Data Management Wrap-up
Wed. December 2 - Wrapup

Selected Project Presentations:

1. WDB: Morris Tirey Hurt Joshua Wu
   SongTing D1
   https://www.github.com/tireymorris/F15D1
2. WDB: Deng Qicong Chee Raymond D2
   https://github.com/alvin319/F15D2

Wrap-up:

- Class Notes
CS411: Intro to Upper Division Electives

April 13, 2015

Dr. Philip E. Cannata - Oracle and UT Adjunct Professor
CS347 will prepare you for this.
CS347 – Data Management

Growth of Data vs. Growth of Data Analysts

Stored Data accumulating at 28% annual growth rate
Data Analysts in workforce growing at 5.7% growth rate

CS347 will prepare you for this.
Talent Remains Key Big Data Challenge

by BOB VI LINO
MAR 18, 2015 5:37PM ET

About two thirds of the IT executives rank big data architects as the most difficult role to fill. Data scientists (48%) and data modelers (43%) round out the top three most difficult positions to fill. More technical big data positions are ranked less difficult to fill.

Not by coincidence, big data companies are introducing online and face-to-face training programs and certifications for Hadoop and other related software platforms.

Still, other big data challenges remain. Variety—the dimension of big data dealing with the different forms of data—hinders organizations from deriving value from big data the most, according to 45% of those surveyed. Speed of data is next in terms of challenges, at 31%, followed by the amount of data, at 24%.

The application of big data is happening in a number of business areas, according to the study, with 81% of organizations viewing operations and fulfillment as priority areas within the next 12 months. This was followed by customer satisfaction (53%), business strategy (52%), governance/risk/compliance (51%) and sales/marketing (49%).

More than 200 IT leaders participated in the February 2015 survey.

CS347 will prepare you for these.
You will learn – SQL, Data Modeling, Web/Database Application Development, the Oracle DBMS Architecture, Transaction Processing, Indexing, NoSQL and Graph Databases, Normalization, and SQL Query Optimization.
Got this in the email last October

Dear Phil,

Please register today for our October online Oracle APEX classes...

<table>
<thead>
<tr>
<th>Course</th>
<th>Comments</th>
<th>Dates</th>
<th>Time</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction to APEX</td>
<td>78</td>
<td>Oct 6 - 10</td>
<td>1 PM - 5 PM EDT</td>
<td>$1,500</td>
</tr>
<tr>
<td>Advanced Oracle APEX</td>
<td>46</td>
<td>Oct 20 - 24</td>
<td>1 PM - 5 PM EDT</td>
<td>$1,500</td>
</tr>
<tr>
<td>HTML 5, CSS &amp; JavaScript (German)</td>
<td>(new)</td>
<td>Oct 6 - 10</td>
<td>12 PM - 5 PM GMT</td>
<td>$1,500</td>
</tr>
<tr>
<td>Introduction to SQL</td>
<td>38</td>
<td>Oct 6 - 10</td>
<td>1 PM - 5 PM EDT</td>
<td>$1,500</td>
</tr>
<tr>
<td>Advanced SQL Queries</td>
<td>74</td>
<td>Oct 20 - 24</td>
<td>1 PM - 5 PM EDT</td>
<td>$1,500</td>
</tr>
</tbody>
</table>

Online, instructor-led, hands-on class taught by APEX and Oracle experts Tyson Jouget, Christian Rokitta and Geoff Wiland.

Register online or call 1.888.803.5607. Call for group and alumni discounts.

Remember: Call us if you need an APEX DBA or Developer!
Continuing Data Management Research

Possible Opportunities:

• WDB Open Source Project

• SQL to SPARQL translation, see next 2 pages for examples

• SIM Query Language to SPARQL translation

  RETRIEVE Project-Title of Sub-Projects of Project, Name of Project-Team of Sub-Projects WHERE Project-Title of Project = “Annual Report Preparation”

• ReL Open Source Project

• All of these are probably publishable and would make very good cs370 projects, also.
5.1 Basic Example

SQL:

select deptno, sal from emp where SAL > 1000

SPARQL:

SELECT v1 "DEPTNO", v2 "SAL"
FROM TABLE(SEM_MATCH('SELECT ?v1 ?v2 WHERE
   ?s1 rdf:type :EMP .
   OPTIONAL { ?s1 :DEPTNO ?v1 } /* Select them */
   OPTIONAL { ?s1 :SAL ?v2 }
   OPTIONAL { ?s1 :SAL ?v3 }
   FILTER(?f1 > 1000)
)

5.2 Join Example

SQL:

select dname, sal from emp e, dept d where e.deptno = d.deptno where sal > 1000

SPARQL:

SELECT v1 "DNAME", v2 "SAL"
FROM TABLE(SEM_MATCH('SELECT * WHERE {
   ?s1 rdf:type :EMP .
   ?s2 rdf:type :DEPT .
   OPTIONAL { ?s2 :DNAME ?v1 }
   OPTIONAL { ?s1 :SAL ?v2 }
   FILTER(?f1 = d.deptno && ?f2 > 1000) }

Dr. Philip Cannata
SQL to SPARQL Translation Examples

5.3 Complete Example

SQL:

```
select deptno, avg(sal) from emp
group by deptno
having avg(sal) > 1000
order by avg(sal)
```

SPARQL:

```
SELECT v1 "DEPTNO", n2 "AVG(EMP.SAL)"
FROM TABLE(SEM_MATCH('SELECT ?v1 (avg(?v2) as ?n1) WHERE {
   ?s1 rdf:type EMP .
   OPTIONAL { ?s1 :DEPTNO ?v1 }
   OPTIONAL { ?s1 :SAL ?v2 }
   OPTIONAL { ?s1 :SAL ?v3 }
   OPTIONAL { ?s1 :SAL ?v4 }
   OPTIONAL { ?s1 :DEPTNO ?v5 }
} GROUP BY ?v5
HAVING( avg(?v3) > 1000)
ORDER BY avg(?v4)
''')
Final Review

- Review the Midterm, Homework (especially homework 8 and 9) and Quizzes (especially Quizzes 5, 6, and 7)
- SQL on the class website
- Normalization
- Query Execution Plans

~ 30%
Chia-Chen Hsu and Chris Timaeus
Midterm Exam Grade = max(Midterm Grade, Final Exam Grade)

\[
AB3 = (25 \times \text{MAX}((V3+8)/140, \text{AE3}))) + (20 \times (Y3/10)) + (15 \times (AA3/10)) + (15 \times (W3/100)) + \text{Round}
\]