WHY AND HOW TO SEARCH 2 YEARS BACK IN YOUR ELASTIC SEARCH LOGS

the last important part of your index lifecycle management

by cloudvyzor.com

In the perfect world

- All my logs do fit into my Elastic Search cluster
- I don’t need old logs, I solve all problems immediately
- All needed metrics are already predefined and pre-calculated

In the real world

- Logs are terabytes in size
- Only 4-8 weeks of logs fit into my Elastic Search cluster
- I may need to GO BACK to the old logs …

3 possible reasons to go back

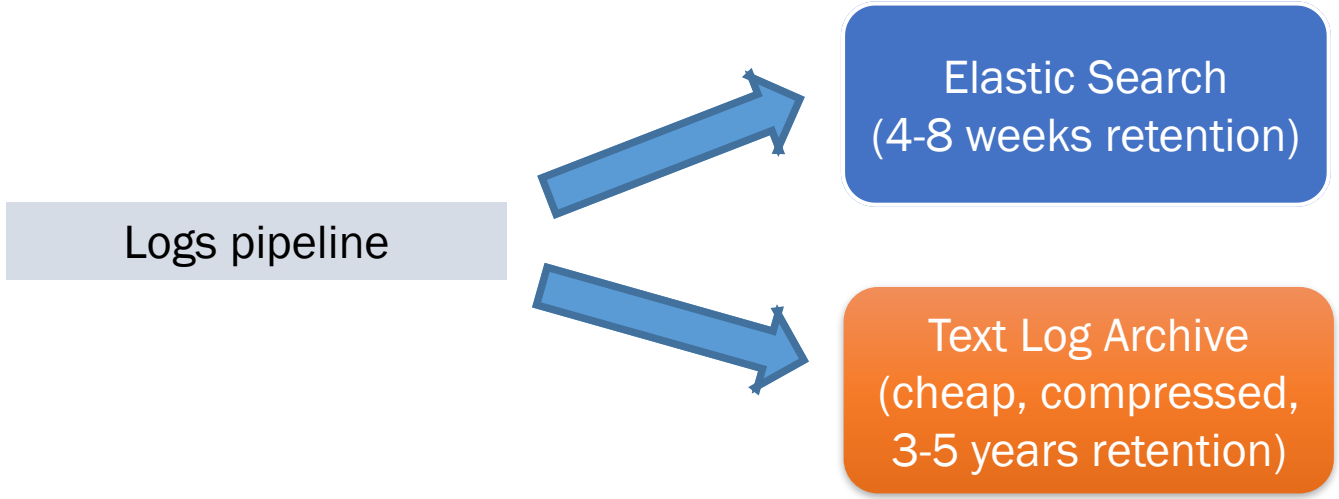
- Support
- Data mining
- Compliance

3 possible reasons to go back

- Support:
  - *“Let’s find when this flaw in business logic got introduced and which customers have been affected.”*
- Data mining
  - *“Let’s calculate this new metric from the last 2 years data”*
- Compliance
  - *“Let’s prove that only authorized engineers had access to production during the last year”*

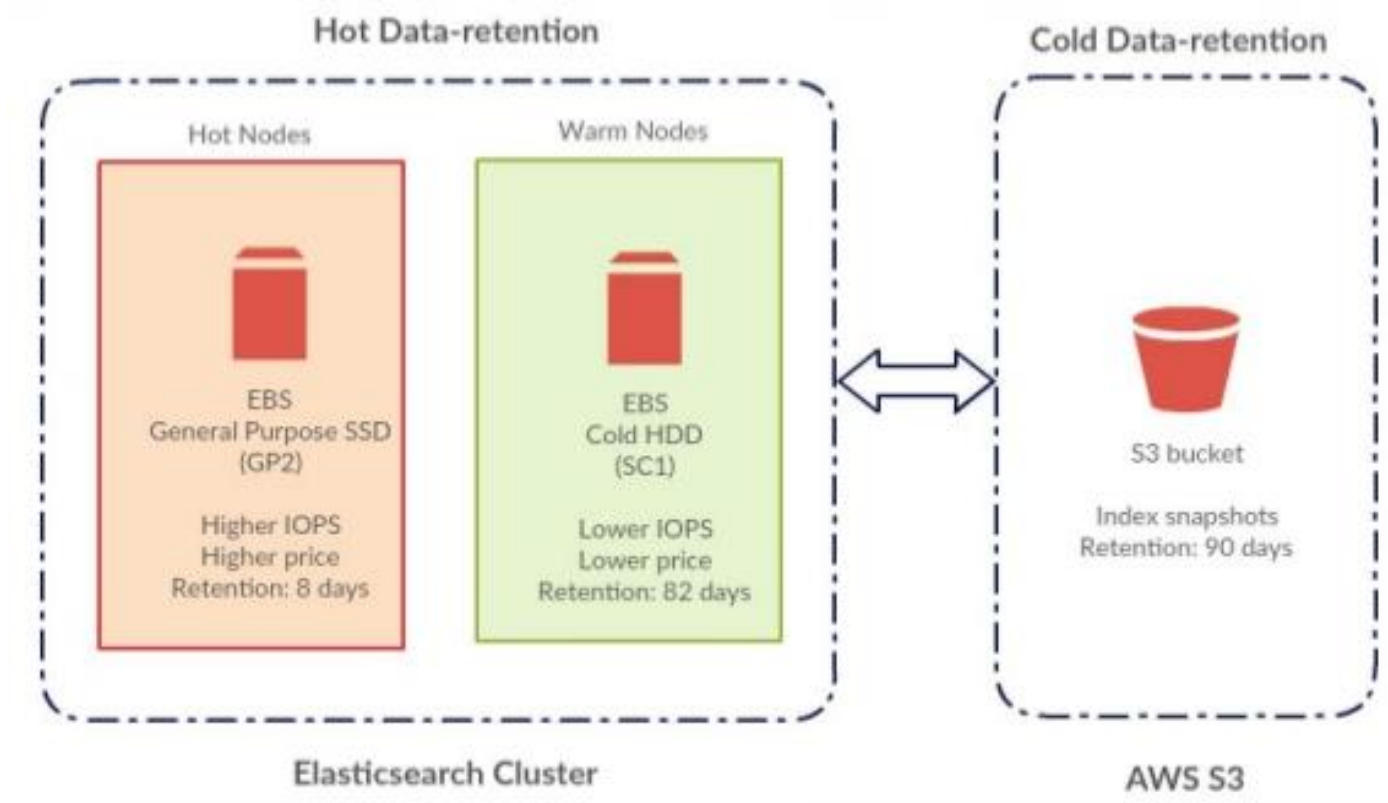
How can I go back?

Some companies do the searchable text log archive



Lazada: https://youtu.be/NAeedJv-S3I?t=1335
Yandex: https://youtu.be/ydwuccVwYBM?t=354

Wait!
But I already do daily index backups!



But can you search it?

Probably: in the easy case

"Show me all logs for this customer for Oct 23th"

- I need to search within the small known time frame
- I know what indexes to mount
- There are just a few of them, I can mount quickly
- So I can search in my ELK cluster quickly

But can you search it?

Not really: in the difficult case

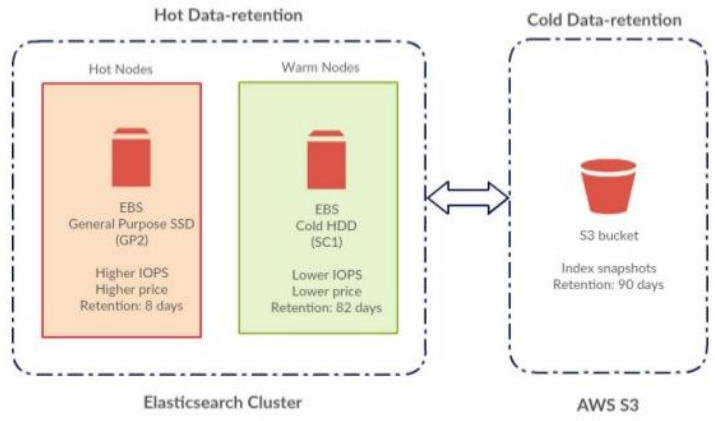
“Let’s search for this error 6 months back to see when it first started”

“Let’s calculate the new metric we didn’t parse before”

- I need to search for some events in all or multiple indexes years back
- I may need full text search
- Some fields have not been parsed yet, need to parse now
- It will be slow to mount hundreds of daily snapshots one-by-one and search

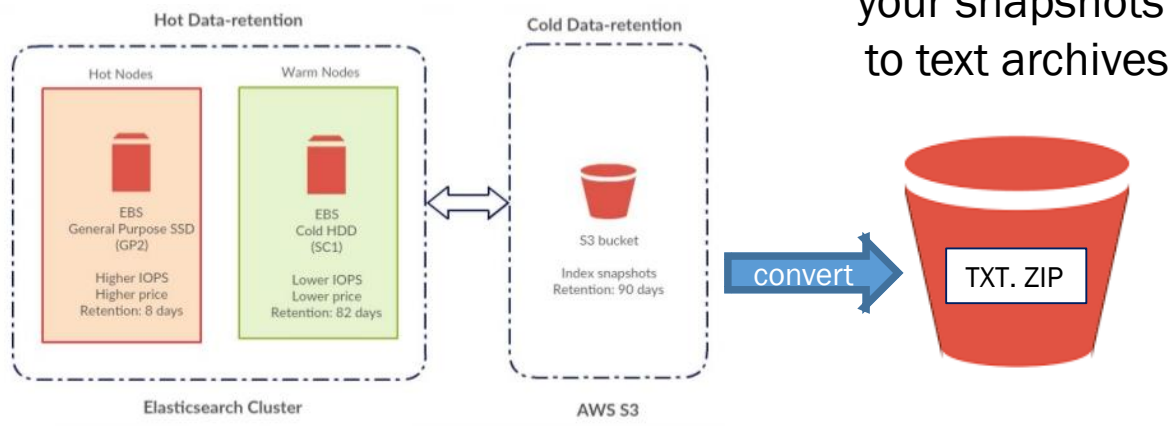
How to be prepared?

1. Keep what you have
2. Add some more steps



How to be prepared?

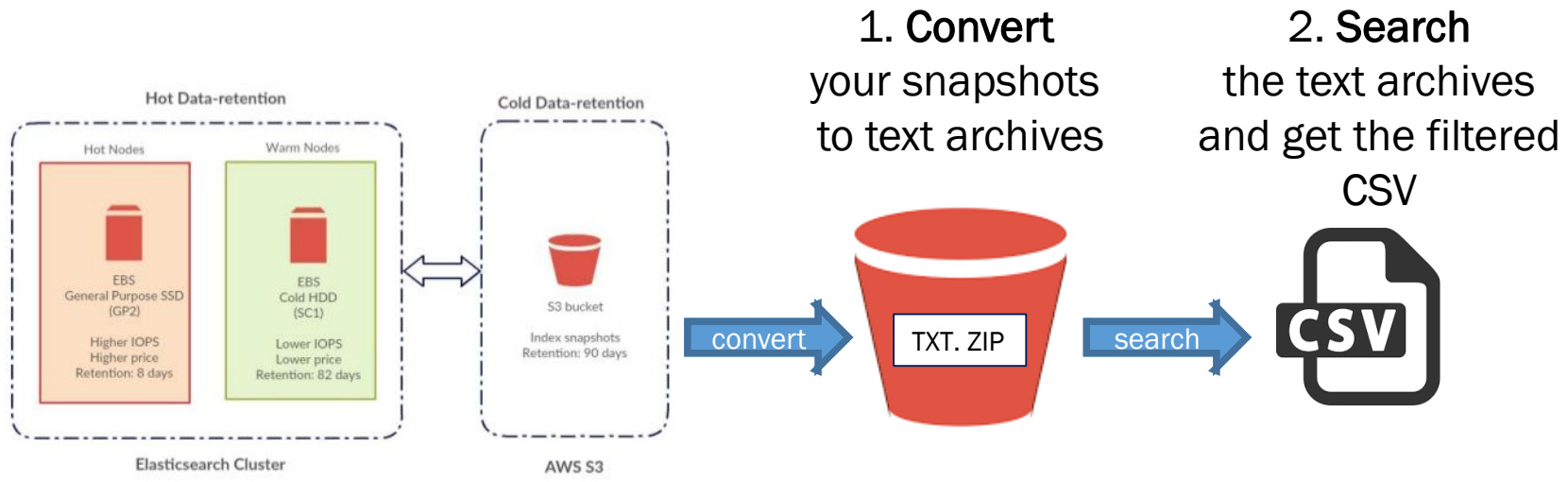
1. Keep what you have
2. Add some more steps



1. **Convert**
your snapshots
to text archives

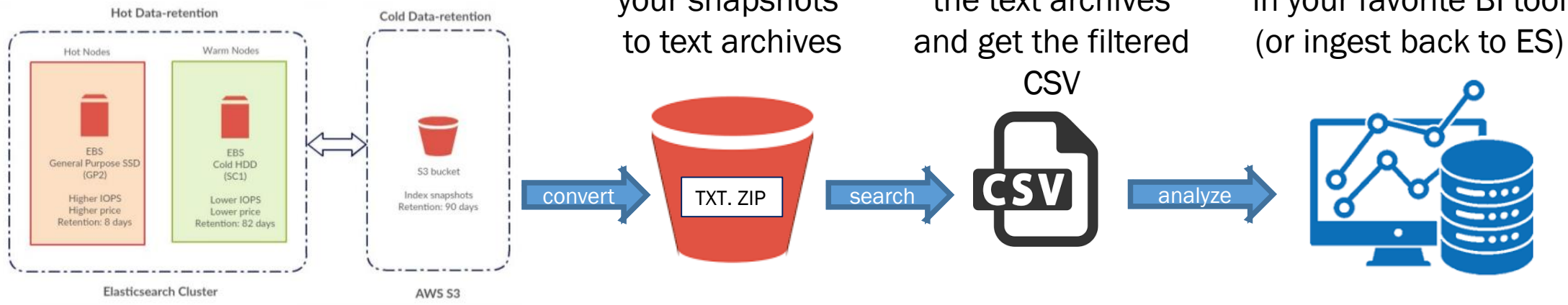
How to be prepared?

1. Keep what you have
2. Add some more steps



How to be prepared?

1. Keep what you have
2. Add some more steps



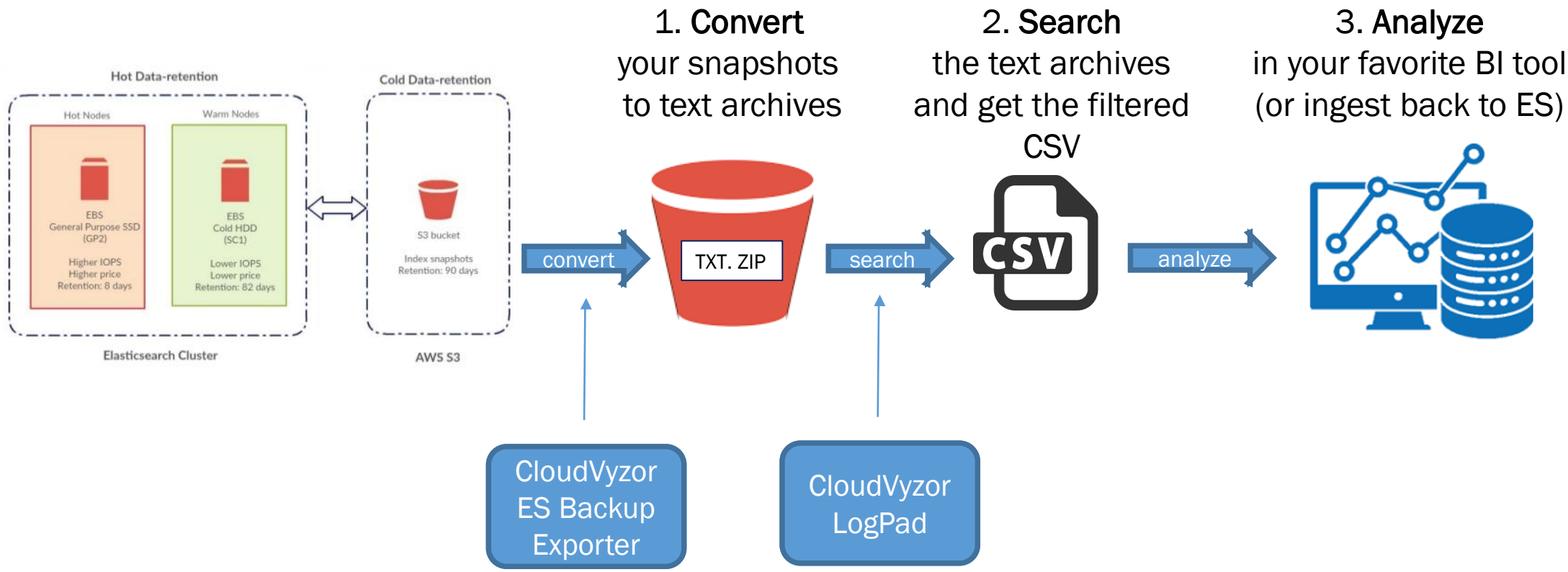
1. **Convert** your snapshots to text archives

2. **Search** the text archives and get the filtered CSV

3. **Analyze** in your favorite BI tool (or ingest back to ES)

How to be prepared?

1. Keep what you have
2. Add some more steps



Case study

2 years of daily ES index snapshots on S3

- *60 TB of snapshots*
- *13B of events*

Task1:

Enlist all engineer access to prod for the past year

Task2:

Get the pivot table of billed transactions for the past year

Step1: Convert snapshots to text file archives

- Tool: CloudVyzor ES Backup Exporter
- Input: AWS S3 bucket, 60 TB, 2 years of snapshots
- Machine: AWS EC2 c5n.18xlarge
  - *72 vCPU, 196 RAM, 100 GB network*

- Time elapsed: 30 hours
- Output: AWS S3 bucket, 3TB of zipped text files
- Spent: 250 USD

Step2: Search and Export

- Tool: CloudVyzor LogPad (on-prem version)
- Input: AWS S3 bucket, 3 TB of compressed text files
- Machine: AWS EC2 c4.xlarge: 4 vCPU, 8GB RAM
- Cost for 1 hour of interactive search & export: 0.5 USD

- Search1: All logons to prod in 2019
  - Search time: 2 minutes, 36GB of logs scanned (AzureAD activity logs)
  - Export time: 15 minutes, 360K events
  - Output: CSV, 36 MB

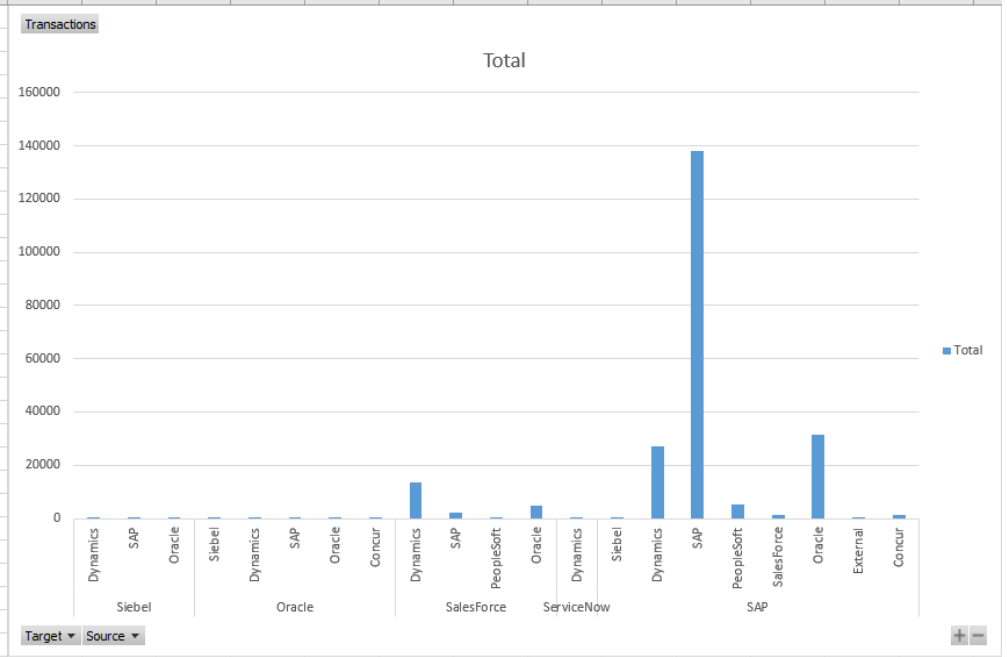
- Search2: All billed transactions in 2019
  - Search time: 5 minutes, 245GB scanned
  - Export time: 11 minutes, 260K events
  - Output: CSV, 40 MB

Step3: Analyze

- Task1: Prod access
  - *you got it: CSV is your final report*

- Task2: Pivot on billed transactions
  - *Tool: MS Excel*
  - *Input: CSV, 260K rows*
  - *Machine: Desktop*
  - *Pivot building time: 10 min*

Row Labels	Transactions
Siebel	**117**
Dynamics	8
SAP	95
Oracle	14
Oracle	**811**
Siebel	3
Dynamics	532
SAP	6
Oracle	239
Concur	31
SalesForce	**20490**
Dynamics	13517
SAP	2159
PeopleSoft	2
Oracle	4812
ServiceNow	**7**
Dynamics	7
SAP	**205155**
Siebel	36
Dynamics	27278
SAP	138032
PeopleSoft	5401
SalesForce	1338
Oracle	31539
External	23
Concur	1508
Grand Total	**226580**



Overall time elapsed

- 1.5 days & 250 USD for the first time
  - if you never had a file archive
  - 30 hours to convert 2 years (60TB) of snapshots
  - 15 min to search and analyze

- 0.5 hour & 1 USD for the next time
  - if you keep your file archive up to date
    - by scheduling the overnight incremental conversion of the last snapshots
    - takes 15 min on 4 CPU machine to convert the snapshots from the last day
  - 15 min to search and analyze

Welcome to try

- CloudVyzor LogPad
- CloudVyzor ES Backup Exporter

And thank you for your attention!