simply a screenshot of the project inside the Eclipse IDE, describes the package organization. There are three main packages in which the whole application is divided:

- The test package contains the tests with all their operations which is possible to execute and configure in the XML file.
- The database package contains, for each database system, the implementations for the tests: the drivers which are required by the operations, enabling them to operate, hiding the specific database implementation. In fact, these drivers are adapters for the database systems' diversity.

- The core package is the hearth of the benchmark application and contains all the framework: from the monitors to the reporters and to everything is needed to execute the tests on the databases.

In fact our application is built as a framework for database benchmarking, ensuring the extension with new tests and databases implementation. Therefore both tests and database have a dedicated package, in which they can be written and plugged-in without the necessity to know the whole application. Furthermore these three packages have few dependencies which in addition are managed only through indirections, such as interfaces and adapters.

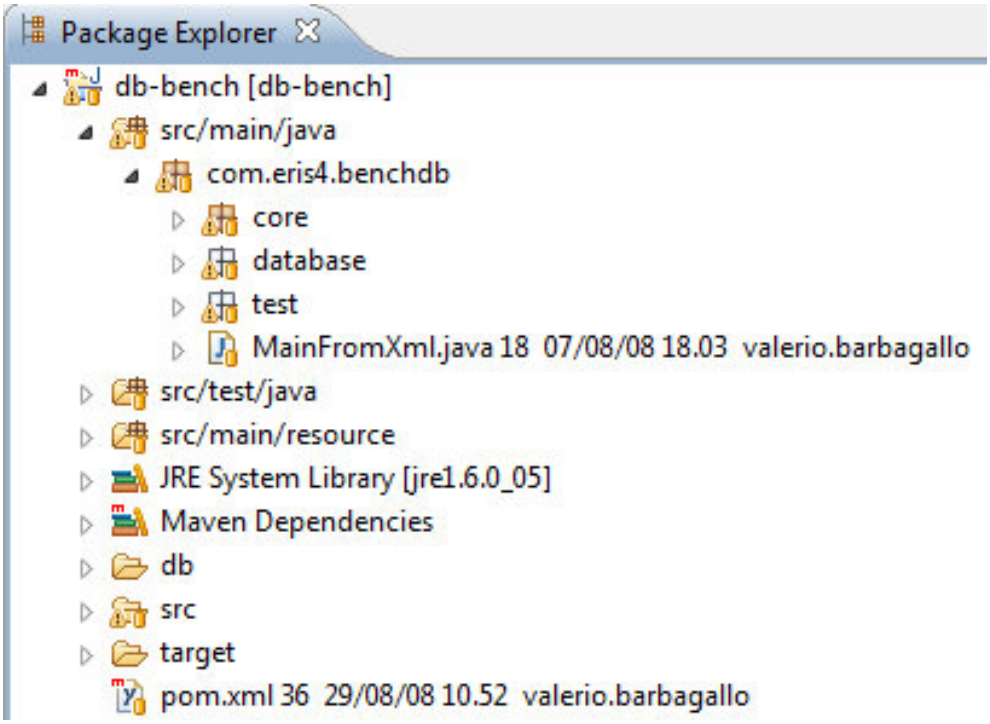

Figure 6.4: Development view: Eclipse's package explorer

Codeline Organization

As regards the codeline organization, the system's source code needs to be stored in a directory structure, managed via some form of configuration management system, built and tested regularly, and released as tested binaries for further testing and use. The way that all of this is achieved is normally called the codeline organization. The figure 6.2 not only shows the division in package of the source code, but it shows also the whole project structure, the codeline organization. Eclipse's users probably have already recognized Maven, a software project management tool which is integrated in Eclipse with thanks to a plug-in. Maven simplifies the build project, provides a uniform build system on different platforms and has a superior dependency management, and it is managed with the pom.xml file, where all the settings are placed, such as the dependencies.

Eclipse's users will also notice the small database symbol which means that the project is hosted on a repository. In particular the application is on a SVN repository: the exact URL for the repository, which is available for everyone in internet, is at `http://portable-db-bench.googlecode.com/svn/↘trunk/`, which is a free SVN repository offered by Google. In fact our project is open source and any contribution is welcome.

Other Technologies

As already explained, Maven manages the depending libraries in a very simple way. The question is now which dependencies we are using. Except all the dependencies involved by the database systems, there are three major

technologies used by our application: log4j, jfreechart and itext.

Log4j is a very common library used by most programmers and it enables a uniform logging system for the whole application. This is low-tech method for debugging which is very useful with concurrent applications. The major advantage of log4j is the possibility to enable logginat runtime, without modifying the application binary using, moreover, different log levels.

Instead, jfreechart is used to create the graphs in the resulting PDF file. It is a free 100% Java library which allows the creation of professional quality charts. A nice feature is the support for many output types, but there is also a great variety of available charts such as bar charts and pie charts, but the chart used by the benchmark application is an `XYLineChart`: a 2D chart with x and y axis. It is suited to analyze the trend in the performance during the benchmark's execution.

Finally, also iText is noteworthy. It is a a free Java library that allows the creation of PDF files on the fly. It is build in the benchmark application, producing the PDF file containing all the results, generating a read-only, platform independent documents, containing also all the graphs/images produced by jfreechart.

Test's Execution

In conclusion for the development view, we illustrate the sequence diagram for a test's execution, with the purpose to increase the comprehension of the life's cycle of a test. This will simplify both the use of the XML file taken as input and the implementation of new tests and databases.

The figure 6.5 is a UML sequence diagram and it shows a simplified

test's execution, because it's not displayed the interaction with monitors and reporters. The focus is on the test, the task wich compose the test and the operations which compose the task, and their life's cycle.

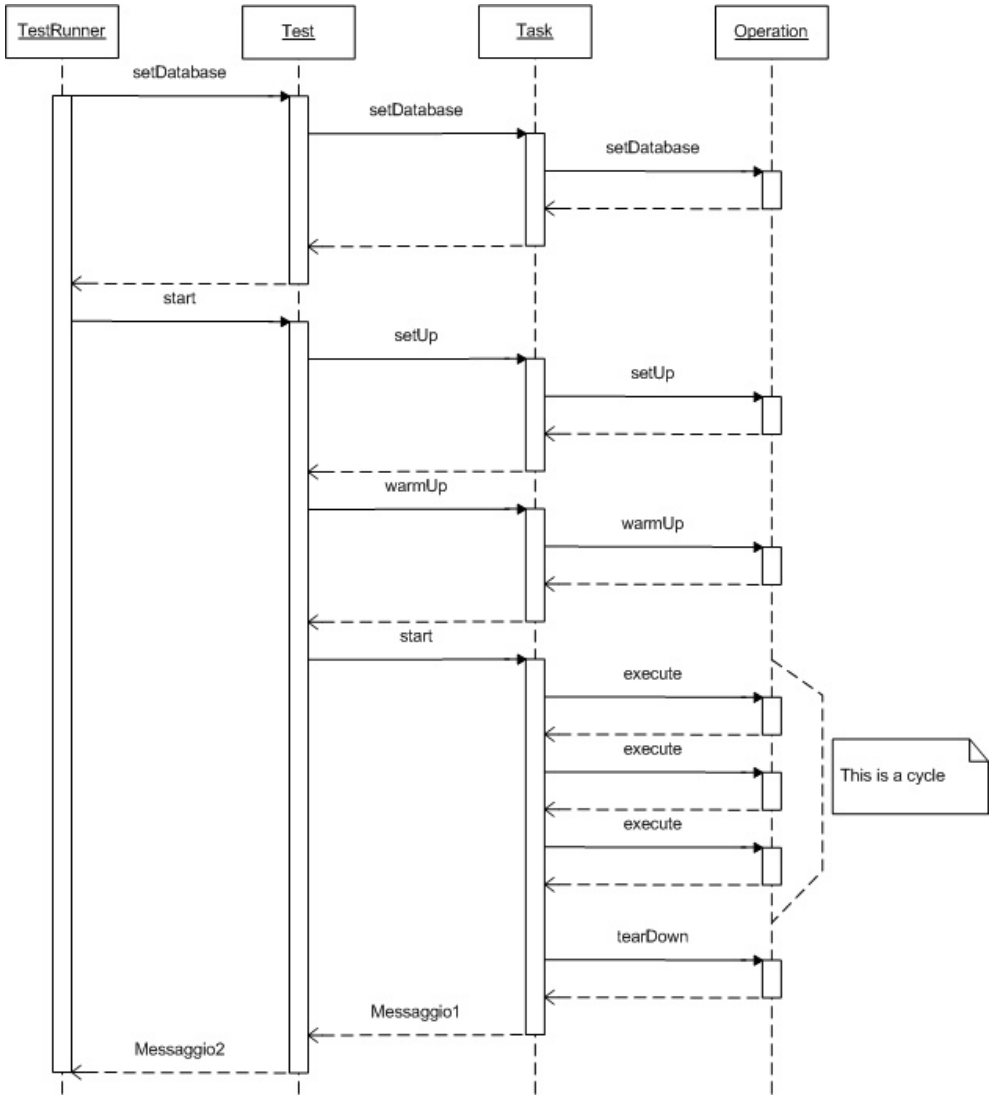

Figure 6.5: Sequence diagram for a test's execution

The sequence diagram how a test is executed and we can divide the execution in five phase:

1. The `TestRunner` is the main class who execute all the tests on all the databases, but before starting a test, a preliminary operation is done: the database is setted in the test so that the `TestRunner` executes the test on a specific database. This operation spread to all components of a test, through the tasks to the operations, so that each operation can take the suitable driver from the specific database.

2. In a second stage the test is started by the `TestRunner`. Then the test sets up all its tasks and consequently all the operations. In other words this phase corresponds to an opening of the connection with the database.

3. Subsequently the test warms up all its tasks and operations to avoid the slow start of the real test, when all performance measures are taken.

4. In this stage the test starts a thread for each task, which then executes repeatedly, and concurrently with the other tasks, the operations.

5. In a final stage, when the stop condition is reached, each task tears down the operations, which in other terms means a close, a disconnect from the database. The driver is detached.

6.3 Application Usage

While in the first section we analyzed the specifications on which the application is built and in the second the focus moved to the design, in this section we assume the application already developed and we analyze the application usage. The main use is of course the execution of the benchmark and

therefore the configuration of the XML. But recalling that the application is structured as a framework, other kinds of use consist of the implementation of new tests and of new databases.

6.3.1 Configure the XML

The benchmark application developed, as clearly showed in the functional view, takes in input an xml file: the bench-properties.xml. It contains the whole configuration for the benchmark's execution and it grants the ease of use so much researched. In fact it is possible to specify and change a lot of parameters such as the several stop conditions for the task, the test's composition, the reporters and more.

Training XML

To take confidence with the XML, we start analyzing a training example, which can't be executed, but explains what every attribute or element is used for. This training XML is showed in the listing 6.1. It is divided into three parts.

The first part is full of definitions which are a mappings between a name and a class, so that we will not need to repeat all the path every time, but we can use a "key", which is a simple name. The definitions are divided into four parts: initializators, operations, test monitors and task monitors. We already know the operations and monitors, although they are classified differently for tests and tasks, but we don't know what is an initializator. It is a special class which is used to fill the database, in order to simulate

scenarios where there is no empty database. For example, we can assume that our operations interact with the database writing and reading a certain object, a `Person`. If we want to simulate a scenario where the database already contains one million of Person, we need a `PersonInitializator`, a class which may be called by the Test before starting each task. Another clarification to mention is that if you declare a monitor as a task monitor, every task will use that kind of monitor, and on the other hand, if you declare it as a test monitor, every test will use it. Although you will probably never change the monitors' definition.

```
<bench>

  <!-- DEFINITIONS -->
  <definitions>
    <!-- initializators -->
    <initializator name="PersonInitializator" class="com.↘
       eris4.benchdb.test.person.initializator.↘
       PersonInitializator"/>
    ...
    <!-- operations -->
    <operation name="WritePersonOperation" class="com.eris4.↘
       benchdb.test.person.operation.WritePersonOperation"/>
    ...
    <!-- test monitors -->
    <testmonitor name="Memory" class="com.eris4.benchdb.core↘
       .monitor.MemoryMonitor"/>
    ...
    <!-- task monitors -->
    <taskmonitor name="Time" class="com.eris4.benchdb.core.↘
       monitor.TimeMonitor"/>
    ...
  </definitions>

  <!-- DATABASE LIST -->
  <databases>
    <database class="com.eris4.benchdb.database.pico4.↘
       Pico4Database"/>
    ...
  </databases>

  <!-- TEST LIST -->
  <tests>
    <test name="Test 1" >
```

```
      <initializator name="Initializator 1" numberOfObjects=↘
         "1000"/>
      <!-- More initializators here -->
      <task name="Task 1"
      iterations="The total number of transaction after the ↘
         task stops"
      time="The maximum time after the task stops"
      tps="The maximum transaction per second">
        <operation name="Operation 1" repetition="1"/>
        <!-- More operations here -->
      </task>
      <!-- More tasks here -->
      <loggraphreporter name="Reporter 1">
        <line x="Time" xTask="Task 1" y="AvgTransaction" ↘
           yTask="Task 1" name="AVG"/>
      </loggraphreporter>
      <lineargraphreporter name="Reporter 2">
        <line x="Time" y="Memory" name="Memory"/>
      </lineargraphreporter>
      <!-- More reporters here -->
    </test>
  <!-- More tests here -->
  </tests>
</bench>
```

Listing 6.1: Training bench-properties.xml

In the second part consists of a list of databases which will be tested. This is not a mapping like for the definitions, but it is simply a declaration of database, specifying the path of the class which extends the base class `Database` of the framework.

The third part is the most important. In fact while the first and the second part are rarely modified, the third is used for the test configuration, containing all those variable parameters which often may be changed. In this part, every test element is a test which is executed on all the databases listed. Each test is composed by different elements:

Initializator : this element describes which initializators the test must execute to fill the database. There are two interesting attribute: the name is a pointer to one initializator defined in the first part with the same

name, while the second attribute specifies the number of objects to insert in the database.

Task : this is a collection of operations which are executed in a serialized way, and concurrently with the other tasks. While the attribute name is used only to create more clear results, the others are the key parameters of the test:

- The tps attribute is used to limit the throughput of each task. This is an optional attribute.
- The iterations attribute stands for the total amount of task's execution before the task is stopped. This is optional too.
- The time attribute is the maximum amount of time before the task is stopped. This is a required attribute and it is used also to avoid that a very slow database system can snooker the benchmark when only a the iteration attribute is used, therefore a sort of timeout.

Operation : this is an operation executed by the test inside a specific task. The name attribute is a pointer to an operation previously defined in the first part of the XML. The repetition attribute tells the number of time that an operation must be repeated for every task's execution.

Loggraphreporter : this reporter uses a linear scale for the x axis and a logarithmic scale for the y axis.

Lineargraphreporter : this reporter uses a linear scale both for the x and the y axes.

Line : this is used inside a reporter to specify what a reporter should draw. This element has several attributes:

- The name of the line.
- The x value to represent: this is a pointer to a monitor defined in the first part.
- The xTask which is a name used as a pointer to a task's name already declared in the current test, specifying to which task's monitor the x value is referred.
- The y value to represent: this is a pointer to a monitor defined in the first part.
- The yTask which is a name used as a pointer to a task's name already declared in the current test, specifying to which task's monitor the y value is referred.

Real Time Prepaid System XML

After a non executable example of the XML file, we now analyze a more complex and real XML file. It describes the main test case developed in our test suite which is described in paragraph 4.3.1. The listing 6.2 shows only the main part of the xml file, the one containing the test's configuration, excluding the first and the second part, which are a simple list of definitions and databases.

```
<test name="Telephone company use case">
  <initializator name="AccountInitializator" numberOfObjects↘
     ="1000000"/>
  <initializator name="MsisdnInitializator" numberOfObjects=↘
     "1000000"/>
```

```
<initializator name="SessionInitializator" numberOfObjects↘
    ="1000000"/>
<task tps="10" time="120000" name="Balance check">
  <operation name="ReadMsisdnOperation" repetition="1"/>
  <operation name="ReadAccountOperation" repetition="1"/>
</task>
<task tps="10" time="120000" name="Account management">
  <operation name="WriteAccountOperation" repetition="1"/>
  <operation name="WriteMsisdnOperation" repetition="2"/>
</task>
<task tps="2000" time="120000" name="Service authorization↘
     and management">
  <operation name="ReadMsisdnOperation" repetition="1"/>
  <operation name="ReadAccountOperation" repetition="1"/>
  <operation name="WriteSessionOperation" repetition="1"/>
</task>
<loggraphreporter name="Balance check transactions">
  <line x="Time" xTask="Balance check" y="Transaction" ↘
     yTask="Balance check" name="Transactions"/>
</loggraphreporter>
<loggraphreporter name="Balance check AVG transactions">
  <line x="Time" xTask="Balance check" y="AvgTransaction" ↘
     yTask="Balance check" name="AVG"/>
</loggraphreporter>
<loggraphreporter name="account management transactions">
  <line x="Time" xTask="Account management" y="Transaction↘
     " yTask="Account management" name="Transactions"/>
</loggraphreporter>
<loggraphreporter name="account management AVG ↘
    transactions">
  <line x="Time" xTask="Account management" y="↘
     AvgTransaction" yTask="Account management" name="AVG"↘
     />
</loggraphreporter>
<loggraphreporter name="Service authorization and ↘
    management transactions">
  <line x="Time" xTask="Service authorization and ↘
     management" y="Transaction" yTask="Service ↘
     authorization and management" name="Transactions"/>
</loggraphreporter>
<loggraphreporter name="Service authorization and ↘
    management AVG transactions">
  <line x="Time" xTask="Service authorization and ↘
     management" y="AvgTransaction" yTask="Service ↘
     authorization and management" name="AVG"/>
</loggraphreporter>
<lineargraphreporter name="Memory usage">
  <line x="Time" y="Memory" name="Memory"/>
</lineargraphreporter>
<lineargraphreporter name="Cpu usage">
  <line x="Time" y="Cpu" name="cpu"/>
```

```
  </lineargraphreporter>
</test>
```

Listing 6.2: Real time prepaid system XML

We already explained how to build this input file in the previous paragraph, therefore here we will focus only on some particular aspects. First of all, recalling the real time prepaid system test case, it was structured in three tasks, and in fact this test is composed by three task. The balance check task and the account management task have both the same limit on the throughput, ten transactions per second: they are not the memory/cpu intensive tasks. On the other hand the service authorization and management task has a very high throughput which is two thousand transactions per second. Another thing to point out is that this test uses three initializators to fill the database with millions of objects, as was stated in the test case's definition. Of course immediately jumps to the reader's attention the great quantity of reporters used by this test: they all represent the test's performance by different point of view, such as the point of view of the different tasks. Furthermore they also take different measures, such as the throughput or the memory usage, and produce different graphs: one graphs for each reporter.

In the following chapter we will show and analyze more in detail the results produced by reporters and by the benchmark in general.

6.3.2 Implementation of New Tests

As already explained, our database benchmark application is built as a framework, allowing the user to implement new tests and new databases. In this section we will analyze how to implement a new test case.

All this explanation beats about a specific example, regarding an object person and the corresponding write and read operations. The listing 6.3 shows the interface of the object person, a very simple POJO with just two private fields: an id and a name. The interface is used because very often in-memory database systems use a proper implementation of the objects to be stored, and therefore the interface the interface provides a proper level of indirection, enabling the dialogue between the operations and the database drivers.

```
public interface Person {
  public long getId();
  public void setId(long id);
  public String getName();
  public void setName(String name);
}
```

Listing 6.3: Person interface

The listing 6.4 is, instead, the person's implementation used by the test to instantiate new person object. This is a very simple Java bean which only implements the person's interface and nothing else, in order to keep it simple and fast.

```
public class PersonImpl implements Person {
  private long id;
  private String name;

  @Override
  public String getName() {
    return name;
  }
  @Override
  public void setName(String name) {
    this.name = name;
  }
  @Override
  public long getId() {
    return id;
  }
  @Override
  public void setId(long value) {
```

```
    this.id = value;
  }
}
```

Listing 6.4: Person implementation

After the definition of the object/s we will use in this test case, we have to implement the operations. We will show an example for a simple read operation and another for a simple write operation. The listing 6.5 shows the read person operation's implementation. The class must extend the base class `Operation` implementing the following methods: `setDatabase`, `setUp`, `warmUp`, `doOperation` and `tearDown`. We have already explained the meaning of this methods in paragraph 6.2.4. An important thing to point out is that when the `setDatabase` method is invoked, the operation knows the database which will be used, and therefore asks to the database the specific driver it needs to work. Furthermore is very interesting also the `doOperation` method, which is the real operation executed on the database, and which will be measured by the benchmark. There is a simple random id generator which is used to read the person from the database (it is a traditional `getById` method). Then the `doOperation` method also execute some integrity test, checking the result value from the database. This is one of the biggest difference with Poleposition where the benchmark doesn't communicate with the database: in Poleposition there is no parameter passing.

```
public class ReadPersonOperation extends Operation {
  private PersonDriver personDriver;
  private int numberOfObject;
  private Random random = new Random();

  @Override
  public void setDatabase(Database database) throws
      NoSuitableDriverException {
    personDriver = (PersonDriver) database.getSpecificDriver
      (PersonDriver.class);
```

```
  }
  @Override
  public void setUp() throws TestDriverException, ↘
      OperationException {
    personDriver.connect();
    numberOfObject = personDriver.getNumberOfPerson();
    if (numberOfObject == 0)
      throw new OperationException("The number of objects ↘
          initializated in the database must not be zero");
  }
  @Override
  public void warmUp() throws OperationException, ↘
      TestDriverException {
    for (int i = 0; i < 10; i++) {
      doOperation();
    }
  }
  @Override
  public void doOperation() throws OperationException, ↘
      TestDriverException {
    int randomId = random.nextInt(numberOfObject);
    Person person = personDriver.read(randomId);
    if (person == null) {
      throw new OperationException("Null person in ↘
          ReadPersonOperation: is the database initialized?")↘
          ;
    } else if (person.getId() != randomId) {
      throw new OperationException("Wrong id in person in ↘
          ReadPersonOperation");
    }
  }
  @Override
  public void tearDown() throws TestDriverException {
    personDriver.close();
  }
}
```

Listing 6.5: Read Person operation

The listing 6.6 shows the write person operation's implementation. It is almost equal to the read person operation, the only difference is in the `doOperation` method, which is, as already said, the heart of the operations.

```
public class WritePersonOperation extends Operation {
  private StringGenerator stringGenerator = new ↘
      StringGenerator();
  private int id;
  private PersonDriver personDriver;
  private int numberOfObject;
```

```
  @Override
  public void setDatabase(Database database) throws ↘
      NoSuitableDriverException{
    personDriver = (PersonDriver) database.getSpecificDriver↘
        (PersonDriver.class);
    id = 0;
  }
  @Override
  public void setUp() throws TestDriverException {
    personDriver.connect();
    numberOfObject = personDriver.getNumberOfPerson();
  }
  @Override
  public void warmUp() throws TestDriverException {
    for (int i = 0; i < 10; i++) {
      doOperation();
    }
  }
  @Override
  public void doOperation() throws TestDriverException {
    Person person = newRandomPerson();
    person.setId(id + numberOfObject);
    personDriver.write(person);
    id ++;
  }
  @Override
  public void tearDown() throws TestDriverException {
    personDriver.close();
  }
  /*
   *  ###############  private methods here ############
   */
  private Person newRandomPerson() {
    Person person = new PersonImpl();
    initPerson(person);
    return person;
  }
  private void initPerson(Person person) {
    person.setName(stringGenerator.getRandomString());
  }
}
```

Listing 6.6: Write Person operation

In the paragraph 6.3.1 we have introduced the database initializator, a special object which is used, before the test starts, to fill the database. The listing 6.7 shows the database initializator for the object person. It must extends the class `DbInitializator` and implement two methods. The `init`↘

method takes as input the database and the number of object to be inserted and execute this operation thanks to a special method provided by the driver, which is very fast and dedicated to the initialization of a huge number of objects. The method `getDescription` is instead used only to produce the PDF output, so that the reader can understand how the database was initialized.

```
public class PersonInitializator extends DbInitializator {

  @Override
  public void init(Database database, int numberOfObjects) ↘
     throws NoSuitableDriverException, TestDriverException {
    PersonDriver personDriver = (PersonDriver) database.↘
       getSpecificDriver(PersonDriver.class);
    personDriver.connect();
    personDriver.init(numberOfObjects);
    personDriver.close();
  }
  @Override
  public String getDescription() {
    return "Number of Person initialized: " + ↘
       getNumberOfObjects();
  }
}
```

Listing 6.7: Person database initializator

Now that we have seen all this classes and all the methods invoked on the driver, the implementation of the driver's interface is extremely simple. In fact the listing 6.8 is just a collection of all the methods previously described. Then the database will need to provide a proper implementation for this driver.

```
public interface PersonDriver {
  public void connect() throws TestDriverException;
  public void init(int numberOfObject) throws ↘
     TestDriverException;
  public int getNumberOfPerson() throws TestDriverException;
  public void close() throws TestDriverException;
  public Person read(long personId) throws ↘
     TestDriverException;
  public void write(Person person) throws ↘
     TestDriverException;
```

```
}
```

Listing 6.8: Person database driver

6.3.3 Implementation of New Databases

While in the previous paragraph we explained how to implement a new test, here we will examine how to implement a new database. The example used regards, as for the test's implementation, the person object, with the read and the write operations. Essentially a database just need to provide and implement a suitable driver for each operation, such as the one reporterted in listing 6.9, which is an implementation of the person driver already described in listing 6.8.

```
public class PersonHsqldbDriver implements PersonDriver {
  private StringGenerator stringGenerator = new ↘
     StringGenerator();
  private final String WRITE_QUERY = "insert into PERSON ↘
     values (?,?)";
  private final String READ_QUERY = "select name from PERSON↘
      where id=?";
  private final String CREATE_TABLE_QUERY = "create table ↘
     PERSON (name varchar,id bigint primary key)";
  private final String COUNT_PERSON = "select count(*) from ↘
     PERSON";
  private String password = "";
  private String databaseURL = "jdbc:hsqldb:file:"+new ↘
     HsqldbDatabase().getFileName()+"/database";
  private String username = "sa";
  private static final String driver="org.hsqldb.jdbcDriver"↘
     ;
  private Connection con;

  @Override
  public void close() throws TestDriverException {
    try {
      con.close();
    } catch (SQLException e) {
      throw new TestDriverException(e);
    }
  }
  @Override
```

```
public void connect() throws TestDriverException {
  try {
    Class.forName(driver);
    con = DriverManager.getConnection(databaseURL,username
        ,password);
  } catch (SQLException e) {
    throw new TestDriverException(e);
  } catch (ClassNotFoundException e) {
    throw new TestDriverException(e);
  }
}
@Override
public int getNumberOfPerson() throws TestDriverException
    {
  int result = 0;
  try {
    PreparedStatement st = con.prepareStatement(
        COUNT_PERSON);
    ResultSet rs = st.executeQuery();
    if (rs.next()){
      result = rs.getInt(1);
    }
  } catch (SQLException e) {
    throw new TestDriverException(e);
  }
  return result;
}
@Override
public void init(int numberOfObject) throws
    TestDriverException {
  try {
    PreparedStatement createStatement = con.
        prepareStatement(CREATE_TABLE_QUERY);
    createStatement.executeUpdate();
    PreparedStatement writeStatement = con.
        prepareStatement(WRITE_QUERY);
    for (int i = 0; i < numberOfObject; i++) {
      writeStatement.setString(1, stringGenerator.
          getRandomString());
      writeStatement.setLong(2, i);
      writeStatement.executeUpdate();
    }
    con.createStatement().execute("SHUTDOWN");//TODO da
        sistemare!!!!
  } catch (SQLException e) {
    throw new TestDriverException(e);
  }
}
@Override
public Person read(long personId) throws
    TestDriverException {
```

```
    Person person = null;
    try {
      PreparedStatement st = con.prepareStatement(READ_QUERY↘
          );
      st.setLong(1, personId);
      ResultSet rs = st.executeQuery();
      person = new PersonImpl();
      if (rs.next()){
        person.setName(rs.getString(1));
        person.setId(personId);
      }
    } catch (SQLException e) {
      throw new TestDriverException(e);
    }
    return person;
  }
  @Override
  public void write(Person person) throws ↘
      TestDriverException {
    try {
      PreparedStatement st = con.prepareStatement(↘
          WRITE_QUERY);
      st.setString(1, person.getName());
      st.setLong(2, person.getId());
      st.executeUpdate();
    } catch (SQLException e) {
      throw new TestDriverException(e);
    }
  }
}
```

Listing 6.9: HSQLDB implementation of PersonDriver, listed in 6.8

After the implementation of the driver, we just need to add the driver in the database object, which must be written only once. The listing 6.10 shows this database implementation and how we can add the new driver directly in the constructor. Note that the method `getFileName` is used by the benchmark to get the file location used to store the database's image or to log the transactions. It's very important to know where the database will be stored in order to measure the database's file size.

```
public class HsqldbDatabase extends Database {
  public HsqldbDatabase(){
    add(PersonDriver.class, PersonHsqldbDriver.class);
  }
```

```
  @Override
  public void clear() {
    DirectoryCleaner.clean(getFileName());
  }
  @Override
  public String getFileName() {
    return "db/hsqldb";
  }
  private String password = "";
  private String databaseURL = "jdbc:hsqldb:file:db/hsqldb/↘
     database";
  private String username = "sa";
  private static final String driver="org.hsqldb.jdbcDriver"↘
     ;
  @Override
  public void shutdown() throws TestDriverException, ↘
     FileNotFoundException,SQLException, IOException, ↘
     URISyntaxException, ClassNotFoundException {
    Class.forName(driver);
    Connection con = DriverManager.getConnection(databaseURL↘
       ,username,password);
    con.createStatement().execute("SHUTDOWN");
  }
}
```

Listing 6.10: HSQLDB database

Another necessary observation is that the driver implemented by the database will not have a unique instance, but it will be instantiated the same number of time of the operations composing the test which use it. Therefore when working with database systems which don't manage concurrent accesses, which happens more often with in-memory databases, it's important to use appropriate patterns such as factory and singleton. If you want to know more about this use, you can just look at the code provided with the benchmark application and look at the implementation of Db4o or Prevayler.

Chapter 7

Results' Analysis

In this chapter we will deal with some of the results obtained by the execution of the test suite, produced in chapter 4, on the database benchmark application, developed and described in chapter 6. The benchmark is executed on a Sun Solaris machine, with 65GB of main memory and 32 processors, with a single 10000 rpm hard disk.

The results obtained come from the execution of 3 test cases on 5 different competiting database systems. The test executed are: the real time prepaid system, the race test based on transactions' number and the reace test based on the test's time. The competitors who take place in the benchmark are: Pico4, Prevayler, HSQLDB, H2, Db4o.

7.1 Race Test: Timed

This test is based on the test case described in paragraph 4.2.1. It is a base test case and therefore it is configured as a race, where all the competiting

database systems run at their maximum speed. In this particular case all the database systems execute continuously a single write operation until a fixed amount of time is passed.

This test, and its results, are not useful to understand the real performance and potentiality of the database and so they are not useful to choose the database for our needs, but it can be used in an early stage to reduce the databases' set which will be analyzed extensively with further tests. In other words we can throw away every databases whose maximum throughput is not enough for our needs.

7.1.1 Total Number Of Transactions

The figure 7.1 shows on the y axis the total number of transactions executed by database systems, and therefore the total number of writes, because one transaction in this test corresponds to a write. On the x axis there is the time, and we can notice how the test, as described in paragraph 4.2.1, lasts for 60 seconds. It's important to point out how the y axis use a logarithmic scale. Note also that the database was initialized with 1 million of object before the test was started.

This graph shows, therefore, the result of the race, and we can see how the fastest database system, in this case, is Pico4, except for the first 30 seconds where HSQLDB was a bit faster. Anyway also this second system is very fast. In 60 seconds Pico4 executed more than 5 millions of writes, with a throughput of about 900 writes per second, while HSQLDB executed "just" 4 millions of writes. On the other hand Prevayler is extremely slow,

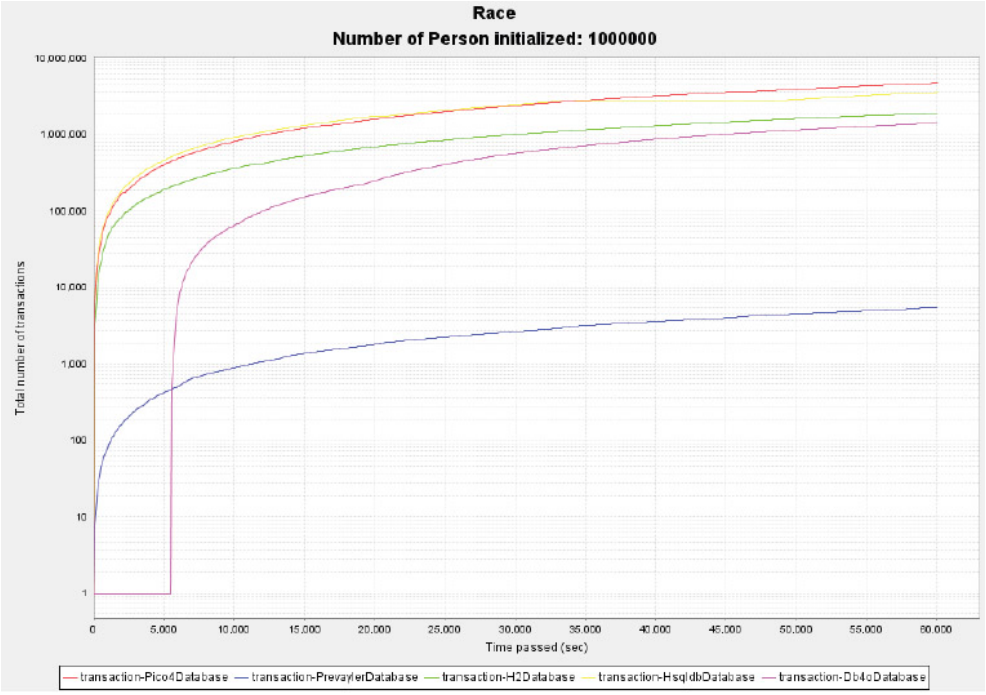

Figure 7.1: Result of the timed race: total number of transactions executed

compared to the others database systems: it executed only 5 thousands of writes, which means a throughput of less than 100 writes per second. So there is a difference about 3 order of magnitude between Prevayler and Pico4. Another interesting observation for this graph is that Db4o has a strange behavior at the start of the test, but then it becomes very fast. This anomaly is put in evidence by the use of this graphs. This is probably caused by a bad implementation of the database driver, or something else, but it clearly tells us that in this case the comparison between the several database systems with Db4o is not appropriate. This strange behavior is exalted by the use of graphs, which therefore are extremely powerful for a better understanding of the test' result. An average value would, instead, have hidden this peculiarity.

7.1.2 Memory Usage

The figure 7.2 is a graphical representation of the main memory used during the test execution. Note that on the x axis there is the total number of transactions, and not the time, in fact not all the lines represented get to the right side of the graph. This has been done because the memory usage doesn't depend on the time, but it depends mainly, in this case, on the number of objects inserted in the database, and therefore in main memory, because we are using in-memory database systems. Another important thing to point out is that this graph represents the memory usage when all the systems are at their 100% usage, because they all execute only writes consecutively without any limitations on the throughput.

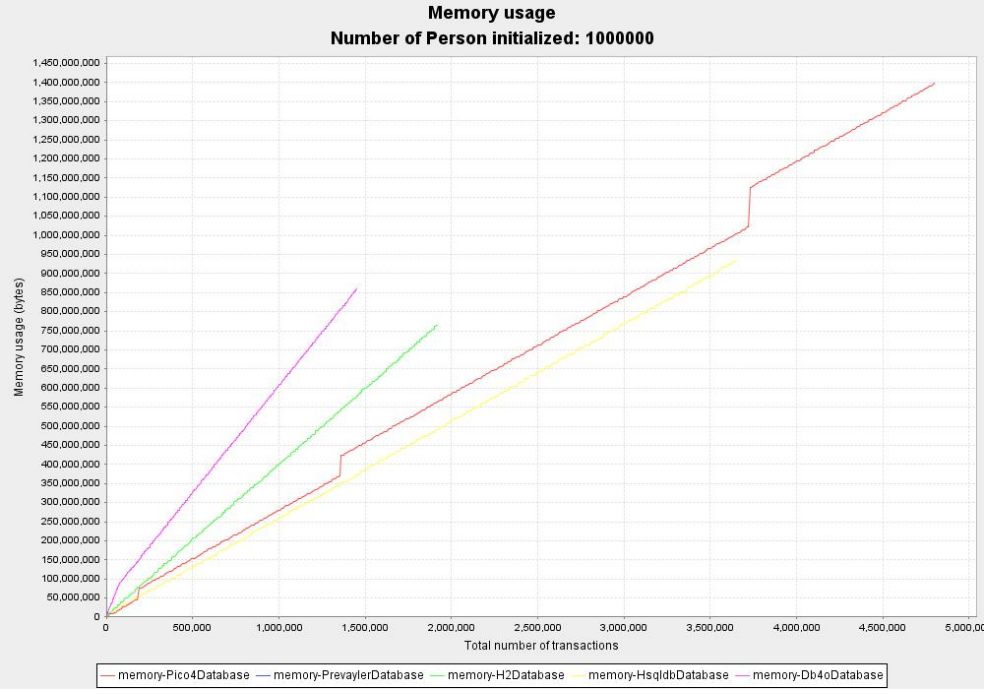

Figure 7.2: Memory usage per write

Analyzing this graph we can see how apparently there are only 4 database systems. This is because Prevayler executes so few transactions that it is hard to represent it on the graph. Then we can see how both Pico4 and HSQLDB use less memory than H2 and Db4o, and this is much more surprising considering that they are also the fastest systems. So we can conclude that their speed doesn't come from a huge memory usage. In addition Pico4 shows a bit strange behavior because, differently from the others, it sometime increments the memory used in blocks, but there too many possible reason for this behavior, such as particular JVM behavior. It would be useful to rerun this test to analyze if this behavior is still present.

7.2 Race Test: Transactions

This test is based on the test case described in paragraph 4.2.2. This is a base test case too, like the timed race test described in the previous chapter. In this case all the competiting database systems must execute a fixed amount of operations. In fact the stop condition for this test is a fixed total amount of transactions, unless they exceed a certain time out and they are stopped also if they are not finished yet. The only operation, which is continuously repeated, is a simple write operation, but then we executed this test a second time with a read operation instead a write. Therefore we will see the results of the two executions. In both executions the database was initialized with 1 million of objects.

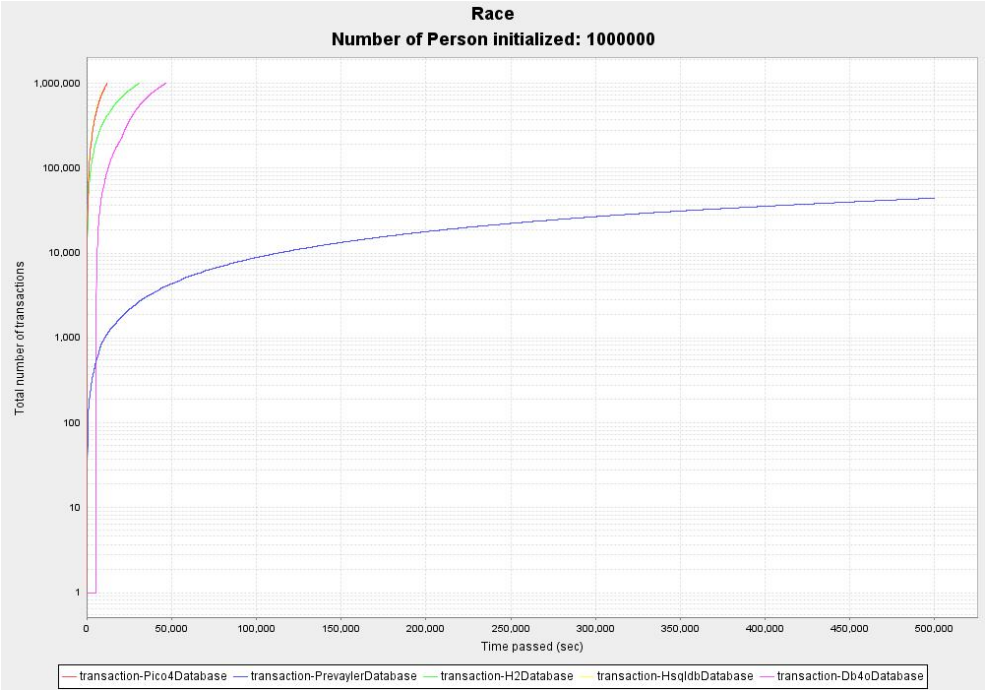

Figure 7.3: Execution 1: total number of writes executed

7.2.1 Total Number Of Transactions

The figure 7.3 represents the result of the first execution, where the only operation executed was a write operation. The graph is the same of the figure 7.1, but this time the database systems stop when they reach the amount of 1 million of transactions executed. This graph shows clearly that Pico4 and HSQLDB are the fastest for this kind of object and operation, and they do it in about 20 seconds. On the other hand Prevayler, also after 500 seconds isn't able to execute 1 million of writes, then it is stopped because it reached the time out.

The figure 7.4 represents instead the result of the second execution, where the only operation executed is a read operation. Surprisingly Prevayler is now

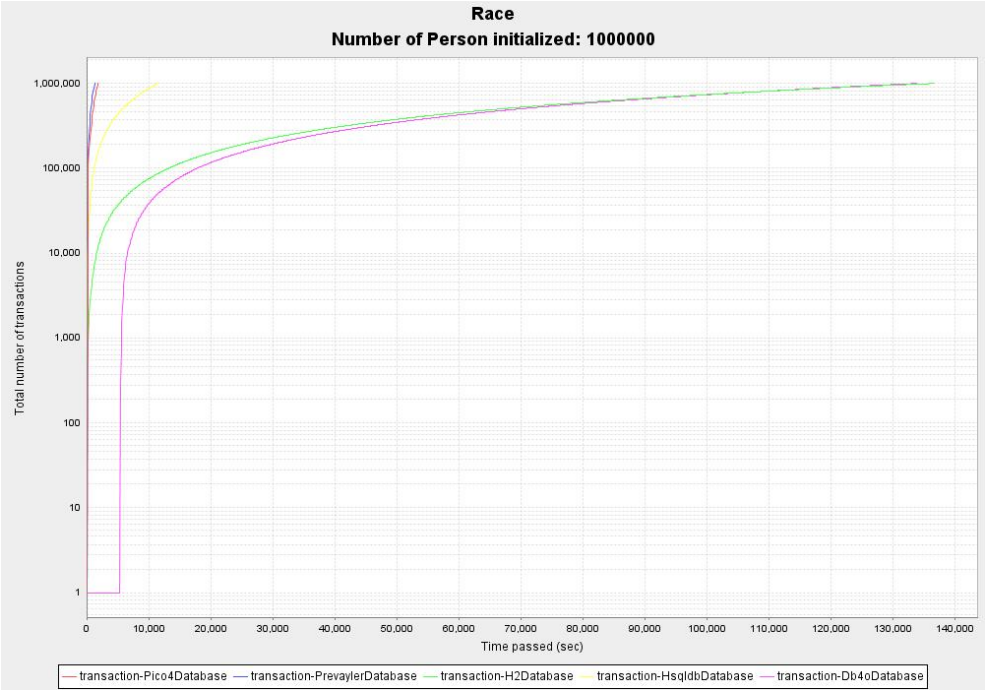

Figure 7.4: Execution 2: total number of reads executed

the fastest with a throughput of about 1 million of reads per second. Soon after there is Pico4 which shows performance almost equal to Prevayler's performance. Then HSQLDB which is still very fast with a throughput of about 100 thousands of reads per second. Finally there is H2 and Db4o which really similar performance and still a reasonable throughput, which is something more than 7 thousands of reads per second.

We can conclude that Prevayler is the fastest system when you need to execute read only operation, but it is dramatically slow for write operations, especially if you database is huge. Instead Pico4 and HSQLDB performed very good, if not the best, in both cases.

7.2.2 Memory Usage

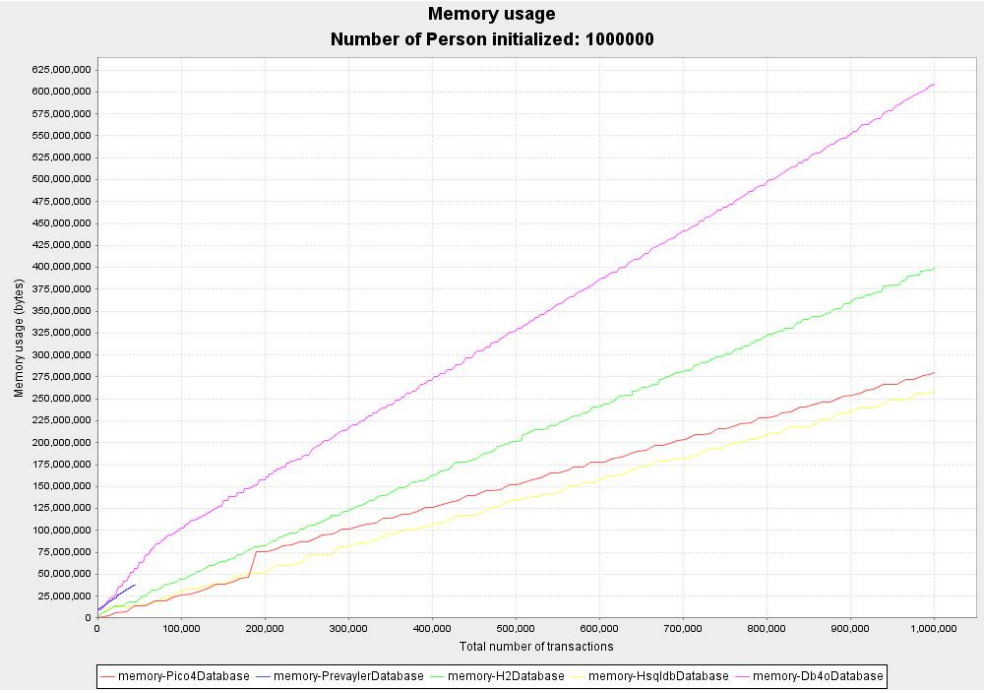

Figure 7.5: Execution 1: memory usage for write operations

The graph in the figure 7.5 represents the memory usage of the first execution of the test, where the only operation executed is a write. The result in fact is the same of the one represented in figure 7.2, except that almost all the lines get to the right side of the graph. But now we can see the performance, in terms of memory used, of Prevayler, which can be placed between H2's and Db4o's performance.

The figure 7.6 represents instead the memory used during the second execution of the test, where there are only read operations. As we could expect Pico4, HSQLDB and Prevayler don't use any memory, in fact the data is already all in main memory, and there is no need to allocate new

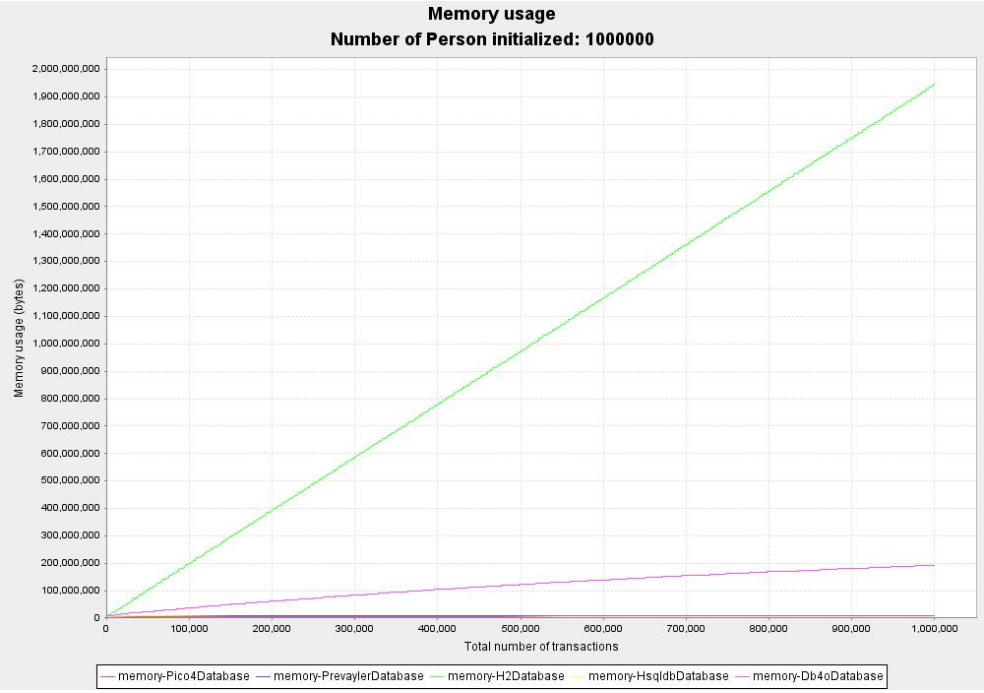

Figure 7.6: Execution 2: memory usage for read operations

memory: there are no writes. Peculiarly H2, but also Db4o, increment the memory used linearly with the increment of the read operations executed. This is a very strange behavior that is put in evidence by this graph.

7.2.3 File Size

The figure 7.7 represents the file size during the first execution of the test, therefore with all write operations. All the databases show a natural increment of the file size. Pico4 and Prevayler use a smaller file size than the other systems, and also the increment, which is linear to the number of writes executed, is less than the others. H2 instead shows a really strange behavior, with a file size which increments and decrease too. A behavior which is

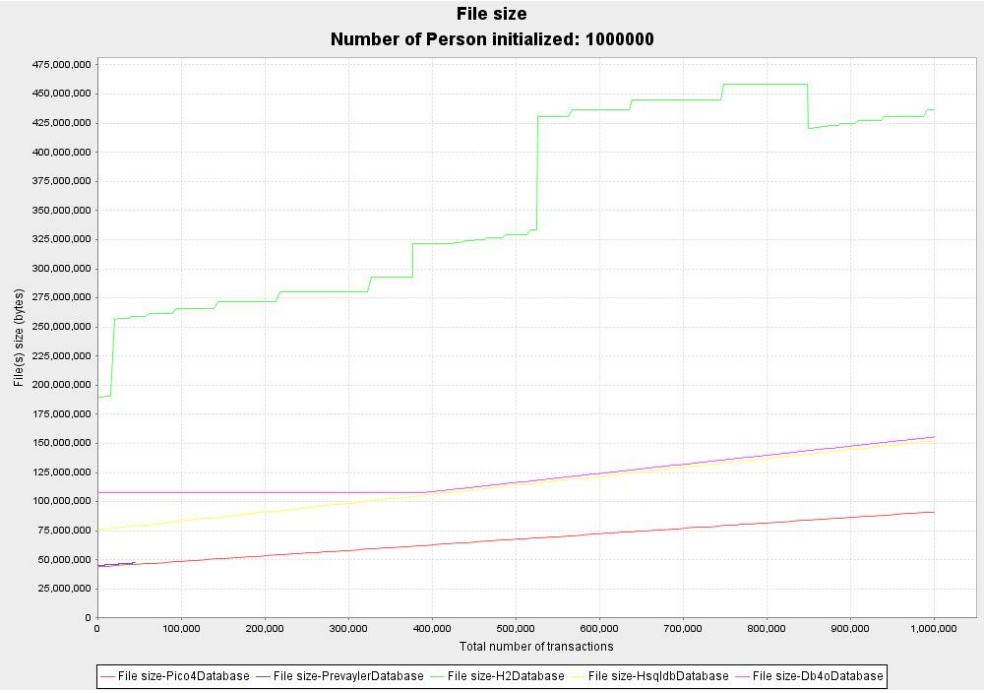

Figure 7.7: Execution 1: file size for write operations

typical of relational databases.

The figure 7.8 represents the file size during the second execution of the test, where all all the operations executed are read operation. As we could expect none of the databases increments the files size. This is of course a normal behavior.

In addition from the last 2 graphs we can see how much memory is used to store a fixed amount of objects. In fact the initial value of the file size is due to the database which is already initialized with 1 million of object. Therefore we can see that Pico4 and Prevayler use less than 50MB to store 1 million of object, while H2 use about 190MB, which is a big difference. If you consider to use such a system on an enterprise application you need to

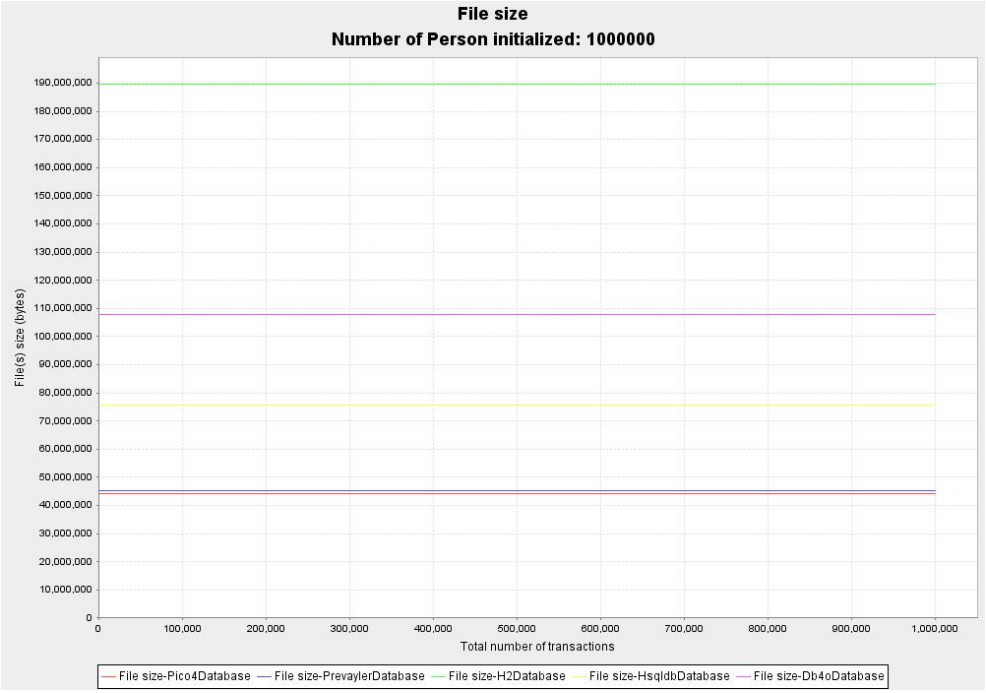

Figure 7.8: Execution 2: file size for read operations

take into consideration the huge amount of bytes used on the hard disk to operate.

7.3 Real Time Prepaid System Test

This test represents the execution of the test case described in paragraph 4.3.1 during the development of the test suite. This is the main test case produced in the test suite and the purpose of this thesis, which was to find the best in-memory database, in terms of performance, for a real time prepaid system, such as a telephone company. This is a load test case, and therefore it is not configured as a race. As already explained, it is composed by three tasks: balance check task, executed 10 times per second; accounts management

task, executed 10 times per second; services authorization and management task, executed 2000 times per second. The database during this test was initialized with 12 millions of objects: 4 millions of accounts, 4 millions of sessions and 8 millions of MSISDN. Now, let's start to analyze some results for this test case.

7.3.1 Throughput

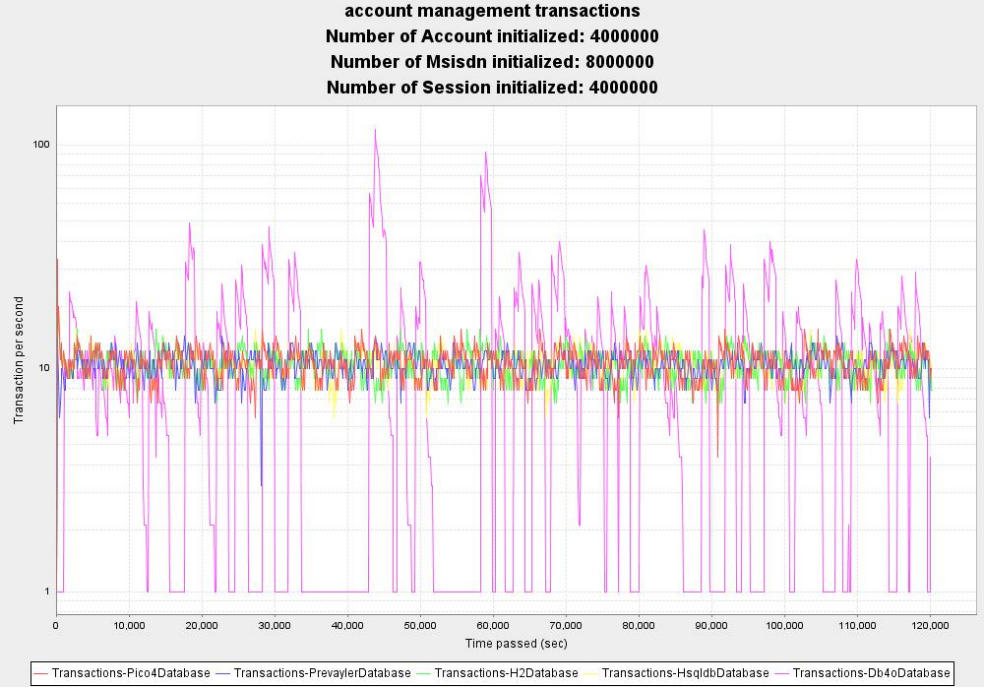

Figure 7.9: Throughput of the account management task

The figure 7.9 shows the throughput of the account management task, which we imposed to 10 transactions per second. In fact we can see how all the databases oscillate around 10. But an important observation for this graphs is that Db4o has very strange oscillation, and the throughput is

not constant, also if the average value is about 10. This fact is very well highlighted by the use of graphs.

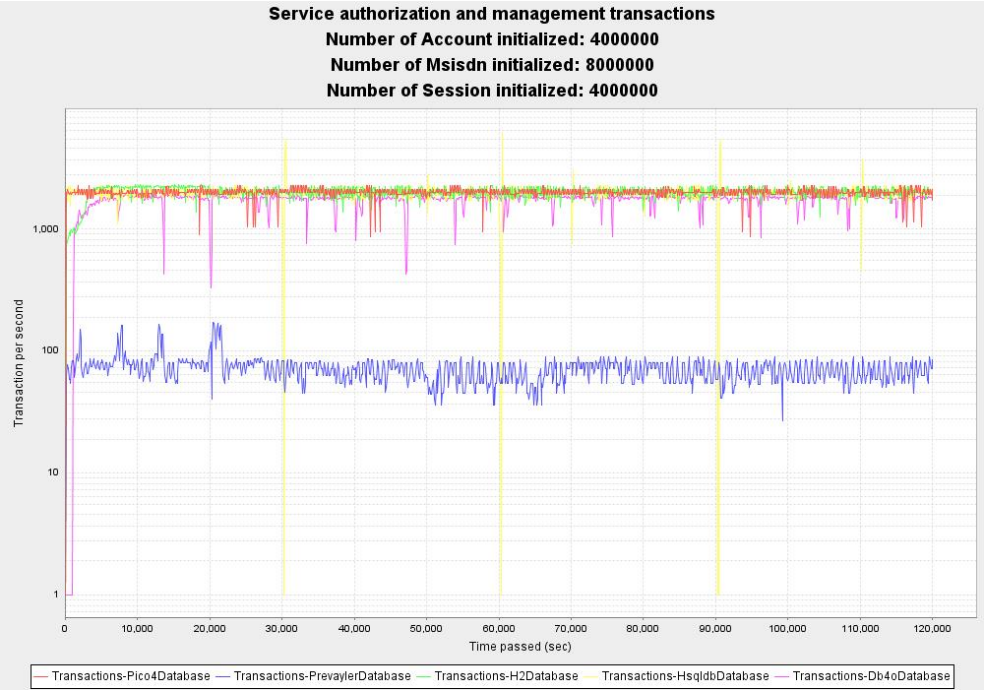

Figure 7.10: Throughput of the service authorization and management task

Instead, the figure 7.10 shows the throughput of the service authorization and management task, which we imposed to 2 thousands of executions per second. This is the most important task in term of performance, in fact this is the usual bottle neck for this application usage. From this graph we can notice how Prevayler is again the worst database system, and this fact confirms the previous tests. In addition also Db4o is slightly under the desired throughput. Another interesting observation is that HSQLDB approximately every 30 seconds has a break, which, although it is very short, can be an important feature for real time systems, such the one we are

simulating with this test.

7.3.2 Memory Usage

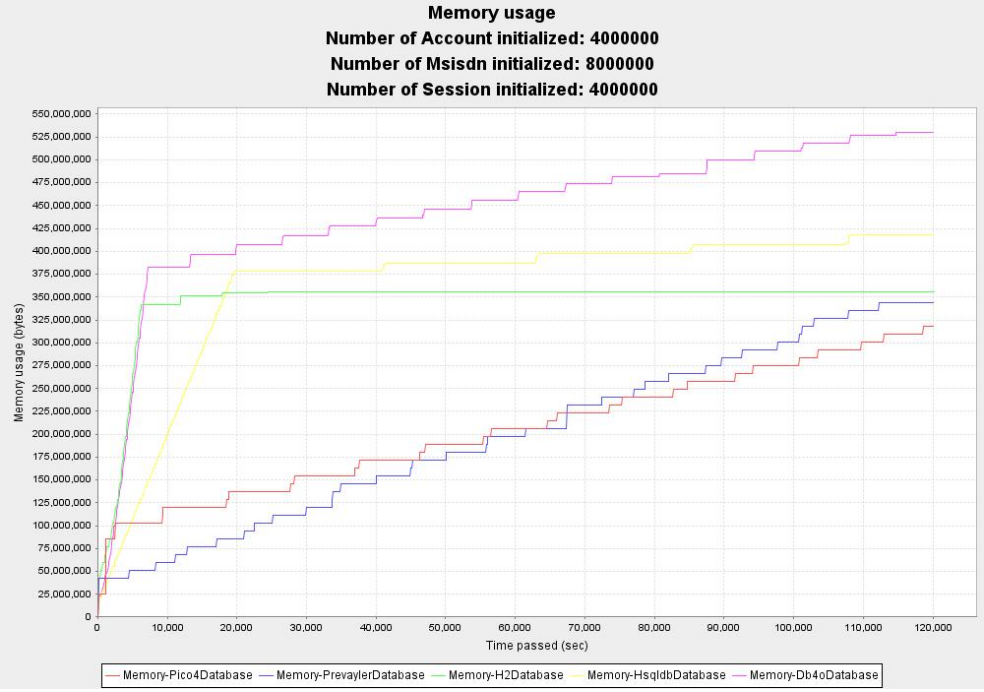

Figure 7.11: Memory usage

The figure 7.11 shows the memory usage during the execution of this load test. It's interesting to note how HSQLDB, although during the transaction test showed in figure 7.5 was using less memory than the other systems, now comes soon after Db4o. This means that is not possible to study database behavior only with simple tests such as the trasanction race test. In this graph Pico4 is still the best database in term of memory used. In fact note that Prevayler can't be compared to the other database systems, because it wasn't able to provide the desired throughput showed in figure 7.10.

7.3.3 CPU usage

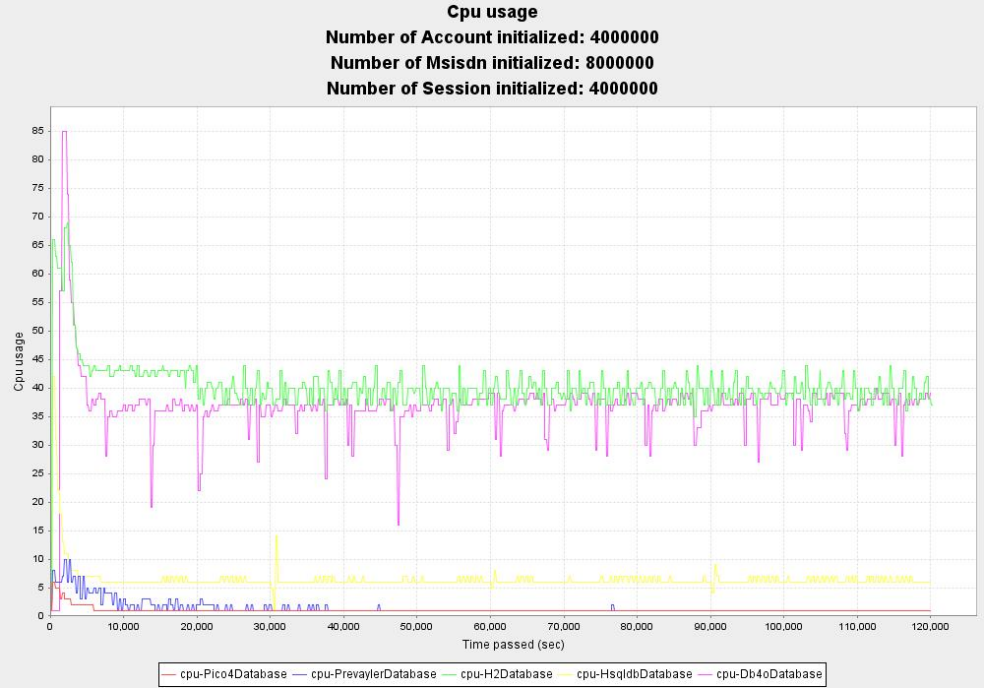

Figure 7.12: CPU usage

The figure 7.12 represents the CPU usage of the competiting systems during the test execution. Take in mind that also this time Prevayler can't be compared to the others because it wasn't able to provide the desired throughput for the service authorization and management task. We can see how Pico4 practically doesn't use the CPU, while HSQLDB use a very limited amount of CPU. Instead Db4o and H2 use about 40% of CPU, which means about half CPU of a systemcomposed by 32 CPU. In fact, as always, the database bottle neck is the hard disk, and it is not the CPU.

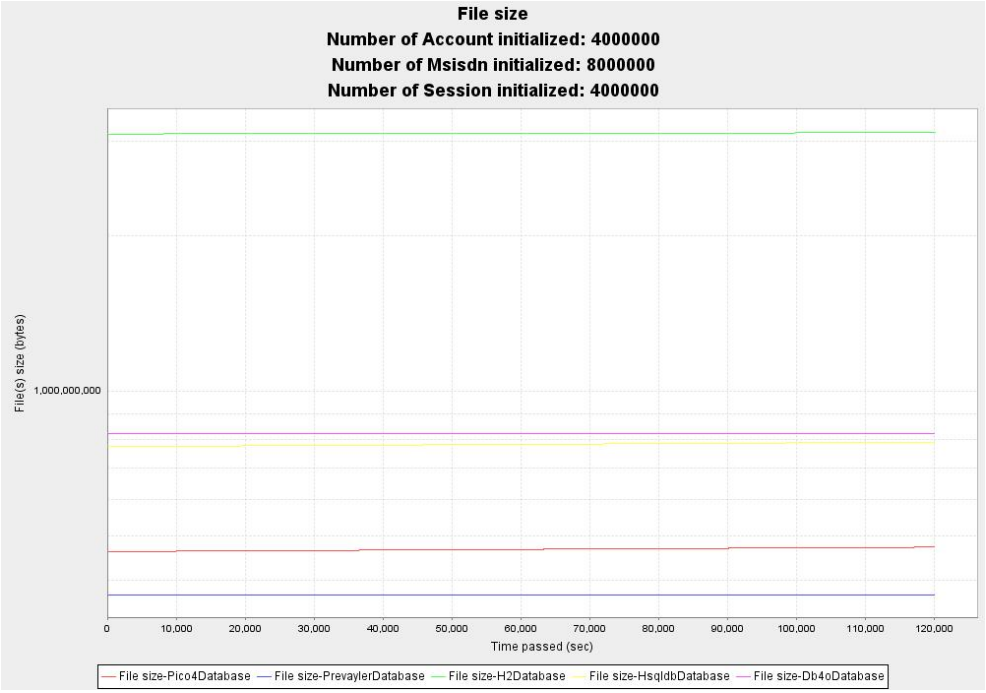

Figure 7.13: File size

7.3.4 File Size

Finally the figure 7.13 shows the file size during the execution of this test. As we can see for all the database systems, the file size's increments are minimally perceptible. This means that this application scenario, although composed by different write operations, doesn't require a huge disk space. But this test lasts only 120 seconds. It would be really interesting to monitor the file size during a test whose duration is longer, for example some hours.

This graph put in evidence also the file size used to initialize the database with 12 millions of objects. Both Db4o and HSQLDB use about 800MB of the disk space, while Pico4 uses less than 500MB and Prevayler less than 400MB. But Prevayler, as already said, wasn't able to provide the desired

throughput. On the other hand there is H2, which uses more than 3GB of the disk space to initialize 12 millions of objects.

7.4 In Summary

The analysis of all these results put in evidence the efficacy of the graphs used, which are produced directly by the benchmark application. They highlight the trend of the performance during the tests' execution. Any oscillation in the performances becomes evident and easy to read and understand. Of course such a behavior would be hidden by simple average values. In addition these graphs are also useful to understand some miss in the tests' implementation.

Before summarizing the results gathered, it's important to point out that these test should also be executed more than just once, as we did, because there are too many transient factors which may produce wrong results. Moreover there is an infinity of possible tuning for the platform software and hardware used for the benchmark. For example we used the following option when executing the "java" command:

```
-Xms50G -Xmn512M -Xmx50G -d64 -XX:+UseParNewGC -XX:+↘
   UseConcMarkSweepGC -XX:CMSInitiatingOccupancyFraction=60 ↘
   -XX:ParallelGCThreads=32 -XX:SurvivorRatio=8 -XX:↘
   TargetSurvivorRatio=90 -XX:MaxTenuringThreshold=31 -↘
   Xloggc:javagc.log -XX:+PrintGCDetails -XX:+↘
   PrintGCTimeStamps -XX:-TraceClassUnloading -XX:+↘
   ExplicitGCInvokesConcurrent -XX:PermSize=512m -XX:↘
   MaxPermSize=512m -Xdebug -Xrunjdwp:transport=dt_socket,↘
   address=1044,server=y,suspend=n
```

In this particular case, this means that we assigned 50GB of main memory to the Java process, that the garbage collector use was the concurrent mark

sweep (CMS), that the permanent size was 512MB, and so on.

Coming to a conclusion, Prevayler surprised us for its performance in read operations, but on the other hand it has a dramatically slow throughput for write operations, which was less than 100 writes per second. But it's really important to note that these performances depend also on the platform used. For example, the same execution, on a laptop with an intel core 2 duo T7500, 3GB of RAM and a 4200rpm hard disk, produced a throughput of about 1 thousand of writes per second. Anyway, Prevayler is not what a real time prepaid system needs, unless there is no durability requirements. Probably these Prevayler's performances are due to a strict synchronization with the hard disk.

Then also Db4o and H2 are not suited for a real time prepaid system. Although they showed better performance than Prevayler, they are average performance, especially comparing them to Pico4 and HSQLDB. Moreover both H2 and Db4o showed a great use of shared resources such as the CPU, the RAM usage and the space on the hard disk. Particularly in regards of this last feature, H2 showed a huge use of the file on the hard disk, with a behavior that is more typical of traditional database management systems.

Instead HSQLDB and Pico4 showed very good performances: their throughput was always very high both in reads and writes, and also the use of shared resources was really limited. But between these two database systems, Pico4 is the leading database: it has a better throughput in read operations and also slightly better in write operations; the CPU usage is very low; and also the file size and the memory used are lower than HSQLDB. Therefore, even if HSQLDB is a great database in this scenario, Pico4 is of course the most

suited.

This is because Pico4 was developed exactly for this kind of application scenario, while HSQLDB and the other systems, are more for general purpose use. In addition Pico4 doesn't support any transaction isolation level, but it was not required by this test case. Therefore this benefits the performances of Pico4.

In conclusion, Pico4 is certainly the best database system for a real time prepaid system application, because it was developed for this kind of application. This means that, when you need high performance, the development of an ad-hoc solution is always the best solution. Of course this is not always possible, and often we don't need the best performance for a specific scenario, but just very good performance, therefore there are a lot of possible "cheaper" solution already available on the market.

Chapter 8

Conclusion

This chapter is the conclusive chapter of the thesis and it is divided into two sections. In the first section we will state again some of the major contributions brought by this thesis and how this thesis can be used by other people. In the second section we will analyze some future research and development for the benchmark application produced during this work.

8.1 Thesis' Contributions

All the work of this thesis is around the topic of in-memory database systems, which has been analyzed from different point of views. This thesis is, in fact, very useful for researchers who are going to study IMDBs. Particularly the first chapter gives a brief introduction on the major features and differences between IMDBs and traditional database management systems. A major difference is that IMDBs have usually better performance, but they have a very slow startup and they are not so brilliant in the administration

departement.

In addition this thesis provides an in-depth analysis of different IMDBs, which can be very useful for a reader who needs to choose, firstly, if he wants to use in-memory or traditional DBMS, and, secondly, which IMDB can be the best for his needs. Although the overview provided by this thesis doesn't state a winner, it asserts a lot of times that there is no the best database, they can be all good for a particular situation. Therefore an in-depth overview is very useful to understand the features of several IMDBs.

In a second part this thesis faces the performance analysis' problem and states how it is hard to correctly understand the benchmark's results. The resulting performance, in fact, may vary for too many factors, but the most important are the test scenario used to analyze the performance, their implementation and the execution platform. Furthermore, it is even more difficult to understand the results if they consist of a single measure. All these results were also confirmed by the benchmark software overview, which helped the comprehension of the performance analysis' problem.

This problem, consequently, led to the development of a proper benchmark application, suited for our needs. This new benchmark application acts as a framework and can be easily extended with new tests' and databases' implementation. Therefore this thesis can be also considered as a user manual for the development of new tests and databases. Furthermore it explain also how to use correctly the application and his configuration file, an XML taken as input by the application which allows the combination and the arrangement of new tests.

Moreover the application offers a very interesting reports. The application

produces, in fact, a detailed PDF file with a lot of graphs, highly configurable by the user. These graphs allow a constant monitoring of the performance and of how they vary over time as the benchmark runs. This is a very important feature also because it shows if the test or the database are not well implemented in the case, for example, the performance are not constant during the test's execution.

8.2 Future Benchmark's Development

The application developed during this thesis is a really powerful database benchmark tool which is based on the idea of flexibility understood as the ability to evolve very quickly. In fact this application is structured as a framework, allowing an easy implementation of new tests and databases. This is the first area of possible future development: the extension of the benchmark with new tests and databases in order to test how the databases work with new specific application scenarios. For example it would be interesting to test all these IMDBs in a mobile environment to see how they behave with limited resources.

This database benchmark application could be extended also with new monitors and reporters. At the moment there are only two different kind of reporter, and they both create a graph, one with a logarithmic and another with a linear scale. But it would be very useful also to have new reporters, such as, for example, a text reporter able to represent the test's results in a simple and schematic form.

In addition, instead of creating new reporters, it is also possible to mod-

ify the whole reporting system. It would be very useful for the benchmark application to produce only a file containing all the data collected by the monitors, and then develop a proper desktop application capable of displaying this data in any graphical representation the user can wish. The main advantage of this approach is that all the data gathered from the benchmark execution will be stored in a data file that can be later used to generate any kind of report without the need to define the report before running the tests. A benchmark execution can last even days and it would be a shame if we find out that we forgot to define a reporter only after the test has already finished.

Finally, one of the first ideas when developing the benchmark application was to implement the databases in a way that they could be able to store whatever object they get with a polymorphic method. In other words, this is a sort of mapping technology, a layer whose task was to translate the object to be stored in a proper way for each database system. This could allow a drastically speed up in the creation of new tests. Although this was realized in a primitive form and it was usable, subsequently it was dismissed for two reason:

- a further layer in the software may slow down the database's performance;
- it's not possible to implement complex test scenarios composed by not pre-existing operations.

However such a technology could be useful for a first and approximate analysis. Therefore this idea can still be taken into consideration for a future

development.

Bibliography

[Bernt 05] Johnsen Bernt. *Database Benchmarks*. `http://weblogs.java.net/blog/bernt/archive/2005/10/database_benchm_1.html`, october 2005.

[Burleson 02] Donald Burleson. *Database benchmark wars: What you need to know*. `http://articles.techrepublic.com.com/5100-10878_11-5031718.html`, august 2002.

[Council 07] Transaction Processing Performance Council. *TPC Benchmark C*. `www.tpc.org/tpcc/spec/tpcc_current.pdf`, june 2007.

[DeWitt 97] Dave DeWitt. *Standard Benchmarks for Database Systems*. `http://www.tpc.org/information/sessions/sigmod/sigmod97.ppt`, 1997.

[Fowler 05] Martin Fowler. *InMemoryTestDatabase*. `http://martinfowler.com/bliki/InMemoryTestDatabase.html`, november 2005.

[Gorine 04] Andrei Gorine. *Building durability into data management for real-time systems.* `http://www.mcobject.com/downloads/bassep04p19.pdf`, september 2004.

[Graves 02] Steve Graves. *In-Memory Database Systems.* `http://www.linuxjournal.com/article/6133`, 2002.

[Gray 93] Jim Gray. *Database and Transaction Processing Performance Handbook.* `http://research.microsoft.com/~Gray/BenchmarkHandbook/chapter1.pdf`, 1993.

[Grehan 05] Rick Grehan. *An Open Source Database Benchmark.* `http://today.java.net/pub/a/today/2005/06/14/poleposition.html`, june 2005.

[Hobbs 03] Darren Hobbs. *Prevayler Pertubations.* `http://darrenhobbs.com/?p=225`, march 2003.

[Miller 03] Charles Miller. *Prevayling Stupidity.* `http://fishbowl.pastiche.org/2003/04/11/prevayling_stupidity/`, april 2003.

[Nick Rozanski 06] Eoin Woods Nick Rozanski. Software system architecture. Addison Wesley, 2006.

[Prevayler 08] Prevayler. *Prevayler Home Page.* `http://www.prevayler.org/wiki/`, march 2008.

[Wuestefeld 01] Klaus Wuestefeld. *Object Prevalence.* `http://www.advogato.org/article/398.html`, december 2001.