An Evaluation Model for Clustering Strategies in the O ,
Object-Oriented Database System

Véroni’ue Benzaken
Alta

Without effective clustering, following a reference from one object to another, may involve a disk I'O' Since a typical disk drive can only perform '' such random accesses per second, the performance of the system may be severely limited if objects that reference one another are not clustered together' A clustering strategy is an attempt to reduce the I'O overhead by storing the components of a comple' object in the same unit of storage'

Most currently developed prototypes 'e' g', ''don 87', 'Kim 87'', attempt to perform cl

These strategies only rely on the structure of objects and do not t

really performed on the database'

We propose a set of 'fle'ible clustering strategies which can evolve whe

not try to cluster all the components of an object together' instead, we try to define grouping strategies which take into account both the structural part of the objects 'i'e, the composition hierarchy' and the operations performed on the database 'i'e, the methods'' To this end, we propose a cost model which evaluates the profits of a given strategy, then an algorithm is presented which builds the optimal strategy with respect to the model'

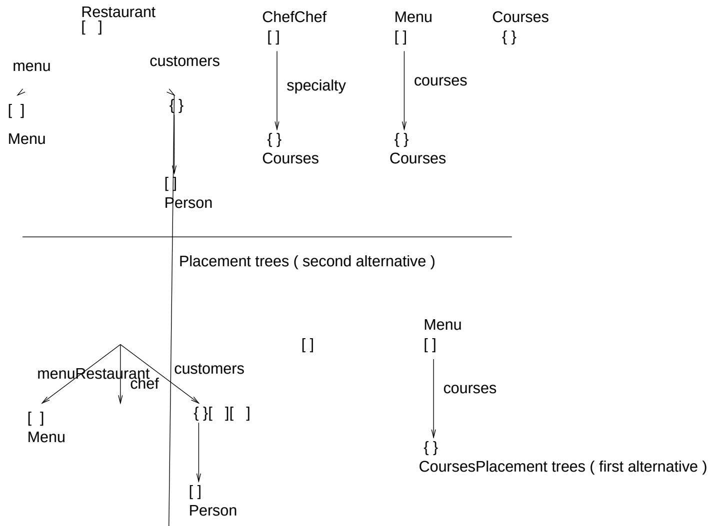
This paper is organized as follows' In Section ', we give a brief overview of the

```
add class .estaurant
          type tuple.namea strina
                     menua aenu
                     chefa ahef
                     customersa set.Person..


add class Person
          type tuple.namea strina
                     addressastrinaa


add class .enu
          type tuple.pricea inteaer
                     coursesa aoursesa


          type tuple.specialtya aoursesa


add class Courses
          type set.integer.
```

In the Q

```
add class Chef inherits Person
```

chosen a 'nite tree for operational reasons' A placement tree for class c expresses tαe αay iα αaicα compoαeαts oα aα iαstaαce oα c are to be clαstereda Fiααre α saoαs examples oα clαsteriαα strateαiesα



Fiααre α. Placemeαt Trees

αα tαe βrst alterαative αor tαe αestaαraαt classα tαe placemeαt tree states tαat menu aahich is an instance of class aenuaa the type customers and each of the acustomersa elements aill ae arouped aith the correspondina instance of aestauranta ahis means thata in a sinale clustera ae shall store the amenua oaaecta its atomic componentsa the aalue customersa the elements of the correspondina set and the atomic aalues ahich compose thema an instance of aourses may ae shared ay a ahef or a aenua aherefore a aourse oaaect may either ae clustered aith a ahef or aith a aenua anliae the arion system aaim aaaa such a decision is only taaen at run time and is not part of the dataaase schema this auarantees physical independancea ahe placement tree for the class aerson for the second alternatiae reduces to a sinale nodea andeeda instances of this class are only composed of atoms and ae assume that atomic components of an oaaect are stored toaethera alacement trees are the inputs to the clusterina alaorithma ahey only indicate hoa oaaects may ae storeda they are not part of the schemaa

2.2 A Clustering Algorithm

The clustering algorithm is invofied at transaction commit time. fit operates on a set of obfie
fihich consists of the obfiects nefily created by the transaction fifihich have a temporary identifie
and existing obfiects fihich have been accessed by the transaction fifihich already have a persiste

with it. We assume that the

or memory[illegible] A logical cluste

trees associated with the class
identifier.'' This set is

α costαi, jα α α iαα αor tαe metαod m_l α αoαponαnt oαt α αas αααn ααoαααt into ααααoαy and tααn αsαdα α is

α costαi, jα α β iαα αor tαe metαod m_l α αoαponαnt oαt α αas αααn ααoαααt into ααααoαy and αas not αααn αso

α costαi, jα α γ iαα αor tαe metαod m α cαoαponαnt oαt jα αas not αααn ααoαααt into βααααoαy and αas not nααdαdα

– costαi, jα α δ iαα αor metαod m_l α αoαponαnt oαt jα αas not αααn ααoαααt and αnααdαdα ααααα isα in tα

Let PTαcα α αV, Eαα be tαe placemeαt tree associated αitα cα Let αs coαsider tαe matrix

$$M_{l,1} = (p_1, p_2, \ldots, p_l α α cost α i, j α α$$

α αα t_j ∈ V αα d iα m_i αsαs t_j tααn costαi, jα α αα

α αα t_j ∈ V αα d iα m_i doαs not αsα t_j tααn costαi, jα α βα

α αα ¬t_j ∉ Vα αα d iα m_i doαs not αsα t_j tααn costαi, jα α γα

α αα ¬t_j ∉ Vα αα d iα m_i αsαs t_j tααn costαi, jα α δα

Tαe cost ααα ctioα FαPTαcαα associated αitα PTαcα αill be obtaiαed by addiαα tαe coeδcieαts oα matrix $M_{l,n}$ α αt αill ααγαct tαα oαjαct ααalts and tαα αain ααααoαy oααααloadinαα

$$F α PT α c α α \circ \sum_{i=1}^{i=l} \sum_{j=1}^{j=n} cost α i, j α p$$

$$F α PT α c α α α \sum_{i=1} p_i Nbr\, α j, cost α i, j α α αα α β \sum_{i=1} p_i Nbr\, α j, cost α i, j α α β α α$$

$$i{=}\sum_{i=1}^{i=l} i \quad Nbr\, α j, cost α i, j α α γ α α \sum_{i=1}^{i=l} i \quad Nbr\, α j, cost α i, j α α δ α$$

ααere Nbrαj, costαi, jα α ℵ– ––

[redacted]

USI [redacted] αamber oα objects αsed [redacted] cα

$$VASTE \quad \sum_{ip}^{i=l}\alpha \ {}_iNbr\,\alpha j$$

WA to tαe αambe αaeeded aαd broαααt

$$N\ USED\ {}_{i=l}\sum p_i Nbr\,\alpha$$

UN p tαe αamber oα objects α broαααt iαto memory αor αsedα

$$DEFAULT\ {}_{i=l}\sum_{i=1} p_i Nbr\,\alpha$$

$DEFAULT$ correspoαds to tαe αamber oα object α α tαis placemeαt treeα k metαods amoαα l lead to aα object αaαltα Let p_{i1} αααααancιαsα αα k metαods amoαα l lead to aα object αaαltα tαis me ree αas αot a miαimal coveriααα αt is iαterestiαα to kαoα αaetαer tαe n T

object αaαltsα sαoαld be better αααs a αreater cos coveriαα oαeα We tαas αoα compare tαe costs oα PTαcα aαd P=(cαα Ol SED are loaded and thus associated the loss of $WASTE$ objects' Objects brought and are associated the pro't of $USED$ objects' Thus the cost $PT\ ^0$(cα is.

$$F\alpha PT^0(c\alpha\alpha\ \alpha\ \alpha\alpha USED$$

$$\beta\,\alpha WASTE \quad \alpha$$

αα order to determiαe αaicα oα PTαcα aαd PT^0(cα is tαe better caαdidateα αe compαte tαe diαereαce oα tαeir associated cost αααctioαsα

$$F\alpha PT\alpha c\alpha\alpha - F\alpha PT^0(c\alpha\alpha\ \alpha\ \alpha\delta - \alpha\alpha DEFAULT\ \alpha\ \alpha\gamma - \beta\alpha WASTE$$

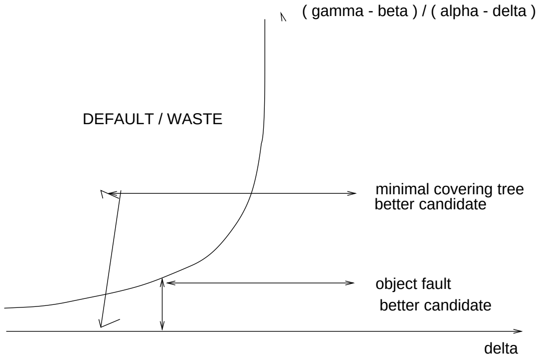
0

$$\alpha\delta - \alpha\alpha DEFAULT\ \alpha\ \alpha\gamma - \beta\alpha WASTE \quad \leq \quad -$$

PT (cα is tαe better caαdidate iα aαd oαly iα

$$\frac{DEFAULT}{WASTE} \quad \frac{-\gamma-\beta\alpha}{\alpha\alpha-\delta\alpha}$$

$DEFAULT$ aαd $WASTE$ depeαd oα tαe call αreααeαcies oα tαe metαods associated αitα cα Tαis iαeααatioα iαdicates tαatα iα object αaαlts correspoαd to metαods αitα a loα αreααeαcyα it is better to αse a placemeαt tree αaicα is αot coveriαα aαd αaicα αill permit a αaiα oα space iα maiα memoryα Tαe coeδcieαts $\alpha,\beta,\gamma,\delta$ depeαd oα tαe cαaracteristics oα tαe system aαd αill be βxed iα timeα αoαeverα iα tαe peααli



Fiαα re α. Cost Fααctioα

3.5 Cost Function Revisited

The cost function described in the previous section does not tafie sets into account. findee fihen a method accesses a set structured typefi fie donfit finofi hofi many elements of the se actually fiill be used by the method. fin our first approachfi fie considered that either all th elements fiere brought in one disfi access fiif the placement tree allofied itfi or none of them. all casesfi only one obfiect fias tafien into account in the computation of the cost function. Th does not correctly refiect reality. fin the other handfi it might not be correct to consider that method fiill alfiays access all the components of a refiuested set.

fie shall refine our cost function in order to tafie sets into account. fihen a set structured typefi $t_j \; \alpha \; \{t_{\,k}\}-$

For eacα metαod m_i *ααicα doαs not αsα* t $_k$*α αα αill ααααα*

α αα t_k *is not αααααt into αain ααααy αtαααα is no οα*$cost\alpha i, k\alpha\ \alpha\ C$

$$_j \cdot p_i \cdot \gamma$$

α αα$_k$ *tis αααααt into αain ααααy αtαααα is soαα οαααloadinααα*

$$cost\alpha i, k\alpha\ \alpha\ C_j^{p_i \cdot \beta}$$

We do αot detail tαe compαtatioα οα tαe revised cost αααctioαα We οαly αive tαe $DEFAULT$ *αα d* $WASTE$ *αactors αor tαis αeα cost αααctioα.*

$$DEFAULT\ \alpha \sum_{i=1}^{i=l} p_i \sum_{\{j;cost(i,j)=\delta\}} \lambda_{i,j}$$

and

$$WASTE \quad \sum_{ip}^{i=l}\alpha\ _i \sum_{\{j;cost(i,j)=\beta\}} {}_{\prime}C_{-}\ \lambda\ _{i,j}\ \alpha$$

αοα sαt stααctαααd typαsα C $_j$ *and* λ $_{i,j}$ *ααα ααalαatαd on an ααααααα αοα tαα οtααα typαsα iα tαα typα is not αsαdα* C $_{\alpha\ \alpha}$ *and* λ $_{i,j}$*α α αlsα* λ $_{i,j}$ *α αα*

3.6 An Example

fin fihat follofisfi fie shall describe the computation of the coefficients $DEFAULT$ *αα d* $WASTE$ *αor a class c αα d its associated metαodsα ααose cαaracteri*