

ČESKÉ VYSOKÉ UČENÍ TECHNICKÉ V PRAZE  
FAKULTA JADERNÁ A FYZIKÁLNĚ INŽENÝRSKÁ

Katedra softwarového inženýrství v ekonomii  
Obor: Inženýrská informatika  
Zaměření: Softwarové inženýrství v ekonomii



# Modely zátěže výpočetních serverů

## Computational requirements modelling

BAKALÁŘSKÁ PRÁCE

Vypracoval: Dmitriy Burdin  
Vedoucí práce: Ing. Jan Doubek  
Rok: 2014

Před svázáním místo téhle stránky 



 s podpisem děkana a do pdf verze oskenované zadání.

### **Prohlášení**

Prohlašuji, že jsem svou bakalářskou práci vypracoval samostatně a použil jsem pouze podklady (literaturu, projekty, SW atd.) uvedené v příloženém seznamu.

V Praze dne .....

.....  
Dmitriy Burdin

## **Poděkování**

Děkuji Ing. Janu Doubkovi za vedení mé bakalářské práce a za podnětné návrhy, které ji obohatily.

Dmitriy Burdin

*Název práce:*

**Modely zátěže výpočetních serverů**

*Autor:* Dmitriy Burdin

*Obor:* Inženýrská informatika

*Druh práce:* BAKALÁŘSKÁ PRÁCE

*Vedoucí práce:* Ing. Jan Doubek  
CISCO Systems s.r.o.

*Abstrakt:* Cílem této práce je efektivně modelovat časovou, výpočetní a paměťovou náročnost distribuovaných úloh. Implementace modelů na základě časových řad výsledků minulých úloh. Výsledkem práce bude studie použitelnosti ekonometrických metod v prostředí velkých data center. Dále návrh algoritmů pro implementaci studovaných metod. Výstupní modely budou použity pro lepší plánování rozvržení výpočetních zdrojů.

*Klíčová slova:* Časová řada, ekonometrie, predikce, analýza, server

*Title:*

**Computational requirements modelling**

*Author:* Dmitriy Burdin

*Abstract:* The goal of this thesis is to effectively make a model of temporal, computational and memory requirements of distributed tasks. Model implementation based on time series results of past tasks. The result of the thesis will be a study of applicability of econometric methods in large scale data centres. In addition, design of algorithms for implementation of the studied methods. Output models will be used for better planning of computational sources arrangement.

*Key words:* Times series, econometrics, forecasting, analysis, server

# Obsah

<b>Úvod</b>	<b>7</b>
<b>1. Teoretická část</b>	<b>8</b>
1.1. Časové řady . . . . .	8
1.1.1. Úvod a charakteristika . . . . .	8
1.1.2. Analýza časových řad . . . . .	9
1.1.3. Predikce . . . . .	9
1.2. Regresní analýza . . . . .	9
1.2.1. Úvod . . . . .	9
1.2.2. Lineární regresní model . . . . .	9
1.2.3. Nelineární regresní model . . . . .	9
1.3. Softwarová aplikace . . . . .	9
1.3.1. Základní předpoklady . . . . .	9
1.3.2. Programovací jazyk a nástroje . . . . .	9
<b>2. Praktická část</b>	<b>10</b>
2.1. Úvod . . . . .	10
2.2. Implementace . . . . .	10
2.2.1. Datová vrstva . . . . .	10
2.2.2. Aplikační vrstva . . . . .	10
2.2.3. Prezentační vrstva . . . . .	10
2.3. Testování aplikace . . . . .	10
2.3.1. Testovací data . . . . .	10
2.3.2. Zpracování a analýza dat . . . . .	10
2.3.3. Predikce . . . . .	10
2.3.4. Vizualizace . . . . .	10
<b>Závěr</b>	<b>11</b>
<b>Literatura</b>	<b>12</b>
<b>Přílohy</b>	<b>13</b>
<b>A. Název přílohy</b>	<b>13</b>

# Úvod

Obrovské a rostoucí množství informací zaplavují dnešní podniky. To se stává hlavně kvůli rostoucímu počtu lidí, firem a zařízení připojených k internetu. Kolem třetiny světové populace má dnes přístup k internetu. Velké množství informace, jinak řečeno - velká data, představuje rychle rostoucí sféru, ve které se pro maximální efektivitu vyžaduje vhodná kombinace softwaru, hardwaru a navíc speciální úpravy s ohledem na oblast působnosti. Tyto fakty vedou k výraznému zaměření na velká data a na nové metody správy a analyzování tohoto proudu informací.

Důležitým prvkem řešení této problematiky je datacentrum. Jsou to specializované prostory pro umístění a zajištění stabilního provozu výpočetní a serverové techniky. Většinou se tato technika skládá z počítačových clusterů, což je několik spolupracujících počítačů propojených počítačovou sítí. Clustery se využívají pro výpočet komplikovaných početních úloh. Jeden z hlavních parametrů kvalitních datacenter je správné plánování distribuovaných úloh pro rychlé zpracování dat. Se stále zvětšujícím množstvím dat se tento problém komplikuje. Proto je potřeba vědět v jakém pořadí zpracovávat určité úlohy a jak dlouho to bude trvat.

Cílem této bakalářské práce bylo navrhnout a vytvořit softwarovou aplikaci pro analytické zpracování dat, se kterými pracují výpočetní servery. Podstatou aplikace je predikce budoucího chování určitých hodnont na základě analýzy minulých výsledků. Protože velká data obvykle obsahují dynamické systémy dat, které se mění s časem, zkoumaná data se předpokládají být ve formě časových řad. Proto se v této práci nejdříve věnuje způsobům zpracování, analýzy a predikce časových řad. Poté následuje popis algoritmu používání vybraných postupů pro vhodnou analýzu a implementace příslušného algoritmu. Na závěr je demonstrováno využití aplikace na příkladových datech.

# 1. Teoretická část

## 1.1. Časové řady

### 1.1.1. Úvod a charakteristika

Jak již bylo zmíněno, časové řady jsou základním zkoumaným prvkem při analýze různých dynamických systémů obsahující chronologicky uspořádaná data. Časové řady jsou vlastně soubory jednoznačně uspořádaných podle času pozorování v příslušném systému. Data ve formě časových řad vznikají v úplně různých odvětvích buď to fyzikální věda, biologie, společenská věda, medicína atd.

Například ve společenských vědách, časové řady jsou užitečné při popsání počtu obyvatel, porodnosti, nemocnosti. V ekonomii teorie časových řad je jednou z nejdůležitějších metod při analýze ekonomických procesů. Mohou popisovat vývoj určitého ukazatele, jako objem výroby, produktivita práce, nezaměstnanost nebo spotřeba surovin. V technice časové řady mohou představovat průběh signálu, spolehlivost nebo intenzitu zatížení elektrického zařízení.

Pro lepší porozumění určitého mechanismu nebo procesu, popsaného časovou řadou, existuje analýza časových řad skládající se z různých metod. Tyto metody pomáhají vytvořit vhodný model popisující chování pozorovaných hodnot. Znalost takového modelu umožňuje kontrolovat činnost a sledovat vývoj určitého systému. Jako důsledek správné provedené analýzy stavá se možným predikovat budoucí chování systému.

Časové řady se skládají ze dvou prvků:

- časový úsek, během kterého byly udělány pozorování
- hodnoty příslušných ukazatelů časové řady

Podle časového úseku časové řady se člení do dvou typů. Jedním je okamžiková časová řada, jejíž pozorování jsou naměřena v jisté časové okamžiky a sčítání hodnot příslušných pozorování nedává smysl. Příkladem může být řada udávající počet zaměstnanců ve firmě na začátku roku. Druhým typem je intervalová časová řada. Pozorování v intervalových časových řadách jsou závislé na délce časového intervalu sledování a v tomto případě už sčítání hodnot ukazatele časové řady dává smysl, neboť hodnotu ukazatele za větší interval lze získat sčítáním hodnot za jednotlivé části příslušného intervalu. Například porodnost ve státě za rok.

Časové úseky jsou obvykle stejnoměrně rozděleny, což znamená, že mezi jednotlivými pozorování stejné časové intervaly. V opačném případě pozorování jsou rozděleny různorodě a tím se analýza časových řad komplikuje.



Obvykle rozlišujeme dva základní modely časových řad:

- Determenistický - model, v kterém časová řada nemá náhodné prvky a je generována známou matematickou funkcí. Důsledkem je srovnatelně jednoduchá analýza časové řady.
- Stochastický - model popisuje náhodný proces a časová řada obsahuje náhodný prvek. Většina běžně se vyskytujících časových řad v praxi mají stochastické modely.

Jedna z dalších důležitých charakteristik časových řad, vyplývající ze stochastického modelu, je jejich stacionarita, případně nestacionarita. Při stacionaritě střední hodnota a rozptyl časové řady se v čase nemění, což nelze říct o nestacionárních časových řadách, ve kterých se objevují změny ve střední hodnotě či rozptylu. Říká se proto, že nestacionární časové řady mají určitý trend vývoje. I když stacionarita je běžným předpokladem většiny metod analýzy časových řad, některé modely se omezují na modely stacionárních procesů.

### **1.1.2. Analýza časových řad**

### **1.1.3. Predikce**

## **1.2. Regresní analýza**

### **1.2.1. Úvod**

### **1.2.2. Lineární regresní model**

### **1.2.3. Nelineární regresní model**

## **1.3. Softwarová aplikace**

### **1.3.1. Základní předpoklady**

### **1.3.2. Programovací jazyk a nástroje**

## 2. Praktická část

### 2.1. Úvod

### 2.2. Implementace

#### 2.2.1. Datová vrstva

#### 2.2.2. Aplikační vrstva

#### 2.2.3. Prezentační vrstva

### 2.3. Testování aplikace

#### 2.3.1. Testovací data

#### 2.3.2. Zpracování a analýza dat

#### 2.3.3. Predikce

#### 2.3.4. Vizualizace

## Závěr

# Literatura

- [1] Fumio Hayashi. *Econometrics*. Princeton. Princeton University Press. 2000.

## A. Název přílohy