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

doloffine Particle social 100010 Changing face of data... Anapol. Inight mant plouply Real Property lies Finance Same Car STRONG! 145.04 HARY DIGH DONE Sec. 1 Pip System 10415 Links References #10 1A008 800 00028 Pegilik Armente Bromium. +2.5 Manufacturing Projects HIRTOGRAM HINT Product Configurator Ampon yes bernes 3.44 Tata Contruncipione Station Die Unauget 0.94 10210 1754 Latitute. Quinteriorit. 516 / N/0114 courses. Engineering. Other Street **LaCourante** SCH. method Dilute Abriet Contract. -1005UNE Quality Corror. \$2,40 **OperTet** append Tatter CONTRACT 100.00 **HCN** \$10H Columbus Louise **East Management** The Internet Client/Server Mainframe Kilobytes Gigabytes Megabytes Cash Ranagement. 199 Anna SU Solemi HOM Cablurge Time and Dipense. VOICE BROJ 200 Plast Asses Danca Accura meaners ADP VIRUNIDOR TOTOL: NOW shat Asserty provide Parch Bring. Courses. Plot the Induction I. Core Adult Matapeneti PL H Luises Pailors Yahasi 10,000 STREET. Reparcher. SHE THREE & VANISH Taba. Append. Aurosa. 10.00 donard. TITLE & AMERICANON GLAGVER ADDRESS OF TAXABLE PARTY. 70.7.04 iones' Service torest. **Wardward** NUMBER OF STREET Coloradi National Sofacana Com Proceeding

Harris East Online Comprehens inCompreh Financia/Force spin istance. 177,755

Did Yawaang

1.7540.04 Bours. 100.00 CONTRACT OF Prised: TOTAL PROPERTY. Mobile, Social, **Big Data & The Cloud** Zettabytes

Travel

Tareo

Navigation Italian, this proces a video -0.00 Tanton Entertainment view Arcent.001 July Software Catina 810.0 1000 mult Coosie Coosie 1.412 MalChing Finglish 2mps Att Palities Giff. Associated in No. descentions. projekt season Modilwirph Productively Fac ID Motive Partners' Ploy 1,000

Pairs All?

Every 60 seconds 98,000+ tweets

695,000 status updates

11million instant messages

698,445 Google searches

168 million+ emails sent

1,820TB of data created

217 new mobile web users

Yottabytes

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

Agility -- adapt with data, real-time processing - Blindly loads data into HDFS.

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

Depth -- allow complex analysis - Well integrated with programming languages

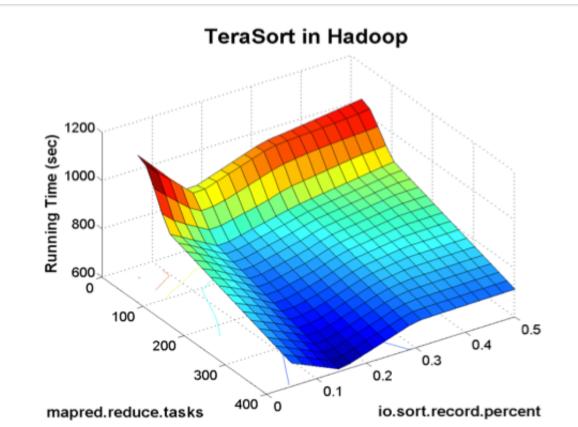
Tuning for Good Performance: Challenges

- 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

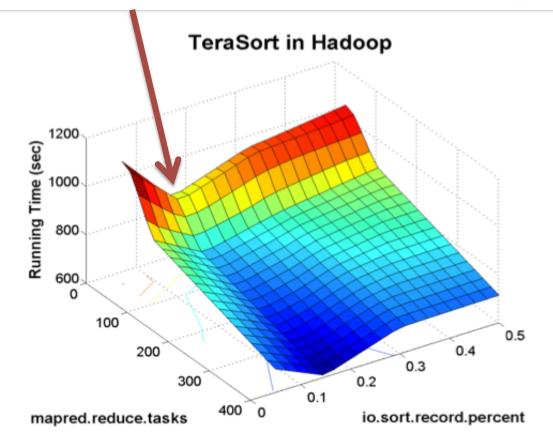
- *mapred.reduce.tasks* = 0.9 * number_of_reduce_slots
- *io.sort.record.percent* = 16 / (16 + average_record_size)



Tuning for Good Performance: Challenges

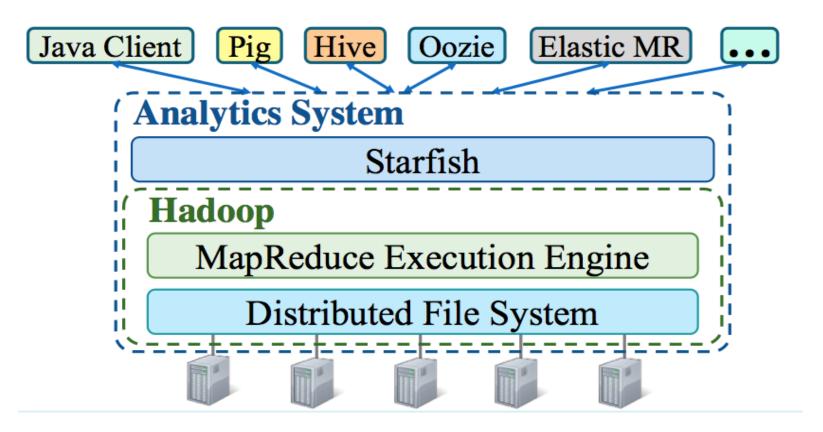
Thumb rule

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

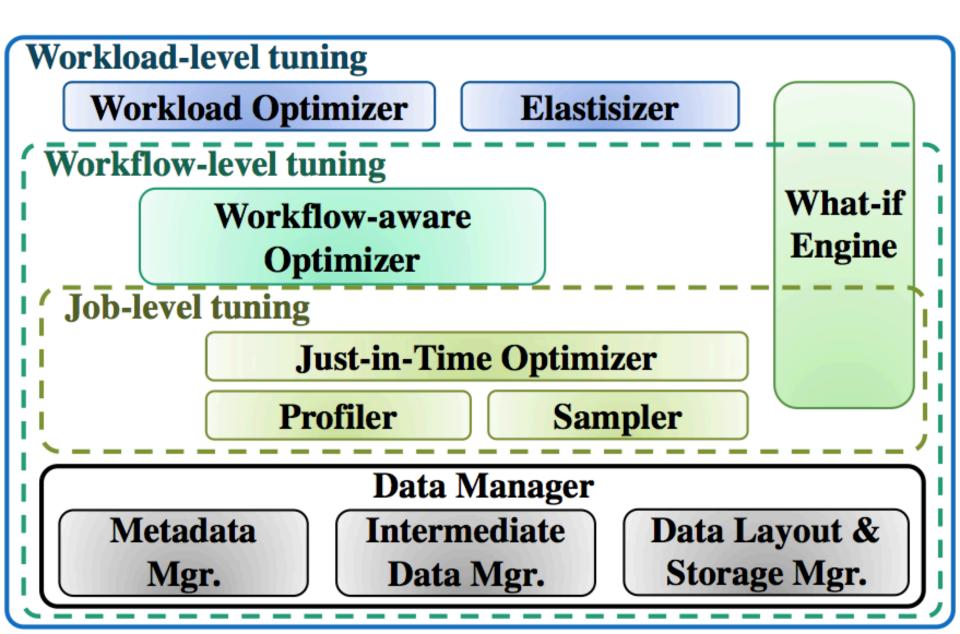


Starfish: A Self-tuning System

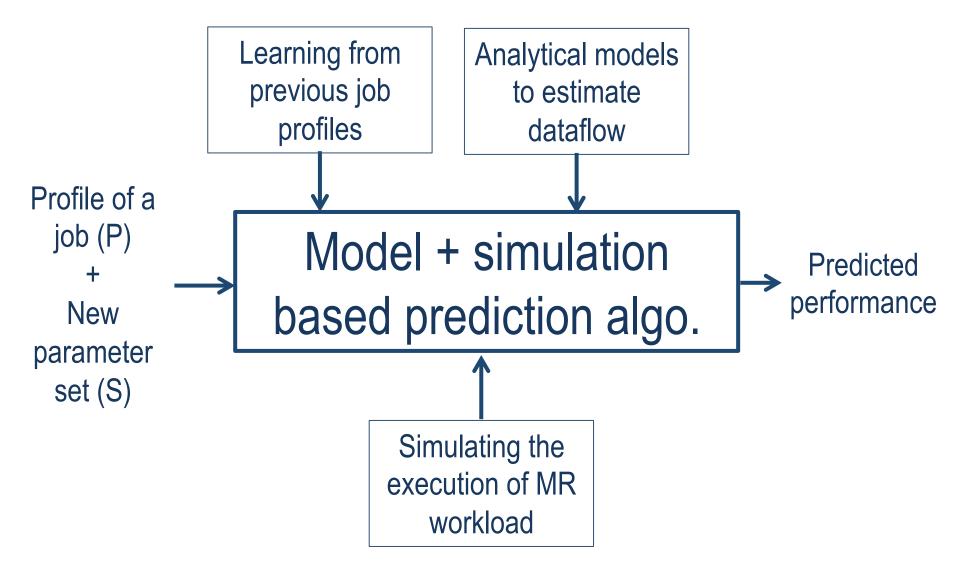
- Builds on Hadoop
- Tunes to 'good' performance automatically



Starfish Architecture



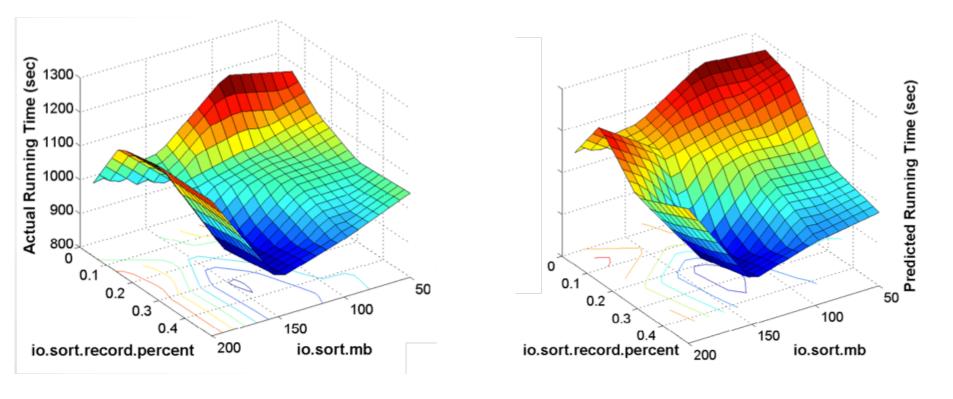
The "What-if" Engine



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

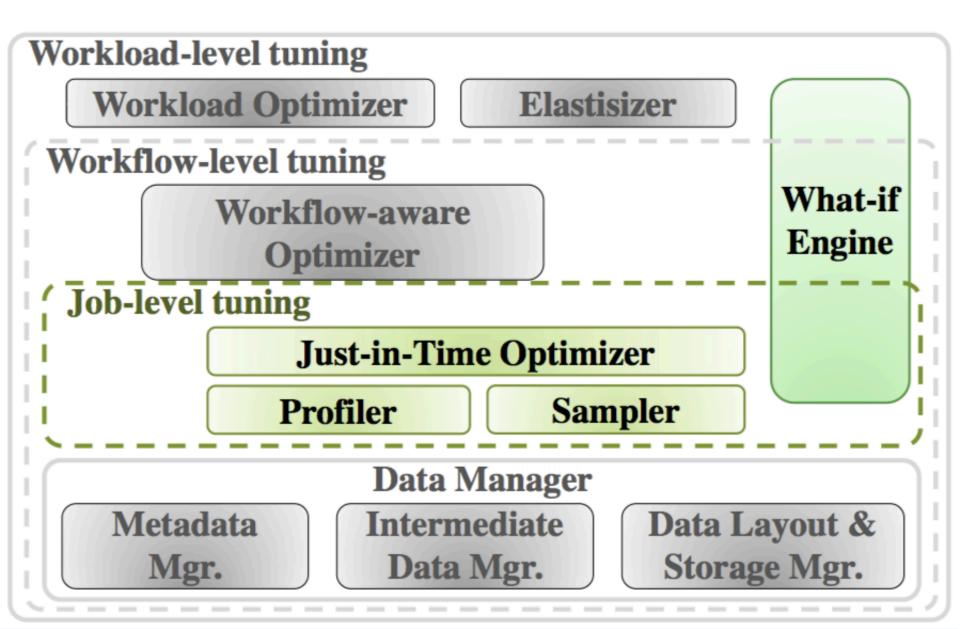
The "What-if" Engine

Ground truth



Estimated by the What-if engine

Starfish Architecture: Job Level



Starfish Architecture: Job Level

Just-in-time optimizer

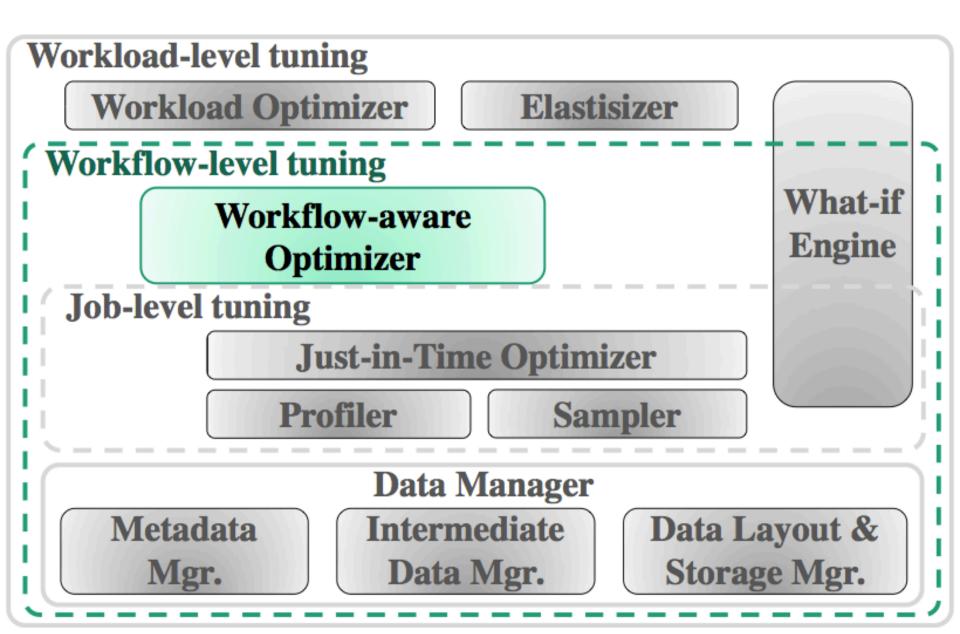
-- Searches the parameter space

Profiler

- -- Collects info. on MapReduce job execution through dynamic instrumentation
- -- Reports timings, data size, and resource utilization

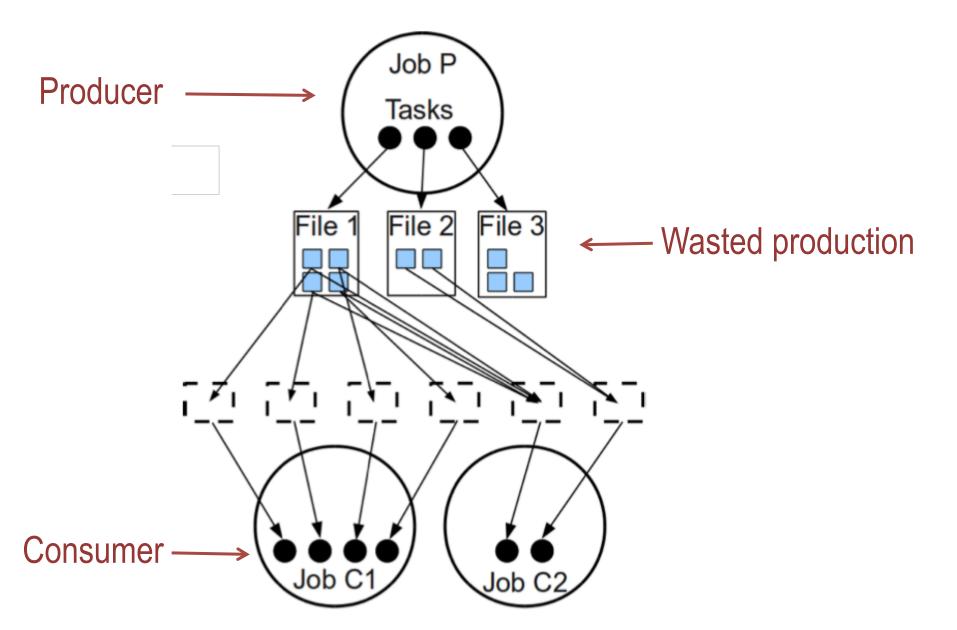
Sampler

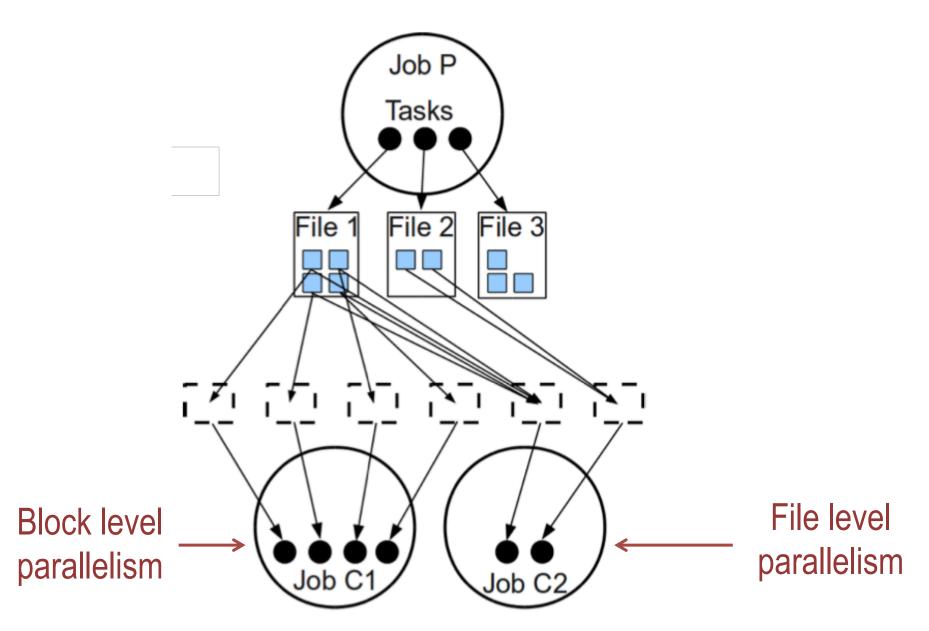
-- Generates profile statistics from training benchmark jobs



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

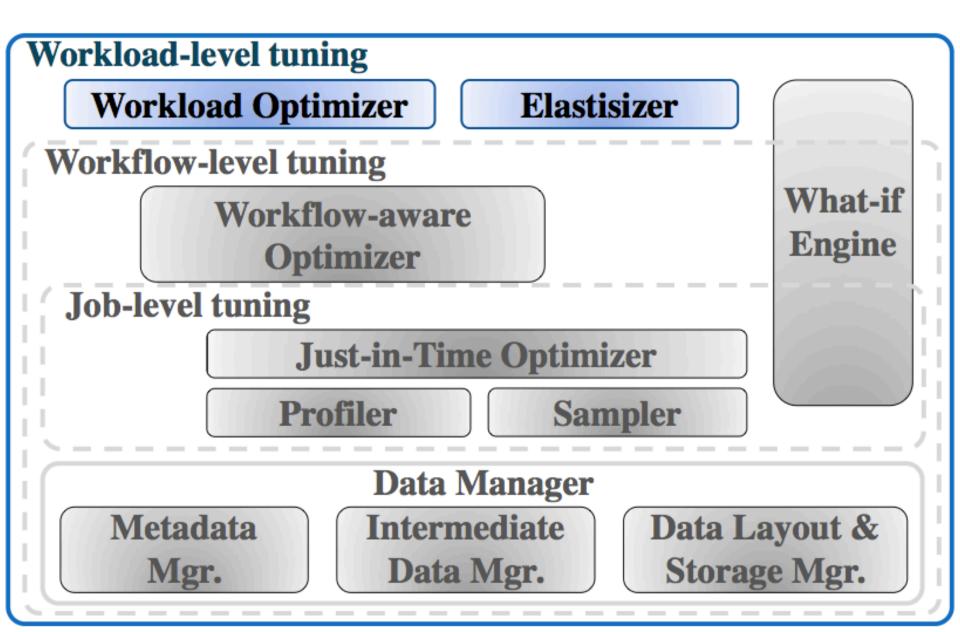




Workflow Aware Optimizer Select best data layout and job parameters



Block placement



Workload Optimizer

- Jumbo operator
- Cost based estimation for best optimization



 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.





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



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.