

**Jeff Darcy**  
**50 Spring St**  
**Lexington MA 02421**  
**(781) 640-4420**  
**jeff@pl.atyp.us**

### **Skills Highlights**

- Large-scale systems, including both HPC and enterprise types at 1000+ nodes
- Parallel and distributed filesystems
- High-performance networking
- SCSI and Fibre Channel storage systems
- Advanced protocol and code verification technology (e.g. Murphi, Coverity)
- Various UNIX (Linux, Solaris, Unixware, AIX, ...) and NT internals
- Highly available and real-time systems

### **Work History**

*01/07 to 05/09*                      *Storage Architect*                                              *SiCortex*

Led practically all software efforts related to storage (especially Lustre) and many in IP networking, for SiCortex's systems from 72 to 5832 processors. Conceived, designed, and implemented FabriCache Lustre-over-RAM functionality, capable of >40GB/s throughput on SiCortex's larger systems. Implemented a flexible API for our unique communications fabric, based on PVFS2/Argonne's BMI, demonstrating 1-2us latency for small messages and ~3.2GB/s (out of 3.5 physical maximum) application-to-application throughput for two nodes. Participated in many pre- and post-sales external engagements, and consulted on various actual or potential strategic relationships (e.g. Isilon, Panasas, Fusion I/O). Wrote or debugged other code throughout the entire system in my usual role as "developer/debugger of last resort" for many of the group's most difficult problems.

*09/02 to 12/06*                      *Developer and Product Architect*                                              *Revivio*

Initially hired as architect for one group out of three implementing a Continuous Data Protection appliance, later became architect for the entire product. The software runs on a real-time version of Linux, translating Fibre Channel on both initiator and target sides plus InfiniBand internally. Defined overall data flow, fault/configuration event handling, performance and availability strategies. Participated in defining overall development process, technology roadmap (including vendor evaluation), and schedules. Deployed various tools and technologies to improve overall software quality. Assumed direct responsibility for various development and debugging tasks that involved multiple subsystems and/or had proven resistant to others' efforts.

*05/98 to 09/02*                      *Principal Software Engineer*                                              *EMC (via Conley acquisition)*

Hired at Conley to prototype and design next-generation virtualizing and self-configuring storage system. After acquisition, became one of three co-designers of what became the HighRoad cluster filesystem, and implemented the first UNIX (Solaris) client. For last year and a half, designed and prototyped a system to extend HighRoad functionality to global scale via a hierarchical caching block subsystem. Also served as

EMC liaison to various academic groups including OceanStore at UC Berkeley and Parallel Data Lab at CMU.

10/97 to 05/98

Principal Software Engineer

Mango

Implemented portions of a distributed filesystem based on software-implemented coherent shared memory in a LAN environment. Participated in design of a second-generation product intended to improve performance and allow operation in WAN environments.

02/96 to 10/97

Senior Software Engineer

Dolphin Interconnect Solutions

Designed, implemented, and analyzed performance of NDIS (NT) and DLPI (SCO UnixWare) network drivers using Dolphin interconnect products. Implemented OS-specific and hardware-interface code in UnixWare and NT device drivers for two of Dolphin's shared-memory interconnect products, as part of two separate product teams in Framingham MA and Oslo (Norway). The PCI-SCI Bridge provides non-coherent shared memory for a wide variety of systems; the SMC (Shared Memory Cluster) hardware adds coherency support and connects directly to the processor bus of an Intel "SHV" system or near-equivalent. Both products use IEEE standard Scalable Coherent Interconnect protocols with link speeds of either 1.6 or 4.0 Gb/s.

05/92 to 02/96 Senior/Principal Software Engineer Clam Associates (later Availant)

Designed, implemented, and debugged AIX disk drivers providing transparent path switchover for dual-controller SCSI disk arrays (IBM 7135/110 and CLARiiON). The success of these products, on which I worked alone approximately 75% of the time, made this product category one of Clam's most profitable.

Led design effort for a major release of Clam's HACMP product, providing high availability for up to eight IBM RS/6000 nodes. Major design elements included a modular "cluster manager" responsible for monitoring and invocation of event-related scripts, and a fully distributed lock manager which incorporated distributed deadlock detection and lock migration to avoid the traffic asymmetry characteristic of similar products. Specific contributions:

Designed, verified and implemented algorithms for membership, consensus and synchronization in the presence of faults.

Designed and implemented a robust communications subsystem for the distributed lock manager.

Modified Clam's distributed lock manager to run in the AIX kernel instead of user space, avoiding context-switch and copying overhead. These changes improved lock manager latency approximately tenfold.

08/89 to 05/92

(Senior) Software Engineer

Encore Computer Corporation

Added parallel/asynchronous I/O facilities to UMAX V, Encore's SVR3-based commercial UNIX, to support Oracle on the "Infinity 90" platform, composed of Series 91 computers linked with Reflective Memory technology (the precursor of Digital's Memory Channel).

Supported and enhanced gang scheduling and other "guest operating system" features in UMAX V on the Series 91, allowing a second operating system to run concurrently with its own processors, memory, and exception handlers. The first targeted guest was microMPX, a version of Encore's venerable MPX real-time operating system, and I worked very closely with that group to attain the low interrupt and context switch latencies on which Encore's business depended.

Participated in the port of UMAX V to the then-new Series 91 computer (4 MC88K processors, VME bus) from the older Multimax (2-32 NS32x32 processors, proprietary Nanobus backplane). Particular areas of responsibility included the networking subsystem and interprocessor interrupt reduction strategies.

Participated in an upgrade of Encore's BSD-based commercial UNIX variants from 4.2 to 4.3 Tahoe on the Multimax platform, with a particular emphasis on networking and NFS/YP (it wasn't NIS yet) code.

4/88 to 08/89

Systems Engineer

Technology Concepts International

Supported, enhanced, and ported versions of TCI's CommUnity product, a DECnet implementation for Macintosh, IBM PC, various UNIX flavors and occasional special-purpose environments such as embedded operating systems running on network hardware. Particular areas of specialization included the file transfer engine and email gateway support.

#### **U.S. Patents**

6389420 (with four others at EMC, granted 2002): File manager providing distributed locking and metadata management for shared data access by clients relinquishing locks after time period expiration.

7124249 (sole inventor at EMC, granted 2006): Method and apparatus for implementing a software cache.

7239581 (with one other at Revivio, granted 2007): Systems and methods for synchronizing the internal clocks of a plurality of processor modules.

7325097 (sole inventor at EMC, granted 2008): Method and apparatus for distributing a logical volume of storage for shared access by multiple host computers.