## WHATS REALY THE DEEERERENCE

Between a VM and a Container?



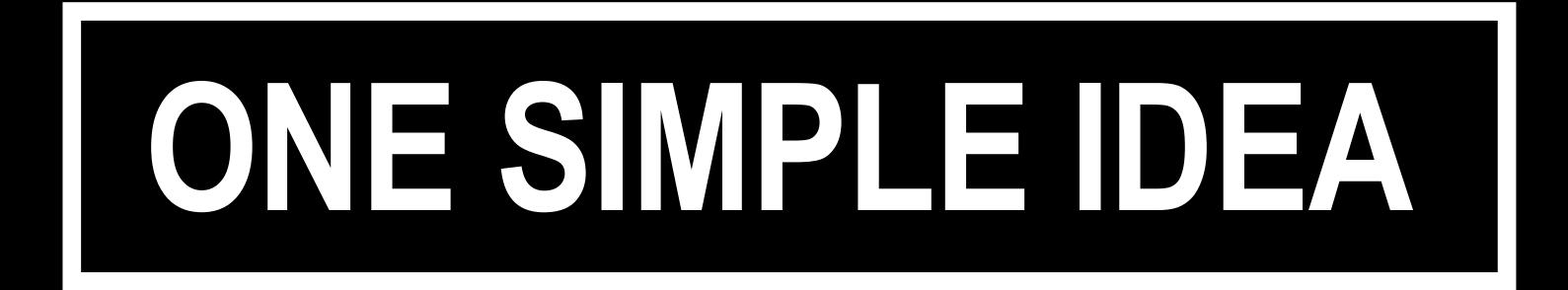












## CHANGED EVERYTHING.





# 1873

-









66

### WE COULDN'T IMPROVE THE PRODUCT SO WE IMPROVED THE TUBE.



- Colgate, 1908



## Colgate Research Center





## THE LAB ASSISTANT









# OMG, SALES DOUBLED!









### TOOTHPASTE TUBE THEORY

1) PRESSURE BUILT UP IN A FINITE BOUNDED SYSTEM WILL BREAK. 2) THERE ARE DIMINISHING RETURNS TO SQUEEZING THE TUBE AFTER A CERTAIN POINT.

## NEEDS TO BE RELEASED SOMEWHERE OR THE SYSTEM









# ADRIAN OTTO

Distinguished Architect, Rackspace Founder, OpenStack Containers Team Founder and PTL, OpenStack Magnum Organizer, Docker Los Angeles













## **PERFORMANCE**





### HISTORY OF VIRTUALIZATION

- 1960's IBM S/360 Mainframes are the 800# Gorilla
  - Single user system designed for batch jobs
- 1963 MIT Project MAC (\$2M grant from DARPA)
  - MAC = Multiple Access Computing: Multics
  - Vendor Choice == GE (Commercial interest in time sharing computer)
  - Whoops! IBM panicked! Created CP-40 for Bell Labs, CP-67.
    - Virtual Machines on the CP-67 using "CP (Control Program)" in 1967!
- 1987 Insignia Solutions "SoftPC"
- 1997 Apple (Connectrix) "VirtualPC"
- 1999 VMWare "VMWare Workstation"





### **APPLICATION VIRTUALIZATION**

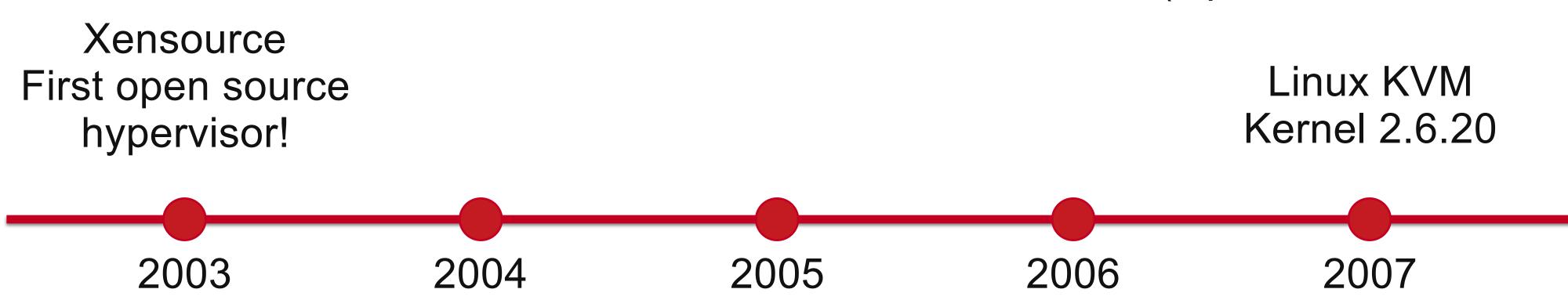
- 1990 Sun Microsystems "Stealth"
  - Address C/C++ Portability problems
  - Renamed Oak -> Webrunner -> Java (1995)
- 1996 Sun Microsystems "Java"
  - Java Development Kit (JDK)
  - Java Runtime Environment (JRE)
    - Java Virtual Machine (JVM)











### **Oracle VirtualBox OSE** (Open Source Edition)





## HISTORY OF CONTAINERS (1/2)

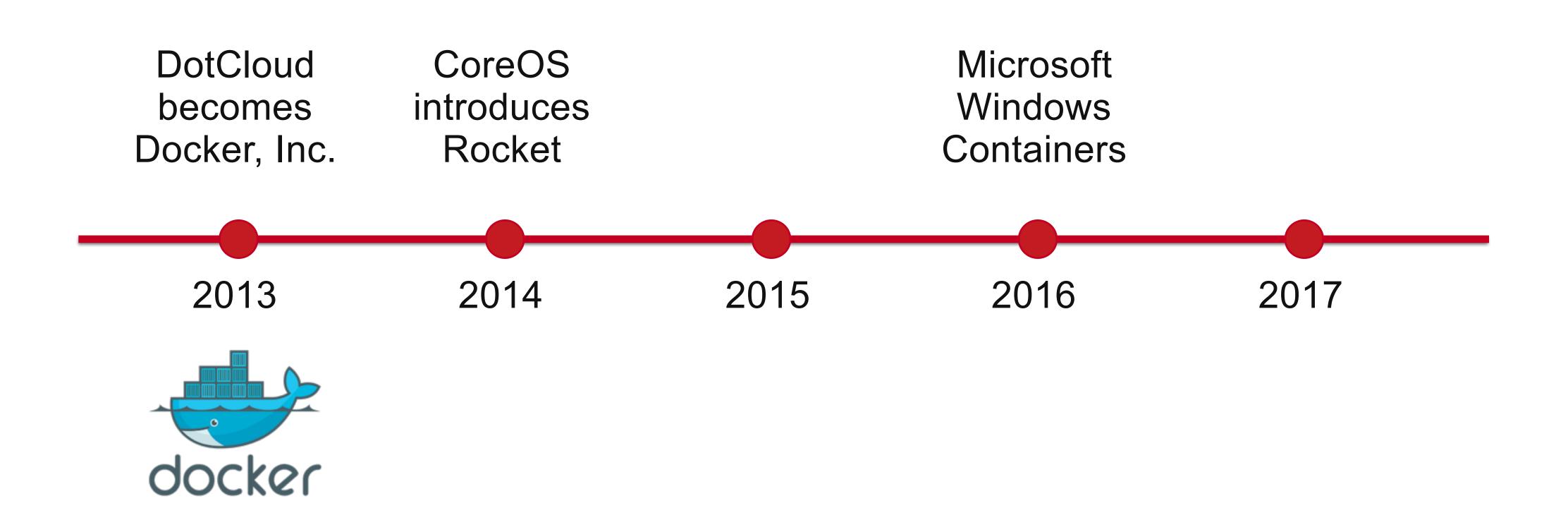
- 1979 UNIX chroot (added to BSD in 1982)
- 2000 FreeBSD Jails (filesystems, users, networks)
- 2001 Linux VServer (VPS Solution)
- 2005 OpenVZ (filesystems, users/groups, process tree, networks, devices, IPC) 2006 Process Containers (Linux Kernel 2.6.24, limit CPU, mem, disk, network IO) 2008 Control Groups (cgroups added to Linux Kernel)

- 2008 LXC (LinuX Containers, CLI and language bindings for 6 languages)
- 2011 Warden, CloudFoundry
- 2013 LMCTFY, Google





### HISTORY OF CONTAINERS (2/2)





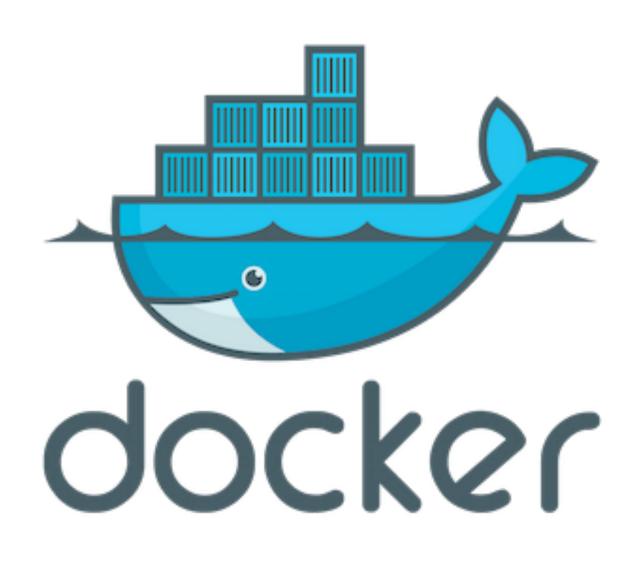




### **EVERYTHING CHANGED IN 2013**

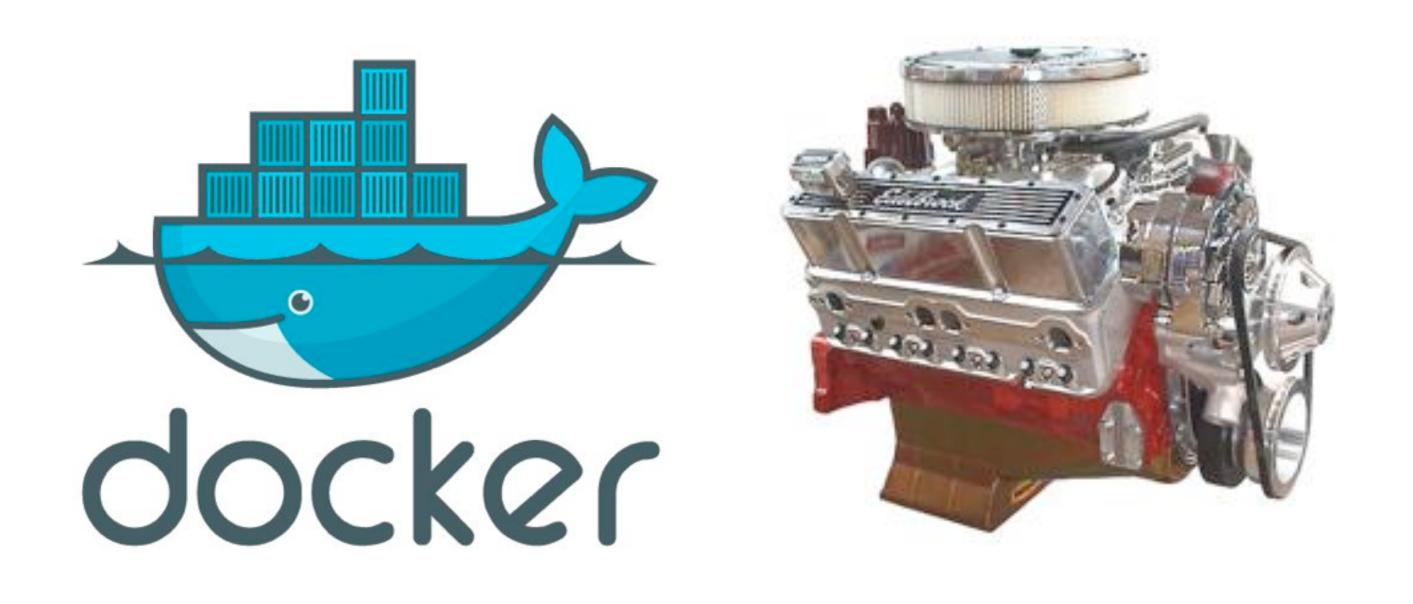
# 

## DOCKER IMAGE













### LINUX CGROUPS

- Kernel Feature
- Groups of processes
- Control resource allocations
  - CPU
  - Memory
  - Disk
  - I/O
- May be nested









### LINUX KERNEL NAMESPACES

- Kernel Feature
- Restrict your view of the system
- Mounts (CLONE\_NEWNS)
- UTS (CLONE\_NEWUTS)
  - uname() output
- IPC (CLONE\_NEWIPC)
- PID (CLONE\_NEWPID)
- Networks (CLONE\_NEWNET)
- User (CLONE\_NEWUSER)
  - See also: privileged/unprivileged modes
- May be nested







### **DOCKER CONTAINER IMAGE**

- NOT A FILESYSTEM
- NOT A VHD
- Basically a tar file
- Has a hierarchy
- Arbitrary depth
- Layered filesystem
  - Top layer can be writable
- Fits into the Docker Registry



### **Base Image**

### **Child Image**

### **Grandchild Image**





### **DOCKER REGISTRY**

- Git Repo Semantics
  - Pull
  - Push
  - Commit
- Hierarchy



### Base Image

### **Child Image**

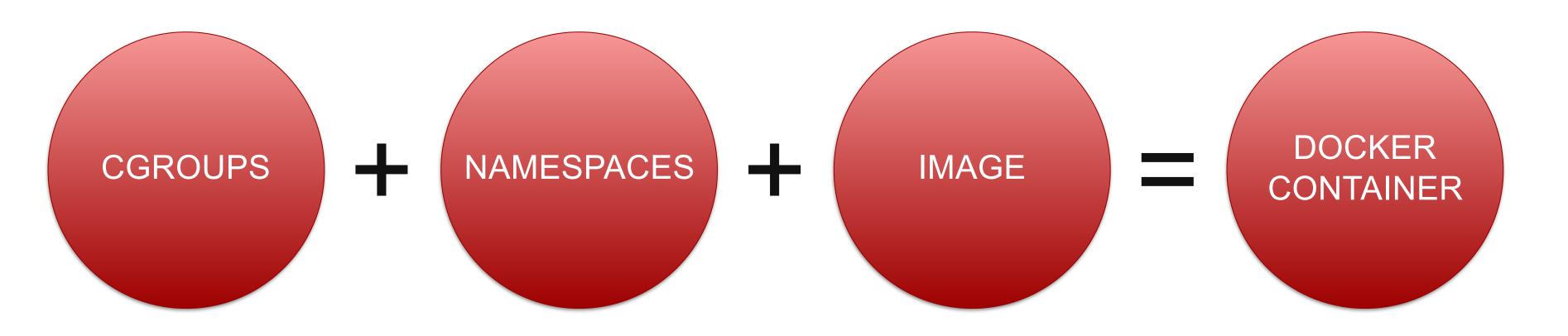
### **Grandchild Image**





### CONTAINER

- Combines several things
  - Linux Cgroups
  - Kernel Namespaces
  - Docker Image
  - Has a lifecycle

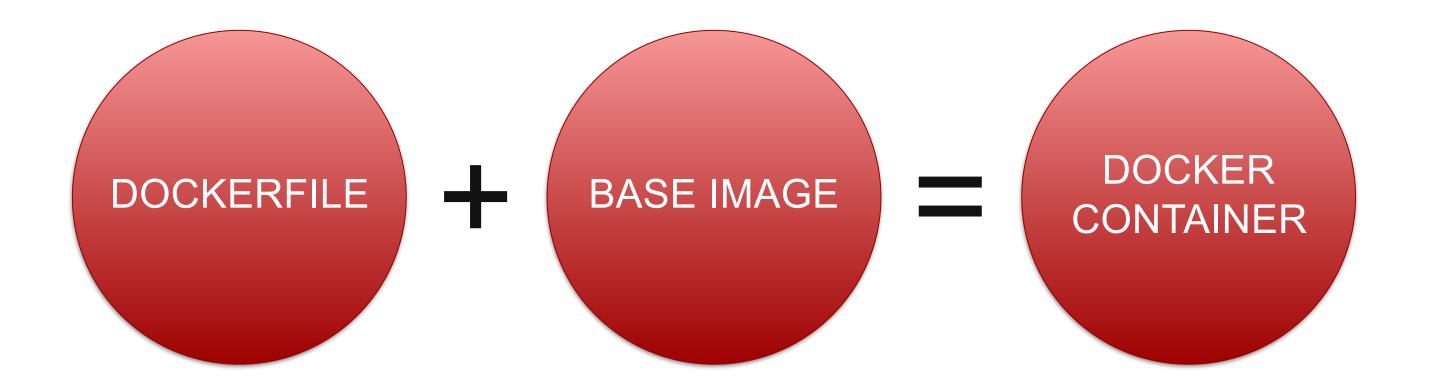






### DOCKERFILE

- Like a Makefile (shell script with keywords)
- Extends from a Base Image
- Results in a new Docker Image
- Imperative, not Declarative









### DOCKERFILE EXAMPLE

FROM centos:centos6 MAINTAINER Adrian Otto <aotto@aotto.com> RUN yum -y install httpd EXPOSE 80 ADD start.sh /start.sh CMD /start.sh



### docker build -t webserver





### DOCKERFILE EXAMPLE

FROM webserver

MAINTAINER Adrian Otto <aotto@aotto.com>

RUN yum -y install mysql-server php EXPOSE 80

ADD start.sh /start.sh

CMD /start.sh



### docker build -t lampstack .







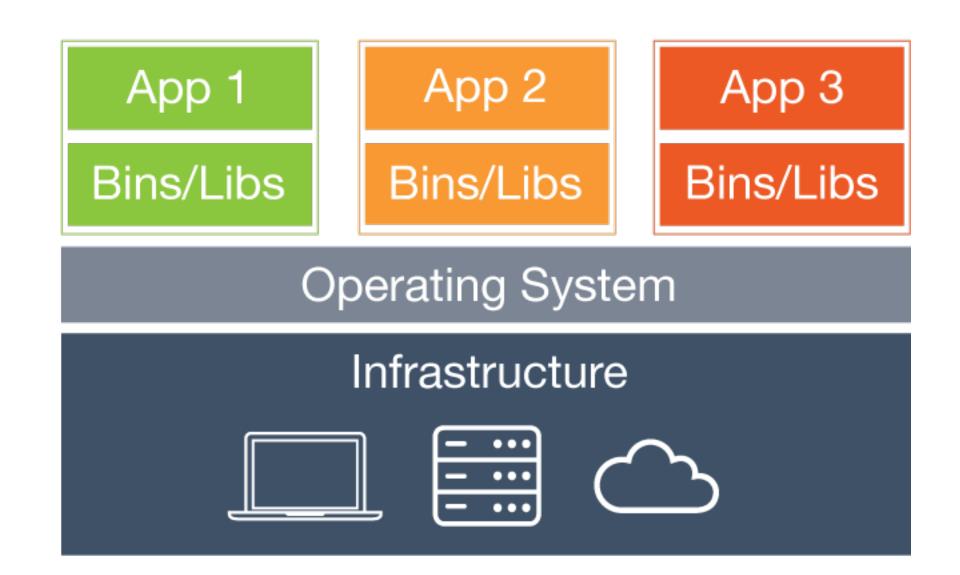


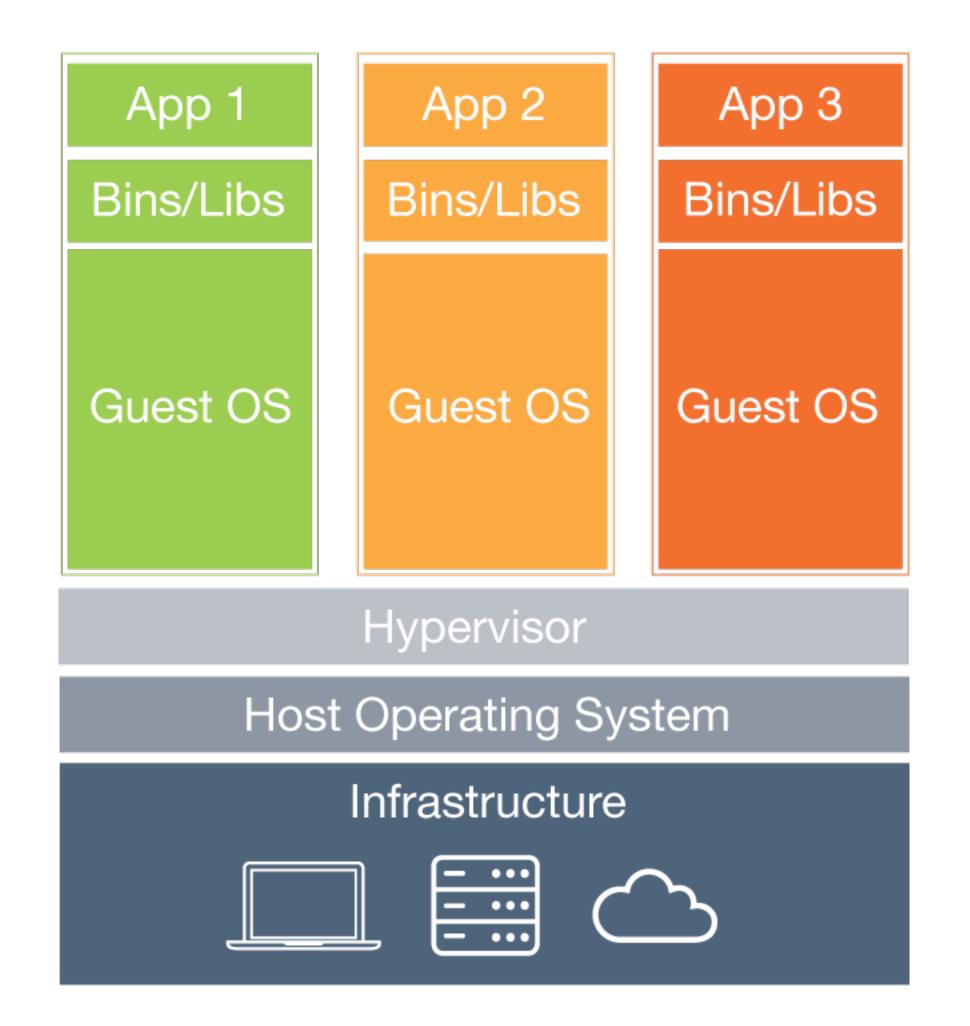
## **PERFORMANCE**







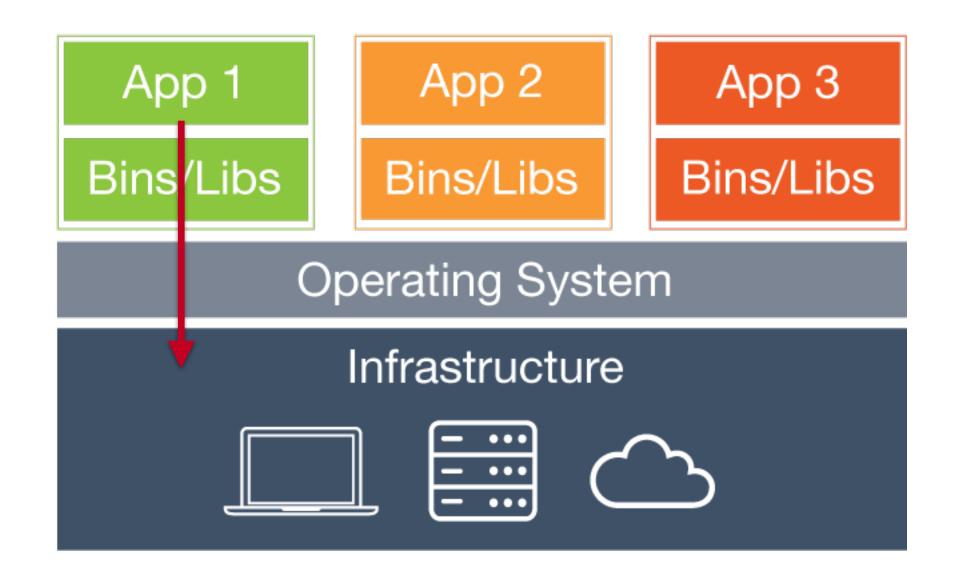


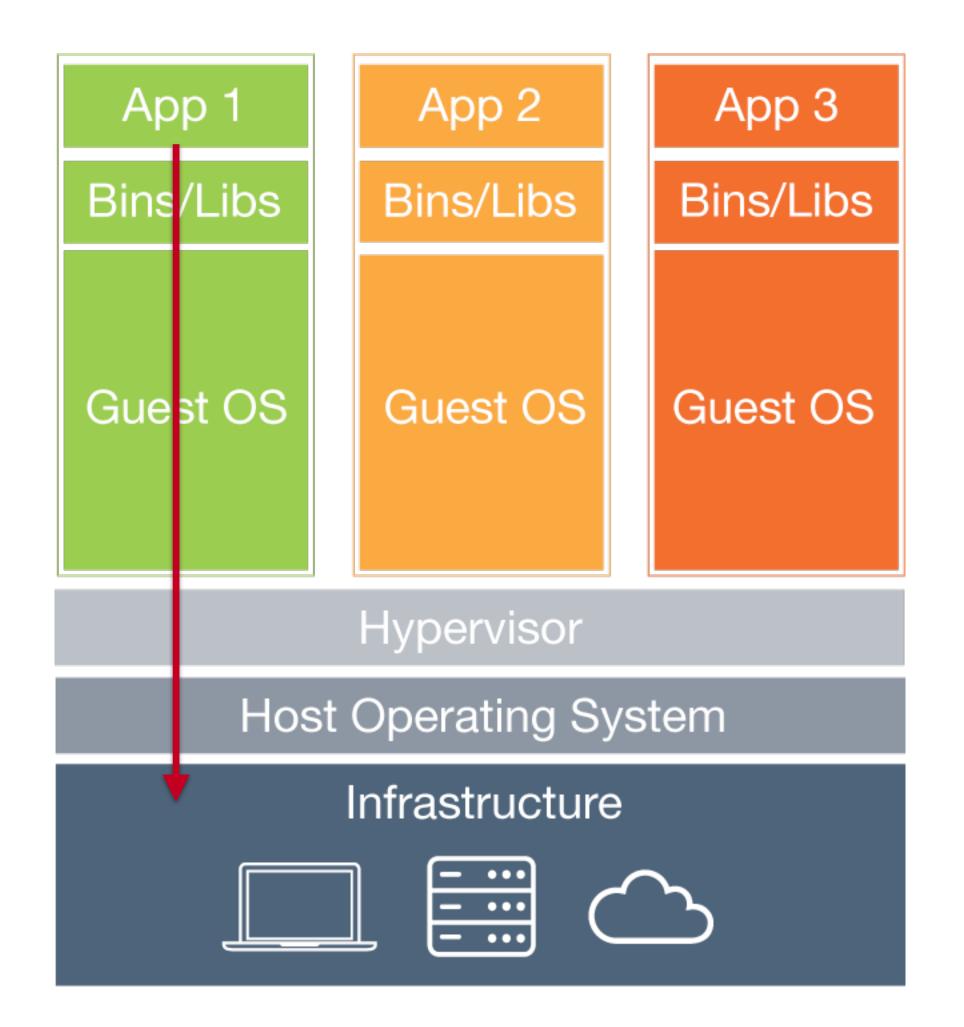








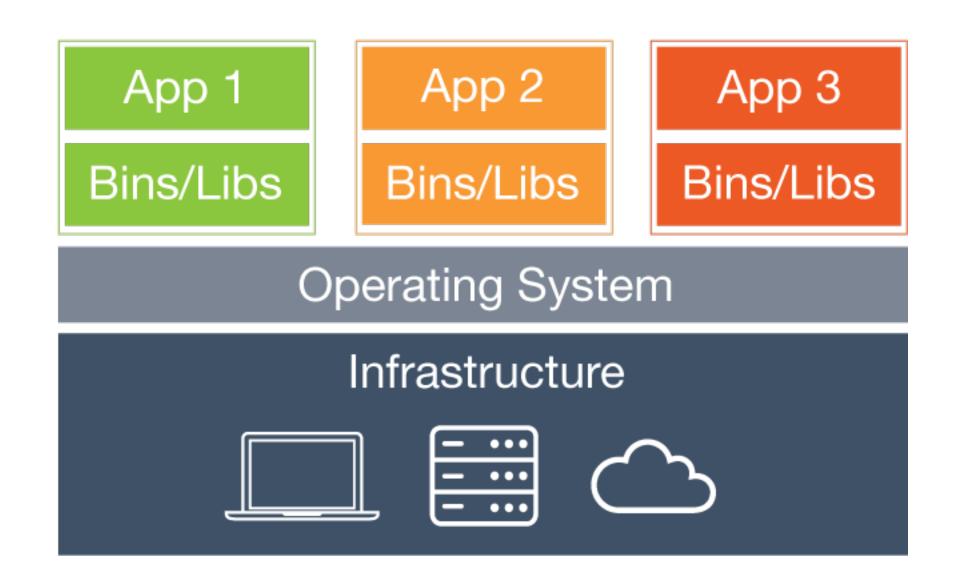


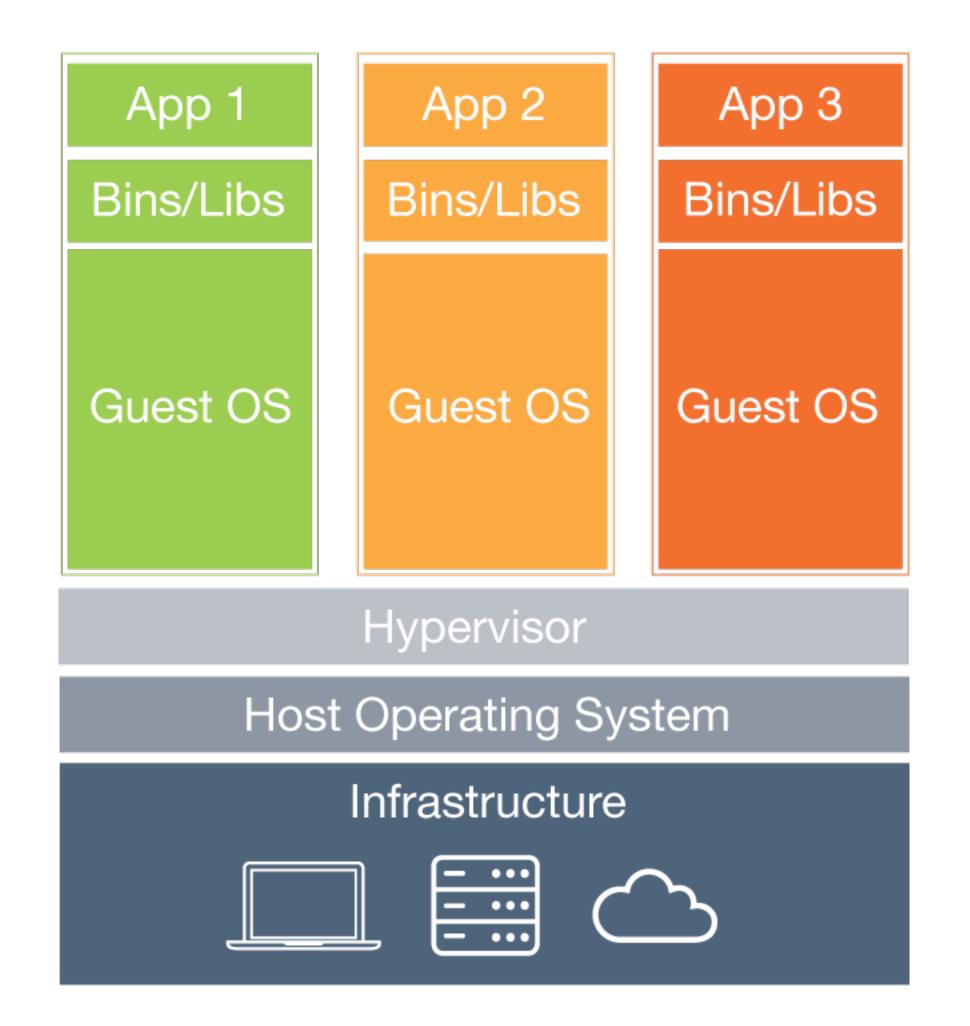


















## CASTILLO DE SAN MARCOS











### VIRTUALIZATION MAPPINGS

### Physical

System

Logical Processor

Advanced Programmable Interrupt Controller (APIC

Physical Address = System mPhysical Address (SP/

	Virtual		
	Partition		
	Virtual Processor		
C)	Virtual APIC + Synthetic Interrupt Controller (SynIC)		
PA)	Guest Physical Address (GPA)		





### LINUX SYSCALL INTERFACE

### Linux Syscall Reference

(Lines	A A entries				
					piero
. **	Name	0			ede 0
1.0	so, welat, avail	0x00			
1.1	101,003	0401	internet, and		
2	up, field	0402	struct pt_rep.*		
- 8	sys, mad	0428	unsigned in fill	the _are had	singlement.
1	tor, write	0x04	service and ine for	none dur _auer hul	sist Cont
5	and share	0x05	const thataver Historie	ine fage.	in main
6	ays, steam	0x05	unsigned in RE		
1	an, waitpid	0x07	HO H	100_000 1000_000	int-options
1	ton, small	0+04	mont that _aust hpathtate	int made	
	aya_link	0409	const charwat hochwise	cond that _war Provides	
18	sys_welleds	0x0a	const that _aset Spationers		
11	Sp. main	0.08	thatand *	the jack light	dar _ant *_ant
12	sys, shally	0404	tions that _ater Tionages		
13	101,079	0x04	trad part that		
14	up, should	Ox De	Norane	Unit models	unsigned dev
15	an dense	0.08	mont that _ater Maname	mode () mode	
18	sys, Mountal	0410	minute charaser	10,20,200	10,90,914
17	Appropriate and a second secon	0411			
18	10,00	0x13	that _ user Missigne	int_intent_inten	
19	tot, heek	0413	unsigned in to	of, tofter	unsigned introngen
20	ton, priorid	0414			-
21	sys, means	0471	that jugar New, rates	that juant Not, raite	14.704.004
30	sys, other own	0438	that _ and have		
-213	sys, which be	0x17	migrady and		
24	sys, general 26	0.08			
-25	sys, stime	0428	ting _ast for		
26	10,0000	the lat	tong required	long put	long-addr
30	tor, dam	0408	service and and		

### **397 CALLS IN KERNEL 3.19**

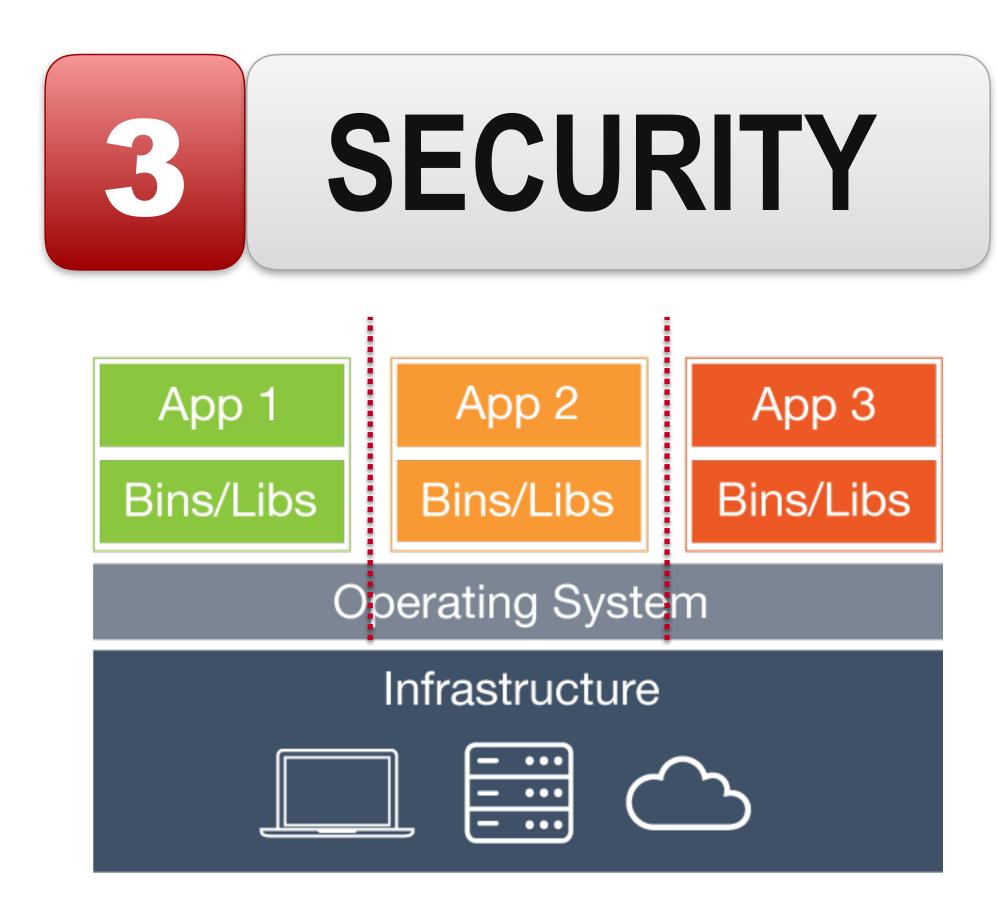
				handler	
43	tot, providité	0411			
14	un propidité	0.12			
11	ten, and	0411	mont thataset		
12	tot, atteast	0x34	thatyear have	une flaga	
10	Apt replemented	0471			
14	555,887	0.01	unspred on No	unsigned instand	anapted long ang
	sys_famil	0x37	unconed on RE	unsigned ine and	nucleur publicate
16	nan implemented	0438			
12	ton, settingid	0439	14,7,14	64,7,994	
54	ngt implemented	(balla			
18	tor, ciderate	0430	of delial_unstance exact*		
44	up, wheek	ilula:	int mask		
61	tor, direct	0x3d	Hanama		
8	ton, wrist	(a)a	smighted dev	struct workdworkt "workd	
	Seath and	0.2	succession of the state	unsigned int needs	
54	tor, provid	048			
45	10,0000	0+41			
17	sys, secold sys, spectral	0+42	-	-	-
	and a general			etd, significes	oid, significan
14	ton, spetrack	Dv84			
10	ton, sontheast	Over 1	Int Internation		
79	sys, antread 14	0+41	the product of	10,25,162	
71	sus, servegid 18	De47	integrate spat	10,94,1494	
72	sys, signational	Ov48	Les KaleryD	ine haenryd.	shipspot, track
79	ton, signaturing	0.49	No.		
24	sys, settentisa re	Orda	thatwat hatte	une la m	
	up, sections	Code Code	uncipital int maturity uncipital int	enventionis_avan frites	
77	to proces	Code .	resturie Introductor	enver römis _aaar fröm	
	up, get marklay	Code:	struct timesal	unan fra	
79	un, artifrendiker	0.4	_aser for	_and traces	
	to any taxa	0/10	and To		
-	ton, artigroups 24	0/11	111-240-01-24	Speaker .	
	ton, old, select	D-11	struct	"program	
			end		
	un, symiak	0413	Hand Charaskit	hond thatauer	
-	and hear	0+54	thatyaar Historia	unut _shijternijstet _aser*toital	
**	sys_matters	0411	rond that _alar Salt	04_44.54	re baha
**	tor, and th	0/18	minut charaset Tribuny		
	spr. saupon	0417	nond charaver hoscolitie	re seat, fags	
	sys, wheel	0/18	int mapping	ve regist	analised on cond
**	sys, stil, wald?	0.01	stragend re-	engli milliman, direct	anoped of

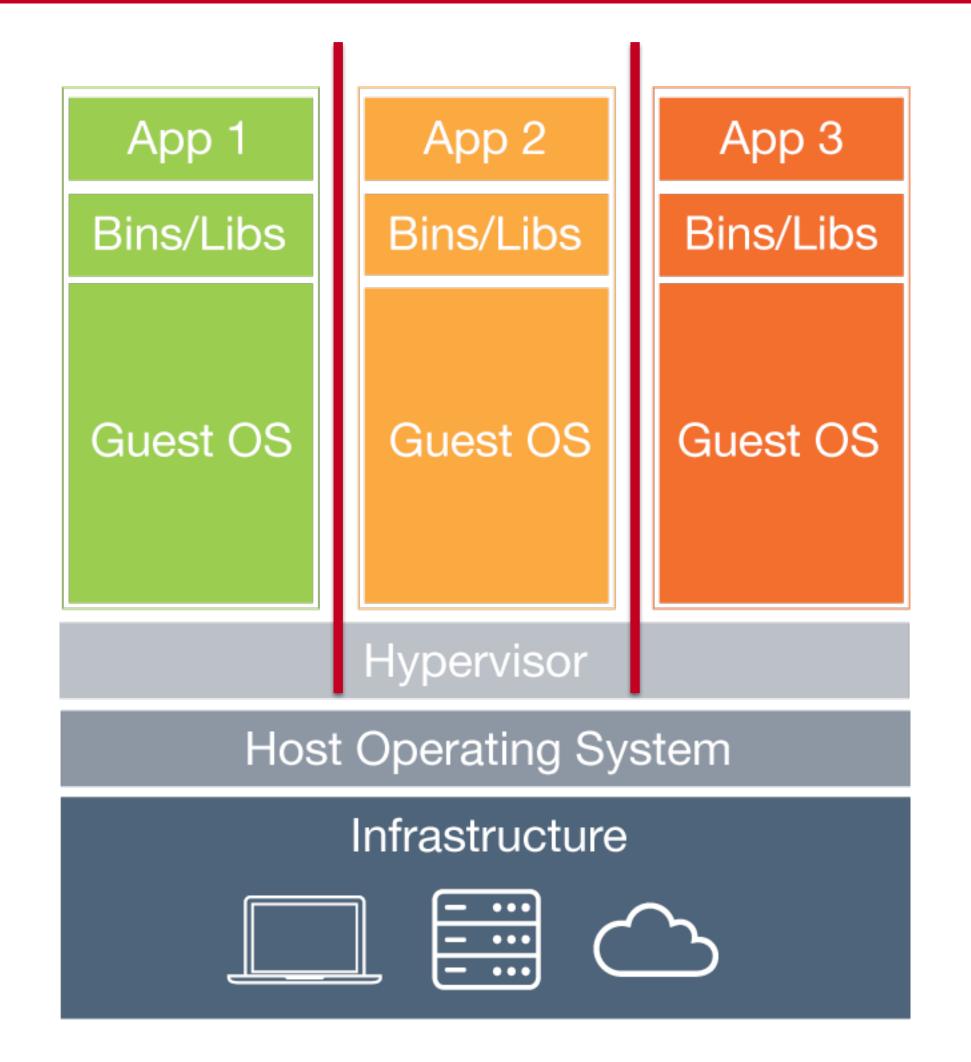
			Bearch:
			Definition 0
			terrel operate 2018
			Rentell works (2046)
			anth alpha harnal anna 576
			followed writere 2011
			fulned_action (#18
			huispeach00
			Relapenarité la
			karnal weber (275)
			hcisperaction
			Reinameter 2120
			Reinameta 3162
•	anan jajanga "		anth/alpha/barnet/entry/8121
			Purpersa 241
			ternel/posts-timen_c815
			hy/namela/2062
			Ecoperation 2007
			terrel/with 27
			Revolution 110
•			Purrent period (67
			karnel/timet.c1337
	unsigned tong Taga	vent_aver have	Pulmartengian at 2118
			Paymentersport of 3171
			terrel/ulditure?
			kernel/width.c312
			karnal times ili
	long data		karnal/procectill
			kernel/timer.cl/life

			karnal/sid/dic/212
			karnet/uid06.c227
			kernet webuilt?
			and and a re-
			Reflection parts of 111 M
-			Purinetta (613
-			Automatics ADD
			karnal (space 523)
			Antimic report (11)
			karnel/sport/MER
			Respondence in the
			hcisseluci25
			Notice 105
			karnel timete 1948
			karnetingsuc 2020
			Antonio (1995)
			and max herein special 200
			karnal regeate (6/57
			kernet regester 2643
			kernel uddituriti
			terrel viditure
			and cliff hereit signals of
			kernel signals 2002
			karnel/spi.e1245
			karnel/spice/3275
			karnal/spice1216
			karmal/syste0412
			karnel Kinesi 201
		-	Annual Street 281
			terminulation por
			terretry/discus?
	-	-	
			Purseinerse 201
			Nonemala 2408
			Bulance 142
			Reparate 201
			Representation of the second s
	-		non/samp@in.c.1790
	101.00 101		kernel/spud 368
			Reconcepted and
			the second second













### **CONTAINTER ISOLATION TECHNIQUES**

- SELinux / App.
- Secure Compt
- Container Nes
- Docker Auth P
- User Namespa
- Encrypted File
- Address Space
- Hardware Sec











## **PERFORMANCE**





# 

### ONE FANATICAL PLACE | SAN ANTONIO, TX 78218 US SALES: 1-800-961-2888 | US SUPPORT: 1-800-961-4454 | WWW.RACKSPACE.COM

Copyright © 2016 Rackspace | Rackspace® Fanatical Support® and other Rackspace marks are either registered service marks or service marks of Rackspace US, Inc. in the United States and other countries. Features, benefits and pricing presented depend on system configuration and are subject to change without notice. Rackspace disclaims any representation, warranty or other legal commitment regarding its services except for those expressly stated in a Rackspace services agreement. All other trademarks, service marks, images, products and brands remain the sole property of their respective holders and do not imply endorsement or sponsorship.







