



**SDC** 

The SDC logo features the letters "SDC" in a large, bold, purple, sans-serif font. To the right of the letters is a graphic element consisting of a yellow square partially overlapping a grey square.

STORAGE DEVELOPER CONFERENCE

SNIA  SANTA CLARA, 2017

The SNIA logo graphic is a small yellow square.An abstract graphic on the left side of the slide features a central white circle surrounded by a network of grey lines and nodes. Various colored hexagons (blue, yellow, orange, green) are scattered around the network. The background is a light grey with a subtle hexagonal pattern.

# **An Effective and Efficient Performance Optimization Method by Design & Experiment: A Case Study with Ceph Storage**

**Lay Wai Kong, Ph.D.**

**Six Sigma Master Black Belt  
Intel Corporation**

# Objective:

- ❑ Design of experiments (DOE)
  - ❑ Who, What, Why, When, Where
  - ❑ Differences from One-factor-at-a-time (OFAT)
  - ❑ Methodology
- ❑ Case study with Ceph storage
  - ❑ HDD & SSD
  - ❑ DOE tuning result
  - ❑ Effectiveness and efficiency by DOE
  - ❑ The benefits of DOE



# Design of Experiments



# Scenario: Does it happen to you

- ❑ A new complex problem with many variables : how to start? How many test needed?
- ❑ Which variables has high impact and which variables not important?
- ❑ Is this the true optimization result from my configuration?
- ❑ I'm guessing this parameter might change the outcome!
- ❑ What is the interaction between all the variables?



# Who should know

- ❑ Whoever trying to solve a complex problem with multiple variables
  - ❑ Engineer
  - ❑ Researcher
  - ❑ Developer



# What is DOE

- ❑ Systematic approach to complex problem-solving.
- ❑ Data collection stage: Applies principles and techniques
- ❑ Ensure generation: valid, defensible and supportable conclusions.



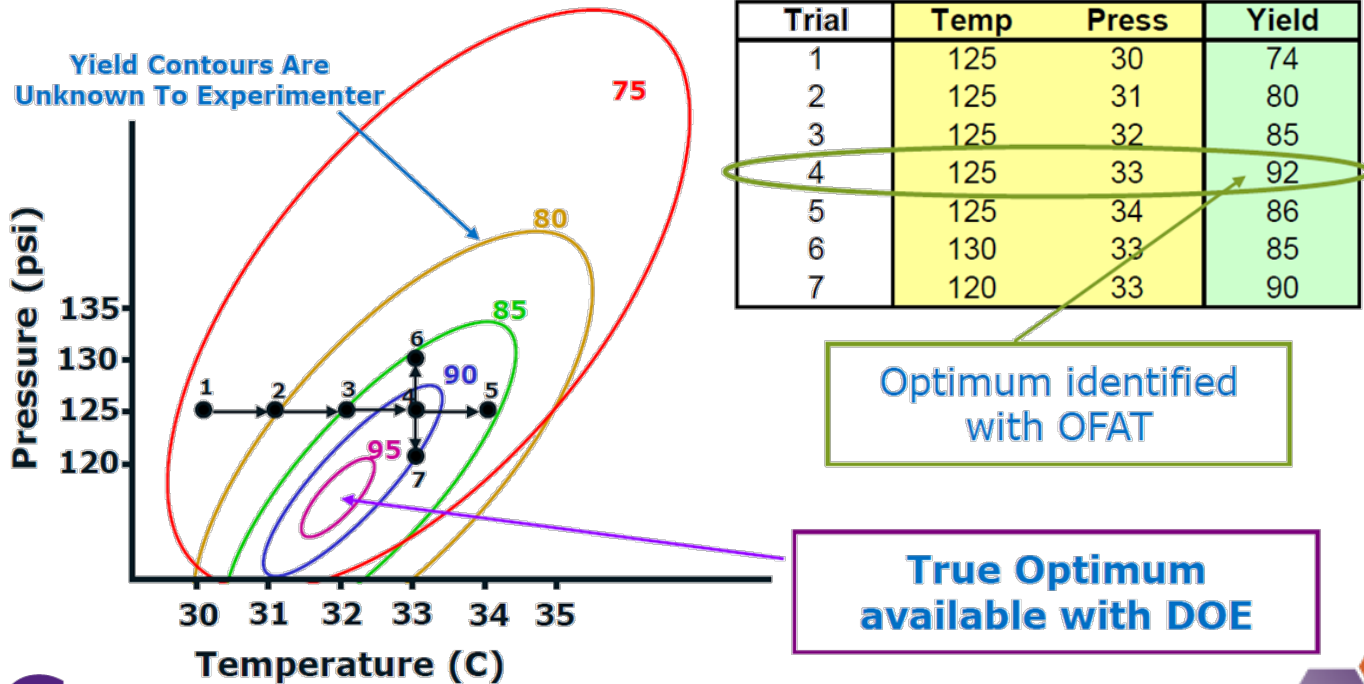
# Why DOE

- ❑ Simply an effective and efficient method.
- ❑ Is one-factor-at-a-time (OFAT) the only way?
  - ❑ Intuitive
  - ❑ Change one factor and make others constant
  - ❑ More like a style instead of planned experiment



# Why DOE : Common scenario with OFAT

The graphic shows yield contours for a process that are unknown to the experimenter.



Optimum identified with OFAT

True Optimum available with DOE

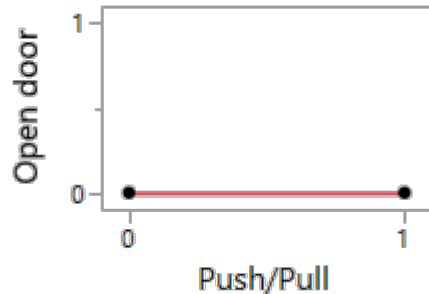
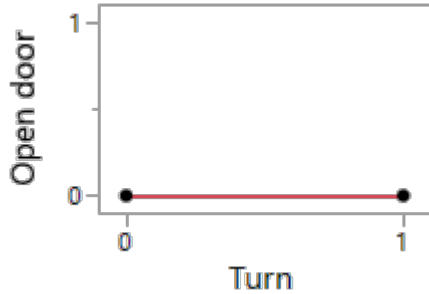




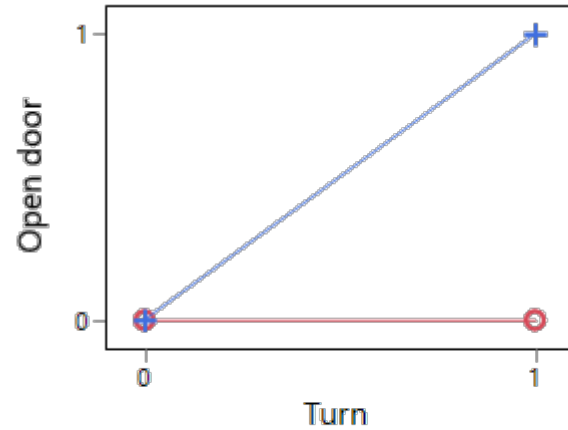
# Why DOE : factor-to-factor interaction

Simple example: open the door (2 factors=turn and push/pull)

### OFAT



### DOE



Groups ○ — Push/Pull=0 + — Push/Pull=1



# When to use

- ❑ Comparative
- ❑ Screening
- ❑ Characterization
- ❑ Modeling
- ❑ Optimization

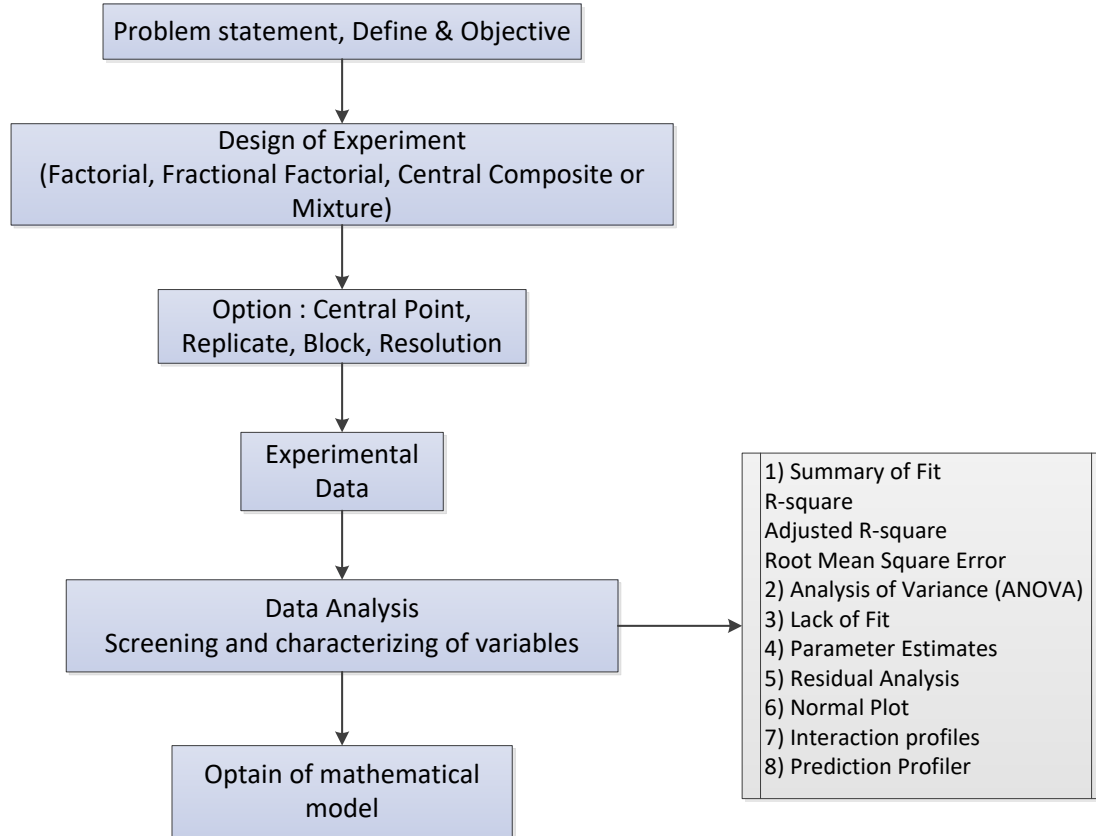


# Where can get

- ❑ Free
  - ❑ R
  - ❑ Action
- ❑ Commercial Statistical packages
  - ❑ SAS (Statistical analysis software)
  - ❑ SPSS (Statistical package for social science)
  - ❑ Statistical, Minitab, Design-Expert, Prisma, etc



# DOE : Recommended Flow & Analysis



# DOE : Full Factorial

- General notation:

$$2^k$$

where,

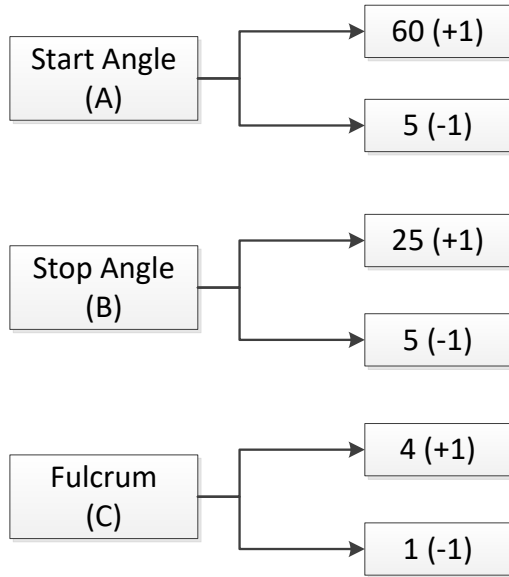
2 is the number of level that will be use for each factor

$k$  is the number of input variables/factors

Total experiment run =  $2^k$



# DOE: Visualize $2^3$ Full Factorial

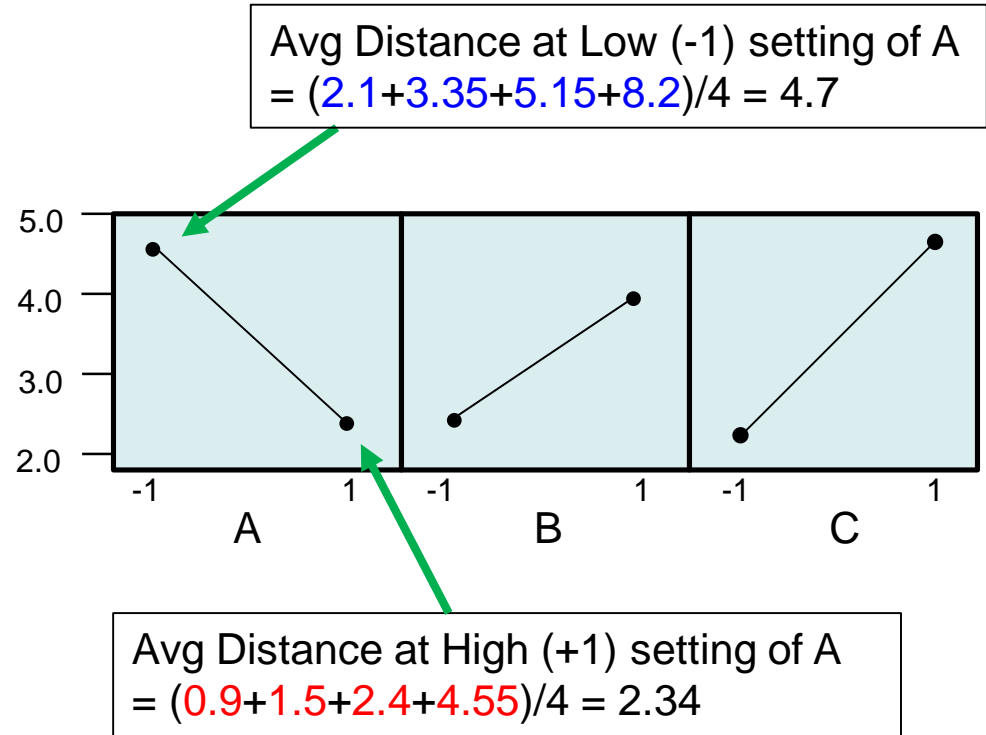


Expt. run#	A	B	C	Distance (Y)
1	-	-	-	2.10
2	+	-	-	0.90
3	-	+	-	3.35
4	+	+	-	1.50
5	-	-	+	5.15
6	+	-	+	2.40
7	-	+	+	8.20
8	+	+	+	4.55



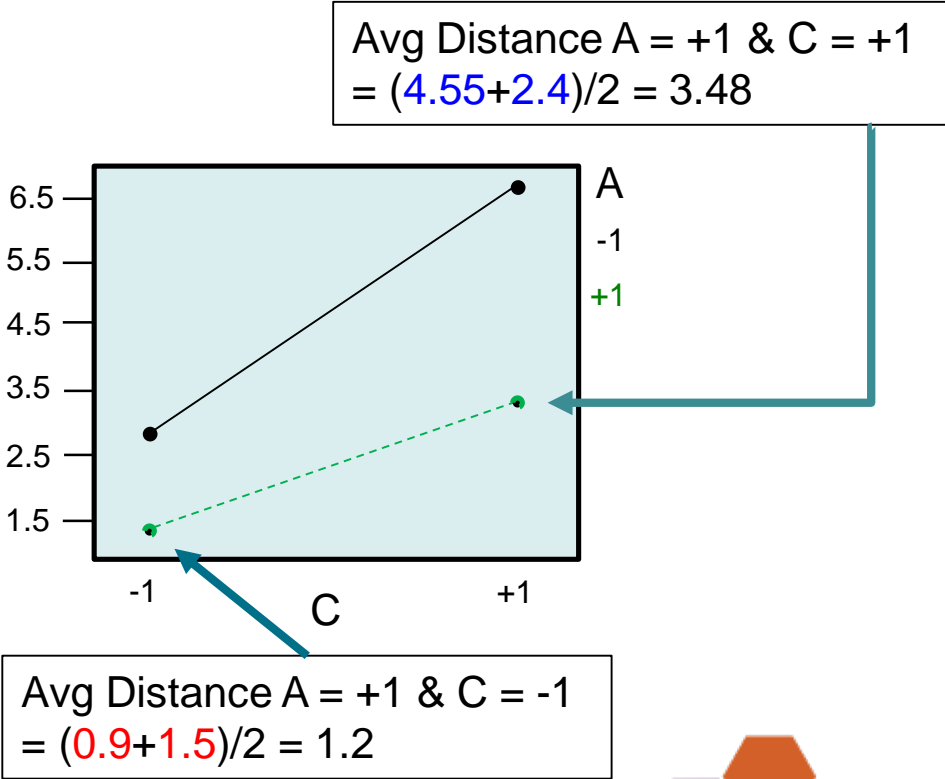
# DOE: Main Effect Plot & Calculation

Expt. run#	A	B	C	Distance (Y)
1	-1	-1	-1	2.10
2	+1	-1	-1	0.90
3	-1	+1	-1	3.35
4	+1	+1	-1	1.50
5	-1	-1	+1	5.15
6	+1	-1	+1	2.40
7	-1	+1	+1	8.20
8	+1	+1	+1	4.55



# DOE: Interaction Effect Plot & Calculation

Expt. run#	A	B	C	Distance (Y)
1	-1	-1	-1	2.10
2	+1	-1	-1	0.90
3	-1	+1	-1	3.35
4	+1	+1	-1	1.50
5	-1	-1	+1	5.15
6	+1	-1	+1	2.40
7	-1	+1	+1	8.20
8	+1	+1	+1	4.55





# DOE: Fractional Factorial

- When  $k \gg 3$
- Resolution & confounding

Resolution		Description
Resolution III	$2_{III}^{k-p}$	No main effect is confounded with any other main effect, but main effects are confounded with two-factor interactions.
Resolution IV	$2_{IV}^{k-p}$	No main effect is confounded with any other main effect, but main effects are confounded with two-factor interactions.
Resolution V	$2_V^{k-p}$	No main effect or two-factor interaction is confounded with any other main effect or two-factor interaction.



# A Case Study with Ceph On All HDD & SSD Storage



# The objective & Plan

- ❑ Performance tuning DOE
  - ❑ Screening factors + improve performance
- ❑ Hypothesis
  - ❑ DOE can screening ceph configuration parameter and suggest a valid optimization setting
- ❑ Experiment & validation
  - ❑ High (SSD) vs low (HDD) performance storage environment
  - ❑ The Tuned performance has to significant higher than Default.
  - ❑ Effective & efficient



# Configuration

Node	Items	Test Environment (HDD)	Test Environment (SSD)
4x Client	Processor	2x Intel® Xeon® CPU E5-2699 v3 @ 2.30GHz	2x Intel® Xeon® CPU E5-2699 v3 @ 2.30GHz
	Memory	128GB	256GB
	OS device	Intel DC S3700 400GB SSD	
	Network	10G Ethernet	
	OS	Centos 7.3	
5x Cluster	Processor	Intel® Xeon® CPU E3-1241 v3 @ 3.50GHz	2x Intel® Xeon® CPU E5-2699 v4 @ 2.20GHz
	Memory	32GB	256GB
	OS device	Intel DC S3700 400GB SSD	
	Public Network	10G Ethernet (for clients access)	
	Private Network	10G Ethernet (for OSD traffic)	
	OS	Centos 7.3	
	Ceph	v10.2.5 (Jewel)	
	OSD	Western Digital* WD4000F9YZ 4TB SATA HDD	Intel DC S3610 1.6TB SSD
	XFS Journal	Intel DC S3700 400GB SSD	Intel DC S3610 1.6TB SSD
	Monitor	3	
	Replicate	3	
	Benchmark Tool	CBT (Ceph benchmarking tool)	



# Factors Selection

Code	Test Environment (HDD)
A1	osd_op_threads
A2	journal_max_write_bytes
A3	journal_max_write_entries
A4	filestore_max_sync_interval
A5	filestore_queue_max_bytes
A6	filestore_op_threads
A7	mon_sync_max_payload_size
A8	mon_sync_fs_threshold
A9	osd_journal_size
A10	filestore_journal_writeahead
A11	filestore_wbthrottle_enable

Code	Test Environment (SSD)
B1	filestore_queue_max_ops
B2	filestore_queue_max_bytes
B3	filestore_max_sync_interval
B4	filestore_op_threads
B5	journal_max_write_entries
B6	journal_max_write_bytes
B7	objecter_inflight_ops
B8	objecter_inflight_op_bytes
B9	ms_dispatch_throttle_bytes
B10	osd_op_num_shards
B11	osd_op_num_threads_per_shard



# DOE Plan

- ❑ Fractional Factorial
  - ❑ Total runs = 266 [( $2^7$  + 5 central point)\*2]
  - ❑ If Full Factorial =  $2^{11}$  = 2048 runs = ~22 days
- ❑ Resolution V
- ❑ Total time for each test environment ~ 3 days.
- ❑ Total time for analysis ~1-2 days



# DOE: Main and Interaction Effect results

#	Test Environment (HDD): Analysis suggestion		
	Factor	Level direction	Optimized value
1	A6	+1	10
2	A4	-1	1
3	A4 * A6	(-1) * (+1)	(1) * (10)

+1 = set to higher level

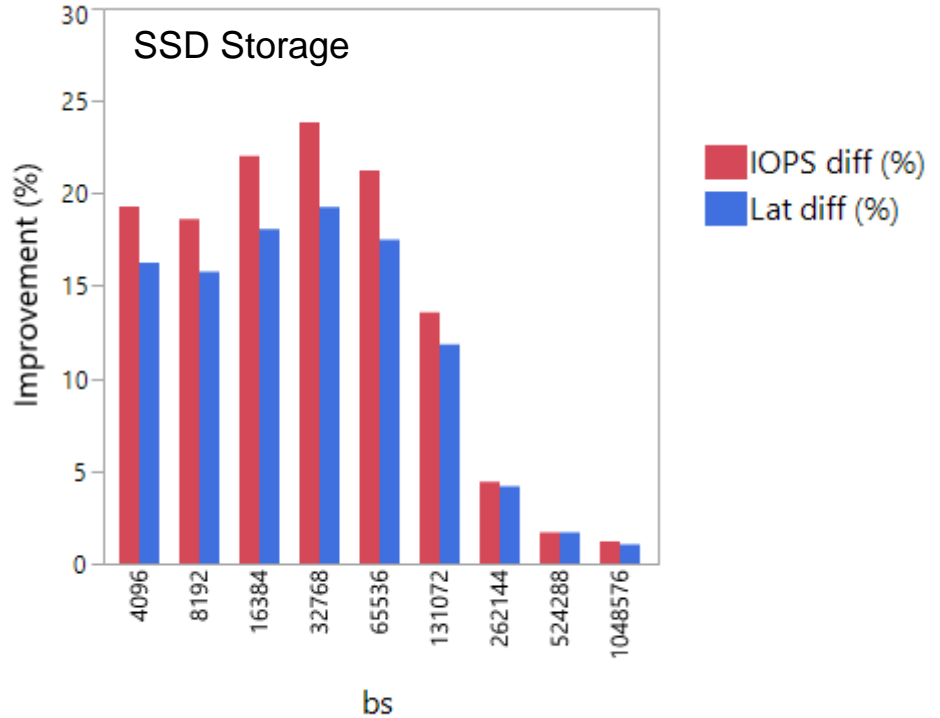
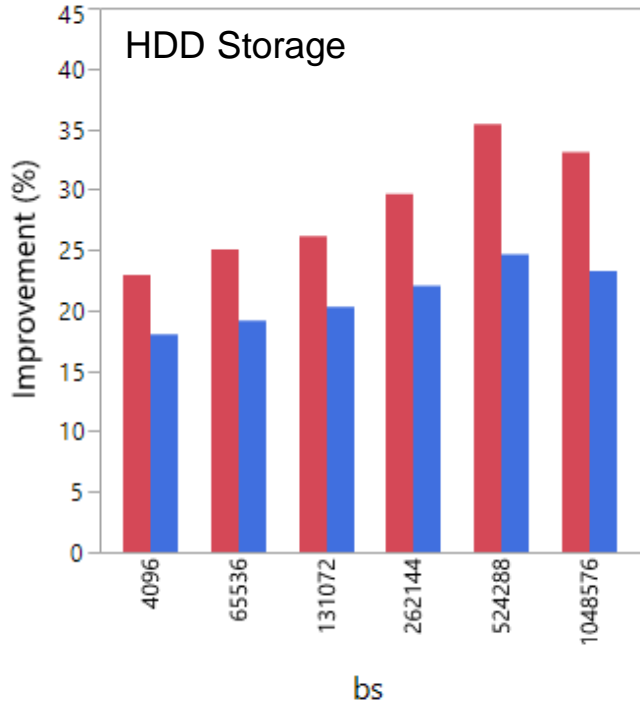
-1 = set to lower level

#	Test Environment (SSD): Analysis suggestion		
	Factor	Level direction	Optimized value
1	B3	+1	20
2	B4	-1	1
3	B5*B10	(+1) * (-1)	(1000) * (4)
4	B2*B11	(+1) * (-1)	(1048576000) * (2)
5	B7*B11	(+1) * (-1)	(102400) * (2)
6	B4*B5	(-1) * (-1)	(1) * (1000)
7	B1*B7	(+1) * (+1)	(5000) * (102400)



# Results:

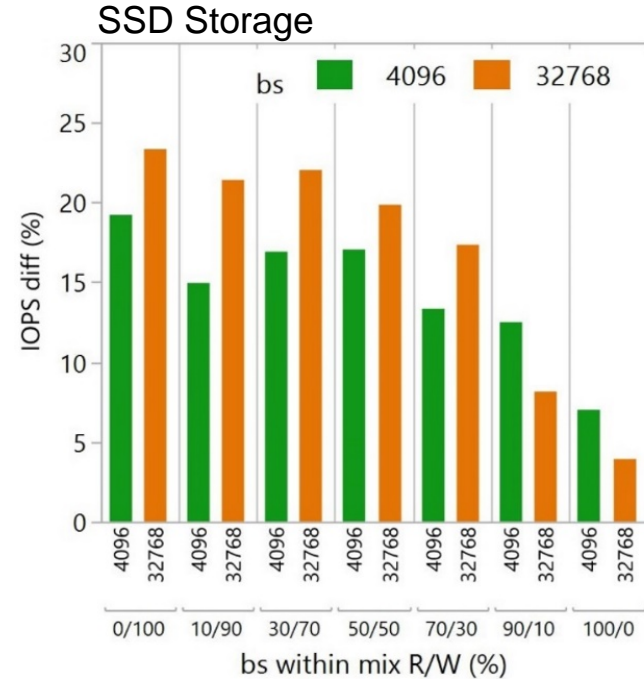
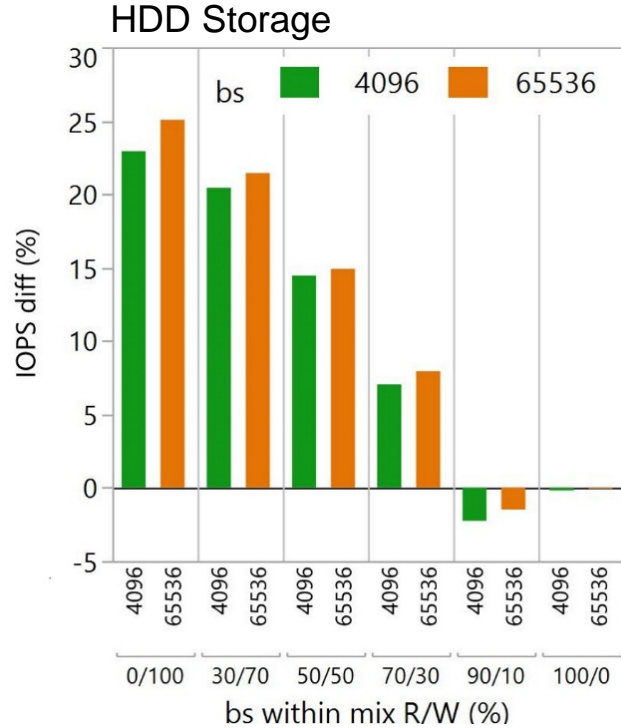
## Improvement (IOPS/Lat vs bs)



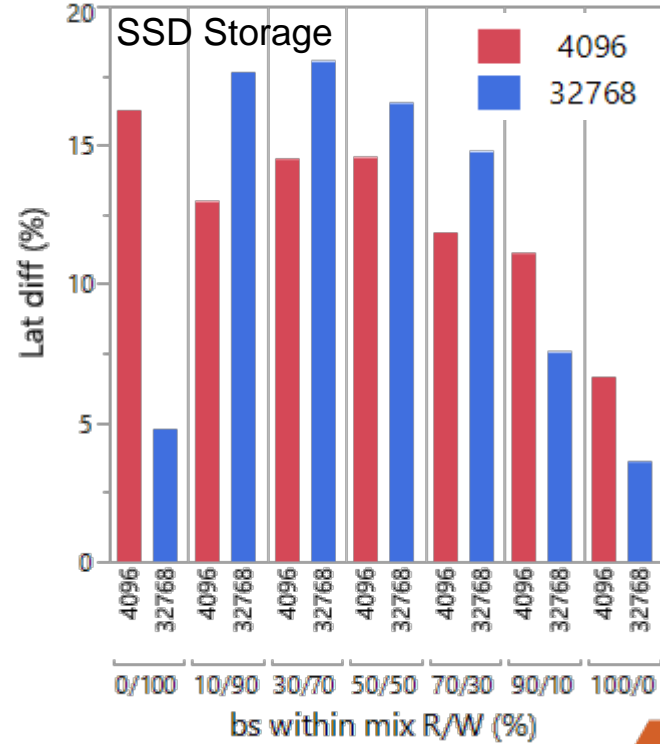
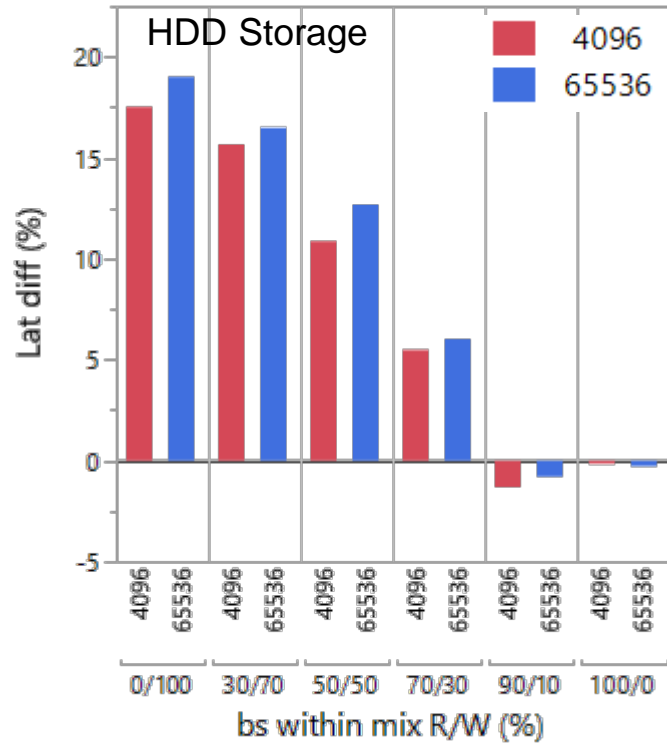


# Results:

## Improvement (IOPS vs bs/Mix R/W %)



# Results: Improvement (Lat vs bs/Mix R/W %)



# Summary

- ❑ Efficient method:
  - ❑ Systematical problem solving approach
  - ❑ In 7-8 days
  - ❑ 2 types of storage cluster
- ❑ Effective in result:
  - ❑ Screening & characterizing factors
  - ❑ Mathematical model for tuning
  - ❑ Improved both test environment (HDD & SSD)
  - ❑ Tuning only configuration parameters without violation in policy



# Thank You

