



86%

62%



EmergiNet

EXTRACTING VALUE FROM DATA



(OPEN-SOURCE) BIG-DATA ANALYTICS IN THE CLOUD

O que são BIG-DATA ANALYTICS, SEARCH, CLOUD, e quem precisa disso?



TURNING THIS

TURNING THIS										
350348467368	A28	GetFolderContents	comcast.comMENU000000000408993	13	51297	000015E36F22	153	09517161197		
1350348467363		SetResumePoint	espn.comTITLE00000000089496	28	0	0000036B4BC9	106	0951942958201	Bal	
1350348467389	A28	GetFolderContents	comcast.comMENU000000000180296	10	297	000013DA3934	12602	09558405259		
1350348467400	A28	GetFolderContents	comcast.comMENU000000000360757	8	4585	000015D3736E	12601	0610223		
1350348467403	A28	GetFolderContents	comcast.comMENU000000000420458	10	1369	00001086A40E	135	87734002334		
1350348467405		AddSavedProgram	tlc.comtitl1620566790150401	14	0	000041941578	167	8993212500037380	Pit	
1350348467413	A28	GetRootContents	10 1570 000003E8948D	163	0581810710102	HLLY	17954	xyz.xyz.xyz		
1350348467420		SetResumePoint	sho.comXPTL0001339676908000	11	0	0000036FDEC2	108	0562717077602	Baltimo	
1350348467429	A25	GetAssociatedFolderContents	SourceId: 14015 6	635	001AC33C1285	136	8773402201534092	Har		
1350348467425	A28	GetFolderContents	comcast.comMENU000000000354872	16	1337	000000ACE4CC	108	05627135848		
1350348467439	A28	GetFolderContents	comcast.comMENU00000000006259	10	1079	000002C2B2F8	108	09546622994		
1350348467447	A28	GetFolderContents	comcast.comMENU000000000165113	8	3075	000013B23A44	127	09589152424		
1350348467455	A28	GetSavedProgramsFolderContents	6 0	21305		xyz.xyz.xyz.123 a_host.com				
1350348467445		AddSavedProgram	cbs.comXPTL0001349938121523	24	0	0000037F4802	111	8773102930108966	Der	
1350348467464	A28	GetFolderContents	comcast.comMENU000000000561743	11	844	00003844A44	14801	0570516		
1350348467472	A28	GetFolderContents	comcast.comMENU0000000005548	10	100	000003834D7	151	09530124005		
1350348467479	A28	GetFolderContents	comcast.comMENU0000000001733	11	843	000038EA02E	133	09549304329		
1350348467485	A28	GetFolderContents	comcast.comMENU0000000001977	15	2471	000011AA95F5	127	01720145753		
1350348467477	A28	GetFolderContents	comcast.comMENU000000000561731	30	2199	000041F0910D	108	05627444980		
1350348467500		ValidatePlayEligibility	comcast.comCCDN1000000003315686	11	100	000001487333	154	09540123872		
1350348467490		SetResumePoint	comedycentral.comXPTL000000000144294	26	0	00004115CC38	114	87731031713		
1350348467512	A28	GetFolderContents	comcast.comMENU00000000001977	15	2456	000015B0961F	13703	0954411		
1350348467505		ValidatePlayEligibility	nbc.comNBCU2012100400000450	28	96	00004178FFA0	187	0956955555903		
1350348467497	A28	GetFolderContents	comcast.comMENU000000000001358	45	2274	000013AB62A3	167	89932126700		
1350348467525		AddSavedProgram	nbc.comNBCU2012100400000670	21	0	001AC326D812	158	2000110997508	Philade	
1350348467511	A28	GetFolderContents	comcast.comMENU000000000408783	38	506	000041393CDE	21305	15557793984		
1350348467549	A28	GetFolderContents	comcast.comMENU000000000177582	8	2364	0000014BDBE4	127	09520466971		
1350348467557	A28	GetFolderContents	comcast.comMENU000000000583113	7	1106	0000000D5A7F	112	87731021901		
1350348467517	A28	GetFolderContents	comcast.comMENU000000000001358	53	2297	000001091A32	127	01721146670		
1350348467554		SetResumePoint	tlc.comtitl1620567970090401	26	0	00000077DA26	190	0952939784303	Washing	

BIG DATA

Big Data is data that is too large, complex and dynamic for any conventional data tools to capture, store, manage and analyze.

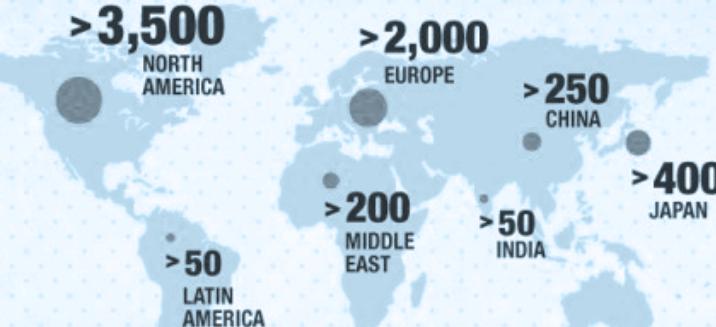
The right use of Big Data allows analysts to spot trends and gives niche insights that help create value and innovation much faster than conventional methods.

57.6% OF ORGANIZATIONS SURVEYED SAY THAT BIG DATA IS A CHALLENGE

72.7% CONSIDER DRIVING OPERATIONAL EFFICIENCIES TO BE THE BIGGEST BENEFIT OF A BIG DATA STRATEGY

50% SAY THAT BIG DATA HELPS IN BETTER MEETING CONSUMER DEMAND AND FACILITATING GROWTH

VOLUME



Amount of Big Data stored across the world (in petabytes)

\$300 billion is the potential annual value to Healthcare

TRANSPARENCY IN CLINICAL DATA AND CLINICAL DECISION SUPPORT

AGGREGATION OF PATIENT RECORDS, ONLINE PLATFORMS AND COMMUNITIES

VARIETY



VELOCITY



PUBLIC HEALTH SURVEILLANCE AND RESPONSE SYSTEMS

ADVANCED FRAUD DETECTION; PERFORMANCE BASED DRUG PRICING

\$47B ACCOUNTS

RESEARCH AND DEVELOPMENT; PERSONALIZED MEDICINE; CLINICAL TRIAL DESIGN

INTO THESE

VALUE



40% PROJECTED GROWTH IN GLOBAL DATA CREATED PER YEAR



5% PROJECTED GROWTH IN GLOBAL IT SPENDING PER YEAR

The estimated size of the digital universe in 2011 was 1.8 zettabytes. It is predicted that between 2009 and 2020, this will grow 44-fold to 35 zettabytes per year. A well defined data management strategy is essential to successfully utilize Big Data.

Sources - ① Reaping the Rewards of Big Data - Wipro Report ② Big Data: The Next Frontier for Innovation, Competition and Productivity - McKinsey Global Institute Report ③ comScore, Radicati Group ④ Measuring the Business Impacts of Effective Data - study by University of Texas, Austin ⑤ US Department of Labour.

DO BUSINESS BETTER



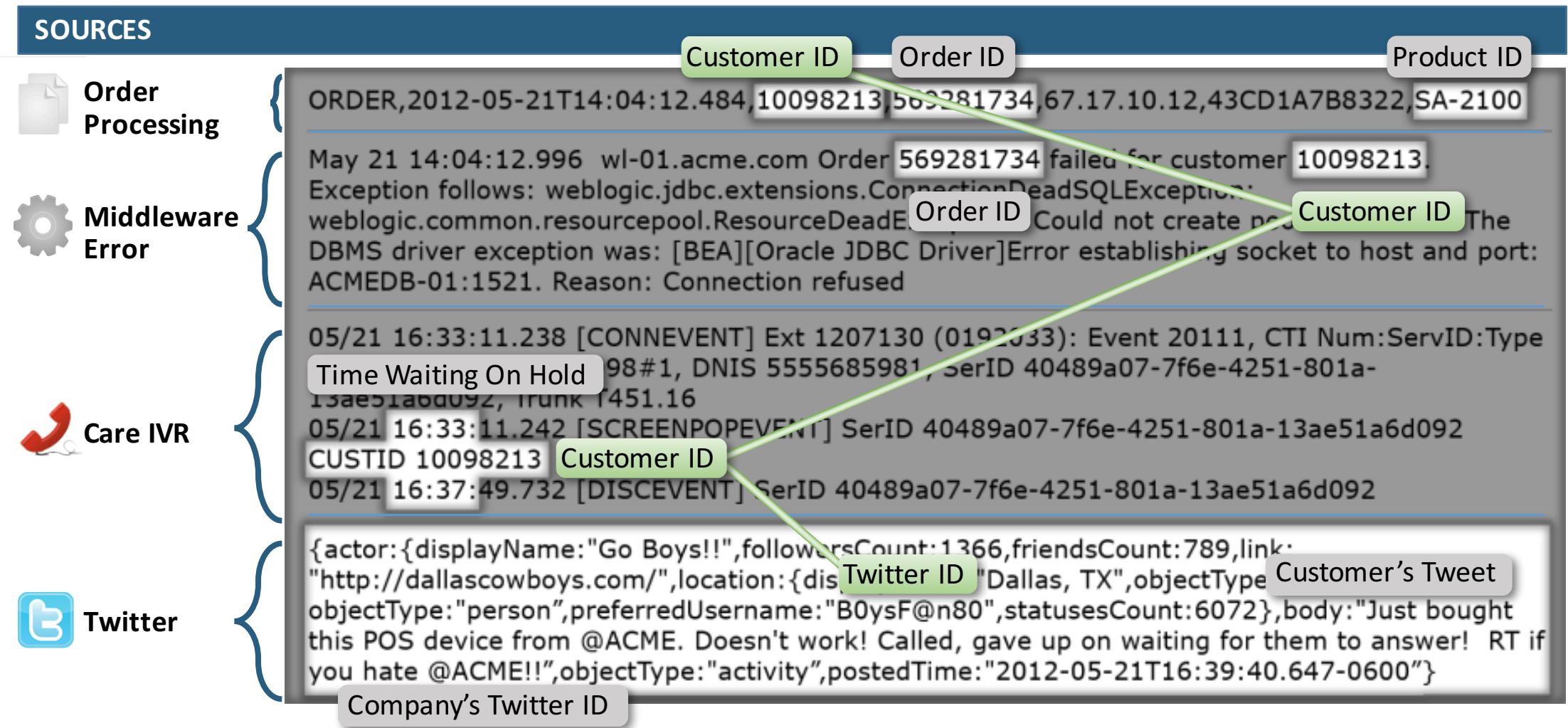
Dados massivos e que já não cabem numa planilha Excel (VVV)

SOURCES

	Order Processing	{ ORDER,2012-05-21T14:04:12.484,10098213,569281734,67.17.10.12,43CD1A7B8322,SA-2100
	Middleware Error	{ May 21 14:04:12.996 wl-01.acme.com Order 569281734 failed for customer 10098213. Exception follows: weblogic.jdbc.extensions.ConnectionDeadSQLException: weblogic.common.resourcepool.ResourceDeadException: Could not create pool connection. The DBMS driver exception was: [BEA][Oracle JDBC Driver]Error establishing socket to host and port: ACMEDB-01:1521. Reason: Connection refused
	Care IVR	{ 05/21 16:33:11.238 [CONNEVENT] Ext 1207130 (0192033): Event 20111, CTI Num:ServID:Type 0:19:9, App 0, ANI T7998#1, DNIS 5555685981, SerID 40489a07-7f6e-4251-801a- 13ae51a6d092, Trunk T451.16 05/21 16:33:11.242 [SCREENPOPEVENT] SerID 40489a07-7f6e-4251-801a-13ae51a6d092 CUSTID 10098213 05/21 16:37:49.732 [DISCEVENT] SerID 40489a07-7f6e-4251-801a-13ae51a6d092
	Twitter	{actor:{displayName:"Go Boys!!",followersCount:1366,friendsCount:789,link: "http://dallascowboys.com/",location:{displayName:"Dallas, TX",objectType:"place"}, objectType:"person",preferredUsername:"B0ysF@n80",statusesCount:6072},body:"Just bought this POS device from @ACME. Doesn't work! Called, gave up on waiting for them to answer! RT if you hate @ACME!!",objectType:"activity",postedTime:"2012-05-21T16:39:40.647-0600"}



Extraindo valor dos dados brutos (filtros, correlações)



WARN
99786[Hide Log List](#) [Clear Display](#) [Reload](#)[Ignore Sender](#) [Insert Marker](#) [Inspector](#)[Filter](#)**SYSTEM LOG QUERIES**[All Messages](#)**DIAGNOSTIC AND USAGE INFORMATION**[Diagnostic and Usage Messages](#)► [User Diagnostic Reports](#)► [System Diagnostic Reports](#)**FILES**[system.log](#)► [~/Library/Logs](#)► [/Library/Logs](#)► [/var/log](#)

Privacidade: um enorme desafio!

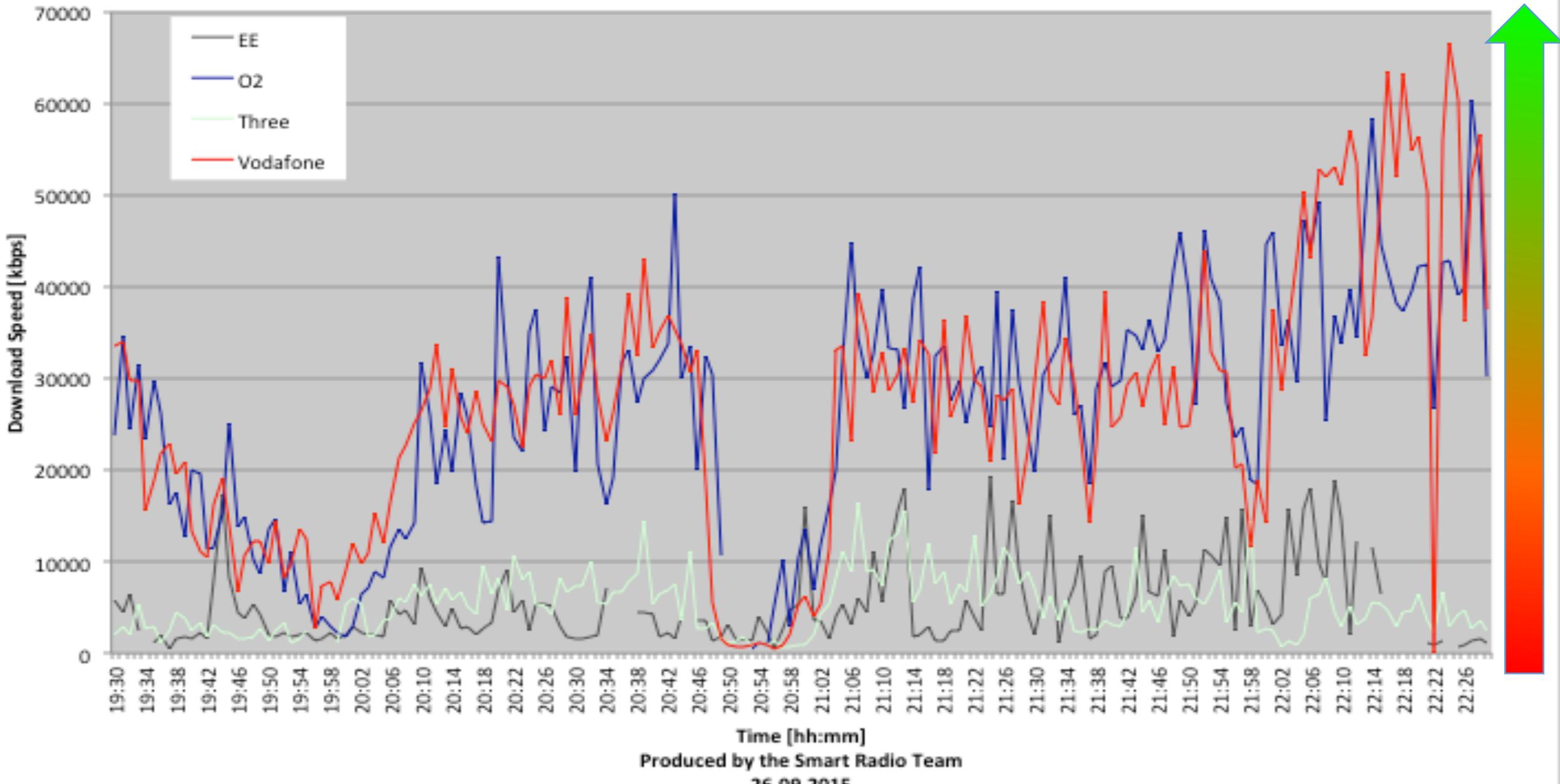
```
Mon Jul 18 08:19:59.281 <kernel> IO80211ScanManager::startScan: Broadcast scan request received from 'airportd' (pid 61) ().
Mon Jul 18 08:19:59.281 <kernel> IO80211ScanManager::startScan: Initiating scan.
Mon Jul 18 08:19:59.343 <kernel> IO80211ScanManager::getScanResult: All scan results returned for 'airportd' (pid 61).
Mon Jul 18 08:19:59.343 AutoJoin: <airportd[61]> Successful cache-assisted auto-join scan request on channel 9 took 0.0628 seconds, returned 0 results
Mon Jul 18 08:19:59.345 <kernel> IO80211ScanManager::startScan: Broadcast scan request received from 'airportd' (pid 61) ().
Mon Jul 18 08:19:59.345 <kernel> IO80211ScanManager::startScan: Initiating scan.
Mon Jul 18 08:19:59.409 <kernel> IO80211ScanManager::getScanResult: All scan results returned for 'airportd' (pid 61).
Mon Jul 18 08:19:59.409 AutoJoin: <airportd[61]> Successful cache-assisted auto-join scan request on channel 10 took 0.0650 seconds, returned 0 results
Mon Jul 18 08:19:59.411 <kernel> IO80211ScanManager::startScan: Broadcast scan request received from 'airportd' (pid 61) ().
Mon Jul 18 08:19:59.411 <kernel> IO80211ScanManager::startScan: Initiating scan.
Mon Jul 18 08:19:59.542 <kernel> IO80211ScanManager::getScanResult: All scan results returned for 'airportd' (pid 61).
Mon Jul 18 08:19:59.543 AutoJoin: <airportd[61]> Successful cache-assisted auto-join scan request on channel 12 took 0.1330 seconds, returned 0 results
Mon Jul 18 08:19:59.545 <kernel> IO80211ScanManager::startScan: Broadcast scan request received from 'airportd' (pid 61) ().
Mon Jul 18 08:19:59.545 <kernel> IO80211ScanManager::startScan: Initiating scan.
Mon Jul 18 08:19:59.677 <kernel> IO80211ScanManager::getScanResult: All scan results returned for 'airportd' (pid 61).
Mon Jul 18 08:19:59.677 AutoJoin: <airportd[61]> Successful cache-assisted auto-join scan request on channel 13 took 0.1331 seconds, returned 0 results
Mon Jul 18 08:19:59.679 <kernel> IO80211ScanManager::startScan: Broadcast scan request received from 'airportd' (pid 61) ().
Mon Jul 18 08:19:59.679 <kernel> IO80211ScanManager::startScan: Initiating scan.
Mon Jul 18 08:19:59.808 <kernel> IO80211ScanManager::getScanResult: All scan results returned for 'airportd' (pid 61).
Mon Jul 18 08:19:59.808 AutoJoin: <airportd[61]> Successful cache-assisted auto-join scan request on channel 52 took 0.1305 seconds, returned 0 results
Mon Jul 18 08:19:59.810 <kernel> IO80211ScanManager::startScan: Broadcast scan request received from 'airportd' (pid 61) ().
Mon Jul 18 08:19:59.911 <kernel> IO80211ScanManager::startScan: Initiating scan.
Mon Jul 18 08:19:59.911 <kernel> IO80211ScanManager::startScan: AutoJoin: <airportd[61]> _event: 0x4d6572637572652d57694669> SCAN_CACED_JOIN (e)
Mon Jul 18 08:19:59.911 <kernel> IO80211ScanManager::startScan: AutoJoin: <airportd[61]> _event: 0x4d6572637572652d57694669> Scan result returned for 'airportd' (pid 61) ().
Mon Jul 18 08:19:59.941 AutoJoin: <airportd[61]> Successful cache-assisted auto-join scan request on channel 60 took 0.1318 seconds, returned 1 results
Mon Jul 18 08:19:59.941 AutoJoin: <airportd[61]> __matchProfile: Attempting to match network ['amazonas-w' - 00:1f:7c:00:f5:57]
Mon Jul 18 08:19:59.942 AutoJoin: <airportd[61]> __matchProfile: [wifi.ssid.<4f322057 696669>] SSID '02 Wifi' (<CFData 0x7fd433621aa0
[0x7fff7dd06440]>{length = 7, capacity = 7, bytes = 0x4f322057696669}) does not match network SSID 'amazonas-w' (<CFData 0x7fd4359a7ce0
[0x7fff7dd06440]>{length = 10, capacity = 10, bytes = 0x616d617a6f6e61732d77}), next
Mon Jul 18 08:19:59.942 AutoJoin: <airportd[61]> __matchProfile: [wifi.ssid.<556e696f 6e496e6e>] SSID 'UnionInn' (<CFData 0x7fd435b7c030
[0x7fff7dd06440]>{length = 8, capacity = 8, bytes = 0x556e696f6e496e6e}) does not match network SSID 'amazonas-w' (<CFData 0x7fd4359a7ce0
[0x7fff7dd06440]>{length = 10, capacity = 10, bytes = 0x616d617a6f6e61732d77}), next
Mon Jul 18 08:19:59.942 AutoJoin: <airportd[61]> __matchProfile: [wifi.ssid.<534b5942 41353844>] SSID 'SKYBA58D' (<CFData 0x7fd435a376a0
[0x7fff7dd06440]>{length = 8, capacity = 8, bytes = 0x534b594241353844}) does not match network SSID 'amazonas-w' (<CFData 0x7fd4359a7ce0
[0x7fff7dd06440]>{length = 10, capacity = 10, bytes = 0x616d617a6f6e61732d77}), next
Mon Jul 18 08:19:59.942 AutoJoin: <airportd[61]> __matchProfile: [wifi.ssid.<4d657263 7572652d 57694669>] SSID 'Mercure-WiFi' (<CFData 0x7fd435a36bc0
[0x7fff7dd06440]>{length = 12, capacity = 12, bytes = 0x4d6572637572652d57694669}) does not match network SSID 'amazonas-w' (<CFData 0x7fd4359a7ce0
[0x7fff7dd06440]>{length = 10, capacity = 10, bytes = 0x616d617a6f6e61732d77}), next
Mon Jul 18 08:19:59.942 AutoJoin: <airportd[61]> __matchProfile: [wifi.ssid.<57694669 56696c6c 61726561 6c>] SSID 'WiFiVillareal' (<CFData 0x7fd435b1b910
[0x7fff7dd06440]>{length = 13, capacity = 13, bytes = 0x5769466956696c6c617265616c}) does not match network SSID 'amazonas-w' (<CFData 0x7fd4359a7ce0
[0x7fff7dd06440]>{length = 10, capacity = 10, bytes = 0x616d617a6f6e61732d77}), next
Mon Jul 18 08:19:59.942 AutoJoin: <airportd[61]> __matchProfile: [wifi.ssid.<54686561 74726520 526f7961 6c>] SSID 'Theatre Royal' (<CFData 0x7fd435a37730
[0x7fff7dd06440]>{length = 13, capacity = 13, bytes = 0x5468656174726520526f79616c}) does not match network SSID 'amazonas-w' (<CFData 0x7fd4359a7ce0
[0x7fff7dd06440]>{length = 10, capacity = 10, bytes = 0x616d617a6f6e61732d77}), next
Mon Jul 18 08:19:59.942 Info: <Wi-Fi Menu Extra[257]> scan cache updated
Mon Jul 18 08:19:59.942 AutoJoin: <airportd[61]> __matchProfile: [wifi.ssid.<4849204d 4144204c 4153202d 5441424c 4153>] SSID 'HI MAD LAS -TABLAS' (<CFData
0x7fd435a15f50 [0x7fff7dd06440]>{length = 18, capacity = 18, bytes = 0x4849204d4144204c4153202d5441424c4153}) does not match network SSID 'amazonas-
w' (<CFData 0x7fd4359a7ce0 [0x7fff7dd06440]>{length = 10, capacity = 10, bytes = 0x616d617a6f6e61732d77}), next
```



Telecom Quality of Experience Olympics 2012-2016



England vs. Wales - Rugby World Cup - Download Performance



Revisão conceitual e Desafios

- Definição e utilidade: BIG-DATA ANALYTICS, SEARCH, CLOUD (Public, Private, On-Premises).
- Estado da arte: banda, armazenamento, processamento, computação distribuída (clusters);
- Aplicações em Infraestrutura e Telecom (call-center, eficiência energética, qualidade, geo-localização, otimização de redes e serviços, ofertas direcionadas, promoções, etc.)
- Privacidade, Escalabilidade, Segurança e Eficiência;
- O Brasil precisa **formar e desenvolver profissionais** com essas **Competências Digitais** (emerginet.org)





EmergiNet

EXTRACTING VALUE FROM DATA

gabriel.bonilha@emergi.net
www.emergi.net