

APLICAȚII INTEGRATE PENTRU ÎNTREPRINDERI

Tema 4

Generarea unor rapoarte statistice referitoare la serviciile medicale oferite în cadrul unui lanț de policlinici

Lansare	Termen de Predare	Pondere
13.01.2015, 23:55	20.01.2015, 23:55	5 (bonus)

Obiective

Obiectivul temelor de casă este reprezentat de implementarea unui sistem informatic performant și scalabil pentru gestiunea resurselor unei întreprinderi de dimensiuni mari. În acest an universitar, organizația pentru care se va proiecta și dezvolta acest pachet de programe este un lanț de policlinici.

Se va realiza implementarea graduală a mai multor module din cadrul aplicației ERP, pe parcursul a patru etape.

În cadrul acestei etape se va dezvolta o aplicație distribuită care va rula pe mai multe mașini, în condițiile în care volumul de date pe care trebuie să îl gestioneze organizația implică timpi de execuție inacceptabili pentru un utilizator care o interoghează.

După rezolvarea temei de casă, studentul va fi capabil să:

- proiecteze o bază de date distribuită, având o structură dinamică și un număr foarte mare de înregistrări;
- utilizeze paradigma de programare MapReduce pentru a gestiona eficient partițiile unei probleme care pot fi procesate concomitent;
- realizeze o aplicație distribuită ce conține mai multe sarcini, dintre care unele vor fi rulate secvențial, iar altele în paralel.

Cunoștințele necesare¹ pentru rezolvarea temei de casă sunt:

- programarea în limbajul Java;
- manipularea unei baze de date distribuite HBase (atât prin shell-ul specific cât și prin API);
- paradigma de programare MapReduce.

Enunț

Întrucât numărul de înregistrări stocat în baza de date relațională determină ca timpul de răspuns pentru operațiile de raportare (ce implică volume mari de date) să nu respecte constrângerea cu privire la operarea în timp real a unei aplicații integrate pentru întreprinderi, la nivelul executiv al lanțului de policlinici s-a luat decizia migrării spre o soluție distribuită, pentru care a decis achiziționarea unui cluster format din mașini larg disponibile în comerț (având un cost redus, fără performanțe deosebite).

Se dorește monitorizarea activității medicilor care oferă servicii medicale în cadrul lanțului de policlinici în vederea evaluării oportunității continuării colaborării cu aceștia. Criteriile avute în vedere sunt numărul de servicii medicale oferite, numărul de pacienți care apelează la cadrul medical respectiv precum și veniturile încasate pe baza consultațiilor / investigațiilor oferite. Acestea vor fi analizate pe baza rapoartelor medicale completate pentru fiecare programare realizată de un pacient în cadrul unității sanitare.

De asemenea, va fi investigat, pentru fiecare specialitate, numărul de medici care activează în cadrul acesteia și sumele de bani pe care le-au generat.

¹ NU este permisă folosirea altor tehnologii decât a celor specificate în cadrul enunțului.

În acest sens, vor fi generate două rapoarte sub forma unor fișiere text:

- raportul detaliat va surprinde activitatea fiecărui medic pe baza rapoartelor medicale completate de acesta;
- raportul sintetic va oferi informații cu privire la fiecare specialitate pentru care au fost oferite servicii medicale.

Un **raport detaliat** conține numele și prenumele medicului, numărul de pacienți unici² pentru care acesta a completat rapoarte medicale, numărul de servicii medicale distincte oferite în cadrul policlinicii³ precum și venitul pe care acesta l-a generat, pentru toate consultațiile / investigațiile medicale:

raport_detaliat

NUME MEDIC _k Prenume Medic _k = număr_pacienti, numar_servicii_medicale, venit

Un raport sintetic conține denumirea fiecărei specialități pentru care au fost oferite servicii medicale în cadrul lanțului de policlinici, numărul de medici care le-au oferit precum și venitul generat⁴.

raport_sintetic

Specialitate _k = număr_medici, venit

Precizări suplimentare

Se recomandă ca pentru implementarea temei de casă să folosiți una dintre mașinile virtuale pusă la dispoziție în cadrul laboratorului sau să folosiți aceleași versiuni / proprietăți de configurare ca în cadrul acesteia, întrucât corectarea se va realiza pe o astfel de platformă. Cu alte cuvinte, veți utiliza Hadoop 2.5.1 / 2.6.0 și HBase 0.98.6.1 / 0.98.8.

Este suficient ca în portarea bazei de date să preluați DOAR acele tabele necesare pentru rezolvarea cerințelor și doar câmpurile care se vor regăsi în cadrul rapoartelor generate. Întrucât procesul de populare implică mai multe comenzi HBase pentru o singură înregistrare, se recomandă și implementarea unui program pentru generarea scriptului care realizează această operație. Nu este necesară includerea acestei aplicații în arhiva temei de casă, este suficient să includeți scriptul pentru popularea bazei de date. Informațiile din baza de date trebuie să fie coerente (serviciile medicale solicitate de un pacient în cadrul unei programări trebuie să corespundă uneia dintre specialitățile pe care le deține medicul; în cadrul acestei teme de casă nu vor fi luate în considerare competențele dobândite de medic, acesta putând oferi toate serviciile medicale corespunzătoare specialității sale).

Nu este necesară includerea unui script pentru crearea unei arhive jar conținând aplicația dezvoltată, întrucât operația este realizată de plugin-ul Hadoop pentru Eclipse, acesta fiind responsabil și de transmiterea sarcinii, permițând totodată consultarea informațiilor din sistemul de fișiere distribuit HDFS.

Nu este permisă utilizarea unui obiect de tip `org.apache.hadoop.hbase.client.Scan` decât ca parametru pentru clasele care realizează sarcini de tip MapReduce, întrucât operațiile realizate de acesta, în contextul unei singure mașini, implică resurse de procesare și memorie care depreciază timpul de execuție al unei aplicații.

² Un pacient va fi luat în considerare o singură dată chiar dacă are mai multe programări la medicul în cauză, respectiv dacă a contractat mai multe servicii medicale realizate de acesta. În cazul în care un pacient se programează pentru servicii medicale la mai mulți medici, el va fi contorizat pentru fiecare dintre aceștia, chiar dacă serviciile medicale sunt identice.

³ Pentru evaluarea numărului de servicii medicale distincte se vor însuma acele consultații / investigații pentru care există înregistrări în cadrul rapoartelor medicale, nu a tuturor serviciilor medicale corespunzătoare specialităților / competențelor pe care le deține.

⁴ În situația în care un medic are mai multe specialități și a oferit servicii medicale unor specialități diferite, el va fi contorizat în cadrul fiecăreia dintre acestea. Dacă un medic nu a realizat consultații / investigații pentru nici o specialitate, el nu va fi luat în considerare, deși activează în cadrul lanțului de policlinici.

În raport vor fi omise informațiile nule, în cazul în care activitatea unui medic / pentru o specialitate nu poate fi contorizată conform înregistrărilor reținute în baza de date.

Barem de corectare și notare

Punctaj	Criterii de acordare
1,25 p	reproiectarea bazei de date conform modelului distribuit <ul style="list-style-type: none">definirea tabelelor și a structurii statice: 30%populare adecvată: 70%
3,50 p	generarea rapoartelor <ul style="list-style-type: none">raport detaliat: 50%raport sintetic: 50%
0,25 p	modularizare <ul style="list-style-type: none">structura aplicației și lizibilitatea codului: 50%comentarii, README: 50%

BONUS. Se pot obține punctaje suplimentare, astfel:

- 0,25p – predarea temei până la 17.01.2014, ora 23:55;
- 0,75p – defalcarea raportului detaliat, pe luni și pe ani, astfel încât acesta să prezinte, pentru fiecare medic, activitatea din cadrul unei luni dintr-un an pentru care există înregistrări în baza de date.

Condiții de realizare și predare

Tema va fi realizată individual și va fi prezentată în cadrul laboratorului, până la sfârșitul semestrului. Neprezentarea temei de casă atrage după sine nepunctarea acesteia. În cadrul prezentării veți specifica succint funcționalitățile pe care le-ați dezvoltat și veți răspunde la întrebări cu privire la diferite soluții adoptate pentru rezolvarea problemelor întâlnite.

Tema va trebui încărcată pe site-ul cs.curs.pub.ro sub forma unei arhive de tip .zip (având denumirea **Grupa34XCX_NumePrenume_Tema4.zip**) care să conțină script-ul pentru crearea și popularea tabelor din baza de date, sursele aplicației și un README în care să prezentați definiția bazei de date (tabele, familii de coloane și structura lor), explicând cum ați proiectat sarcinile de tip map și reduce. Prezentarea se poate realiza numai după ce tema a fost încărcată pe site.

Încărcarea pe site nu este redundantă, temele vor fi comparate prin aplicații specializate pentru a se depista eventualele fraude. În această situație, întreg punctajul pe parcursul semestrului va fi anulat, studenții implicați (atât originalul, cât și copia / copiile) fiind obligați să repete disciplina – cu taxă – în anul universitar următor.