



Fast, Secured Mission Data Management

Michael Hawkshaw
June 2013



© 2014 by CGI. Published by The Aerospace Corporation with permission.

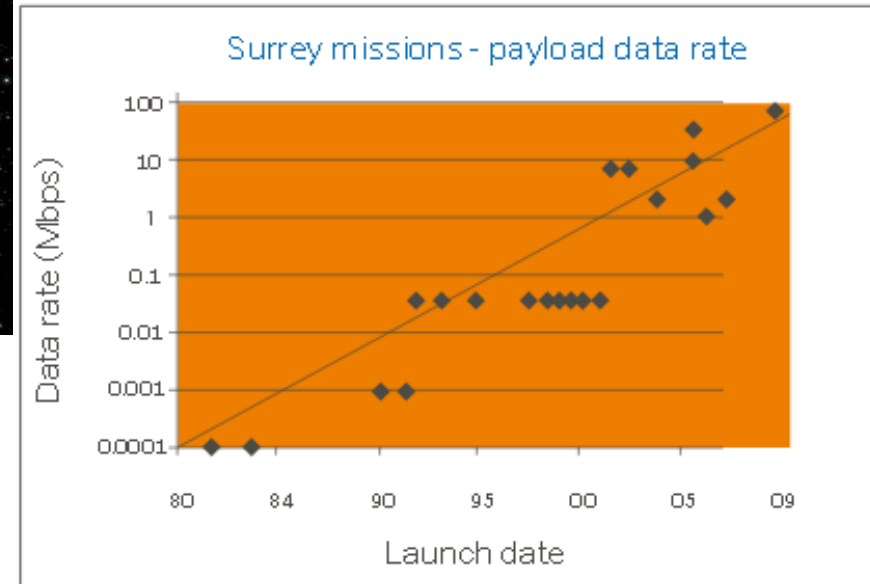
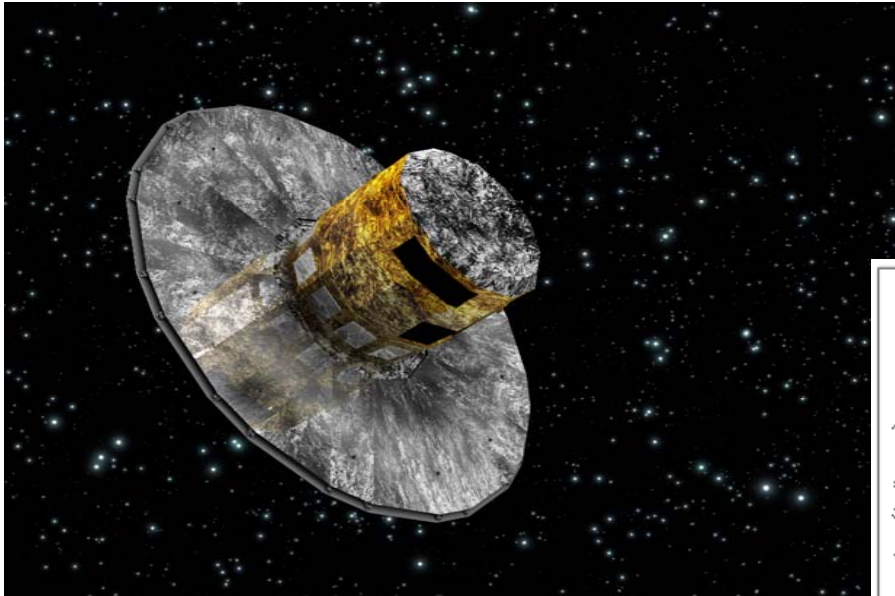
Experience the commitment®

Problem #1



© 2014 by CGI. Published by The Aerospace Corporation with permission.

Problem #2



Problem

- We are collecting a lot of data and that's great
- But we need to store it
 - High initial investment cost
 - Obsolete before mission even starts
 - Not scalable should data needs grow
 - ... and we need to maintain the hardware
- Limits in storage space

X N

CGI

Example: GAIA

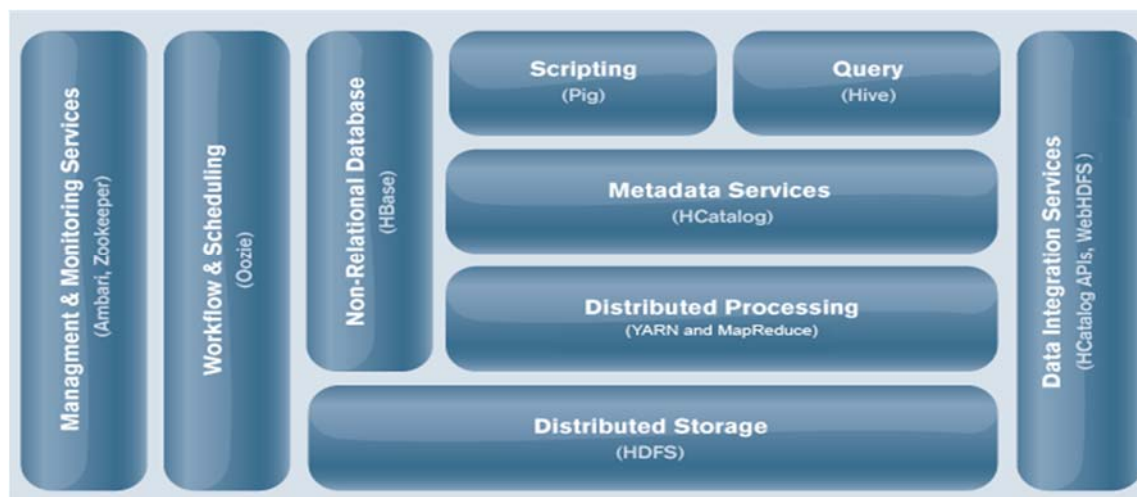
- Estimated 4-6 Tbyte of data over lifetime
 - Housekeeping data
 - Onboard Events
 - Orbital events
 - Telecommands with history
 - Log messages
- Request to have history data accessible for complete lifetime
- Mission is ongoing! Limits of existing archive is reached in 6 months...

We're not the only ones

- NASA
 - Square Kilometer Array, a network of radio telescopes. Will require 700TB/s data rate.
- Facebook
 - Two major clusters in use:
 - a 1100 machine cluster with 8800 cores and ~12 PB raw storage
 - A 300 machine cluster with 2400 cores and ~3 PB raw storage
- Twitter
 - Use Hadoop to store and process tweets
- Yahoo!
 - More than 100,000 CPUs in >40,000 computers
 - Biggest cluster: 4500 nodes giving 18 PB storage

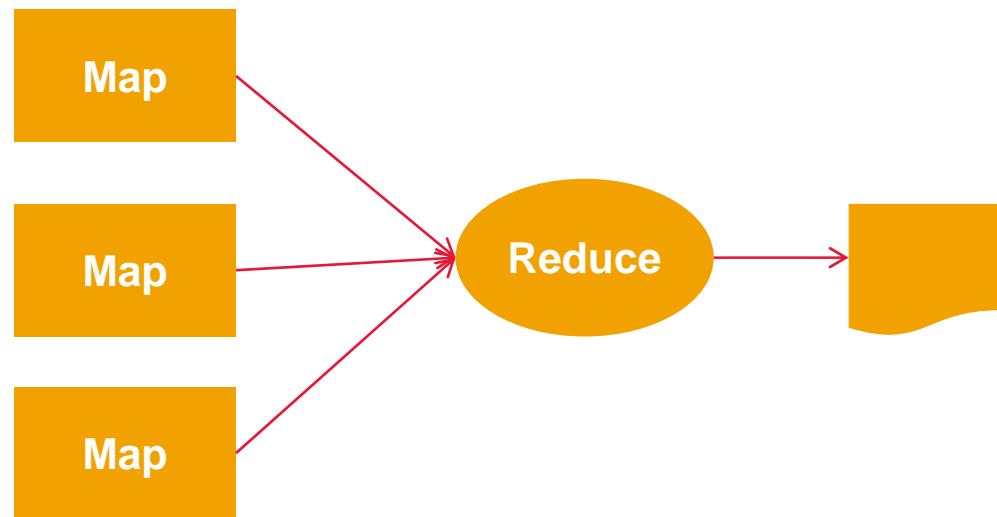
Solution

- Add hardware as you go
- Commodity off-the-shelf hardware
- No expensive SANs
- Access to all data



Solution

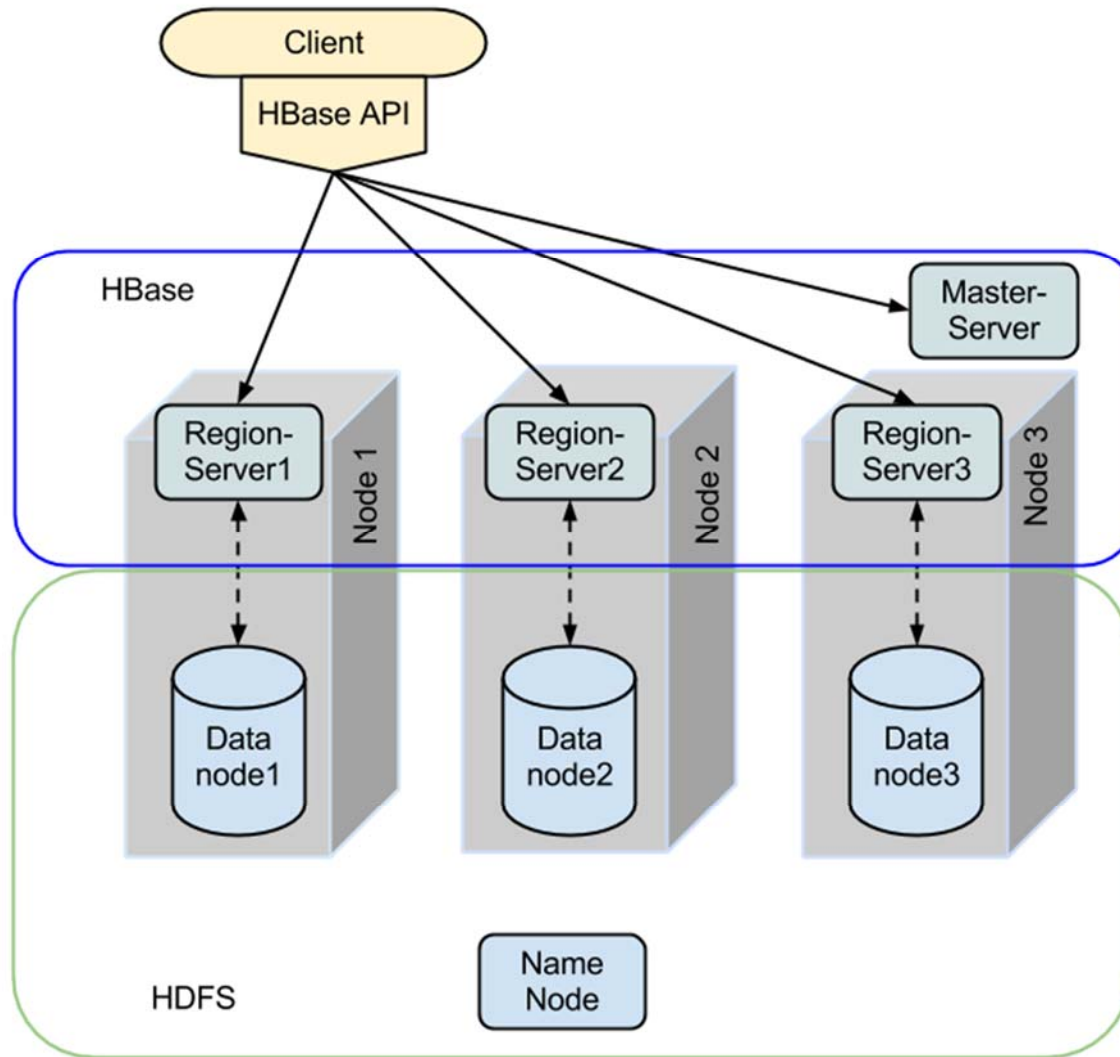
- Parallel processing
- High performance



Security: Data Integrity Between Missions

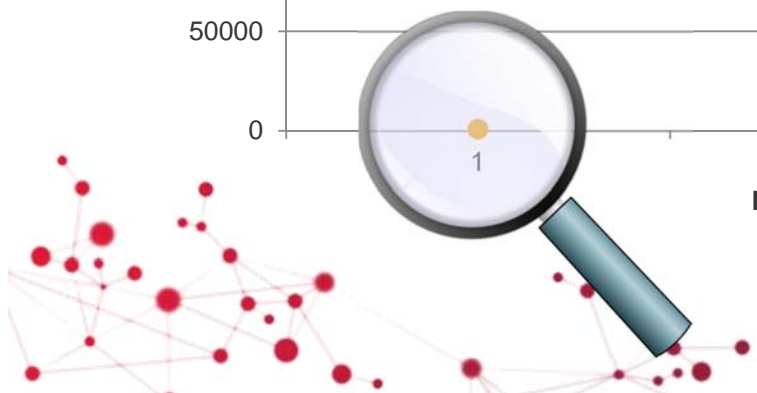
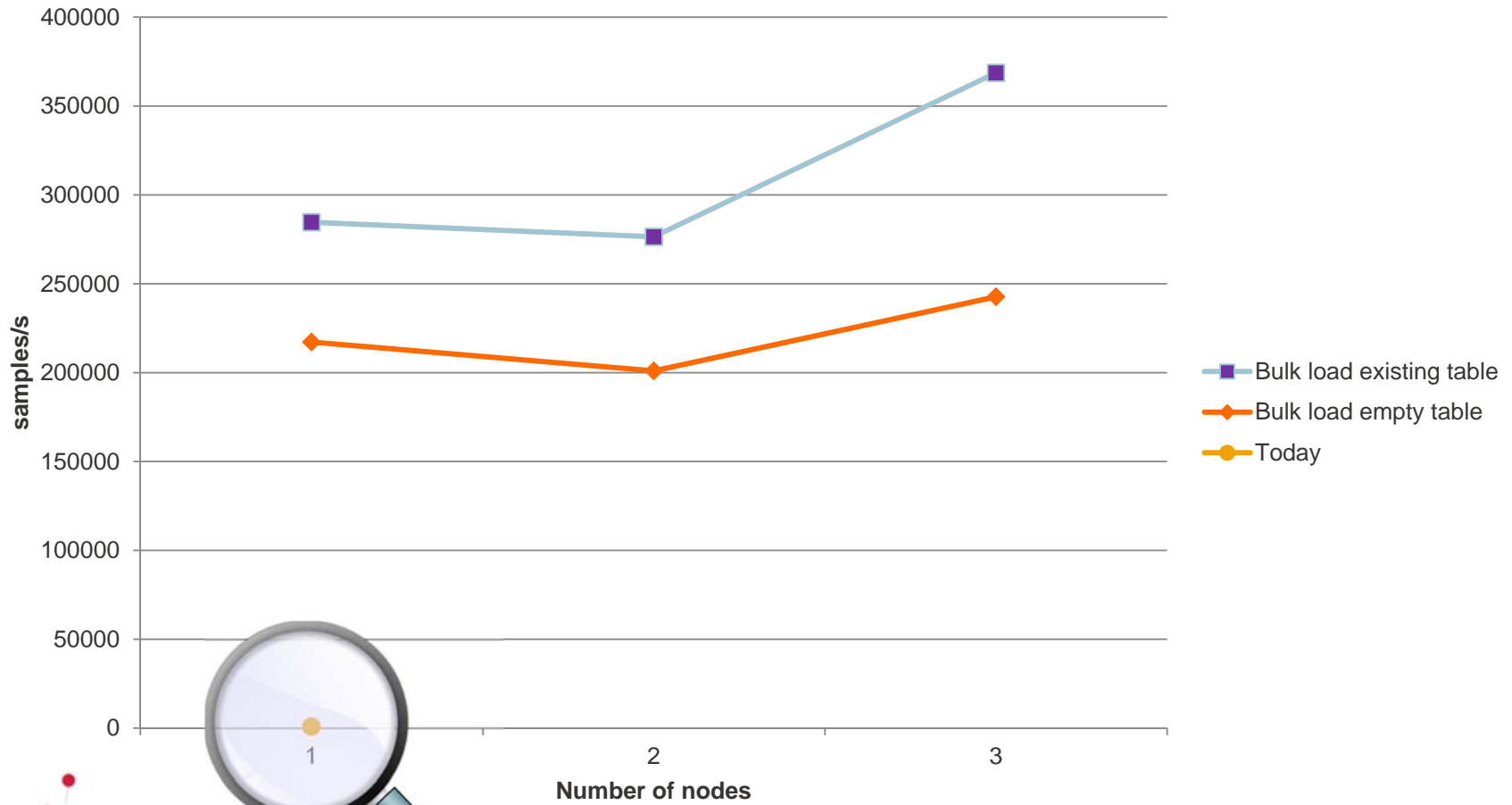
- Complete segregation of data
- Permissions on individual tables
- Each mission has its own user assigned
- Read-only permissions for other mission data

Prototype Architecture

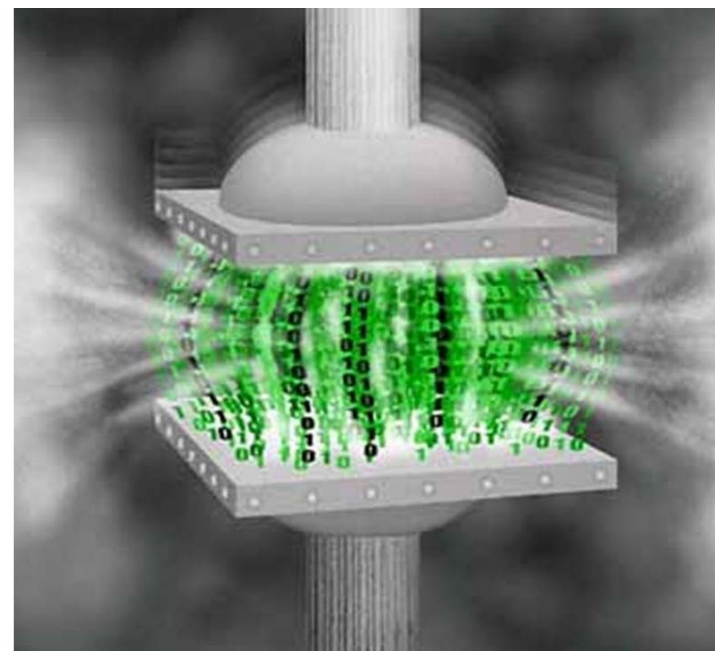
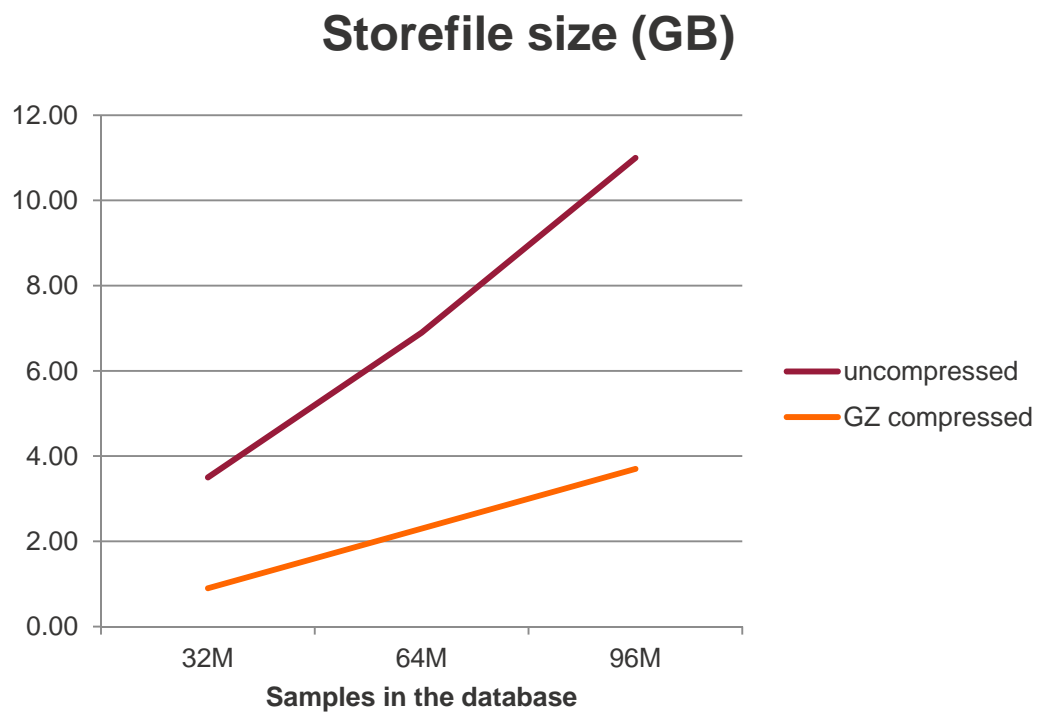


Some Results

Data loading speed



Getting the most out of storage



Benefits

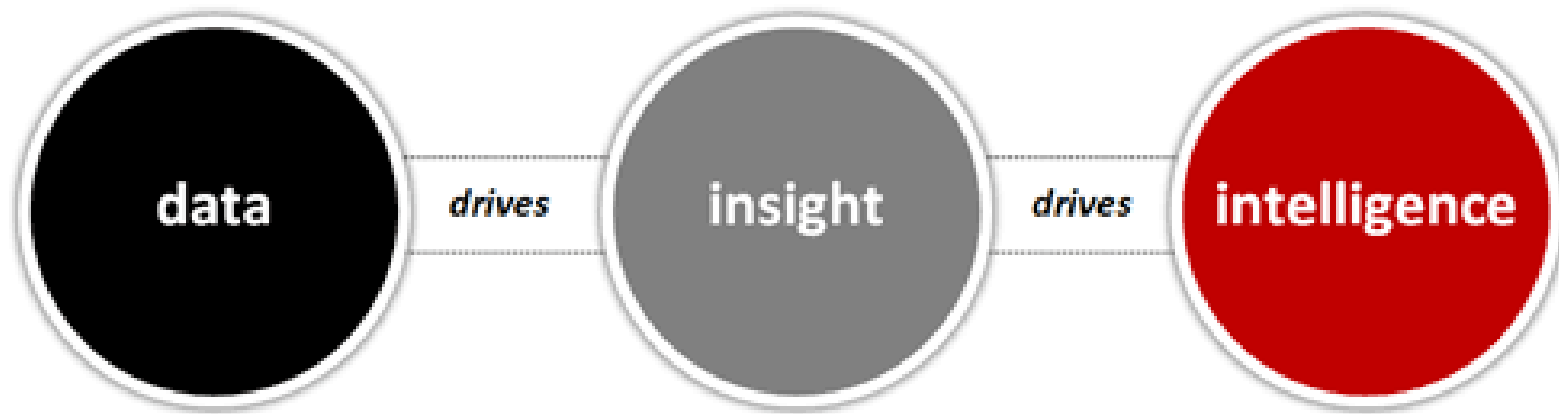
- Scalability
 - Instead of replacing hardware, only add hardware
 - Continuous access to data even during hardware replenish cycles on current assets
- Reliability
 - Replication is built into the solution, not done by additional external processes
- Fast access
 - All of the data across missions is always available
- Secure
 - Table level access control
 - Ensures mission data integrity

Financial Benefits

- Commodity hardware
- Reduce initial expenditure
- Reduce operational expenditure
- “Pay-as-you-go”



The future mission insight



Lessons Learned

- Surprisingly easy to do
- Get keys right
- Optimal Mapper:Reducer ration not clear