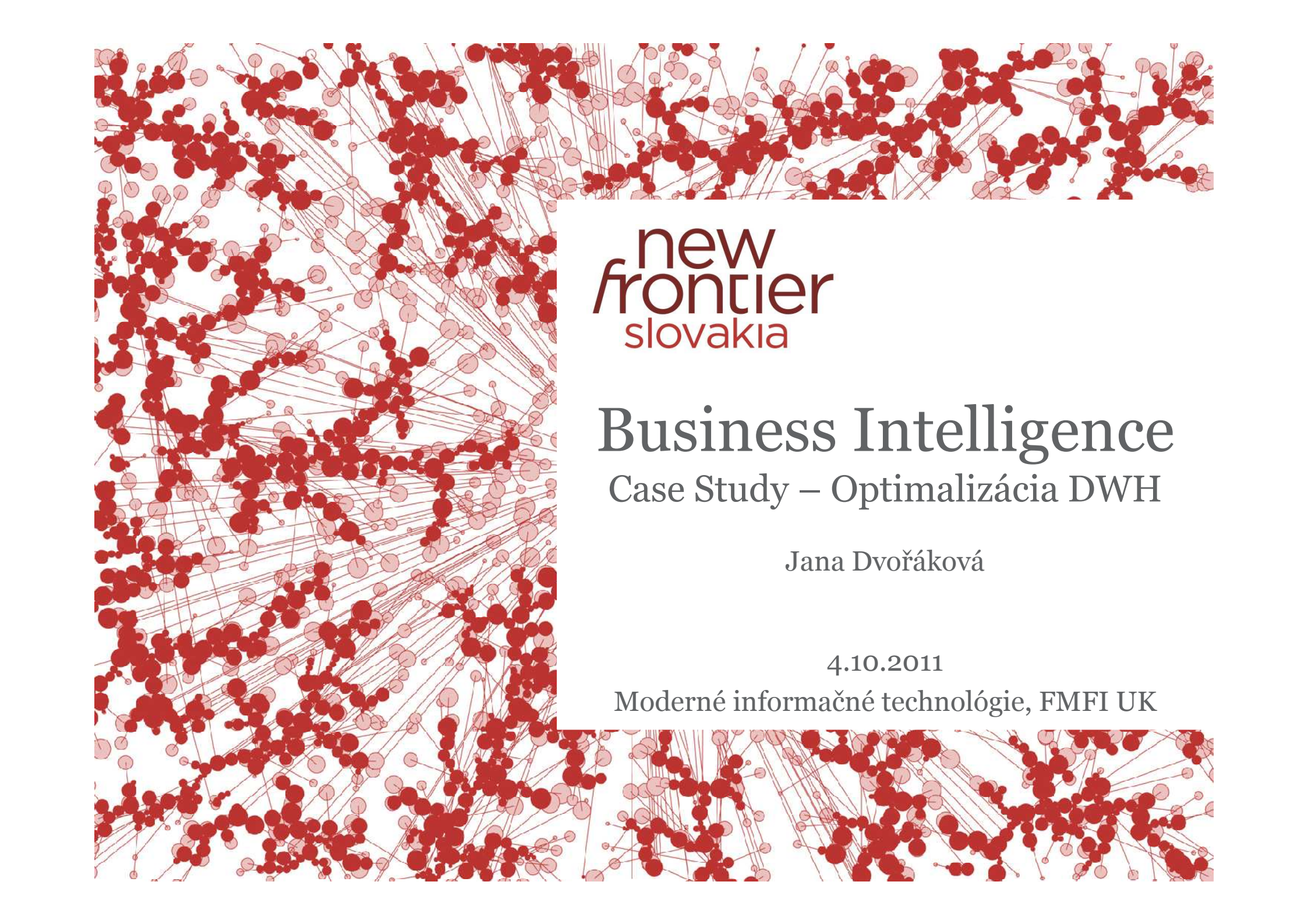


The background of the slide is a complex network of red circles of various sizes connected by thin red lines, creating a dense, web-like pattern. This pattern is visible on the left side and at the bottom of the slide, framing the central text area.

new
frontier
slovakia

Business Intelligence

Case Study – Optimalizácia DWH

Jana Dvořáková

4.10.2011

Moderné informačné technológie, FMFI UK

- o Veľká telekomunikačná spoločnosť
- o DWH
 - 9 TB
 - 24 zdrojových systémov
 - ~ 2500 ETL procesov
 - ~ 530 reportov
 - 500 000 000 riadkov denný prírastok do stage
 - stovky business používateľov

- o Vývojové – DEV
 - ETL vývojári, front-end vývojári
 - Implementácia, unit testovanie
- o Testovacie – TEST 1, TEST 2
 - Integrátori
 - Integrácia, integračné a regresné testovanie
- o Produkčné – PROD
 - Integrátori
 - Integrácia, ostrá prevádzka

Technológia

- o Informatica PowerCenter 8.6 AE
- o Oracle 10g2
- o Cognos v.7
- o HP-UX PA-RISC 11i v1
- o ERwin
- o SPSS Clementine



Priebežná optimalizácia: Zrýchlenie ETL procesov



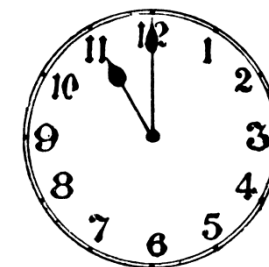
Požiadavky

o Problém: Neskoré dodávanie reportov na dennej báze

- Rozvoj DWH
- Riešenia vyvíjané so slabým dôrazom na čas spracovania
- Žiadne globálne optimalizácie (na úrovni workflowov)

o Jasné požiadavky –

- Zrýchliť denné spracovanie
- Zlepšiť dostupnosť DWH



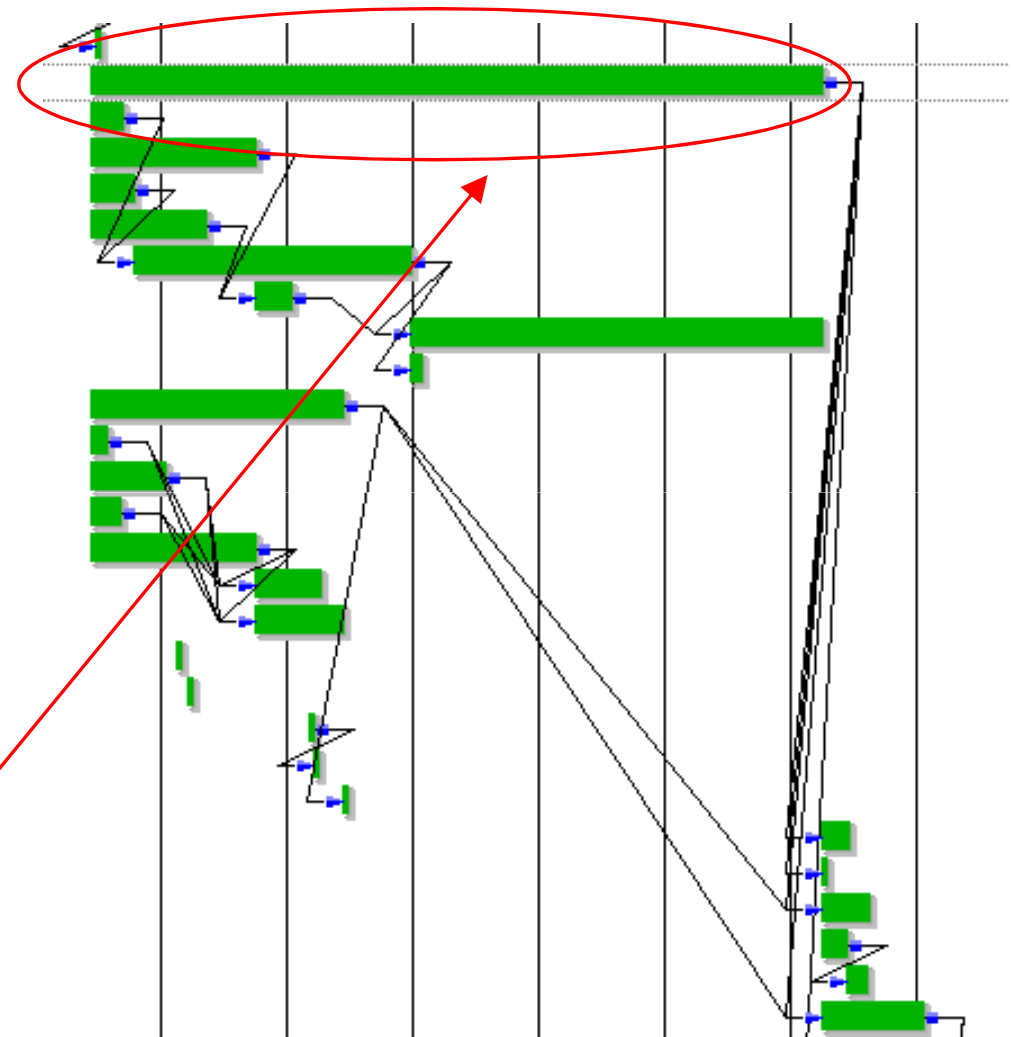
Riešené problémy

- o Identifikácia menených ETL procesov
- o Analýza ETL a hľadanie možností optimalizácie
- o Dôraz najmä na technickej časti riešenia



Identifikácia menených ETL

- o Vstup: Záznamy o ETL spracovaniach na produkčnom prostredí
- o Na úrovni session:
 - Metrika - Čas spracovania / množstvo prenesených dát
- o Na úrovni workflowu:
 - Procesy, na ktoré čakajú iné procesy



Analýza ETL a hľadanie možností optimalizácie

1. Zmena logiky ETL procesov

- Obmedzenie množstva prenesených dát
- Spojenie viacerých ETL do jedného procesu

2. Možnosti prostredia

- Rozdelenie záťaže medzi Informatica server a DB server

3. Zmena parametrov

- Úprava konfigurácie jednotlivých transformácií

Výstup:

- Návrh riešenia pre identifikované ETL
- Dokumentácia, ktorá pomáha pri vývoji nových riešení priamo zapracovať vytvorené optimalizačné postupy

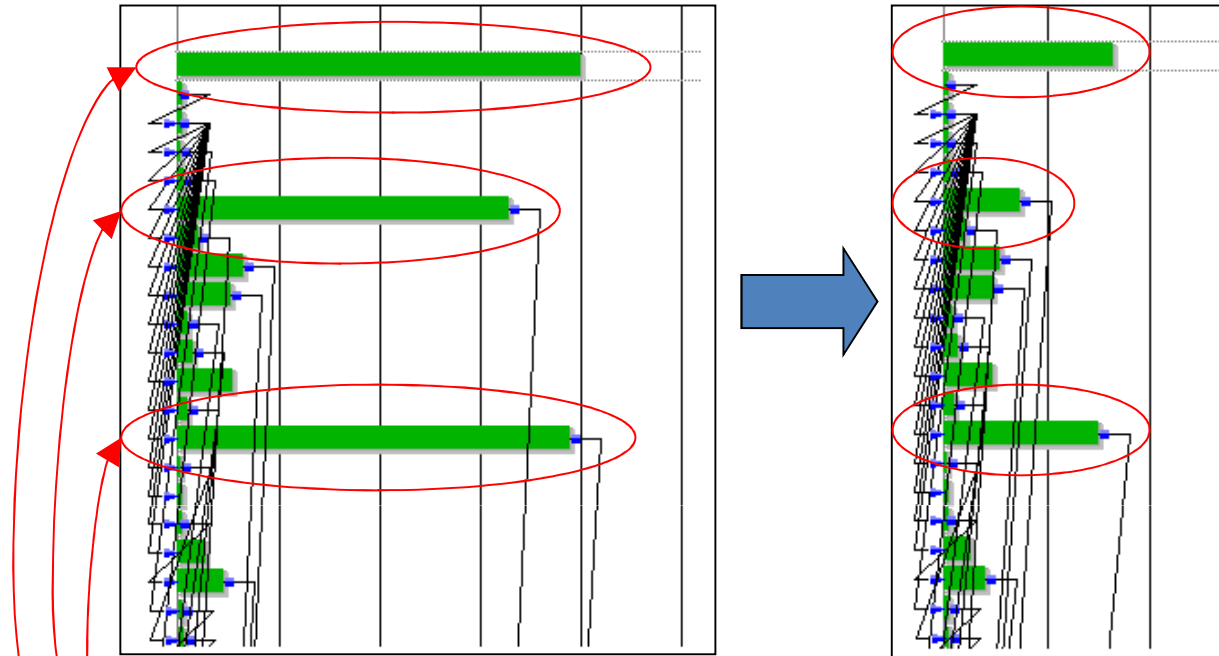
Unit testy

- o Vývojové prostredie (DEV) vyťažené, nie je možné testovať čas behu
- o Riešenie
 - DEV – korektnosť
 - TEST 3 – čas behu, je potrebné preniesť dáta z DEVu

Integračné testy

- o Štandardne na TEST 1, TEST 2 prostrediach

Výsledky



o Zrýchlenie spracovania:
20 – 60 %
pôvodného času

	Pôvodný čas	Nový čas	Zrýchlenie	Počet prenesených záznamov
ETL proces	03:53:20	01:30:22	39 %	~ 33 400 000
ETL proces	03:17:07	00:43:55	22 %	~ 14 800 000
Workflow	04:00:25	01:38:08	41 %	

Diskusia

new
frontier
slovakia

