# Enhancing Live Migration Process for CPU and/or memory intensive VMs running Enterprise applications

Benoit Hudzia
CEC Belfast / SAP Research
08/2011
With the contribution of Aidan Shribman and Petter Svard

# Agenda

- Background: Enterprise Applications and Live Migration
- Warm Up
- Delta Compression
- Page Priority
- Future Works

# Background

Migrating Enterprise Class applications

# Enterprise application and Live Migration

## Issues

- **Enterprise class application:**
  - Bigger than average resource requirement
  - Average SAP ERP 16GB + per VM with 32 GB of swap more than common
  - OLTP system such as ERP are very sensitive to time variation.
  - Rely heavily on precise scheduling capabilities, triggers, timers and on the ACID compliance of the underlying

- **Challenge when migrating such application:**
  - Disconnection of services:
    - Gigabit Ethernet timeout ≈ 5 seconds (>500 MB memory left in stop and copy phase )
    - Downtime is workload dependent
  - Disruption of services:
    - Migration progressively increasing the amount of resource dedicated to itself => gradually degrade performance of the coexisting systems / VMs.
  - Difficulty to maintain consistency and transparency
  - Unpredictability and rigidity

SAP RESEARCH

# Warm Up for Live Migration

Increasing the flexibility of Live Migration

SAP RESEARCH

# Warm Up

## Increasing flexibility

Extended adaptive Pre-copy phase without triggering actual migration

Increased flexibility :

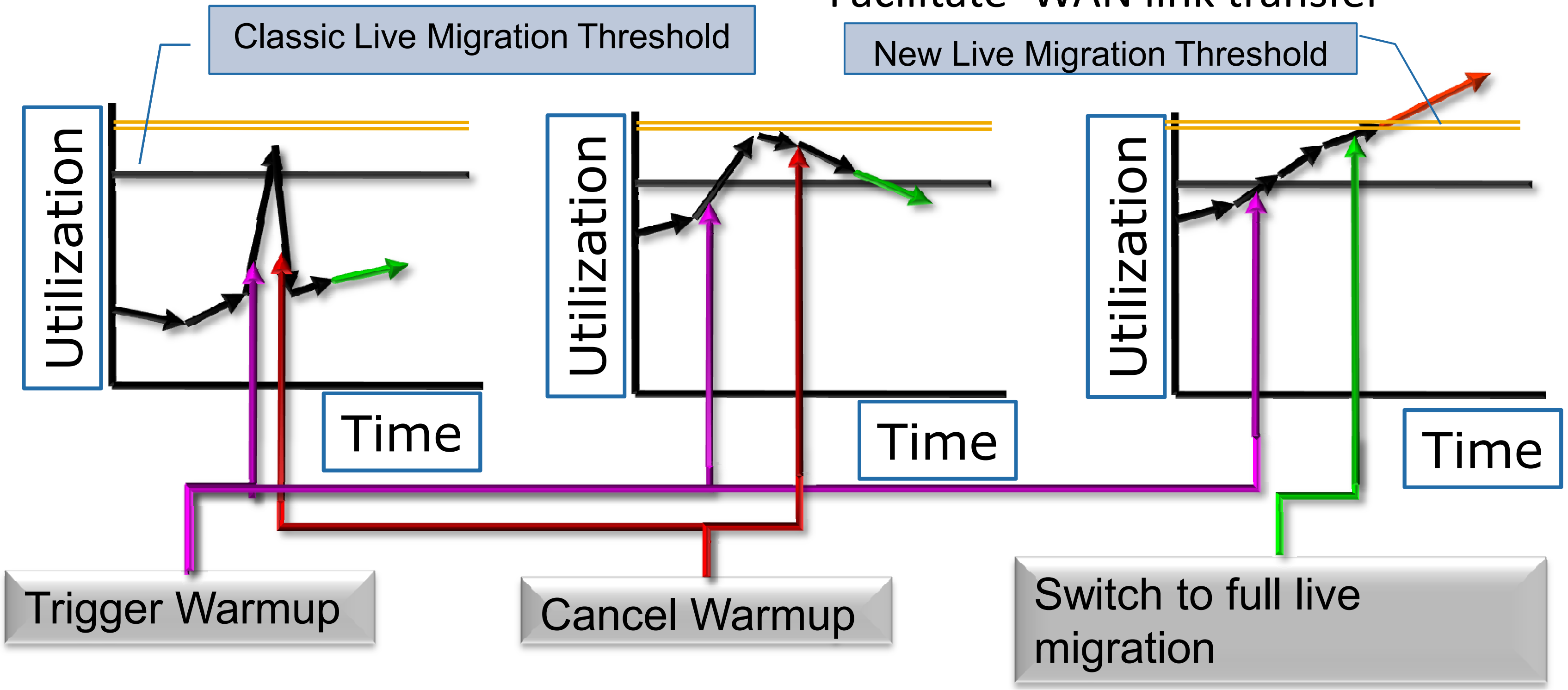
- “just in time“ triggering of live migration
- Reduce down time

Dynamic adaptive bandwidth allocation

- Manual and automatic

Allow “hot standby”

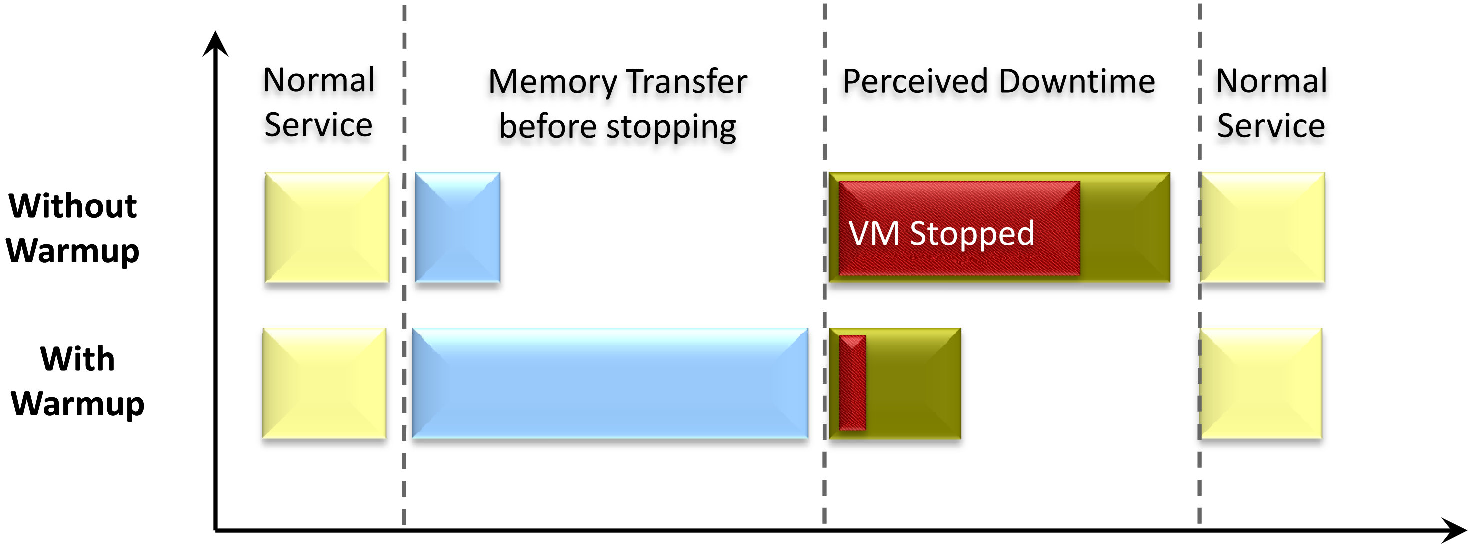
Facilitate WAN link transfer


Classic Live Migration Threshold
New Live Migration Threshold
Utilization
Time
Utilization
Time
Utilization
Time
Trigger Warmup
Cancel Warmup
Switch to full live migration

SAP RESEARCH

# Experimental Results: Warm-up Summary

## SAP Sales and Distribution Benchmark



VM size : 4GB

SMP : 2 vCPU

Users : 150

Load ~= 80%

| | CPU | Avg Response Time |
|---|---|---|
| Baseline | 60% | 2.18 sec |
| Warm-up | 73% | 2.16 sec |

Downtime under load: <1 sec
Success ratio : ~99%

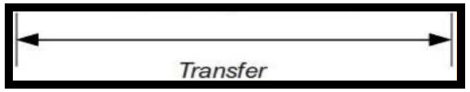
# Delta Compression of Page

Limiting the impact of resending Page

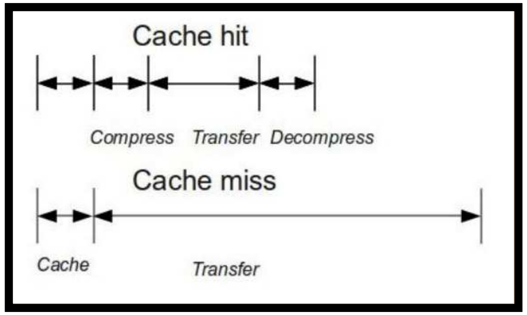
# Dirty Page Delta Compression

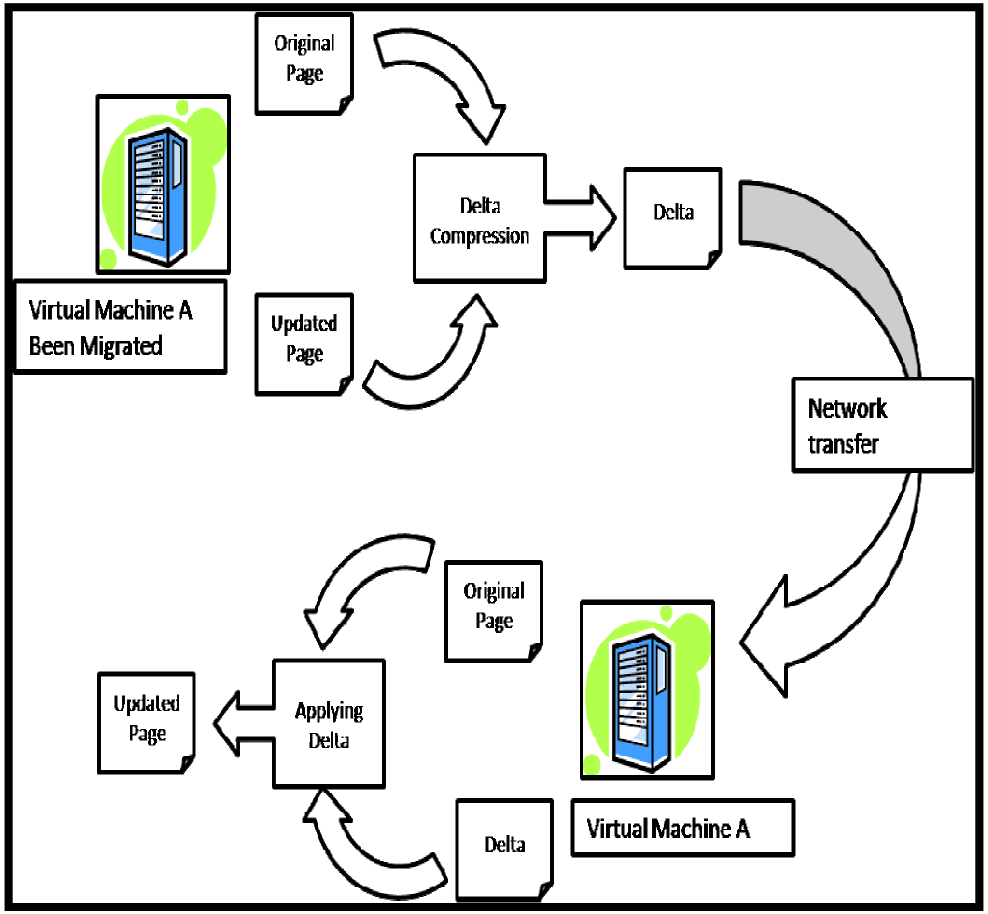
- Cache page with highest dirtying rate during send operation
- Compression Algorithm:
  - XBRLE : XOR +binary run length encoding

**Vanilla (no compr.)**



**Delta compression**



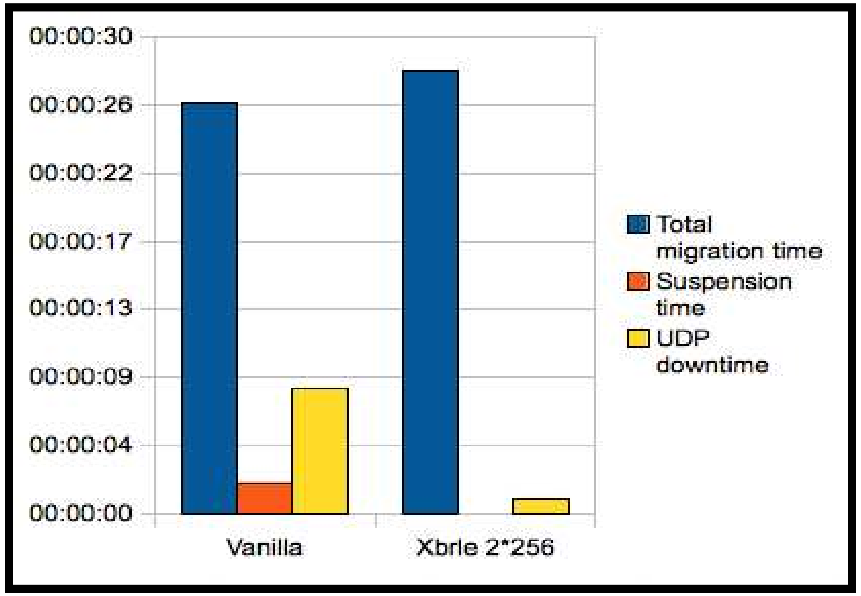


# Evaluation

## Benchmark

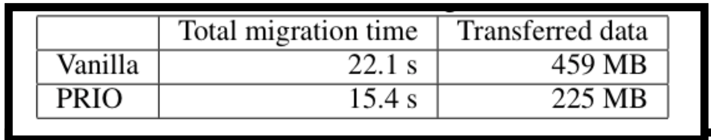
### Memory write benchmark (lm_bench)

- 1 GB RAM, 1 vcpu VM
- Near ideal case
- Downtime reduced by a factor of 100
- Throughput increased by 63 %

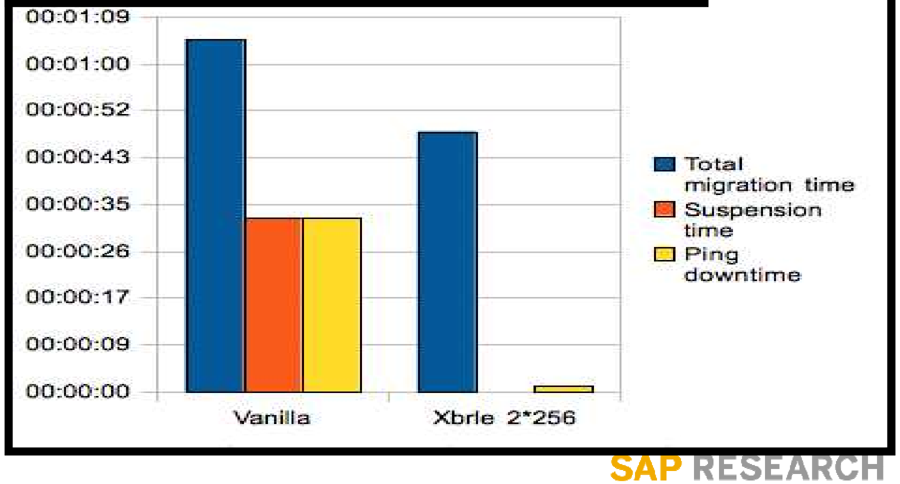


### Transcoded HD Video (VLC)

- 1 GB RAM, 1 vcpu VM
- Real-world, non-ideal case
- UDP downtime reduced from 8 s to 1
- Migration is transparent using XBRLE
- 31% faster, 51% less data sent



| | Total migration time | Transferred data |
|---|---|---|
| Vanilla | 22.1 s | 459 MB |
| PRIO | 15.4 s | 225 MB |



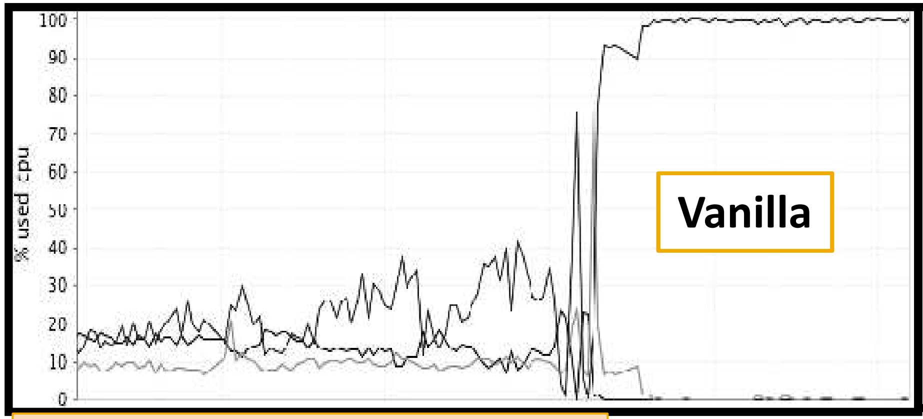
SAP RESEARCH

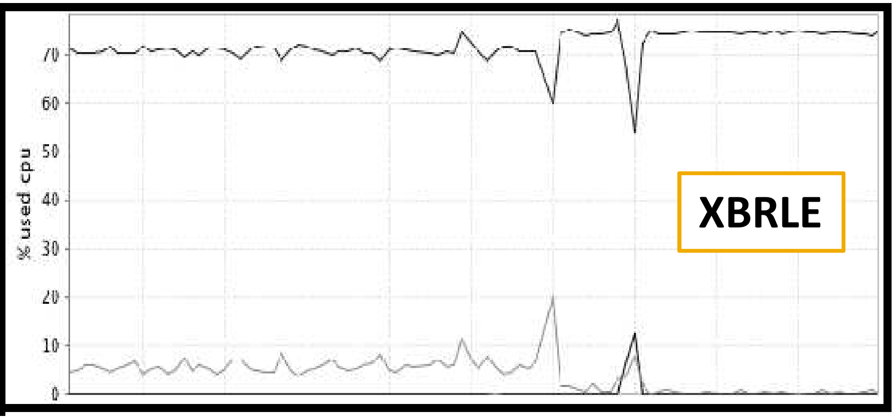
# Evaluation- SAP ERP

Sales and Distribution benchmark, load 100%

- Non-responsive on resume with vanilla algorithm
- Survived using XBRLE
- >0.5s of downtime = risk of damaging the system

- Measured downtime was 0.2s for XBRLE and 2s for vanilla
- Live Migration Cpu usage directly impact ( limit ) the available resource for the ERP





HW:4x 3,0GHz Xeon dual-core 32GB RAM
16TB Raid 5, 6Gbits/s trunked NFS server
1000Mbit/s Network

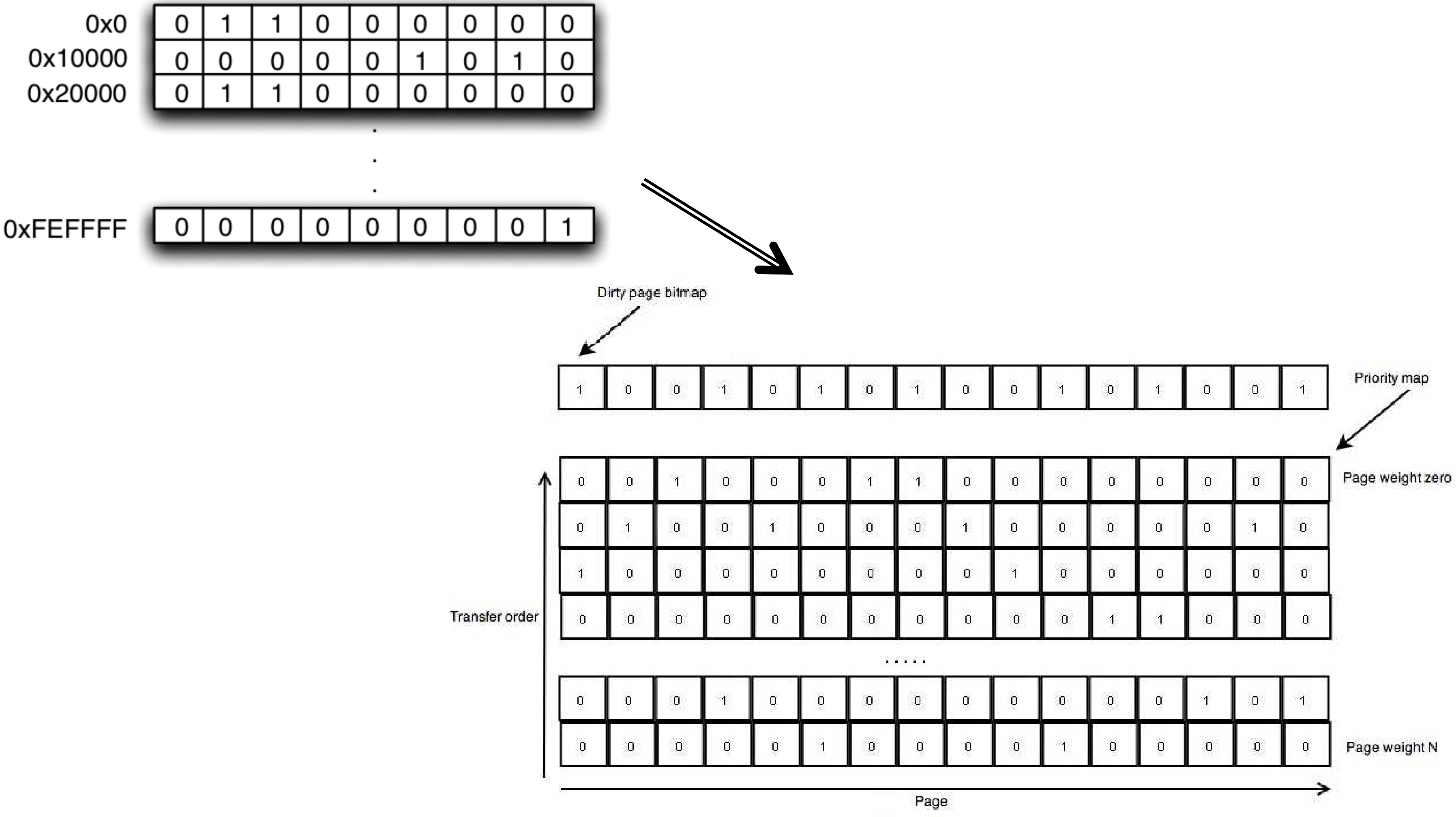
VM:8 GB RAM, 4 vcpus VM
App: SAP ERP 7.0 / S&D Benchmark

# Page Prioritization

Dynamic page transfer reordering

# Dynamic page transfer reordering

## Prioritizing page sends ( similar to writable working set concept in Xen)



| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0x0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0x10000 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 |
| 0x20000 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| . | | | | | | | | | |
| . | | | | | | | | | |
| . | | | | | | | | | |
| 0xFEFFFF | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

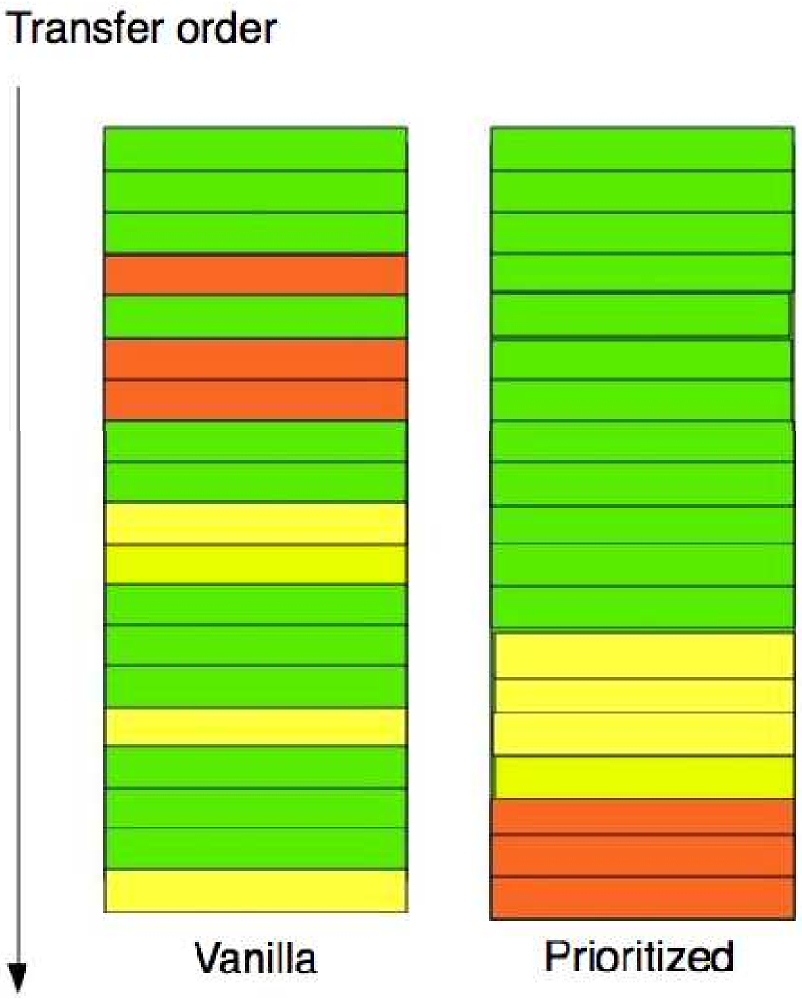
| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

SAP RESEARCH

# Dynamic page transfer reordering

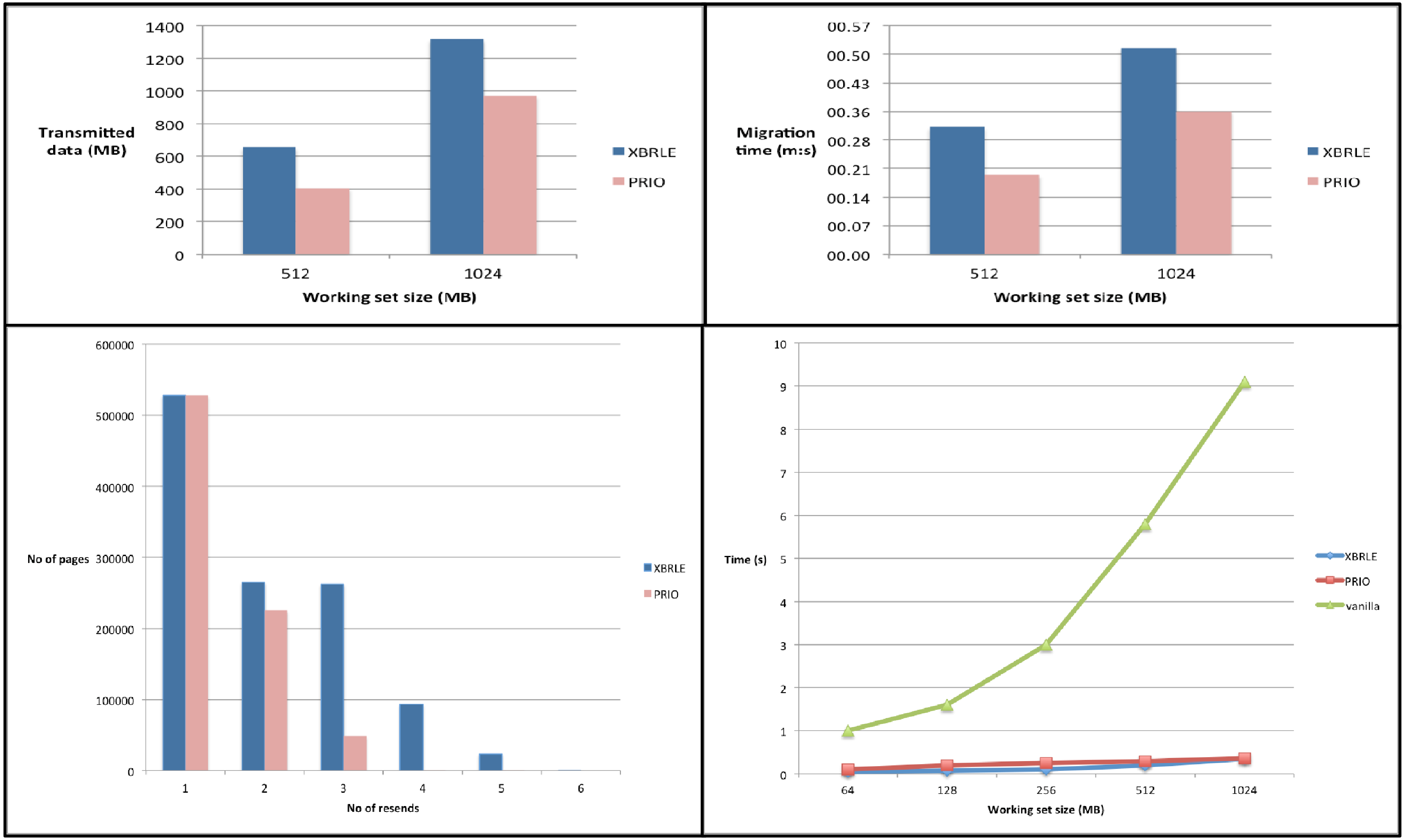
Prioritizing page sends



SAP RESEARCH

# Evaluation

Prio vs XBRLE : reveal Cache miss and compression efficiency Issue



SAP RESEARCH

# Optimizing Compression

Making XBRLE more efficient

SAP RESEARCH

# XBZRLE

## Increase compression speed /efficiency

- Only compress unmodified data using word aligned encoding and only encodes runs of zeros

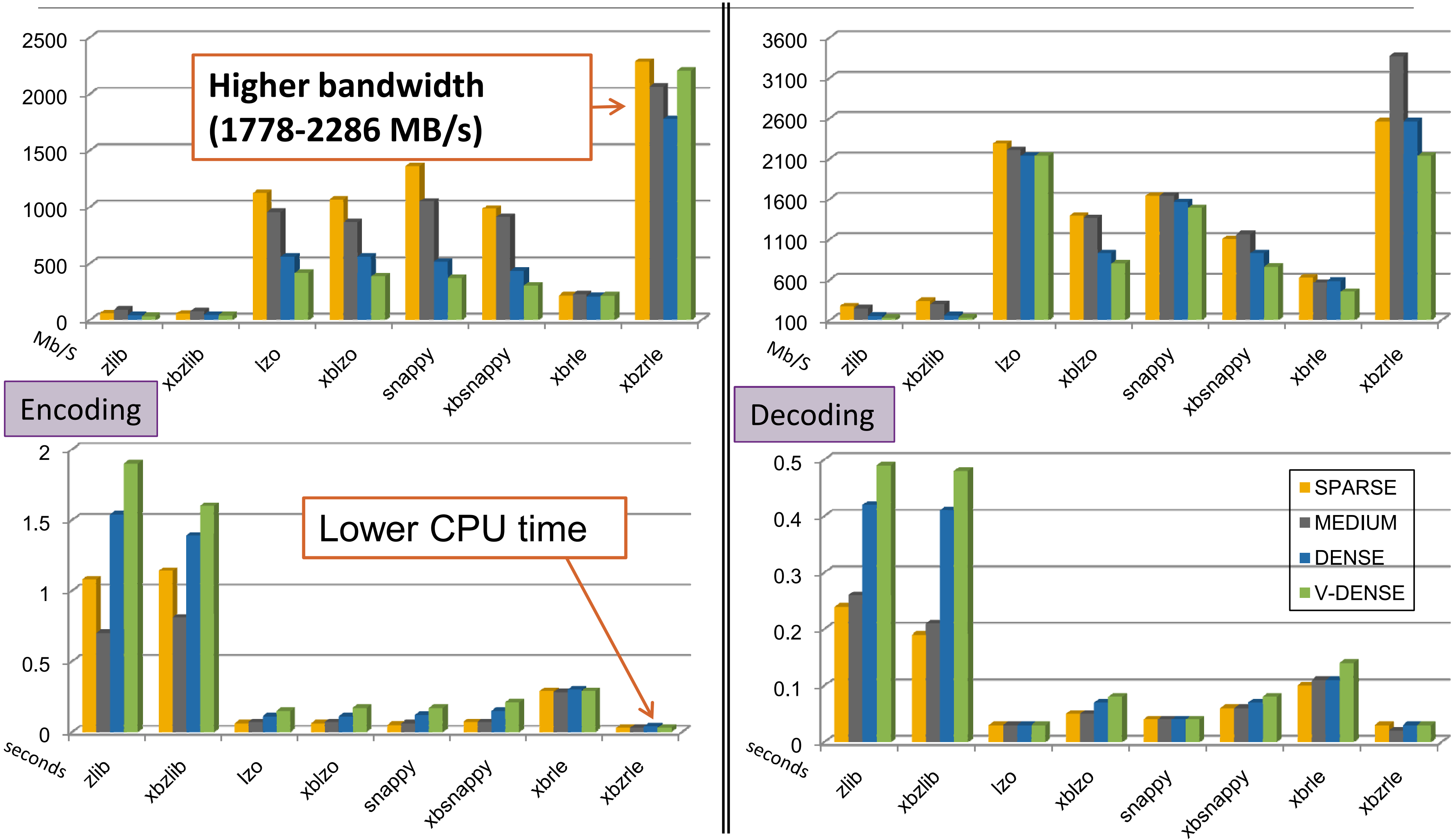
- For encoding page diffs XBZRLE is:
  - Compression :
    - 20% more efficient than XBRLE
    - 20% less efficient than LZO/Snappy.
  - Speed:
    - Overall 2.5x-5x faster than XOR + LZO/Snappy
    - 11x-9x faster  than the original XBRLE

- Doesn’t solve the impact of cache miss

SAP RESEARCH

# Performance comparison

## Synthetic benchmark representing enterprise workload



**Higher bandwidth (1778-2286 MB/s)**

Encoding

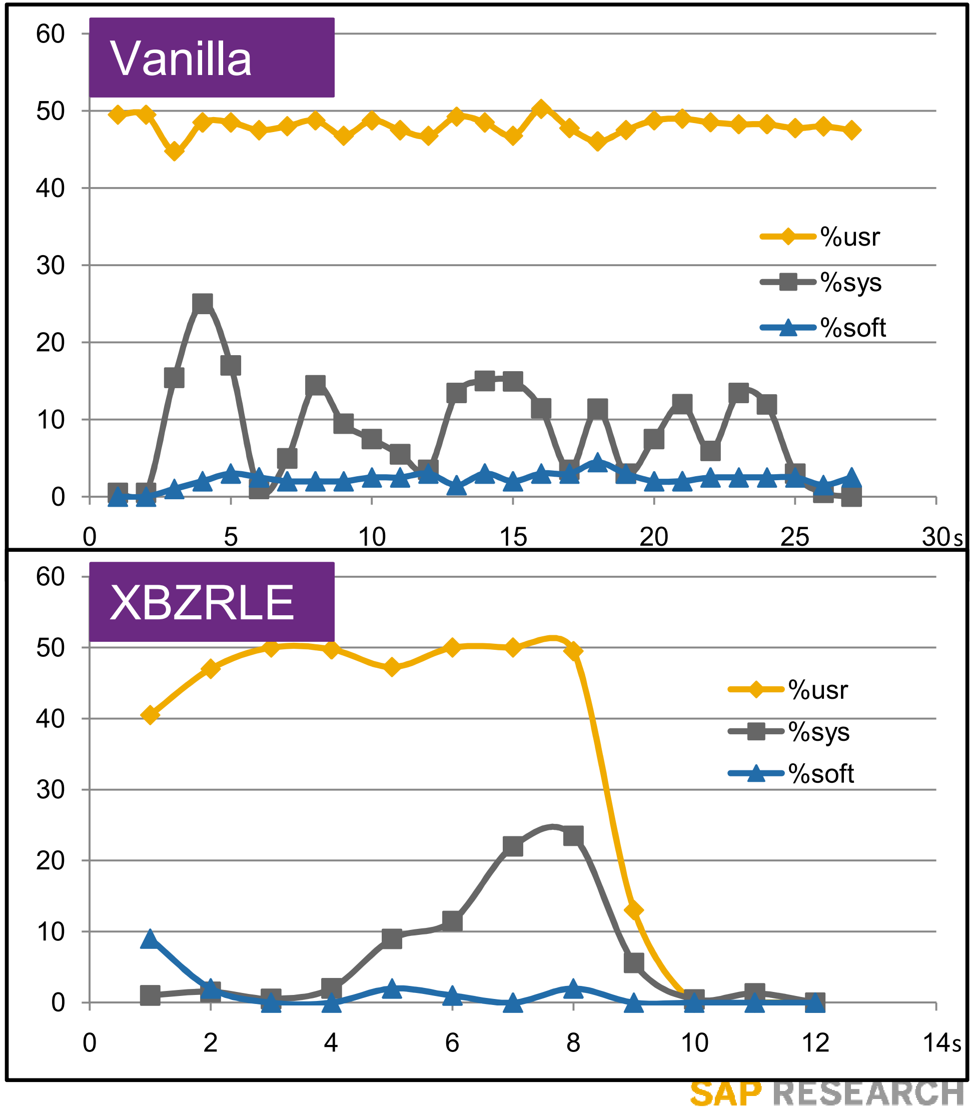
Decoding

Lower CPU time

SAP RESEARCH

# Performance comparison
# Live Migration Benchmark

- Compute capacity used for live migration :
  - **xbzrle** : 50%
  - **vanilla**:  between 30%-60%
- Live Migration:
  - **xbzrle** : terminate in seconds
  - **Vanilla** :not able to complete in the allocated time



SAP RESEARCH

# Future Work

SAP RESEARCH

# Future Works

- **Dynamically disable XBZRLE algorithm if the cache miss ratio is to important**
- **Combine Page priority algorithm and XBZRLE:**
  - Cache page with highest dirtying rate
  - Eliminate unnecessary cache check
  - Eliminate page compression with low potential return

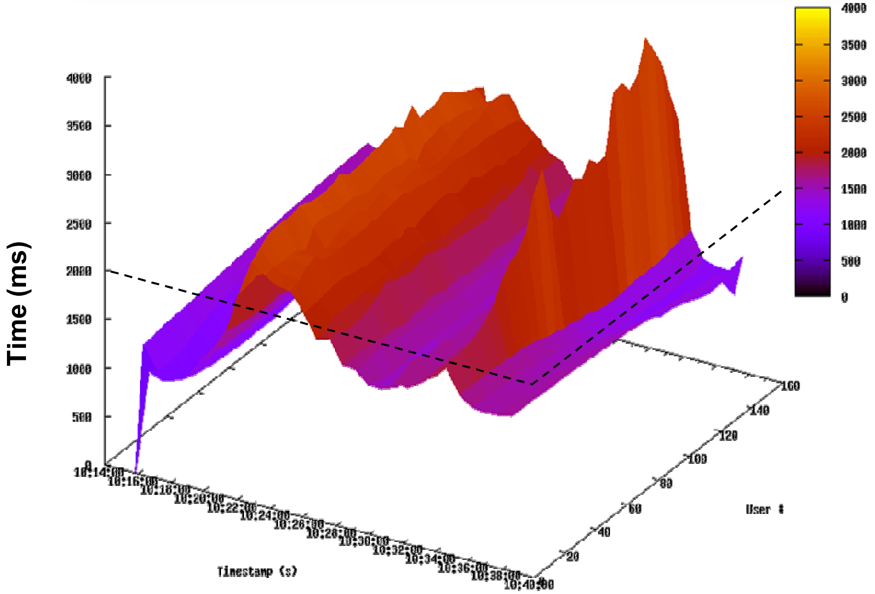
# Thank You!

Contact information:

Dr. Benoit Hudzia
Senior Researcher
benoit.hudzia@sap.com

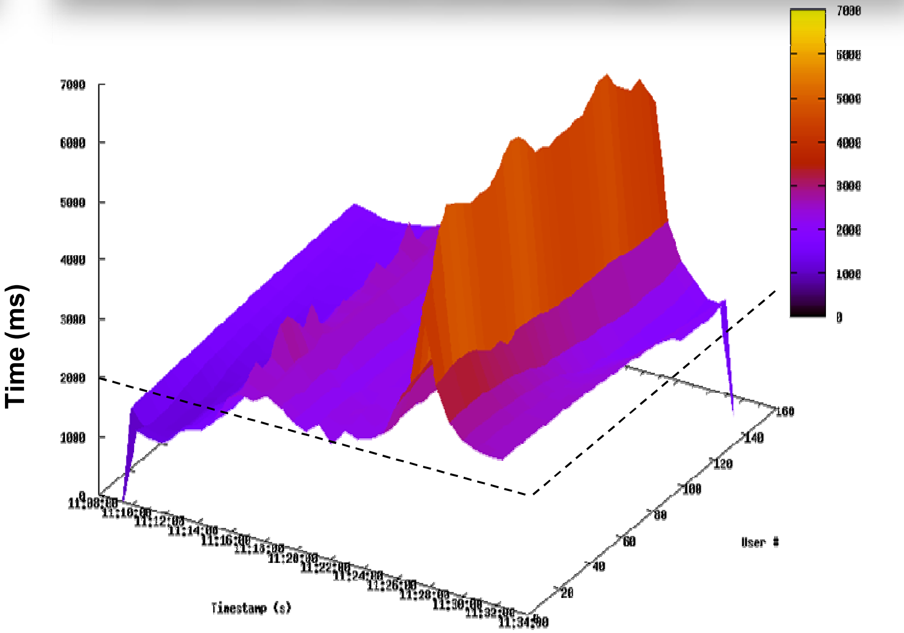
# Experimentations Results: S&D Benchmark with/out warm-up

## Response Time (baseline)



- - - - - - - - - - - 2 s response time threshold

## Response Time (warm-up)



- - - - - - - - - - - 2 s response time threshold

VM size : 4GB

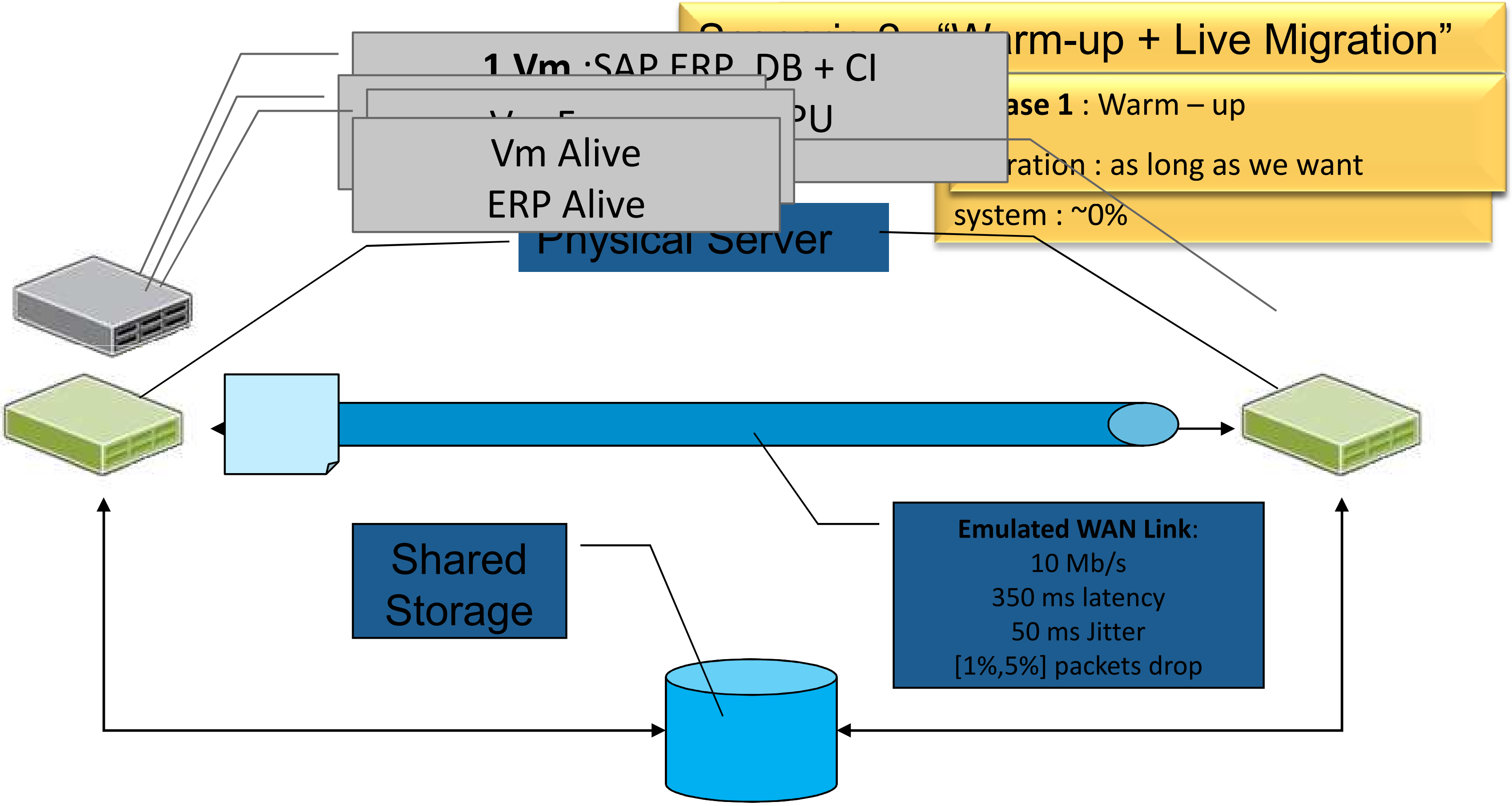
SMP : 2 vCPU

Users : 150

| | CPU | Avg Response Time |
|---|---|---|
| Baseline | 60% | 2.18 sec |
| Warm-up | 73% | 2.16 sec |

Downtime under load: <1 sec
Success ratio : ~99%

# Live Migration over emulated WAN Link



…rm-up + Live Migration”

1 Vm :SAP ERP, DB + CI

…PU

…ase 1 : Warm – up

…ration : as long as we want

system : ~0%

Vm Alive

ERP Alive

Physical Server

Shared Storage

**Emulated WAN Link:**
10 Mb/s
350 ms latency
50 ms Jitter
[1%,5%] packets drop

# © 2011 SAP AG. All rights reserved

This document is a preliminary version and not subject to your license agreement or any other agreement with SAP. This document contains only intended strategies, developments, and functionalities of the SAP® product and is not intended to be binding upon SAP to any particular course of business, product strategy, and/or development. Please note that this document is subject to change and may be changed by SAP at any time without notice.

SAP assumes no responsibility for errors or omissions in this document. SAP does not warrant the accuracy or completeness of the information, text, graphics, links, or other items contained within this material. This document is provided without a warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, or non-infringement.

SAP shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials. This limitation shall not apply in cases of intent or gross negligence.

The statutory liability for personal injury and defective products is not affected. SAP has no control over the information that you may access through the use of hot links contained in these materials and does not endorse your use of third-party Web pages nor provide any warranty whatsoever relating to third-party Web pages.