

Prístupy k zdieľaniu informácií v distribuovaných projektoch

MIROSLAV VNUK

*Slovenská technická univerzita
Fakulta informatiky a informačných technológií
Ilkovičova 3, 842 16 Bratislava*

Abstrakt. Práca rozoberá a analyzuje rôzne možné prístupy k uchovávaniu a zdieľaniu informácií, ktoré boli nadobudnuté počas riešenia rôznych na sebe viac či menej závislých projektov, riešených v distribuovanom prostredí. Tieto informácie kategorizuje a na nich popisuje klady a zápory jednotlivých prístupov. Na záver popisuje reálne systémy, ktoré implementujú uvedené prístupy.

Úvod

Dnešný svet patrí informatike a informáciám. V tomto svete majú informácie vysokú cenu, keď sa nachádzajú na správnom mieste v správnych rukách. Nastáva však problém, ako žiadanú informáciu vhodne rozdistribúovať medzi ľuďmi, ktorí majú o nej vedieť. Touto problematikou sa zaoberá táto práca a pokúša sa vniesť pohľad na prístupy k zdieľaniu informácií (nielen) v distribuovaných projektoch.

Informácie

Na začiatok by som chcel uviesť význam slova informácia v akom je v práci použitý. Význam informácie je chápaný v zmysle užitočných významných poznatkov nadobudnutých riešením rôznych projektov. Môže sa jednať o poznatky nielen iba z určitej časti riešeného projektu, ale o ucelené poznatky nadobudnuté počas riešenia projektu počas celého jeho životného cyklu. Na základe takto definovaného významu informácie je možné samotné informácie deliť do troch skupín [3].

Prvá skupina predstavuje *informácie nadobudnuté počas riešenia projektov*. Zahŕňa technickú dokumentáciu k projektu, ďalej rôzne použité študijné materiály použité k vyriešeniu projektu a časť riadiacej dokumentácie – záznamy z porad (hlavne čo sa týka technickej realizácie projektu).

Druhá skupina, do ktorej možno klasifikovať *informácie reprezentuje poznatky získané z manažovania projektov*. V tomto prípade je to najmä riadiaca dokumentácia vytváraná k príslušnému projektu zahŕňajúca najmä informácie o časovom plánovaní jednotlivých úloh, stave ich plnenia, deadlines. Ďalej zahŕňa rozdelenie príslušných počtov ľudí na príslušné typy úloh vrátane ich zaradenia na konkrétne dielčie úlohy. Táto skupina informácií by mala obsahovať aj informácie týkajúce sa financií celého projektu (napríklad: návratnosť investícií, finančné prostriedky vynaložené na riešenie dielčích úloh, ...).

Tretia posledná kategória, do ktorej je možné informácie klasifikovať je predstavuje *informácie získané po ukončení príslušného projektu*. Jedná sa o rôzne post analýzy a auditu úspešnosti projektu zamerané či už na menšie časti, alebo na celý projekt.

Delenie prístupov k zdieľaniu informácií

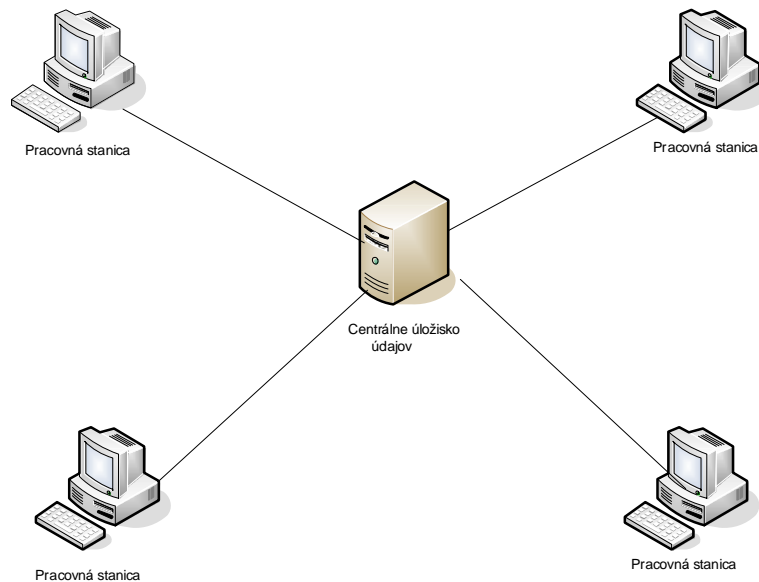
Takto kategorizované informácie je potrebné zdieľať medzi rôznymi riešiteľmi projektov (čím sa uľahčí riešenie „už raz“ riešených častí, aj keď boli riešené na inom projekte). Samotné zdieľanie týchto informácií je možné realizovať rôznymi spôsobmi, ktoré sa líšia neskôr uvedenými vlastnosťami.

Prístup k zdieľaniu informácií pri riešení projektov v distribuovanom prostredí (rôzne tímy riešia rôzne projekty v rôznych, vzdialených lokáciách) je možné rozdeliť do troch skupín. Prvý prístup je možné nazvať *centrálny model* a ako už názov napovedá centralizovane uchováva informácie. Druhý prístup vychádza z P2P (peer-to-peer) architektúry, z toho je odvodené aj jeho meno – *P2P model*. Tretí posledný prístup k zdieľaniu informácií reprezentuje *hybridný model*, ktorým reprezentuje istú kombináciu prvých dvoch uvedených prístupov. V tejto práci bude uvedený jeden konkrétny hybridný prístup, s konkrétnymi vlastnosťami, pričom vo všeobecnosti v tomto prístupe návrhu môžu prevažovať viac či menej vlastností prvého alebo druhého prístupu.

Centrálny model

Na opis vlastností uvedeného prístupu, reprezentovaného týmto modelom, sa dá pozerieť z rôznych pohľadov - abstraktný pohľad, nepozerať na samotnú technickú implementáciu a jej vlastnosti, ktoré sa môžu v konkrétnych systémoch líšiť, ale popisuje všeobecné vlastnosti prístupu, ktoré pokrývajú takmer všetky konkrétne systémy založené na príslušnom modeli. Na druhú stranu je to konkrétny pohľad na jednotlivé systémy alebo na skupinu rovnakých riešení. Predstavuje v podstate opak abstraktného pohľadu.

Na začiatok sa budem venovať abstraktnému pohľadu na vymenované prístupy a začnem centrálnym modelom.

**Obr. 1.** Centrálny model zdieľania dát

Aj v tomto prípade sa dá na výhody a nevýhody tohto prístupu pozerať z viacerých hľadísk: *zdieľanie dát*, *riadenie prístupu* a *štruktúra dát*. Z pohľadu *zdieľania dát* vyplýva jednoznačná výhoda a tou je oddelenie informácií od používateľa a ich centrálné uchovávanie. Jednou z výhod z toho vyplývajúcich je nezávislosť informácií príslušného používateľa od aktuálneho stavu jeho počítača (či je vypnutý, zapnutý, funkčný, nefunkčný). Ďalšou výhodou centrálného uchovávanie je fakt, o ktorom vedia všetci, ktorí systém používajú, a tým je jednoznačné miesto, kde požadované informácie hľadať a tým sa v podstate jedná o štandard, ktorí všetci používajú.

Nakoľko všetko má aj isté nevýhody, aj tento prístup z pohľadu zdieľania dát má isté nevýhody. Jednou z nich je relatívne dlhý časový úsek medzi nadobudnutím informácie a jej zdieľaním medzi používateľmi, čím hrozí v istom zmysle strata, alebo znehodnotenie informácie. Ďalšou nevýhodou je istá strata kontroly používateľa nad ním vytvorenými informáciami, ktorá sa dá riešiť zavedením istých prístupových práv.

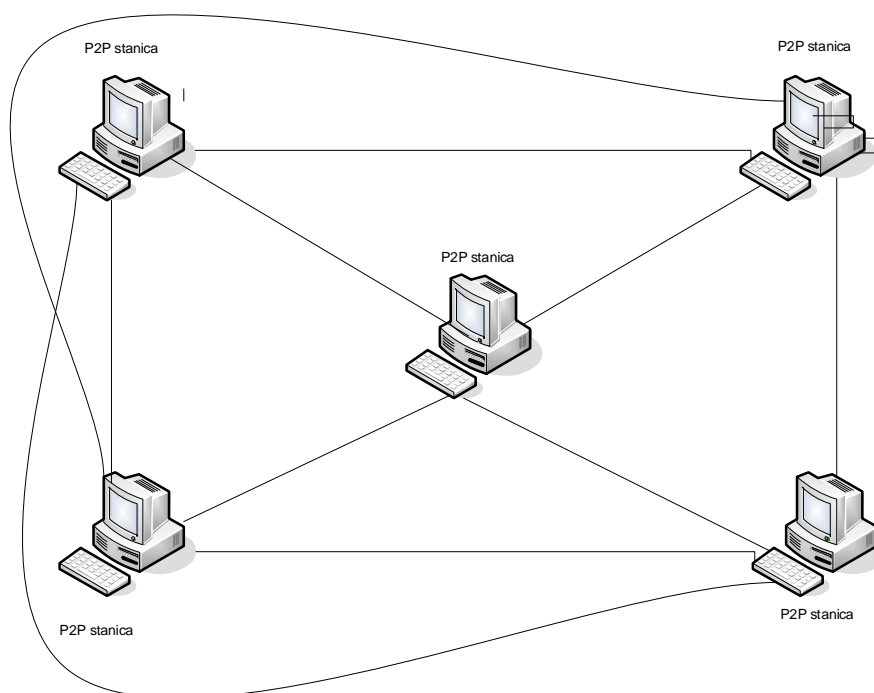
Z pohľadu riadenia prístupu vzišla azda len jedna už spomenutá nevýhoda, a tou je riadenie prístupu kvôli spätnej kontrole nad informáciami používateľom, ktorý príslušnú informáciu vytvoril. Z pohľadu štruktúry informácií má tento prístup asi najvýznamnejšie žiadané vlastnosti, a tými sú pevná organizácia štruktúry informácií, optimalizovaná najmä na rýchlosť, použitie rôznych druhov filtrov a vyhľadávacích schém. S tým sú ale spojené určité nevýhody, najmä komplikovanosť návrhu dátovej štruktúry použitej na uchovávanie rôznych informácií. Ďalšia nevýhoda je spôsobená samotnou pevnou štruktúrou informácií, a tou je nedostačujúca variabilita a tým

nemožnosť zaznamenávať všetky typy informácií, čím môže v určitom zmysle dôjsť k strate informácie.

Z uvedených výhod a nevýhod vyplýva použitie tohto prístupu na uchovávanie informácií, ktoré sa príliš často nemenia a nemajú „príliš veľkú“ variabilitu, a to sú *informácie získané z manažovania a po ukončení projektov.*

P2P (peer-to-peer) model

Ďalším možným, už spomínaným prístupom k zdieľaniu informácií, je model P2P (peer-to-peer).



Obr. 2. P2P model zdieľania dát

Z pohľadu zdieľania informácií sa tento prístup ukazuje ako vhodnejší aj oproti predchádzajúcemu prístupu v minimalizovaní časového úseku medzi nadobudnutím informácie a jej zaznamenaním a určením pre zdieľanie pre ostatných používateľov. Ukazuje sa, že osoba, ktorá nadobudne nejakú informáciu, ju radšej a hlavne rýchlejšie zdieľa u seba ako v nejakom centrálnom dátovom sklade (rozdiel oproti *centrálnemu modelu*). Z hľadiska riadenia centrálnemu prístupu je tento model oproti predchádzajúcemu hlavne jednoduchší na realizáciu, keďže netreba centrálnu určiť a riadiť prístup. Každý používateľ je zodpovedný za svoje informácie, ktoré zdieľa

a kto má k nim prístup. Pri zmene informácie znovu nenastáva žiadny problém, jednoducho si svoje informácie prepíše, keďže každý sám si spravuje svoje informácie.

Z hľadiska štruktúry dát je tento prístup komplikovanejší ako predchádzajúci, nakoľko jednotliví prispievatelia informácií nepoužívajú jednotnú štruktúru ukladania a zdieľania informácií. Z toho v podstate plynú už uvedené výhody a ďalej vymenované nevýhody. Ako výhodu je ešte možné označiť vlastnosť zdieľať informácie s veľkou variabilitou a tým je možné citlivejšie určovať informačnú úroveň príslušnej zdieľanej informácie. Nevýhoda spôsobená vysokou variabilitou a nejednoznačnou štruktúrou je znemožnenie kvalitného a plnohodnotného vyhľadávania. Na vyhľadávanie je potrebné použiť rôzne špecializované agenty, ktoré sa každý zameriavajú na určitý typ informácií. Tým je samozrejme predĺžená dĺžka vyhľadávania a ochudobnená možnosť použitia filtrov a vyhľadávacích schém.

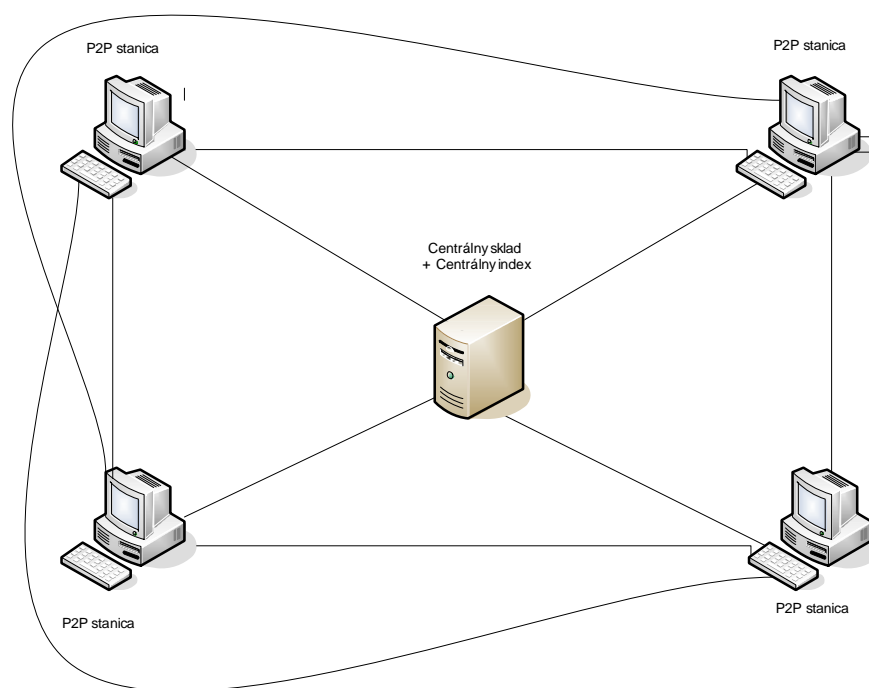
Z uvedených výhod a nevýhod vyplýva, že tento prístup k zdieľaniu informácií je najvhodnejšie použiť na zdieľanie informácií vykazujúcich veľkú variabilitu a časté obmieňanie. Z uvedenej klasifikácie typov informácií je tento systém najvhodnejšie použiť na zdieľanie *informácií získaných počas riešenia projektov*.

Hybridný model

Posledným možným, aj keď nie až tak jednoznačným prístupom z hľadiska zdieľania dát, je hybridný model. Nejednoznačnosť modelu vyplýva z možnosti existovania množstva hybridných modelov kombinujúcich predchádzajúce prístupy, pričom každý môže mať iné vlastnosti. Následne je uvedený jeden z možných modelov v ktorom je snaha o čo najideálnejšiu kombináciu predchádzajúcich modelov, s ohľadom na uvedenú klasifikáciu typov informácií. Pokiaľ by sa zmenili typy informácií, možno by bolo vhodné zvoliť iný spôsob kombinácie modelov.

Konečná podoba systému obsahuje tri časti: *centrálny sklad* informácií (časť prevzatá z *centrálneho modelu*), stanice prepojené P2P obsahujúce *distribuované sklady* informácií (časť prevzatá z *modelu P2P*) a *centrálny index*. Centrálny index je v podstate dátový sklad riešený rovnako ako *centrálny model*, ale majúci formu slovníka, ktorého jednotlivé položky odkazujú najmä na informácie uložené v dátových skladoch na P2P staniciach. Najvýznamnejšou nevýhodou tohto popísaného prístupu je komplikovanejšia fyzická realizácia.

Tento typ hybridného modelu využíva výhody centrálneho modelu na zdieľanie informácií, ktoré sa nemenia príliš často a majú relatívne pevnú dátovú štruktúru (*informácie získané z manažovania projektov a informácie získané po ukončení projektov*). Tieto výhody kombinuje s výhodami P2P modelu, čiže zdieľanie informácií s variabilnou štruktúrou, čo v tomto prípade predstavujú *informácie získané počas riešenia projektov*.



Obr. 3. Hybridný model zdieľania dát

Použitie v praxi

Popísané prístupy v zdieľaní informácií neexistujú len v teoretickej rovine, ale existujú ale aj ako reálne systémy. Samozrejme, nemusia mať všetky vymenované vlastnosti uvedených prístupov.

Uvedené prístupy vôbec nemusia byť použité pri zdieľaní informácií z projektov v distribuovanom prostredí, ale môžu byť použité na zdieľanie akýchkoľvek informácií v distribuovanom prostredí, ale v tomto prípade poslúžia ako vhodné ukážky z praxe.

Ako príklad centrálného modelu zdieľania dát môže byť prezentovaný systém FTP (File Transfer Protocol), v ktorom sa komunikácia uskutočňuje iba medzi FTP serverom a klientom. Ak si potom dvaja používatelia chcú zdieľať súbor s využitím protokolu FTP, najskôr musí jeden z nich pomocou svojho FTP klienta nahráť tento súbor na FTP server a cieľový prijímateľ si ho zas musí pomocou svojho FTP klienta stiahnuť. Priame spojenie medzi FTP klientmi nie je možné. [2]

Príkladom P2P modelu by mohol byť systém Gnutella. V ňom jednotliví používatelia siete Gnutella predstavujú rovnocenné uzly. Ak chce používateľ – uzol A získať prístup do siete, musí nájsť aspoň jeden ďalší uzol siete - uzol B. Na hľadanie funkčných uzlov môže poslúžiť zoznam dodávaný priamo s aplikáciou alebo

vyhľadávací systém uzlov. Po nadviazaní spojenia pošle uzol B uzlu A svoj vlastný zoznam a uzol A sa pokúsi nadviazať spojenie aj s nimi. Pri vyhľadávaní v sieti potom pošle uzol A požiadavku všetkým uzlom, s ktorými je aktívne spojený, a tie smerujú jeho požiadavku ďalej. Teoreticky by sa táto požiadavka mala dostať postupne ku všetkým klientom, ale v praxi väčšina z nich neobsiahne ani polovicu všetkých uzlov. Ak používateľ nájde viac uzlov, ktoré poskytujú ten istý súbor, môže využiť segmentové sťahovanie a sťahovať súčasne časti súboru z rôznych uzlov, čím sa zvýši celková rýchlosť. [2]

Posledný príklad zdieľania informácií reprezentovaný hybridným modelom predstavuje systém DC++, ktorý sa spolieha na špeciálne servery, po anglicky nazývané huby (rozbočovače), ktoré súžia ako zhromaždisko používateľov. Jednotlivé huby medzi sebou nekomunikujú. Hub udržiava zoznam všetkých prihlásených používateľov a nimi poskytovaných, zdieľaných súborov a zabezpečuje vyhľadávanie medzi týmito súborami. Výmena súborov sa vykonáva už priamo medzi jednotlivými používateľmi.[2]

Záver

Ako vidieť uvedené prístupy k zdieľaniu informácií nemusia sa na sto percent odrážať v konkrétnych technických realizáciách. Majú poslúžiť na načrtnutie hlavných prúdov v prístupe k zdieľaniu informácií. To, ktorý si príslušná osoba, tím, firma vyberie, závisí od charakteru zdieľaných informácií a od požiadaviek, ktoré sa od toho očakávajú. Na to potom nadväzuje úspech či neúspech príslušného prístupu v danom prostredí a z toho čiastočne vyplývajúci úspech alebo neúspech riešených projektov.

Použitá literatúra

1. Kevin C. Desouza, J. Roberto Evaristo: Managing Knowledge in Distributed Projects. In *Communications of the ACM*, Vol. 47, No. 4 (April 2004), 87-91
2. Ondík M.: P2P – história a súčasnosť. *PC Revue*, No. 3 (2005) 12-15.
3. Schindler, M. and Damm, D. Security issues of a knowledge medium for distributed project work. *Intern. J. of Project Management* 20 (2002), 37-47.

Annotation

Approaches to sharing information in distributed projects

This document analyses and describes various approaches in storing and sharing informations generated by projects solved in distributed environment. It categorizes informations and describes pros and cons of various approaches of sharing these informations. In the end it describes real systems, using analyzed approaches.