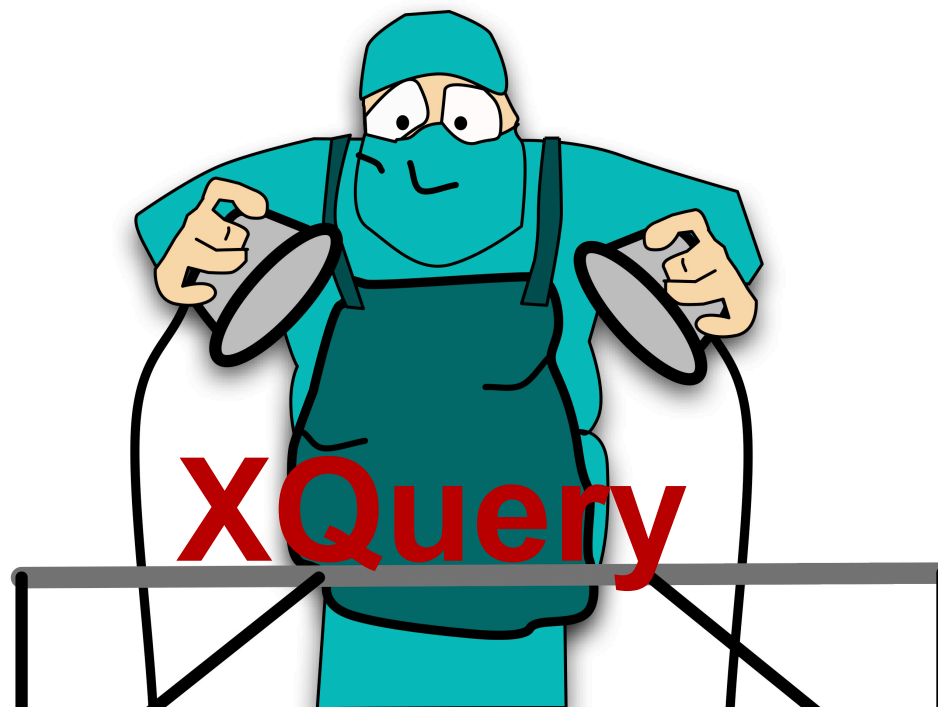
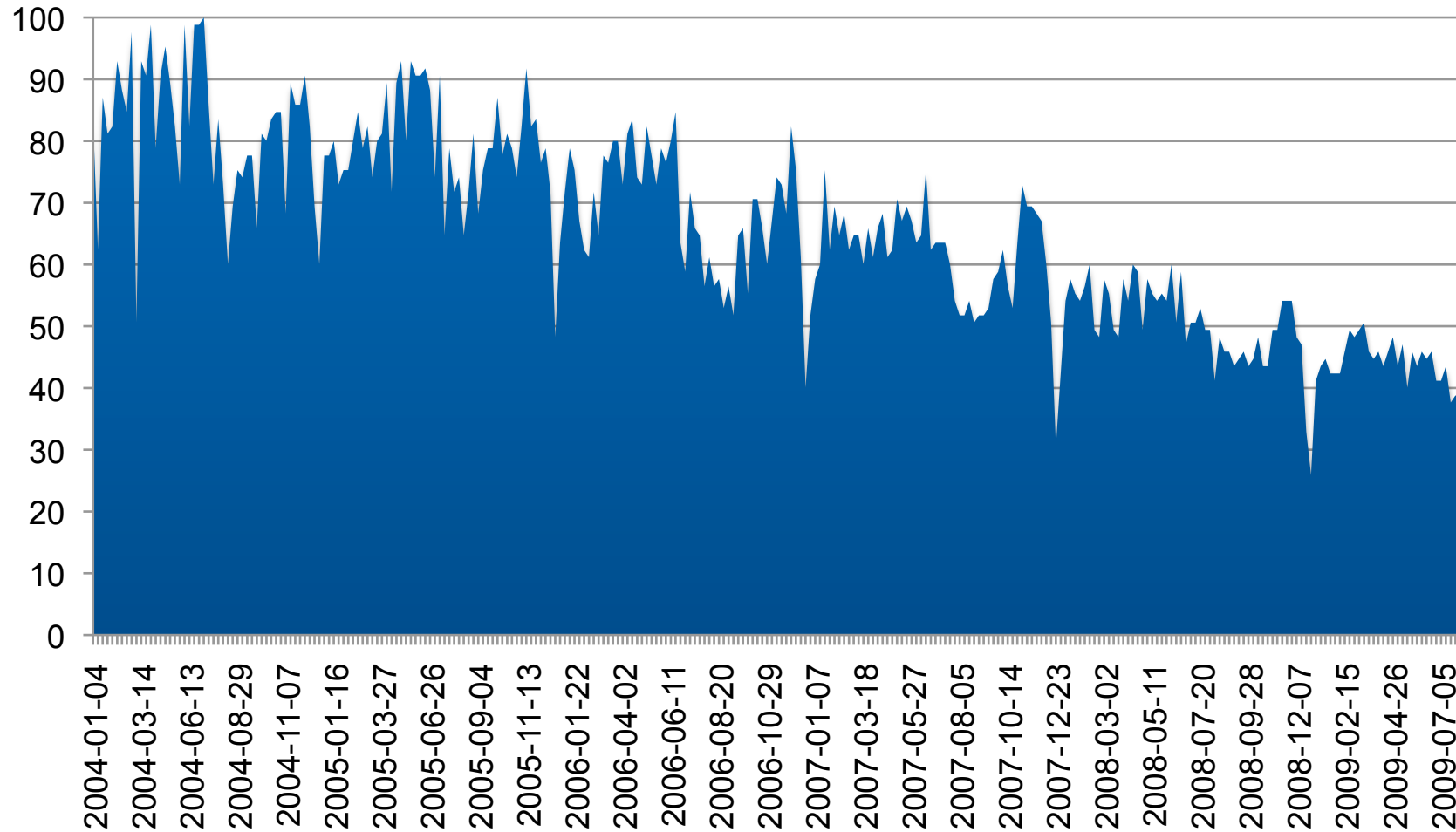


XQuery Reloaded

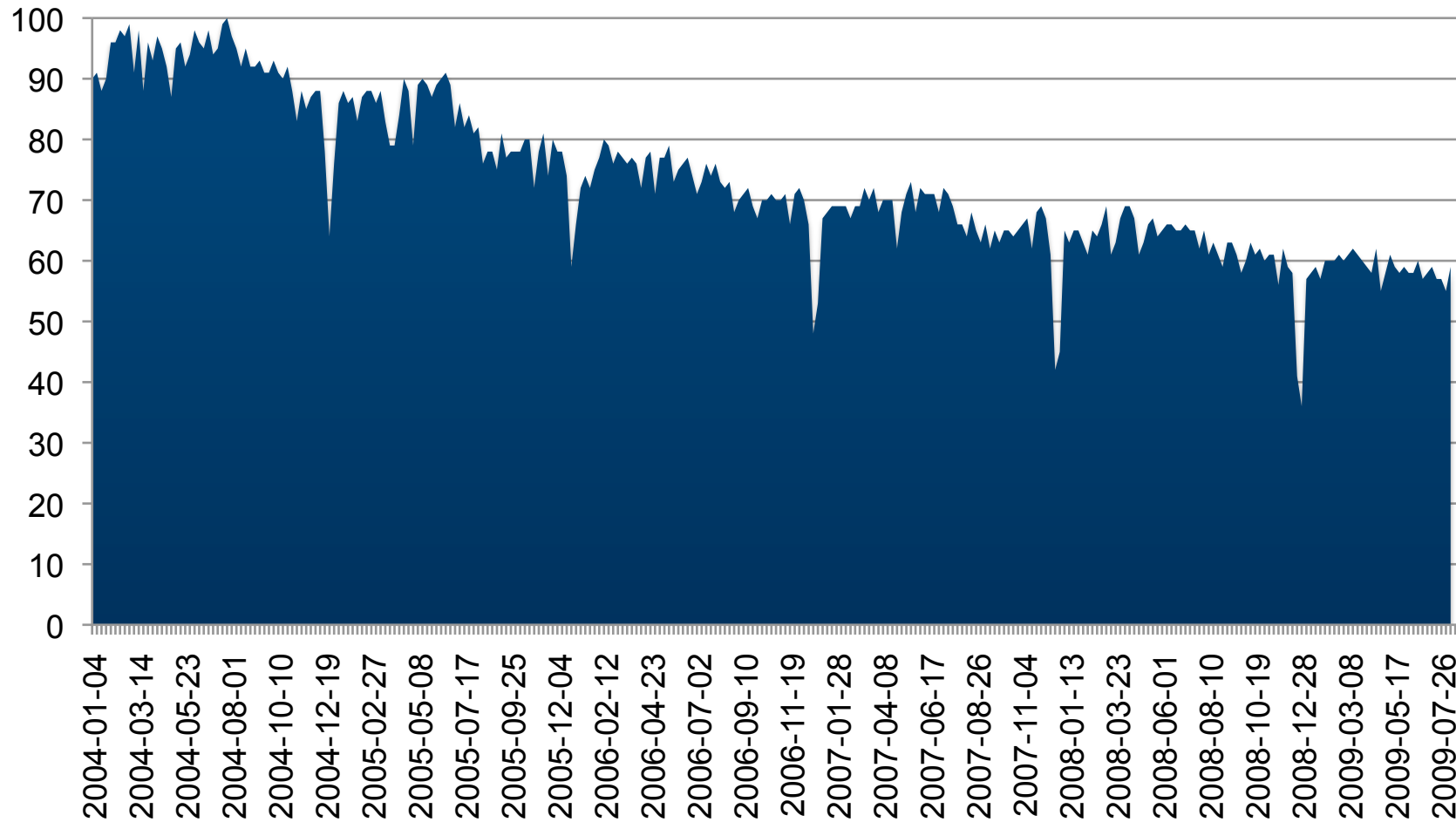


Roger Bamford, Vinayak Borkar, Matthias Brantner, Peter M. Fischer,
Daniela Florescu, David Graf, Donald Kossmann, Tim Kraska,
Dan Muresan, Sorin Nasoi, Markos Zacharioudakis

Google searches for “XQuery” (normalized)



Google searches for “SQL” (normalized)



XQuery folklore

- XML and XQuery are slow
- XQuery is complicated
- Legacy of XML, Namespaces, Schema, Xpath
- Bad products
- No people
- ...

XQuery folklore

- XML and XQuery are slow
 - partly true, but products are catching up
 - highly optimizable (like SQL; better than [uncz06])
- XQuery is complicated
 - is skiing more complicated than snowboarding?
 - try to process Web Services Feeds with Java!
- Legacy of XML namespaces, Schema, XPath
 - yes, but no alternative (relevance!)
- Big products
 - huge investments: research projects, big players
- No people
 - courses at all top places (ETH, Stanford, etc.)

Some true, some not true, all solvable

[<http://www.ibm.com/developerworks/xml/library/x-xqmyth.html>]

This talk

Clear your mind and get rid of prejudices

- Explain XQuery in a nutshell
- XQuery today and tomorrow
- Introduce Zorba – the MySQL for XQuery
- Show some fancy usages of Zorba

*Not in this talk: How to design an XQuery processor.
If you are interested in that, talk to one of us after the talk*

The roots of XQuery...

- Origins go back to the QL workshop 1998
- Standardized by the W3C (recommendation since 2007)
- Started out as a query language
- Closed data model, composition
 - results of expressions can be input for expressions
- Compliant with other standards
 - XML, XML Schema, XPath, Web Services, ...
- Example query:

```
for $empl in //employees
let $name := $emp/name
where $x/salary > 5000
order by $name
return $name
```

Today, XQuery is much more....

XQuery = Query + Update + Fulltext +
Scripting + Streaming + Libraries

■ X

- XQuery is the only language for XML, but that does not mean that XML is all it can do
- CSV, JSON, HTML, ...
- spectrum: structured data to unstructured text

■ Query

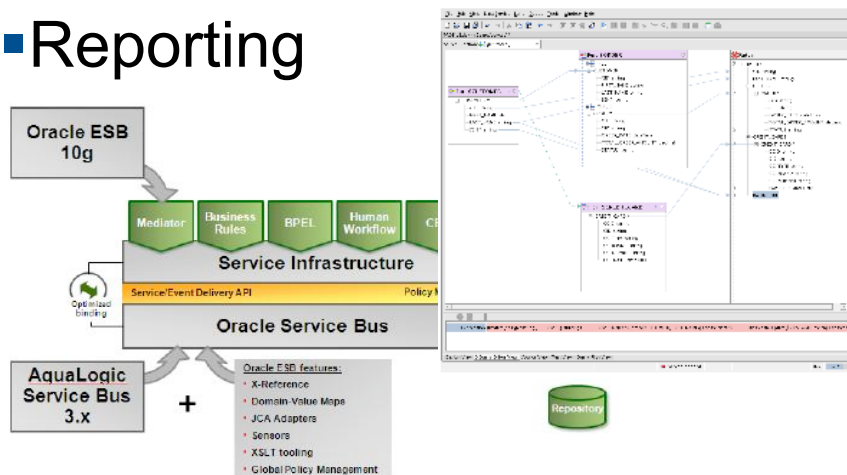
- XQuery has joins, group-by, sorting, etc., but that does not mean that it is only good for the DB
- by now, full-fledged programming language
- modules for structured programming

→ The name „XQuery“ is a disnomer!!!

XQuery is alive *...why else should so many people from different companies care to have his/her name on the paper*

Most successful in the **middle-tier**

- Data integration
- Configuration
- Reporting



but also in the **database world**

ORACLE

Microsoft

IBM

(Oracle has 8000+ customers reporting XQuery bugs)

XQuery in the future: Gartner's Top Ten disruptive technologies for 2008 to 2012

- Cloud computing and cloud/web platforms
- Multi-core and hybrid processors
- Virtualisation and fabric computing
- Social networks and social software
- Web mashups
- User interface
- Ubiquitous computing
- Contextual computing
- Augmented reality
- Semantics

What is needed:

A programming language for the Web

- **Machine-to-Machine Communication** between/inside distributed systems across company boundaries
- **Machine-to-Human Communication** over the browser, event-based interaction
- **Variety of workloads:** Updates, OLTP/OLAP/Streaming queries, structured and semi-structured data, varying and evolving over time



IMHO **XQuery** is the best starting point

*XQuery is not perfect,
but solves many of the problems*

XQuery in the future: Gartner's Top Ten disruptive technologies for 2008 to 2012

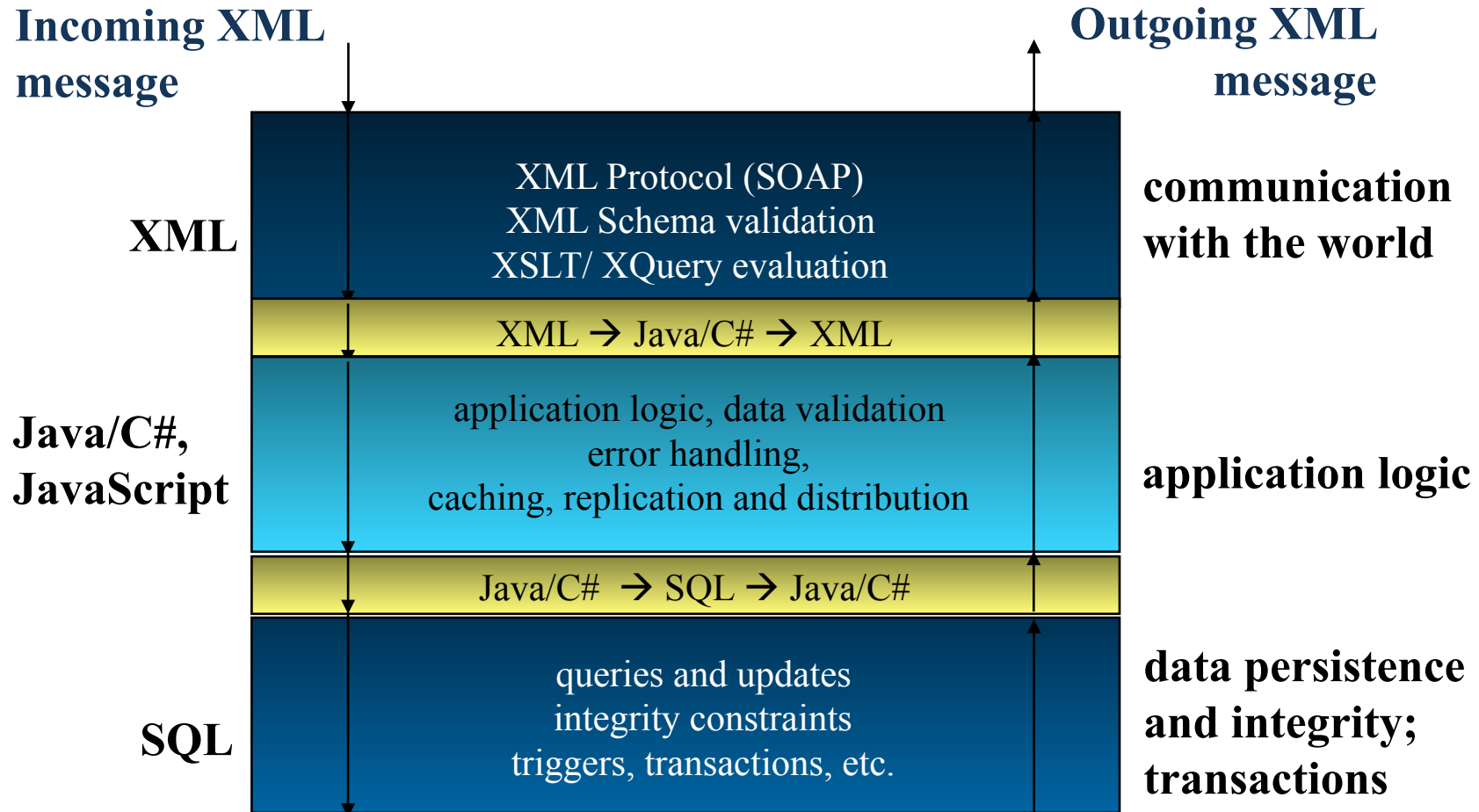
- **Cloud computing and cloud/web platforms**
- **Multi-core and hybrid processors**
- Virtualisation and fabric computing
- Social networks and social software
- **Web mashups**
- User interface
- Ubiquitous computing
- Contextual computing
- Augmented reality
- Semantics

XQuery mashups

„Gartner predicts that web mashups, which mix content from publicly available sources, will be the dominant model (80 percent) for the creation of new enterprise applications.“

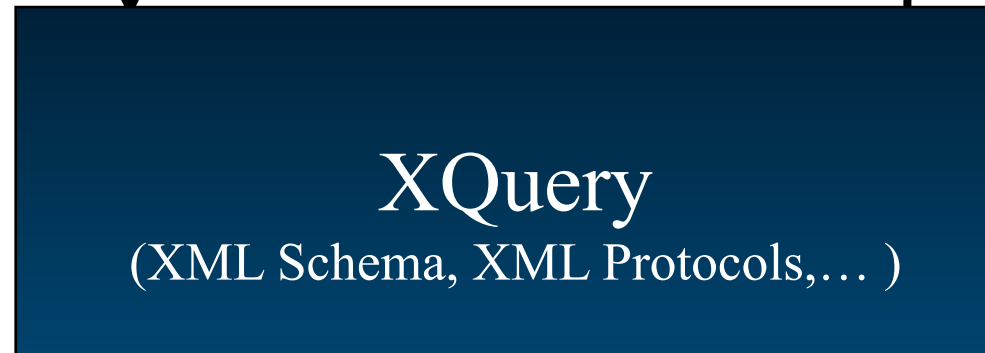
- Requires to integrate different source
- XQuery is made for the web
 - Works natively with XML and JSON
 - Has native support for REST and WebServices
- XQuery is a programming language

XQuery and cloud computing/Web platforms



XQuery and cloud computing/Web platforms

Incoming XML
message



Outgoing XML
message

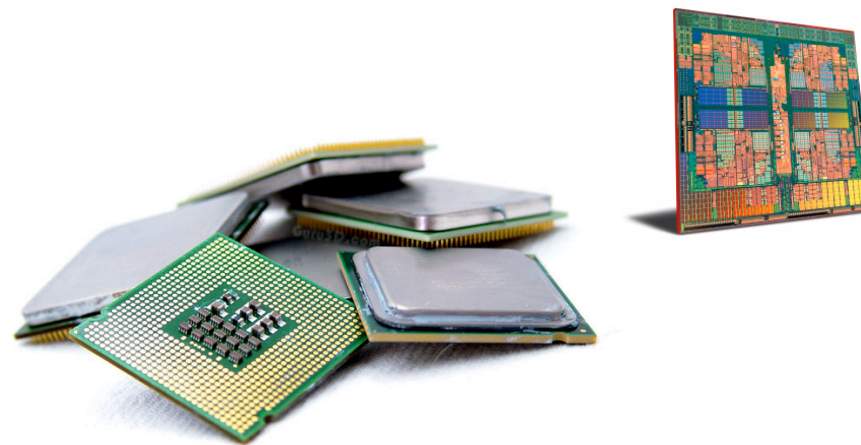


**communication
with the world,
application logic,
data persistence
and integrity;
transactions**

- No impedance mismatch
- Reduce the numbers of “hops”
- XQuery made for the web-standards

XQuery and new hardware / multi-cores

- Requires a highly parallelizable programming language
- Declarative (functional) programming model
- XQuery is highly optimizable and parallelizable
(as well as SQL is highly optimizable and parallelizable)
- Made for bulk processing



If you do not buy all that: Still, XML is out there and it will not disappear

- XML is best choice for *communication data*
 - general: web services (SOAP, WSDL), REST, RSS
 - specific: XBRL, HL7, ebXML, RosettaNet, ...
- XML is best choice for *meta-data and code*
 - configuration files, XMI (Eclipse), XForms (apps), XMP (photography), XAML / InfoPath (UIs), ...
- XML is best choice for *documents*
 - XHTML, SVG, OpenXML (Office), UBL (business)

XQuery is the cheapest way to process XML (and JSON)

What is missing is a MySQL for XQuery

Zorba

- Intended to be the MySQL for XQuery
- Developed by the **FLWOR Foundation**
- A team of 5 full-time programmers and several volunteers
- Contributing companies/organisations

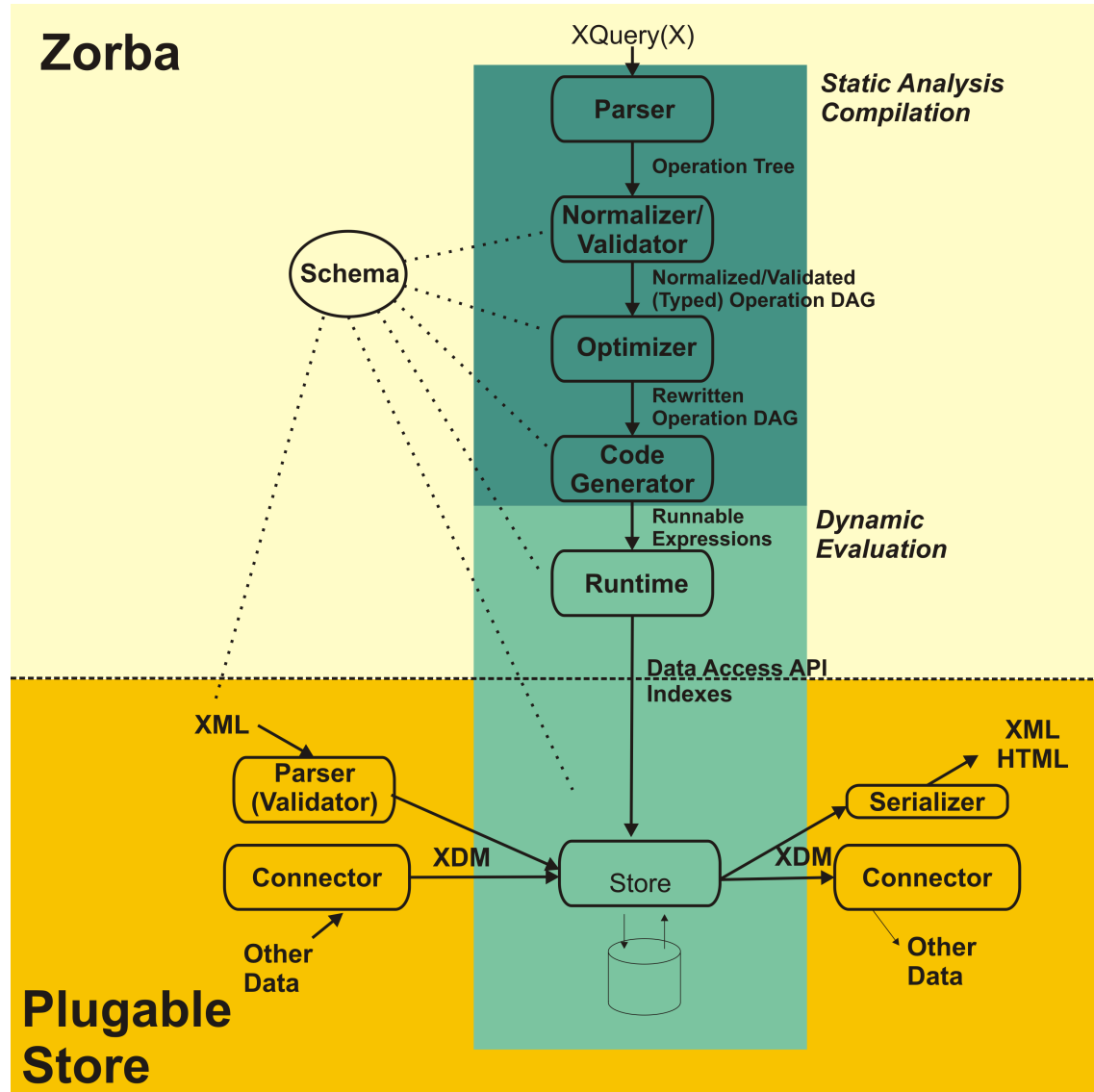
**ETH**Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

<http://www.zorba-xquery.com/>

What makes Zorba different?

- Not a “research” project!!!
- Open-source “Apache License”
- Zorba is a query processor like MySQL
- Exchangeable store
 - comes with an in-memory store
 - feel free to implement other stores (like 28msec)
- Zorba is designed for compliance first!
- Feature complete (except for full-text)
 - full-standard and not just the researchy features
- C++, interfaces to all main languages (Java, PHP, ...)

Zorba Architecture

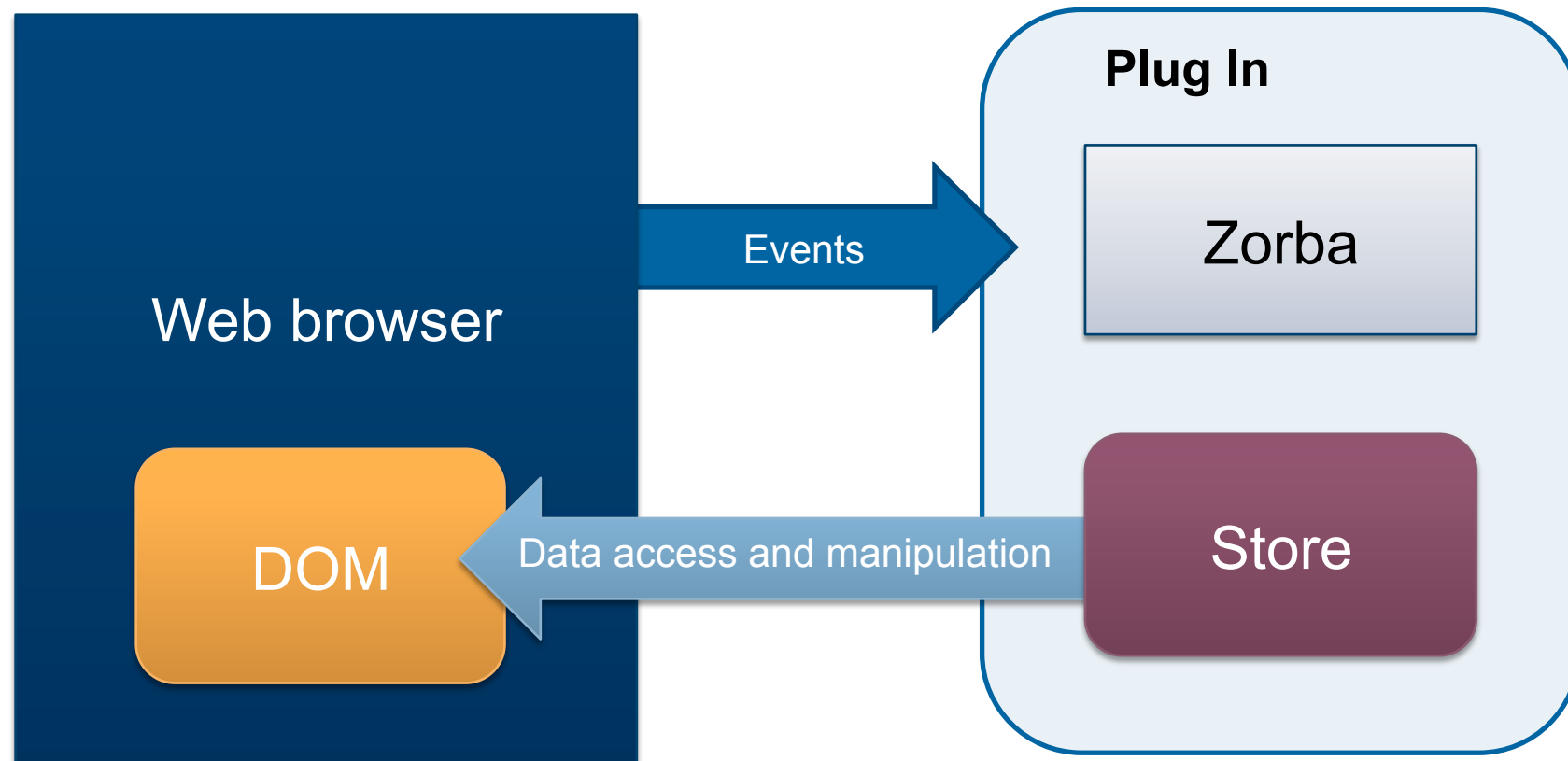


Zorba libraries/features

- XQuery 1.1, Update, Scripting
- Debugger
- Eval
- JSON
- REST/HTTP
- XQDoc
- PDF
- E-Mail
- Collections
- Tidy

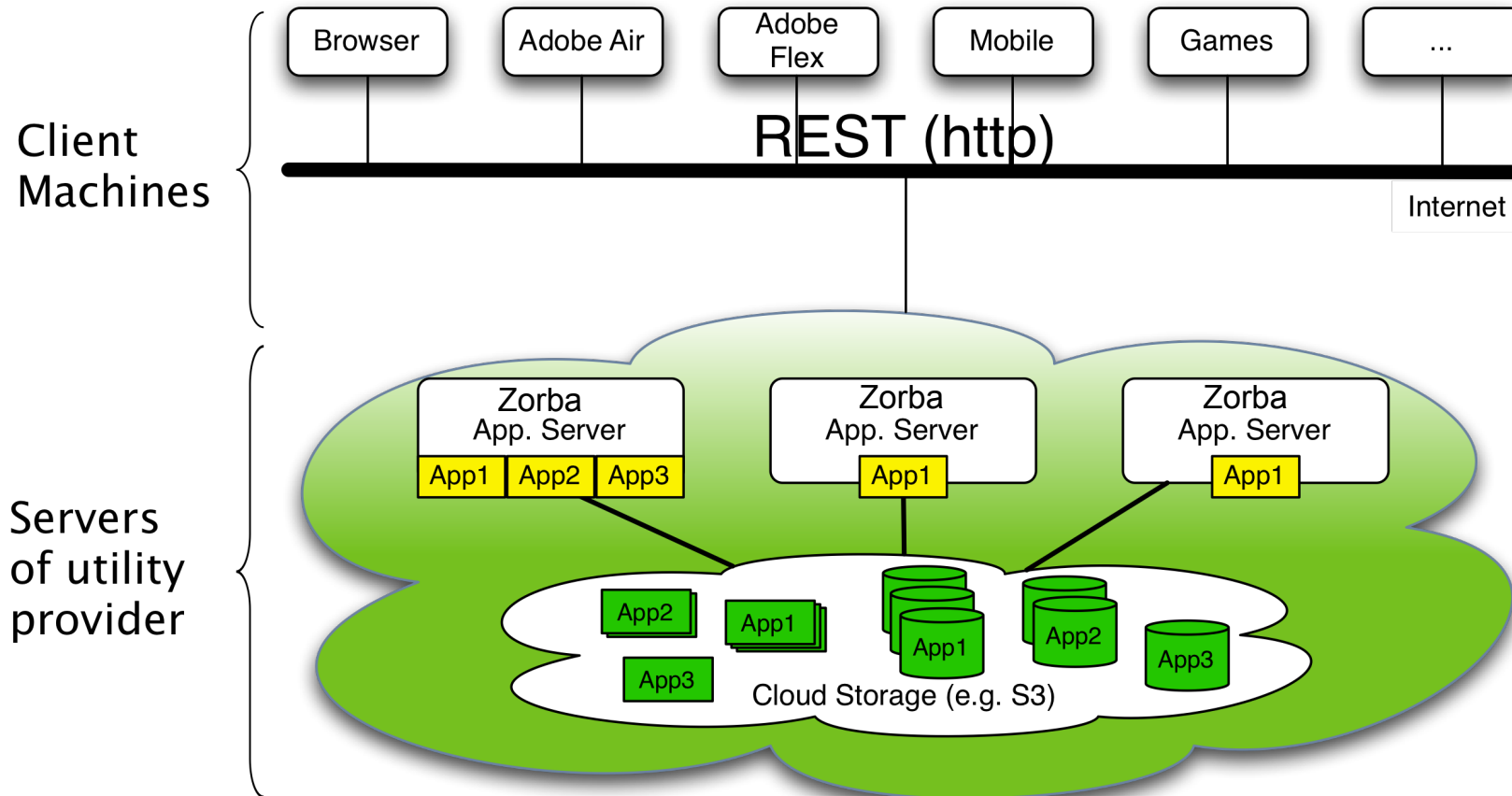
....

Cool stuff with Zorba: XQuery in the browser



<http://www.xqib.org>

Cool stuff with Zorba: 28msec WebServer



<http://www.28msec.com>

Cool stuff with Zorba: Xadoop

- Hadoop + Zorba
- Position XQuery as an alternative to Pig/Hive/etc.
- Well-suited for semi-structured data
- Use cases
 - pre-processing
 - financial data
 - analysis of blogs/wikipedia etc.
 -
- Pre-packed on EC2

So far, just a toy project with open ending...

XQuery Benchmarking Service

<http://xqbench.org>

- Supports several benchmarks
 - XMark
 - TPC-W like scenario
- Automatic test infrastructure
- Browse results
- **Allows to post: Your queries + Your documents**

Main goal: put pressure on the vendors and give use cases for the open-source community

Conclusion

- Zorba as the MySQL for XQuery
- XQuery combines a database and a programming language
- XQuery is ready for the next step

FLWOR Foundation: <http://www.flworfound.org>

Zorba XQuery Processor: <http://www.zorba-xquery.com>

XQuery in the Browser: <http://www.xqib.org>

XQuery Benchmarking Service: <http://www.xqbench.org>

XQuery Eclipse Plug-In: <http://www.xqdt.org>

Xadoop (not yet online): <http://www.xadoop.org/>