



IRSA conversion from to Oracle:

Trials and tribulations

Steve Groom

Ramon Rey

Claude Felizardo



About IRSA

- Archive for NASA's infrared and sub-mm missions
 - IRAS, 2MASS, Spitzer, WISE, Planck, others
 - Other contributed and "value-added" datasets
- Images, catalogs, spectra
- Data Discovery / Inventory search
- Data search/retrieval services e.g. Gator
- Visualization services e.g. FinderChart
- Custom product creation: cutouts, mosaics

- *IRSA shares technical heritage and infrastructure with NStED, KOA, KSAS and others, most of these are included in this effort*



Database holdings

- As of June 16, we operate:
 - 70 databases ...
 - ... consisting of ~3300 tables
 - ... containing ~38 billion records (rows)
 - ... occupying >30TB of storage
 - ... before tempspace/mirroring/RAID/etc
 - ... and growing daily

With recent additions from Spitzer, WISE, and others, the volume of IRSA's holdings has grown by an order of magnitude in the past three years

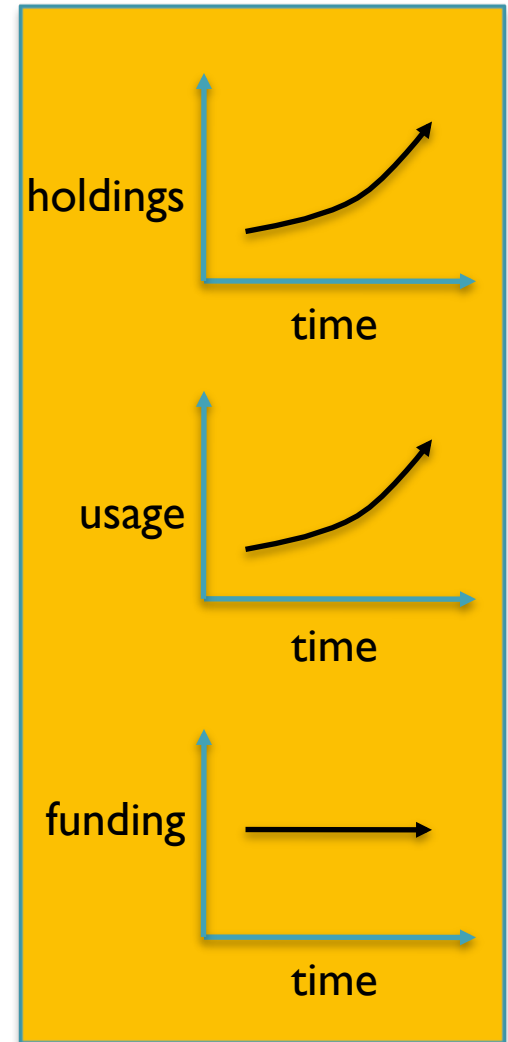
What's a DBMS?

- DBMS = Database Management System
 - Software to manage and search large tables of information
- IPAC has used Informix DBMS software for years
- Support costs have been steadily rising, and licensing terms have evolved
- Costs have outgrown our budget – Informix is no longer affordable
- Need a more affordable solution, preferably a more sustainable one



The problem restated

- Dealing with very large data...
- ...while keeping costs under control



Migration priorities

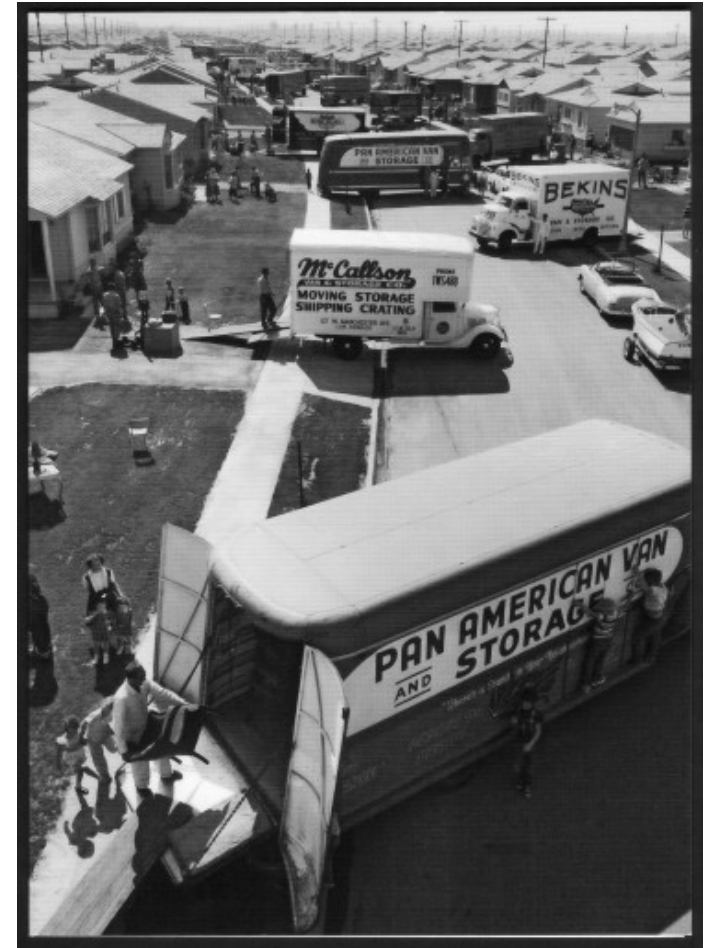
- Most of our database interaction is done through shared “search engine” code – minimizes code impacts to migration
- Applications moving to ODBC/JDBC interfaces to reduce impact of future applications
- With our large systems and short staff, out-of-the-box functionality is very important

- Oracle selected as replacement for Informix
 - Available under Caltech site license at no cost
 - Offers enterprise features key for very large datasets
 - Management tools
 - Partitioning and other large data features

- *While perhaps not the ideal solution, the best one for our current situation.*

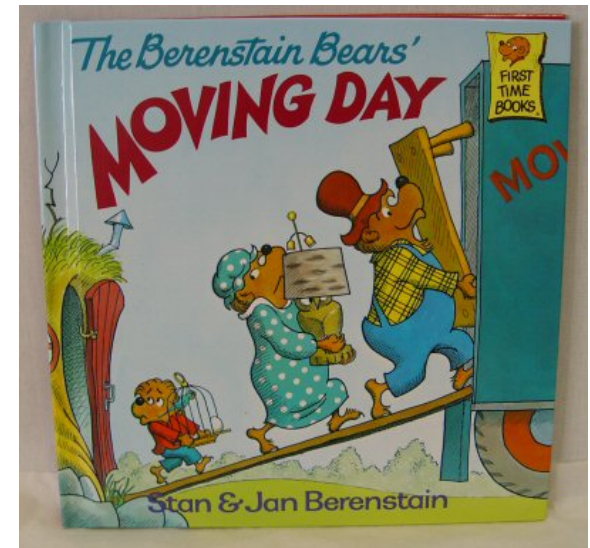
Migration outline

- Setup/configure hardware
- Setup/configure DBMS
- Port apps
- Copy the data
- Validation/Test
- Flip the switch



Moving ahead

- Setup/configure hardware
 - straightforward
- Setup/configure DBMS
 - straightforward
- Port apps
 - Not much code to touch - isisql
 - JDBC ok as-is
 - Some esql/C still to be addressed



Getting from A to B

- Getting data out of one system and into the other
 - What the IT business calls “ETL”: Extract, Transform, Load
- Generally, dump data out of one system to ASCII files, then load those files into new system
- In some cases, load new system from original material



Moving Data: Type Mapping

- Mapping Informix data types to Oracle counterparts
 - smallint -> NUMBER(5)
 - integer -> NUMBER(10)
 - smallfloat -> FLOAT(63)
 - float -> FLOAT(126)
 - date/time types are the most problematic:
 - datetime year to second -> DATE
 - datetime hour to second -> VARCHAR2(8)
 - datetime year to fraction(3) -> TIMESTAMP(3)

Other Changes

- Resolving conflicts with reserved words
 - DATE , USER
- We've been collecting data for a while, it's not as organized as we'd like
- Inconsistencies in existing data
 - “why is this one different?”
- Old stuff
 - “do we still need this?”



Validating Data

- Software testing
 - Treat like software regression tests
- Statistical/random sampling
 - Random queries against large data sets
- Brute force: direct data comparison
 - Dump the whole thing and diff it
 - Size limits
 - Differences in dump formats
 - Leading/trailing zeros, spacing
 - Numeric differences due to floating point precision

The World Doesn't Stand Still

- While we're working on this, IRSA is still in full operations
- New data arriving all the time
- Need a coordinated cutover:
 - Freeze content
 - Make final update
 - Check result
 - Cutover software
 - Unfreeze
- WISE & Spitzer will move later

Progress to date

- Servers ready
- DBMS ready
- Software mostly ready
- Data largely moved
 - Several databases need major refresh
- Thorough testing and tuning still to come
- KOA approaching cutover readiness

Lessons Learned

- Data organization and conversion has been the big challenge, not so much software
- Cleanup will be beneficial for the long haul
- Software porting will simplify future moves. New software being written in more portable style.
- Somewhat concerned about the compatibility mechanisms we're putting in place – will these become troublesome?

Lessons Learned – cont.

- Validation should have occurred before we started modifying/moving DD's
- Conversion mappings were complex and often special-cased

Future

- Investment in data cleanup worthwhile
- Inevitably, we'll have to move again someday
 - No technology lasts forever
- Architectural evolution also needed
 - We're growing faster than our monolithic model can sustain
 - We aren't the only ones with this problem
 - The move we make now buys us some time for open source solutions to improve

It's Happening, Slowly



Thanks

- This has been a large team effort, with contributions from:
 - John Burke
 - Claude Felizardo
 - John Good
 - Anastasia Laity
 - Ramon Rey
 - Angela Zhang
 - Advice from Tom Handley, Ed Jackson
 - Systems support from Wendy Burt, Eugene Hacoceans, and the rest of ISG