

## Data Mining on an OLTP System (Nearly) for Free

Authors:

E. Riedel, C. Faloutsos, G. Ganger, D. Nagle

Group:

- Egor Kozmenko
- Mike Stasiowski
- Mulianto The

## Overview

- Introduction to the Problem
- Key Contribution
- Relevant Prior Work
- Methodology
- Results
- Our Opinion of the Paper

## Problem

- Operation support vs. decision support (OLTP vs. OLAP)
- Having special servers for OLAP is expensive
- Decision support systems are 10 times more expensive (e.g. \$12M vs. \$839K by TPC)

## Key Contributions

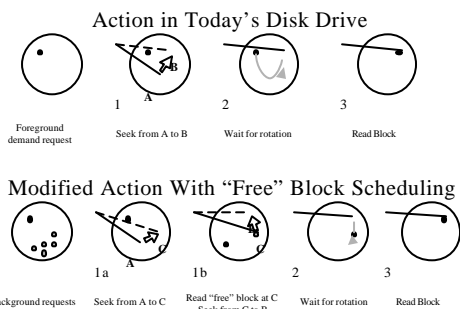
- New scheme that allows performing data mining task (e.g. association rules mining) as a background load of an OLTP database without decreasing the performance of the system
- Uses Active Disk technology

## Relevant Prior Work: Active Disk Technology

- Computation at the disk level
- Filter and aggregate data at the storage devices
- Offload server systems and network resources

## Methodology

- Background vs. foreground processing
- Foreground: normal OLTP activity
- Background (data mining):
  - Accessing idle disk (when foreground disk queue is empty)
  - Reading background blocks “for free” while waiting for disk rotation
- Crucial: ordering of background blocks is irrelevant



## Results

- The experiments show that the introduced scheme allows several MB/s data mining throughput without a significant impact on the OLTP response time.
- In other words, we can do data mining on an OLTP server, and the users (or customers) won't notice anything!

## Problem

- Background blocks residing of the “edge” of the disk - expensive to seek
- Possible solutions:
  - Keep data close to the “middle” of the disk
  - Upgrade “expensive” background requests to foreground

## Our Opinion

- We give this paper a rate of 8. That is:
  - 10 for those 11 million dollars we might save using this scheme
  - -2 for not specifying how actually to plug data mining algorithms into this scheme