

# Persistence of the DSM-System Plurix



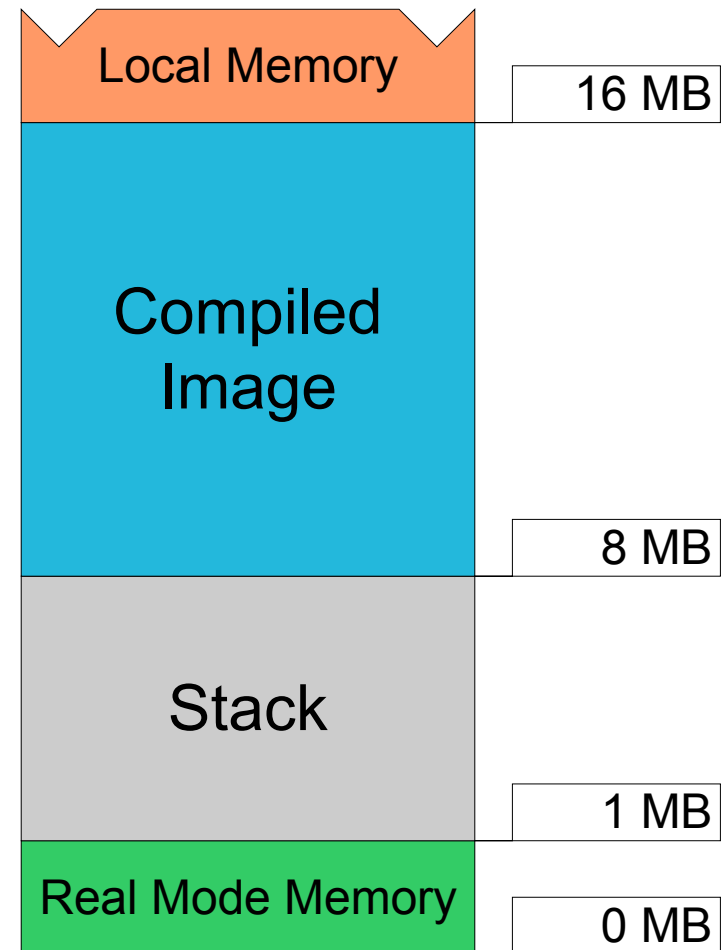
**Stefan Frenz**  
**Distributed Systems**  
**University of Ulm**

**Talk in Rennes**  
**2004/03/04**

- Memory Layout Overview
  - Memory after Booting
  - Installing / Joining
  - Running System
- Current Pageserver
  - Plain Integration
  - Linear Segment
- Future Prospects
  - Server-Extraction
  - Distribution

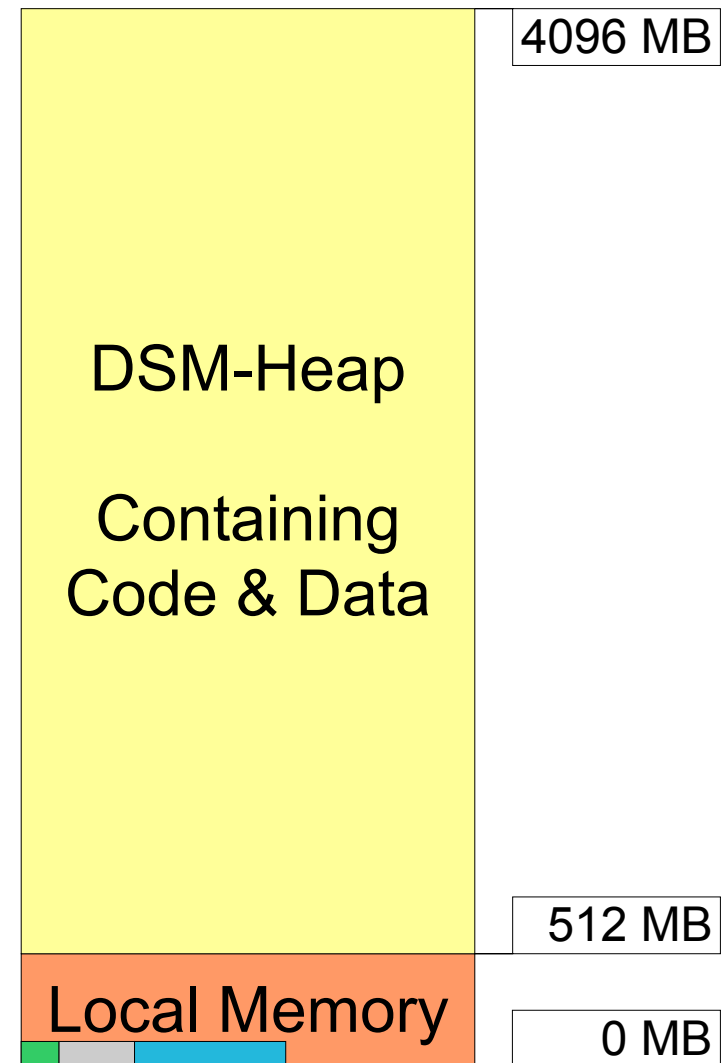
# Memory after Booting

- Immediate start after BIOS from disk, floppy, USB key drive
- Load image to compiled address
- Prepare processor
- Call java-method
- Prepare paging and local heap
- Initialize boot-devices
- Search cluster



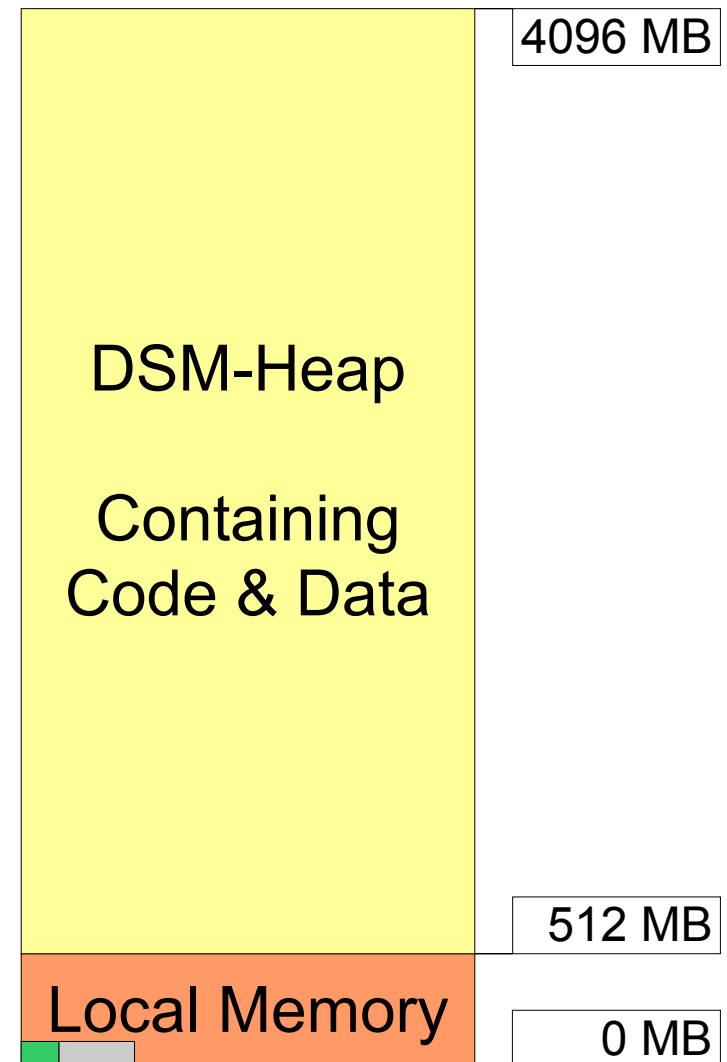
# Installing / Joining

- First node:
  - Installs DSM-Heap & cluster-management
  - Moves all objects into the Heap
  - Becomes owner of all pages
  - Frees up local memory
- Joining node:
  - Retrieves required kernel-pages
  - Joins as new cluster-node
  - Frees up local memory
- All objects reside in the DSM



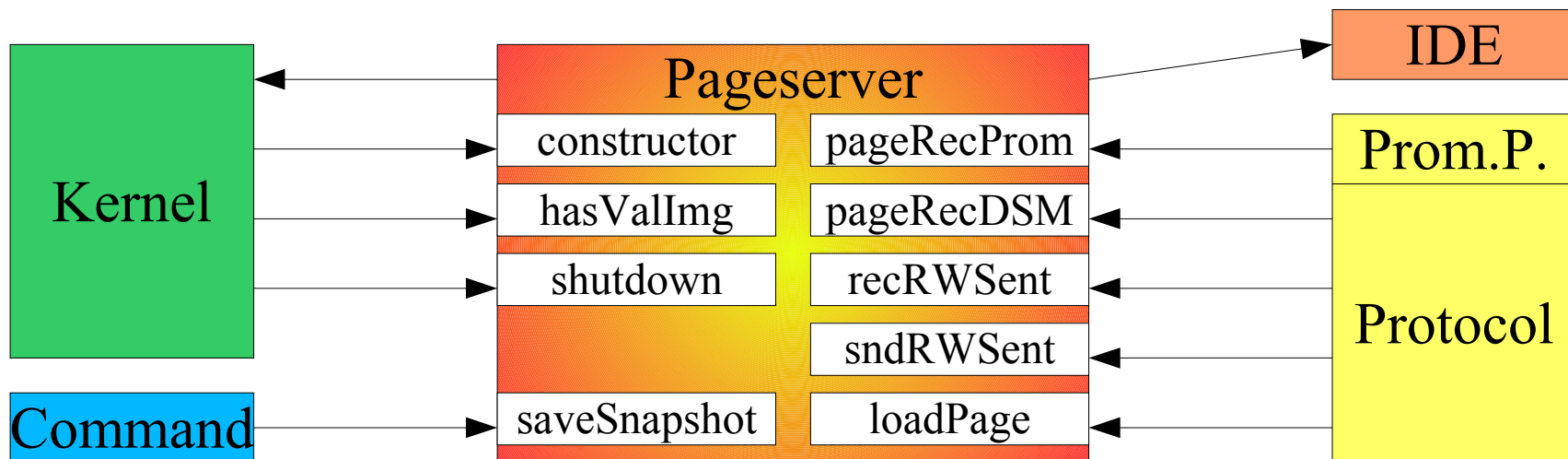
# Running System

- Code and data reside in DSM-heap
- Running code changing the DSM is encapsulated in transactions
- Local memory is used for:
  - Kernel-state & page-tables
  - Device-states & -data
- After commit of a transaction the protocol publishes addresses of changed pages
- Pageserver logs addresses of changed pages



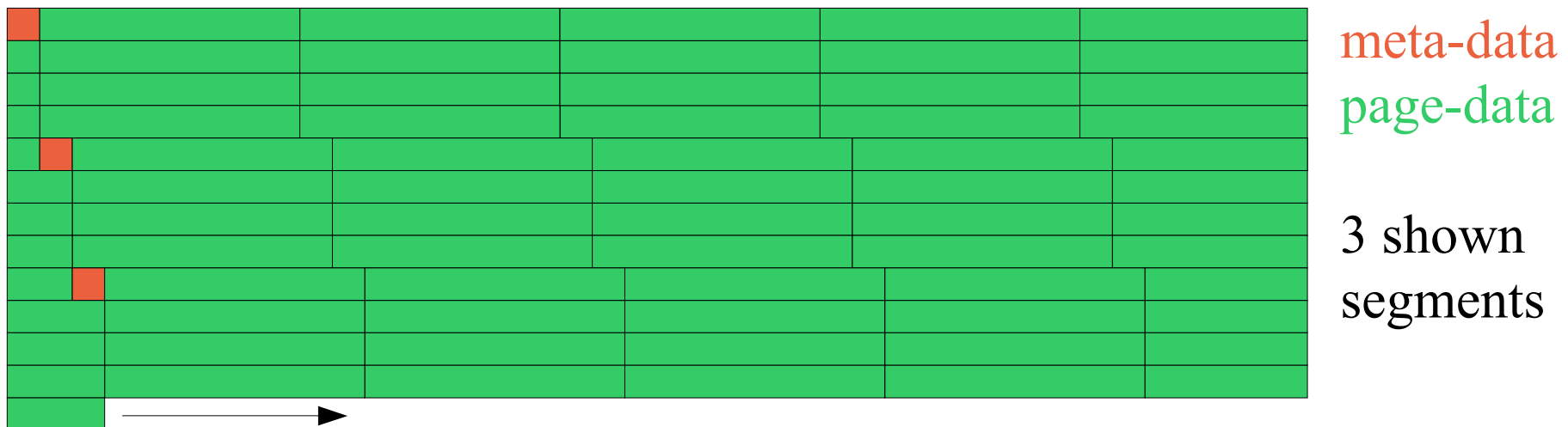
# Plain Integration

- Pageserver saves complete DSM-heap using Linear Segment
- Lean interface with 9 methods, called only from:
  - kernel during boot-up and initialization (3 methods)
  - standard DSM-protocol at run-time (4 methods)
  - promiscuous protocol at run-time (1 method)
  - user-command at run-time (1 method)
- Kernel and IDE-Driver required



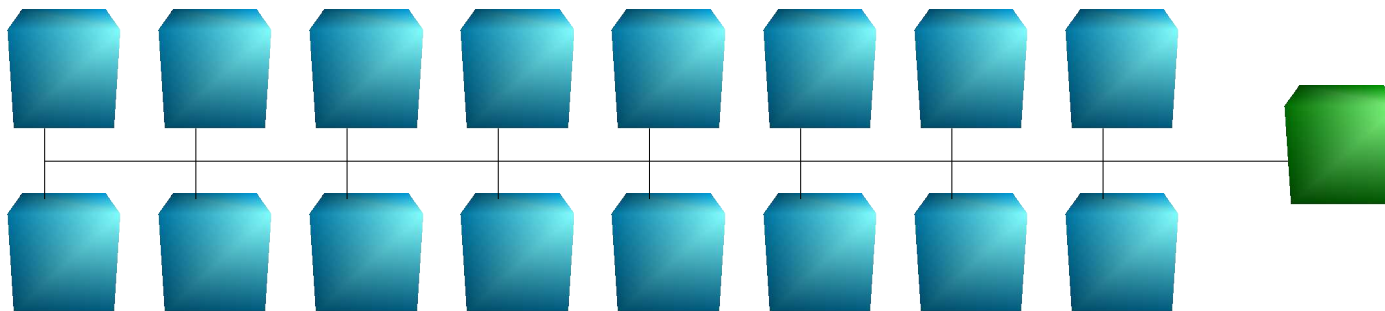
# Linear Segment

- Minimum overhead (meta-data : page-data = 1 : 160)
- Complete recovery possible without additional data
- Segments with at compile-time fixed size
- Fits best to hard disk requirements
- Normal throughput: ~45 MB/s
- Direct hardware-access



# Server-Extraction

- Always running server-node with integrated Pageserver
- Fast cluster-restart with cached data in Pageserver
- Separated from cluster and its DSM-image
- Crashing cluster will not crash Pageserver
- Running without user-transaction
- Straightforward behavior



# Distribution

- Higher performance: shorter intervals with increased nodes
- Redundancy: higher reliability

