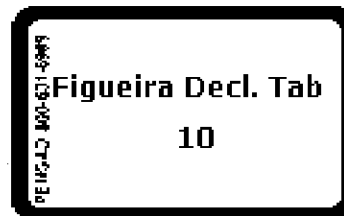


To: "Jay Yagnik" <jyagnik@google.com>, "Michele Covell" <covell@google.com>
 From: "Jeff Faust" <jfaust@google.com>
 Cc:
 Bcc:
 Received Date: 2007-04-25 00:19:17 GMT
 Subject: Fwd: Fingerprinting: cost for 1 hour of reference content



You guys might be interested.

----- Forwarded message -----

From: Jeff Faust <jfaust@google.com>
 Date: Apr 24, 2007 4:11 PM
 Subject: Re: Fingerprinting: cost for 1 hour of reference content
 To: Bo <botao@google.com>
 Cc: Franck Chastagnol <franckc@google.com>, david king <dgking@google.com>

I published some rough numbers here:

<http://trix.corp.google.com/pub?key=pfZfCQU-ItYUc7HCh6x39fQ>

This is just a storage estimate. The main component of machine cost will be GFS chunk servers which correlates directly to the storage estimates but I'm not sure how to account for those.

-Jeff

On 4/24/07, Bo <botao@google.com> wrote:

> capex is
 > opteron \$660
 > RAM/GB \$90
 > Disk/TB \$200
 >
 > btw, our pso folks perhaps know about tv recording. they have such a system
 > that ingest into google backend.

>
 >
 > On 4/24/07, Franck Chastagnol <franckc@google.com> wrote:
 >> Hi Jeff and Bo,

>>
 >> I'm trying to get to the business the hardware and maintenance cost of
 > adding 1 hour of video to the reference fingerprint database.
 >> The idea is that we may bill some of the companies wanting to block
 > content on YT. And we would use this per hour cost for as price tag.
 >> I would like to get conservative numbers (better edge on the higher side
 > in terms of costs).
 >> Assumption is 300,000 lookup daily, with 5min long lookup videos.

>>
 >> Jeff: could you tell us how much hard disk space and how much RAM are
 > needed to handle 1 hour of ref fingerprint ?
 >> It is ok to be conservative and for performance reason say that everything
 > has to be in RAM.

>>
 >> Bo: would you know the capex for a linux server in a datacenter at google
 > ?

>> and what are the specs of such a server in terms of hard drive space and
 > RAM ?

DATE: 12/10/08 EXHIBIT# 19
 DEPONENT: Chastagnol
 CASE: Viacom, et al., v. YouTube, et al., The Football
 Association Premier League, et al., v. YouTube, et al.,
 Case Nos. 07-CV-2203 and 07-CV-3582
 A. Ignacio Howard, CLR, RPR, CSR No. 9830

>>
>>> For your reference, I have pasted below an email thread we had which was
> related.
>>
>>> Let me know any question
>>
>>> Thanks,
>>> Franck
>>
>>
>>> On 4/5/07, Jeff Faust <jfaust@google.com> wrote:
>>>> It's hard for me to come up with good numbers until the design for the
>>>> LSH engine is finalized. There are two distinct paths we can take for
>>>> that design. It also looks like there are some sublinear costs. One
>>>> design makes the cost per reference second much higher initially with
>>>> the incremental cost as reference material is added very low eventually
>>>> increasing to something like linear cost per reference video second.
>>>>
>>>> I'll try to get a handle on this as soon as possible and give you some
>>>> real answers.
>>>>
>>>> -Jeff
>>>>
>>>> On 4/4/07, Jeremy Doig <jeremydo@google.com> wrote:
>>>>> we chatted about this today briefly - do we really need to keep the
> video ?
>>>>> I'm intrigued by the idea of a piece of software that looks at what is
>>>>> coming in the video card and drops out fingerprints every [interval].
> if you
>>>>> assume all broadcast content is copyright in some fashion, all you
> need is a
>>>>> tribune feed on top of that [or something more accurate].
>>>>> i think we should be careful about saying we can do something like
> this
>>>>> though. and could only be done when the algo is stable.
>>>>>
>>>>>
>>>>> On 4/4/07, Jay Yagnik <jyagnik@google.com> wrote:
>>>>>> [[+Jeff, Jeremy]]
>>>>>>
>>>>>> Jeff has been playing with the designs for disk vs. memory systems
> lately.
>>>>>> I'll let him comment on the estimates.
>>>>>>
>>>>>>
>>>>>> On 4/4/07, Franck Chastagnol <franckc@google.com> wrote:
>>>>>>> Hi Jay,
>>>>>>>
>>>>>>>> So the business asked if we could record some TV channels 24/7 and
> add
>>>>>>>> this content to our reference fingerprint DB.
>>>>>>>>> Besides the technical issues (like commercials, etc...) we wanted
> to
>>>>>>>>> also have them understand the infrastructure cost of this.
>>>>>>>>>
>>>>>>>>>> So assuming:
>>>>>>>>>>> - 1Gb of disk storage costs X dollars

>>>>> - 1Gb of memory costs Y dollars
>>>>> - 300,000 lookups per day, with an average video length of 5min
>>>>>
>>>>> Then my question is: what is the infrastructure cost for every
> hour of
>>>> reference content we add to the video fingerprinting reference DB ?
>>>>>
>>>>> Again, I'm not looking at detailed number. Just a rough estimate
> (edging
>>>>> on the high side rather than low side).
>>>>>
>>>>> Thanks for all your help,
>>>>> Franck
>>>>>
>>>>>
>>>>>
>>>>
>>>>
>>>
>>
>>
>
>
