Current Efforts

The major ongoing Mero activity is a T1 project. T1 is a horizontally scalable object store, providing fault tolerant block device tuned for a few selected applications: Lustre, Hadoop, *etc*. Production release of T1 is expected at the end of 2011, but this estimate is subject to change due to changes in the staffing plans and schedules.

Total T1 size is estimated as 120KLOC, of which 15% are already implemented. A more detailed "medium term plan" covers 20KLOC (50 man months, assuming productivity of 5LOC per hour and standard TSP working hours). This plan includes:

- SNS repair and
- non-blocking request handler

On a high level, current T1 activities are concentrated on implementing SNS. Some important parts of SNS are already implemented:

- parity de-clustered layout mapping function,
- redundancy calculation algorithms (further tuning and benchmarking is needed),
- full stripe I/O path from a client to server storage.

At the moment, team works on the infrastructure necessary for SNS repair and non-blocking request handler (see T1 sprint 8 overview for details):

- fop-related tasks. These introduce infrastructure for "generic fop methods". To recall, a fop is a file operation packet, that contains fields describing a file system operation. Generic fop methods, described in the request handler HLD referenced above, allow common tasks such as authorization, authentication, resource pre-allocation, *etc*. to be done by the generic code (rather than in a per-fop-type way). This can be thought of as a more uniform and generic variant of HABEO_ flags in Lustre 2.0 MDT;

- request handler stubs. These tasks produce stub implementations for external interfaces that request handler uses. Although the implementations are non fully functional, the interfaces must be right. Full implementations will be supplied later;

- fol task. This task adds an ability to register undo- and redo- actions with a fol record type. Undo and redo actions are used by distributed transaction recovery, which is in turn used by SNS repair;

- SNS repair auxiliary tasks. These tasks adds interfaces that SNS repair uses to control its resource consumption;

- replicator protocol translator task. This component builds Mero fops from Lustre change-log entries, and

- a couple of miscellaneous tasks.

Agreed upon Future Directions

The overall plan is to implement most difficult features first, and to postpone various optimizations and niceties for later. The following components are high priority:

- SNS (full I/O support, repair, recovery),
- distributed transaction manager (DTM),
- non-blocking availability (NBA).

Suggestions for Strategic and Tactical Initiatives

- systematic simulation of the system behavior with a discrete event simulator
- Trinity-controlled testing and benchmarking

Main Architectural Challenges

- networking: lnet, portals or oncrpc?
- containers: merging of meta-data tables. What data-structures are needed?
- how does our architecture works at exascale?