F4 Vysoká škola ekonomická v Praze
Fakulta informatiky a statistiky

fis.vse.cz

Návratnost investic do Big Data

Ota Novotný, ota.novotny@vse.cz

Návratnost investice

ROI (%) = výnosy / investice * 100

TCO

Metoda hodnocení nákladových variant. Prostřednictvím TCO se vyjadřují kompletní náklady na investici a její provoz, zohledňující nejen pořizovací cenu, ale také výdaje vznikající vlastnictvím hodnocených statků.

- Přímé náklady na pořízení investice (cena)
- Poplatky spojené s provozem/údržbou
- Náklady spojené s vyškolením uživatelů investice
- Náklady spojené s nečinností investičního celku z příčiny opravy, inovace atp.

Zdroj: https://managementmania.com/cs/total-cost-of-ownership

Big Data TCO

Just look at the cost of managing 1TB of data -- estimates for Hadoop range anywhere from $500 to $2,000; estimates for a high-end data warehouse can range from $20,000 to $200,000.

Zdroj: https://tdwi.org/Articles/2014/03/18/BusinessCasefor-Hadoop.aspx?Page=2

Big Data TCO

For an enterprise class Hadoop cluster, a mid-range Intel server is recommended. These typically cost $4,000 to $6,000 per node with disk capacities between 3TB to 6TB depending desired performance. This means node cost is approximately $1,000 to $2,000 per TB.

Zdroj: Hadoop Business Case: A Cost Effective Queryable Data Archive/Storage Platform statslice.com

BIG DATA TECHNOLOGY TCO BENCHMARK

Bootstrap research

Zdroj: http://get.treasuredata.com/Big_Data_Tech_TCO.html

Big Data TCO – Bootstrap Research

On-Premises Hadoop

On-Premises Appliance

Analytic Database

Cloud Big Data Warehouse

Cloud Big Data Analytics Service

Big Data TCO – Bootstrap Research

Hadoop

- Hardware cost = $1000/TB with a minimum $5000 purchase. 10% discount @ 500 TB. Purchase volume is ½ of raw data volume i.e., 250 TB purchase for 500 TB of raw data.
- Assumes 26% CAGR
- Hardware Maintenance and Support = $100/TB/YR; 10% discount @ 500 TB.
- Power/Space/Cooling = $301/TB/YR
- System Management Software = $3500/node, based on 16 TB node clusters
- Hadoop Management (SysAdmin and DBA) – 2 FTE @ 5 TB; 4 FTE @ 50 TB; 5 FTE @ 500 TB; 6 FTE @ 1PB.

FTE= $150,000/YR

Big Data TCO – Bootstrap Research

Database Appliance

- Appliance Cost = $300,000 @ 5 TB with no upgrades; $400,000 @ 50 TB with $468,000 in upgrades; $1.25M @ 500 TB with $1.9 M in upgrades
- Assumes 26% CAGR
- Appliance Maintenance Cost = 23.5% of appliance expenditure starting in Year 2
- Power/Space/Cooling = $150/TB/YR
- Management (SysAdmin and DBA) = 1 FTE @ 5 TB; 2 FTE @ 50 TB; 3 FTE @ 500 TB; 4 FTE @ 1 PB. FTE= $150,000/YR

Big Data TCO – Bootstrap Research

Analytic Database

- Software License in Perpetuity = $175,000 @ 5TB with $385,000 in upgrades; $400,000 @ 50 TB with $624,000 in upgrades; $2M @ 500 TB with no cost for upgrades
- Assumes 26% CAGR
- Software Maintenance Cost = 21% of software license expenditure starting in Year 1
- Hardware Cost = $400/TB with minimum cost of $5,000. Purchase volume is ½ of raw data volume i.e., 250 TB purchase for 500 TB of raw data.
- Hardware Maintenance Cost = $100/TB/YR
- Power/Space/Cooling = $301/TB/YR
- Management (SysAdmin and DBA) = 1 FTE @ 5 TB; 2 FTE @ 50 TB; 3 FTE @ 500 TB; 4 FTE @ 1 PB.

FTE= $150,000/YR

Big Data TCO – Bootstrap Research

Amazon Redshift Reserve

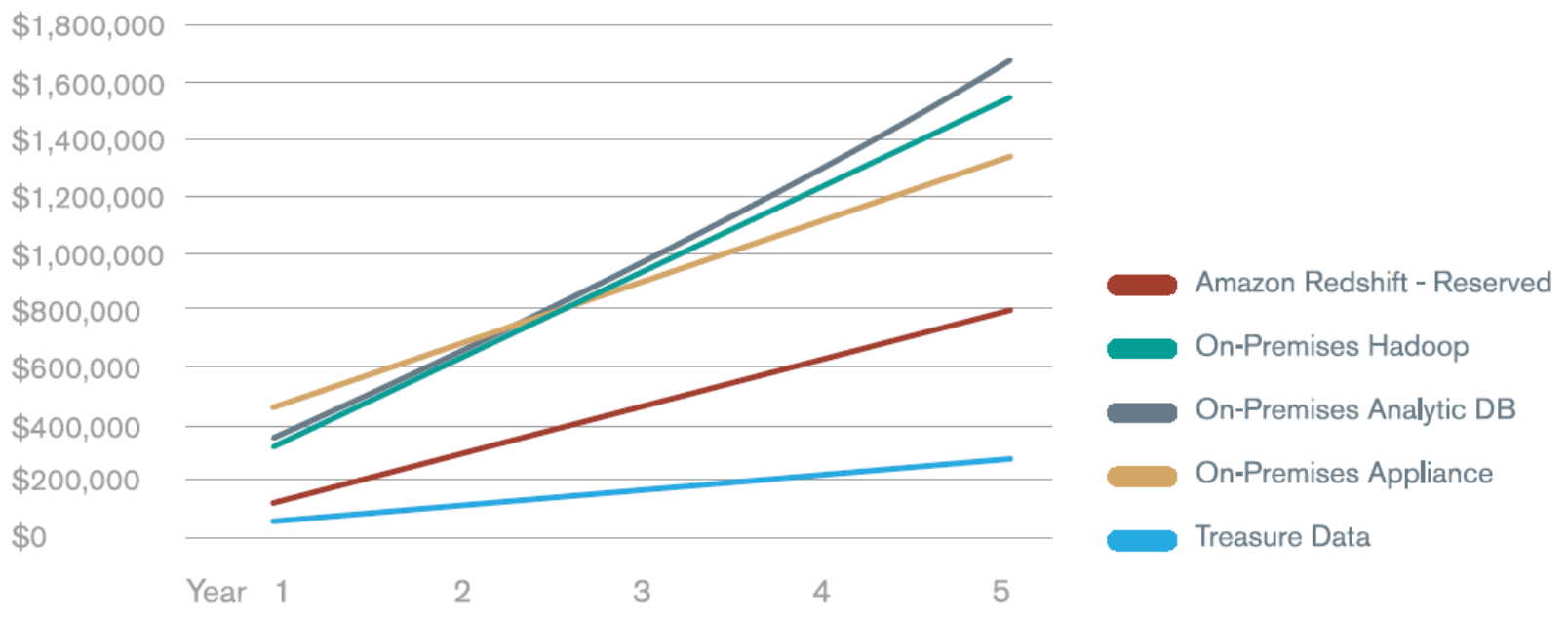
- Leasing Cost = $999/TB/YR
- Assumes 26% CAGR
- Management (DBA) = 1 FTE @ 5, 50, 500 TB; 1.5 FTE @ 1 PB. FTE=$150,000/YR

Big Data TCO – Bootstrap Research

Treasure Data Big Data Analytic Service

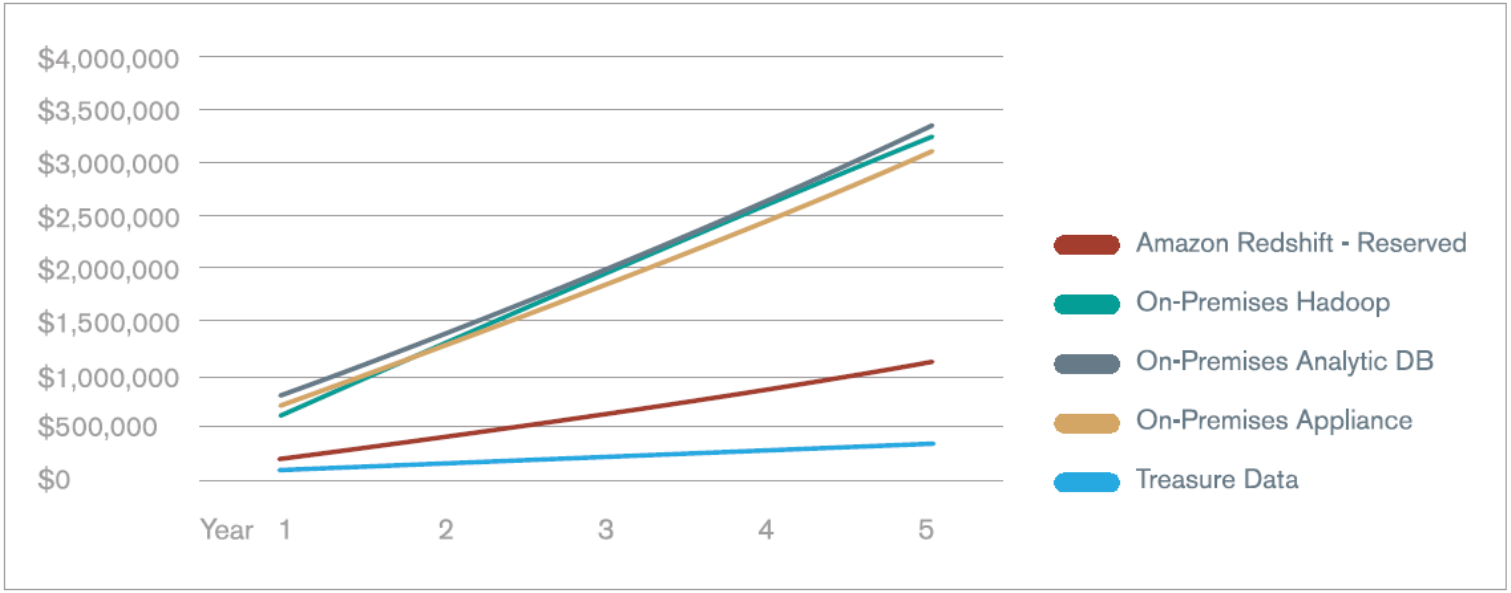
- Leasing Cost = 0-10 TBs @ $3000/YR; 10-30 TBs @ $7500/YR; 50 TBs @$12,500; 500 TB @ $100,000
- Assumes 26% CAGR
- Management (DBA) = 1/3 FTE @ 5, 50, 500 TB; .5 FTE @ 1 PB. FTE=$150,000/YR

5 TERABYTES



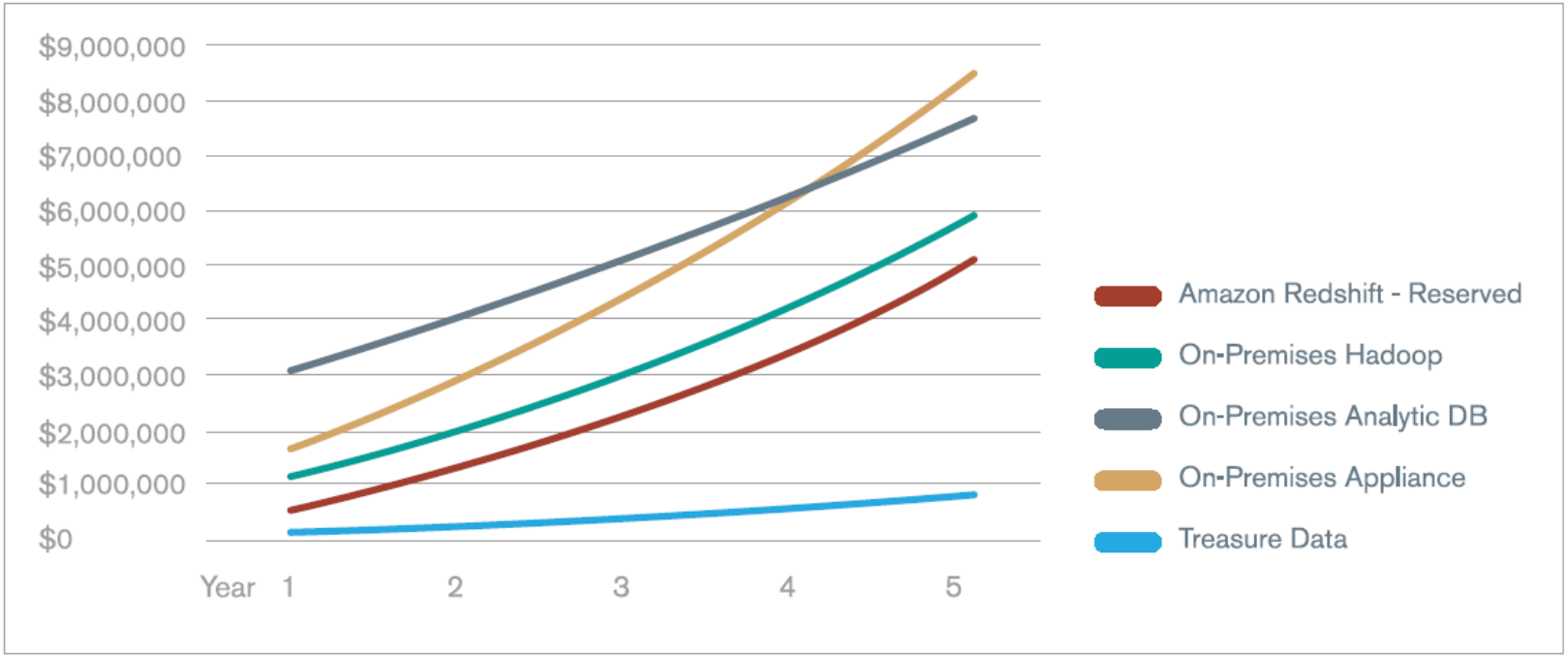
Solution	Year 1	2	3	4	5
Amazon Redshift - Reserved	$154,995	$311,988	$470,979	$632,967	$798,951
On-Premises Hadoop	$310,500	$616,000	$921,500	$1,228,400	$1,537,100
On-Premises Analytic DB	$368,750	$642,200	$930,350	$1,275,550	$1,685,150
On-Premises Appliance	$453,000	$676,500	$900,000	$1,123,500	$1,347,000
Treasure Data	$53,000	$106,000	$159,000	$216,500	$274,000

50 TERABYTES



Solution	Year 1	2	3	4	5
Amazon Redshift - Reserved	$199,950	$412,887	$642,807	$893,706	$1,171,578
On-Premises Hadoop	$642,025	$1,268,857	$1,903,397	$2,548,848	$3,201,512
On-Premises Analytic DB	$804,025	$1,329,497	$1,919,137	$2,581,668	$3,343,572
On-Premises Appliance	$707,500	$1,207,280	$1,757,580	$2,364,640	$3,049,820
Treasure Data	$62,500	$125,000	$187,500	$262,500	$337,500

500 TERABYTES



Solution	Year 1	2	3	4	5
Amazon Redshift - Reserved	$649,500	$1,428,870	$2,372,076	$3,597,075	$5,082,813
On-Premises Hadoop	$1,128,750	$2,130,415	$3,196,942	$4,498,433	$5,902,154
On-Premises Analytic DB	$3,070,250	$4,092,565	$5,154,562	$6,417,063	$7,742,094
On-Premises Appliance	$1,775,000	$3,014,625	$4,460,200	$6,315,938	$8,499,163
Treasure Data	$150,000	$310,000	$481,000	$689,100	$910,510

Big Data TCO – další náklady

- ETL sw a práce s tím spojená
- Vizualizace
- Analytici, datoví inženýři

Big Data výnosy

- ***Macy's Inc.*** *and real-time pricing*. The **retailer adjusts pricing in near-real time for 73 million (!) items**, based on demand and inventory, using technology from **SAS Institute**.

- ***Tipp24 AG****, a platform for placing bets on European lotteries, and prediction*. That led to a 90% decrease in the time it took to build predictive models.**SAP** is in the process of acquiring KXEN. "That's probably a great move by SAP to fill a predictive analytics gap they've long had," Laney said.

Zdroj: http://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell

Big Data výnosy

- ***Wal-Mart Stores Inc.*** *and search*. The mega-retailer's latest search engine for Walmart.com includes semantic data. Wal-Mart says adding semantic search has improved online shoppers completing a purchase by 10% to 15%. "In Wal-Mart terms, that is billions of dollars," Laney said.
- ***Fast food*** *and video*. This company is training cameras on drive-through lanes to determine what to display on its digital menu board. When the lines are longer, the menu features products that can be served up quickly; when the lines are shorter, the menu features higher-margin items that take longer to prepare.

Zdroj: http://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell

Big Data výnosy

- ***PredPol Inc.*** *and repurposing*. The **Los Angeles** and **Santa Cruz police departments**, a team of educators and a company called **PredPol** have taken an algorithm used to predict earthquakes, tweaked it and started feeding it crime data. The software can predict where crimes are likely to occur down to 500 square feet. In LA, there's been a 33% reduction in burglaries and 21% reduction in violent crimes in areas where the software is being used.
- ***Tesco PLC*** *and performance efficiency*: The supermarket chain collected 70 million refrigerator-related data points coming off its units and fed them into a dedicated data warehouse. Those data points were analyzed to keep better tabs on performance, gauge when the machines might need to be serviced and do more proactive maintenance to cut down on energy costs.

Zdroj: http://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell

Big Data výnosy

- ***American Express Co.*** *and business intelligence*. AmEx started looking for indicators that could really predict loyalty and developed sophisticated predictive models to analyze historical transactions and 115 variables to forecast potential churn. The company believes it can now identify 24% of Australian accounts that will close within the next four months.
- ***Express Scripts Holding Co.*** *and product generation*. Express Scripts, which processes pharmaceutical claims, realized that those who most need to take their medications were also those most likely to forget to take their medications. So they created a new product: Beeping medicine caps and automated phone calls reminding patients it's time to take the next dose.
- ***Infinity Property & Casualty Corp.*** *and dark data*. Laney defines *dark data*as underutilized information assets that have been collected for single purpose and then archived. But given the right circumstances, that data can be mined for other reasons. Infinity, for example, realized it had years of adjusters' reports that could be analyzed and correlated to instances of fraud. It built an algorithm out of that project and used the data to reap $12 million in subrogation recoveries.

Zdroj: http://searchcio.techtarget.com/opinion/Ten-big-data-case-studies-in-a-nutshell

Big Data Use cases

Customer Sentiment
Customer Experience (CX)
Predictive Analytics
ETL (DSA) workload management

Marketers, Sales managers,
Customer service managers, Human Capital Management, Retail organizations, Risk managers, Finance

Návratnost investice

ROI (%) = výnosy / investice * 100

Návratnost investic do Big Data

Ota Novotný, ota.novotny@vse.cz