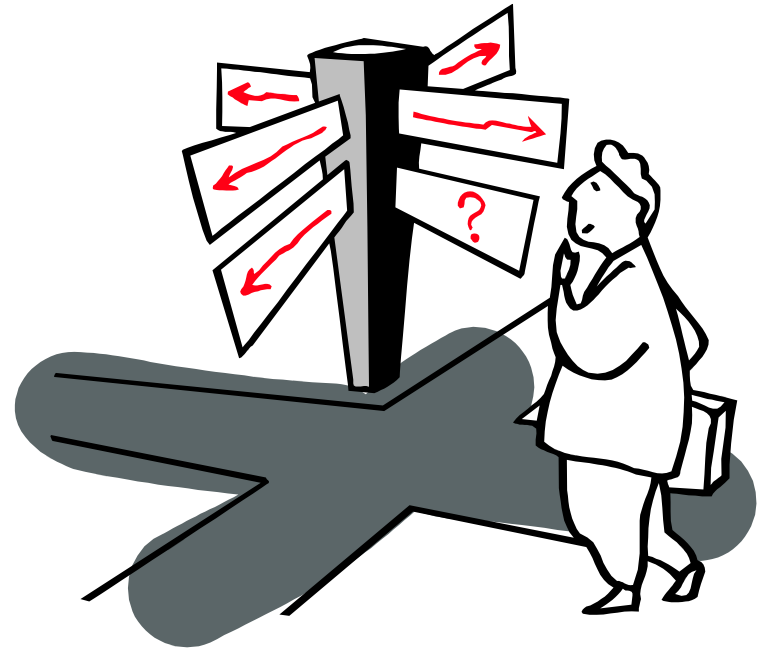


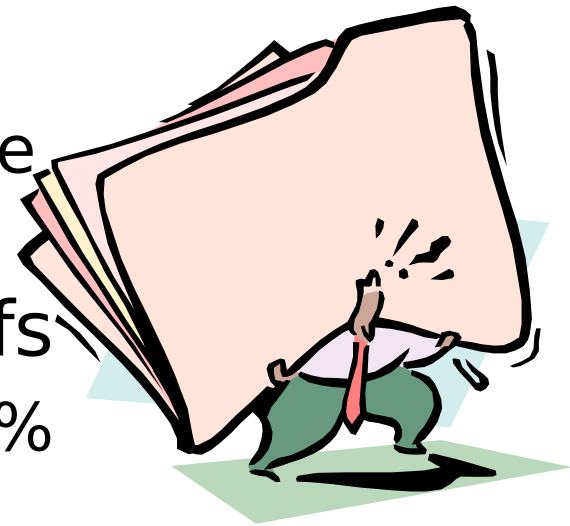
HIRLAM/HARMONIE archiving in MARS

- Why we need / want to use it
- General about MARS
- MARS components
- Problems and solutions
- Activities and Plans



Why we need / want MARS (at ECMWF)

- HIRLAM Retrieves and Reads MARS data at ECMWF
 - (Initial and boundary data)
 - ECMWF is special case – nowhere else
- “Archiving” is done in fs or ecfs
 - Enormous load at ECMWF 30-40%
 - Number of files enormous
 - Difficult to clean, thin or migrate





Hirlam archiving at ECMWF

- Currently, most of data in the working directory are archived into the ECFS with cycle structure
 - `ec:/$uid/hirlam/$exp_name/$yyyy/$mm/$dd/$hh/`
- Previously experiment results were stored in \$SCRATCH disc.

Aggressive cleaning is applied to reduce storage need on \$SCRATCH.

Instead, the same data are moved to ECFS

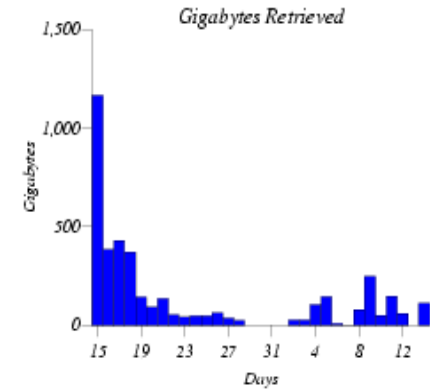
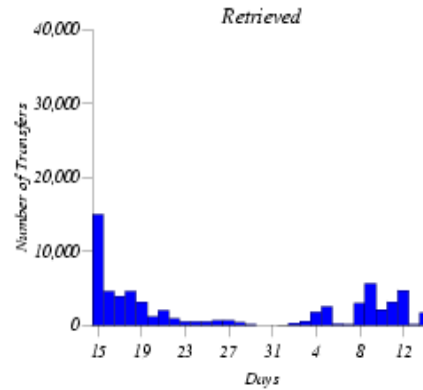
- ECFS volumes:
 - HIRLAM: +200TB – 8 Million files
 - Users: 500TB – 16 Million files
 - Total: 1000TB – 30 Million files
- ECFS Data transfers per day:
 - HIRLAM: 1-2TB – 20-40k files
 - Total: 4-6TB – 70-90k files
- MARS volumes:
 - 3000+ TB – 20000 Million fields

Hirlam ECFS transfers

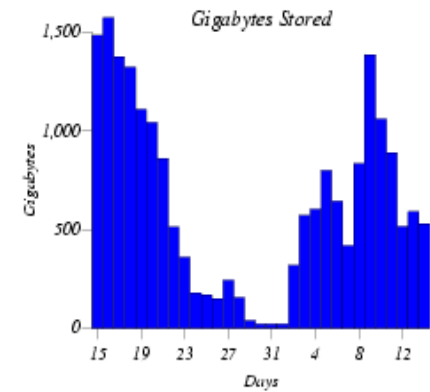
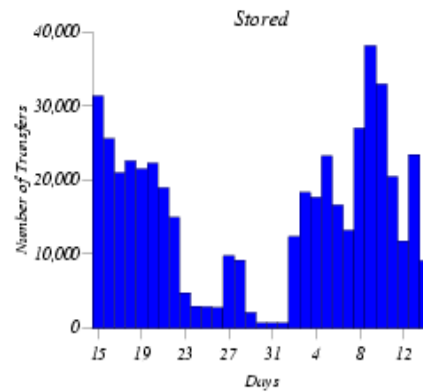
Hirlam Daily Total Transfers

15 December 2006 to 14 January 2007

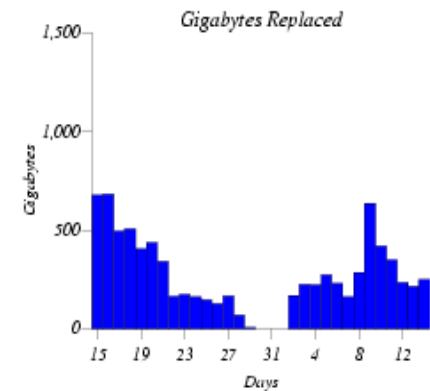
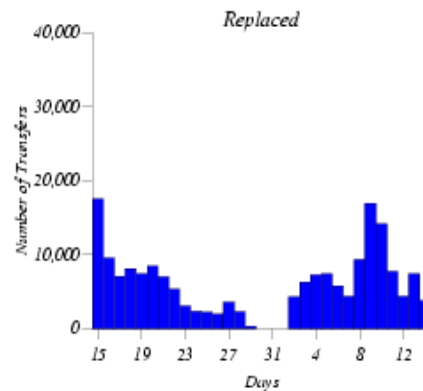
retrieved



stored



replaced





Hirlam-A activities on ECMWF platform (2)

- HIRLAM system activities:
 - ☐ HIRLAM Reference system is maintained at ECMWF
 - preparation for system upgrade and code tests
 - technical and meteorological evaluation
 - validation experiment; derivation of structure functions etc.
- Hirlam-related research activities
 - code development, impact/sensitivity studies
 - GLAMEPS: HIRLAM-ALADIN eps system development
 - supplemental numerical experiments related to operational activities

Why we need / want MARS

(ii)

- Cumbersome to find data and extract what you

want Differing residence, paths and organisation depending on time of archive (and migration)

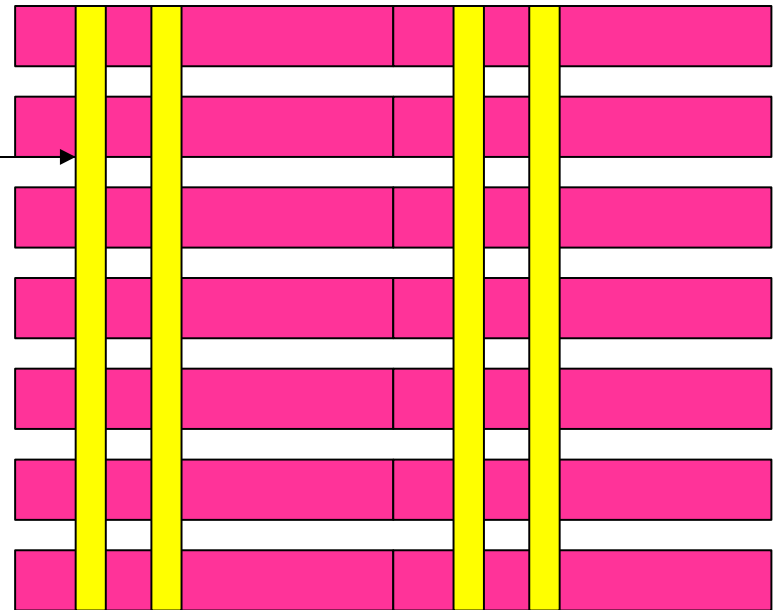
- Access only at file or directory (tar) level

- There are general interfaces and tools

- Extraction of sub-sets and/or collation of data
- COMPUTE (maths on fields! - + * / exp sqrt log etc.)
- Webmars
- MARSINT (although projections...?)
- Data services (to be provided)

- Hardware independent – different media

- Data server – disk, tape, etc.
- Management and migration tools



MARS

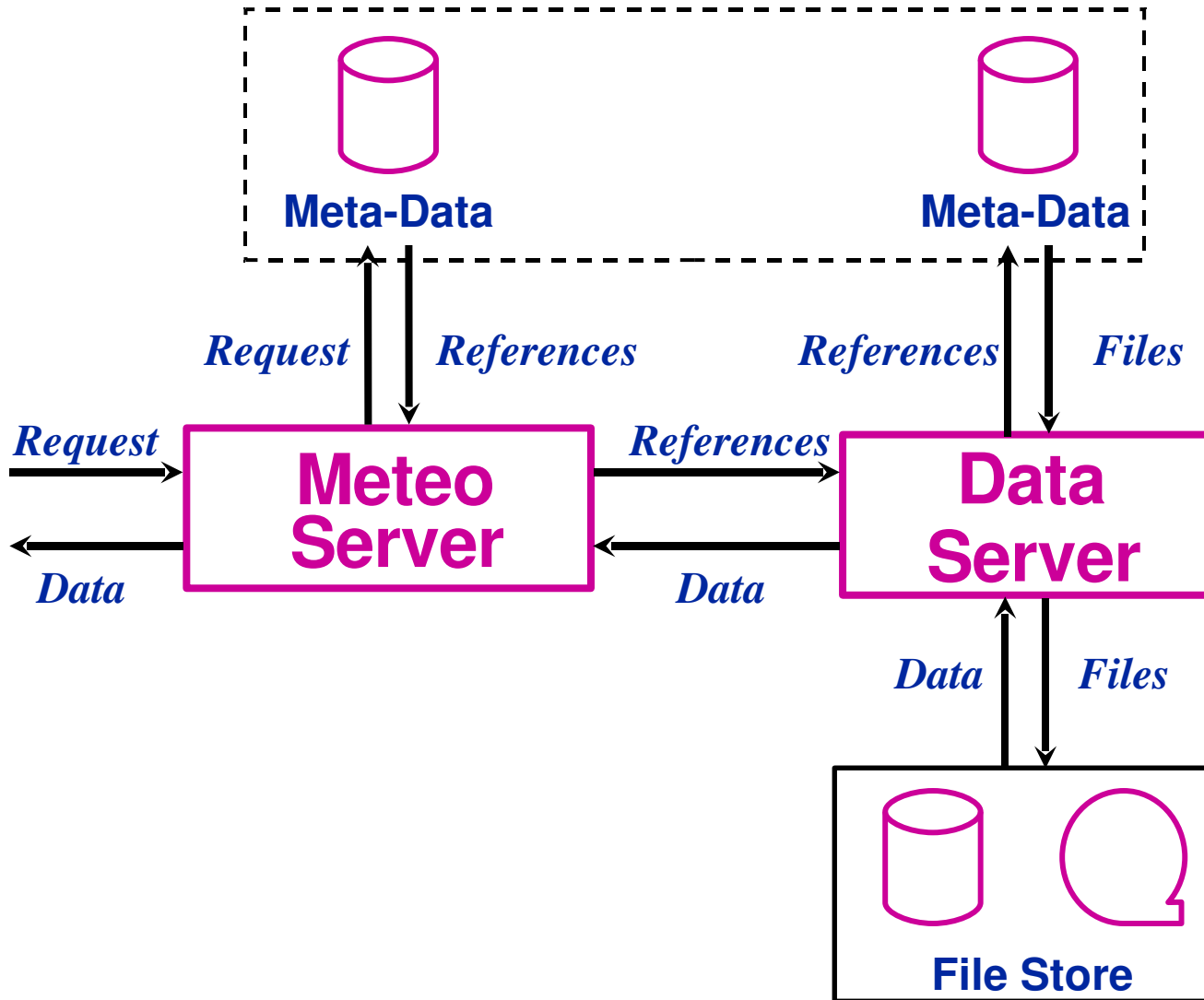
(courtesy of Manuel Fuentes, ECMWF
and of course Baudouin Raoult)

- Meteorological Archival and Retrieval System
- **Manages** meteorological fields
 - Operational data
 - Research data
 - Special projects
 - Member State's own models
- Not a file system
 - Users are not aware of the location of the data
- An archive, not a database
 - Metadata online
 - Data offline

MARS Components:

- Clients (retrieve)
 - Syntax
 - Analyse request (number of fields)
 - Enquire server 1, 2,
 - Transfer data
 - Computations
 - Deliver data (target file)
- MARS server
- Data server
- MARS hardware
 - Meta data
 - Cache , pre-archive space
 - File store
 - Disk
- Clients (archive)
 - Syntax
 - Check data
 - Transfer data (pre-archive file)
- Other Server functions:
 - Flush
 - Migrate
 - Clean
- Scheduling, queueing
- SMS control

MARS Server Architecture



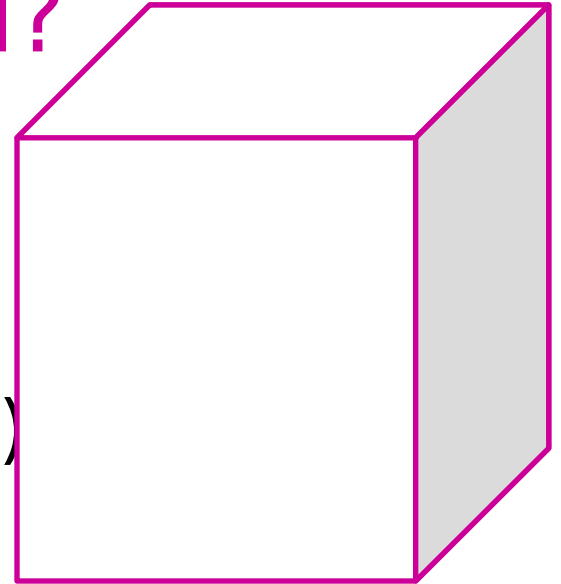
A meteorological language

Retrieve,

date = 20010101/to/20010131,
parameter =
temperature/geopotential,
type = forecast,
step = 12/to/240/by/12,
levels = 1000/850/500/200,
grid = 2/2,
area = -10/20/10/0

What is a field?

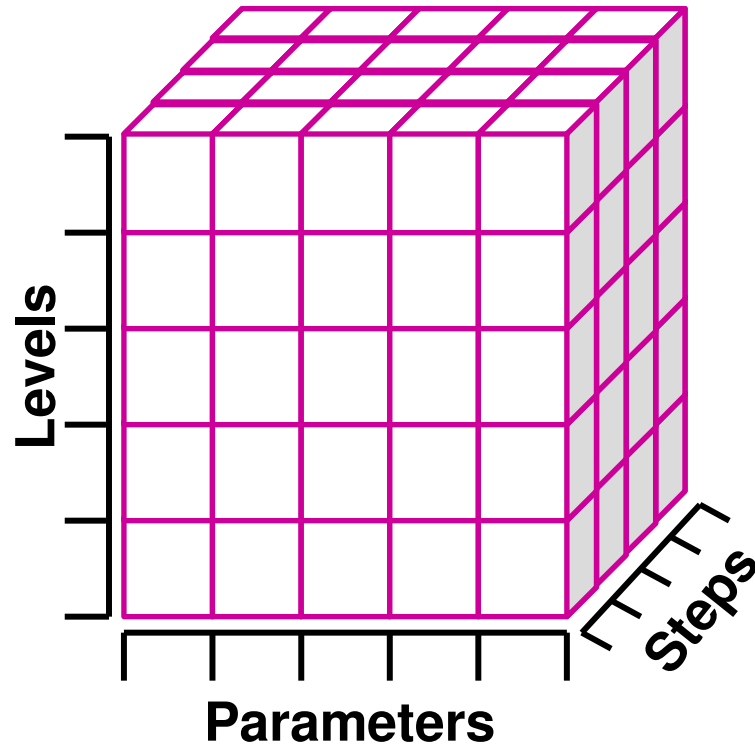
- Class (Operations)
- Type (Forecast)
- Stream (Daily Archive)
- Version (1)
- Date (1978-01-12)
- Base Time (12 UTC)
- Step (240 hours)
- Parameter (Temperature)
- Level (1000 hPa)



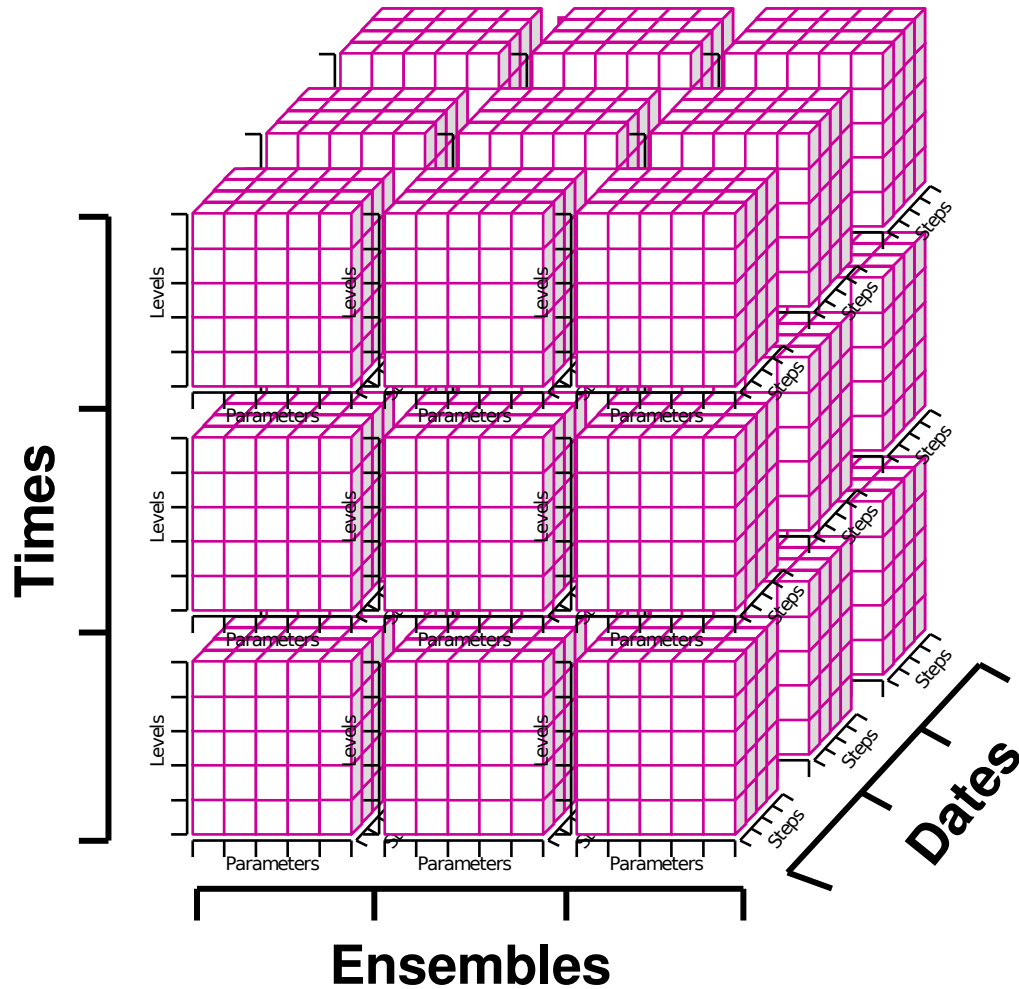
Archive objects

- Grouping
 - One forecast
 - One month of analysis
- Grouping
 - Hundreds of Thousands of fields
 - Size of a tape
 - Reduce number of tapes
- Natural Collocation
 - Minimise tape mounts
- 1 archive object = 'field selection page' in webmars

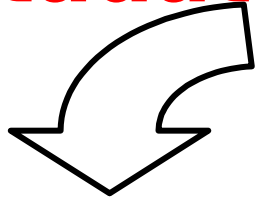
Archive objects are cubes



Archive objects are hyper-



MARS Metadata



MARS Server Meta-Data "Shapes"

Levels

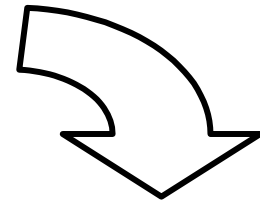
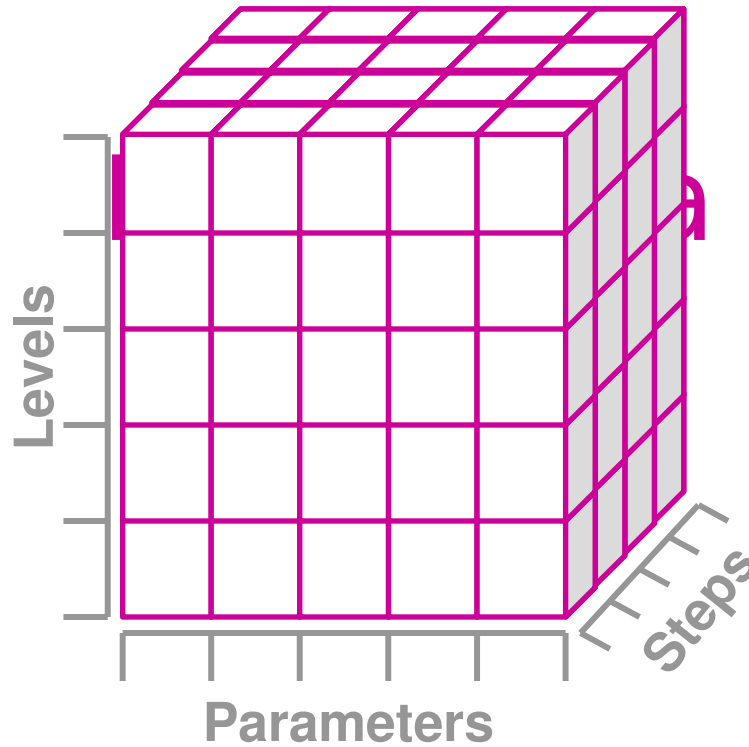
850	500	200	100	50
-----	-----	-----	-----	----

Parameters

Z	T	U	V	W
---	---	---	---	---

Steps

12	24	36	48
----	----	----	----

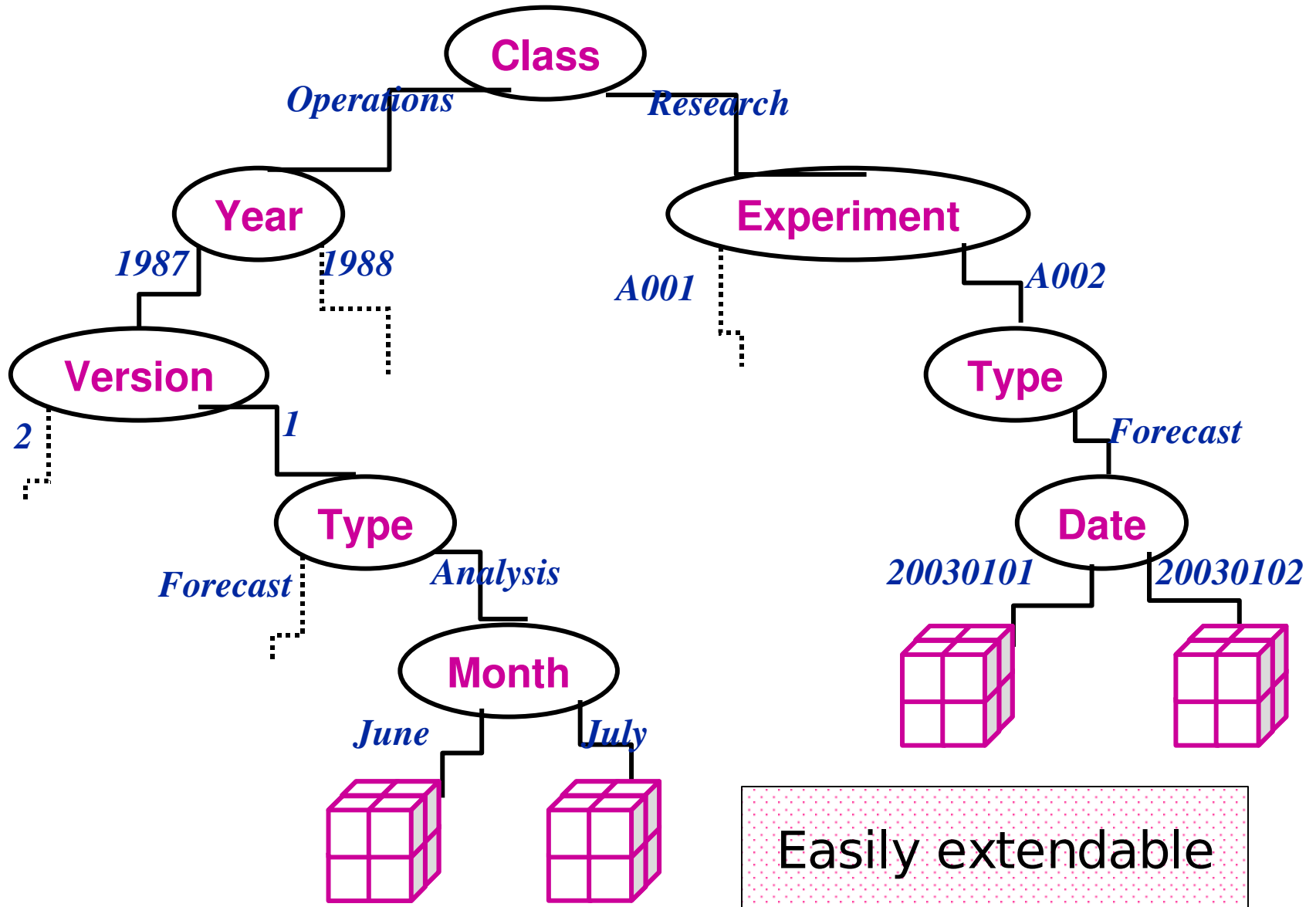


Data Server Meta-Data "Layouts"

Offse	Length	Files
t 0	10000	A
1000	10000	A
0	15000	B
1500	15000	B
...



MARS tree



Issues for HIRLAM

archiving

- Changes to existing GRIB coding
 - Changes in your model Or modify GRIB headers before archiving : **We wrote a filter**
 - Insert vertical coordinate in model level fields
 - Re-pack non standard bit length
 - Split multi-level surface fields into files
 - Create script with archive field list (right order)
 - Sub-centre trick (98) for extension in Sect 1 with parameters and MARS names
- Changes for users of data:
 - Be able to decode new GRIBs Or convert into current format before they can use data : **filter**
- Data access policy (restrictions, ownership, quotas)
- New types of data (eg. EPS) require coordination with MARS analysts (ECMWF)
- Use of Fields DataBase (or not)

MARS keywords/GRIB header to clarify

- CLASS = HIRLAM (?)
- STREAM = OPER/ENFO(?)
- TYPE = AN/FC/EF
- EXPVER = 0001 (operations)/ aaaa
(research) : create unique exp id not to
overwrite existing exps
- LEVTYPE, some type of levels currently
are HIRLAM specific : surface multi-level
– 191, 192, ...

Benefits of archiving in MARS

- Homogeneous access to data (MARS language)
- GRIB data is better handled by MARS rather than ECFS or filesystems
 - Granularity, Data collocation, Incremental archiving
- (Back-archive (?)) **not for SMHI research**
- Support for HIRLAM data by all ECMWF tools
 - Webmars
 - EMOSLIB/Interpolation (?)
 - Metview
 - Verify
- Fase migration to GRIB Edition 2 –

Activities / Plans

- MARS at SMHI
 - Research (HIRLAM Climate, HIRLAM NWP, Ocean, HARMONIE): NSC Linköping **Ongoing Project SMHI-NSC**
 - Operations later ? **NSC manages SMHI data storage**
- MARS for HIRLAM (at ECMWF)
 - Build on SMHI research implementation
- MARS for HARMONIE (at ECMWF) ?
- **MARS for ALADIN ?**
 - General development questions:
 - Emoslib interpolation extensions
 - Open source ?
 - Maintenance and updating

NSC installations

Name	Type	#CPUs	Usage	In use since
Darkstar		Cluster	88	SAAB Mar 2006
Mozart	Shared memory		64	SNIC Feb 2006
Hyperion		Cluster	16	SAAB Jan 2006
Storage Oct 2005	Storage solution		6	SMHI, SNIC
Dunder	Cluster	104	SMHI	Sep 2005
Tornado	Cluster	264	SMHI, MISU at SU	Aug 2005
Blixt/ Pavel		Cluster	174	SMHI Feb 2005
Green	Cluster	74	IFM at LiU	Sep 2004
Hooke	Shared memory		8	SAAB Dec 2003
Navier	Shared memory		32	SAAB Dec 2003
Bluesmoke Aug 2003		Cluster	100	Swegrid, SNIC
Match	Cluster	8	SMHI	Jun 2003
Maxwell	Cluster	80	SAAB	Jan 2003
Dayhoff	Cluster	32	IFM at LiU	Dec 2002
Monolith Nov 2002		Cluster	400	SNIC, SMHI
Bris	Cluster	36	SMHI	Oct 2002