

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Programa de Doctorat:

AUTOMÀTICA, ROBÒTICA I VISIÓ

Proposta de Tesi Doctoral

**DISSENY I MODELITZACIÓ D'UN SISTEMA DE GESTIÓ
MULTIRESOLUCIÓ DE SÈRIES TEMPORALS**

Aleix Llusà Serra

Direcció:

Teresa Escobet Canal
Sebastià Vila-Marta

Juny de 2012

Primera edició: juny de 2012.

Primera versió: 1.0.1 (composta 12 de setembre de 2012).

Amb el suport de la Universitat Politècnica de Catalunya (UPC).

Amb la col·laboració del departament de Disseny i Programació de Sistemes Electrònics (DiPSE) i la Càtedra de Programari Lliure (CPL).



Copyright (C) 2012 Aleix Llusà Serra.

Aquest document està sotmès a una llicència de Reconeixement-CompartirIgual 3.0 No adaptada de Creative Commons. Per veure una còpia de la llicència, visiteu <http://creativecommons.org/licenses/by-sa/3.0/deed.ca> o envieu una carta a Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

Aleix Llusà Serra

Departament de Disseny i Programació de Sistemes Electrònics de la Universitat Politècnica de Catalunya (DiPSE-UPC)

Escola Politècnica Superior d'Enginyeria de Manresa (EPSEM), Av. de les Bases de Manresa, 61-73, 08242 Manresa (Barcelona), CATALUNYA

aleix@dipse.upc.edu

El codi font L^AT_EX del document es troba a <http://escriuny.epsem.upc.edu/projects/rrb/>

Índex

1. Introducció	4
1.1. Estructura del document	5
1.2. Objectius i contribucions esperades	5
1.3. Justificació	6
2. Estat de l'art	8
2.1. Sèries temporals	8
2.1.1. Adquisició de sèries temporals	9
2.1.2. Anàlisi de sèries temporals	12
2.1.3. Emmagatzematge i gestió de sèries temporals	13
2.1.4. Conclusió	16
2.2. Sistemes de gestió de bases de dades	17
2.2.1. Sistemes relacionals	17
2.2.2. Altres sistemes	19
2.2.3. Dades temporals	20
2.2.4. Conclusió	21
3. Planificació	22
3.1. Treball realitzat	22
3.2. Treball futur	23
3.3. Mitjans	25
A. Direcció	35
A.1. Teresa Escobet Canal	35
A.2. Sebastià Vila-Marta	35

1. Introducció

Aquesta recerca s'estructura al voltant dels sistemes d'emmagatzematge i tractament de dades com a sèries temporals. Concretament se centra en els sistemes de gestió de bases de dades (SGBD) que s'ocupen de sèries temporals (SGST). Els SGST han de tenir funcionalitats adequades per gestionar i explotar correctament la informació de les sèries temporals.

La recerca en anàlisi de sèries temporals ha augmentat en la darrera dècada, tal com explica Fu, [30]. Actualment, és un camp de gran interès, sobretot pel que fa a processar grans volums de dades amb bon rendiment, tant des de la vessant de temps d'execució com, modernament, de la vessant de consum d'energia. Com es mostra a l'apartat 2 d'aquesta memòria, hi ha multitud de metodologies i algoritmes que proposen solucions a aquests problemes. Tot i així, la recerca en aquests temes segueix avançant a la recerca de noves solucions. En general podem dir que les metodologies i algoritmes basats en sèries temporals són bones candidates per a ésser resoltes de forma ordenada i integrada mitjançant SGBD. Per a assolir aquest objectiu és necessari disposar d'un model per a sèries temporals a on es reculli de forma general el problema de tractament de sèries temporals, aleshores molts d'aquests estudis de la recerca de sèries temporals es podran aplicar en la implementació de SGBD que segueixin el model de SGST.

Recentment s'ha observat que hi ha necessitat de coneixement entre els SGBD i les aplicacions de les sèries temporals, [74, 85]. Hi alguns sistemes que es poden considerar com a SGST, com per exemple *RRDtool*, [58], o *Cougar*, [33], però en ells no s'ha definit clarament un model de SGST. Dels treballs publicats en els que es mostra la necessitat d'estudiar conceptes de model es destaquen el treball de Dreyer, Dittrich i Schmidt, [25], en el qual presenten l'estructura bàsica que han de tenir els SGST; els estudis de Bonnet, Gehrke i Seshadri, [5], per a xarxes de sensors; i els exemples de consultes de Zhang i altres., [85], per algunes de les propietats de les sèries temporals.

En els SGBD es contempla l'estudi formal dels seus conceptes mitjançant models basats en teories matemàtiques, del qual n'és una referència el model relacional, [12]. Dins del model relacional hi ha hagut un estudi profund relacionat amb els intervals temporals, considerats com a dades temporals, [20], que resol el problema dels històrics temporals en els SGBD. Les sèries temporals necessiten un estudi similar ja que no poden ser tractades com a intervals temporals per pertànyer a una categoria diferent de dades temporals, [1, 65].

1.1. Estructura del document

Aquest document és una proposta de recerca per a definir un model de sistema de gestió per a sèries temporals. En una primera part, que segueix a aquesta introducció, es presenten els objectius i les justificacions de la recerca. En una segona part s'estudia el context de la recerca i l'estat actual en els àmbits de les sèries temporals i els SGBD. En una tercera part es resumeix el treball dut a terme fins a l'actualitat i la planificació del treball futur per a assolir els objectius proposats. Al final, s'annexa el perfil professional de la directora i del director de la tesi.

1.2. Objectius i contribucions esperades

Aquesta recerca té per objectiu l'estudi de les necessitats específiques que comporta l'emmagatzemat i gestió de dades amb naturalesa de sèrie temporal i la proposta d'un model de SGBD que satisfaci aquestes necessitats. Aquest objectiu es divideix en els següents subobjectius més concrets:

1. Estudi de les aplicacions en que les dades són sèries temporals amb la finalitat de determinar quines són les propietats i problemes comuns que planteja la seva gestió i emmagatzematge.
2. Estudi dels models de SGBD existents. Segons es desprèn de la formalització de Date, [12], el model principal és el model relacional, el qual es fonamenta en dos conceptes: relacions i tipus de dades.
3. Una àrea de treball important en els SGBD és la incorporació de nous tipus de dades complexos. És important estudiar com es modifica el model de dades d'un SGBD quan s'afegeix un nou tipus de dades complex. Les sèries temporals es poden d'entendre com a tipus complex ja que presenten diferents propietats característiques i necessiten operadors addicionals. Els SGBD permeten que els usuaris defineixin nous tipus de dades, [71], però no hi ha un estudi teòric dels tipus de dades en els SGBD. Date, [12], descriu abastament les relacions però no els tipus de dades. Els tipus de dades s'han d'estudiar i modelar per a poder-los tractar i generar operadors, oimés els tipus complexos ja que requereixen un estudi més complet i possiblement s'hagin de modelar com un propi SGBD. Una referència d'estudi és el cas dels intervals temporals, [20].
4. Disseny d'un model de SGBD per a les sèries temporals. D'aquesta manera els SGBD podran tractar dades amb instants de temps que mostrin l'evolució de variables en funció del temps. El model consisteix en la definició de l'estructura de les sèries temporals i les operacions bàsiques que necessiten.

L'assoliment d'aquest objectiu té dues parts:

1. Introducció

- a) Disseny d'un model per a la gestió bàsica de les sèries temporals, el qual anomenem model de SGBD per a sèries temporals (SGST). L'estructura d'aquest model és similar a l'utilitzat en els intervals temporals, [20]. Prenent com a base el model de SGST, el qual és un model general per a les sèries temporals, s'hi poden incloure altres models per a propietats més específiques de les sèries temporals.
 - b) Disseny d'un model específic en base del model de SGST. Concretament es dissenya un model pels SGST multiresolució (SGSTM). En el model de SGSTM s'hi poden incloure propietats de les sèries temporals relacionades amb la resolució que s'han observat en les aplicacions pràctiques de les sèries temporals: regularització, canvis de resolució mitjançant agregacions, reconstrucció de forats, etc.
5. Implementació de referència dels models de SGST i SGSTM. Per una banda, aquesta implementació, a nivell acadèmic, ha de servir com a exemple per a futurs desenvolupaments de sistemes de gestió, acadèmics o productius. Per altra banda, ha de servir per a exemplificar-ne els seu funcionament amb unes dades de prova.

1.3. Justificació

Una de les grans mancances en l'àmbit de les sèries temporals és no tenir un model que faci d'enllaç entre les diferents aplicacions i les diferents implementacions, el que s'anomena com a necessitat de coneixement entre els SGBD i les sèries temporals, [74, 85]. Un model permetria, entre d'altres, estudiar les propietats abstractes de les sèries temporals i poder comparar diferents sistemes de gestió. L'estudi d'un model de SGBD per a sèries temporals té dues vessants.

Per una banda, després d'una mirada general a algunes aplicacions de les sèries temporals, s'observa que hi ha molts algorismes específics i dissenyats a mida per al seu tractament. A més, s'observa que en el tractament de les sèries temporals es duen a terme un conjunt de tasques comunes: canvis de resolució, reconstrucció de forats, reducció del volum de dades, etc. Tot i que hi ha sistemes dissenyats específicament per a tractar-hi, com per exemple *RRDtool*, [58], o *Cougar*, [31], no s'ha identificat un model general per tractar aquests problemes.

Per altra banda, en el model relacional de SGBD es troba una formalització per a tractar qualsevol tipus de dades. Disposar d'un model matemàtic consolidat, com és el cas del relacional, ha estat una fita important en l'àmbit dels SGBD. No obstant, el model relacional només presenta els conceptes bàsics dels SGBD, quedant per resoldre la formalització dels tipus de dades. Concretament dins d'aquest àmbit destaca la publicació del model per a intervals temporals, [20], el qual s'utilitza per a formalitzar els històrics en els SGBD.

1. Introducció

En l'àmbit dels SGBD també s'hi inclou l'estudi d'implementacions que permetin obtenir un bon rendiment. Un cop s'ha formalitzat el model es poden avaluar les implementacions més adequades per a cada aplicació. Per exemple en el cas de les sèries temporals es pot aplicar la gestió mitjançant *data streams*, [3].

El model relacional, [10], va permetre situar els SGBD relacionals com a preeminents, sobretot els que tenien en comú un llenguatge anomenat SQL (*Structured Query Language*). Tot i que semblava que els SGBD eren un àmbit consolidat, recentment el corrent NoSQL (*Not Only SQL*), [28, 72], ho ha posat en dubte. En aquest nou enfoc han aparegut SGBD que milloren el rendiment dels SGBD SQL, però no queda massa clar els models emprats. En alguns casos segons comenta Date, [13, cap. 14,27], sembla que sigui una reducció del model relacional, com per exemple els conceptes d'arbres i d'objectes; en altres casos s'aporten conceptes nou però les metodologies emprades són encara joves per avaluar-ne la potència. Date i Darwen, [18], també s'oposen a la preeminència dels SGBD SQL i proposen conceptes i llenguatges purament relacionals.

Així doncs, tenint en compte que segons Date el model relacional és complet, consolidat i que no hi ha cap de tant potent, la modelització de SGST s'ha de definir pròpera als conceptes relacionals, tal com s'ha efectuat pel cas dels intervals temporals, [20]. Així els SGST podran aprofitar els avantatges i l'experiència del model relacional, així com també podran aprofitar els estudis d'implementacions eficients, com per exemple els *data streams* per a dades amb naturalesa de seqüència, [4].

El model relacional contempla la seva extensió mitjançant nous tipus de dades, tot i que el considera un aspecte independent i només en defineix la seva possible necessitat. Per a tipus de dades senzills és suficient amb definir una possible representació amb restriccions, per exemple definir telèfons com a subconjunts dels enters. Ara bé, per a tipus de dades complexos cal estudiar-los prèviament i modelar-los amb les mateixes eines matemàtiques amb les que es modelen els SGBD. En els SGBD podem trobar tipus de dades complexos que han rebut una gran atenció, com per exemple els històrics mitjançant intervals temporals o els sistemes d'informació geogràfica. Observant aquestes condicions, es conclou que les sèries temporals necessiten un estudi similar.

2. Estat de l'art

En aquest capítol s'estudia el context i l'estat actual de la recerca en emmagatzematge i gestió de sèries temporals. En tractar-se d'una recerca entre dos àmbits es divideix aquest capítol en dos apartats.

En un primer apartat s'estudien les sèries temporals i les seves aplicacions, principalment pel que fa a adquirir-les, emmagatzemar-les i analitzar-les. En un segon apartat s'estudien els sistemes de gestió de bases de dades, principalment les teories que els formalitzen entre les quals hi destaca el model relacional.

2.1. Sèries temporals

Una sèrie temporal és un conjunt de valors cadascun dels quals té associat un instant de temps diferent. Tradicionalment s'anomenen sèries temporals tot i que també s'accepta la denominació de seqüències temporals, per exemple a [36].

Les sèries temporals s'emmarquen dins l'àmbit més genèric del que es coneix com a *dades temporals*. Les dades temporals són col·leccions de dades arbitràries que estan associades a la dimensió temps. Dins del concepte de dades temporals s'hi encabeixen col·leccions de dades de diversa natura. En funció de com un valor queda vinculat amb el temps, es poden diferenciar dues categories, [1].

1. La primera la formen les sèries temporals tal i com s'han definit prèviament, en la qual la dada està associada a un instant de temps.
2. La segona, que anomena *bitemporal data*, la formen col·leccions de dades en que cada element té dos atributs temporals: el rang de validesa, que indica l'interval de temps en que la dada és vàlida, i el temps de transacció, que indica quan es va desar la dada a la base de dades.

Aquestes dues categories de dades temporals, tot i tenir aspectes en comú, no poden ser tractades amb els mateixos sistemes, [65].

Les sèries temporals s'utilitzen en camps molt diversos i amb objectius molt diferents. L'ús generalitzat és per a l'anàlisi i la comprensió del comportament temporal de variables. L'evolució d'una sèrie temporal es pot representar amb un model. Aquests models, en l'àmbit de l'enginyeria, permeten realitzar tasques relacionades amb validació de dades, diagnòstic i prognòsis. Per exemple, trobem aplicacions de sèries temporals en el

2. Estat de l'art

camp de l'avaluació de la degradació de components, [83], anàlisi de l'estat dels sensors d'un vaixell, [61], validació i reconstrucció de dades en xarxes de distribució d'aigua, [63], classificació de valors econòmics, [27], optimització de la planificació semafòrica, [48], estimació del temps de viatge en autopistes, [70], o transmissió d'informació en xarxes de sensors, [39, 81].

Aquest apartat té com a objectiu mostrar l'estat de l'art dels principals processos vinculats en el treball amb sèries temporals. A tal efecte s'organitza en tres subapartats.

El primer subapartat se centra en l'adquisició de dades. El primer requeriment d'una sèrie temporal és l'adquisició de dades. Els sistemes de monitoratge s'encarreguen de recollir dades dels sensors, periòdicament o en base a esdeveniments. Els problemes que es donen durant l'adquisició generen defectes específics en les sèries temporals que cal analitzar i tractar convenientment.

El segon subapartat tracta de l'anàlisi de sèries temporals, que és la formalització de les tècniques que s'utilitzen per extreure informació. A vegades aquesta extracció també es coneix com descobriment de coneixement i s'emmarca dins de l'intel·ligència artificial.

El tercer subapartat es dedica als sistemes d'emmagatzematge de sèries temporals. L'emmagatzematge de les dades i la implementació de les tècniques d'anàlisi té lloc en els sistemes de gestió de bases de dades. Aquests s'encarreguen de l'organització correcta de la informació i de respondre a les operacions de consulta. Les sèries temporals necessiten un tractament específic per part d'aquests sistemes.

2.1.1. Adquisició de sèries temporals

Els sistemes de monitoratge són una part important d'interacció entre un procés i els usuaris, entenent com a procés qualsevol sistema físic, químic, ambiental, etc. del qual es pugui recollir informació continuada, ja sigui de forma periòdica o en funció d'esdeveniments. Principalment, aquests sistemes s'encarreguen de recollir dades, conèixer l'estat actual del procés i informar a l'usuari. Els sistemes de monitoratge constitueixen la part principal dels sistemes SCADA (*Supervisory Control And Data Acquisition*). Un SCADA és el sistema encarregat de recollir i centralitzar les dades de manera periòdica en el temps.

El monitoratge es pot dividir en diferents blocs principals. Un monitor adquireix dades dels sensors. Les dades poden ser valors de mesures o estats del procés adquirits com a esdeveniments. Fent referència a la classificació de dades temporals de Aßfalg, [1], en general les mesures es poden entendre com a sèries temporals i els esdeveniments com a dades bitemporals.

En el cas de sistemes controlats o automatitzats, les dades adquirides poden ser utilitzades per comandar o modificar el funcionament del procés. Aleshores s'incideix en diferents nivells des de llaços de control modificant directament un accionament, fet que no sol ser habitual ja que els llaços de control solen realitzar-se els sistemes electrònics

2. Estat de l'art

que resideixen prop dels sistemes controlats, fins a gestió de modes de funcionament i coordinació entre màquines, fet més habitual.

L'ús generalitzat dels sistemes de monitoratge és el de proporcionar informació de l'estat actual del procés. També disposen de la possibilitat de generar alarmes senzilles com per exemple que no s'han pogut adquirir les dades o que el sensor ha assolit un valor crític. Per a usuari ens referim tant a un usuari humà com a un altre sistema supervisor dotat amb intel·ligència artificial.

Per a càlculs més complicats amb les dades, els sistemes de monitoratge utilitzen sistemes de gestió de bases de dades (SGBD). Mitjançant els SGBD, s'emmagatzemen les dades en bases de dades i posteriorment l'usuari les consulta per observar els històrics o per obtenir informació i elaborar coneixement a partir de les dades emmagatzemades.

Fins ara s'ha presentat una visió centralitzada de l'adquisició de dades. Ara bé, els sistemes de monitoratge internament poden tenir estructura distribuïda quan els sensors tenen suficient capacitat de processament, com per exemple les xarxes de sensors. En aquests casos els monitors cedeixen parts al sensors, sobretot pel que fa als SGBD que passen a tenir un paper més rellevant en la comunicació.

Un dels camps recents on l'adquisició de sèries temporals hi juga un paper fonamental és el de les xarxes de sensors. L'abaratiment del maquinari permet monitorar el procés amb grans quantitats de sensors intel·ligents, [39, 81], els quals tenen processador i ràdio incorporats però tenen recursos limitats pel que fa a transmissió, energia i processament i estan sotmesos a la incertesa dels sensors. Així doncs, el problema de les xarxes de sensors rau en estudiar l'ús eficient d'aquests recursos, pel qual actualment trobem dues propostes. Una solució consisteix en transmetre la informació a un node central comprimint-la tant amb agregacions (estadístics) com amb aproximacions, [22]. Una altra solució consisteix en tenir les dades distribuïdes en diferents sensors i decidir com s'ha de resoldre cada consulta tenint en compte que el processament local és més barat que la comunicació, [5, 32, 81].

Problemes en el monitoratge

Els sistemes de monitoratge habitualment presenten problemes derivats de la recollecció de dades. Principalment distingim tres problemes.

1. El primer problema és la gestió d'una quantitat enorme de dades.

Un sistema de monitoratge recull una gran quantitat de dades. Ara bé, l'usuari només en pot observar una petita part sincronitzat (*online*) amb el procés i les dades emmagatzemades esdevenen massa grans per a ser processades posteriorment, [43]. No obstant, les dades han de ser analitzades ja que contenen informació interessant per a les aplicacions de les sèries temporals descrites a l'apartat anterior. S'observa que en el context de monitoratge les dades recollides es poden considerar com a sèries temporals ja que abstractament són una col·lecció de mesures.

2. Estat de l'art

2. El segon problema és el de la necessitat de censurar les dades, és a dir validar que les dades siguin correctes i en cas contrari rebutjar-les o reconstruir-les.

Quevedo i altres., [63], mostren la quantitat d'informació que hi ha en els sistemes complexos de telecontrol. Aquesta informació s'obté de diversos sensors distribuïts pel camp de mesura. En el moment de recollecció de dades apareixen dos problemes: valors que en un instant de temps prefixat no s'han pogut recollir i valors que són falsos. En el procés de gestió de dades no es poden emmagatzemar les dades amb aquests dos tipus de problema ja que aleshores els registres històrics serien inconsistents. Així doncs, cal comprovar que les dades emmagatzemades són correctes, mitjançant un procés de validació, i modificar-les en el cas que siguin falses, mitjançant un procés de reconstrucció que estimi els valors correctes. Per exemple, Quevedo i altres. apliquen aquests processos de validació i reconstrucció a xarxes de distribució d'aigua.

3. El tercer problema es dóna quan el període de mostreig no és regular, és a dir que les dades no es recullen de manera uniforme en el temps, però les aplicacions no ho contemplen o volen treballar amb dades a intervals regulars, també anomenat dades equi-espaiades.

Una causa de la irregularitat es deu a que els sistemes de monitoratge informàtics sovint no són capaços de complir amb exactitud el temps de mesura sinó que presenten una certa variació, ja sigui deguda a retards en els sensors, les comunicacions o la planificació del monitoratge amb altres tasques concurrents del sistema operatiu. Aquesta causa, però, es pot atenuar si els sensors envien el temps de mesura juntament amb el valor mesurat. Aleshores, el problema recau en la sincronització dels rellotges dels sensors.

Una altra causa es deu a que l'adquisició de dades prové de processos sotmesos a sistemes de control, els quals prenen el control de l'adquisició de dades. És a dir, el sistema de monitoratge ha d'obeir a les restriccions de temps imposades pels llaços de control. Aquestes restriccions són especialment crítiques en els sistemes de control en temps real ja que, aleshores, el sistema de monitoratge no pot imposar restriccions de temps diferents de les que s'han calculat per als llaços de control. Lozoya, Velasco i Martí, [51], mostren que s'ha de vigilar amb les entrades i sortides de les tasques periòdiques als sistemes en temps real. L'actuació dels sistemes de control es degrada quan no es té en compte que les operacions d'entrada i sortida estan subjectes a fluctuacions degudes al mostreig i a latències. Aquest problema afecta als sistemes de monitoratge en dues vessants. Per una banda, els sistemes de monitoratge tenen una part de l'adquisició controlada per les aplicacions de control en temps real i per tant el període de mostreig resultant que veu el monitor no és regular. Per altra banda, les aplicacions que analitzen les dades obtingudes del monitoratge poden veure com la seva actuació es degrada si no consideren que l'adquisició de dades és irregular, el qual és similar a la regressió que s'observa, [51], quan en el disseny d'un controlador discret es considera que es mostreja i s'actua

periòdicament però en la implementació amb un sistema en temps real aquest pot fluctuar la periodicitat.

En conclusió, per tal de gestionar la complexitat derivada de la recollida de dades i també la complexitat de les consultes posteriors per part de l'usuari, els sistemes de monitoratge es recolzen en sistemes de gestió de bases de dades per gestionar l'emmagatzematge de les dades i la recuperació d'informació.

2.1.2. Anàlisi de sèries temporals

L'anàlisi de sèries temporals consisteix en l'aplicació de metodologies i d'algoritmes que permeten tasques com per exemple l'extracció de característiques o obtenció de models. Aquestes tècniques es recullen en el que es coneix amb el nom de mineria de sèries temporals (*time series data mining*). La mineria de dades, en la qual s'inscriu la de sèries temporals, és l'estudi d'algoritmes específics per a extreure patrons de comportament de les dades i s'inclou com un pas del procés general de descobriment de coneixement a les bases de dades (*knowledge discovery in databases*), [29, 49].

Actualment, les sèries temporals es consideren com un dels deu problemes prioritaris en la mineria de dades, [80]. Tal com esmenta Fu, [30], en un article recent, la recerca en mineria de sèries temporals s'ha incrementat en la darrera dècada. L'objectiu principal és reduir la mida de les sèries temporals per tal de disminuir el temps de processat de les dades. Fu resumeix l'estat actual de la mineria de sèries temporals de forma exhaustiva i conclou que encara queden molts problemes per investigar i resoldre. La recerca en tasques de mineria ha estat intensa però es necessita millorar la representació de sèries temporals, ja que es considera el pas que redueix la mida de les dades.

Segons Keogh i Kasetty, [40], les quatre tasques que centren l'atenció de la recerca actual de sèries temporals són l'indexat (*indexing*), que treballa amb una estructura comprimida de les dades; l'agrupament (*clustering*), que agrupa les dades segons la similitud entre elles per tal de descobrir patrons; la classificació (*classification*), que etiqueta les dades segons les característiques que presentin; i la segmentació (*segmentation*), que parteix una sèrie temporal en subsequències. A més, Keogh i Kasetty comparen alguns algoritmes experimentals duts a terme en aquests camps per diversos autors. Recomanen a la comunitat de mineria de sèries temporals que segueixi el seu estudi com a punt de referència per avaluar el rendiment d'algoritmes similars.

Un pas comú previ a les quatre tasques anteriors és el de representació de la sèrie temporal. Les sèries temporals són discretes, són valors en punts de temps discrets, i la representació és el model de funció que aproxima la sèrie temporal a la seva naturalesa contínua original. La mineria de sèries temporals aprofita la representació per reduir la mida de les sèries temporals. La representació de sèries temporals a trossos lineals (PLR, *Piecewise Linear Representation*), [42, 43], és la més habitual actualment per ser més propera als usuaris ja que la visió de l'ésser humà segmenta les corbes en línies rectes. Després de definir la PLR, Keogh i altres., [44, 45], exploren altres representacions

de sèries temporals per tal de reduir la mida d'una sèrie temporal i poder-la indexar més fàcilment. Proposen dues tècniques eficients en el càlcul: la *Piecewise Aggregate Approximation* i la *Adaptive Piecewise Constant Approximation*, ambdues basades en la representació a trossos constants de la sèrie temporal. D'aquestes dues tècniques, Keogh i altres. conclouen que mantenen una bona aproximació a la sèrie temporal i que a més tenen molt menys cost de càlcul que altres de més complicades, com ara la *Discrete Fourier Transform*, la *Singular Value Decomposition* o la *Discrete Wavelet Transform*.

2.1.3. Emmagatzematge i gestió de sèries temporals

Els sistemes de gestió de bases de dades (SGBD) són els sistemes informàtics que s'encarreguen d'emmagatzemar informació i de permetre a l'usuari consultar-la. A la secció 2.2 es descriu com es formalitzen els SGBD, en aquest apartat ens centrarem en les necessitats que tenen les sèries temporals dels SGBD.

Les sèries temporals es diferencien d'altres tipus de dades en que els seus valors són dependents d'una variable: el temps. Com a conseqüència, qualsevol SGBD que hi vulgui tractar no ho pot fer de manera independent pels valors i pel temps; ha de conservar la coherència temporal.

Per poder aplicar les tècniques d'anàlisi de les sèries temporals de manera eficient cal disposar de SGBD específics. Durant l'última dècada, el maquinari informàtic ha millorat tant des del punt de vista tecnològic com de l'econòmic, [22], el qual ha facilitat l'adquisició de dades, per exemple amb xarxes de sensors, i alhora ha ampliat la capacitat per emmagatzemar les dades. Per tant, el volum de dades a tractar en els SGBD cada cop esdevé més crític.

En els SGBD, el problema de grans quantitats de dades també es troba en altres camps com demostren Mylopoulos i altres., [55], sobre la necessitat de grans bases de dades de coneixement. Els SGBD que tracten amb aquestes dades s'anomenen *very large databases* (VLDB) i han de construir, accedir i gestionar la quantitat de dades de manera eficient.

Ogras i Ferhatosmanoglu, [60], consideren que les aproximacions que fan les VLDB estan pensades per a bases de dades estàtiques. No obstant, observen que les sèries temporal normalment són dinàmiques, és a dir de naturalesa contínua i de mida no fitada. Conseqüentment, conclouen que les solucions tradicionals, les quals analitzen a posteriori i sense tenir en compte l'ordre, no es poden aplicar a causa de l'arribada seqüencial i contínua de les dades. Com a solució proposen resumir dinàmicament les sèries temporals amb les tècniques de compressió que s'apliquen en altres aplicacions on hi ha bases de dades grans.

Dreyer, Dittrich i Schmidt, [25], proposen desenvolupar SGBD que implementin operacions específiques per les sèries temporals, aleshores els anomenen sistemes de gestió de bases de dades per sèries temporals (SGST, *time series database management systems*). Consideren que els altres SGBD no són adequats per tractar sèries temporals, tot i que

2. Estat de l'art

després de comparar els SGBD per dades temporals i els SGST, [65], troben que hi ha aspectes comuns entre els dos sistemes. Els SGST estan optimitzats per gestionar les dades segons les operacions de temps i rotació, les quals són molt comunes en la gestió de les sèries temporals. A més també cal controlar el creixement de la base de dades i la consulta ha de ser flexible i d'alta velocitat, [7]. No obstant, fins a on coneixem, les propietats d'un model de SGST no s'han investigat més enllà ja que la recerca s'ha concentrat en tasques de mineria de dades. Per exemple Last i altres., [49], estudien una metodologia general per descobrir coneixement en els SGST, tant pel que fa a patrons temporals com a regles temporals, i breument noten l'existència del model [25] pels SGST.

Altres estudis proposen tractar les sèries temporals com a tipus que tenen ordre, per exemple seqüències o matrius.

Seshadri, [67], proposa que les sèries temporals són un subconjunt de les seqüències i per tant el model i les operacions per les seqüències, [68], serveixen per les sèries temporals. Bonnet, Gehrke i Seshadri, [5], utilitzen el model de seqüències en SGBD distribuïts per xarxes de sensors, aleshores l'estratègia de comunicació inclou agregacions de les sèries temporals en els sensors, [23]. També es relaciona el model de seqüències de les sèries temporals amb els *data streams*, [3, 38, 60]. Els *data streams* són dades que arriben contínuament i amb ordre temporal i es modelen com una seqüència on només s'hi poden afegir elements. Aleshores les consultes poden ser contínues, és a dir cada cop que arriba una dada nova s'actualitza incrementalment la informació. Per les sèries temporals s'utilitza en el càlcul de correlacions i prediccions de forma incremental, [82], i en la cerca de patrons, [4].

En els SGBD per matrius *arrays* destaquen els anomenats sistemes de gestió de bases de dades científiques, camp en el qual les sèries temporals hi tenen un paper de primer ordre, [85]. Stonebraker i altres., [74], estudien les necessitats d'aquests sistemes sobretot en l'àmbit de la ciència. Kersten i altres., [47], proposen un sistema molt semblant però a més integren el seu llenguatge, anomenat SciQL (*SQL for science applications*), amb la sintaxi de SQL (*Structured Query Language*, vegeu apartat 2.2.1). Zhang i altres., [85], exemplifiquen detalladament l'ús de SciQL en les sèries temporals per a algunes de les seves propietats: regularitat, interpolació i cerca de correlacions.

Implementacions actuals

Hi ha hagut varies implementacions de sistemes específics per a sèries temporals. Algunes són només l'aplicació d'un algoritme d'anàlisi per un problema concret de sèries temporals però altres són més elaborades i es defineixen com a SGBD per a sèries temporals. En aquest apartat resumim algunes aplicacions que considerem que implementem conceptes dels SGST.

Calanda Dreyer, Dittrich i Schmidt, [25], proposen els requeriments de propòsit específic que han de complir els SGST i basen el model en quatre elements estructurals

2. Estat de l'art

bàsics: esdeveniments, sèries temporals, grups i metadades, a banda de les bases de dades per sèries temporals. Implementen un SGST anomenat Calanda, [24, 26, 27], que té operacions de calendari, pot agrupar sèries temporals i respondre consultes simples i ho exemplifiquen amb dades econòmiques. A [65] es compara Calanda amb els SGBD temporals que operen amb sèries temporals.

T-Time Abfalg, [1], mostra un sistema que pot cercar similituds calculades com a distàncies entre sèries temporals. Principalment, dues sèries temporals es marquen com a similars si la seva distància és menor que un llindar en cada interval. A partir d'aquest mètode dissenya algoritmes eficients que implementa en un programa anomenat T-Time, [2].

iSAX Camerra i altres., Shieh i Keogh, [7, 69], estudien l'anàlisi i l'indexat de col·leccions massives de sèries temporals. Descriuen que el problema principal del tractament rau en l'indexat de les sèries temporals i proposen mètodes per calcular-lo de manera eficient. El mètode principal que proposen està basat en l'aproximació a trossos constants de la sèrie temporal, [44]. Ho implementen en una estructura de gestió de dades que anomenen *indexable Symbolic Aggregate approxImation* (iSAX), [41]. Les representacions de sèries temporals que s'obtenen amb aquesta eina permeten reduir l'espai emmagatzemat i indexar tant bé com altres mètodes de representació més complexos.

TSDS Weigel i altres., [78], noten la necessitat de mostrar les dades en tot el seu rang temporal i no només en un subconjunt com normalment s'ofereixen. Desenvolupen el paquet informàtic *Time Series Data Server* (TSDS), [79], a on es poden introduir les dades de sèries temporals per posteriorment consultar-les per rangs temporals o aplicant-hi filtres i operacions.

RRDtool RRDtool, [58], és un SGBD molt usat per la comunitat de programari lliure. Projectes en diversos camps l'utilitzen com a SGBD, en els quals hi ha sistemes de monitoratge professionals, també en l'àmbit de programari lliure, com Nagios/Icinga, [37, 56], o el Multi Router Traffic Grapher (MRTG), [59]. Aquests monitors transfereixen a RRDtool la responsabilitat de gestionar l'emmagatzematge i d'operar amb les dades, i així es poden centrar en l'adquisició de dades i la gestió d'alarmes. En l'evolució de RRDtool hi ha dues millores destacables. En primer lloc, Oetiker, [57], va separar el sistema de gestió de RRDtool de MRTG i el va dissenyar amb una estructura característica de Round Robin. En segon lloc, Brutlag, [6], va estendre RRDtool amb algoritmes de predicció i detecció de comportaments aberrants.

Actualment, s'està estudiant l'eficiència i rapidesa de RRDtool a processar sèries temporals. Carder, [8], ha dissenyat una aplicació, *rrdcached*, que millora el rendiment de RRDtool amb la qual s'aconsegueix fer funcionar simultàniament sistemes amb grans quantitats de bases de dades RRDtool, [62]. Markovic i Vandamme, [54], han dissenyat una adaptació de RRDtool anomenada *JRobin*. Finalment, és destacable l'ús emergent de RRDtool en entorns d'experimentació, com és el cas de

2. Estat de l'art

Zhang i Figueiredo, [84], i Chilingaryan i altres., [9], que hi emmagatzemen dades experimentals per posteriorment predir o validar-les.

Cougar Fung, Sun i Gehrke, Gehrke i altres., [31, 33], proposen Cougar com un SGBD per xarxes de sensors (*sensor database systems*). El sistema té dues estructures, [5]: una basada en relacions per les característiques dels sensors i una basada en seqüències per les dades dels sensors, les quals són sèries temporals. Les consultes es processen de manera distribuïda: cada sensor és un node amb capacitat de processament que pot resoldre una part de la consulta i fusionar-la amb les altres. D'aquesta manera es minimitza l'ús de comunicacions però l'estructura i estratègia de comunicació dels nodes esdevé una part crítica a configurar, [23].

TinyDB Un altre prototip de SGBD per xarxes de sensors desenvolupat paral·lelament a Cougar és TinyDB, [52, 53]. A part de les característiques descrites per Cougar, aquest sistema modifica i s'implica en parts del procés d'adquisició de les dades com és el temps, la freqüència o l'ordre de mostreig. Per exemple donada una consulta que vol correlacionar les dades de dos sensors, el sistema indica als sensors implicats que han d'adquirir amb la mateixa freqüència.

SciDB Stonebraker i altres., [74], estudien els SGBD científiques amb models de dades basats en matrius. Estan desenvolupant SciDB, [76], un SGBD productiu i optimitzat per treballar amb matrius.

SciQL Kersten i altres., [47], descriuen SciQL, un llenguatge per a SGBD científiques basades en matrius. Hi ha un prototip en desenvolupament de SciQL, [46].

2.1.4. Conclusió

Els SGST actuals bàsicament resolen alguns problemes d'anàlisi de sèries temporals. Però no solen atendre la relació entre la base de dades i el sistema de monitoratge, és a dir la manera com s'adquireixen les dades. En aquest pas intermig hi ha un sèrie de problemes, com per exemple forats, dades falses o irregularitat en els temps de mostreig, que cal gestionar correctament. Concretament un dels problemes que no s'atén és el de mostreig irregular ja que es considera que les mostres estan a intervals regulars (equi-espaiades) encara que els sistemes de monitoratge informàtics sovint no són capaços de complir-ho amb exactitud sinó que presenten una certa variació en els temps de mesura.

RRDtool n'és una excepció ja que, per ser un sistema productiu, el processament de dades i emmagatzematge és més proper als sistemes de monitoratge. No obstant, està centrat en un tipus de dades particulars, les magnituds i els comptadors, i no té tantes operacions generals per les sèries temporals com els altres SGST.

També Cougar i TinyDB que exploren l'encaix dels SGBD en entorns distribuïts de xarxes de sensors. Proposen noves estratègies de comunicació amb l'objectiu d'ajustar el consum d'energia.

SciQL, un model recent per SGBD basat en matrius, és el que més es pot considerar com a SGST, ja que s'està desenvolupant per complir-ne algunes propietats.

2.2. Sistemes de gestió de bases de dades

Segons Date, [12], “una base de dades és un contenidor informàtic per a una col·lecció de dades”. El sistemes informàtics que tracten amb bases de dades s'anomenen sistemes de gestió de bases de dades (SGBD, *Data Base Management Systems*) i tenen els objectius d'emmagatzemar informació i permetre consultar i afegir aquesta informació per part dels usuaris. Per complir aquests objectius, els SGBD ofereixen a l'usuari diferents operacions com per exemple crear una base de dades, afegir dades o consultar informació a partir de les dades emmagatzemades.

Els SGBD es poden descriure mitjançant teories matemàtiques que reben el nom de model de dades, per tant es poden veure els SGBD com una implementació d'un model de dades. Segons Date, “un model de dades és una definició abstracta, auto continguda i lògica dels objectes, de les operacions i de la resta que conjuntament constitueixen la màquina abstracta amb la que els usuaris interaccionen. Els objectes permeten modelar l'estructura de les dades. Les operacions permeten modelar el comportament”. Ara bé, Date avisa que el concepte model de dades també s'usa per a definir una estructura persistent de dades concreta i, per tant, cal distingir adequadament entre els dos conceptes. Tal com fa Date, en aquest document parlarem de model de dades, o simplement de model, en el primer sentit de màquina abstracta.

Un model de SGBD que ha s'ha consolidat i esdevingut un referent és el model relacional (*relational*). L'èxit d'aquest model és degut a que es fonamenta en teories matemàtiques consolidades: la lògica de predicats i la teoria de conjunts, [12].

Date, [12], diferencia amb detall els conceptes de model i d'implementació. El model d'un SGBD és el model matemàtic tal com s'ha descrit anteriorment, en canvi un SGBD és la implementació d'un model de dades, per exemple *PostgreSQL*, [34]. En aquest context una base de dades és una instància d'un SGBD, per exemple la base de dades dels estudiants.

Aquesta diferència entre implementació i model aporta independència de dades (*data independence*), [16]. En altres paraules, els models no han de tenir detalls d'implementació ni parlar d'objectius de rendiment. Date i Pascal, [21], detallen algunes confusions actuals sobre la independència entre el model i la implementació.

2.2.1. Sistemes relacionals

El model relacional va ser proposat per Codd, [10], per a formalitzar els SGBD, els quals quan es basen en aquestes teories s'anomenen relacionals (SGBDR). A partir de llavors

2. Estat de l'art

els SGBDR han anat evolucionat fins a aconseguir una gran solidesa, amb Date, [12, 13, 16], com a principal divulgador.

Les implementacions més populars de SGBDR són les que s'anomenen *SQL* ja que tenen en comú el llenguatge *Structured Query Language*. Ara bé, els SGBD *SQL* es desvien considerablement del model relacional: permeten files duplicades, tenen ordre en les columnes, permeten valors nuls, etc, sent aquest últim un tema actualment discutit, [11].

Les diferències entre els SGBD *SQL* i el model relacional han contribuït a que hi hagi hagut una sèrie de malentesos i errors, alguns dels quals han estat avaluats i desmentits per Date i Pascal en diverses publicacions, [13, 21].

Date, [13, cap. 2], considera que no hi cap implementació comercial que segueixi fidelment el model relacional, tot i que esmenta algunes implementacions prometedores com *Dataphor* o la seva proposta tecnològica *TransRelational*, [15]. A banda, també cal destacar *Rel*, [77], com un SGBDR bastant consolidat.

Actualment Date i Darwen, [18], estan treballant en el '*Third Manifesto*' com a proposta per a obtenir SGBDR purament relacionals. Destaquen que, en el model relacional, els tipus de dades i les relacions són necessaris i suficients per representar qualssevol dades a nivell lògic. Defineixen dos principis bàsics dels SGBDR: l'*Information Principle* o *The Principle of Uniform Representation*, [16], segons el qual una base de dades només conté variables relacions, i el principi d'ortogonalitat entre la teoria de tipus i el model relacional, [13, cap. 6], segons el qual relacions i tipus de dades són independents i per tant els atributs de les relacions admeten qualsevol tipus. Segons aquest punt de vista, els tipus de dades són el conjunt de coses de les que podem parlar mentre que les relacions són proposicions certes sobre aquestes coses.

En la proposta per a obtenir SGBDR purament relacionals Date i Darwen, [17, 19], classifiquen com a *D* els llenguatges que segueixin els principis del *Third Manifesto*. Concretament, com a exemple d'un llenguatge *D* estan definint les regles de *Tutorial D*, que ha de servir pels estudis del model relacional a nivell acadèmic. Aquest llenguatge ja s'utilitza en alguns SGBDR, com per exemple a *Rel*, [77].

El model relacional ha incorporat conceptes d'altres disciplines. En destaca sobretot la incorporació de conceptes dels models d'orientació a objectes com és el cas de l'herència. Aleshores s'entén que els SGBDR també es puguin anomenar SGBD objecte/relacionals (*object/relational*), [14]. Tot i així, Date, [13, cap. 6], manifesta i avisa de l'ús de la mateixa terminologia amb significat diferent entre el model relacional i l'orientació a objectes, sobretot pel que fa als termes valor i variable. La seva hipòtesi a aquestes diferències és que el model relacional és un model de dades i el model d'orientació a objectes és més proper a un model d'emmagatzematge.

Extensió del model

El model relacional ha evolucionat però no es considera que hi hagi hagut cap revolució des de la seva aparició, [13, cap. 19]. Consideren que el model relacional és bastant complet i que segueix evolucionant en la comprensió de les teories i els conceptes que hi intervenen, com per exemple la recent àlgebra relacional 'A', [17, ap. A]. En aquest context d'evolució, es contemplen les investigacions que poden estendre el model relacional, és a dir, aconseguir abstraccions més generals de les dades, [13, cap. 25].

En el sentit d'extensió, també cal contemplar la definició de nous tipus de dades, els quals estenen els SGBD en funcionalitat. Aquests nous tipus de dades poden afegir estructures i operadors que ja siguin expressables amb l'àlgebra relacional. No obstant, un bon model d'un tipus de dades serveix per augmentar el nivell d'abstracció en el tractament dels conceptes relacionats amb aquestes dades, [20].

Com s'ha dit anteriorment, la teoria de tipus i el model relacional són ortogonals: el model relacional requereix que hi hagi un 'sistema' de tipus de dades però diu molt poc de la naturalesa d'aquest sistema, si bé el model relacional defineix que com a mínim hi ha d'haver el tipus booleà i el tipus relació, [18]. Pel que fa a implementar el tipus de dades en els SGBD, els quals aleshores també s'anomenen SGBD objecte/relacionals, destaquen les primeres propostes fetes per Stonebraker, [71], per tal que els usuaris puguin definir els seus propis tipus de dades i les de Seshadri, [66], que estudia la definició de tipus de dades complexos per tal que es puguin tractar eficientment.

2.2.2. Altres sistemes

Les crítiques als SGBDR, sobretot les degudes als SGBD *SQL*, han contribuït a voler explorar altres models de SGBD, [73]. Aquests models presenten diferents maneres de representar les dades: llistes, seqüències, enllaços, matrius, etc.

Tot i així, Date, [13, cap. 21–25], considera que els nous models de SGBD, a vegades anomenats post-relacionals, no estan fundats tant sòlidament en teories matemàtiques i la lògica de predicats com el model relacional i pronostica que ens els propers cent anys els SGBD encara estaran basats en el model relacional. Considera la possibilitat, tot i que remota, que es pugui definir un model més potent que el relacional però que no hi ha cap indicatiu que cap definició dels nous model tingui la mateixa potència que el relacional. Per tant, aconsella que per ara els SGBD no s'allunyin del model relacional.

Recentment ha aparegut un nou corrent en l'àmbit dels SGBD que s'anomena NoSQL (*Not Only SQL*) amb l'objectiu de sobrepassar les limitacions dels SGBDR, [28, 72]. A l'espera que Date valori aquest corrent, cal tenir present els seus apunts sobre sistemes relacionals contra sistemes no relacionals, [13, part 7], sobretot pel que fa al concepte de l'*Information Principle* i que SQL no és un bon referent pels SGBDR. És a dir, s'ha d'entendre NoSQL com una crítica a les implementacions comercials actuals del model relacional, una crítica que pot estar motivada per l'ús de SQL per part d'aquest

productes. No es pot entendre, però, el NoSQL com una crítica al model relacional ja que els objectius parlen de millorar el rendiment dels SGBD, cosa només atribuïble a les implementacions però no al model. Precisament, actualment del model relacional destaca la proposta de Date i Darwen d'un llenguatge, *Tutorial D*, que no és SQL.

El corrent de NoSQL també critica l'adequació dels productes actuals. Els SGBD NoSQL apunten els SGBDR actuals per voler ser *one size fits all*, [73, 75], però que cada aplicació té els seus requisits i per tant una mateixa implementació no pot ser bona per a tots el camps. En aquesta mateixa línia els models pels SGBD prenen més sentit que mai ja que permeten mantenir una definició comuna per a moltes implementacions.

Segons es desprèn de Date, [13], fins a l'actualitat només hi ha hagut un model consolidat pels SGBD: el model relacional. Ara, en el corrent NoSQL també es parla de nous models de SGBD, [28, 74]. Date ha avaluat que alguns nous models recuperen intents fallits en el passat tot i que es poden representar amb el model relacional, per exemple els SGBD XML basats en estructures d'arbre, [13, cap. 14], o els ODMG basats en objectes, [13, cap. 27]. Tot i així, en un futur cal estar atents per si alguns d'aquests models joves de SGBD arriben a consolidar-se i poden esdevenir tant potents com el model relacional.

2.2.3. Dades temporals

Una de les extensions importants del model relacional s'ha produït amb l'estudi de les dades temporals, [20]. Amb el model de dades temporals basat en el model relacional s'obtenen SGBDR capaços d'emmagatzemar i consultar dades històriques.

El model de les dades temporals, [20], es basa en relacions i intervals per representar les dades històriques. Cada relació s'estén amb un atribut que és un interval temporal que indica el rang temporal de validesa de les proposicions. A més el model també defineix les operacions necessàries per tractar amb les dades temporals. Aquestes operacions són extensions de l'àlgebra relacional. El principal objectiu del model de dades temporals és poder tenir suport per a les dades històriques en els SGBDR. Date, [13, cap. 28], compara aquest model per dades temporals amb altres aproximacions que s'han fet no basades en el model relacional.

El model de dades temporals representa les dades històriques amb el temps vàlid i temps de transacció, [20, cap. 15], el que es coneix com a dades bitemporals. Així doncs, els SGBD per dades bitemporals no es consideren adequats com a SGBD per sèries temporals ja que els primers estan pensats per a històrics, es descriuen amb intervals temporals, i els segons per anàlisi d'observacions seqüencials, es descriuen amb instants temporals. Schmidt i altres., [65], arriben a aquesta conclusió després de comparar els SGBD bitemporals amb els de sèries temporals, tot i que per a desenvolupaments futurs observen que hi aspectes temporals comuns entre els dos sistemes i es pregunten si es podran trobar sistemes que els englobin a tots dos o cadascun necessitarà sistemes específics.

2.2.4. **Conclusió**

Actualment l'àmbit informàtic de SGBD se centra en les implementacions, com ho demostra el nou corrent NoSQL concentrat en trobar models d'implementacions que tinguin bon rendiment. A tal efecte la recerca es concentra en temes de garantia de propietats ACID (*atomicity, consistency, isolation, durability*), d'optimització de consultes, d'emmagatzematge de grans volums de dades, de consultes via web, de distribució de bases de dades, de reduir la despesa en energia, etc. [72, 75], la qual cosa és excel·lent per a disposar d'un SGBD adequat a cada aplicació. Härder, [35], descriu diferents models d'implementació per als SGBD, ja que indica que per obtenir bon rendiment la implementació d'un SGBD s'ha d'estudiar per cada aplicació. És més, una implementació d'un SGBD que vulgui obtenir un bon rendiment en una determinada aplicació potser no pot implementar el model de dades complet sinó que només una part, com per exemple en els sistemes encastats, [64].

Per altra banda, l'àmbit matemàtic de SGBD, amb el model relacional com a màxim exponent, se centra en els conceptes teòrics, és a dir respon a la pregunta de què són els SGBD. Recerca en millorar-ne la comprensió, en obtenir la màxima potència i facilitat de cara a la gestió de dades per part de l'usuari o en obtenir nous models. Tot i així, actualment encara no s'ha trobat cap altre model que tingui la mateixa potència que el relacional. Cal destacar que tot i que el model relacional té conceptes madurs i consolidats, i que a més han tingut èxit amb els SGBD SQL, s'obre una nova perspectiva amb l'evolució de conceptes que proposa el *Third Manifesto*, especialment amb *Tutorial D* i les implementacions que comencen a prendre cos a nivell acadèmic.

3. Planificació

3.1. Treball realitzat

La tesi de màster, [50], va consistir en l'estudi de *RRDtool*, [58], un sistema de gestió de bases de dades (SGBD) específic per a sèries temporals. Fruit d'aquest estudi es va formalitzar el model de *RRDtool* i es va dissenyar i implementar un prototip software que responia a la formalització. Aquest treball va permetre concloure que:

- Els SGBD aplicats a sèries temporals tenen aspectes propis que els converteixen en objecte d'estudi de *per se*.
- El concepte de multiresolució és interessant en moltes aplicacions reals, especialment quan existeixen restriccions d'espai per emmagatzemar dades.
- El model proposat per a *RRDtool* és susceptible de ser millorat com a mínim en els següents aspectes:
 - Generalització. El model que es va presentar estava fortament lligat a *RRDtool*. Generalitzar el model de forma que encabeixi altres concepcions permetria enriquir-lo.
 - Incorporació d'operacions. El model presentat únicament feia referència a les dades. Era un model de dades. Per completar el model cal també considerar les operacions sobre les dades.
 - Contextualització. Cal interrelacionar i descriure el model el context d'altres models existents per a SGBD. Específicament cal comparar-lo amb el model relacional usat en els SGBD convencionals.
- És convenient estudiar la feina feta per altres autors en l'àmbit de l'emmagatzemat i gestió de dades provinents de sèries temporals.

Arrel de la feina anterior, s'ha treballat en els següents aspectes:

- S'ha realitzat una tasca de recerca bibliogràfica i estudi dels treballs existents en l'àmbit de la gestió i emmagatzemat de dades provinents de sèries temporals. El resultat d'aquesta tasca s'ha reflectit a l'apartat 2 d'aquest mateix document.

3. Planificació

- S'ha estudiat en profunditat el model relacional. Atesa la preeminència d'aquest model en els SGBD actuals, s'ha considerat imprescindible tenir-ne un bon coneixement que permetés estudiar les seves deficiències per la gestió de sèries temporals i en quina forma poden ser superades.
- S'està refent el model de SGBD per sèries temporals. S'ha dividit el model en dues parts ben diferenciades:
 1. La primera és el model general. Aquest defineix la gestió de sèries temporals enteses com a col·lecció de dades mesurades en diferents instants de temps. Es basa en els conceptes de temps, mesura i sèrie temporal.
 2. La segona és el model de multiresolució. Aquest model explica la forma d'emmagatzemar una sèrie temporal amb diferents resolucions temporals.

3.2. Treball futur

Per tal d'assolir els objectius detallats a la secció 1.2, a continuació es proposen les tasques a realitzar. A la figura 3.1 es detalla el pla de treball amb temps estimat.

1. Estudi d'aplicacions de les sèries temporals. Per a assolir l'objectiu 1 estudiarem recerca actual de sèries temporals. Consultarem articles i llibres que tinguin les sèries temporals com a temàtica principal. També cercarem l'existència de programari que tingui en els seus objectius el tractament de sèries temporals.
2. Estudi del model relacional. Per a l'objectiu 2 estudiarem el model relacional com a referent pels models de SGBD. Ens basarem en l'estudi dels llibres de Date, [12, 13, 16]. Usarem *rel*, [77], com a implementació de referència ja que incorpora el llenguatge *Tutorial D*, el qual és utilitzat per Date en els seus exemples.
3. Estudi de la gestió d'interval temporals. Per a l'objectiu 3 estudiarem la recerca que ha conduït a incorporar els interval temporals en els SGBD per a gestionar històrics. Ens basarem en l'estudi que relaciona el model relacional amb els interval temporals de Date, Darwen i Lorentzos, [20].
4. Disseny d'un model de SGBD per sèries temporals. Per a la primera part de l'objectiu 4 dissenyarem un model per als SGBD de sèries temporals formalitzat amb expressions algebraiques. A partir de l'estudi de l'estat de l'art de les sèries temporals, observarem les propietats interessants de ser modelitzades i les operacions que precisen els SGBD per sèries temporals.
5. Disseny d'un model de SGBD multiresolució per sèries temporals. Per a la segona part de l'objectiu 4 dissenyarem un model que contempli la multiresolució de les sèries temporals. Aquest model utilitzarà propietats del model anterior per les

3. Planificació

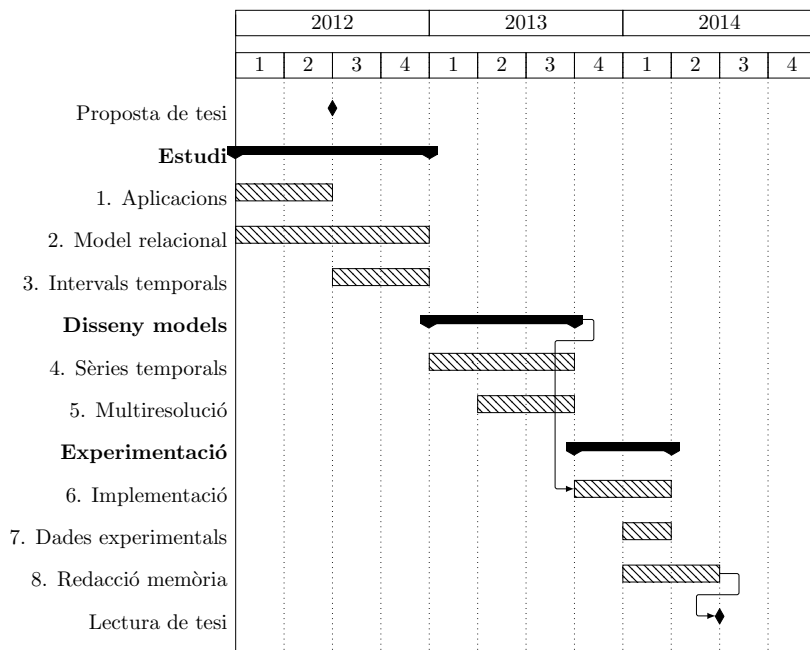


Figura 3.1.: Planificació del treball

sèries temporals. La multiresolució té l'objectiu d'emmagatzemar les sèries temporals de forma compacta, així es preveu que alguns conceptes de la recerca en *data streams* poden prendre-hi sentit.

6. Implementació de referència. Per a l'objectiu 5 s'implementaran els models anteriors utilitzant un llenguatge de programació adequat per a models, com per exemple Python o Prolog. Es prioritzarà la implementació correcta del model enfront a una implementació que contempli un bon rendiment.
7. Experimentació amb dades. Per a complementar l'objectiu 5 es provarà la implementació amb dades experimentals per alguna aplicació concreta.

3.3. Mitjans

La recerca es duu a terme amb el suport de la Universitat Politècnica de Catalunya (UPC) mitjançant una beca FPU-UPC adscrita al departament d'Enginyeria del Disseny i Programació de Sistemes Electrònics (DiPSE).

No es preveu un ús de mitjans més enllà de l'accés als recursos bibliogràfics i d'eines informàtiques de programació i gestió de documentació.

Bibliografia

- [1] Johannes Abfalg. “Advanced Analysis on Temporal Data”. Tesi doctoral. Fakultät für Mathematik, Informatik und Statistik der Ludwig Maximilians Universität München, 19 de maig de 2008. URL: <http://edoc.ub.uni-muenchen.de/8798/> (visitat el 14 d’abril de 2011). (Vegeu les pàgines 4, 8, 9, 15).
- [2] Johannes Abfalg i altres. “T-Time: Threshold-Based Data Mining on Time Series”. A: *Proceedings of the 24th International Conference on Data Engineering*. ICDE’08. Cancun, Mexico: IEEE, abril de 2008, pàgines 1620–1623. (Vegeu la pàgina 15).
- [3] Brian Babcock i altres. “Models and issues in data stream systems”. A: *Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. PODS’02. Madison, Wisconsin: ACM, 2002, pàgines 1–16. DOI: 10.1145/543613.543615. URL: http://infolab.usc.edu/csci599/Fall2002/paper/DML2_streams-issues.pdf (visitat el 17 d’abril de 2012). (Vegeu les pàgines 7, 14).
- [4] Yijian Bai i altres. “Efficient support for time series queries in data stream management systems”. A: *Stream Data Management*. Edició a cura de Nauman Chaudhry, Kevin Shaw i Mahdi Abdelguerfi. Volum 30. The Kluwer International Series on Advances in Database Systems. Kluwer Academic Publishers, 2005. Capítol 6. DOI: 10.1007/b106968. URL: <http://www.cs.ucla.edu/~zaniolo/papers/eslTS.pdf> (visitat el 19 d’abril de 2012). (Vegeu les pàgines 7, 14).
- [5] Philippe Bonnet, Johannes Gehrke i Praveen Seshadri. “Towards Sensor Database Systems”. A: *Proceedings of the Second International Conference on Mobile Data Management*. Volum 1987. MDM’01. Hong Kong: Springer-Verlag, gener de 2001, pàgines 3–14. DOI: 10.1007/3-540-44498-X_1. URL: <http://www.cs.cornell.edu/johannes/papers/2001/MDM2001-sensor.pdf> (visitat el 16 d’abril de 2012). (Vegeu les pàgines 4, 10, 14, 16).
- [6] Jake D. Brutlag. “Aberrant Behavior Detection in Time Series for Network Monitoring”. A: *Proceedings of the 14th Systems Administration Conference*. LISA’00. New Orleans, Los Angeles: USENIX Association, desembre de 2000, pàgines 139–146. URL: <http://www.usenix.org/events/lisa00/brutlag.html>. (Vegeu la pàgina 15).

Bibliografia

- [7] Alessandro Camerra i altres. “iSAX 2.0: Indexing and Mining One Billion Time Series”. A: *Proceedings of the 10th IEEE International Conference on Data Mining*. ICDM'10. Sydney, Australia: IEEE, desembre de 2010, pàgines 58–67. URL: http://www.cs.ucr.edu/~eamonn/iSAX_2.0.pdf (visitat el 15 de març de 2011). (Vegeu les pàgines 14, 15).
- [8] Dale Carder. *RRDtool Scalability, performance in the large scale*. URL: <http://net.doit.wisc.edu/~dwcarder/rrdcache/> (visitat el 22 de març de 2011). (Vegeu la pàgina 15).
- [9] S. Chilingaryan i altres. “Advanced data extraction infrastructure: Web based system for management of time series data”. A: *Journal of Physics: Conference Series*. 17th International Conference on Computing in High Energy and Nuclear Physics (CHEP'09) 219.4 (2010). DOI: 10.1088/1742-6596/219/4/042034. (Vegeu la pàgina 16).
- [10] E. F. Codd. “A relational model of data for large shared data banks”. A: *Communications of the ACM* 13.6 (juny de 1970), pàgines 377–387. DOI: 10.1145/362384.362685. (Vegeu les pàgines 7, 17).
- [11] Christopher J. Date. “A critique of Claude Rubinson’s paper nulls, three - valued logic, and ambiguity in SQL: critiquing Date’s critique”. A: *SIGMOD Rec.* 37.3 (setembre de 2008), pàgines 20–22. DOI: 10.1145/1462571.1462574. (Vegeu la pàgina 18).
- [12] Christopher J. Date. *An Introduction to Database Systems*. 7a edició. Boston, MA, USA: Addison-Wesley, 2000. ISBN: 0-201-38590-2. (Vegeu les pàgines 4, 5, 17, 18, 23).
- [13] Christopher J. Date. *Date on Databases: Writings 2000-2006*. 1era edició. Berkely, CA, USA: Apress, 2006. ISBN: 159059746X. (Vegeu les pàgines 7, 18–20, 23).
- [14] Christopher J. Date. “Foundation matters”. A: *Proceedings of the 28th international conference on Very Large Data Bases*. VLDB'02. Hong Kong, China: VLDB Endowment, 2002, pàgina 2. URL: <http://dl.acm.org/citation.cfm?id=1287369.1287371>. (Vegeu la pàgina 18).
- [15] Christopher J. Date. *Go Faster! The TransRelational Approach to DBMS Implementation*. 1era edició. Ventus Publishing ApS, 2011,2002. ISBN: 978-87-7681-905-7. URL: <http://bookboon.com/en/textbooks/it-programming/go-faster> (visitat l'1 de maig de 2012). (Vegeu la pàgina 18).
- [16] Christopher J. Date. *The Relational Database Dictionary, Extended Edition*. 1era edició. Berkely, CA, USA: Apress, 2008. ISBN: 1430210419, 9781430210412. (Vegeu les pàgines 17, 18, 23).
- [17] Christopher J. Date i Hugh Darwen. “Databases, Types and the Relational Model”. A: *Databases, Types and the Relational Model*. 3era edició. Pearson Education, 2006. Capítol 4,5,A,I. ISBN: 0321399420. URL: <http://www.thethirdmanifesto.com/> (visitat l'1 d'abril de 2012). (Vegeu les pàgines 18, 19).

Bibliografia

- [18] Christopher J. Date i Hugh Darwen. *The Third Manifesto*. www.thethirdmanifesto.com, octubre de 2011. URL: <http://www.dcs.warwick.ac.uk/~hugh/TTM/TTM-2011-10-30.pdf> (visitat l'1 d'abril de 2012). (Vegeu les pàgines 7, 18, 19).
- [19] Christopher J. Date i Hugh Darwen. "Tutorial D". A: *Database Explorations: Essays on The Third Manifesto and related topics*. Trafford publishing, 2010. Capítol 11. ISBN: 978-1-42693-723-1. URL: <http://www.dcs.warwick.ac.uk/~hugh/TTM/DBE-Chapter11.pdf> (visitat l'1 d'abril de 2012). (Vegeu les pàgines 18, 20).
- [20] Christopher J. Date, Hugh Darwen i Nikos A. Lorentzos. *Temporal Data and the Relational Model. A detailed investigation into the application of interval and relation theory to the problem of temporal database management*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2002. ISBN: 1-55860-855-9. DOI: 10.1016/B978-155860855-9/50047-7. (Vegeu les pàgines 4–7, 19, 20, 23).
- [21] Christopher J. Date i Fabian Pascal. *Database Debunkings. Dispelling persistent prevalent database management fallacies*. 2011–2000. URL: <http://www.dbdebunk.com/> (visitat l'1 d'abril de 2012). (Vegeu les pàgines 17, 18).
- [22] Antonios Deligiannakis, Yannis Kotidis i Nick Roussopoulos. "Dissemination of compressed historical information in sensor networks". A: *The VLDB Journal* 16.4 (octubre de 2007), pàgines 439–461. DOI: 10.1007/s00778-005-0173-5. (Vegeu les pàgines 10, 13).
- [23] Alan Demers i altres. "The Cougar Project: a work-in-progress report". A: *SIGMOD Record* 32.4 (desembre de 2003), pàgines 53–59. DOI: 10.1145/959060.959070. URL: <http://www.cs.cornell.edu/johannes/papers/2003/SigmodRecord2003-Sensor.pdf> (visitat el 16 d'abril de 2012). (Vegeu les pàgines 14, 16).
- [24] Werner Dreyer, Angelika Kotz Dittrich i Duri Schmidt. "An Object-Oriented Data Model for a Time Series Management System". A: *Proceeding of the 7th International Working Conference on Scientific and Statistical Database Management*. Virginia, USA: IEEE Computer Society, setembre de 1994, pàgines 186–195. DOI: 10.1109/SSDM.1994.336948. URL: <http://www.ubilab.org/publications/index.html> (visitat el 30 de novembre de 2011). (Vegeu la pàgina 15).
- [25] Werner Dreyer, Angelika Kotz Dittrich i Duri Schmidt. "Research perspectives for time series management systems". A: *SIGMOD Rec.* 23.1 (març de 1994), pàgines 10–15. DOI: 10.1145/181550.181553. URL: <http://www.ubilab.org/publications/index.html> (visitat el 30 de novembre de 2011). (Vegeu les pàgines 4, 13, 14).
- [26] Werner Dreyer, Angelika Kotz Dittrich i Duri Schmidt. *The Implementation of the CALANDA Time Series Management System – A Novel Approach to the Construction of Special-Purpose DBMS*. 1995. URL: <http://www.ecofin.ch/aktuelles/presseartikel/zImplementation.pdf> (visitat l'1 de desembre de 2011). (Vegeu la pàgina 15).

Bibliografia

- [27] Werner Dreyer, Angelika Kotz Dittrich i Duri Schmidt. “Using the CALANDA Time Series Management System”. A: *SIGMOD Rec.* 24.2 (maig de 1995). DOI: 10.1145/568271.223902. URL: <http://www.ubilab.org/publications/index.html> (visitat el 30 de novembre de 2011). (Vegeu les pàgines 9, 15).
- [28] Stefan Edlich. *NoSQL Databases*. 2012–2009. URL: <http://nosql-database.org/> (visitat l'1 de maig de 2012). (Vegeu les pàgines 7, 19, 20).
- [29] Usama Fayyad, Gregory Piatetsky-shapiro i Padhraic Smyth. “From Data Mining to Knowledge Discovery in Databases”. A: *AI Magazine* 17.3 (1996), pàgines 37–54. URL: <http://www.aaai.org/ojs/index.php/aimagazine/article/download/1230/1131> (visitat l'1 de març de 2012). (Vegeu la pàgina 12).
- [30] Tak chung Fu. “A review on time series data mining”. A: *Engineering Applications of Artificial Intelligence* 24.1 (febrer de 2011), pàgines 164–181. DOI: 10.1016/j.engappai.2010.09.007. (Vegeu les pàgines 4, 12).
- [31] Wai Fu Fung, David Sun i Johannes Gehrke. “COUGAR: the network is the database”. A: *Proceedings of the 2002 ACM SIGMOD international conference on Management of data*. SIGMOD'02. Madison, Wisconsin: ACM, 4–6 de juny de 2002, pàgines 621–621. DOI: 10.1145/564691.564775. URL: <http://www.cs.cornell.edu/johannes/papers/2002/sigmod2002-cougardemo.pdf> (visitat el 16 d'abril de 2012). (Vegeu les pàgines 6, 16).
- [32] Johannes Gehrke i Samuel Madden. “Query processing in sensor networks”. A: *Pervasive Computing, IEEE* 3.1 (2004), pàgines 46–55. DOI: 10.1109/MPRV.2004.1269131. (Vegeu la pàgina 10).
- [33] Johannes Gehrke i altres. *Cougar design and implementation*. Cornell University. 2002. URL: <http://www.cs.cornell.edu/boom/2002sp/extproj/www.cs.cornell.edu/database/cougar/cougardesigndoc.pdf> (visitat el 16 d'abril de 2012). (Vegeu les pàgines 4, 16). *COUGAR: the network is the database*. URL: <http://www.cs.cornell.edu/bigreddata/cougar/index.php>.
- [34] The PostgreSQL Global Development Group. *PostgreSQL: The world's most advanced open source database*. 2012–1996. URL: <http://www.postgresql.org/> (visitat el 28 de maig de 2012). (Vegeu la pàgina 17).
- [35] Theo Härder. “DBMS Architecture — Still an Open Problem”. A: *Proceedings Datenbanksysteme in Business, Technologie und Web. 11 Fachtagung des GI-Fachbereichs "Datenbanken und Informationssysteme"*. Volum 65. Lecture Notes in Informatics LNI-BTW'05. Bonn: Gesellschaft für Informatik (GI), març de 2005, pàgines 2–28. ISBN: 3-88579-394-6. URL: <http://subs.emis.de/LNI/Proceedings/Proceedings65/article3661.html> (visitat el 21 de maig de 2012). (Vegeu la pàgina 21).
- [36] Magnus Lie Hetland. “A survey of recent methods for efficient retrieval of similar time sequences”. A: *Data mining in time series databases*. Edició a cura de Mark Last, Abraham Kandel i Horst Bunke. Series in Machine Perception and Artificial Intelligence 57. Singapore: World Scientific, 2004. Capítol 2, pàgines 23–41. (Vegeu la pàgina 8).

Bibliografia

- [37] Icinga. *Icinga - Open Source Monitoring, Nagios fork*. URL: <http://www.icinga.org/> (visitat el 25 d'octubre de 2010). (Vegeu la pàgina 15).
- [38] H. V. Jagadish, Inderpal Singh Mumick i Abraham Silberschatz. "View maintenance issues for the chronicle data model (extended abstract)". A: *Proceedings of the fourteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems*. PODS'95. San Jose, California, United States: ACM, 1995, pàgines 113–124. DOI: 10.1145/212433.220201. URL: <http://www.cs.yale.edu/~avi/home-page/publication-dir/PODS95.pdf> (visitat el 19 d'abril de 2012). (Vegeu la pàgina 14).
- [39] Neha Jain i Dharma P. Agrawal. "Current Trends in Wireless Sensor Network Design". A: *International Journal of Distributed Sensor Networks* 1.1 (2005), pàgines 101–122. DOI: 10.1080/15501320590901865. (Vegeu les pàgines 9, 10).
- [40] Eamonn Keogh i Shruti Kasetty. "On the Need for Time Series Data Mining Benchmarks: A Survey and Empirical Demonstration". A: *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD'02. Edmonton, Alberta, Canada: ACM, juliol de 2002, pàgines 102–111. URL: http://www.cs.ucr.edu/~eamonn/sigkdd_bench.pdf (visitat el 15 de març de 2011). (Vegeu la pàgina 12).
- [41] Eamonn Keogh i Jessica Lin. *iSAX (Symbolic Aggregate approXimation)*. URL: <http://www.cs.ucr.edu/~eamonn/iSAX/iSAX.htm> (visitat el 15 de març de 2011). (Vegeu la pàgina 15).
- [42] Eamonn Keogh i Michael Pazzani. "An enhanced representation of time series which allows fast and accurate classification clustering and relevance feedback". A: *Proceedings of the 4th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD'98. New York: ACM, agost de 1998, pàgines 239–243. URL: <http://www.cs.ucr.edu/~eamonn/kdd98.pdf> (visitat el 15 de març de 2011). (Vegeu la pàgina 12).
- [43] Eamonn Keogh i Padhraic Smyth. "A probabilistic approach to fast pattern matching in time series databases". A: *Proceedings of the 3rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD'97. Newport Beach, California: ACM, agost de 1997, pàgines 24–20. URL: <http://www.cs.ucr.edu/~eamonn/kdd97.ps> (visitat el 15 de març de 2011). (Vegeu les pàgines 10, 12).
- [44] Eamonn Keogh i altres. "Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases". A: *Knowledge and Information Systems*. KAIS 3.3 (agost de 2001), pàgines 263–286. URL: http://www.cs.ucr.edu/~eamonn/kais_2000.pdf (visitat el 15 de març de 2011). (Vegeu les pàgines 12, 13, 15).
- [45] Eamonn Keogh i altres. "Locally adaptive dimensionality reduction for indexing large time series databases". A: *Proceedings of ACM SIGMOD International Conference on Management of Data*. SIGMOD'01. Santa Barbara, California, USA: ACM, maig de 2001, pàgines 151–162. URL: <http://www.cs.ucr.edu/~eamonn/>

Bibliografia

- sigmod_apca_2001.pdf (visitat el 15 de març de 2011). (Vegeu les pàgines 12, 13).
- [46] Martin Kersten. *SciQL, Scilens project*. scilens.project.cwi.nl. 2011. URL: <http://www.scilens.org/Resources/SciQL> (visitat el 20 d'abril de 2012). (Vegeu la pàgina 16).
- [47] Martin Kersten i altres. “SciQL, a query language for science applications”. A: *Proceedings of the EDBT/ICDT 2011 Workshop on Array Databases*. AD'11. Uppsala, Sweden: ACM, 2011, pàgines 1–12. DOI: 10.1145/1966895.1966896. URL: <http://www.cwi.nl/~zhang/papers/arraydb11.pdf> (visitat el 20 d'abril de 2012). (Vegeu les pàgines 14, 16).
- [48] Mark Last, Gil Avrahami i Abraham Kandel. “Using data mining techniques for optimizing traffic signal plans at an urban intersection”. A: *International Journal of Intelligent Systems* 26.7 (juliol de 2011), pàgines 603–620. DOI: 10.1002/int.20473. (Vegeu la pàgina 9).
- [49] Mark Last i altres. “Knowledge discovery in time series databases”. A: *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics* 31.1 (febrer de 2001), pàgines 160–169. DOI: 10.1109/3477.907576. (Vegeu les pàgines 12, 14).
- [50] Aleix Llusà Serra. “Estudi i modelització dels SGBD Round Robin pel tractament de sèries temporals”. Tesi de màster. Universitat Politècnica de Catalunya: Màster Universitari en Automàtica i Robòtica, juny de 2011. (Vegeu la pàgina 22).
- [51] Camilo Lozoya, Manel Velasco i Pau Martí. “The One-Shot Task Model for Robust Real-Time Embedded Control Systems”. A: *IEEE Transactions on Industrial Informatics* 4.3 (juliol de 2008), pàgines 164–174. DOI: 10.1109/TII.2008.2002702. (Vegeu la pàgina 11).
- [52] Samuel R. Madden i altres. *TinyDB. A declarative database for sensor networks*. 2003–2002. URL: <http://telegraph.cs.berkeley.edu/tinydb/index.htm> (visitat el 17 d'abril de 2012). (Vegeu la pàgina 16).
- [53] Samuel R. Madden i altres. “TinyDB: an acquisitional query processing system for sensor networks”. A: *ACM Transactions on Database Systems* 30.1 (març de 2005), pàgines 122–173. DOI: 10.1145/1061318.1061322. URL: <http://db.cs.berkeley.edu/papers/tods05-tinydb.pdf> (visitat el 17 d'abril de 2012). (Vegeu la pàgina 16).
- [54] Sasa Markovic i Arne Vandamme. *JRobin: A Java implementation of RRDtool*. URL: <http://www.jrobin.org/> (visitat el 22 de març de 2011). (Vegeu la pàgina 15).
- [55] John Mylopoulos i altres. “Building knowledge base management systems”. A: *The VLDB Journal* 5.4 (1996), pàgines 238–263. DOI: 10.1007/s007780050027. (Vegeu la pàgina 13).
- [56] Nagios. *Nagios - The Industry Standard In IT Infrastructure Monitoring*. URL: <http://www.nagios.org/> (visitat el 25 d'octubre de 2010). (Vegeu la pàgina 15).

Bibliografia

- [57] Tobias Oetiker. “MRTG The Multi Router Traffic Grapher”. A: *Proceedings of the 12th Systems Administration Conference*. LISA’98. Boston, Massachusetts: USENIX Association, desembre de 1998, pàgines 141–148. URL: <http://www.usenix.org/publications/library/proceedings/lisa98/oetiker.html>. (Vegeu la pàgina 15).
- [58] Tobias Oetiker. *RRDtool, Round Robin database*. 1998–2011. URL: <http://oss.oetiker.ch/rrdtool/> (visitat el 10 d’octubre de 2011). (Vegeu les pàgines 4, 6, 15, 22).
- [59] Tobias Oetiker. *The Multi Router Traffic Grapher*. URL: <http://oss.oetiker.ch/mrtg/> (visitat el 15 d’octubre de 2010). (Vegeu la pàgina 15).
- [60] Umit Ogras i Hakan Ferhatosmanoglu. “Online summarization of dynamic time series data”. A: *The VLDB Journal* 15.1 (2006), pàgines 84–98. DOI: 10.1007/s00778-004-0149-x. (Vegeu les pàgines 13, 14).
- [61] Carl A. Palmer, Nicholas A. Mackos i Michael J. Roemer. “Approach to Monitor and Assess the Quality of Sensor Data in Support of Calibration and Condition Based Maintenance for Turbine Powered Navy Vessels”. A: *ASME Conference Proceedings* 1 (2007), pàgines 981–989. DOI: 10.1115/GT2007-28251. (Vegeu la pàgina 9).
- [62] David Plonka, Archit Gupta i Dale Carder. “Application Buffer-Cache Management for Performance: Running the World’s Largest MRTG”. A: *Proceedings of the 21st Systems Administration Conference*. LISA’07. Dallas, Texas: USENIX Association, novembre de 2007, pàgines 63–78. URL: <http://www.usenix.org/events/lisa07/tech/plonka.html>. (Vegeu la pàgina 15).
- [63] Joseba Quevedo i altres. “Validation and reconstruction of flow meter data in the Barcelona water distribution network”. A: *Control Engineering Practice* 18.6 (juny de 2010), pàgines 640–651. (Vegeu les pàgines 9, 11).
- [64] Gunter Saake i altres. “Downsizing Data Management for Embedded Systems”. A: *Egyptian Computer Science Journal (ECS)* 31.1 (gener de 2009), pàgines 1–13. URL: <http://www.witi.cs.uni-magdeburg.de/~rosenmue/publications/downsizing09ecs.pdf> (visitat el 21 de maig de 2012). (Vegeu la pàgina 21).
- [65] Duri Schmidt i altres. “Time Series, A Neglected Issue in Temporal Database Research?” A: *Proceedings of the International Workshop on Temporal Databases: Recent Advances in Temporal Databases*. Workshops in Computing. Zürich, Switzerland: Springer, setembre de 1995, pàgines 214–232. URL: http://www.ecofin.ch/aktuelles/presseartikel/zNeglected_Issue.pdf (visitat l’1 de desembre de 2011). (Vegeu les pàgines 4, 8, 14, 15, 20).
- [66] Praveen Seshadri. “Enhanced abstract data types in object-relational databases”. A: *The VLDB Journal* 7.3 (agost de 1998), pàgines 130–140. DOI: 10.1007/s007780050059. (Vegeu la pàgina 19).

Bibliografia

- [67] Praveen Seshadri. “Management of Sequence Data”. Tesi doctoral. University of Wisconsin, 1996. URL: <http://www.cs.cornell.edu/home/praveen/papers/thesis.ps.gz> (visitat el 16 d’abril de 2012). (Vegeu la pàgina 14).
- [68] Praveen Seshadri, Miron Livny i Raghu Ramakrishnan. “SEQ: A model for sequence databases”. A: *Proceedings of the Eleventh International Conference on Data Engineering*. ICDE’ 95. Març de 1995, pàgines 232–239. DOI: 10.1109/ICDE.1995.380388. URL: <http://www.cs.cornell.edu/home/praveen/papers/seq.de95.ps.Z> (visitat el 16 d’abril de 2012). (Vegeu la pàgina 14).
- [69] Jin Shieh i Eamonn Keogh. “iSAX: Indexing and Mining Terabyte Sized Time Series”. A: *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD’08. Las Vegas: ACM, agost de 2008, pàgines 623–631. URL: <http://www.cs.ucr.edu/~eamonn/iSAX.pdf> (visitat el 15 de març de 2011). (Vegeu la pàgina 15).
- [70] Francesc Soriguera Martí. “Highway travel time estimation with data fusion”. Tesi doctoral. Barcelona: Universitat Politècnica de Catalunya, octubre de 2010. URL: http://www.fundacioabertis.org/pdf/Editarium_FSoriguera.pdf (visitat el 30 d’abril de 2012). (Vegeu la pàgina 9).
- [71] Michael Stonebraker. “Inclusion of New Types in Relational Data Base Systems”. A: *Proceedings of the Second International Conference on Data Engineering*. ICDE’86. Los Angeles, California, USA: IEEE Computer Society, febrer de 1986, pàgines 262–269. URL: http://pages.cs.wisc.edu/~nil/764/ORDB/34_stonebraker86inclusion.pdf (visitat el 8 de novembre de 2011). (Vegeu les pàgines 5, 19).
- [72] Michael Stonebraker. “SQL databases v. NoSQL databases”. A: *Communications of the ACM* 53.4 (abril de 2010), pàgines 10–11. DOI: 10.1145/1721654.1721659. (Vegeu les pàgines 7, 19, 21).
- [73] Michael Stonebraker i Jason Hong. “Saying good-bye to DBMSs, designing effective interfaces”. A: *Communications of the ACM* 52.9 (setembre de 2009), pàgines 12–13. DOI: 10.1145/1562164.1562169. (Vegeu les pàgines 19, 20).
- [74] Michael Stonebraker i altres. “Requirements for Science Data Bases and SciDB”. A: *Fourth Biennial Conference on Innovative Data Systems Research*. CIDR’ 09. Asilomar, CA, USA: www.cidrdb.org, gener de 2009. URL: <http://www.scidb.org/Documents/SciDB-CIDR2009.pdf> (visitat el 20 d’abril de 2012). (Vegeu les pàgines 4, 6, 14, 16, 20).
- [75] Michael Stonebraker i altres. “The end of an architectural era: (it’s time for a complete rewrite)”. A: *Proceedings of the 33rd international conference on Very large data bases*. VLDB’07. Vienna, Austria: VLDB Endowment, 2007, pàgines 1150–1160. URL: <http://www.vldb.org/conf/2007/papers/industrial/p1150-stonebraker.pdf> (visitat el 26 de març de 2012). (Vegeu les pàgines 20, 21).
- [76] Mike Stonebraker i altres. *SciDB, A Science-Oriented DBMS*. 2012–2010. URL: <http://www.scidb.org/> (visitat el 20 d’abril de 2012). (Vegeu la pàgina 16).

Bibliografia

- [77] Dave Voorhis. *Rel. An Implementation of Date and Darwen's Tutorial D database language*. 2012–2004. URL: <http://dbappbuilder.sourceforge.net/> (visitat l'1 de maig de 2012). (Vegeu les pàgines 18, 23).
- [78] Robert Weigel i altres. “TSDS: high-performance merge, subset, and filter software for time series-like data”. A: *Earth Science Informatics* 3.1 (2010), pàgines 29–40. ISSN: 1865-0473. DOI: 10.1007/s12145-010-0059-y. (Vegeu la pàgina 15).
- [79] Robert Weigel i altres. *TSDS (Time Series Database Server)*. 2011 – 2009. URL: <http://tsds.net/> (visitat el 2 de novembre de 2011). (Vegeu la pàgina 15).
- [80] Qiang Yang i Xindong Wu. “10 Challenging Problems in Data Mining Research”. A: *International Journal of Information Technology and Decision Making*. IJITDM 5.4 (desembre de 2006), pàgines 597–604. URL: <http://www.cs.uvm.edu/~icdm/10Problems/index.shtml> (visitat el 22 de maig de 2011). (Vegeu la pàgina 12).
- [81] Yong Yao i Johannes Gehrke. “The Cougar Approach to In-Network Query Processing in Sensor Networks”. A: *Sigmod Record* 31.3 (setembre de 2002), pàgines 9–18. DOI: 10.1145/601858.601861. URL: <http://www.cs.cornell.edu/johannes/papers/2002/sigmod-record2002.pdf> (visitat el 16 d'abril de 2012). (Vegeu les pàgines 9, 10).
- [82] Byoung-Kee Yi i altres. “Online data mining for co-evolving time sequences”. A: *Proceedings. 16th International Conference on Data Engineering*. IEEE, 2000, pàgines 13–22. DOI: 10.1109/ICDE.2000.839383. (Vegeu la pàgina 14).
- [83] Ming Yu i altres. “Prognosis of Hybrid Systems With Multiple Incipient Faults: Augmented Global Analytical Redundancy Relations Approach”. A: *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on* 41.3 (maig de 2011), pàgines 540–551. ISSN: 1083-4427. DOI: 10.1109/TSMCA.2010.2076396. (Vegeu la pàgina 9).
- [84] Jian Zhang i Renato Figueiredo. “Adaptive Predictor Integration for System Performance Prediction”. A: *Proceedings of the 21st International Parallel and Distributed Processing Symposium*. IPDPS'07. Long Beach, California: IEEE, març de 2007. URL: <http://www.cecs.uci.edu/~papers/ipdps07/pdfs/IPDPS-1569011105-paper-1.pdf> (visitat el 3 d'octubre de 2010). (Vegeu la pàgina 16).
- [85] Ying Zhang i altres. “SciQL: bridging the gap between science and relational DBMS”. A: *Proceedings of the 15th Symposium on International Database Engineering & Applications*. IDEAS'11. Lisboa, Portugal: ACM, 2011, pàgines 124–133. DOI: 10.1145/2076623.2076639. URL: <http://www.cwi.nl/~zhang/papers/ideas11.pdf> (visitat el 20 d'abril de 2012). (Vegeu les pàgines 4, 6, 14).

A. Direcció

En tractar-se d'una recerca situada entre dos àmbits, l'anàlisi de les sèries temporals i els sistemes de gestió de bases de dades, compta respectivament amb la direcció de dos doctors d'aquest àmbits: la Teresa Escobet Canal i el Sebastià Vila-Marta.

A.1. Teresa Escobet Canal

És professora del departament d'Enginyeria del Disseny i Programació de Sistemes Electrònics de la Universitat Politècnica de Catalunya. És membre del Programa de Doctorat en Automàtica, Robòtica i Visió.

A.2. Sebastià Vila-Marta

És professor del departament d'Enginyeria del Disseny i Programació de Sistemes Electrònics de la Universitat Politècnica de Catalunya. A continuació s'incorpora el seu currículum.