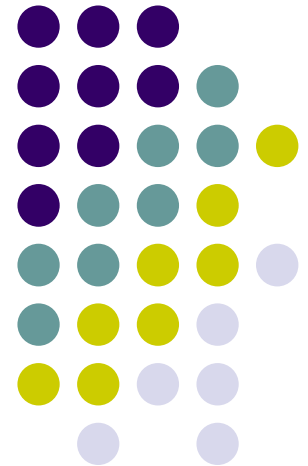# Scalable, Reliable Marshalling and Organization of Distributed Large Scale Data Onto Enterprise Storage Environments

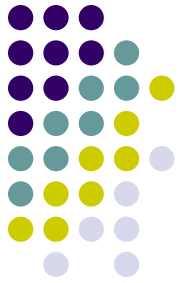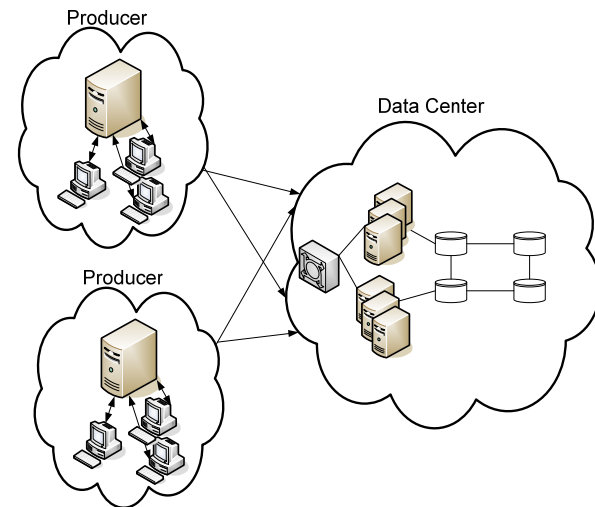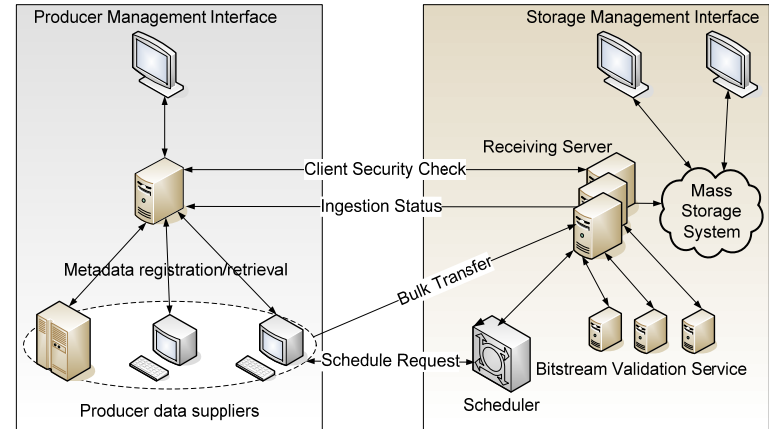Joseph JaJa, Mike Smorul,

Fritz McCall, Yang Yang
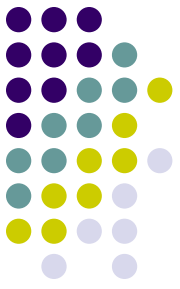
Producer-Archive Workflow Network (PAWN)

- PAWN is:
  - A way to create self describing information packets containing objects along with all necessary provenance, structural, and contextual metadata.
  - An application of archival technique to mass storage ingestion
  - A way to enforce minimal standards on groups ingesting data.

Architecture

- Producer
  - Arrange, annotate
  - Manage clients
- Data Center
  - Validate transmission
  - Transform into appropriate form
- Web service calls for light communication
- Multipart messages for bulk data

Distributed Scalable Ingestion

- Multiple distributed archival receiving stations.

- Flexible dataflow and agreement per producer

- Independent trust relationships between producers and archive

- Data Grid is geographically distributed with multiple entrance points

- Client driven resource reservation to efficiently distribute ingestion load