



# OpenSolaris Storage Summit 2008

## Project Celeste

**Glenn Scott**  
**Sun Microsystems Laboratories**

# What Is Celeste?

- High Availability Data Store
  - ▶ Distributed,
  - ▶ Peer-to-peer,
  - ▶ Masterless,
  - ▶ Mutable

# High Availability Data Store

- Celeste strives for availability of data over long periods of time, without external backup and little human intervention.
- Clients create, write, read, update, and delete data files.
- Data files have basic access control.
- Celeste is *not* a file system.
  - > But can build a file system on top of Celeste.

# Celeste is: Distributed

- Arbitrary combination of concentrated and dispersed topology of nodes.
- Nodes are whole data centers, servers, desktops, laptops, ...
- Nodes contribute storage and processing.
- Nodes may intermittently participate.

# Celeste is: Peer-to-Peer

- No central authority.
- Any node may be used for client interactions.
  - > Clients can simultaneously multiplex operations across more than one node.
- The system uses a DHT for messaging routing.
- DHT routing tables are dynamically updated.
- Nodes choose neighbours based on “reputation.”

# Celeste is: Masterless

- No node is ever entirely trusted.
- Nodes are expected to fail at any time.
- Malicious nodes are expected.
- No node is individually responsible for an operation.
- Synchronous operations are achieved through quorums of a set of nodes.
- Quorum protocol detects cheaters and liars.

# Celeste is: Mutable

- Arbitrarily update data files at any time.
- Every update is recorded as a delta.
- Every access operates on the latest delta.
- Previous versions are accessible to clients.
- Files can be deleted (real delete).
- Deletion works despite intermittent connectivity and leaves behind proof.

# Celeste is: More

- Celeste objects can proxy the physical world.
  - > People, Places, Things.
- Celeste objects have programmatic behaviour.
  - > Objects respond to messages, change their state, and may invoke other objects.
- Example uses:
  - > Automatic transcoding of data as it is read.
  - > Dynamic data generation.
  - > Anticipating data requirements by correlating physical world activities.

# Celeste

- [www.opensolaris.org/os/project/celeste/](http://www.opensolaris.org/os/project/celeste/)