



# A Service for Data-Intensive Computations on Virtual Clusters

Executing Preservation Strategies at Scale

Rainer Schmidt, Christian Sadilek, and Ross King

[rainer.schmidt@arcs.ac.at](mailto:rainer.schmidt@arcs.ac.at)

## Planets Project

- “Permanent Long-term Access through NETworked Services”
- Addresses the problem of digital preservation
  - driven by National Libraries and Archives
- Project instrument: FP6 Integrated Project
- 5. IST Call
- Consortium: 16 organisations from 7 countries
- Duration: 48 months, June 2006 – May 2010
- Budget: 14 Million Euro
- <http://www.planets-project.eu/>

## Outline

- Digital Preservation
- Need for eScience Repositories
- The Planets Preservation Environment
- Distributed Data and Metadata Integration
- Data-Intensive Computations
- Grid Service for Job Execution on AWS
- Experimental Results
- Conclusions

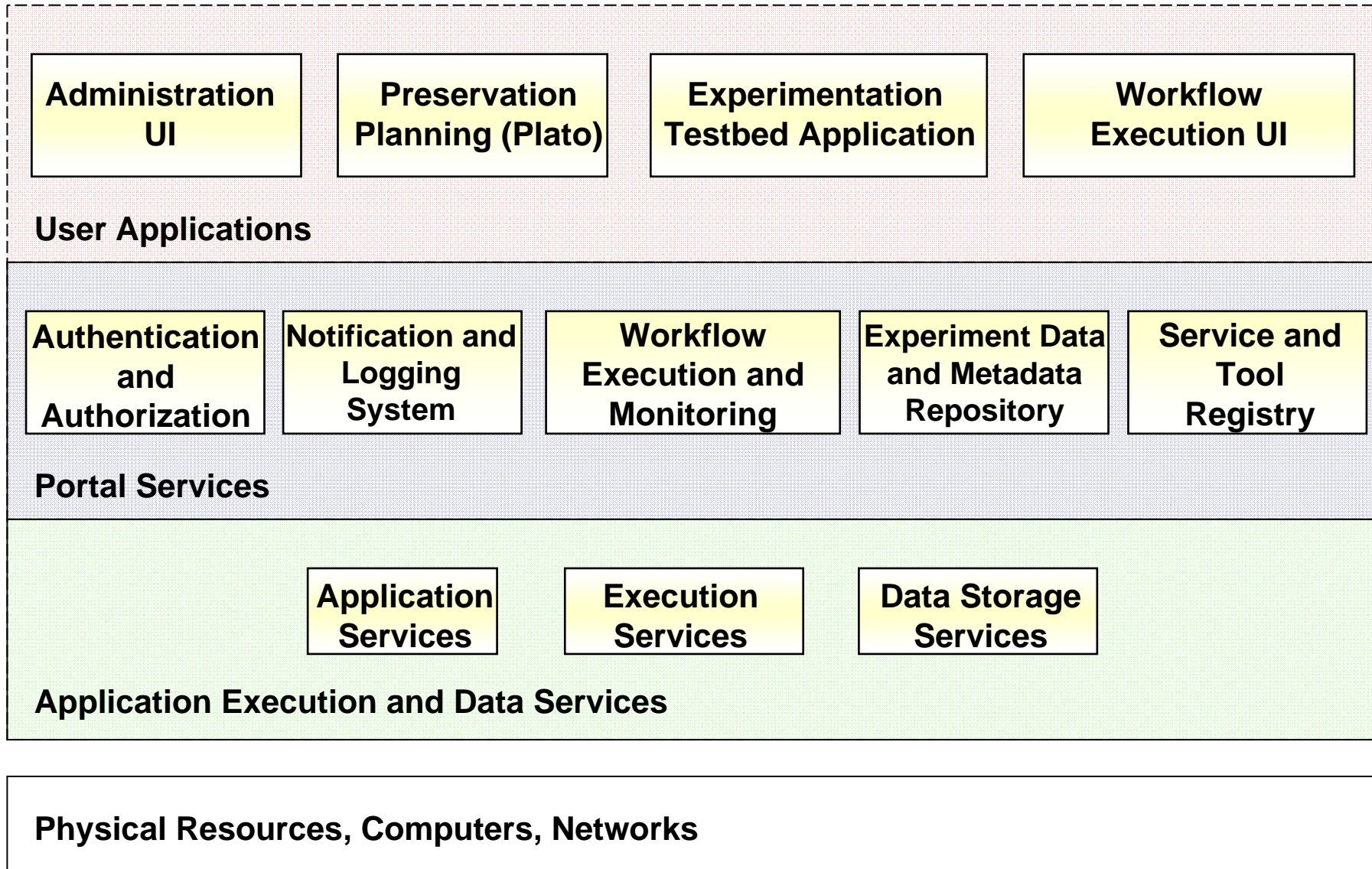
## Drivers for Digital Preservation

- Exponential growth of digital data across all sectors of society
- Large quantities of often irreplaceable information
  - Research and scientific data
- Data becomes significantly more complex and diverse
  - Digital information is fragile
- Ensure long term access and interpretability
  - Requires data curation and preservation
- Development and assessment of preservation strategies
  - Need for integrated and largely automated environments

## The Planets Environment

- A collaborative research infrastructure for the systematic development and evaluation and of preservation strategies.
- A decision-support system for the development of preservation plans
- An integrated environment for dynamically providing, discovering and accessing a wide range of preservation tools, services, and technical registries.
- A workflow environment for the data integration, process execution, and the maintenance of preservation information.

# Service Gateway Architecture



## Data and Metadata Integration

- Support distributed and remote storage facilities
  - data registry and metadata repository
- Integrate different data sources and encoding standards
  - Support range of protocols and description schemas
  - Connect distributed data repositories and archives.
- Development of common digital object model
  - Dynamic mapping of objects from different institutions (memory inst., research collections, content networks)
  - Recording of provenance information, preservation, integrity, and descriptive metadata

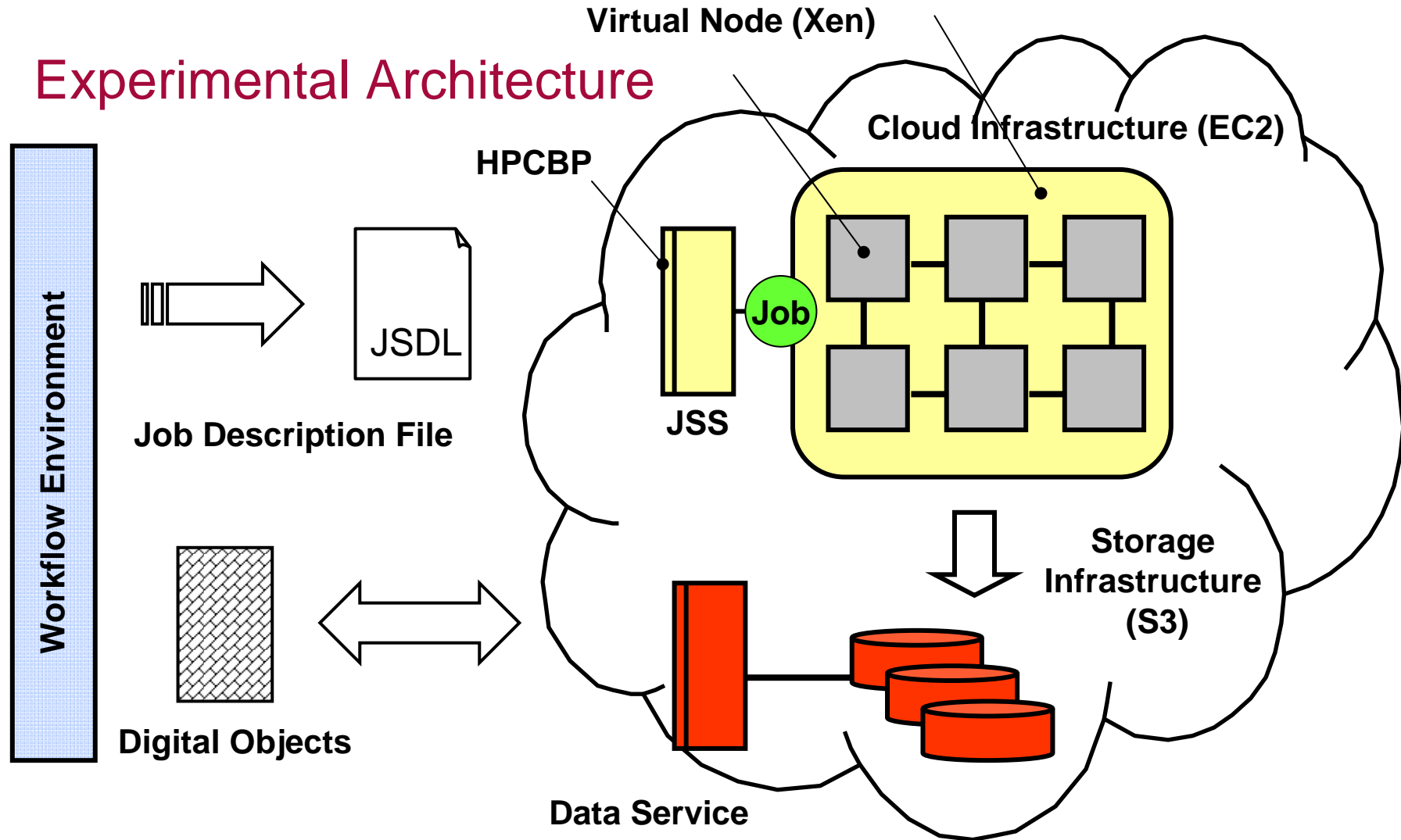
## Data Intensive Applications

- Development Planets Job Submission Services
  - Allow Job Submission to a PC cluster (e.g. Hadoop, Condor)
  - Standard grid protocols/interfaces (SOAP, HPC-BP, JSDL)
- Demand for massive compute and storage resources
  - Instantiate cluster on top of (leased) cloud resources based on AWS EC2 + S3, (or alternatively *in-house* in a PC lab.)
  - Allows computation to be moved to data or vice-versa
- On-demand cluster based on virtual machine images
  - Many Preservation Tools are/rely on 3rd party applications
  - Software can be preinstalled on virtual cluster nodes

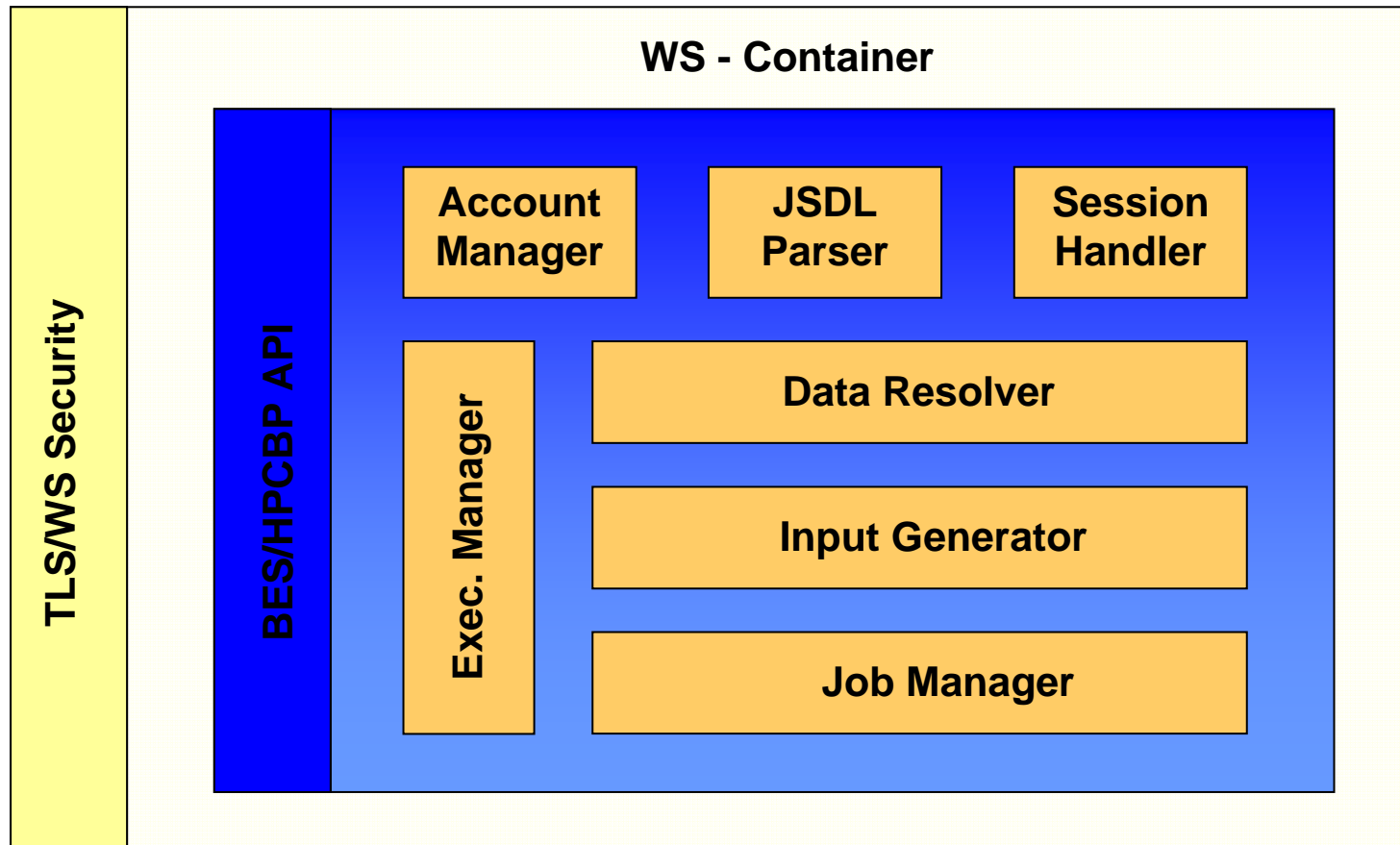
# Virtual Cluster and File System (Apache Hadoop)



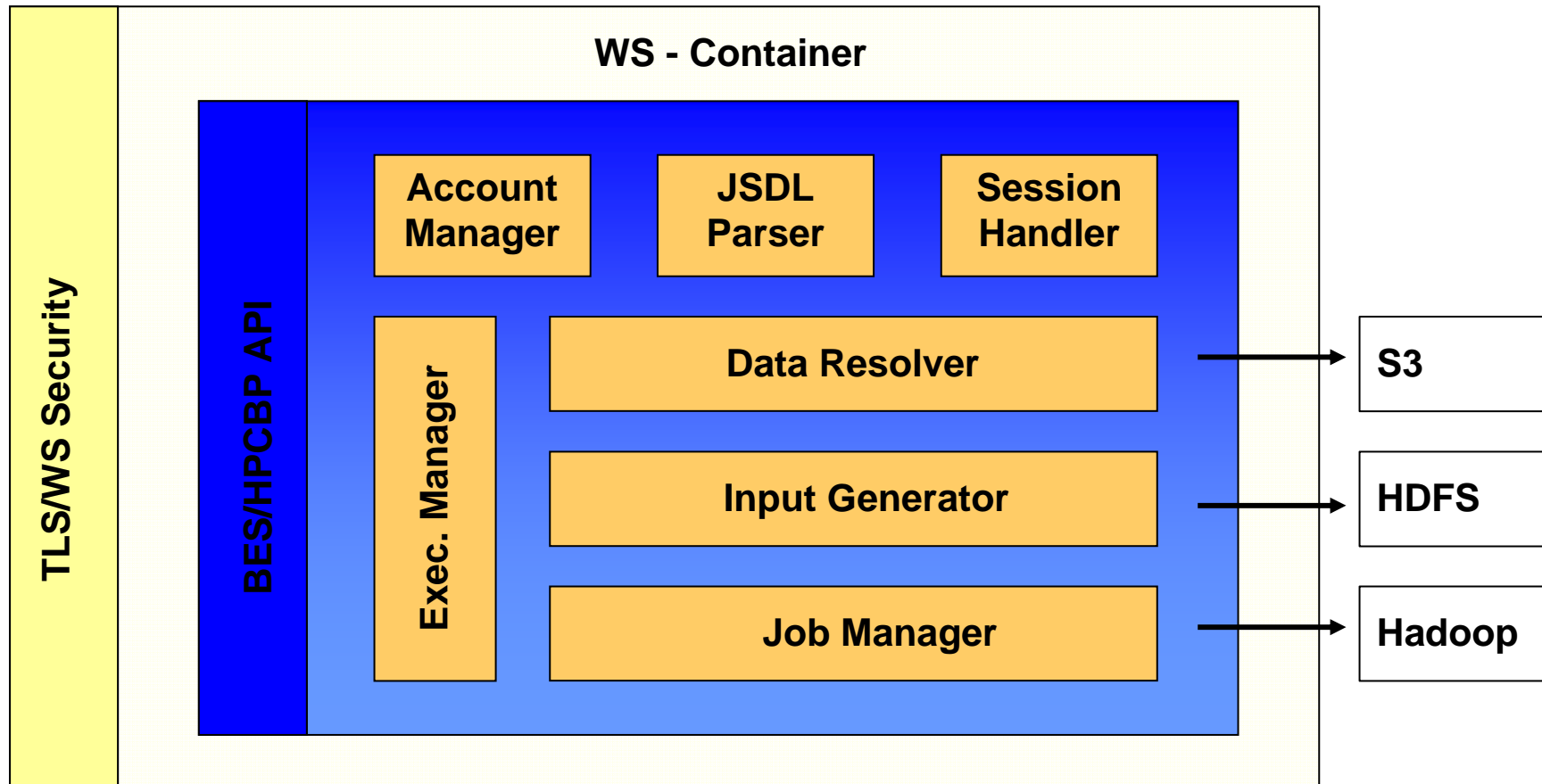
## Experimental Architecture



# A Light-Weight Job Execution Service



# A Light-Weight Job Execution Service



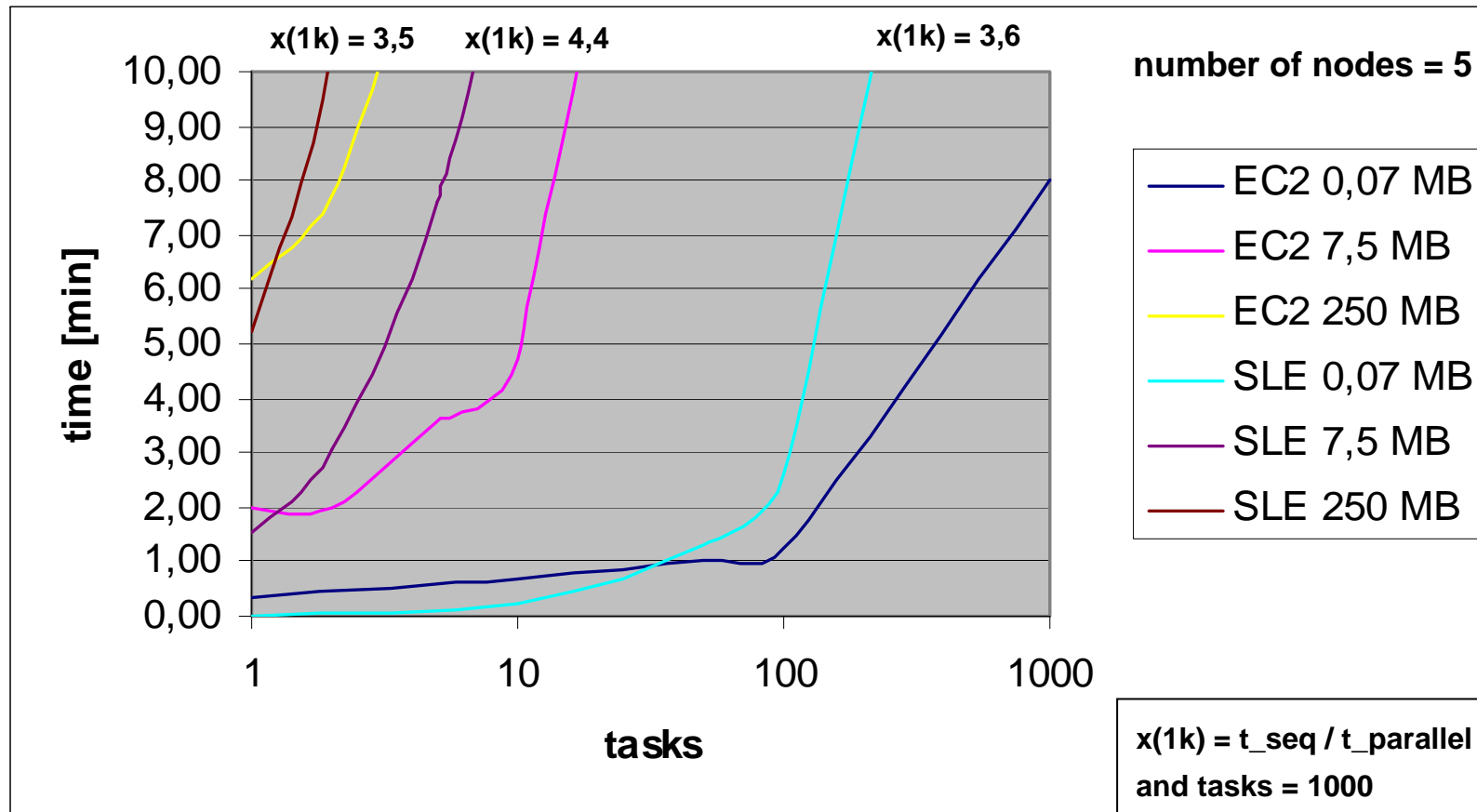
## The MapReduce Application

- Map-Reduce implements a framework and prog. model for processing large documents (Sorting, Searching, Indexing) on multiple nodes.
  - Automated decomposition (split)
  - Mapping to intermediary pairs (map), optionally (combine)
  - Merge output (reduce)
- Provides implementation for data parallel + i/o intensive applications
- Migrating a digital object (e.g web page, folder, archive)
  - Decompose into atomic pieces (e.g. file, image, movie)
  - On each node, process input splits
  - Merge pieces and create new data collections

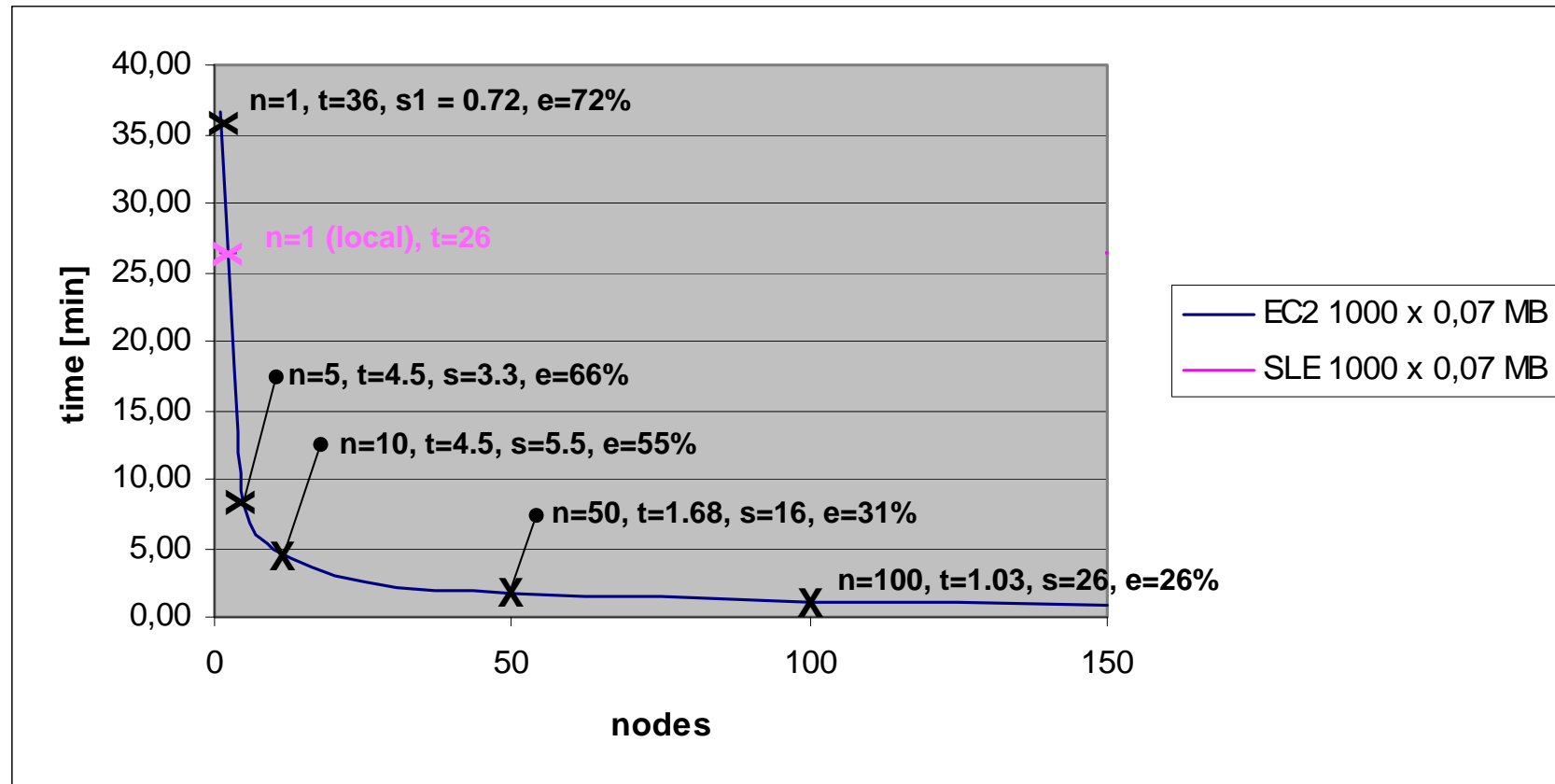
## Experimental Setup

- Amazon *Elastic Compute Cloud (EC2)*
    - 1 – 150 cluster nodes
    - Custom image based on Fedora 8 i386
  - Amazon *Simple Storage Service (S3)*
    - max. 1TB I/O per experiment
    - All measurements based on a single core per virtual machine
  - Apache Hadoop (v.0.18)
    - MapReduce Implementation
  - Preinstalled command line tools
  - Quantitative evaluation of VMs (AWS) based on execution time, number of tasks, number of nodes, physical size
-

# Experimental Results 1 – Scaling Job Size



## Experimental Results 2 – Scaling #nodes



## Results and Observations

- AWS + Hadoop + JSS
  - Robust and fault tolerant, scalable up to large numbers of nodes
  - ~32,5MB/s download / ~13,8MB/s upload (cloud internally)
- Also small clusters of virtual machines reasonable
  - $S = 4,4$  for  $p = 5$ , with  $n=1000$  and  $s=7,5\text{MB}$
  - Ideally, size of cluster grows/shrinks on demand
- Overheads:
  - SLE vs. Cloud ( $p=1$ ,  $n=1000$ ) 30% due to file system master
  - In average, less then 10% due to S3
  - Small overheads due to coordination, latencies, pre-processing

## Conclusion

- Challenges in digital preservation are diversity and scale
  - Focus: scientific data, arts and humanities
- Repositories Preservation systems need to be embedded with large-scale distributed environments
  - grids, clouds, eScience environments
- Cloud Computing provides a powerful solution for getting on-demand access to an IT infrastructure.
  - Currently integration issues: Policies, Legal Aspects, SLAs
- We have presented current developments in the context of the EU project Planets on building an integrated research environment for the development and evaluation of preservation strategies