|  |
| --- |
| Univerzita Pardubice |
| Datové sklady |
| Semestrální práce |
|  |
| **Matěj Trakal** |
| **23.4.2012** |

|  |
| --- |
|  |

Návštěvnost webového portálu

Společnost se zabývá vývojem webového portálu. Jelikož se snaží minimalizovat náklady, je pro ně důležité sledovat trendy na poli operačních systémů a používaných prohlížečů, jelikož by zbytečně mohli dodávat podporu pro systémy, které nikdo nevyužívá.

Dále je velice zajímá trend mobilních zařízení, ze kterých se na webový portál může přistupovat, jelikož optimalizace pro tato zařízení dělá jiné oddělení. Otázkou tedy je, zda se mobilní trh bude nadále vyvíjet a jestli se má tento vývojový tým do budoucna posilovat o nové tváře.

# Základní struktura datového skládku



Obrázek 1: Diagram tabulek

Jako dimenze bylo použito času (s přesností na měsíce), následně operační systém a prohlížeč a jejich verze. Jako fakta počet přístupů na webový portál a počet nových návštěvníků ve stejném období.

Tabulka faktů je tedy složena právě z těchto dimenzí a ukazuje nám, jaká je návštěvnost portálu pro vývojové oddělení a částečně pro vrcholový management. U vývojářů zobrazuje trendy, kam se ubírají návštěvníci a u managementu dokážeme z počtu návštěvníků určit úspěšnost webu.

# Výsledné dotazy nad datovým skládkem

## Vedení požaduje analýzu, jak se zvedá návštěvnost za poslední tři roky, které jsou v db pro jednotlivé OS.

select datum.rok, DECODE(GROUPING(os.os),1,'-- Za všechny',os.os) OS , sum(visitors)

from datovysklad ds, datum, os

where

 ds.datum = datum.datum\_id and

 datum.rok >= (select max(rok) from datum)-2

 and ds.os = os.os\_id

group by datum.rok, cube(os.os)

order by os,rok;

Je znát značný nárůst systému Android, následně Macintosh a postupné slábnutí Linuxu.

## Vedení zajímá trend, jak se mění četnost využívání prohlížečů a celkové návštěvnosti, jelikož je možné, že bude potřeba aktualizovat redakční systém, na kterém aktuálně funguje naše aplikace a webové rozhraní.

select

 DECODE(GROUPING(datum.rok),1,'-- Za všechny roky',datum.rok) ROK,

 DECODE(GROUPING(browser.browser),1,'-- Za všechny prohlížeče',browser.browser) PROHLIZEC,

 sum(visitors) NAVSTEV

from datum

left join datovysklad ds on datum.datum\_id = ds.datum

left join browser on browser.browser\_id = ds.browser

group by rollup(datum.rok, browser.browser)

order by prohlizec, rok;

Z grafu je vidět, že návštěvnost našeho portálu se neustále zvedá, opravdu hodně s prohlížeči za poslední dva roky zamíchal Chrome, tedy tam bude třeba zaměřit optimalizaci. Je znát postupný úpadek Firefoxu, tedy tým vývojářů pro Chrome nejspíše sestavíme s lidí vyvíjejících právě pro Firefox.

## Vedení zajímá, jaký bude trend a jaká bude návštěvnost v roce 2013 pro mobilní přístup, pokud se nezmění tempo růstu (pro výpočet berme v potaz poslední dva roky v DB).

WITH tab AS (

select

 rok rok,

 mobilni.mobilni mobilni,

 sum(visitors) visitors

from datum

left join datovysklad ds on datum.datum\_id = ds.datum

left join mobilni on mobilni.mobilni\_id = ds.mobilni

where mobilni.mobilni = 'ANO'

group by rok, mobilni.mobilni

order by rok

)

select rok, mobilni, round(visitors,0) from tab

model UNIQUE SINGLE REFERENCE

 partition by (mobilni)

 dimension by (rok)

 measures (visitors)

 rules (

 visitors[2012] = (visitors[2011]/visitors[2010])\*visitors[2011],

 visitors[2013] = (visitors[2012]/visitors[2011])\*visitors[2012]

 )

Z grafu je jasně vidět, že trend mobilních prohlížečů se dá předpokládat jako silně rozvíjející a má rozhodně význam na těchto platformách zapojit více vývojářů, aby již s programováním pro mobilní aplikace byli seznámeni, a může se zrovna postupně začít vyvíjet.

## Vedení zajímá, zda již klesá podíl OS Windows XP, který začíná být zastaralý, jelikož by již rádi ukončili podporu IE6, který je v této verzi OS jako defaultní.Statistiku požadují za posledních 10 měsíců.

with maxDatum as (select max(mesic) mesic, rok from datum where rok = (select max(rok) rok from datum) group by rok),

tab AS (

select

 datum.mesic,

 os.os,

 osversions.version,

 sum(visitors) visitors

from maxDatum,datum

left join datovysklad ds on datum.datum\_id = ds.datum

left join os on os.os\_id = ds.os

left join osversions on osversions.version\_id = ds.os\_version

where

 datum.mesic >= maxDatum.mesic-10 and

 datum.rok = maxDatum.rok and

 os.os = 'Windows' and

 osversions.version = 'XP'

group by datum.mesic, os.os, osversions.version

order by datum.mesic

)

select mesic, os, version, round(visitors,0) from tab;

Z grafu je vidět prázdninový pokles, kdy je známý snížený počet návštěvností, jinak je graf postupně klesající. Tento trend se dá odhadovat I do budoucna a postupně budeme moci snižovat počet vývojářů, kteří provádějí optimalizace pro tento operační systém. Dále je bude možné přesunout do nově vznikajících projektů.

## Zobrazte, jak se mění počet návštěvníků z Windows za poslední rok.

CREATE MATERIALIZED VIEW sem05 REFRESH WITH ROWID AS

with tab AS (

select

 datum.mesic,

 os.os,

 sum(visitors) visitors

from datum

left join datovysklad ds on datum.datum\_id = ds.datum

left join os on os.os\_id = ds.os

where

 os.os = 'Windows'

group by datum.mesic, os.os

order by datum.mesic

)

select mesic, os, round(visitors,0) visitors from tab;

## Zobrazte v posledním roce, jak se mění využití webového portálu s tím, že výsledky uvedete v procentech a s přesností na měsíce. Tato statistika zajímá marketingové oddělení, zda se má pokusit vytvořit reklamu na novou funkci na portálu, která je zaměřena právě na mobilní aplikace. Pokud by byl trend klesající, byl by zájem o opětovné podpoření návštěvnosti.

select datum.mesic,

round(percent\_rank() over (order by sum(ds.visitors)),3)\*100 procentualne

from datum

left join datovysklad ds on datum.datum\_id = ds.datum

where datum.rok = (select max(rok) from datum)

group by datum.mesic

order by datum.mesic

Z grafu vyplývá, že od zavedení nové funkce v 8. měsíci návštěvnost opět začíná klesat. Tedy marketingové oddělení zkusí vice zacílit reklamou na majitele mobilních telefonů, aby opět zvedli návštěvnost svého portálu.

## Získejte pro vývojáře data, jaké se používají verze prohlížečů. Chtějí je seřadit od nejpoužívanějších po nejméně, zajímá je prvních 15 výsledků a pouze letní měsíce (červenec, srpen), když jsou nejvíce vybírané dovolené (předpokládá se zvýšená návštěvnost z domácích počítačů).

with subdotaz as (

select round(dense\_rank() over (order by sum(ds.visitors) desc),3) poradi, browser.browser prohlizec, browserversions.version verze, sum(ds.visitors) pocet\_navstev

from datum

left join datovysklad ds on datum.datum\_id = ds.datum

left join browser on browser.browser\_id = ds.browser

left join browserversions on browserversions.version\_id = ds.browser\_version

where datum.rok = (select max(rok) from datum) and datum.mesic between 7 and 8

group by browser.browser, browserversions.version

order by poradi

)

select \* from subdotaz where poradi <= 15;



## Zobrazte trend verzí prohlížeče Chrome, za poslední rok, z grafu by mělo být patrné, kdy nastoupila nová verze prohlížeče a kdy pomalu zanikají jeho starší verze.

with tab AS (

select

 datum.rok rok,

 datum.mesic mesic,

 browser.browser\_id brID,

 browser.browser br,

 browserversions.version\_id verID,

 browserversions.version ver,

 sum(visitors) visit

from datum

left join datovysklad ds on datum.datum\_id = ds.datum

left join browser on browser.browser\_id = ds.browser

left join browserversions on browserversions.version\_id = ds.browser\_version

where

 datum.rok >= (select max(rok) from datum) and

 browser.browser = 'Chrome'

group by datum.rok, datum.mesic, browser.browser\_id, browser.browser, browserversions.version\_id, browserversions.version

order by browser.browser, browserversions.version, datum.rok, datum.mesic

)

select \* from tab

where visit > 50;

Výsledků z vymyšleného zadání je enormní množství (stovky), jelikož se často mění čísla verzí.

## Jaký bude trend na poli prohlížečů Chrome a Firefox, pokud budeme uvažovat poslední roční statistiky a chceme dopočítat další rok dopředu.

with tab AS (

select

 datum.rok rok,

 datum.mesic mesic,

 browser.browser\_id browser\_id,

 browser.browser browser,

 sum(visitors) visitors

from datum

left join datovysklad ds on datum.datum\_id = ds.datum

left join browser on browser.browser\_id = ds.browser

left join browserversions on browserversions.version\_id = ds.browser\_version

where

 datum.rok = (select max(rok) from datum) and

 browser.browser = 'Chrome' or

 browser.browser = 'Firefox'

group by datum.rok, datum.mesic, browser.browser\_id, browser.browser

order by datum.rok, datum.mesic, browser.browser

)

select rok, mesic, browser,round(visitors,0) from tab

model UNIQUE SINGLE REFERENCE

 partition by (browser)

 dimension by (rok,mesic)

 measures (visitors)

 rules (

 visitors[2012,mesic] = (visitors[cv(rok)-1,cv(mesic)] / visitors[cv(rok)-2,cv(mesic)])\*visitors[cv(rok)-1,cv(mesic)]

 )

order by rok, mesic, browser

# Získání podkladových dat

Vstupní data jsem exportoval z reálné databáze na Google Analytics, pomocí externí aplikace pro získávání těchto dat do tabulkového procesoru Excel. Jedná se o aplikaci Excellent Analytics, která dokáže z původních dat exportovat již souhrn, dle zadaných kritérií, jelikož surová data ani nejsou poskytována. Z této aplikace jsem dokázal získat CSV soubor, který jsem importoval do mého databázového prostoru a data převedl do normálních forem, aby se s nimi dalo lépe pracovat.

<http://excellentanalytics.com/>