



Big Data, Analytics and the Mainframe

Paul Pendle EMC Corporation

March 15, 2012 7:00AM-8:00AM Session Number: 3278



Big Data – What is it?

"Data not easily handled by the traditional data processing mechanisms of today"

Characterized by:

- High volume and arrival rate
 - Billions of data events per hour
- Semi-structured or unstructured
 - Not easily analyzed or parsed
- Wide disparity
 - Text, video, graphics, sensor data, RFID, web, etc

"The three Vs"





Mainframe Data Feeds

					1																									8	3()	1.150	b	y	/1	te	Э	S																												
Г		-					-						22			-	10	22		ž.		i.													-							5.	14		*	ic :			2		-						2					5					٦
-	8	0	0	0	10	0	0 0	0	0	0 1	0 0	0	0	0	0 1	0 1	0 0	0	0	0 1	0 0	10	0 0	0	0	0	0	6 1	1	1		0	0	1		1	1 0	0	8	1	8	11	0	0	Ó	ġ I	1	1	ġ.	9 (1 8	ú	Ó	Ó (10	đ	Ó	ġ	ð Ó	Ó	ů	8 1	10	0	Ò	0 1	Ô
1	2	1	A	5	5	7.	0.9	1 12	11;	12 1	13 14	1 11	15	17	(8)	8.1	9,9	12	12	14.3	15. 2	5.2	9.8	128	21	2)	II :	11 1	81.3	29.2	1.1	28	21	4	113	0.4	d a	1.65	4	47.1	1	1 5	6 51	53	21	54.5	6 5	6.5	M	99	8 B)	12	10	(A)	5.5	R)	11	И.	1 7	1.72	71	Ir i	11	\$ 77	78	0.4	10
1	1	1	1	1	1	1	11	1	1	1	11	1-1	1	1	1	11	11	1	1	1	1.1	1.1	1.1	1	1	1	1	1	1	1 1	1	1	1	1	1	1)	1.1	1	1	1	1	U	1	1	1	1	IJ	1.1	1	IJ	1	1	1	1	11	1	Ł	1	1)	1	1	1	IJ	1	1	1	E
2	2	2	2	2	2	2	2 2	2 2	2	2 ;	2 2	2 2	2	2	2 1	2 1	2 2	2	2	2	2 1	12	2 2	2	2	2	2	2	2	11	1 2	2	2	2	2	2 1	2 2	2	2	2	2	11	1 2	2	2	2	2 7	1 2	2	2 2	1 2	2	2	2 1	22	2	2	2	2 2	2	2	2	2 2	2	2	2	2
3	3	3	3	3	3	3	3 3	1 3	3	3 :	3 3	3	3	3	3 3	3 3	3 3	3	3	3	3 3	13	3 3	3	3	3	3	3	1	3.3	13	3	3	3	1	1	3 3	3	3	3	3	1 1	1 3	3	3	3	1 3	1 1	3	3 3	13	3	3	3	13	3	3	3	3 3	3	3	3	3 3	1 3	3	3 :	3
4	4	4	4	4	4	4	8 4	14	4	4	6.6	4	4	4	4	4.1	1.4	4	4	4	4 4	1.4	14	4	4	4	4	4	4	4.1	1.4	4	4	4	4		4.4	4	4	1	4	1	14	4	4	4		14	4	4	14	4	4	4	14	4	4	4	4 4	4	4	4	4 4	4	4	4	4
5	5	5	5	5	5	5	5 5	i 5	5	5 !	5 5	5	5	5	5 !	5 5	5.5	5	5	5	5 5	1 5	5 5	5	5	5	5	5 1	5 1	5 5	5	5	5	5	5	5 1	5 5	5	5	5	5	5 1	5	5	5	5	1	15	5	5 1	1 5	5	5	5	5 5	5	5	5	5 5	5	5	5	5 5	5	5	5 !	5
	-	6	6	6	16	6	6 6	5 5	5	6 1	5 5	6	6	6	6 1	6 6	5.6	6	6	6	6 6		5 6	5	6	6	6	5		5 6	6	6	6		6 1	6 8	5 6	6	6	5	6	5 6	6	8	6	6 1	. 1		6	6 6	6	8	6	6 1	5 6	6	6	6	6 6	1	6	5 1	6 6	6	6	6 1	6
7	1	1	1	1	7	7	11	17	7	1	11	T	7	7	71	1	11	7	7	7	11	1	1	7	7	1	1	1	1	11	1	1	1	1	,	1	11	1	1	1	,	1	11	1	1	1	1)	11	1	1	11	1	1	1	11	1	1	1	11	1	1	1	11	1	7	7	1
	8		8	8	8	8	8 8	8	8	8 1	8 8	8	8	8	8 1	8 1	8 8	8	8	8	8 8	1 8	3 8	8	8	8	8	8.1		8 8	1	8	8	8			8.8	8	8	8	8		18	8	8	8		1	1	8 2	1.8	8	8	1	1 8	8	8	8	8 8	1	8	8	8 8	1 8	8	8	8
-	-	9	-	9	9	9	9 9	9	9	9 1	9.9	9	9	9	9 9	9 8		9	9	g	9 9	1 9	3 9	9	9	9	9	9		9 9	1 5	9			9 1	9 1	9 9	1 5	5	\$	9		1.8	-	8	9		1 1	9	9 1	1 3	-	1	1 1	1.1	4	-	9		1 9	1		9 9	19	9	9	9
1	2	1		-		3	1 1	11	IL	181	11	1.15	11	n.	18 1		1 3	22	10	1:)	5.1	12	2.0	1.21	N	21	IJ	1 1	4	8.1	6.31	31	R	40	0	6.4	12.4	1 45	41	41	0	())	0.01	31	41	51 3		6 31	- 58	21 1	0 61	0	-	H	13 18	1	68	11	10.1	1.1	11	Я.	15.1	1.7	1	21	





Mainframe data feeds (3270)



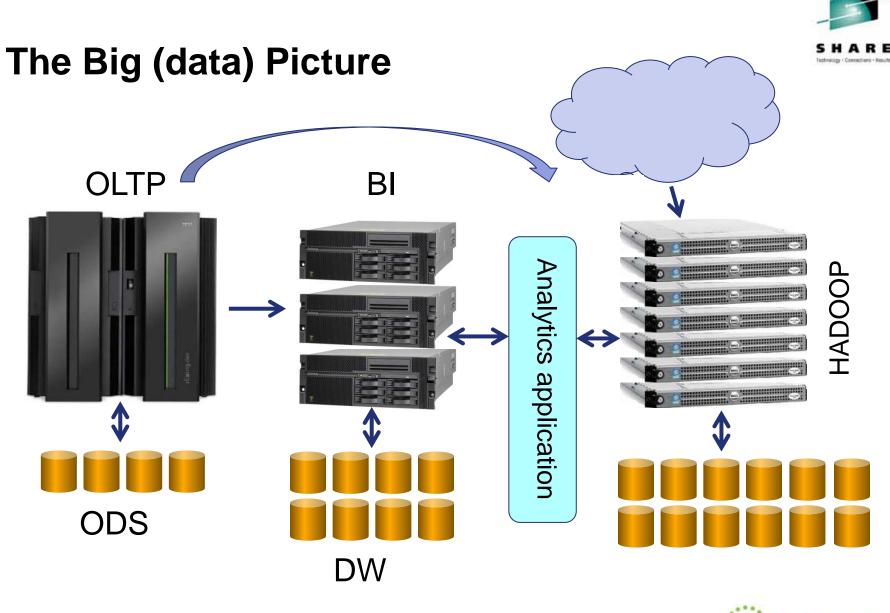




Mainframe Today

- Still highly structured
- Built for low-cost per transaction
- Small transactions, high volume
- RDBMS
 - Limited capability for unstructured data
 - XML/LOB not fully assimilated in DB2



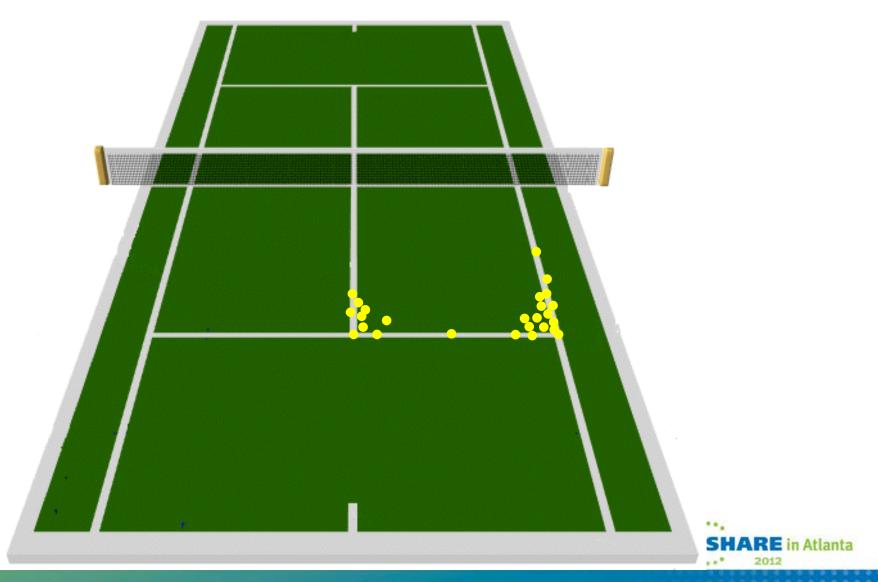


ARE in Atlanta



Australian Open Tennis

7





A lot of unanswered questions

- Maturity of new software (Hadoop)?
 - Reliability security
- Integration with existing systems
- How reduce "time to analysis"
- How to get business value from the data
 - Competitive advantage
 - Fraud detection
 - Risk analysis/avoidance

