The Internet of Things: The Death of a Traditional Database?

Keith G. Jeffery

Director IT & International Strategy, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Harwell Science and Innovation Campus, Chilton, Didcot, Oxfordshire OX11 0QX UK

Abstract

Traditional database research has developed technology to ensure that the database — even when distributed — represents the world of interest with integrity and a consistent state. Important concepts have been developed and proven. However, the internet of things challenges all this. Very large numbers of nodes handle volumes that are vast, the speed is fast and the data/information space is global — indeed with space data — universal. This poses challenges. What does the concept of a state mean when the information map of the real world of interest is represented across millions of nodes, many of which are updating in real-time? What does a transaction look like when the data being updated is spread across hundreds or thousands of nodes with differing update policies? Worse, how does one roll back or compensate a transaction? We have already seen database research applied to semi-structured data, to streamed data, and real-time applications. Is it possible for these techniques to be applied to the internet of things? The internet of things opens up more opportunities for security compromises. How do we develop trust band security techniques across multiple policies? How do we prevent the unauthorized use of private information yet permit authorized use? We need dynamic trust, security, and privacy management. Do we need a new theoretical framework?

Keywords

Database, Future internet, Integrity, Process, State, Transaction, Workflow.

1. Introduction

There is much activity in Europe and the world on predicting the future of information and communication technology (ICT). There are roadmapping exercises for R and D in various domains to meet that predicted future. The EC has set up expert groups and/or projects covering GRIDs, CLOUDs, Service-Oriented Architectures, quantum and bio-computing, new materials, human-computer interaction, and cognitive technology among others. There is much discussion of Web2.0 and beyond. The 'Internet of Things' (http://en.wikipedia.org/wiki/Internet_of_Things) is a strong theme with a recent EC (European Commission) conference (May 2009) dedicated to it. The formation of the FIA (Future Internet Assembly) underpins the groundswell of enthusiasm for this idea, and Issue 77 of ERCIM News [1] has Future Internet Technology as the special theme, with a foreword by Viviane Reding, EC Commissioner for Information Society and Media, emphasizing the importance. Europe is establishing an e-Infrastructure and the US is establishing its Cyberinfrastructure.

Database researchers (with a few notable exceptions) have not been very prominent in these discussions. This is surprising, as the movement toward take-up of these new technologies by the business world pioneered in the research field will require, at the least, interoperation with the existing database technology, and most likely a further wholesale evolutionary or revolutionary development of the database technology, to adapt to the new environment. Database research has moved to include semi-structured data and its processing and managing of data streams. There is work on schema matching and mapping for interoperation (sometimes in the context of Dataspaces), and on domain ontologies. There is still ongoing work on web-database interfaces, modeling, and systems development. Work on performance or query optimization with new algorithms continues, as does optimized storage architecture—including P2P (Peer to Peer).

Where are the advances in database research matching—and/or contributing to — the huge advances in (among others) social networking, content creation and repurposing, gaming, sensor systems, robotics, autonomic systems, visualization, user interaction, systems and software development, and service-oriented architecture?

2. A Vision

The vision has its roots in [2] with subsequent refinements [3,4] leading to an analysis and synthesis performed in 2008 and updated in 2009 by ERCIM (www. ercim org). It is based on the architecture proposed for the UK e-Science program [2] and is represented in Figure 1.

Let us imagine a possible state in 20 years' time. The problems facing Europe — and the world — (from continent through country to individual person scale) are large, complex, and require unprecedented scientific, mathematical, and IT skills for their solution.

There is a fast, reliable, inexpensive e-infrastructure providing all communication services. Persons are connected to the e-infrastructure via personal computer devices that are continuously online. The networking components of the e-infrastructure invisibly provide optimal connectivity in terms of performance, reliability, cost, and security. The e-infrastructure physically senses, detects, records, and curates everything, using all the computers, storage devices, networks, and sensors. Subject to security, privacy, ownership and commercial rights all computational, storage, detector, and communication facilities are available to everyone. Detectors and subsystems will occur in all environments, across all industries and social services, as also in the home environment. Subsystems are embedded within the e-infrastructure — for example control systems for utilities — including personaltransport. Other subsystems will be robotic for agriculture, manufacturing, healthcare, and other applications. This e-infrastructure vision has major implications:

1. There is a continuing and accelerating need for ever faster, smaller, cheaper, and more energy-efficient (and less heat-producing) devices. At some point biologically-inspired systems will dominate and will compete/cooperate with quantum-based technologies.
2. New 'intelligent materials' will be developed, which will allow artifacts to be constructed 'internet-ready'. These will range from agricultural products through to manufactured products.
3. The open availability of everything simplifies the physical access and improves the performance, including reducing latency, but will demand ever-increasing performance, scalability, reliability, and self-management.
4. The middleware of the e-infrastructure bears heavy responsibilities: (a) for providing the self-* characteristics (self-managing, self-tuning, self-repairing) of a reliable e-infrastructure; (b) for identification, authorization, trust, security, privacy, and access control; (c) for hiding the complexity through virtualization and abstraction, thus providing homogeneous access to and utilization of heterogeneous facilities.

The i-infrastructure relies on the underlying e-infrastructure and converts the data (structured, semi-structured, and unstructured) to information. The i-infrastructure provides the processing capabilities to collect, structure, manage, describe, and manipulate the information. It provides computational modeling/ simulation facilities to generate new information. The processing capabilities will be Service-Oriented Knowledge Utilities (SOKUs) which are discoverable/ composable and dynamically tunable, based on properties described by their metadata. There is a massive amount of content: From structured verified data and information through to personally authored social networking artifacts, and from data streams generated by detectors through to entertainment and education material. The volumes of data and information will preclude shipping data to processors with appropriate software; rather we shall need to ship software to the data.

The k-Infrastructure manages knowledge; allowing differing semantic descriptions over a formal syntax in the i-layer. This is the domain where humans or

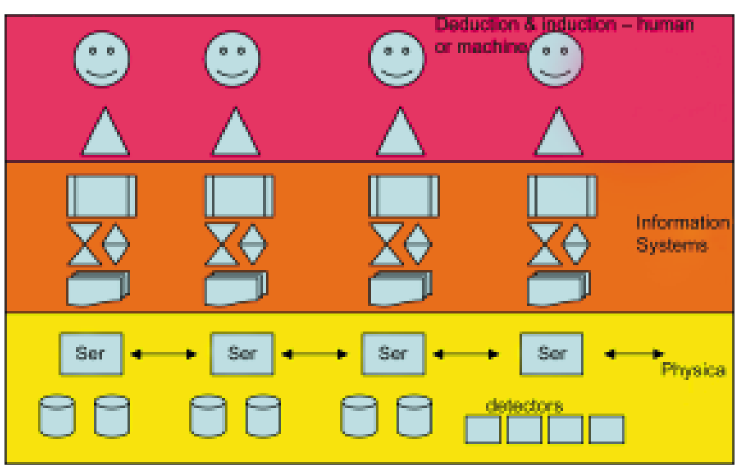

Figure 1: 3-Layer architecture.

data mining extract knowledge from information by deduction or induction, where that knowledge is codified and stored for use in optimizing the e- and i- layers, and for interfacing to intelligent applications and intelligent user interfaces in the overlying application layer.

Relying on this e-, i-, and k-infrastructure are applications. They also will be constructed from SOKUs. The SOKUs will have functional characteristics and their nonfunctional characteristics (including performance, security, and use-condition aspects) will be determined by a well-defined interface to the e-infrastructure. Such architecture allows extensive re-use of well-tried components and the rapid development of applications, using them and additional new services specifically for a particular application. The applications will range from games and edutainment through to B2C (Business-to-customer) and B2B (business-to-business) transactions within an E2E (enterprise-to-enterprise) environment and on to advanced R and D activities. Decision-making will be based not just on structured information and knowledge induction and deduction utilizing information, but also on simulations. These applications will be available (under appropriate conditions determined by the restrictive metadata) to everyone. Some applications will be general and widely applicable — ranging from entertainment and games through cooperative working/socializing to information management and analysis. These are likely to be pre-composed and optimized for efficiency. Some applications will be highly specialized for particular industrial/commercial sectors or for social sectors such as healthcare and environment; these will be constructed dynamically at demand-time.

The end-user will interact with the applications via a set of personalized devices — including robots — providing services. Each device-based service will have associated role-based profiles (metadata covering mainly nonfunctional requirements) to interact with the e-infrastructure. This provides the context for user-application interactions mediated by SOKU agents. The end-user device services will be 'intelligent' and will 'learn' from experience. They will act on behalf of the user in a majority of the cases. The end-user will not know (or care) where and how her requirements are met, as long as the agreed service levels are achieved. The use of service level agreements negotiated by agents on behalf of the user, their enforcement, and dealing with dissatisfaction and recompense will raise new challenges.

With everyone and everything available (subject to policy restrictions encoded as restrictive metadata) operations like scheduling meetings, travel planning, project management (milestones, deliverables, resources), resource control, prediction and planning, cooperative working, workflow task management, and information distribution will all be integrated seamlessly as seen by the end-user, whether professional or social. In the domestic environment, home management systems — including robots — will interact with the external systems for the utilities including transport, supplies of food, consumer goods, entertainment, information (news, weather, socioeconomic parameters) and education, and providing the user with options for choice.

It should be noted that the concept of complete virtualization of ICT services is taken to its logical (or extreme?) conclusion in the concept of Cloud computing, where some or all of the ICT services for an organization are outsourced to a large systems supplier. The offering from the Cloud supplier may be IaaS (Infrastructure as a Service), PaaS (Platform as a Service), AaaS (Applications as a Service), or EaaS (Enterprise as a Service).

3. The R and D Required

From the above it is clear that there are several key areas of computer science/information systems engineering that relate to database technology and need to be addressed by R and D. All of them involve metadata [5]. These are:

1. SOKU engineering is based on metadata, with formal syntax and declared semantics, to describe intelligent software services that are mobile, so they can be shipped to any processing node.
   a) Schema metadata for integrity control.
   b) Navigational metadata for location.
   c) Supportive metadata to assist in discovery, composition, and execution.
      i) Descriptive metadata of the functional aspects for discovery, composition, execution, and monitoring.
      ii) Restrictive metadata covering the nonfunctional aspects, for discovery, composition, execution, and monitoring.
      iii) Supportive metadata is related to a whole domain, to assist (by semantic interpretation) in discovery, composition, execution, and monitoring.

2. Data, information, knowledge engineering over structured, semi-structured, and unstructured data, with temporal and certainty/probability properties based on metadata — with formal syntax and declared semantics — to describe sources.
   a) Schema metadata for integrity control.
   b) Navigational metadata for location.
   c) Supportive metadata to assist in discovery, binding, and execution.

i) Descriptive metadata of the functional aspects for discovery, binding, execution, and monitoring.
ii) Restrictive metadata covering the nonfunctional aspects for discovery, binding, execution, and monitoring.
iii) Supportive metadata related to a whole domain, to assist (by semantic interpretation) in discovery, binding, execution, and monitoring.

3. Agent (representing a person in role — a special kind of SOKU) management based on metadata — with formal syntax and declared semantics — to describe the sources.
   a) Schema metadata for integrity control.
   b) Navigational metadata for location.
   c) Supportive metadata to assist in discovery, binding, and execution.
      i) Descriptive metadata of the functional aspects for discovery, binding, execution, and monitoring.
      ii) Restrictive metadata covering the non-functional aspects for discovery, binding, execution, and monitoring.
      iii) Supportive metadata related to a whole domain, to assist (by semantic interpretation) in discovery, binding, execution, and monitoring.

From the point of view of database research this means that the traditional database management systems and applications based upon them will be replaced. The replacement will be the composition/orchestration of appropriate SOKUs, initiated after intelligent dialogue with the end-user SOKU agent. This agent acts on the appropriate data and information to produce and utilize knowledge, all within the defined service levels, trust/security/privacy and rights. It is possible that a composed service will offer a complete encapsulated solution across the e-infrastructure, i-infrastructure, and k-infrastructure produced in response to the agent-improved user request and executed in a self-* environment, to assure service and quality levels (including price if appropriate) as required by the end-user.

4. The Research Challenges

To achieve such a flexible, effective, efficient, and economic environment, various research challenges are outstanding. There are many research challenges relating to computer and network architecture, performance optimization, self-* middleware, and novel operating system architecture, among others. Considering only the database technology, the major ones are:

4.1 Metadata

a) The need for metadata related to services, data/information/knowledge, agents.
b) What is data, what is metadata?
c) Types of metadata and their use.
d) Representation and structure - syntax.
e) Semantics (meaning).

4.2 Management of State

a) Detection of state across millions of individual nodes.
b) Maintenance of state across many nodes.
   i) Transactions and locking.
   ii) Roll-back and compensation.

4.3 Data Representativity

a) Data structures representing real-world inter-relation ships.
   i) Data attribute value encoding (character set, media encoding), types, lengths.
   ii) Data attribute value language.
   iii) Fully connected graphs — the death of the hierarchy.
   iv) The time-machine: Temporal duration of the inter-relationships.
   v) Certainty, probability of the inter-relationships.
b) Interoperation — reconciliation of different data structures representing a similar real-world domain.
c) Data location/locality and replication for business continuity.

4.4 Data Quality, Veracity, and Permanency

a) Detection of quality against metadata parameters, for example, precision, accuracy.
b) Provenance.
c) Temporal recording.
d) Data curation across media and policy evolution.

4.5 Trust, Security, and Privacy

a) Policies declared, enforced, and monitored through restrictive metadata.
b) Policy reconciliation for interoperation.

4.6 Management of Service Levels and Quality of Service

a) Policies declared, enforced, and monitored through restrictive metadata.
b) Service level negotiation (e.g., lower price for lower performance).

4.7 Systems Design, Development, Maintenance, and Decommissioning

a) Based on strong separation of:
   i) Services (processes).
   ii) Data, information, and knowledge.
   iii) Agents.
b) Assuming self-composition, self-managing and adjusting, and self-maintaining properties.
c) Assuming mobile code properties.

All of these contain fascinating and difficult R and D challenges. Number 2 challenges the conventional database to the core. Number 3 demands a different approach to data modeling. Number 7 precludes object-orientation. Space does not permit consideration of each one in detail, therefore, two are chosen to illustrate the challenge and these are now considered.

5. Management of State

In a conventional database, the management system state is maintained in near-real time by ACID (Atomicity, Consistency, Isolation, Durability) transactions, which lock areas of the database to prevent update conflicts. As transactions become more complex in such an environment (more instructions related to the update, more database Tables affected) the duration of locking — with subsequent performance degradation — becomes unacceptable, and hence, other techniques are used. Roll-back allows for the possibility that a conflict occurs leaving the database in an inconsistent state and the roll-back reschedules the updates to run sequentially. Compensation initiates a transaction that restores the database to a correct state.

In a distributed database environment a transaction may affect Tables geographically remote from each other with communication latency. Two-phase commit protocols are designed to ensure a consistent state across the distributed database — even if this state no longer corresponds with the real world. Similarly compensating transactions can restore a distributed database to an internally-consistent state, which may not represent the real world. Use of such transaction controls typically takes from microseconds to minutes. In an environment when millions of nodes are self-updating locally from sensors (including audio, video) and human input, it is clear that conventional database technology does not work. The best we can achieve is the local state internally consistent and consistent with (a small slice of) the world of interest, and reconciliation across the distributed database as and when required using lazy techniques, unless of course we re-think on the notions of state and transactions.

6. Data Representativity

The problem is to represent the real world inside the database system, so that changes caused by events triggered by external factors — humans or machines in the real world — are reflected in the database. The database has to represent values of attributes of entities. However, these attributes can range from simple characters (even then there is the problem of representation in legacy systems, although today solved by Unicode) through numbers with precision and accuracy to a whole text or multimedia objects or even a binary representation, impenetrable, without more knowledge of the representation. It is thus necessary to describe the attributes by descriptive metadata, ensure integrity using schema metadata, locate it by using navigational metadata, and restrict usage as appropriate, using the restrictive metadata. The structure of the information constructed from the data is a key aspect. Commonly humans structure 'things' into hierarchies — for example groups within departments within faculties or schools within a university. However, few universities have such a structure and groups may 'belong' to >1 department, faculties or schools may not exist, and research centers may either exist independently of any department or may be 'owned' by several. This would indicate that hierarchies do not represent the real world, and in fact the relationships between hierarchic levels in a data structure have semantics, temporal duration, and probability — the latter is especially important when dealing with incomplete and inconsistent information.

This should not surprise us, as we are all familiar with the non-hierarchic 'two parents one child' graph (which has semantics, temporal duration, and probability) and the generalization to many-to-many cardinality. Another example is the classification of species and subspecies: Simple hierarchies do not match the real world state and again relationships have semantics, temporal duration, and probability. Much of the problem with relational database implementations is, people try to force a real world into a hierarchic representation (using primary and foreign keys), whereas, using base Tables and 'linking relations' allows a natural n:m relationship (with semantics, temporal, and probability attributes) to be expressed. One data model with these properties for the domain of CRIS (Current Research Information Systems) is CERIF (www.eurocris.org/cerif). In fact this data model has been used in conventional organizational settings as well. As a by-product such information structure maps go directly to hypermedia structures such as WWW.

Such a structure has many advantages, for example, it is common in some applications for the base Tables to be relatively static (append not update), whereas, the 'linking relations' are frequently updated — indicating rapid changes in inter-relationships of 'things'. Is this not exactly what is required for the 'internet of things?' Alternatively, the base Tables may be updated frequently, whereas, the 'linking relations' may remain static — examples include detector arrays where streams of data are collected according to a plan — and the plan is encapsulated in the structural relationships between the data streams — at least for a given temporal duration.

Such a representation allows the 'time machine' to operate; the state can be recreated at any time-slice by retrieval across the 'linking relations'. How many times do we wish to know the 'state of the world of interest at a given time or over a given interval? As a side effect, storing the temporal information in 'linking relations' is much less expensive than storing temporal information with base attributes, as in the conventional temporal relational model [6], and allows standard SQL processing without recourse to the temporally-extended TSQL. As a by-product such a representation assists greatly with provenance.

Such a representation allows for probability on the relations between 'things' (entities or objects) to be expressed within the 'linking relations'. This is particularly helpful in many applications where relationships are deduced by humans, inferred by discussion, or are speculative in a scientific research environment. Of course the 'intelligence' of processing the probabilities depends on the capabilities of the system, from simple relational calculus through to full fuzzy logic capabilities.

Interoperation — usually to provide a homogeneous view of information to an end-user from heterogeneous sources — clearly requires rich metadata about the attributes and syntax (structure), to be able to match schemas (and more detailed metadata) and to map them to each other. Invariably, additional knowledge processing is required, and this may be supplied by humans, but progressively more of this task is done by computer systems, using a variety of techniques including graph theory (matching structures), lexical matching with thesauri, and knowledge-based reasoning (using domain ontologies related to the attributes and syntax of the domain of discourse). The current 'half-way-house' of part-machine, part-human reconciliation of databases is the research area commonly referred to as 'dataspaces'. To improve the machine support of this process it is thus necessary to describe the attributes by descriptive metadata, ensure the integrity using schema metadata, locate it using navigational metadata, and restrict the usage as appropriate, using restrictive metadata — exactly as stated in the first paragraph of this section.

7. Conclusion

The database research community has in the past made great advances with major results (e.g., relational database technology — although the pivotal research was done 40 years ago) being taken up and developed further by the IT industry and their products used throughout commerce and industry generating wealth and value. Similarly the technology has underpinned advances in domains such as, environmental monitoring, healthcare, and education, generating improvements in the quality of life.

Dave deWitt [7] proposed that database technology was a 'roadkill on the information superhighway' and the subsequent development of web-based systems — although backended by database technology — has in some ways proved him right.

The 'internet of things' provides database research with a new opportunity and new challenges. Conceptually the scale and complexity are almost overwhelming and architecturally, are actually overwhelming. The speed and required low latency for many applications are beyond the current capabilities and demand architecture beyond the current database server clusters and distributed databases. Although database technology research has addressed semi-structured and multimedia information and datastreams, homogeneous access to — and processing of — heterogeneous sources is not yet solved. There has been research on representation of temporal properties, but no generally accepted representation has emerged (despite standardization). There has been research on incomplete and uncertain information, but again no generally accepted consensus.

There are challenges (not dealt with above) in the representation of morality in the processing of information with appropriate privacy and security. A more widespread, open, and intelligent environment will produce new malware and malicious attacks; therefore, safeguards will have to be designed.

In the 'internet of things' can we — the database research community — continue to 'make do and mend' with conventional relational technology in a client-server environment or has the time come to find new architectural solutions to the challenges? Is the key to success in the (much improved) metadata of SOKUs, providing services over data/information sources and acting as agents?

References

1. http://ercim-news.ercim.org/content/blogcategory/102/777/.
2. K. Jeffery, "Knowledge, Information and Data" Paper submitted to Director General of Research Councils proposing the programme that became e-Science; Sep. 1999; available at http://www.semanticgrid.org/docs/KnowledgeInformationData/KnowledgeInformationData.html.
3. K. Jeffery, "GRIDs, Databases and Information Systems Engineering Research", in Advances in Database Technology - EDBT 2004 Mar. 2004 pp. 3-16.
4. K. Jeffery, "Systems Development in a GRIDs Environment" in Conceptual Modelling in Information Systems Engineering" Jul. 2007 pp. 279-94.
5. K. Jeffery, "Metadata": In Information Systems Engineering: State of the Art and Research Themes Jun. 2000.
6. R. Snodgrass, "Temporal Databases" in Theories and Methods of Spatio-Temporal Reasoning in Geographic Space, Springer LNCS vol. 639, Sep. 1992.
7. D. DeWitt. DBMS: Roadkill on the Information Superhighway? International Conference on Very Large Databases (VLDB 95), 1995, Invited Talk: Not published in conference proceedings.

AUTHOR

Keith Jeffery is currently Director IT and International Strategy, STFC (Science and Technology Facilities Council), based at Rutherford Appleton Laboratory in UK. Keith previously had operational responsibility for IT with 360,000 users, 1100 servers, and 140 staff. Keith holds three honorary visiting professorships, is a Fellow of the Geological Society of London and the British Computer Society, is a Chartered Engineer, a Chartered IT Professional, and an Honorary Fellow of the Irish Computer Society. Keith is currently president of ERCIM and euroCRIS, and serves on international expert groups, conference boards, and assessment panels.

E-mail: keith.jeffery@stfc.ac.uk

DOI: 10.4103/0256-4602.55272; Paper No TR 217_09;