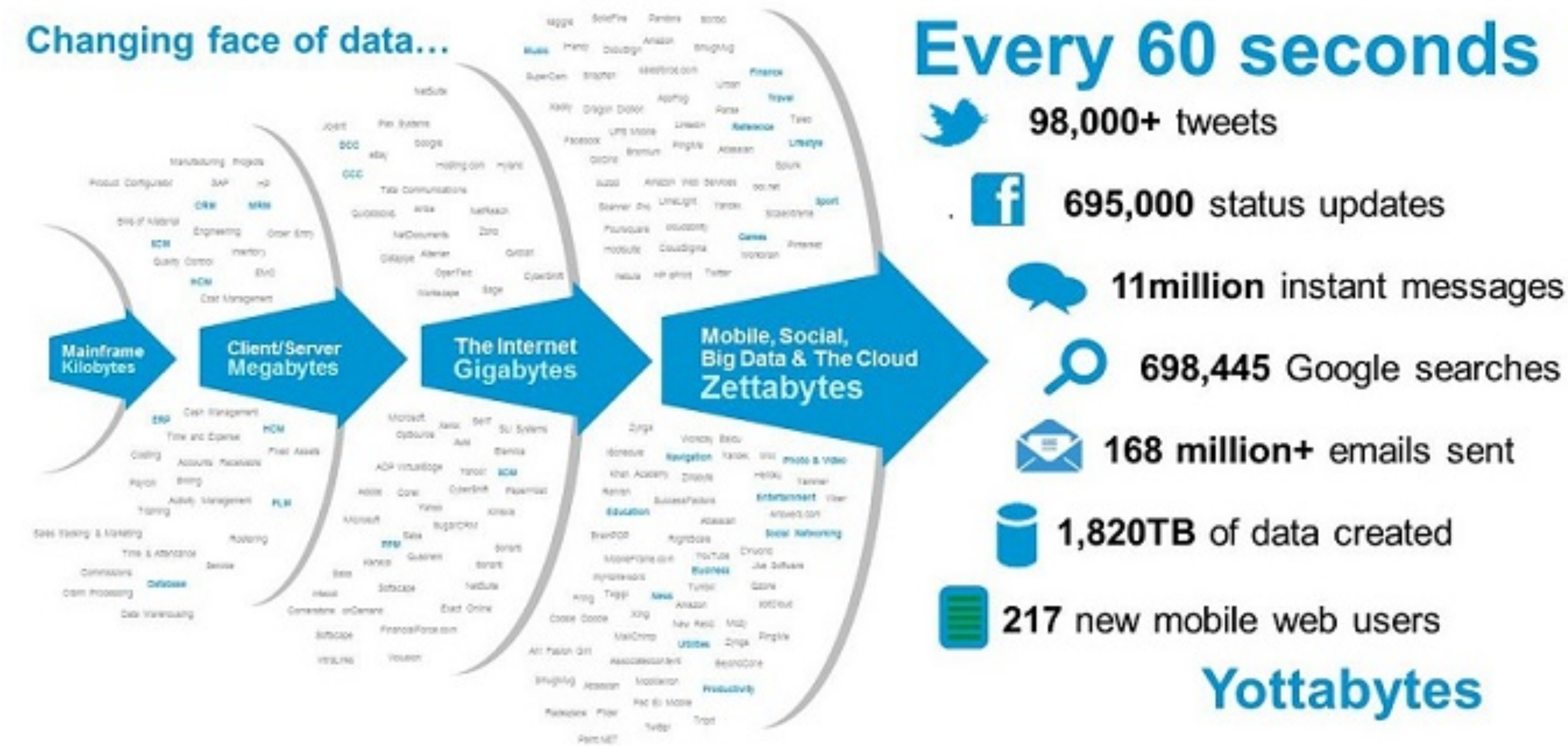
Starfish: A Self-tuning System for Big Data Analytics

Herodotos Herodotou, Harold Lim, Gang Luo, Nedyalko Borisov, Liang Dong, Fatma Bilgen Cetin, Shivnath Babu

Department of Computer Science
Duke University

Presented by Nirupam Roy

The Growth of Data



MAD: Features of Ideal Analytics System

Magnetism
-- accept all data

Agility
-- adapt with data,
real-time processing

Depth
-- allow complex analysis

Hadoop is MAD

Magnetism
-- accept all data

- Blindly loads data into HDFS.

Agility
-- adapt with data, real-time processing

- Fine-grained scheduler
- End-to-end data pipeline
- Dynamic node addition/ dropping

Depth
-- allow complex analysis

- Well integrated with programming languages

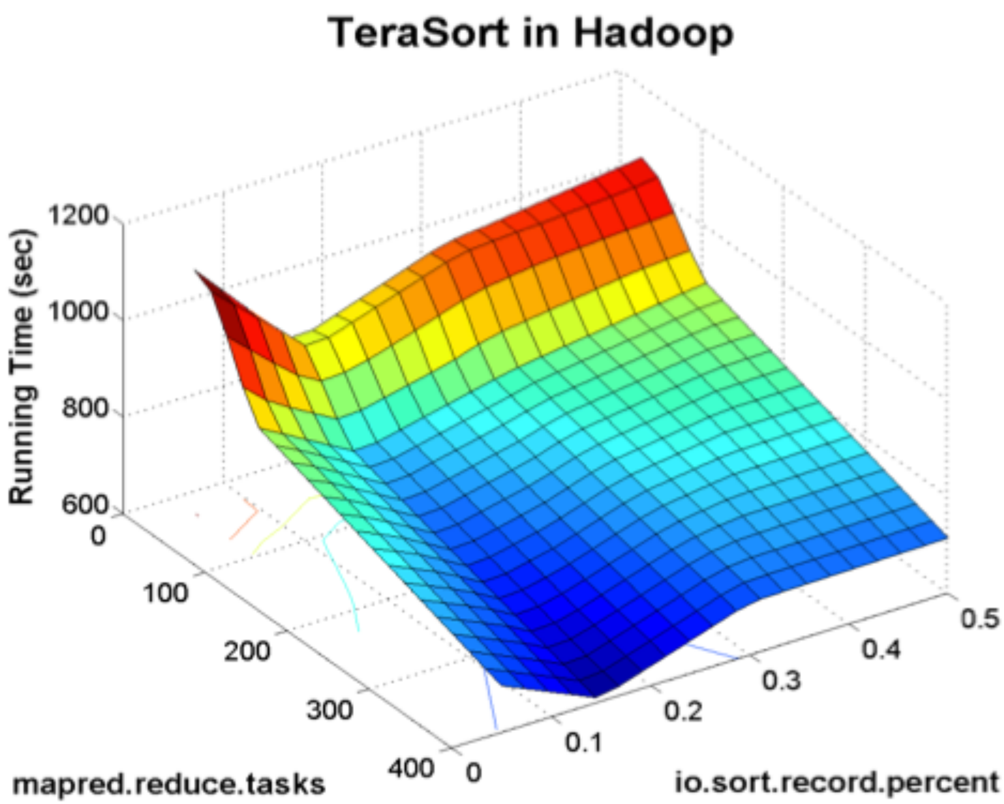
Tuning for Good Performance: Challenges

- Multiple dimensions of performance
    -- time, cost, scalability …
- Multiple levels of abstraction
    -- job-level, workflow-level, workload-level …
- Tons of Parameters
    -- more than 190 parameters in Hadoop.

Tuning for Good Performance: Challenges

Thumb rule

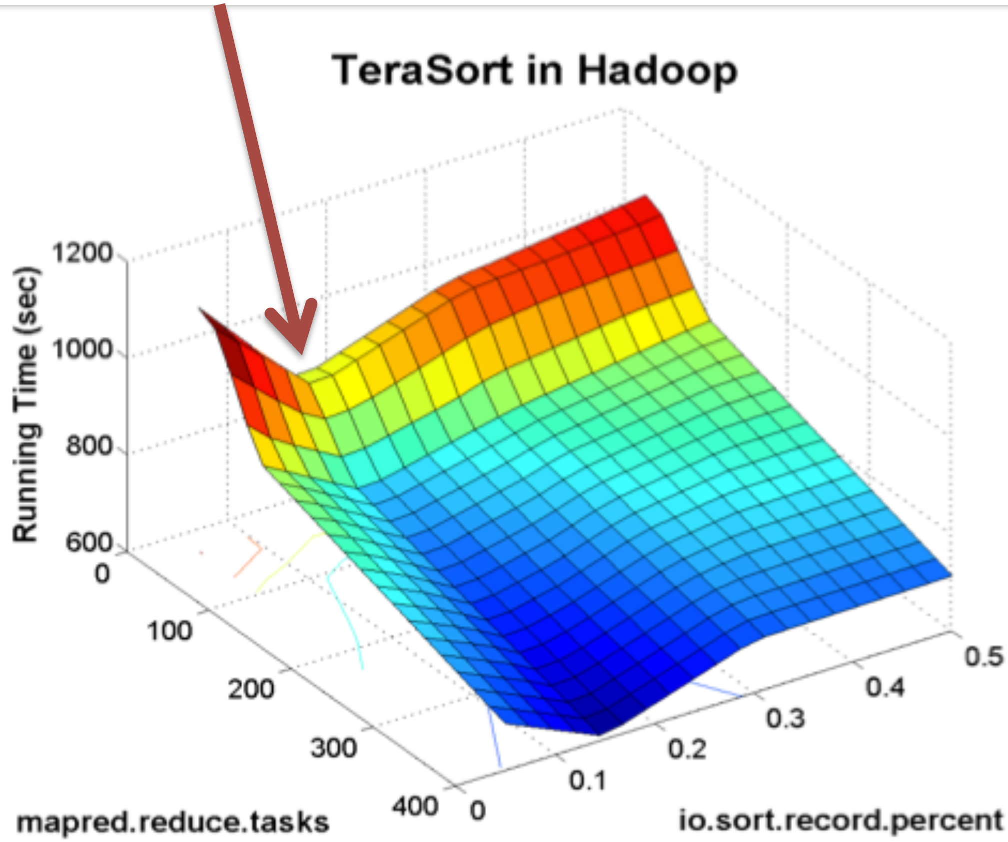
- *mapred.reduce.tasks* = 0.9 * number_of_reduce_slots
- *io.sort.record.percent* = 16 / (16 + average_record_size)



Tuning for Good Performance: Challenges

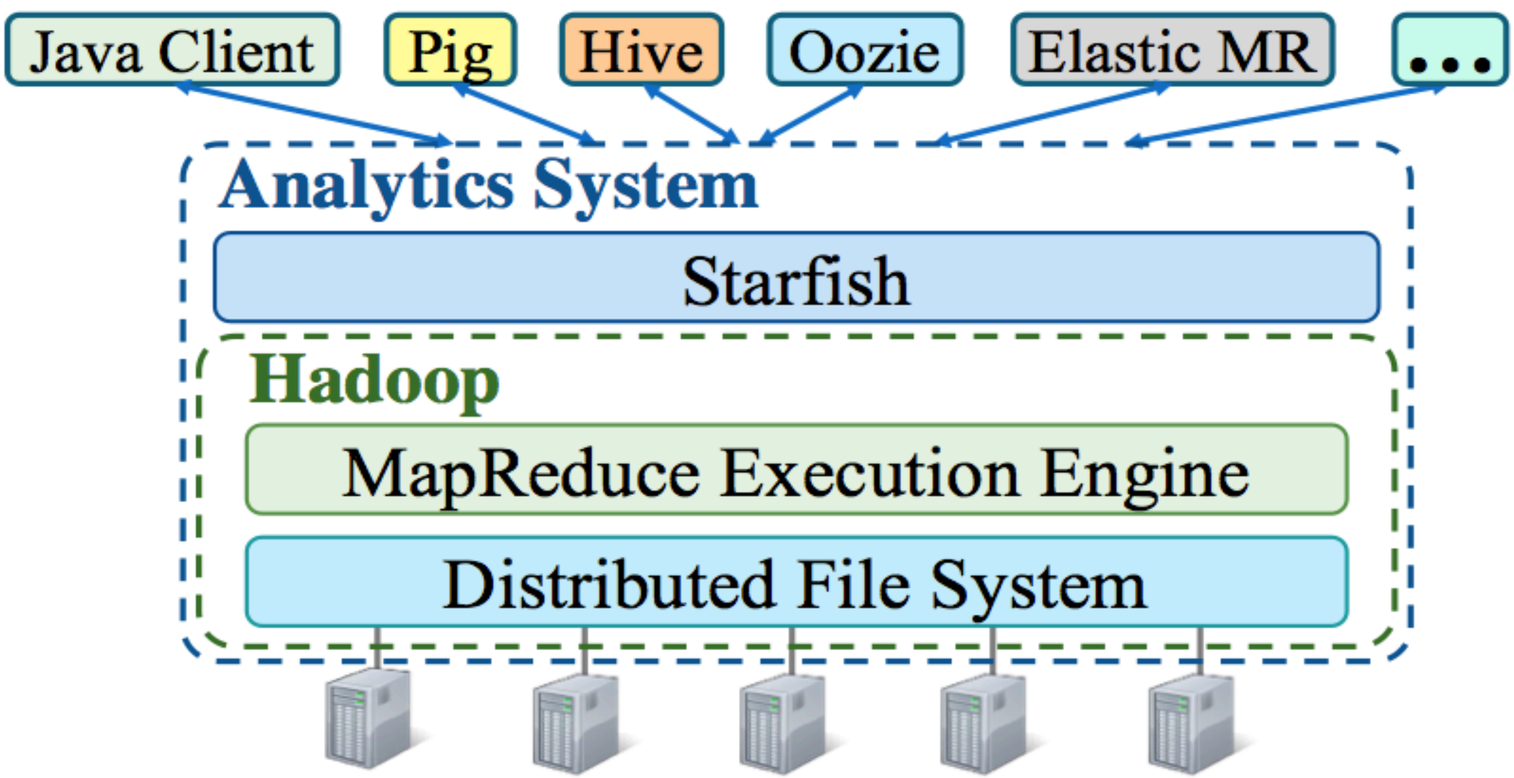
Thumb rule

- *mapred.reduce.tasks* = 0.9 * number_of_reduce_slots
- *io.sort.record.percent* = 16 / (16 + average_record_size)

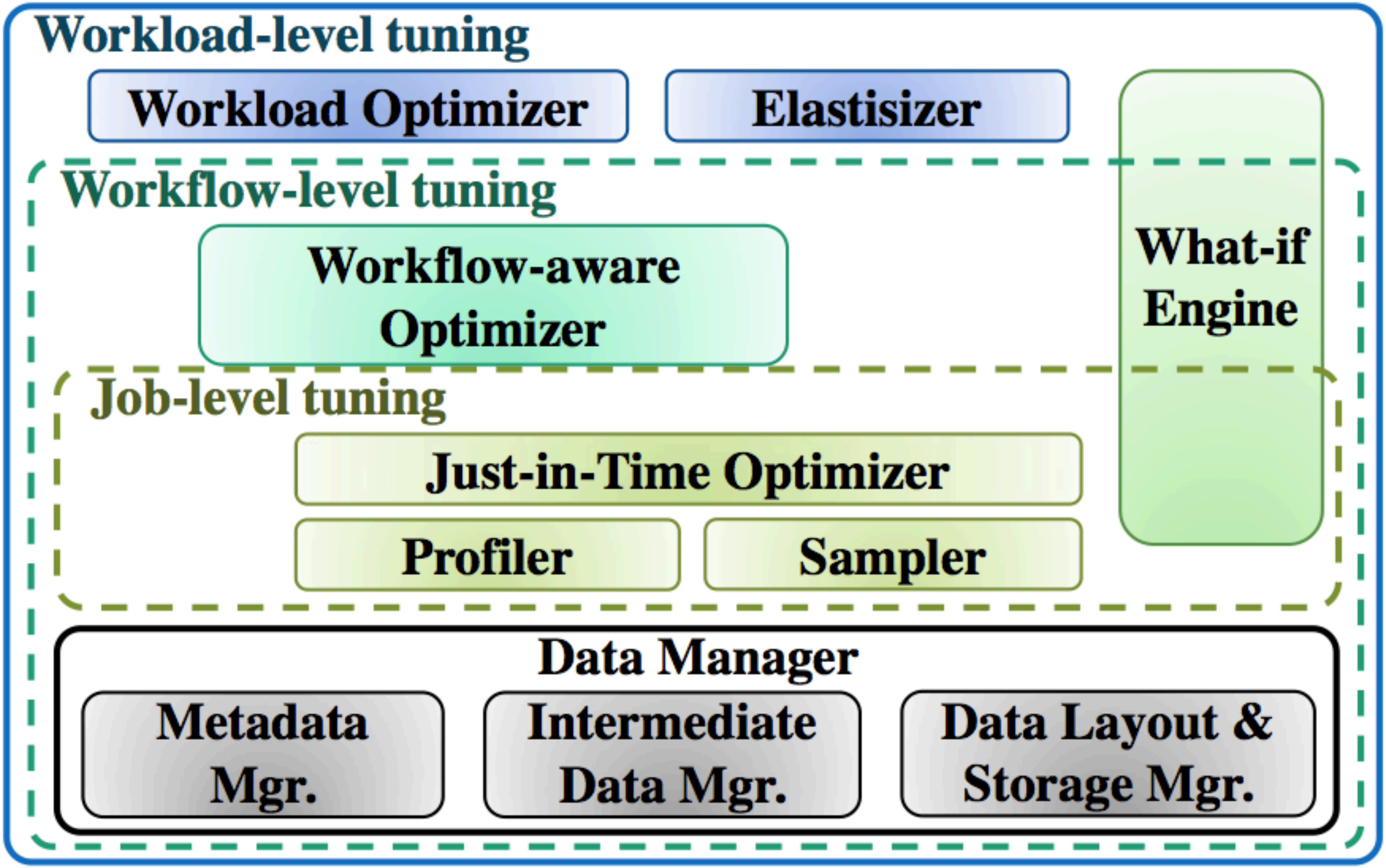


Starfish: A Self-tuning System

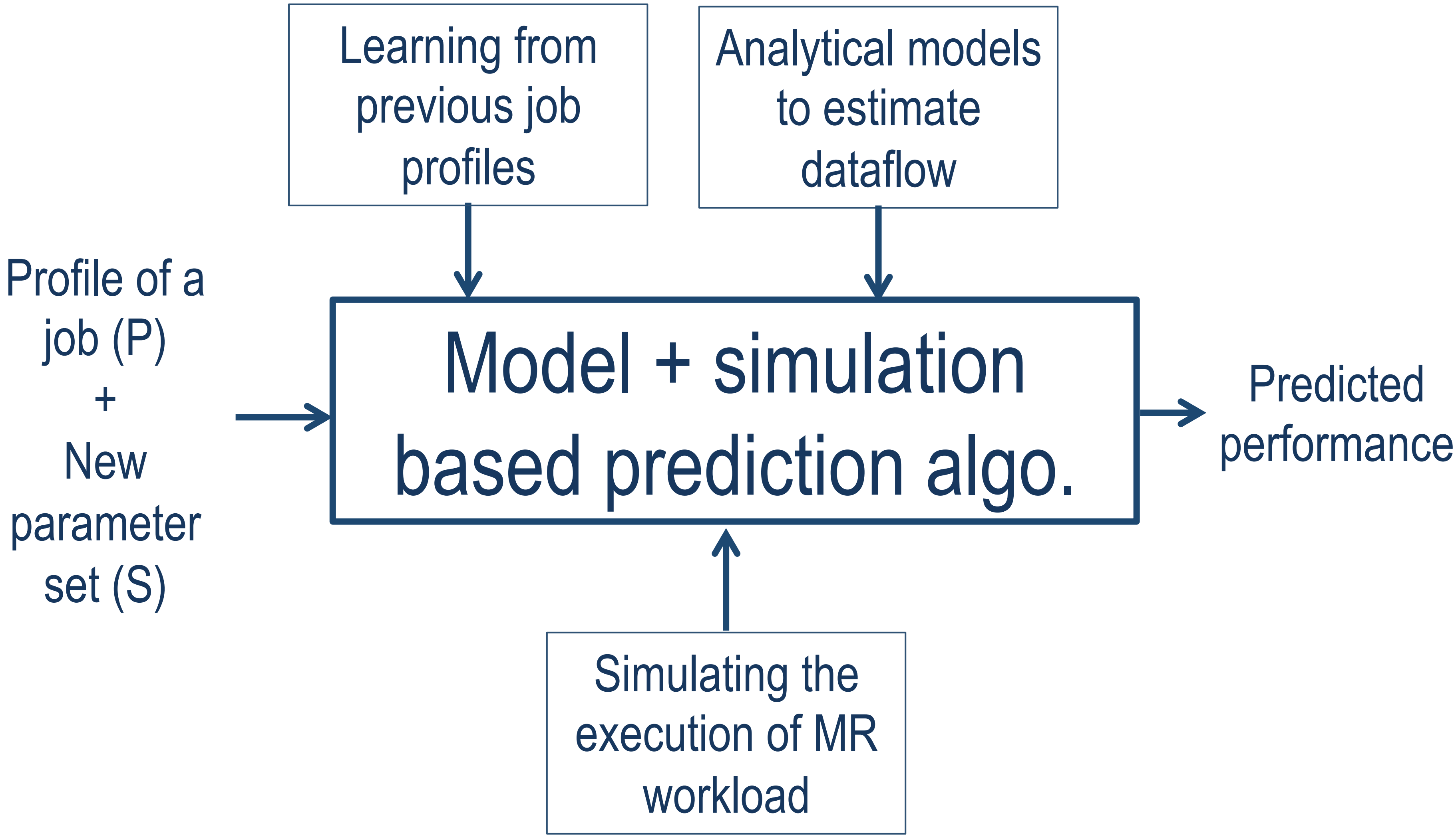
- Builds on Hadoop
- Tunes to ‘good’ performance automatically



Starfish Architecture

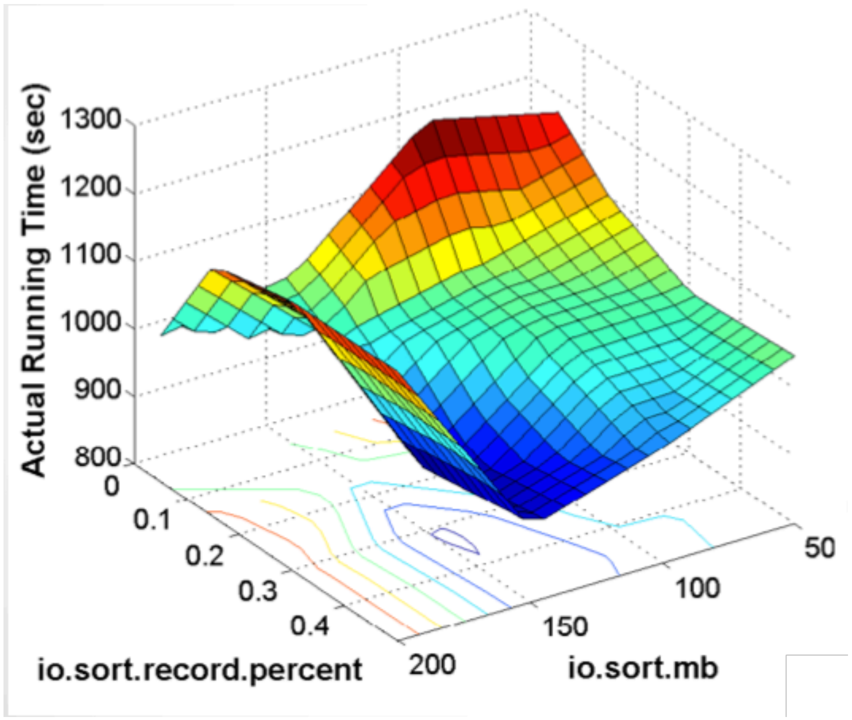


The "What-if" Engine

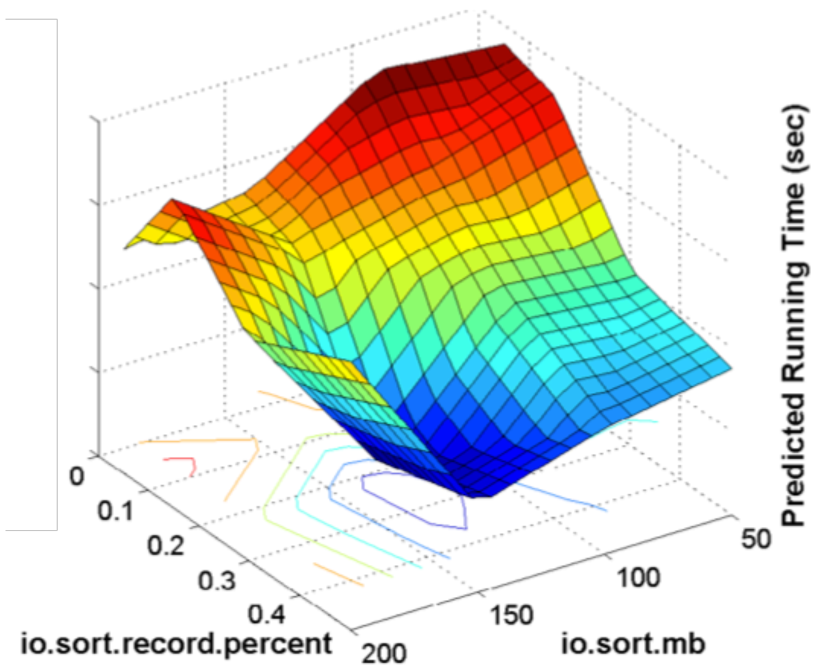


[Ref:] A What-if Engine for Cost-based MapReduce Optimization. H. Herodotou et.al.

The "What-if" Engine

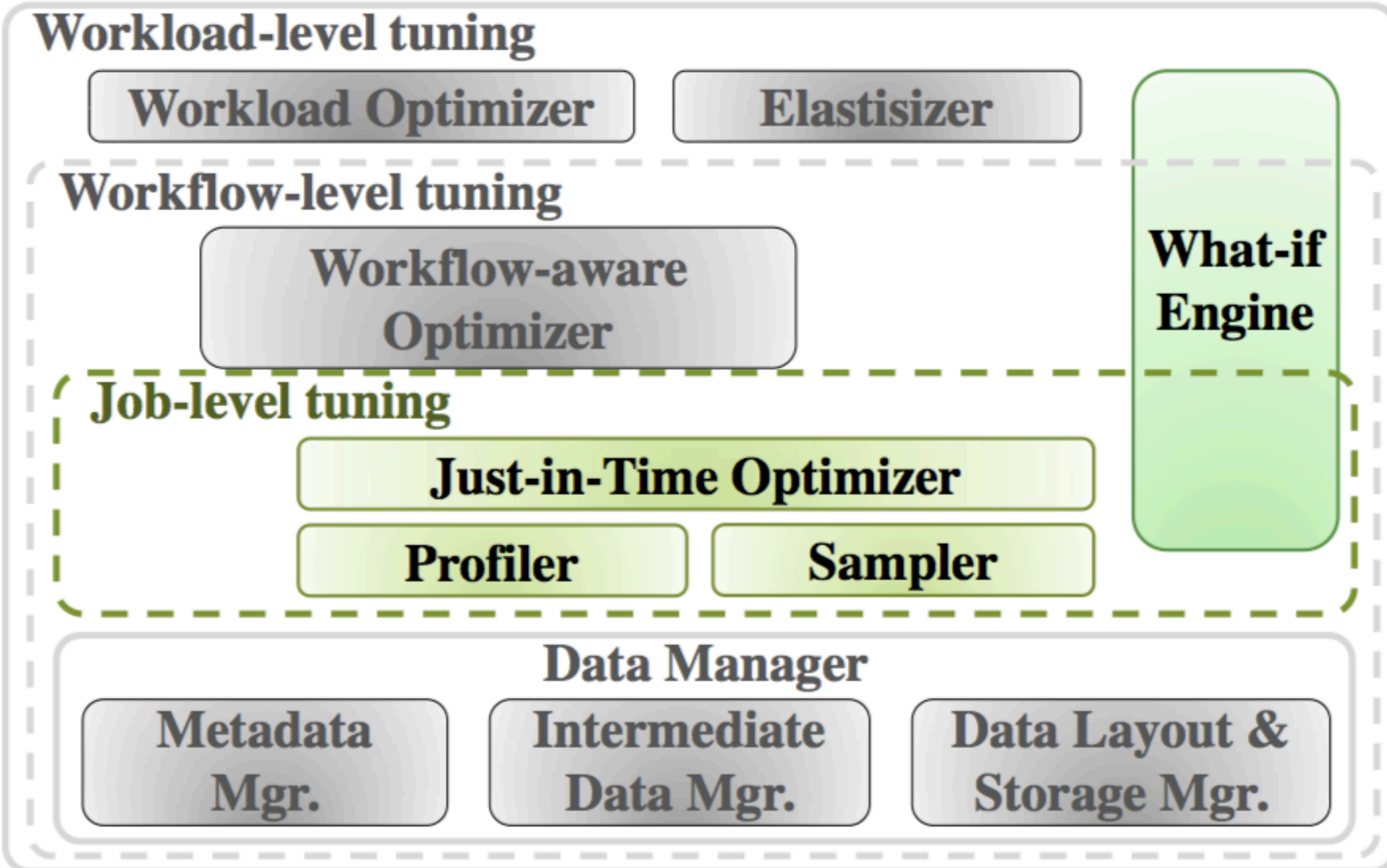


Ground truth



Estimated by the What-if engine

Starfish Architecture: Job Level



Starfish Architecture: Job Level

Just-in-time optimizer

-- Searches the parameter space

Profiler

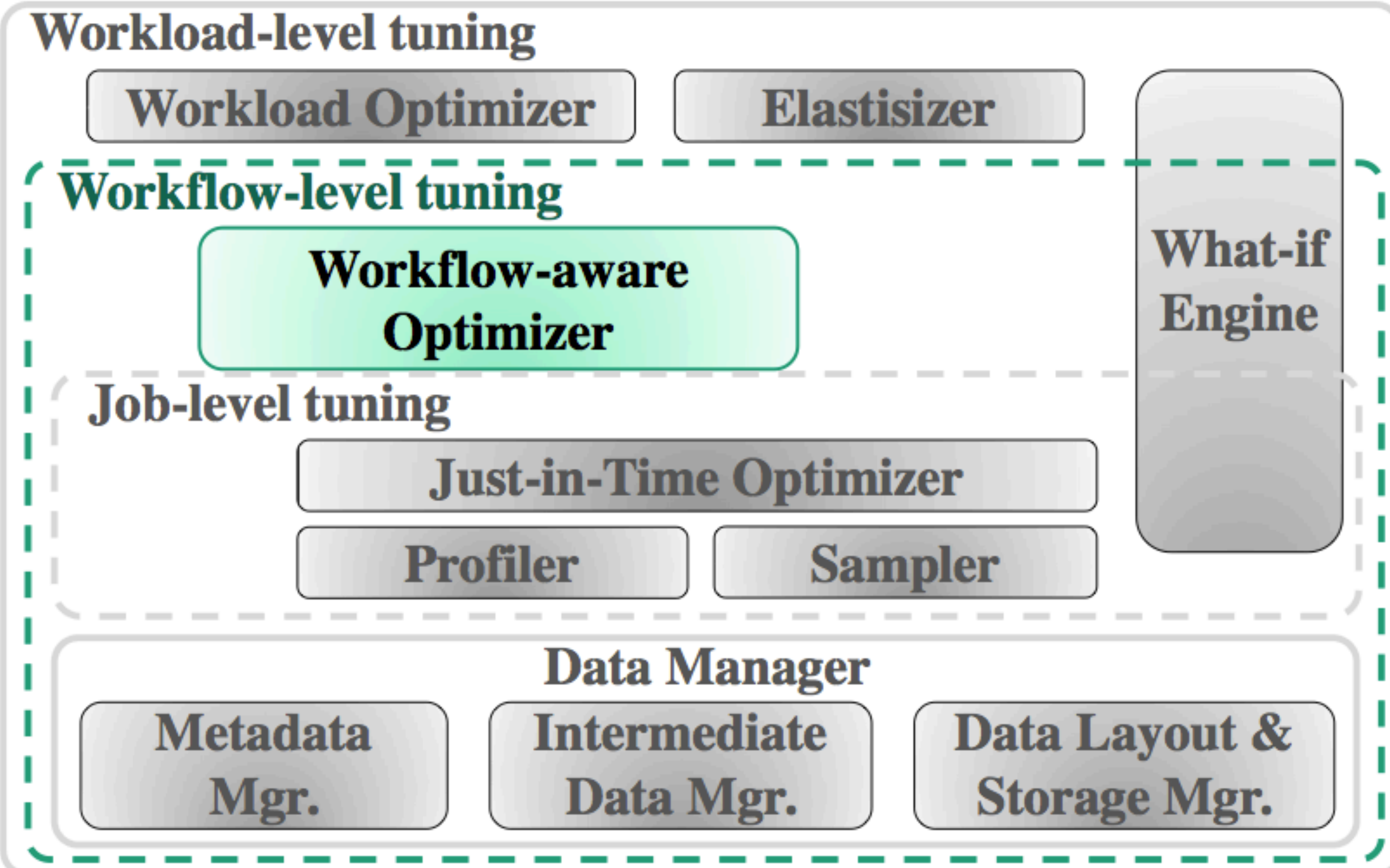
-- Collects info. on MapReduce job execution through dynamic instrumentation

-- Reports timings, data size, and resource utilization

Sampler

-- Generates profile statistics from training benchmark jobs

Starfish Architecture: Workflow Level



Starfish Architecture: Workflow Level

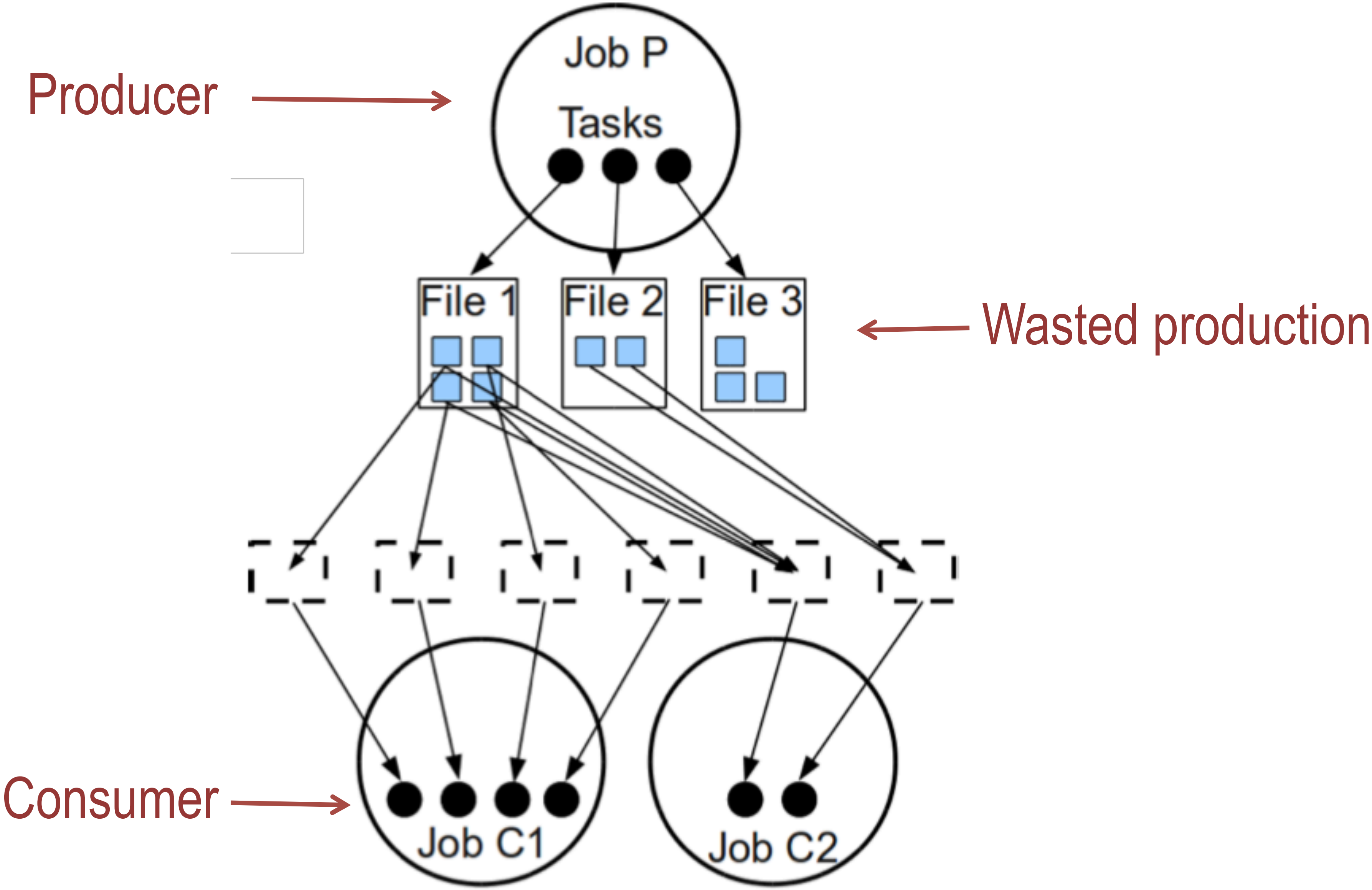
Scheduler to balanced distribution of data

-- deals with skewed data, add/drop of nodes,
tradeoff between balanced data v/s data-locality

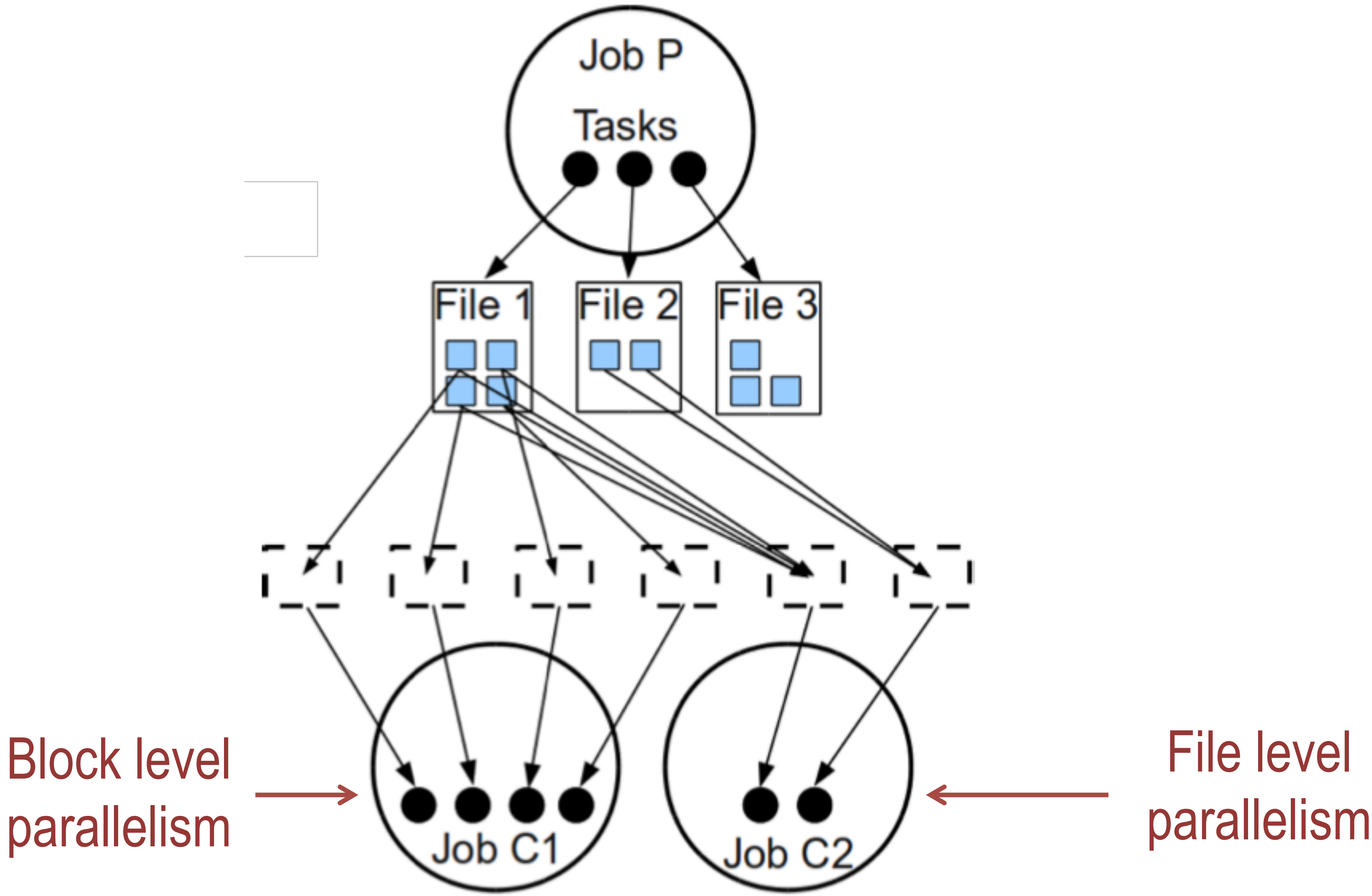
Block placement policy for data collocation

-- Local-write v/s round-robin

Starfish Architecture: Workflow Level



Starfish Architecture: Workflow Level



Starfish Architecture: Workflow Level

Workflow Aware Optimizer

Select best data layout and job parameters

Running time?

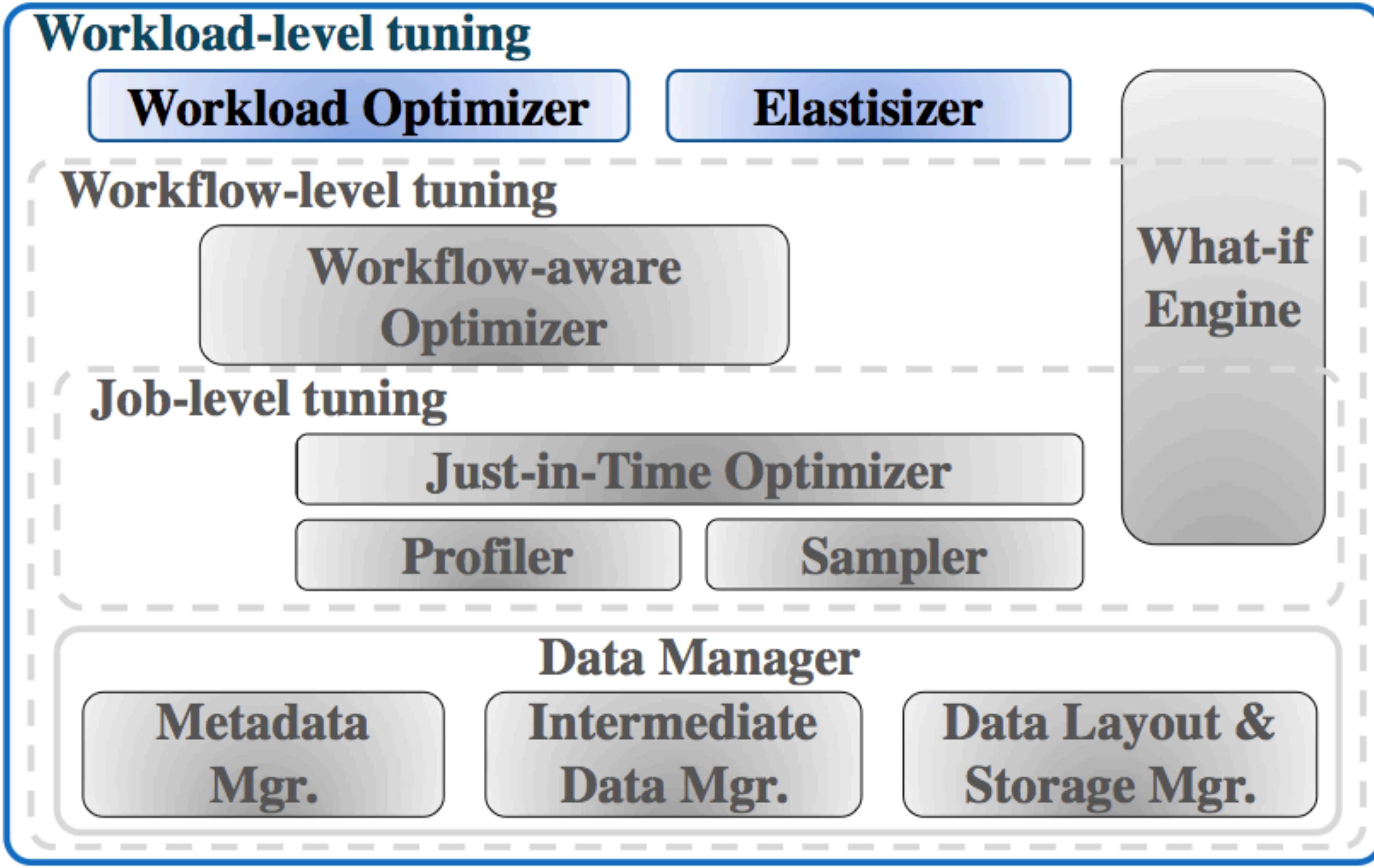
Data layout?

What-if simulation

- MR job execution
- Task scheduling
- Block placement

Compare cost & benefits

Starfish Architecture: Workload Level



Starfish Architecture: Workload Level

Workload Optimizer

- Jumbo operator
- Cost based estimation for best optimization

Elastisizer

- Determine best cluster and Hadoop configurations

Starfish: Summary

- Optimizes on different granularities
    -- Workload, workflow, job (procedural & declarative)
- Considers different decision points
    -- Provisioning, optimization, Scheduling, Data layout

Starfish: Piazza Discussion

Top criticisms (till 1:30pm, 17 reviews):

1) Limited evaluation: 10
2) Not explained well: 7
3) Profiler overhead/better search algo: 5

* What is the effect of wrong prediction?

* What-if engine requires prior knowledge.

Thank you.

Starfish

Self-tuning Analytics System

http://www.cs.duke.edu/starfish/

Photo courtesy: Starfish group, Duke University

Going MAD with Big Data

Magnetic system

Agile system and Analytics

Deep Analytics

Data Life Cycle Awareness

Elasticity

Robustness

Backup: What-if Engine 1

Starfish's What-if Engine can answer any *what-if question* of the following general form:[1]

> Given the profile of a job $j = \langle p, d_1, r_1, c_1 \rangle$ that runs a MapReduce program p over input data d_1 and cluster resources r_1 using configuration c_1, what will the performance of program p be if p is run over input data d_2 and cluster resources r_2 using configuration c_2? That is, how will job $j' = \langle p, d_2, r_2, c_2 \rangle$ perform?

Backup: What-if Engine 2

Algorithm for predicting MapReduce workflow performance

Input: Profile of jobs in workflow, Cluster resources, Base dataset properties, Configuration settings
Output: Prediction for the MapReduce workflow performance

For each (job profile in workflow in topological sort order) {
    Estimate the virtual job profile for the hypothetical job (Sections 3.1, 3.2, and 3.3);
    Simulate the job execution on the cluster resources (Section 3.4);
    Estimate the data properties of the hypothetical derived dataset(s) and the overall job performance;
}

Figure 1: Overall process used by the What-if Engine to predict the performance of a MapReduce workflow.