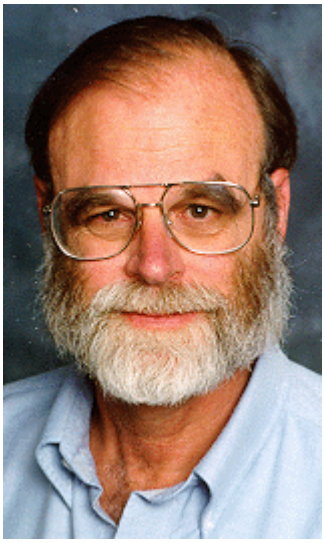
the Availability Digest

# Jim Gray – In Memoriam

July 2008

Jim Gray has been a computing legend since the 1970s. Perhaps his most visible contribution has been in the field of transaction processing. Jim's work in this field is what now powers network applications from ATMs to Internet shopping to enterprise mission-critical applications. He was the recipient of the prestigious Turing Award in 1998 for "seminal contributions to database and transaction processing research."



Jim Gray, from his home page
http://research.microsoft.com/~Gray/

Jim Gray was 63 when he was lost at sea eighteen months ago. On Sunday, January 28, 2007, he set sail in his forty-foot sailboat, *Tenacious*, enroute to the Farallon Islands, a wildlife refuge just 27 miles off the shore of Northern California, to spread the ashes of his mother. By nightfall, Jim had vanished without a trace. His disappearance triggered one of the most massive search-and-rescue efforts in history.

This article is not about the search for his whereabouts, though it certainly covers that. Rather, it is about Jim Gray, the scientist.

Mentoring was one of Jim's strengths. He mentored such people as Werner Vogels, CTO of Amazon, and Sergey Brin, cofounder of Google. And he mentored me.

I first met Jim when he was at Tandem Computers working on fault-tolerant, high-capacity databases. Though he was a decade my junior, he was more of a father figure to me. I often shared my work with him. He touted my book on *Performance Analysis of Transaction Processing Systems* as one of the most readable books on performance analysis that he had come across. When he first read *Breaking the Availability Barrier*, which I coauthored with two others, he said "Wow! What a great book. Very readable, very direct. I loved the laws," a statement that he allowed us to publish.

This support was all the more important because of the stature that Jim had achieved in the technology world.

Steve Silberman, contributing editor to Wired Magazine, wrote a wonderful story about Jim Gray, his achievements, and the search for him. Dated July 24, 2007, it was published in Wired Magazine in its Issue 8, Volume 17.[1] Much of the material in this article is taken from Steve's work.

[1] http://www.wired.com/techbiz/people/magazine/15-08/ff_jimgray?currentPage=all

## Jim Gray – The Technologist

Born in 1944, Jim enrolled in UC Berkeley in 1961 as a mathematics student. However, he discovered computers when he took a course in numerical analysis, the only mathematics course at Berkeley that offered access to a computer. He went on to earn the university's first PhD in computer sciences in 1969.

Jim joined Bell Labs for a short stint and then moved to IBM Research Labs. There he joined a team involved in research on Ted Codd's recently published concepts describing a relational database. The team was tasked with trying to turn Codd's theories into functional software.

Jim focused mainly on a new concept called "transactions." A transaction was a bounded set of database operations that were atomic – they either were all applied to the database successfully, or none were. The questions he was trying to answer included:

- When should changes be firmly committed to the database?
- What if one of a pair of transactions failed halfway through?
- How can many users access and update the same data simultaneously without corrupting the database?

In his research papers, he likened transactions to a marriage contract. When is the right time to say "I do"? His solutions are embedded in our network world, from the corner ATM to enterprise applications, to online shopping, and on and on.

Jim and his colleagues created one of the first relational databases, which they called System R. System R was the first implementation of Structured Query Language (SQL), which has since become the standard relational data query language.[2]

Most techies considered business data processing to be beneath them. For instance, how exciting was it to help Bank of America reconcile checking accounts? But not a young entrepreneur named Larry Ellison. Working at Ampex on a data-storage system named Oracle for the CIA, Larry started a fledgling company using System R's technical papers as a blueprint to develop a relational database for the minicomputer market. He brought his product to market under the name of Oracle.

Years later, in an obvious reference to Oracle, Jim said during an interview that his life had been a "researcher's dream – you have a lot of fun, you do something innovative, and then people make billions of dollars off of it."

However, Jim was no stranger to productizing advanced theoretical concepts. For instance, he contributed heavily to the development of one of IBM's first relational databases – DB2. DB2 remains today a popular database marketed and fully supported by IBM.

In the 1980s, Jim moved to Tandem Computers to continue his work on high-availability systems and transaction processing, contributing to Tandem's Transaction Monitoring Facility (TMF) and NonStop SQL. TMF and NonStop SQL provide the transaction-processing engine that is recognized today as one of the most powerful and scalable in the industry.

Early implementations of relational databases suffered greatly from performance issues. To improve performance, Gray met with 25 colleagues from various commercial and university institutions to coauthor an anonymous article that defined a measurement for computer performance.[3] The Debit/Credit benchmark that was proposed by this group evolved to become the TPC-C standard defined by the Transaction Processing Performance Council. TPC-C and its

[2] James Gray, et al., A History and Evaluation of System R, *Communications of the ACM*, Vol. 24, No. 10; October, 1981.
[3] Anon, et. al., A Measure of Transaction Processing Power, *Datamation*, V 31.7; April 1985.

successors are widely used today for comparative measures of performance by database and computer vendors.

These benchmarks verified the exponential improvement in the power of transaction-processing systems over the following years. A 1998 $10 million Tandem system had the capacity to process about 200 transactions per second. This system comprised 34 CPUs and 86 disks. Just seven years later, a $400 Toshiba desktop computer processed over 8,000 transactions per second![4]

Jim left Tandem and joined Digital Equipment Corporation to continue his work in transaction processing. It was during his Tandem and Digital tenures that he coauthored with Andreas Reuter his book entitled *Transaction Processing: Concepts and Techniques*.[5] This book is arguably the authoritative treatise on transaction processing, even today.

Jim joined Microsoft in 1995 as a Technical Fellow. He convinced Microsoft to open a research center in San Francisco so that he and his wife, Donna Carnes, would not have to move to Redmond, Washington. The center became Microsoft's Bay Area Research Center. It was there that Jim immersed himself in the study of massive databases.

One of Jim's first projects at Microsoft was the development of a web site that he called TerraServer. TerraServer used data from the commercial satellite industry to bring high-resolution imagery of the planet Earth online. Previously available only to intelligence agencies and weather forecasters, detailed pictures of anywhere on earth became available to the masses seven years before the introduction of Google Earth. TerraServer became at the time the largest database on the Internet.

Jim mingled freely with a wide range of scientists and technologists. An avid sailor, he hosted on his boat many astronomers, oceanographers, geologists, geneticists, and those from many other disciplines. He had an interest in each of these fields, and he collaborated with these technologists on a broad array of projects.

For example, he teamed with astronomer Alex Szalay of Johns Hopkins University to port a massive star-mapping project – the Sloan Digital Sky Survey – to the web. This made astronomical data available via the Internet to professional astronomers, to backyard stargazers and to students. Since its debut in 2001, this data has become the most widely used astronomical resource in the world, supporting many discoveries of dwarf galaxies, dark matter, other planets, and sonic waves triggered by the Big Bang.

Jim once told me of his problems in sharing large databases that had to be distributed among the scientific community. It rapidly became clear to him that it was faster to mail a hard disk (three days) than it was to transmit a terabyte database over even the fastest communication channel (two weeks). This was a major hurdle to the sharing of very large databases.

In 1998, Jim was the recipient of the prestigious Turing Award for "seminal contributions to database and transaction processing research and technical leadership in system implementation from research projects to commercial products." Awarded by the Association of Computing Machinery (ACM) and often referred to as the Nobel Prize of computing, the Turing Award is presented once a year to "an individual selected for contributions of a technical nature made to the computing community. The contributions should be of lasting and major technical importance to the computer field."

---

[4] Jim Gray, Thousands of *DebitCredit* Transactions-Per-Second: Easy and Inexpensive, Microsoft Research Technical Report MSR-Tr-2005-39; April 1, 2005.

[5] Jim Gray, Andreas Reuter, *Transaction Processing: Concepts and Techniques*, Morgan Kaufmann Publishers Inc.; 1993.

## The Search

When Jim failed to show up the evening of his departure to the Farallon Islands, the Coast Guard initiated a search for him. At this point, it was a standard search, since boats are often reported missing only to show up hours later or at some other port. However, as time went on, the search for Jim became one of the most massive search-and-rescue missions in history. Before it was over:

- the Coast Guard had scoured 132,000 square miles of ocean by air and sea.
- teams of scientists and Silicon Valley power players had steered satellites and NASA planes over a swath of ocean from southern California to Oregon and 300 miles out to sea.
- oceanographers and engineers from the U.S. Navy and NASA's Jet Propulsion Lab had joined in.
- thousands of volunteers around the world had searched satellite images for any sign of the *Tenacious*.

As all sailors should, Jim carried an EPIRB (Emergency Position Indicating Radio Beacon) onboard. An EPIRB will emit a signal when in the water to guide search-and-rescue operations to the scene. No EPIRB signal was ever heard, nor was a Mayday call from Jim. There was no sign of flotsam from the boat floating in the water or washed up on shore. In short, the *Tenacious* had disappeared without a trace.

The Coast Guard initially searched 40,000 square miles of ocean without success. After three days, they announced that they were calling off the search. At this point, the Friends of Jim, a large group of technologists and corporate executives, came out of nowhere to start their own search. Microsoft let it be known that money for the search was no object. The Coast Guard then reversed its decision and greatly expanded its search, using computer simulations to determine where the boat might have drifted in the event of the incapacity of its skipper.

The Coast Guard pressed into service C-130 aircraft, helicopters, and patrol boats to widen the search. They investigated the logs and hulls of any ships that might have struck the *Tenacious*. The Canadian Space Agency flew one of their satellites with cloud-piercing radar over the search area. This was followed by scans from DigitalGlobe's QuickBird and GeoEye's Ikonos.

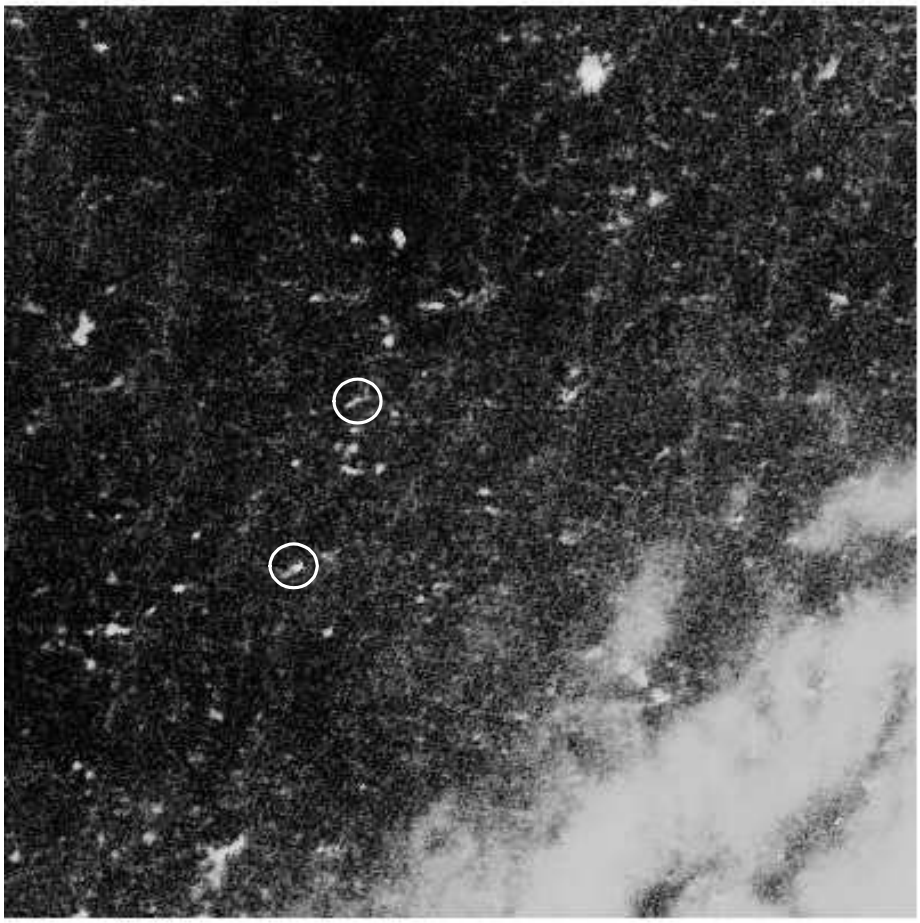
The crunching of all of this satellite data was going to be a massive effort. Amazon turned Mechanical Turk to the task. Mechanical Turk is an Amazon service that allows online workers to cooperate in massive tasks. Five days after Jim's disappearance, Werner Vogels, CTO of Amazon and a close friend of Jim's, posted a request on his blog entitled "Help find Jim Gray." More than 12,000 volunteers signed up and set about scanning satellite images of the search area. The question was, "Is that a white cap or a boat." Over the weekend, 30,000 square miles of ocean had been scanned. Twenty specks were tagged as "likely," and one was tagged as "highly likely." But the data was stale. If one of these specks were the *Tenacious*, where would it have drifted?

A team of ocean modelers was assembled. Using Coast Guard radio buoys and other sensing devices, they attempted to predict the drift trajectory. All leads were tracked by flights but with no results.

Two weeks later, Jim's wife, Donna Carnes, chartered a fishing boat. Accompanied by a marine search expert, they set out for the Farallon Islands to interview the resident naturalist. Access to the islands, an important wildlife reserve, was limited; but the office of Nancy Pelosi, House Speaker of the U.S. Congress, arranged clearance so that they could talk to the naturalist. As they approached the islands, they had to traverse a maze of debris, including many large logs,

any of which could have pierced *Tenacious'* hull. But the interview yielded nothing except a possible sighting of *Tenacious* the afternoon of Jim's disappearance.

In Spring, Donna Carnes employed a private company to comb the ocean floor with side-scan sonar and remote-operated underwater vehicles. Three-and-a-half months later, nothing had turned up. Jim's family ended its search on May 31.



A satellite image from Mechanical Turk. Can you find *Tenacious?*

On May 31, 2008, one year after the end of the search, the University of California, Berkeley, hosted a conference that was a tribute to Jim

## A Tribute

Jim Gray was one of the brightest technologists of our time, and his work is felt every day by almost everyone. He was also one of the most humble. When conference hosts introduced him as a database guru, he would simply say "I'm just a programmer."

Jim also had a great sense of humor. In his article referenced above in which he described the rapid increase in transaction processing performance, he concluded with, "The next article in this series, scheduled for April Fools' Day 2025, will show that a $1 wrist watch can run the world economy as of 1990. Since cell phones are already at a gigabyte of storage and approach a GHz processor, such an article may be possible – we hope we are around to write it."

He was as bullish on massive databases as he was on transaction-processing performance. In a 1999 posting on the Microsoft web site announcing his receipt of the Turing Award,[6] he was quoted as saying, "'A second goal … is to turn computers into virtual librarians that can store and summarize massive amounts of data and present this data to users in a convenient way. Someday, you're going to have a personal digital assistant that is watching and listening to everything that is going on around you, It's going to have a few petabytes of information to work with. And you're going to say, 'I was talking to this guy, he was somewhere in San Francisco, he had a beard, it was about 30 years ago. Now what did he say?' And this thing is going to hunt around and come back and give you a clip … telling you this. It sounds like science fiction, but it's actually doable.'"

If Jim is really gone – and many still do not accept that – he leaves behind his wife, Donna, his daughter, Heather, and his grandson, Sam. But as Steve Silberman concluded in his wonderful article on Jim, "Even his disappearance proved a sort of mentoring, providing a template for networked search-and-rescue that might save countless lives in the future."

[6] http://www.microsoft.com/presspass/features/1999/05-14turing.mspx