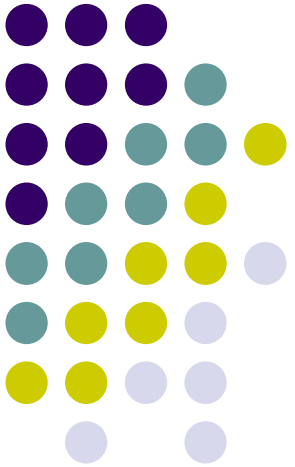
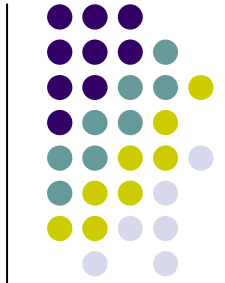
Provozní data a jejich vyhodnocováni

Jiří Bořík

borik@civ.zcu.cz



Agenda

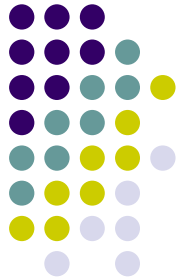


- Definice provozních dat
- Zdroje provozních dat
- Předzpracování a uložení
- Metody vyhodnocování
- Vizualizace výsledků

Motivace

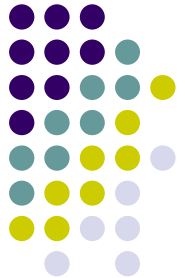
- Podobnost s reálným světem, sledujeme vitální funkce systému
- Různorodé proměnné veličiny, zobrazené pohromadě
- Přístrojová deska auta
- Nepřímý indikátor
- Cestující na zastávce
- Hledáme anomálie

Provozní data



- Co jsou to provozní data
  - Veškeré veličiny, které souvisejí s provozem systému
  - Obsahují informaci obvyklý/neobvyklý stav
- Zdroje provozních dat
  - Logy, monitorovací systémy, výkonové charakteristiky, počty událostí, uživatelů, dotazů…
- Příklady
  - Mrtg – základní výkonové údaje (load, traffic, mail queue…)
  - Windows - NSClient++ (Nagios klient)
  - Kerberos - počty přihlášení k jednotlivým strojům
  - Webové servery – počty přístupů k jednotlivým aplikacím
  - Oracle - Automatic Workload Repository (AWR)

Sběr a předzpracování dat



- Způsob sběru
  - Manuálně
    - pro dlouhodobější statistiky a periodické přehledy
    - testování vyhodnocovacích a zobrazovacích metod
  - Automaticky
    - Operativní vyhodnocení anomálií, včasné varování
- Předzpracování dat
  - Agregace
  - Normalizace
  - Identifikace zdroje, kategorizace údaje

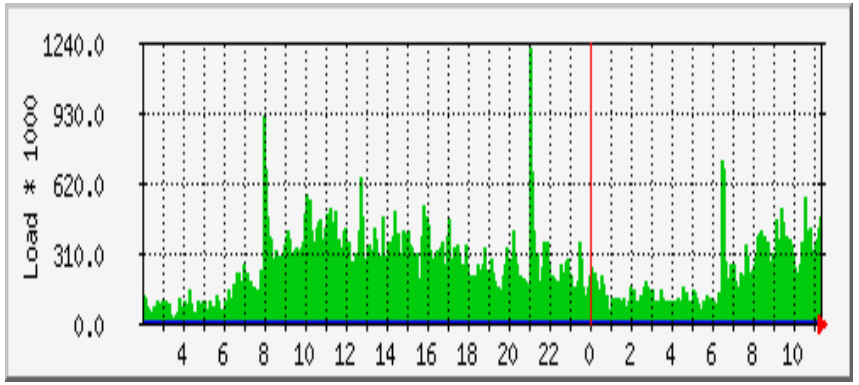
Uložení dat

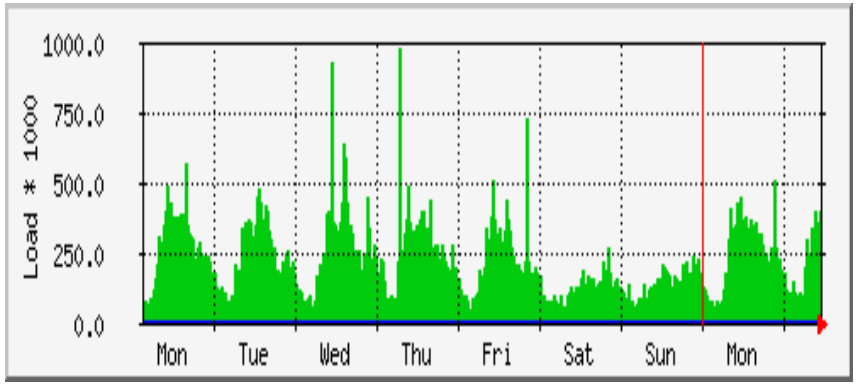
- Jednotná báze, rychlost a variabilita výpočtu = SQL
- Základní atributy:
  - Časový údaj (timestamp a délka měřené periody)
  - Hodnota měřené (testovaná) veličina
  - Původ veličiny (server, služba, kategorie)
  - Další charakteristiky

utm	dutm	sval	host	service
1294670400	300	410	axor	load
1294670400	300	28	fred	spam
1294670400	300	0	fred	virus
1294670400	300	100	axor	ups
1294670400	300	32	axor	ups2
1294670400	300	75	fred	mail
1294670400	300	32	fred	mqueue
1294670400	300	1429	fred	load
1294670700	300	1290	fred	load
1294670700	300	100	axor	ups
1294670700	300	32	axor	ups2
1294670700	300	20	fred	spam
1294670700	300	0	fred	virus
1294670700	300	80	fred	mail
1294670700	300	33	fred	mqueue
1294670700	300	429	axor	load

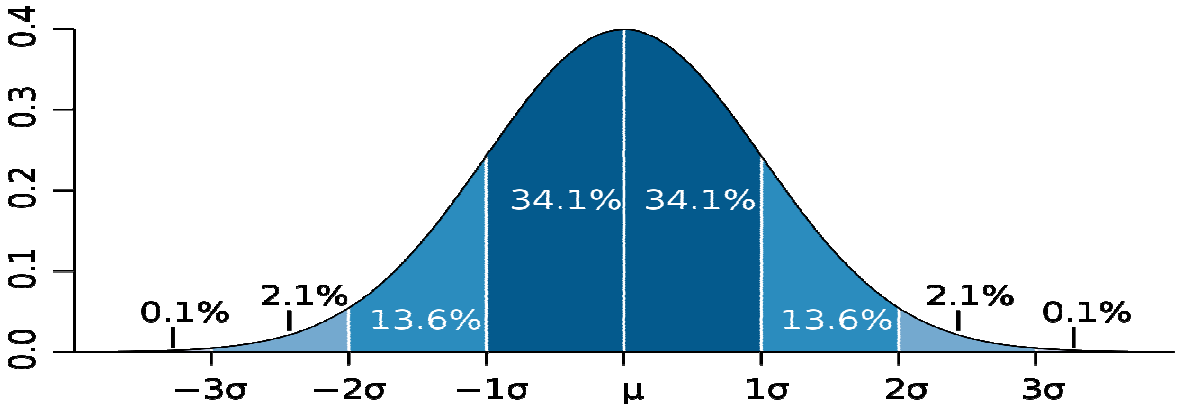
Zpracování dat

- Cíl: detekce anomálií
- Předpoklad: znalost obvyklého průběhu
- Odchylky na obě strany jsou podezřelé
- Lze vyhodnotit pomocí statistických parametrů (průměr a směrodatná odchylka)



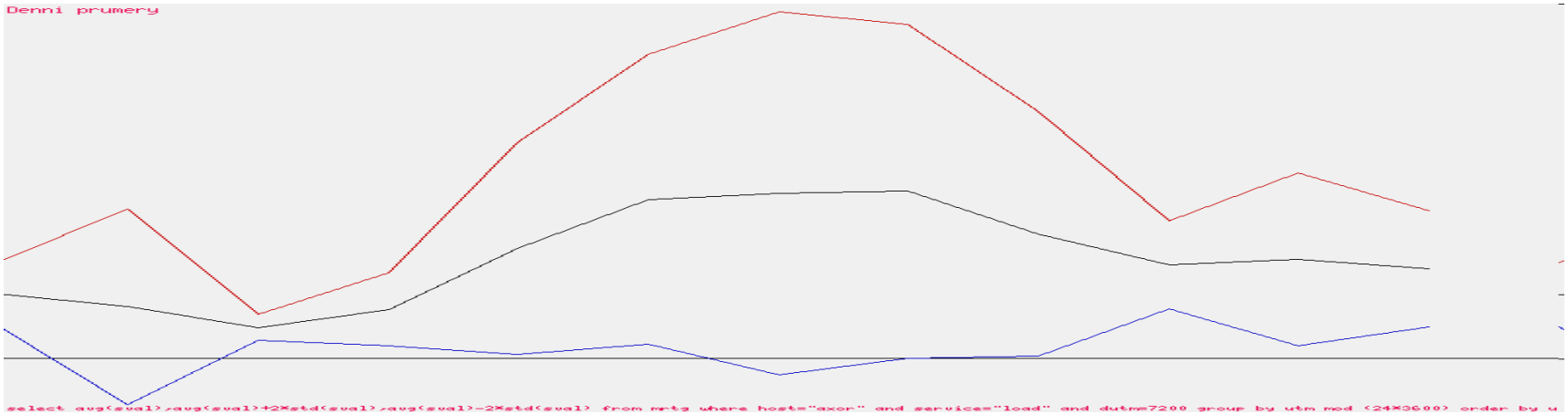


Průměr a odchylka



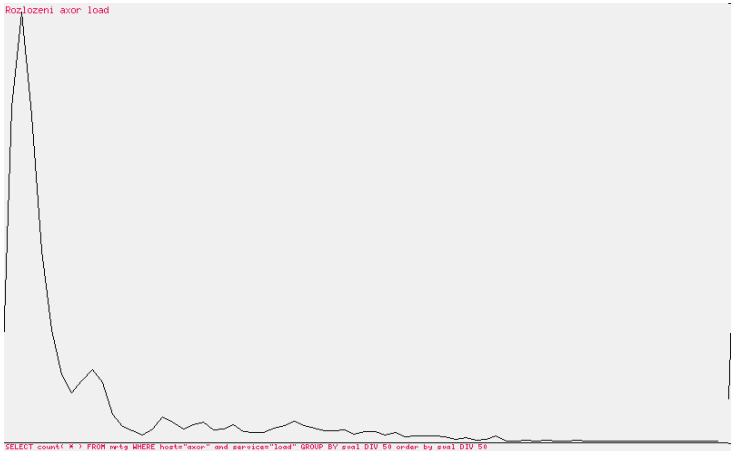
- Platí pro veličiny s normálním rozdělením
- V intervalu μ ± 2σ leží víc než 95 procent vzorků, zbylé okraje nás zajímají

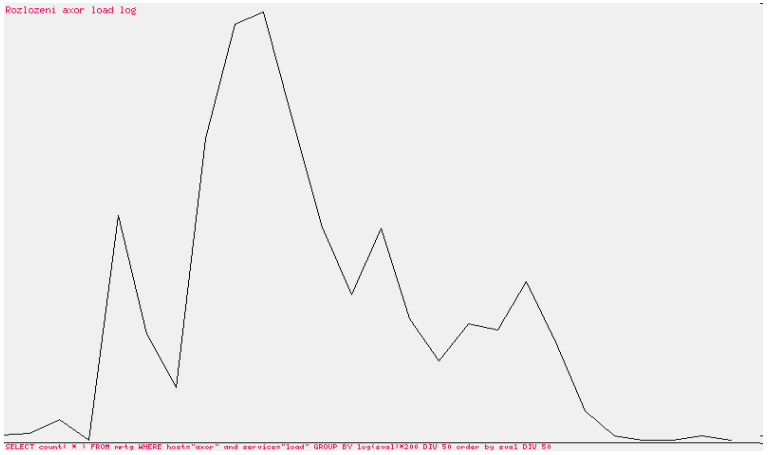
Použití pro provozní data



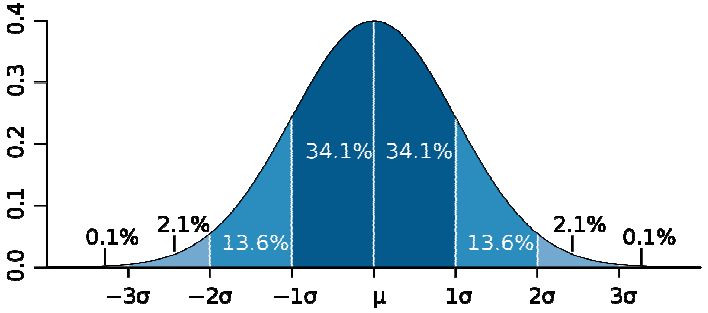
- Průměr a odchylka vytváří pásmo očekávaných hodnot
- Pásmo sahá dole až za nulovou osu – nelze vyhodnotit „dolní“ anomálie

Nedostatky řešení

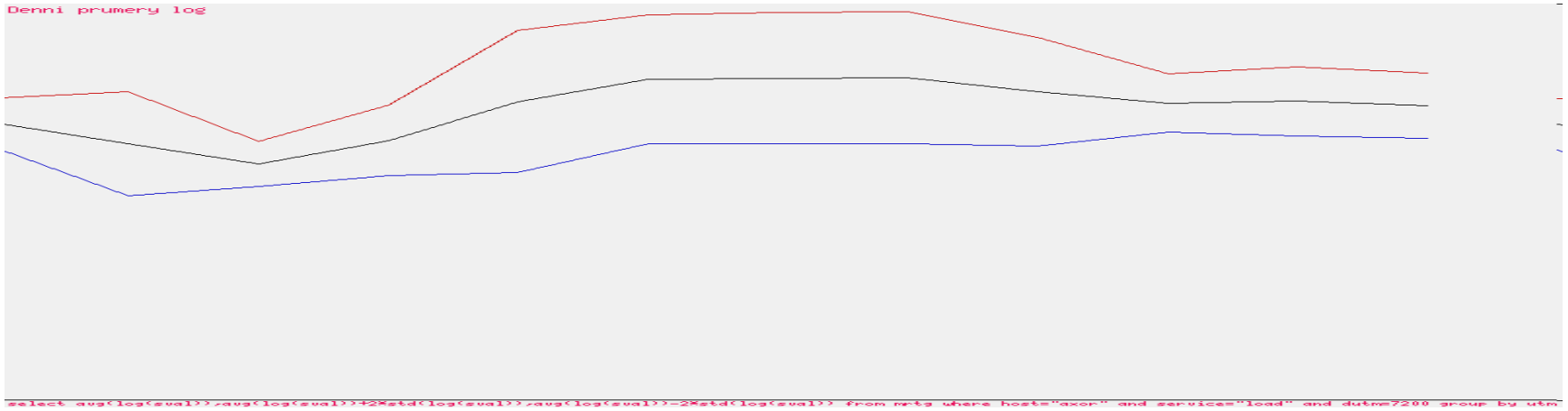




- Data nemají normální rozdělení
- Zdola 0, shora (téměř) neomezeno
- Rozdílná váha 10, 1010
- Použít logaritmické měřítko

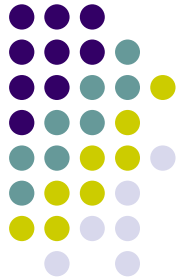


Upravený průběh



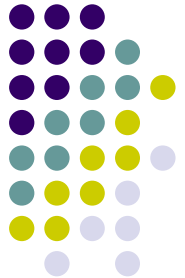
- Volba výpočtu průměru
  - celý interval,
  - denní perioda,
  - týdenní perioda

Základy SQL



```
select m.service,log(sval),a-2*s,a-s,a+s,a+2*s
from mrtg m,( select avg(log(sval)) a, std(log(sval)) s,
    host, service,utm mod (7*24*3600) utmd
        from mrtg where dutm=7200
        group by host,service,utm mod (7*24*3600)) p
where m.host=p.host and m.service=p.service and
    m.utm mod (7*24*3600)=p.utmd and dutm=7200
order by m.host,m.service,m.utm
```

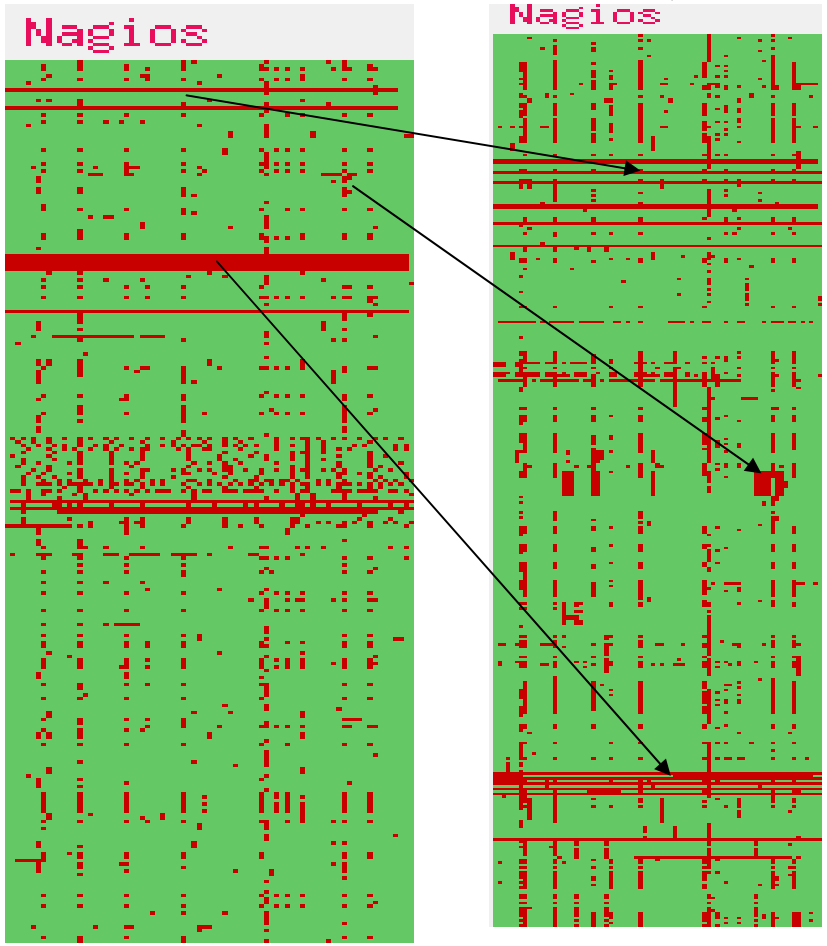
Zobrazení výsledků



- Hlavní účel - globální přehled
- Běžné grafy mají malou hustotu informací
- Použito plošné kompaktní zobrazení,
  - osa x čas,
  - osa y jednotlivé veličiny,
  - barva – hodnota veličiny nebo indikátor anomálie (pseudo 3D)

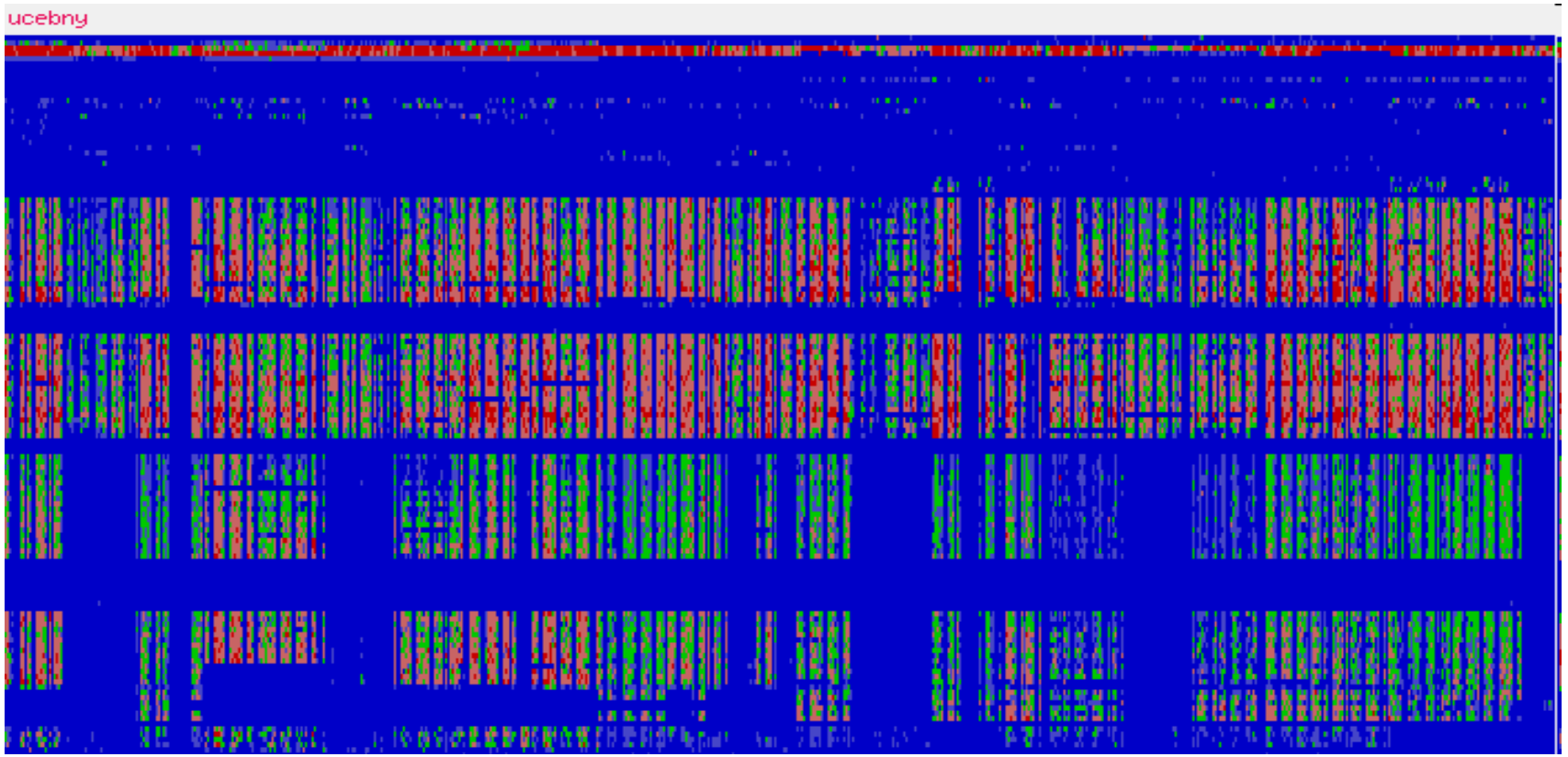
Příklad zobrazení: Nagios

- Pouze dvě hodnoty
- Celek/detail
  - Host status
  - Service status
- Agregace se zdůrazněním poruchy
- Denní stav (cca 80 dní)

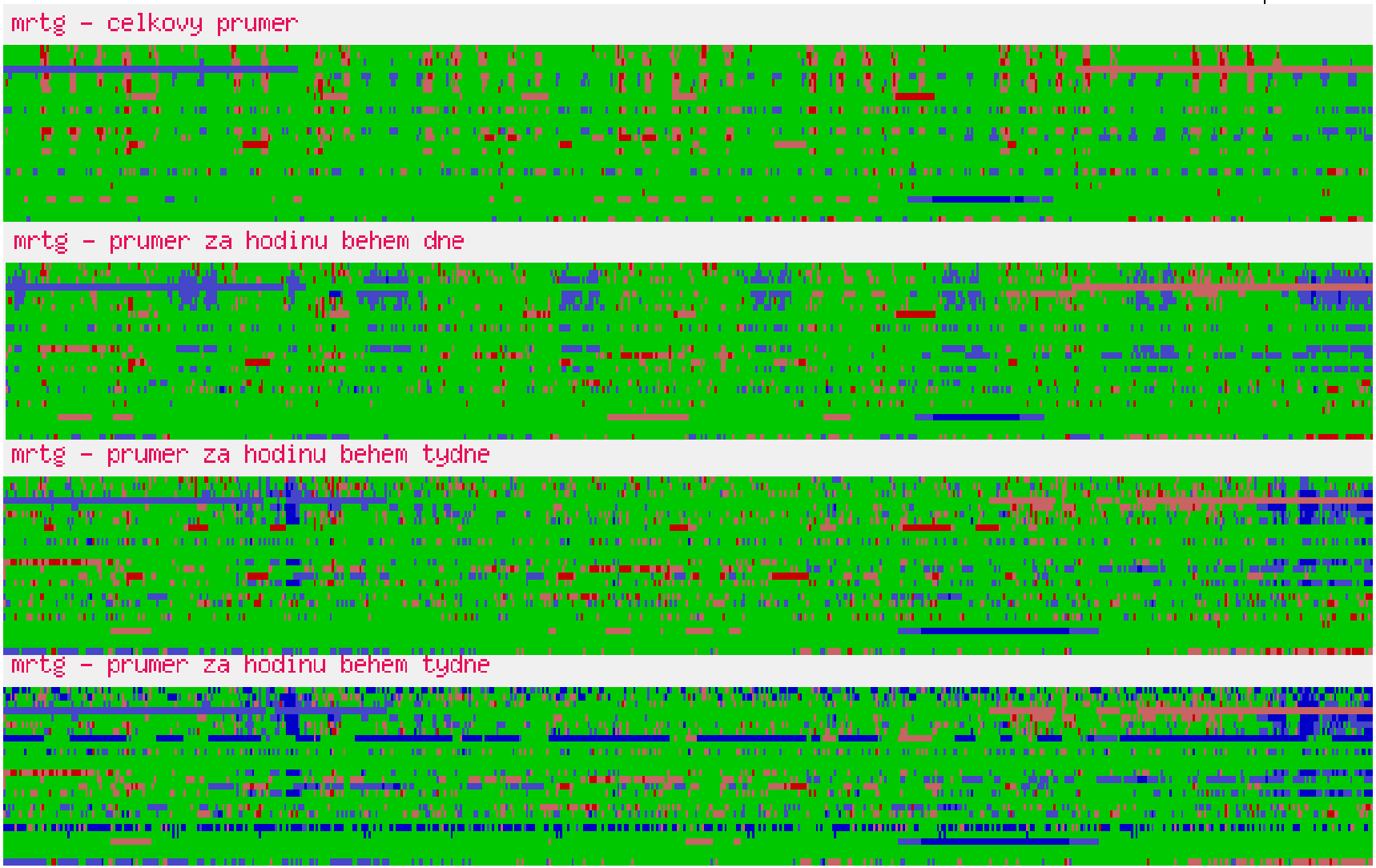


Příklad zobrazení: učebny

- Zdroj: Kerberos logy
- Parametr: počet přihlášení na jednom stroji

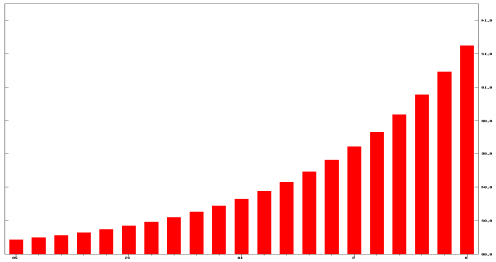


Příklad zobrazení: servery



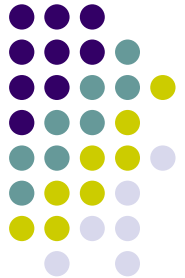
Perský koberec

Další vylepšení



- Klesající význam minulosti
  - Promazávání starších vzorků
  - Klouzavý průměr
- Grupování – pohledy z různých směrů na související veličiny
- Rozklik celek/detail, podrobné zobrazení průběhů, vazba na nagios, mrtg a zdrojové logy…
- Filtrace malých změn, derivace průběhů (pixelů v obrázku), zajímají nás velké a náhlé změny

Závěr



- Metoda je vhodná pro provozní systémy, s pravidelným periodickým chodem, kde odchylka parametrů není běžná
- Nehodí se pro testovací a experimentální stroje (vývoj, ladění, testy aplikací) a pro nárazově využívané služby bez pravidelného průběhu provozních veličin

Dotazy?