



Inktomi: Scaling the Internet

"The traffic on the Internet is exploding, growing at 10X every year—any machine designed by Moore's Law will run into a wall. Inktomi's architecture was designed to sit in the middle of the traffic. We map the heat."

Kevin Brown
Director Marketing, Inktomi

Veteran software executive and new CEO of Inktomi, Dave Peterschmidt was pondering the company's upcoming formal strategic planning meeting, its first such session. Founded in February 1996, the company was only six months old. Peterschmidt had joined in July 1996, just two months ago, bringing years of experience from the software industry. In his previous job, he was the Chief Operating Officer at Sybase and had over 6000 people reporting to him. When he joined Inktomi there were only 12 employees. In addition, the majority of the Inktomi team had deep technical backgrounds but only limited work experience.

The company was facing a critical decision for the two-day offsite: How could it best leverage its core technology? As Peterschmidt realized, "We had a core technology that could provide a platform to write scalable applications. We could solve big problems. The Internet market was exploding all around us. We had aspirations of playing a big role in this emerging market. How could we build an organizational structure and process to match the technology and context?"

Peterschmidt began drafting some thoughts in preparation for the session.

Company Origins¹

In 1994, the US government was looking for a means of achieving supercomputing power without paying supercomputing prices—costs well beyond the budget of many government labs. The Department of Defense's Advanced Research Projects Agency (ARPA) provided funding to UC Berkeley professor, Dr. Eric Brewer, to explore possibilities beyond traditional high-performance and proprietary massively parallel approaches. The charter: To obtain supercomputing performance from a cluster of commodity PCs and workstations, a "network of workstations" (NOW).

¹ Adapted from <http://www.inktomi.com/about/index.html>

Doctoral candidate Richard Bergin prepared this case under the supervision of Professors Marco Iansiti and Myra Hart as the basis for class discussion rather than to illustrate either effective or ineffective handling of an administrative situation.

Copyright © 1999 by the President and Fellows of Harvard College. To order copies or request permission to reproduce materials, call 1-800-545-7685 or write Harvard Business School Publishing, Boston, MA 02163. No part of this publication may be reproduced, stored in a retrieval system, used in a spreadsheet, or transmitted in any form or by any means—electronic, mechanical, photocopying, recording, or otherwise—without the permission of Harvard Business School.

The “network of workstations” was designed to dramatically lower the costs of supercomputing. First, NOW would use off-the-shelf workstations, initially a SUN Solaris workstation. Second, NOW would use standard operating systems, SGI UNIX, Digital Unix, and Windows NT. **Exhibit 1** provides a brief summary of the NOW project.

Three experienced professors from UC Berkeley were the principal researchers for the NOW project, joined by a dozen graduate students. The project was generously funded through ARPA, the National Science Foundation, and corporate sponsors including Intel, Digital Equipment Corporation, and Sun. Two members of the team were Brewer and Gauthier. Brewer was a new professor in computer science, with a freshly minted Ph.D. from MIT. Gauthier had just turned 23. Before entering the Ph.D. program at Berkeley, he received his undergraduate degree from Dalhousie, Canada, and he was awarded the Canadian national science prize.

It was clear to all that these core technologies could be applied to a broad range of emerging applications. Research was already underway for Fast Communication (such as active messaging and systems area network mapping), Distributed Operating Systems (such as Cooperative Caching, Network RAM), and Scalable File Service. The main NOW projects were focused on the “glue” that held the cluster of workstations together. See **Exhibit 2** for a description of a NOW cluster compared to a traditional server architecture.

Brewer and Gauthier wanted to find a commercially viable application, something that was highly scalable. At the time, Lycos, Webcrawler, and Yahoo! had gone commercial and were drawing a lot of attention in the press, and so the two men settled on a search engine application for the emerging World Wide Web. Given the rapid expansion of users and documents on the Internet, the search application was an ideal showcase for the technology, they believed.

Gauthier wrote an application known as “Project Inktomi.”² Posted on a public server at Berkeley, the application quickly won a loyal following on university campuses. The Inktomi prototype search engine was designed from the beginning to economically scale with the rapid growth of the Internet, without compromising high performance, reliability, or fault tolerance. The resulting architecture, based on clusters of inexpensive commodity workstations or building blocks, represented an important breakthrough: superior price-performance, virtually unlimited scale, and the flexibility to deploy capacity on a pay-as-you-go basis.

Advantages of Clustering

According to Gauthier, “The main difference between traditional server farms and clustering is that the workstations in a server farm do not cooperate with one another. In contrast, the clustered workstations are tightly coupled. Usually communication between machines is considered the death of any system. While working at Berkeley on the ARPA Network of Workstations project, Brewer and I worked on solving this problem for web searching.”

Gauthier explained the four advantages of clustering:

1. **Low cost.** A clustered system is typically 2-5 times cheaper for the same level of performance. Inktomi’s clustered workstations are Sun Solaris machines that cost \$15,000 for a 400 MHz processor, 128 MB of RAM, and 9 GB of disk capacity. Each machine in the Inktomi cluster stores and searches a proportional slice of the database. In contrast, other search engine companies typically utilize expensive computers that are limited to a fixed capacity. Each machine

² The company's name, pronounced "INK-tuh-me," is derived from a Lakota Indian legend about a trickster spider character. Inktomi is known for his ability to defeat larger adversaries through wit and cunning.

duplicates a copy of the entire database. However, the WWW is so large that it now takes 300-500 GB to store a good portion of it. Thus workstations have to be custom built at a cost \$50,000 per unit.

2. **Absolute scalability.** There is no absolute limit to the growth in Inktomi's system size. Capacity can be added without diminishing system performance. The reason that this is important is that the World Wide Web (WWW) is growing exponentially with the number of documents that are connected (the number of URLs) and the number of users that have accounts. While the total size of the WWW was not known accurately, estimates from IDC placed its size at 160 million URLs and the growth rate at one million new URLs per day. Inktomi's cluster had 35 million URLs when the firm was launched in February 1996 and it had already grown to 54 million by June 1996. Their database only took 0.25 seconds to search. This represents approximately 34% of the WWW. To handle the increase in number of documents and requests, in 1996 Inktomi had increased the number of server's in its cluster from 4 to 20.

In contrast, AltaVista had run into a wall, having reached a limit to continued growth due to diminishing returns to scale. The largest Symmetric Multiprocessor (SMP) computer produced by AltaVista's parent, Digital Equipment Corporation, could only store 30 million URLs and required several seconds to search the database. The firm duplicated the database across each machine every night. To increase the size of the database so that a workstation can search, the company has to order ever-increasing systems to add to the network. AltaVista has divided its database into four SMP machines, each with 500 gigabytes hard drive, to alleviate the problem, but this was only a temporary solution.

3. **Incremental scalability.** Inktomi's modular cluster can be grown incrementally, in small bites of investment. When you are short on cash and growing rapidly it is important to match capacity additions and capital spending to growth in revenue. The capacity of the cluster of workstations can be increased by simply adding one more Sun Solaris workstation. The workstations are bought straight off-the-shelf with no need to customize. The processor and the 9 GB hard drive are sufficient to store and search over one million URLs within 250 milliseconds.
4. **Fault Tolerance.** The system is very robust and continues to run even if there is a partial failure.

Founding the Company

Based on the success of their prototype application and the promise of this new scalable architecture, Mike Stonebraker (professor of computer science at Berkeley and founder of Ingres Software) advised the team to form a company with a few other Ph.D.'s and an exclusive technology license from UC Berkeley. They thus founded Inktomi Corporation in February 1996 as a limited liability corporation (LLC). Brewer was the interim CEO and Gauthier the Chief Technical Officer. They were joined by several other Ph.D. students and Kevin Brown, an "MBA type" from the Haas School of Business, hired to start marketing.

The first major decision the fledging company faced was what business model to adopt. Gauthier described the three alternatives they discussed:

1. **OEM infrastructure providers to the search engine companies.** At the time, AltaVista would be our primary competitor. They provided search services to

companies such as Yahoo! Based on AltaVista's success, Lycos, Infoseek, and Webcrawler began to offer commercial services to the other search engine companies. We liked this option because it would allow us to focus on our strengths and avoided our weaknesses. There would be no consumer branding and no end-user marketing. The real advantage was that we would not compete with our customer. Our competitors could not scale, and thus competed with their customers.

2. **Builds own brand.** This was a serious undertaking. Eric and I had no experience in this realm and we would be the seventh new entrant into the space. We would have to build not only the brand, but also an advertising sales department. At the time there were no "Double-Clicks" who could sell advertising on your behalf.
3. **Buy rights to an existing consumer brand.** We toyed with the idea of purchasing a famous fictional character. We considered the third alternative only briefly. We had the option to license a major consumer brand. The brand appealed to a broad demographic, and had lots of good "Internet/web" synergy. It would have had the following problems: it would have been expensive for us to license this brand; we would have become a brand/portal instead of an OEM technology provider.

The founders debated the options during December 1995 and January 1996, in the hallways of Berkeley's computer science department, and decided to go with the first option. Brown described the decision:

Our core competence is in technology. We understood that the computer market was no longer dominated by vertically integrated players (as it was in DEC's days)—players moved in to dominate particular segments and leverage their competence across a greater volume. For example, Intel dominated microprocessors while Microsoft owned the operating system and software applications. The Internet market will evolve in the same way. Where could Inktomi play? We decided to get "media partners" on the front end and concentrate on the back end. Our goal was to be the ubiquitous infrastructure provider.

Gauthier described how the model operated:

We have a very crisp break between our company and the customer. The OEM model works through the Inktomi Data Protocol (IDP), which submits a request from a search front-end page. The request is processed in our machine room in San Mateo, and the top 10 matches are sent back within 250 milliseconds. The customer then can take the data string, format the page and add advertising.

This system granted enormous freedom to the customer—in essence they could scale their portal sites. There would be a division of labor according to the expertise of the Inktomi team and the customer, and there would be a sharing of the risk and reward. See **Exhibit 3** for a diagram of the Inktomi-OEM business model.

Brewer's cousin, David—a CPA and a lawyer—incorporated the company. The small team signed a lease and began transforming the technical prototype into a production-ready system. The office was in the second floor of an old building, located above the Sees candy store on Shattuck Avenue, a dodgy part of Berkeley. Next door was a telemarketing company who referred to themselves as "Punks on Pizza." Everyone at the firm next door, according to Brown, had "two-foot high orange hair and bones through their noses."

Financing

Inktomi's search engine had drawn strong grass roots interest at colleges across the United States and with developers in Silicon Valley. It also had received very favorable press coverage. For example, in the December 1995 issue of Wired magazine, Inktomi was listed as a "wired" search engine³ and competitor Lycos was described as "tired." In May 1996, the same publication printed a review of search engines currently available to index the web, featuring Inktomi prominently⁴.

As soon as the small venture settled into its small, grungy offices it began receiving cold calls from venture capitalists, angels, and large corporations offering to help provide some of the funds for a commercial venture. The second challenge for the founders was which source of funding they should pursue.

They sought the advice of David Brewer and other Berkeley associates who had recently formed ventures. It seemed an easy choice. Angels would allow them to retain control and would also provide far less dilution. On the flip side, the risk of angels was they would not provide commercial advice for structuring the business or the financial advice on how to prepare for an initial public offering. Venture capitalists would remove some control and dilute the founders substantially.

The venture raised \$3 million of private funding from a group of high net worth individuals from the Southeast. They had no technology background, but wanted to get into an Internet play.

Commercial Potential

Inktomi landed its first partner, HotWired, in March 1996. HotWired had been launched in October 1994, as an "online-cousin" to Wired magazine. The deal was a real coup for the young company. Wired had "named, defined, then (literally) trademarked the Revolution that it chronicled, a revolution that was whipping through our lives like a Bengali typhoon... HotWired pioneered and soon popularized notions like site membership and weekly informed member E-mailings. Its *Wired*-inspired look and *attitude* meanwhile informed the content and presentations of countless other Web publications that came after it. But of all its innovation, perhaps the most important (...was that) it *broke the Web* as an advertising medium. The banner advertisement, which underpins the revenue hopes of almost every commercial content company of the Web today, was a HotWired creation."⁵

HotWired and Inktomi teamed to build HotBot, the world's first truly scalable search engine. Inktomi would provide the product, the technology, and the service support for the "crawler" that would search the Web and the search cluster that would search the database and reply with the top 10 sites. HotWired would design the user interface, the branding, the marketing, and the ad sales. In the first incarnation of the deal, the two companies would share the advertising revenues.

The service launched in May 1996, and soon generated a significant following and numerous awards from magazines such as PC Computing and PC Magazine. According to Brown, however, all was not smooth:

It took us only 90 days to go live with our first product. We received a licensing fee and a portion of the advertising revenues in return for maintaining the search engine. However, the learning curve was very steep and we got off to a rocky start. Strange things come across the Web. The speed and performance were under-whelming. It was not enough to have a fast application located in a

³ Wired/Tired. Wired Magazine, December, 1995

⁴ Steinberg, Steve G. Seek and Ye Shall Find (Maybe). Wired magazine, May 1996

⁵ Reid Robert (1997) Architects of the Web. Pages 281-282, John Wiley & sons, New York

machine room in Berkley. The bandwidth was not reliable. One slow router between you and the rest of the world could kill the response time. Another concern was whether search engines would be a viable business. There was a lot of press about the shift to “portals” and that “search” would only be one of the attractions to the site.

Building the Organization

Aware of their lack of industry experience, the two founders stepped up their efforts to find people that had already “done it.” Brian Totty was on Brewer’s list, and he became the first person with industry knowledge to join in February 1996. He had studied at MIT with Brewer and then went on to earn a Ph.D. at University of Illinois. Upon graduation, he joined Silicon Graphics and rose to become one of SGI’s supercomputing scientists.

The second “experienced” person to join was Peterschmidt, who became the CEO, replacing Brewer. As Peterschmidt recalled:

I had just stepped down from being the COO of Sybase, where I had grown revenues from \$100 million to \$1 billion in four years. As a favor for a buddy, I went to talk to Brian Totty and Eric Brewer. I then realized that this was a deep software infrastructure play. Inktomi has the potential to scale with the Internet. It had the right architecture and an environment that allowed applications to be written. This is the Cisco equivalent of Internet software. I joined the company in July 1996.

The clear challenge for me was to build an organization that could scale with the technology. This was very heavy technology and the team was smart enough to crack any technology problem on the Web. We had to build solid teams of people around the core processes, like technology, sales and marketing, and finance. First, because of the large number of technology graduates, we had to build some development teams. Once the core teams were established, we could pile on people.

The firm’s culture was formed by the academic traditions of Berkeley: intellectual curiosity and a willingness to change. Already Totty and Gauthier had begun to spin off ideas that had commercial opportunity. Brown recalled the first summer prior to Peterschmidt’s arrival; “Our first summer was spent scrambling to make our search engine robust. Teams of people were constantly reassigned to the next hot spot. Meanwhile, Brian started seriously forming the cache server idea [described below]. It was pretty chaotic.”

Introducing a Technology Development Process

One of the problems Peterschmidt discovered at Sybase was that the company had grown very quickly but its processes overall in the organization, and particularly in product development, had not grown commensurately. One thousand people would be doing things in development in ways that had been done when only a dozen people were there. Not surprisingly, the result was inefficiency, missed schedules, and bad quality.

In his role as Sybase’s COO, Peterschmidt brought in a small technology consulting firm — Pitiglio, Rabin, Todd and McGrath (PRTM), whose specialty was in advising companies on how to set up product development processes. The approach emphasized the whole product process for creating a pipeline that would be managed from idea conception through delivery. Peterschmidt had experience in trying to retrofit these ideas into Sybase. The change process was hard and had, in a sense, come too late. The organization had already developed its own routines and methods for product development. He determined that this would not happen at Inktomi:

A month after I joined Inktomi I brought in PRTM to help design the product development process — to avoid the problems that developed at Sybase. Of course, we had to adapt their traditional “stage gate” process to Internet speed. We found that the window of opportunity opens and closes very quickly and there are too many opportunities to concentrate force and mass on. We had to avoid hip shooting. I had seen too many companies that were 40 miles wide and only one micron thick. At Inktomi, an idea will be debated for a couple of days, and then someone will say, “let’s write a phase 0 spec.” This outlines the idea, the market, and the technical feasibility. We need this when developing fairly heavy infrastructure software.

The [company] was fairly skeptical of hiring expensive consultants. With the exception of two people, no one on the development team had any industry experience. They all were new graduates, many with advanced degrees, but no commercial experience. But one characteristic that I noticed at the first meeting I had with people here was that in addition to being very smart, they were very open to ideas. And so if there was an area or something that they didn’t know about, they were curious to learn about it, including how do commercial companies develop software products. There was a willingness to take this input, a willingness to experiment, to try.

Added Brown:

Some folks were a little confused about the value of consultants who had not worked on Web applications, but we thought we might as well give it a shot. We were having a fairly tiring summer starting our first commercial service and had just gotten an inquiry from Nippon Telegraph (NTT) to provide search services in Japan. One thing was sure, we did not want to just sit around here in dirty jeans, scrambling from one crisis to the next. No one questioned Dave’s authority. He had been in charge of 6,000 people and in total we were only 15. There was unanimous willingness to march to his orders.

The objective of the product development process was to take an idea from a rough concept to reality in an economical way:

1. The first step was to widen the search process. This would allow Inktomi to expand its knowledge base and increase the number of new product ideas. At weekly pizza-lunch sessions, the team would kick around concepts. If any of the ideas showed promise, they would be moved into “phase 1.”
2. At that point, the focus would be narrowed, to concentrate scarce resources on a few promising ideas. A cross-functional core team would be assembled to investigate market potential and prepare a Marketing Requirements Document (MRD). Against these requirements, the team would also draft an outline of the resource needed to code the product and complete a system test. This would be recorded in the Product Requirements Document (PRD).
3. The final challenge would be to ensure that the project was completed on time and delivered the specifications promised. This required a very clear and routinized decision-making process. An executive team would approve the ideas when they moved from concept to development and give the final go-ahead to launch after system testing.

Preparing for the Offsite Meeting

By late August, Peterschmidt decided that the time had come to hold a two-day offsite meeting to discuss Inktomi's future strategy—in particular, how the company's core technology could best be leveraged. Weighing heavily on his mind was whether the search business was going to survive, either at all or in its present form.

Inktomi was the seventh entrant in an extremely visible segment of the Internet industry. The company's business model to be an OEM to the search engines directly competed with AltaVista backed by Digital's deep pockets. Even more significant was the fact that "search sites were morphing into what came to be known as portal sites," recalled Dennis McEvoy, who subsequently joined his Sybase colleague Peterschmidt at Inktomi, as VP Engineering and Product Development. McEvoy continued:

That meant, these sites had 15 or 20 applications—services—they were offering to users, to try to attract and retain eyeballs. "Search" was vulnerable to becoming a commodity; it would no longer be the only reason for going the site—so that would make it a very good candidate for outsourcing. So we wondered if there were opportunities in these other services for Inktomi—what we could supply. Or, were there opportunities for Inktomi to partner with other suppliers to even build a coalition so that someone could go and develop one of these sites.

Inktomi began to recognize, in other words, that if it had just "search" to offer, then the customer it was selling to would end up as a general contractor putting all the applications together. Or the customer might go to an Excite, for example and take from them the multiple services they had up and running.

The Web Caching Opportunity

In fact, the two founders, Brewer and Gauthier, had always seen Inktomi as more than just a search engine company. And in June, the team had come up with a new idea that could leverage the network of workstations (NOW) modular cluster technology: a network cache application. This could become a second business for the company.

Web caching was an emerging solution to solve Internet performance and scalability problems, which were a direct consequence of both exploding Internet use and a limited supply of bandwidth. The key issue was moving data. The Internet was originally designed to work using a voice network. However, the volume of data traffic that was sent through the network soon overtook voice traffic. As well, data traffic exhibited radically different characteristics than did voice traffic. Usage patterns were not predictable, as connections (between users and sites) could last for hours—or minutes. Traffic was not periodic, moreover, and so its distribution would not "smooth out" over an extended period. And even when heavy demand might be predicted, accessing data generally still meant frustrating delays for users, Internet Service Providers, and telecommunication companies alike.

The backbone of the Internet was controlled by global telecommunications centers that also controlled how quickly infrastructure-related Internet technologies would be deployed. The idea was for Inktomi to develop technology to help customers expand their networks incrementally by providing network cache software that stored copies of frequently accessed Web pages "closer" than the original site—thereby eliminating the need for users to go back across the backbone itself. Essentially, Inktomi would develop software that would manage requests for those Web. Ideally, this would reduce congestion, save customers money, and allow them to optimize their network architecture through enhanced performance and efficiency.

Customers, in this case, would be network operators, telecommunications companies, cable networks, and large ISP's—by and large a different group from those Inktomi's search engine technology targeted. It would also have to build a sales organization to accommodate these large companies—as well as establish partnerships with OEMs, VARs, and systems integrators to manage the channel. In addition, Inktomi would be entering a new realm of competition: Cisco Systems, Netscape, Microsoft, Novell—all these heavy-hitters were pursuing network cache solutions aimed at the corporate market.⁶

Of critical importance, pursuing the cache business implied another business model: Revenues would be generated differently from the search engine business.

Inktomi's search engine business received a cut of advertising revenue and per-page view pricing. In contrast, the cache technology idea would follow an enterprise software license model: generating revenues through software licenses, maintenance and upgrade fees, and consulting fees. Sales would involve a pilot phase before customers deployed network caching in their production environments. The majority of the revenue would be from software license fees based on the number of CPUs deployed in the network. In addition, customers could be charged for a nominal quarterly maintenance fee. Upgrade fees could be assessed for major software releases. An advantage for Inktomi, given its NOW technology on which caching would be built, was that larger caches generally worked better than smaller caches, at least to a point: The more data a cache contained the higher the likelihood a user's request could be served directly from the cache.

The cache technology idea had great appeal according to Brown:

The team, especially Dave and the founders, thought big. They had a clear vision that this company could leverage the technology in multiple directions. If we want to grow big, we have to target big companies.

The team played around with other applications that involved big databases and high user traffic. Ideas considered at the time were an Internet ad server, some data mining applications, and a dynamic publishing system. Only caching really fit well in the sense that it demanded large scale (which is good for Inktomi's clustering/parallel technology) and involved Internet-centric knowledge.

The combination of how "search" was evolving and the opportunities provided by caching—both in leveraging Inktomi's core technology *and* as a potential application for other large Internet companies—created the possibility of a new business for Inktomi. That would establish an additional revenue stream and help manage earnings—particularly if the so-called advertising model that was to generate revenues on the Internet in fact didn't.

However, there were a number of strong disadvantages. The cache market was small (less than \$20 million in 1995) and dominated by public providers of free software code. Most current implementations relied on publicly available code, such as Harvest, Squid and the CERN caches⁷. At the same time, the market was predicted to grow explosively. Given the rapid growth of the World Wide Web, it was estimated that Internet Service Providers would be forced to migrate to commercially available cache software that would support both the latest releases of Hypertext Transport Protocol (HTTP) and further development of the software.

⁶ Source: Rimer, D.H. [Inktomi Corporation](#). Hambrecht & Quist Inc. July 6, 1998

⁷ Source: Wade, James P., Jennifer St. Germain, and Thomas Bain, [Inktomi Corporation, Initiated research coverage of Pioneer in Scalable Network Applications](#). B.T. Alex. Brown, July 21, 1998

Organizational Implications

The market and competitive risk, however, paled in comparison to the organizational challenge a second business implied. Inktomi was only six months old and less than half of the 15 people at the firm had any work experience prior to joining the company. The search engine business, whatever its longer-term future, was still booming, and the earlier difficulties the company had experienced launching the first customer HotWired were being overcome. Inktomi was fast establishing a premier reputation in this arena—entering a different business could blur its image.

Then there was the very recent introduction of a new product development approach. While the expressed reaction was positive, dealing with this highly structured process might be hard when push came to shove. Inktomi was a traditional startup, a pizza-and-Coke sort of culture. Though greatly modified from what had been implanted at Sybase, with its thousands of employees, the PRM methodology involved core project teams, sign-offs at various development points, formal procedures, and lots of meetings.

Creating a new business itself, irrespective of the new processes, raised large organizational issues. Rough estimates indicated that developing the cache technology software would take about 18 months. But it was not simply developing the product. A sales organization would have to be built, one that would address very different customers—and potential partners—from those targeted by the search engine business. Whom would the company hire for this? Berkeley/MIT Ph.D.'s in computer science? Veterans from other companies? Who would head a cache technology business? How would all these people fit into the existing culture? Was there an existing culture in a company that was only six months old?

Peterschmidt began jotting down notes for the strategy meeting, thinking to himself:

If we are to go into hypergrowth, as a second business implies, we can't get ourselves into the situation of having executives with no experience. But, how do you mix the wisdom of experienced industry veterans with the youth and vigor of recent college graduates? What are our criteria for hiring new people?

The whole strategy of getting into a second business feels incredibly risky. We might have to cut the core development team [doing search engines] in half. The whole thing could blow up. A split culture could develop, the "A team" against the "B team." What would provide an integrating mechanism for the company? Our technology ties the pieces together, but is this big enough for the culture? The image we would present to the outside world would be very confusing, as well. Should we spin off a separate company rather than keep a second business in house? Should we wait and simply pursue the search engine business harder and let the new product development process take hold, pursuing the cache business later? Or can we grow the company by dividing it into parts right away? How do we scale the Internet?

Exhibit 1 Summary of the Network of Workstation (NOW) project⁸

The Berkeley Network of Workstations (NOW) project seeks to harness the power of clustered machines connected via high-speed switched networks. By leveraging commodity workstations and operating systems, NOW can track industry performance increases. The key to NOW is the advent of the killer switch-based and high-bandwidth network. This technological evolution allows NOW to support a variety of disparate workloads, including parallel, sequential, and interactive jobs, as well as scalable web services, including the world's fastest web search engine, and commercial workloads, such as NOW-Sort, the world's fastest disk-to-disk sort.

Project Overview

The Berkeley NOW project is building system support for using a network of workstations (NOW) to act as a distributed supercomputer on a building-wide scale. Because of the volume production, commercial workstations today offer much better price/performance than the individual nodes of Massively parallel program's (MPP's). In addition, switch-based networks such as ATM will provide cheap, high-bandwidth communication. This price/performance advantage is increased if the NOW can be used for both the tasks traditionally run on workstations and these large programs.

In conjunction with complementary research efforts in operating systems and communication architecture, we hope to demonstrate a practical 100 processor system in the next few years that delivers at the same time (1) better cost-performance for parallel applications than a massively parallel processing architecture (MPP) and (2) better performance for sequential applications than an individual workstation. This goal requires combining elements of workstation and MPP technology into a single system. If this project is successful, this project has the potential to redefine the high-end of the computing industry.

To realize this project, we are conducting research and development into network interface hardware, fast communication protocols, distributed file systems, and distributed scheduling and job control. The NOW project is being conducted by the Computer Science Division at the University of California at Berkeley.

The core hardware/software infrastructure for the project will include 100 SUN Ultrasparcs and 40 SUN Sparcstations running Solaris, 35 Intel PC's running Windows NT or a PC UNIX variant, and between 500-1000 disks, all connected by a Myrinet switched network. Most of this hardware/software has been donated by the companies involved. In addition, the Computer Science Division has been donated more than 300 HP workstations which we are also planning on integrating into the NOW project.

We propose to build hardware and software to enable a network of workstations (NOW) to act as a single large-scale computer. Because of volume production, commercial workstations today offer much better price/performance than the individual nodes of MPP's; in addition, switch-based networks such as ATM will provide cheap, high-bandwidth communication. This price/performance advantage is increased if the NOW can be used for both the tasks traditionally run on workstations and large programs. We hope to demonstrate a practical 100 processor system in the next few years that delivers at the same time (1) better cost-performance for parallel applications than a massively parallel processing architecture (MPP) and (2) better performance for sequential applications than an individual workstation (by using more of the resources of the network). If projects like NOW are successful, they have the potential to redefine the high-end of the computing industry.

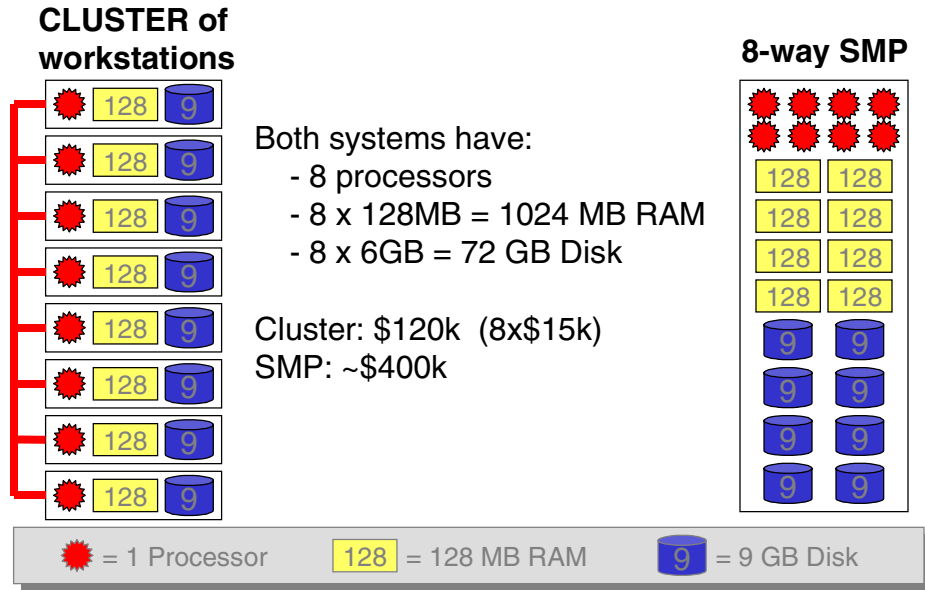
⁸ Source: <http://now.cs.berkeley.edu/>

Exhibit 1 (continued) Summary of the Network of Workstation (NOW) project

To realize the potential of NOWs, we need to move two MPP technologies into the workstation community: low latency networking and global system software that treats a collection of processors, memory, and disks as if they were a single machine. Our approach is to leverage off-the-shelf technology as much as possible — workstation hardware, standard workstation operating systems on each node, and local area network ATM switches. To this, we will add communications protocol software and a global system layer that together provide low overhead communication, a single view of operating system services across the cluster, parallel file I/O, and robustness to individual node failures. We will demonstrate our results by using our system for the everyday computing needs, both sequential and parallel.

Exhibit 2 NOW cluster compared to traditional server architecture.

Cluster vs. SMP:

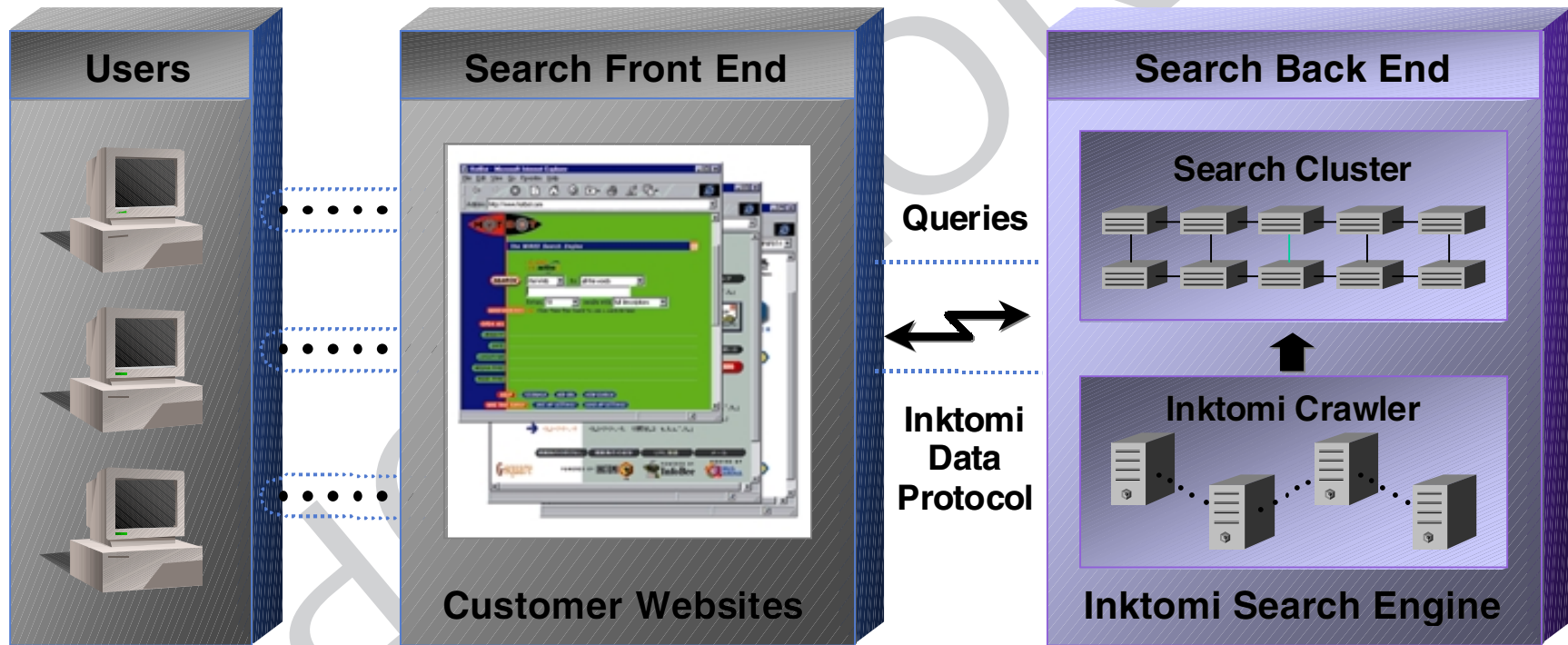


Source: Company documents

COPY

Exhibit 3 Inktomi-OEM Business Model

Inktomi Search Engine



Source: Company documents