FILED
U.S. DISTRICT COURT
NORTHERN DIST. OF TX.
FT. WORTH DIVISION

2009 AUG 11 AM 9: 51

CLERK OF COURT

**IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF TEXAS
FORT WORTH DIVISION**

AMERICAN AIRLINES, INC.,	§	
Plaintiff,	§	
	§	
V.	§	Civil Action No. 4-08-CV-626-A
	§	
YAHOO! INC., and	§	
OVERTURE SERVICES, INC. d/b/a YAHOO!	§	
SEARCH MARKETING,	§	
Defendants.	§	
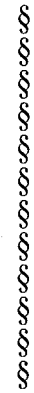
**APPENDIX IN SUPPORT OF DEFENDANTS' REPLY IN SUPPORT OF
THEIR MOTION TO RECONSIDER PORTION OF AUGUST 4, 2009
CORRECTED ORDER ON MOTION TO COMPEL**

EXHIBIT	DOCUMENT	PAGE
A	Excerpts from deposition of Catherine Cameron taken May 29, 2009	3

CERTIFICATE OF SERVICE

The undersigned attorney certifies that a true copy of the foregoing document has been served on counsel of record as follows:

Dee J. Kelly
Dee J. Kelly, Jr.
Lars L. Berg
Kelly Hart & Hallman, LLP
201 Main Street, Suite 2500
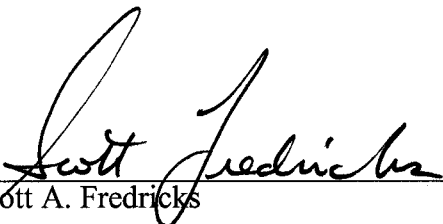Fort Worth, TX 76102

Via Hand Delivery

Frederick Brown
Jason Stavers
Gibson, Dunn & Crutcher LLP
One Montgomery Street, Suite 3100
San Francisco, CA 94104

Via FedExpress

Howard S. Hogan
Gibson, Dunn & Crutcher LLP
1050 Connecticut Avenue, N.W.
Washington, DC 20036

Via FedExpress

Date: August 11, 2009

Scott A. Fredricks

A

1

1   IN THE UNITED STATES DISTRICT COURT

2   FOR THE NORTHERN DISTRICT OF TEXAS

3   FORTH WORTH DIVISION

4   )
     AMERICAN AIRLINES, INC.,   )
5   )
              Plaintiff,   )
6   vs.   ) No. 4:08-CV-626-A
)
7   YAHOO! INC. and OVERTURE   )
     SERVICES, INC. d/b/a YAHOO!   )
8   SEARCH MARKETING,   )
)
9   Defendants.   )
)

10

11

12

13

14

15

16   DEPOSITION OF

17   CATHERINE CAMERON

18   _____

19   May 29, 2009

20

21

22

23

24   REPORTED BY: CAROLYN M. MANN, CSR 10066 [#419917]

25

1   SAN FRANCISCO, CALIFORNIA

2   FRIDAY, MAY 29, 2009

3   9:28 A.M.

4   --oOo--

5   CATHERINE CAMERON

6   _____

7 called as a witness, who, having been first duly sworn,

8 was examined and testified as follows:

9   --oOo--

10   EXAMINATION BY MR. HOGAN

11   MR. HOGAN: Q. Miss Cameron, my name is

12 Howard Hogan. I represent the plaintiffs, American

13 Airlines, in this lawsuit. Could you please state your

14 name for the record.

15   A.   Catherine Cameron.

16   Q.   And this deposition is being held as part of

17 the discovery process in a lawsuit. Do you understand

18 that?

19   A.   Yes.

20   Q.   The court reporter is transcribing your

21 testimony, and that transcript may be used in court. Do

22 you understand that?

23   A.   Yes.

24   Q.   If you don't understand any of my questions,

25 please ask me and I will try to clarify.

1   A.   On September 30th, 2002, I started working for

2   Overture.

3   Q.   What was your initial job with Overture?

4   A.   Data analyst.

5   Q.   And have you been with Overture after it was

6   acquired by Yahoo! and now working with Yahoo!?

7   A.   Yes.

8   Q.   All right. So before we move on to your time

9   with Overture and Yahoo!, I guess the first question is

10   when did you find time to get your MBA?

11   A.   When I was working at Ralphs Grocery Company,

12   I worked there while I got my MBA. I worked on my MBA

13   full-time for a year while I was working at Ralphs

14   part-time, and then I finished up my MBA part-time while

15   I was working at Ralphs full-time.

16   Q.   Were you ever asked to leave any of the

17   positions you've just listed involuntarily?

18   A.   No.

19   Q.   All right. So if you can, can you list the

20   positions that you've had at Overture and Yahoo! from

21   data analyst in 2002 to the present.

22   A.   Yes. I was a data analyst from 2002 until

23   December 2003, when I was promoted to senior data

24   analyst. In 2005, I was promoted to manager of internal

25   reporting. In 2007, I was promoted to senior manager of

1 internal reporting, which is what I do today.

2   Q.   What are your responsibilities as senior

3 manager for internal reporting?

4   A.   I manage a department of 20 people. They're

5 engineers and data analysts who create reports, pull

6 data, analyze data trends, investigate and fix problems

7 with the data, and provide the source of truth for

8 revenue data for Yahoo!.

9   Q.   What do you mean by "provide a source of truth

10 for revenue data"?

11   A.   So within Yahoo!, there are many different

12 people, analysts, spreadsheets, but when it comes to

13 wanting to know what the actual gross revenue numbers

14 are, that's my team. For search data, I have a sister

15 team in Bangalore of about 30 people, and we

16 transitioned the bulk of the data reporting work to

17 Bangalore starting last year.

18   Q.   So the 20 people that you mentioned, those are

19 20 people in Burbank that are separate from the

20 Bangalore --

21   A.   Yes.

22   Q.   -- sister team?

23   A.   Yeah, and the 20 people in Burbank, we work on

24 display data and we do some of the work still on search.

25   Q.   What do you mean by "display data"?

1 everything, but it's too expensive.

2   Q.   We'll get back to the tape data in a moment,

3 but let's talk about the live data for a second. Could

4 you explain to me your considerations in evaluating the

5 size of the data with respect to whether to keep it or

6 not?

7   A.   Sure. So data that is very small, for

8 example, revenue for the company for a particular day is

9 very tiny and it's very interesting, so we can keep that

10 for a long period of time. Data that is every search a

11 person types in, which can create a billion rows of data

12 a day, is very large, not very interesting, very

13 difficult to query, and so we keep that for a short

14 period of time.

15   Q.   Can you define "small," you know, in terms of

16 size?

17   A.   Sure. Less than a hundred thousand rows a

18 day, if you think of like an Excel spreadsheet.

19   Q.   What does that translate to in terms of bytes

20 of data?

21   A.   I do not know. I am not a hardware and

22 storage expert. I'm a, the data expert in terms of what

23 gets stored, how long it gets stored, and how, what it's

24 used for.

25   Q.   So what would be considered a large set of

1 data?

2   A.   The largest sets we have, as I said, are every

3 search, which is a billion rows a day, or impressions,

4 which is about 5 billion rows a day. That's very large.

5 Our databases are very large.

6   Q.   What are the, what's the rough size of

7 Yahoo!'s total databases?

8   A.   We have many databases. So I think you're

9 just interested in the ones pertaining to search.

10   Q.   Well, that may or may not be the case. My

11 question was broader, what is the overall size.

12   A.   Okay. I don't know about most of the areas

13 outside of search. There's a whole particular area of

14 the company that uses vast amounts of data that tracks

15 user behavior, which is, I go to this page, I stay here

16 for this period of time, then I read this news article,

17 then I go back to this page. It's not search-related

18 or, or revenue-related data.

19   The data I deal with is the, talking about

20 with the search data, is we have currently a data

21 warehouse that's 1.5 petabytes, a petabyte is 1,024

22 terabytes. It's one of the largest data storage

23 warehouses in the world. It's been written up in a

24 couple of cases.

25   Even going back to 2003, we used an Oracle

1 database at the time that was 270 terabytes, and I

2 remember it being the largest database in the world next

3 to France Telecom, and I believe it exceeded that. We

4 have several other large Oracle databases.

5   In addition to the 1.5 petabyte storage, which

6 is not Oracle, we have another database we use for

7 reporting for search called Edward, E-D-W-A-R-D, which

8 is 170 terabytes.

9   Q.   So a moment ago, you distinguished between

10 search data and user data.

11   A.   Yes.

12   Q.   Which of the three sets of databases that

13 you've just mentioned -- well, let me stop and back up.

14 Make sure I've got the three.

15   A.   Sure.

16   Q.   One is a not-Oracle database that's 1.5

17 petabytes?

18   A.   Yes. That is called Sage, S-A-G-E. That will

19 make it easier. Sage.

20   Q.   Where is that located?

21   A.   We just moved it. It's in the United States.

22 It was in the SP1, yes, Space Park 1.

23   Q.   What is Space Park 1?

24   A.   It's a location in the Bay Area here. We also

25 have a replica of it in MUD, M-U-D, which is in Dallas,

1   which is for backup purposes. So if you think about it,

2   if something were to happen in California where we could

3   not get to that data or that database would fail, we can

4   get to it in Dallas, switch over.

5   Q.   What is the purpose of the Sage database?

6   A.   The purpose of the Sage database is to take

7   all the click stream data that comes in, so every click,

8   every search, all the impressions, all the ads that are

9   shown, and put it into meaningful tables and data feeds

10 for reporting for billing and paying and for several

11 other systems.

12   Q.   What applications are supported by the Sage

13 database?

14   A.   There are many applications. The most

15 important, to me, is reporting for internal recording.

16 Sage feeds Edward, which is the main reporting database.

17   Q.   Any other applications that --

18   A.   Oracle Financials, which does our billing and

19 paying. It feeds the revenue reordering system, which

20 determines the order that ads are shown up on a page.

21 It feeds account monitoring, which is how much money

22 advertiser accounts have used. And it feeds several

23 others that are lesser importance that I'm not familiar

24 with, but they're not main systems for Yahoo!.

25   The other system it feeds that you need to

1 know about is called The Grid.

2   Q.   What is The Grid?

3   A.   So there's a big fad called "cloud computing"

4 in the Internet world and in general, where you take

5 massive amounts of data and instead of having one

6 processor or four processors work on it, you can take

7 hundreds of processes -- processors and work on the

8 data, and it will do very fast calculations.

9   Q.   And the Sage database is, has hundreds of

10 processors to allow it --

11   A.   No.   The Grid does.

12   Q.   So what is the difference between Sage -- oh,

13 I see.   Sage feeds The Grid?

14   A.   Right.

15   Q.   The Grid has the hundreds of databases?

16   A.   Yeah.   Sage has its own processors, it's just

17 that it's an older system, and the way that we go with

18 technology is about every two years you can increase

19 processing power and technology takes these huge leaps.

20 So Sage is the, the current technology, and then The

21 Grid is the future.

22   Q.   So let me understand this.   The data itself

23 resides in Sage; Grid extracts data from Sage to be

24 processed.   Is that how it works?

25   A.   Grid's -- so most of the data processing is

1   Q.   So Yahoo! has several data colos and this is

2   in one, the Edward 2 is in one of them?

3   A.   Uh-huh. So I think, looking back, "data

4   colo," everyone seems to know what that means, so I'm

5   thinking it's an industry term. I've been with Yahoo!

6   so long that it's sometimes hard to separate.

7   Q.   So would all the databases we've talked

8   about -- Sage, Edward, Edward 2, and I guess for lack of

9   a better word, Sage 2 -- would each have a data colo for

10 them?

11   A.   What you have is a data colo is a physical

12 site. So we had a big one in Santa Clara. And it's a

13 physical location where you can put databases, data

14 processors, servers, storage, that has plenty of power

15 and electricity. You can read about it in the trade

16 magazines. Everybody, the states are all saying, We

17 want these data colos. So the more data you have, the

18 bigger the locations you need to store them in.

19   And then if you are smart, you have a backup

20 location somewhere else physically located for when the

21 main colo, should it go down.

22   Q.   So for Sage, the data colo is in California

23 but the backup is in MUD, Texas?

24   A.   It's called M-U-D. It's in Dallas, but --

25   Q.   M-U-D isn't a name of a place --

1   A.   Yeah.

2   Q.   -- it's the name of the data colo.

3   A.   Yes, yes.

4   Q.   Okay.

5   A.   We moved them around due to leases expiring

6   and getting better deals -- and you can read about this

7   in the trade magazines -- and increasing the size of our

8   data, that they use a tremendous amount of power.

9   Q.   All right. So just to finish up with

10   Edward 2, the data colo is called AC4, and it's -- where

11   is it physically --

12   A.   I'm sorry, AK4. It's located on the East

13   Coast. I don't know the exact city.

14   Q.   And is Edward 2 actively being used, or it's

15   just a work in progress?

16   A.   It is actively being used.

17   Q.   Who is actively using it?

18   A.   My team has built the reports on it. And it's

19   actively being used to monitor the health of the

20   newspaper consortium on the Exchange, and we have a few

21   test accounts running, and we're seeing how they're

22   doing.

23   Q.   All right. So right now Edward 2 is basically

24   a database being designed to facilitate the exchange of

25   banner ads between newspaper Web sites and ads for

1   A.   No.   They were given the request for data from

2 our legal team in a manner in which an engineer can

3 understand it.

4   Q.   Okay.   I'd like to talk about the spreadsheets

5 that were produced --

6   A.   Sure.

7   Q.   -- in the case.   Is it correct that the search

8 data, the first bulk of search data that Yahoo! produced

9 was for September 22nd through February -- sorry,

10 September 22nd, 2007, through February 2009?

11   A.   So I don't know the order of -- when you say

12 "first," I don't know the order that you received.

13   Q.   Let's put that aside, go backwards,

14 chronological order.

15   A.   Okay.   So if you're asking about the data that

16 we provided, and you say, What have we provided or what

17 information have we given you, and you're specific about

18 it, it's a lot easier for me to answer it.

19   Q.   Well, I understood that there was a block of

20 data that was, the most recent chronological was --

21   A.   Yes.

22   Q.   -- September 22nd, 2007, through

23 February 2009.

24   A.   Yes.

25   Q.   And this data comes from the Sage database; is

1   that right?

2   A.   Yes.

3   Q.   What -- do you know why September 22nd, 2007,

4   is the starting point for that data?

5   A.   Yes. Because the retention for that

6   particular block of data, it rolls off after

7   approximately 18 months.

8   Q.   Does the data for that block, the

9   September 22nd, substantively differ from any of the

10 previous segments of data chronologically?

11   A.   The query we would use and the table structure

12 would be the same.

13   Q.   Then going back, I believe the next period for

14 which data has been produced is October 2005 through

15 March 2007?

16   A.   Yes.

17   Q.   Do you know why Yahoo! has not produced data

18 for the April 2007 to September 22, 2007 period?

19   A.   Yes. The reason is because it had rolled off

20 from Sage due to the size of the data and the limitation

21 of storage on Sage.

22   Q.   Do you know when it rolled off?

23   A.   It rolls off after 18 months. So if you look

24 back from today -- actually, we're keeping it a little

25 longer now because of this particular case. We would

1   see data going back in history for 18 months. And then

2   if we had looked at it from last month, it would have

3   been 18 months. And then the month before, it would

4   have been 18 months.

5   Q.   Do you know if the data for April 2007 through

6   September 22nd, 2007, was preserved on backup tapes?

7   A.   It is.

8   Q.   Is that going to be produced?

9   A.   We are working on it.

10   Q.   Okay. Do you have an estimate of when it will

11   be produced?

12   A.   They will give me an ETA -- "they" meaning the

13   data -- it's a concerted monumental effort requiring

14   coordination across many data teams. And as of this

15   morning, I was told I would get an ETA at the end of

16   next week, an ETA meaning estimated time of arrival.

17   Q.   Okay. The -- chronologically going back,

18   then, the next block would have been, I think it's

19   October 2005 to March 2007, is that your recollection?

20   A.   Yes.

21   Q.   And would that data overlap the Live_Stor and

22   the Sage phases?

23   A.   The data that, the data block you had just

24   talked about was data on Live_Stor.

25   Now, one thing I need to be clear on is that