Daisy: The Center for Data-intensive Systems at Aalborg University

Hua Lu Torben Bach Pedersen Simonas Šaltenis
Bent Thomsen Lone Leth Thomsen Kristian Torp
Daisy, Aalborg University
{luhua,tbp,simas,bt,lone,torp}@cs.aau.dk

1. INTRODUCTION

The history of the Center for Data-intensive Systems (Daisy) (daisy.aau.dk) at Aalborg University dates back to the late 1980'es where Christian S. Jensen was hired as a Ph.D. student and outposted to the University of Maryland. After having graduated in 1991, he was hired as an assistant professor and started to build up a database group. The main research topic was initially temporal data management [38, 65], including temporal data models, temporal query languages such as TSQL2, and temporal query processing techniques, in close collaboration with Richard T. Snodgrass at the University of Arizona. In the mid 1990'es, new faculty members and Ph.D. students joined, followed by accelerated growth towards the end of the 1990'es. During this period, two additional research topics emerged, namely data warehousing/multidimensional databases and spatio-temporal data management. In 2000, the database group was merged with the programming language group to form the Database and Programming Technologies (DPT) group. This paper focuses on the data management related research within DPT, excluding purely programming language oriented research, and is authored by the permanent faculty members involved in this research. During the next decade, the research agenda widened as the group grew to also cover mobile service infrastructure, business intelligence and data mining, multimedia data management, as well as general data management infrastructure. The Center for Data-intensive Systems (Daisy) was founded in 2007. After two decades at Aalborg University, Christian S. Jensen joined Aarhus University in 2010. Since then, the research has been led by Professor Torben Bach Pedersen. The 28 current Daisy members can be seen on the web page.

2. RESEARCH CONTEXT

Research Approach and Context. The overall research philosophy is based on three principles. First, we aim to perform *use inspired* research which is motivated by the (real or anticipated) practice and use of the resulting technologies. Doing this means interacting regularly with practitioners, e.g., users or other stakeholders, and using real-world data when available. We believe that doing so has a positive effect on the relevance and impact of our research results. Second, the aim is that research results should have *general applicability*. Thus, the results should be more abstract than just providing a solution to a single specific problem. The task of providing this will be done by practitioners themselves, ensuring a productive division of labor. Third, our research is inherently *constructive and experimental* in nature. Thus, we typically design technical artifacts and most often build more or less elaborate prototypes, which are used for experiments, the results of which guide subsequent improvement iterations.

The research philosophy also reflects the (funding- and otherwise) context that we operate in. Our group has a good number of full, associate, and assistant professors financed by the department. Post docs and Ph.D. students must however mostly be funded by external grants. The funding bodies at Danish and European level provide considerably more funding for application-oriented projects than for purely research topic focused projects. Thus, it is necessary to use the participation in application-oriented projects (which we do believe is good for the research) to produce interesting general results. Doing this is only possible if the deep domain knowledge required for the applications is leveraged over a number of projects, gaining synergy between the individual activities. To ensure this, we focus on a few advanced *key application areas*, namely mobile services, intelligent transport systems, energy, and logistics. As we will see next, most of our projects do indeed have the nature described above. Another characteristic of Aalborg University is the use of *Problem-Based Learning (PBL)*, where students work in groups guided by a supervisor for half of the time on all semesters. We thus try to utilize the student projects in our research and collaboration.

Recent and On-going Projects. The *MIRABEL* FP7 project (www.mirabel-project.eu) has introduced a data-driven approach to smart grids, based on the novel concept of a *flex-offer*, an atomic unit of intended electricity use and flexibility, e.g., "1.5 kwh for my dishwasher over a 2 hour period anytime between 8PM and 7AM." By capturing, aggregating, and scheduling these, a large part of the energy consumption (40% for households, growing to 80% when EVs and heat pumps are introduced) can be shifted in time and thus match much larger quantities of renewable energy from wind mills, solar panels, etc. Realizing this vision involves significant data management challenges in the areas of aggregation, near real-time DWs, tight integration with forecast data, large-scale distri-

buted data management, etc. [3].

The *Intelligent Sound* (www.intelligentsound.dk) project involved interdisciplinary collaboration with signal processing researchers. The project aimed to support advanced and efficient queries on music data by automatically extracting music metadata (melody, rhythm, timbre, etc.) directly from the sound signal. Modeling and querying frameworks, and effective indexing and query processing techniques, were developed, enabling efficient advanced querying (playlist generation, etc.) on even very large music databases.

The *Daisy Innovation* project (daisy.aau.dk/dain) funded by the European Regional Development Fund (ERDF) collaborates with companies in the North Jutland Region within data-intensive systems, through education activities, technical advice for companies, networks for knowledge exchange, and smaller, focused sub-projects on specific topics such as business intelligence, data mining (where results are already in TARGIT products), database testing, mobile services, and intelligent transport systems. The focus of the project is as much on innovation as it is on research, yet a number of papers resulted from the collaboration.

Emerging computing application areas, e.g., transportation, involve the monitoring of continuous variables which yields massive update loads that existing systems cannot contend with. Thus, the *Sensload* project (sensload.cs.aau.dk) explored techniques for selective shedding of updates and it developed main-memory-optimized data structures and algorithms that increase the update throughput of database systems.

The *BagTrack* project (daisy.aau.dk/bagtrack) funded by The Danish Advanced Technology Foundation aims to build a global IT solution that significantly improves the worldwide aviation baggage handling quality. Daisy develops data management techniques, specifically data cleansing, continuous query processing, data warehousing, and data mining, for massive amounts of baggage RFID data.

The *REDUCTION* (www.reduction-project.eu) FP7 project aims to lower the environmental footprint from the transportation of passengers. Daisy provides accurate computation and estimates of travel times and fuel consumption of individual trips to support for improved fleet management and trip planning. This requires efficient handling of large GPS data sets with a huge number of trajectories, based on a completely open-source software stack. The fuel reductions are evaluated through two case studies using real-world Danish data.

The *Streamspin* project (streamspin.cs.aau.dk) aimed to "be for mobile services what Youtube is for video." The Streamspin system: (i) enables user-generated services by enabling programmers to create service templates from which non-programmers can create services, (ii) offers support for basic aspects of services such as authentication, security, and privacy as well as the ability to flexibly push content to users, (iii) enables tracking of users with varying accuracy, (iv) enables service sharing, and (v) enables the scalable delivery of services. The system is regularly used in teaching mobile system development and has been used by several projects.

In support of the regional development strategy for North Jutland of using location and context information for SmartCities, Intelligent transport and Infotainment, the *SmartCampusAAU* (smartcampus.cs.aau.dk) ERDF project built a generic infrastructure that extends Streamspin to a software platform (Android, iPhone, Windows Mobile) for combined indoor positioning (WiFi and Bluetooth) and outdoor GPS positioning. Demonstrators were built with the companies MVC-Data (secure door locks), Folia (campus services) and the Utzon Centre (educational). The follow-up SmartCampus 2.0 project commercializes the results through Folia which now has product offerings on creating indoor maps on top of Google Maps.

Next, we will describe recent (mid 2000s and up) and current data management related research topics.

3. SPATIO-TEMPORAL DATA MANAGEMENT

This work is generally motivated by the increasing abundance of spatio-temporal data, often in the form of GPS data obtained from a variety of so-called moving objects, as well as the increasing mobile use of services and the Internet.

Frequent Updates and Indexing. The group devoted a substantial effort to the development of spatial and spatio-temporal indexing techniques. Several new indexing techniques have been proposed, and benchmarking of indexes has been pursued [8]. Recently the group turned its attention to emerging application areas of computing technologies that involve the monitoring of continuous variables, such as positions of moving objects. Such monitoring yields massive update loads that existing systems are unable to contend with. We explored a number of spatial indexing approaches to enable the support of such application areas. One such approach is shedding of index updates. We proposed a framework [69] that renders an underlying disk-based R-tree index adaptive to the incoming workload. Query latency is low in frequently queried areas, but it is allowed to deteriorate in infrequently queried areas, in order to free resources to process massive update loads. Alternatively, we explored the best ways to use main memory for update operation buffering in R-trees [2].

For very high update rates solutions involving secondary storage are not practical. An extensive experimental study was performed to compare main-memory variants of the R-tree with variants of a simple, uniform grid [74]. In addition to main-memory, another resource keeps increasing in current computer systems, namely the parallel processing capabilities of chip multi-processors (CMPs). We explored how to harness this parallelism to support high rates of spatial updates. This involves the non-trivial challenge of avoiding contention between long-running queries and frequent updates. One approach is to maintain two copies of an index, a static index for queries and a live one for updates, and to perform frequent refreshing of the query index by copying of the live index [73, 72]. Alternatively, a more fine grained concurrency control based on hardware-assisted atomic updates as well as object-level copying is employed in PGrid [75].

Spatio-Textual Search and Ranking of Spatial Web Objects. Web users and content are increasingly being geo-positioned, and increased focus is being given to serving local content in response to textual web queries. This development calls for spatial keyword queries that take into account both the locations and textual descriptions of content. We studied the efficient, joint processing of multiple top-k spatial keyword queries [79] and moving top-k spatial keyword query processing [80]. Incorporating the effects of the nearby objects to a returned query result, a so called prestige-based relevance, was also explored [5]. Related to this research is the research on automatic mining of the semantically significant locations

from the GPS traces of movement [4].

Skyline Queries. In the context of multi-dimensional data management, we have conducted series of research on skylines that go beyond the limits of conventional skyline queries. A flexible framework [48] is proposed to efficiently resolve arbitrary user-specified size constraints on skyline queries. Another generalized framework [89] guides the extension of conventional skyline queries. The conventional skyline dominance concept is adapted to rank assorted user preferences [49]. In addition we have studied how to upgrade disadvantaged points to skyline points at a low cost [47]. We have also developed skyline algorithms for distributed sites [6], peer-to-peer networks [10, 7], and data streams [52]. Further, we have integrated the skyline dominance concept to spatial queries to support various spatial decision makings [51] as well as spatial object ranking [88].

ITS. Within spatio-temporal research Intelligent Transport Systems (ITS) is a focus area. In particular, usage of trajectories to determine turn-times in intersections, estimating traveltime, and finding eco-friendly routes has been studied [42, 41].

We have access to several large GPS data sets and have built a large software infrastructure for handling GPS and trajectory data to make the results available to collaborators.

The group also worked on other topics within spatio-temporal data management such as modeling, indexing, and query processing proposals for spatial-network constrained objects and tracking of moving objects.

4. MOBILE SERVICES

Daisy has engaged in a range of research activities that target technologies for outdoor as well as indoor mobile services. These activities have produced not only novel theoretical findings, but also useful practical prototypes.

Middleware for Mobile Services. As described in Section 2, the Streamspin platform focused on enabling user-generated location-based services by building a scalable middleware infrastructure for mobile services. An overview of Streamspin is given in a journal paper [39]. Streamspin has served as prototyping infrastructure for many activities, including outdoor GPS-based tracking of mobile objects, WiFi- and Bluetooth based indoor positioning, as well as seamless indoor-outdoor positioning [32, 31, 28]. Streamspin has been extended to produce a middleware infrastructures for indoor positioning.

Indoor Positioning. The foundations for the work on WiFi-based indoor positioning uses the so-called finger printing technique where a radiomap is constructed for building by measuring signal strengths at given locations inside it and storing these in a database. These fingerprints are later used to position a mobile user by measuring received signals and looking up the position in the database. The approach developed in Daisy exploits a weighted graph approach to achieve efficiency and accuracy [29, 30]. Similarly to WiFi, it is possible to use Bluetooth signals if a Bluetooth infrastructure is available. Such an approach and comparative studies with WiFi have been conducted reporting a slightly higher precision from Bluetooth over WiFi, but at the expense of a much more fine-grained and costly infrastructure [31]. In addition, we have conducted studies on hybrid indoor positioning [1] with both WiFi and Bluetooth technologies. Particularly, a limited number of expensive Bluetooth hotspots are deliberately deployed at preselected indoor positions such that the indoor space as well as the original WiFi radio map is partitioned into small parts. As a result, the computation is reduced for the online WiFi based positioning, and the overall positioning accuracy is enhanced. The combination of indoor WiFi based and outdoor GPS based positioning and a Bluetooth based door lock protocol for easy access to buildings requiring authentication of users has also been studied [78]. The protocol has been formally verified and made public for scrutiny.

Managing Indoor Space and Indoor Moving Objects. Indoor spaces are characterized by unique entities such as walls, doors, rooms, etc. that not only allow but also constrain movements. Such characteristics, plus the fact that alternative positioning technologies other than GPS are much more suitable in indoor spaces, call for new data management technologies for indoor spaces and moving objects in indoor spaces. Daisy has been one of the pioneers in the line of this new research frontier. We have designed a distance-aware indoor space model [46] that supports efficient spatial queries including shortest path searching in indoor spaces, as well as a unified space model [33] for large mixed spaces like an airport. Other topics in the indoor setting include: indoor moving object tracking [36], indoor trajectory indexing techniques [37], indoor-distance aware queries [81], and query processing for indoor moving objects [82, 83, 50].

Location Privacy in Mobile Services. Further, Daisy has been active in the area of location-related privacy protection for different kinds of mobile services. Relevant contributions apply to location privacy aware spatial queries in client-server location-based services [87, 86], privacy-preserving online route planning [70], geo-social networks [77, 18, 17], and health care emergency services [71]. We have also explored privacy in location based systems where the goal is to detect and maintain the proximity/separation information between private moving object positions [76].

5. DATA WAREHOUSING AND BI

Data warehousing and business intelligence tools and techniques motivated by advanced realworld business intelligence applications is another major research topic, divided in a number of subareas.

ETL Frameworks. The group has developed its own *programmatic* approach to Extract-Transform-Load (ETL) that unlike tradtional graphical ETLs aims to provide very high ETL programmer productivity through high-level, code-based tools. First was the Python-based *pygramETL* [66] that provides a number of high-level constructs for dimensional ETL concepts such as star and snowflake schemas and slowly-changing dimensions (SCDs). This was later complemented with a version supporting multi-core parallel execution [67] and the *ETLMR* framework that runs on MapReduce environments and thus provides very good scalability [44, 45].

Near-real-time DW and BI. The group has developed the RiTE (Right-Time Etl) middleware system that combines INSERT-like data availability with bulk loading speeds through the use of a main memory based *catalyst* that provides data on-demand and is transparent to the data producers and consumers [68]. A system for performing effective OLAP on data streams has also been developed [85]. Recently, real-time BI on energy data scenarios led to the TimeTravel system that supports efficient seamless querying of past and (fore-

casted) future data using a hierarchical model-based storage scheme [40], and to techniques for subscription-based forecast queries [16].

Bitmap Indexing. A novel compressed bitmap index called *Postion List Word Aligned Hybrid (PLWAH)* was developed [15]. Compared to the previous state-of-the-art techniques WAH, it often only takes half the storage space and provides 60-70% better query performance, through the use of special CPU instructions available on modern processors. The PLWAH technique is currently patent pending and is being commercialized through the Algorhyme startup company.

Warehousing and OLAPing Complex Data. This major topic covers the 3XL system that utilizes the object-relational features of PostgreSQL for efficiently warehousing OWL Lite data [43], the Multidimensional Integrated Ontologies framework for designing semantic web DWs [57], online integration of cubes with XML and object data [84], the *relevance cube (R-cube)* framework for contextualizing DWs with text documents [61, 60], warehousing smart grid data [64] and aggregating smart grid inspired so-called *flexibility objects* [63], and finally warehousing multi-granular dimensional data [34]

Spatio-temporal Data Warehousing and Data Mining. This research covers pattern mining and privacy-preserving data mining and data collection on moving object trajectories [26, 24, 27], continuous moving object location and density prediction on road networks [19], spatio-temporal data generation [21], spatio-temporal prediction in mobile networks [62] and their use in location-based advertising [20] and social ride- and cab-sharing [22, 25, 23].

Sentinels. This is a novel type of data mining pattern capturing cause-effect relationships between changes in cube measure values. Sentinels can warn users of possible future changes in key so-called target measures such as revenue based on earlier changes in earlier so-called source measures. The concepts of sentinels was formalized [53] and a number of increasingly efficient algorithms developed [56] which were incorporated into the commercial TARGIT BI Suite [54, 55].

6. FURTHER TOPICS

Multimedia Data Management. This was a major topic during the last half of the 2000s, but is now no longer actively pursued. The research focused on providing effective and scalable techniques for different types of multimedia data. Specific topics included querying frameworks and playlist generation for music databases [13, 14]; effective indexing techniques for large music databases [35]; effective query processing techniques for large music databases [12]; similarity search for high-dimensional multimedia data such as time series and images; data mining, especially subspace clustering, for high-dimensional multimedia data; text mining and question-answering systems; and social network mining.

Data Management Infrastructure. This covers performance and correctness test of database applications [58, 9] and automated evaluation of database schemas. The work is being done in close collaboration with industry partners.

7. FUTURE PERSPECTIVES AND OPPORTUNITIES

Finally, we turn to the research agenda for the coming years. Motivated by the increasing proliferation of IT services processing ever larger and more complex data sets on diverse platforms, the main overall research theme will be *data-intensive services - mobile, ubiquitous, cloud, and beyond*. The group further plans to build on its strengths in *spatio-temporal data management and mobile services*, including spatial and spatio-temporal indexing and query processing, and support for integrating indoor and outdoor spaces. Finally, another focus area will be *cloud intelligence* [59, 11]; i.e., BI in, for, and with the cloud, which is also explored in the new Cloud Intelligence workshop series co-located with VLDB (eric.univ-lyon2.fr/cloud-i).

Energy data management will be explored in the TotalFlex project (www.totalflex.dk) and in further funded projects, and within the Energy Data Management workshop series co-located with EDBT (www.endm.org). Similarly, new funded projects are expected within mobile services (for example e-health services), logistics, and intelligent transport systems.

Positions will be regularly available at most levels. A number of Ph.D. positions are available each year, especially in the brand new Erasmus Mundus Joint Doctorate *IT Technologies for Business Intelligence - Doctoral Collegege (IT4BI-DC)* (it4bi-dc.ulb.ac.be) in collaboration with ULB (Zimanyi), TUD (Lehner), UPC (Abello), and PUT (Wrembel), with application deadline December 21, 2012. Post doc, assistant, and associate professor positions will be announced annually.

8. ACKNOWLEDGMENTS

We would like to thank all the previous members of our team, our sponsoring agencies, and our collaborators, friends and families for all supporting the research described above. Furthermore, we would like to thanks the Research Centers column editors Ugur Cetintemel and Alkis Simitsis.

9. REFERENCES

[1] A. Baniukevic, D. Sabonis, C. S. Jensen, and H. Lu. Improving wi-fi based indoor positioning using bluetooth add-ons. In *MDM*, pages 246–255, 2011.

[2] L. Biveinis, S. Šaltenis, and C. S. Jensen. Main-memory operation buffering for efficient r-tree update. In *PVLDB*, pages 591–602, 2007.

[3] M. Böhm, L. Dannecker, A. Doms, E. Dovgan, B. Filipic, U. Fischer, W. Lehner, T. B. Pedersen, Y. Pitarch, L. Siksnys, and T. Tusar. Data management in the mirabel smart grid system. In *EDBT/ICDT Workshops*, pages 95–102, 2012.

[4] X. Cao, G. Cong, and C. S. Jensen. Mining significant semantic locations from gps data. *PVLDB*, 3(1):1009–1020, 2010.

[5] X. Cao, G. Cong, and C. S. Jensen. Retrieving top-k prestige-based relevant spatial web objects. *PVLDB*, 3(1):373–384, 2010.

[6] L. Chen, B. Cui, and H. Lu. Constrained skyline query processing against distributed data sites. *TKDE*, 23(2):204–217, 2011.

[7] L. Chen, B. Cui, H. Lu, L. Xu, and Q. Xu. isky: Efficient and progressive skyline computing in a structured p2p network. In *ICDCS*, pages 160–167, 2008.

[8] S. Chen, C. S. Jensen, and D. Lin. A benchmark for evaluating moving object indexes. In *PVLDB*, pages 1574–1585, 2008.

[9] C. Christensen, S. Gundersborg, K. Linde, and K. Torp. A unit-test framework for database applications. In *IDEAS*, pages 11–20, 2006.

[10] B. Cui, L. Chen, L. Xu, H. Lu, G. Song, and Q. Xu. Efficient skyline computation in structured peer-to-peer systems. *TKDE*, 21(7):1059–1072, 2009.

[11] J. Darmont, T. B. Pedersen, and M. Middelfart. Cloud intelligence: what is really new? In *Cloud-I*, page 10, 2012.

[12] F. Deliege, B. Chua, and T. B. Pedersen. High-level audio features: Distributed extraction and similarity search. In *ISMIR*, pages 565–570, 2008.

[13] F. Deliege and T. B. Pedersen. Fuzzy song sets for music warehouses. In *ISMIR*, pages 21–26, 2007.

[14] F. Deliege and T. B. Pedersen. Using fuzzy lists for playlist management. In *MMM*, pages 198–209, 2008.

[15] F. Deliege and T. B. Pedersen. Position list word aligned hybrid: optimizing space and performance for compressed bitmaps. In *EDBT*, pages 228–239, 2010.

[16] U. Fischer, M. Böhm, W. Lehner, and T. B. Pedersen. Optimizing notifications of subscription-based forecast queries. In *SSDBM*, pages 449–466, 2012.

[17] D. Freni, C. R. Vincente, S. Mascetti, C. Bettini, and C. S. Jensen. Preserving location and absence privacy in geo-social networks. In *CIKM*, pages 309–318, 2010.

[18] G. Ghinita, C. R. Vincente, N. Shang, and E. Bertino. Privacy-preserving matching of spatial datasets with protection against background knowledge. In *ACM GIS*, pages 3–12, 2010.

[19] G. Gidófalvi, M. Kaul, C. Borgelt, and T. B. Pedersen. Frequent route based continuous moving object location- and density prediction on road networks. In *GIS*, pages 381–384, 2011.

[20] G. Gidofalvi, H. Larsen, and T. B. Pedersen. Estimating the capacity of the location-based advertising channel. *IJMC*, 6(3):357–375, 2008.

[21] G. Gidofalvi and T. B. Pedersen. St-acts. In *ACM GIS*, pages 155–162, 2006.

[22] G. Gidofalvi and T. B. Pedersen. Cab-sharing: An effective, door-to-door, on-demand transportation service. In *ITS Europe*, 2007.

[23] G. Gidofalvi and T. B. Pedersen. Instant social ridesharing. In *ITS World*, 2008.

[24] G. Gidofalvi and T. B. Pedersen. Mining long, sharable patterns in trajectories of moving objects. *Geoinformatica*, 13(1):27–55, 2009.

[25] G. Gidofalvi, T. B. Pedersen, T. Risch, and E. Zeitler. Highly scalable trip grouping for large scale collective transportation systems. In *EDBT*, pages 678–689, 2008.

[26] G. Gidofalvi, H. Xuegang, and T. B. Pedersen. Privacy-preserving data mining on moving object trajectories. In *MDM*, pages 60–68. MDM, 2007.

[27] G. Gidofalvi, H. Xuegang, and T. B. Pedersen. Privacy-preserving trajectory collection. In *ACM GIS*, 2008.

[28] R. Hansen, C. S. Jensen, B. Thomsen, and R. Wind. Seamless indoor/outdoor positioning with streamspin. In *MobiQuitous*, 2008.

[29] R. Hansen and B. Thomsen. Using weighted graphs for computationally efficient wlan location determination. In *MobiQuitous*, pages 1–5, 2007.

[30] R. Hansen and B. Thomsen. Efficient and accurate wlan positioning with weighted graphs. In *MOBILIGHT*, pages 372–386. 2009.

[31] R. Hansen, R. Wind, C. S. Jensen, and B. Thomsen. Seamless indoor/outdoor positioning handover for location-based services in streamspin. In *MDM*, pages 267 –272, 2009.

[32] R. Hansen, R. Wind, C. S. Jensen, and B. Thomsen. Algorithmic strategies for adapting to environmental changes in 802.11 location fingerprinting. In *IPIN*, pages 1 –10, 2010.

[33] S. H. Hussein, H. Lu, and T. B. Pedersen. Towards a unified model of outdoor and indoor spaces. In *ACM GIS*, 2012.

[34] N. Iftikhar and T. B. Pedersen. Schema design alternatives for multi-granular data warehousing. In *DEXA*, volume 6262, pages 111–125, 2010.

[35] C. A. Jensen, E. M. Mungure, T. BachPedersen, K. Sørensen, and F. Deliege. Effective bitmap indexing for non-metric similarities. In *DEXA*, pages 137–151, 2010.

[36] C. S. Jensen, H. Lu, and B. Yang. Graph model based indoor tracking. In *MDM*, pages 122–131, 2009.

[37] C. S. Jensen, H. Lu, and B. Yang. Indexing the trajectories of moving objects in symbolic indoor space. *LNCS*, 5644:208–227, 2009.

[38] C. S. Jensen and R. T. Snodgrass. Temporal data management. *TKDE*, 11(1):36–44, 1999.

[39] C. S. Jensen, C. R. Vicente, and R. Wind. User-generated content: The case for mobile services. *IEEE Computer*, 41(12):116–118, 2008.

[40] M. E. Khalefa, U. Fischer, T. B. Pedersen, and W. Lehner. Model-based integration of past & future in timetravel. *PVLDB*, 5(12):1974–1977, 2012.

[41] B. B. Krogh, O. Andersen, and K. Torp. Trajectories for novel and detailed traffic information. In *IWSG*, 2012.

[42] H. Lahrmann and K. Torp. Travel times, congestion levels, and delays at intersections calculated on the basis of floating car data. In *ITS World*, 2010.

[43] X. Liu, C. Thomsen, and T. B. Pedersen. 3xl: Supporting efficient operations on very large owl lite triple-stores. *Inf. Syst.*, 36(4):765–781, 2011.

[44] X. Liu, C. Thomsen, and T. B. Pedersen. Etlmr: A highly scalable dimensional etl framework based on mapreduce. In *DaWaK*, pages 96–111, 2011.

[45] X. Liu, C. Thomsen, and T. B. Pedersen. Mapreduce-based dimensional etl made easy. *PVLDB*, 5(12):1882–1885, 2012.

[46] H. Lu, X. Cao, and C. S. Jensen. A foundation for efficient indoor distance-aware query processing. In *ICDE*, pages 438–449, 2012.

[47] H. Lu and C. S. Jensen. Upgrading uncompetitive products economically. In *ICDE*, pages 977–988, 2012.

[48] H. Lu, C. S. Jensen, and Z. Zhang. Flexible and efficient resolution of skyline query size constraints. *TKDE*, 23(7), 2010.

[49] H. Lu and L. Xu. Identifying the most influential user preference from an assorted collection. In *SSDBM*, pages 233–251, 2010.

[50] H. Lu, B. Yang, and C. S. Jensen. Spatio-temporal joins on symbolic indoor tracking data. In *ICDE*, pages 816–827, 2010.

[51] H. Lu and M. L. Yiu. On computing farthest dominated locations. *TKDE*, 23(6):928–941, 2011.

[52] H. Lu, Y. Zhou, and J. Haustad. Efficient and scalable continuous skyline monitoring in two-tier streaming settings. *Inf. Syst.*, 38(1):68–81, 2013.

[53] M. Middelfart and T. B. Pedersen. Discovering sentinel rules for business intelligence. In *DEXA*, pages 592–602, 2009.

[54] M. Middelfart and T. B. Pedersen. Using sentinel technology in the targit bi suite. *PVLDB*, 3(2):1629–1632, 2010.

[55] M. Middelfart and T. B. Pedersen. Implementing sentinels in the targit bi suite. In *ICDE*, pages 1187–1198, 2011.

[56] M. Middelfart, T. B. Pedersen, and J. Krogsgaard. Efficient discovery of generalized sentinel rules. In *DEXA*, pages 32–48, 2010.

[57] V. Nebot, R. Berlanga, J. Perez, M. Aramburu, and T. B. Pedersen. Multidimensional integrated ontologies: A framework for designing semantic data warehouses. *JODS*, XIII:1–36, 2009.

[58] K. Pedersen, K. Torp, and R. Wind. Simple and realistic data generation. In *PVLDB*, pages 1243–1246, 2006.

[59] T. B. Pedersen. *Research challenges for cloud intelligence (invited talk)*. BEWEB, 2010.

[60] J. Perez, M. Aramburu, R. Berlanga, and T. B. Pedersen. *R-Cubes*. IEEE Press, 2007.

[61] J. Perez, R. Berlanga, M. Aramburu, and T. B. Pedersen. Contextualizing data warehouses with documents. *DSS*, 45(1):77–94, 2008.

[62] S. Samulevicius, T. B. Pedersen, T. B. Sorensen, and G. Micallef. Energy savings in mobile broadband network based on load predictions: Opportunities and potentials. In *VTC Spring*, pages 1–5, 2012.

[63] L. Siksnys, M. E. Khalefa, and T. B. Pedersen. Aggregating and disaggregating flexibility objects. In *SSDBM*, pages 379–396, 2012.

[64] L. Siksnys, C. Thomsen, and T. B. Pedersen. Mirabel dw: Managing complex energy data in a smart grid. In *DaWaK*, pages 443–457, 2012.

[65] R. T. Snodgrass. *Developing Time-Oriented Database Applications in SQL*. Morgan Kaufmann, 1999.

[66] C. Thomsen and T. B. Pedersen. pygrametl: A powerful programming framework for extract-transform-load programmers. In *DOLAP*, pages 49–56, 2009.

[67] C. Thomsen and T. B. Pedersen. Easy and effective parallel programmable etl. In *DOLAP*, pages 37–44, 2011.

[68] C. Thomsen, T. B. Pedersen, and W. Lehner. Rite: Providing on-demand data for right-time data warehousing. In *ICDE*, pages 456–465, 2008.

[69] K. Tzoumas, M. Yiu, and C. S. Jensen. Workload-aware indexing of continuously moving objects. *PVLDB*, (1):1186–1197, 2009.

[70] C. R. Vicente, I. Assent, and C. S. Jensen. Effective privacy-preserving online route planning. In *MDM*, pages 119–128, 2011.

[71] C. Vincente, M. Kirkpatrick, G. Ghinita, E. Bertino, and C. S. Jensen. Towards location-based access control in healthcare emergency response. In *SPRINGL*, pages 22–26, 2009.

[72] D. Šidlauskas, C. S. Jensen, and S. Šaltenis. A comparison of the use of virtual versus physical snapshots for supporting update-intensive workloads. In *DaMoN*, pages 1–8, 2012.

[73] D. Šidlauskas, K. Ross, C. S. Jensen, and S. Šaltenis. Thread-level parallel indexing of update intensive moving-object workloads. *LNCS*, 6849:186–204, 2011.

[74] D. Šidlauskas, S. Šaltenis, C. Christiansen, J. Johansen, and D. Saulys. Trees or grids? indexing moving objects in main memory. In *ACM GIS*, pages 236–245, 2009.

[75] D. Šidlauskas, S. Šaltenis, and C. S. Jensen. Parallel main-memory indexing for moving-object query and update workloads. In *SIGMOD*, pages 37–48, 2012.

[76] L. Šikšnys, J. Thomsen, S. Šaltenis, and M. Yiu. Private and flexible proximity detection in mobile social networks. In *MDM*, pages 75–84, 2010.

[77] L. Šikšnys, J. Thomsen, S. Šaltenis, M. Yiu, and O. Andersen. A location privacy aware friend locator. *LNCS*, 5644:405–410, 2009.

[78] E. R. Wognsen, H. S. Karlsen, M. Calverley, M. N. Follin, B. Thomsen, and H. Huttel. A secure relay protocol for door access control. In *SBSeg*, 2012.

[79] D. Wu, M. Yiu, G. Cong, and C. S. Jensen. Joint top-k spatial keyword query processing. *TKDE*, 24(10):1889–1903, 2012.

[80] D. Wu, M. Yiu, C. S. Jensen, and G. Cong. Efficient continuously moving top-k spatial keyword query processing. In *ICDE*, pages 541–552, 2011.

[81] X. Xie, H. Lu, and T. B. Pedersen. Efficient distance-aware query evaluation on indoor moving objects. In *ICDE*, 2013.

[82] B. Yang, H. Lu, and C. S. Jensen. Scalable continuous range monitoring of moving objects in symbolic indoor space. In *CIKM*, pages 671–680, 2009.

[83] B. Yang, H. Lu, and C. S. Jensen. Probabilistic threshold k nearest neighbor queries over moving objects in symbolic indoor space. In *EDBT*, pages 335–346, 2010.

[84] X. Yin and T. B. Pedersen. Evaluating xml-extended olap queries based on a physical algebra. *JDM*, 17(2):85–116, 2006.

[85] X. Yin and T. B. Pedersen. *What Can Hierarchies Do for Data Streams?* LNCS. BIRTE, 2007.

[86] M. Yiu, C. S. Jensen, J. Møller, and H. Lu. Design and analysis of an incremental approach to location privacy for location-based services. *TODS*, 32(2), 2010.

[87] M. L. Yiu, C. S. Jensen, X. Huang, and H. Lu. Spacetwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services. In *ICDE*, pages 366–375, 2008.

[88] M. L. Yiu, H. Lu, N. Mamoulis, and M. Vaitis. Ranking spatial data by quality preferences. *TKDE*, 23(3):433–446, 2011.

[89] Z. Zhang, H. Lu, B. C. Ooi, and A. K. H. Tung. Understanding the meaning of a shifted sky: A general framework on extending skyline query. *VLDB Journal*, 19(2):181–201, 2010.