



HUMUSOFT s.r.o.

Paralelní výpočty ve finančnictví

Jan Houška

HUMUSOFT s.r.o.

houska@humusoft.cz



Výpočetně náročné úlohy

- **distribuované úlohy**
 - mnoho relativně nezávislých úloh
 - snížení zatížení klientské pracovní stanice
 - využití výpočetního výkonu nepoužívaných stanic
- **paralelní úlohy**
 - úlohy s velkou náročností na paměť
 - obtížně řešitelné nebo neřešitelné na jednom počítači



Výpočty v reálném čase

- **důležitá je nejen hodnota výsledku, ale také čas jeho získání**
 - výsledek dodaný po termínu je hodnocen jako by byl nepřesný
- **intra-day analýzy**
- **měsíční uzávěrky**
 - zkrácení doby výpočtu umožňuje zahrnout novější data
- **poskytování pokročilých informací vybraným klientům**
 - on-line analýzy podle požadavků klienta

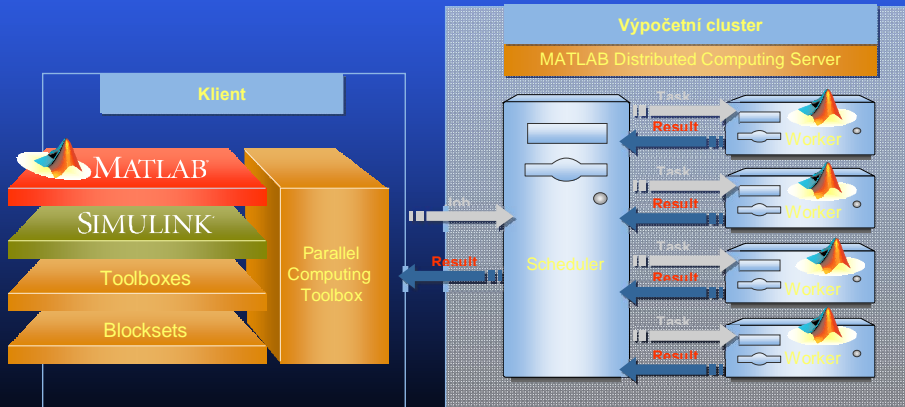


Metody typu Monte Carlo

- **definice množiny možných vstupních hodnot**
 - obvykle velmi rozsáhlá množina
- **deterministický výpočet s náhodně vybranými vstupy**
 - mnohonásobné opakování stejného výpočtu
 - žádná závislost mezi jednotlivými běhy
 - zvýšení počtu opakování zvyšuje věrohodnost výsledku
- **vyhodnocení výsledků**
 - hledání společných vlastností výsledků
 - hledání závislostí mezi vstupy a výstupy



Distribuované výpočty v MATLABu



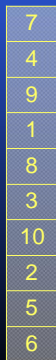


Paralelní cyklus `parfor`

```
for x = 1:10  
    disp(x)  
end
```



```
parfor (x = 1:10)  
    disp(x)  
end
```



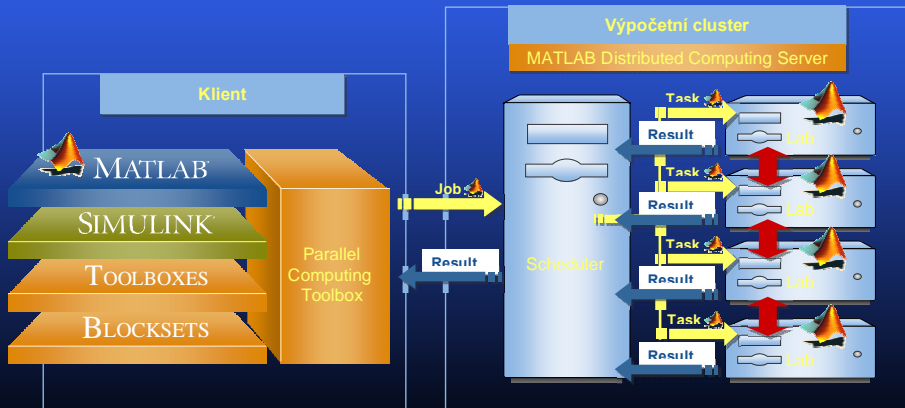


Tipy pro psaní distribuovaného kódu

- **identifikujte kód vhodný k distribuci**
 - často se pozná podle smyčky `for`
 - může jít nahradit příkazem `parfor`
- **distribuce má vždy nějakou režii**
 - úlohy by neměly být příliš malé
 - úloha by měla běžet nejméně 30 sekund
 - více v rozsáhlejších clusterech
 - je vhodné sloučit více malých opakujících se úloh do jedné větší
- **parametry jsou vždy předávány úlohám jako celek**
 - pozor na velké objemy dat při předávání parametrů
 - je výhodné využít sdílený systém souborů



Paralelní aplikace



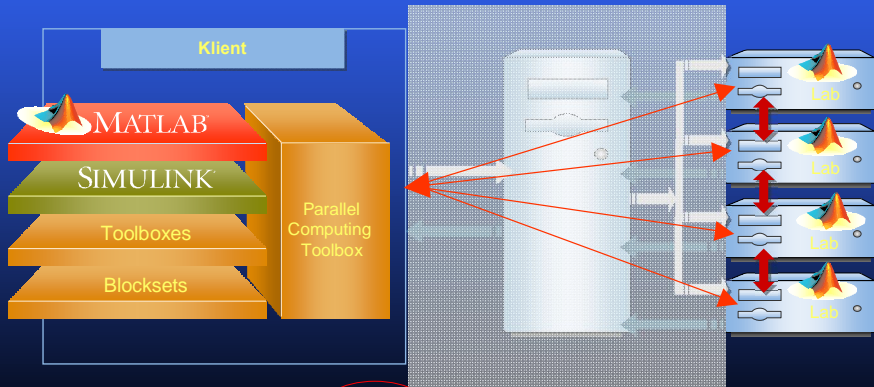


Paralelní aplikace

- **komunikace založená na zprávách**
 - protokol MPI
- **interaktivní režim zajišťuje většinu komunikace automaticky**
 - příkaz `pmode`
- **přímá podpora v mnoha vestavěných funkcích**
 - více než 150 funkcí
 - podpora knihoven ScaLAPACK



Interaktivní paralelní režim výpočtu (pmode)



P>>> E = D'



Distribuovaná pole

- **datová pole vytvořená na výpočetním clusteru**
 - každý uzel clusteru
- **mechanismus distribuce je pro uživatele transparentní**
 - `A = rand(10, 10, distributor())`
 - pole distribuované po sloupcích
 - `sum(A)`
 - sčítání po sloupcích
 - každý uzel sečte “svoje” hodnoty
 - `sum(A, 2)`
 - sčítání po řádcích
 - první uzel sečte všechny hodnoty, ostatní uzly mu je pošlou



Parallel Computing Toolbox 3.3

- **nová verze úspěšného produktu**
 - dříve Distributed Computing Toolbox
- **interaktivní paralelní příkazová řádka**
 - režim `pmode`
- **distribuovaná data**
 - rozsáhlé pole distribuované mezi výpočetními uzly
 - každý uzel zpracovává svou část pole
- **nové prvky jazyka MATLABu**
 - paralelní cyklus `parfor`
- **podpora 64-bitových platforem**



MATLAB Distributed Computing Server

- **dříve MATLAB Distributed Computing Engine**
- **serverová část výpočetního clusteru The MathWorks**
- **podpora různých schedulerů**
 - interní scheduler
 - Microsoft Compute Cluster Server
 - schedulery dalších výrobců
- **heterogenní cluster**
 - uzly 32 nebo 64 bitů, různé operační systémy
- **centrální správa licencí**
 - výpočetní uzly **nepotřebují** licenci MATLABu
- **dynamické licencování závislých produktů**
 - produkty dostupné na klientu jsou automaticky dostupné na výpočetních uzlech



Pracovní stanice Humusoft HeavyHorse

- **vysoce výkonný počítač pro paralelní výpočty**
- **modely s čtyřmi až šestnácti procesory**
 - založeno na technologii AMD Opteron
- **64-bitový operační systém**
 - Windows Vista 64-bit
 - Windows XP 64-bit Edition
 - OpenSuSE Linux 64-bit
 - Ubuntu Linux 64-bit
- **optimalizovaný pro použití s MATLABem**
 - možnost předinstalace MATLABu a Parallel Computing Toolboxu
- **usnadňuje vstup do světa paralelních výpočtů**



Pracovní stanice Humusoft HeavyHorse

HUMUSOFT s.r.o.

Model 16x354



- čtyři čtyřjádrové procesory AMD Opteron 8354, 2.2 GHz
- 32 GB RAM DDR II, 667 MHz
- grafická karta ATI Radeon
- pevný disk 500 GB
- operační systém podle přání
- možnost předinstalace aplikačního software
- příznivá cena