ORACLE in AMS

E.Grancher, N.Segura
CERN IT/DB
V.Choutko, A.Klimentov
AMS collaboration

VLDB Workshop

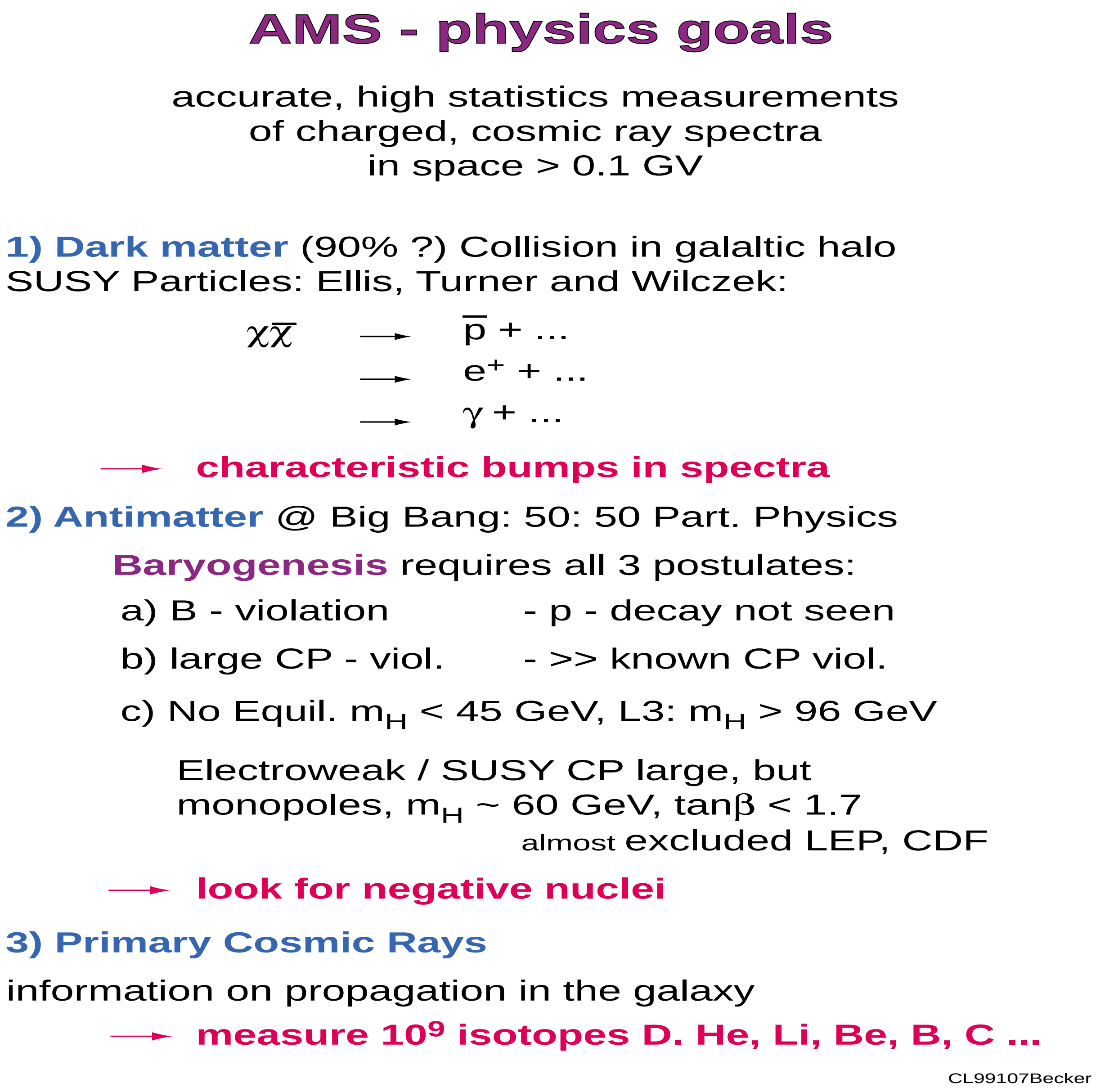
CERN

July 11 - July 12, 2001.

Outline

- AMS experiment
- Data Volumes
- Decision Factors for Database and Data store choice
- Tests with Oracle (event tags, time dependent values)
- Data Production

AMS : Physics Goals



AMS - physics goals

accurate, high statistics measurements
of charged, cosmic ray spectra
in space > 0.1 GV

1) Dark matter (90% ?) Collision in galaltic halo
SUSY Particles: Ellis, Turner and Wilczek:

$$\begin{aligned} \chi\bar{\chi} &\longrightarrow \bar{p} + \ldots \\ &\longrightarrow e^{+} + \ldots \\ &\longrightarrow \gamma + \ldots \end{aligned}$$

⟶ **characteristic bumps in spectra**

2) Antimatter @ Big Bang: 50: 50 Part. Physics

Baryogenesis requires all 3 postulates:

a) B - violation - p - decay not seen

b) large CP - viol. - >> known CP viol.

c) No Equil. $m_H < 45$ GeV, L3: $m_H > 96$ GeV

Electroweak / SUSY CP large, but
monopoles, $m_H \sim 60$ GeV, $\tan\beta < 1.7$
almost excluded LEP, CDF

⟶ **look for negative nuclei**

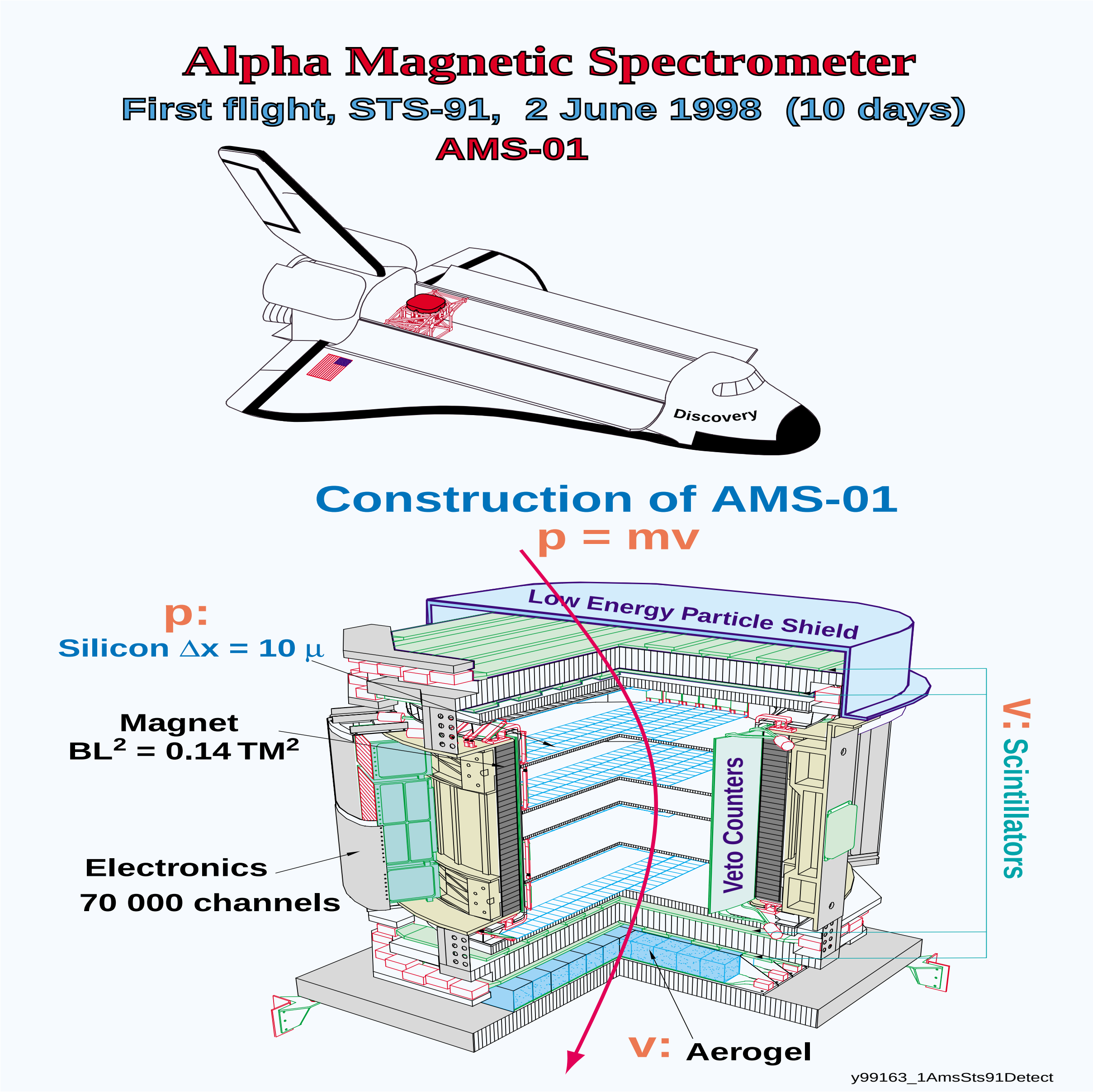
3) Primary Cosmic Rays

information on propagation in the galaxy

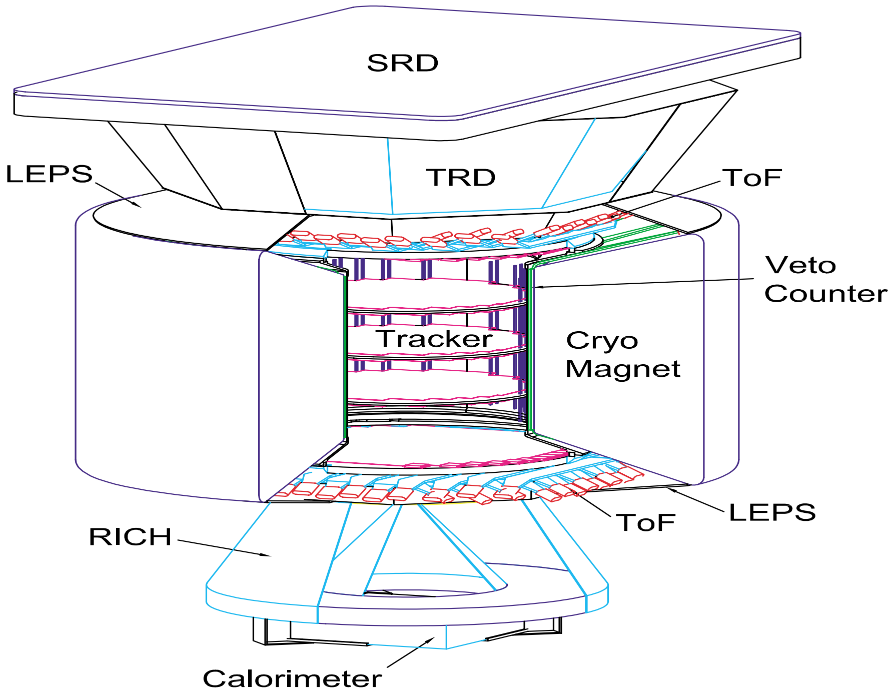
⟶ **measure 10^9 isotopes D. He, Li, Be, B, C ...**

CL99107Becker

STS-91 Precursive Flight



AMS-02
3 Years in Space

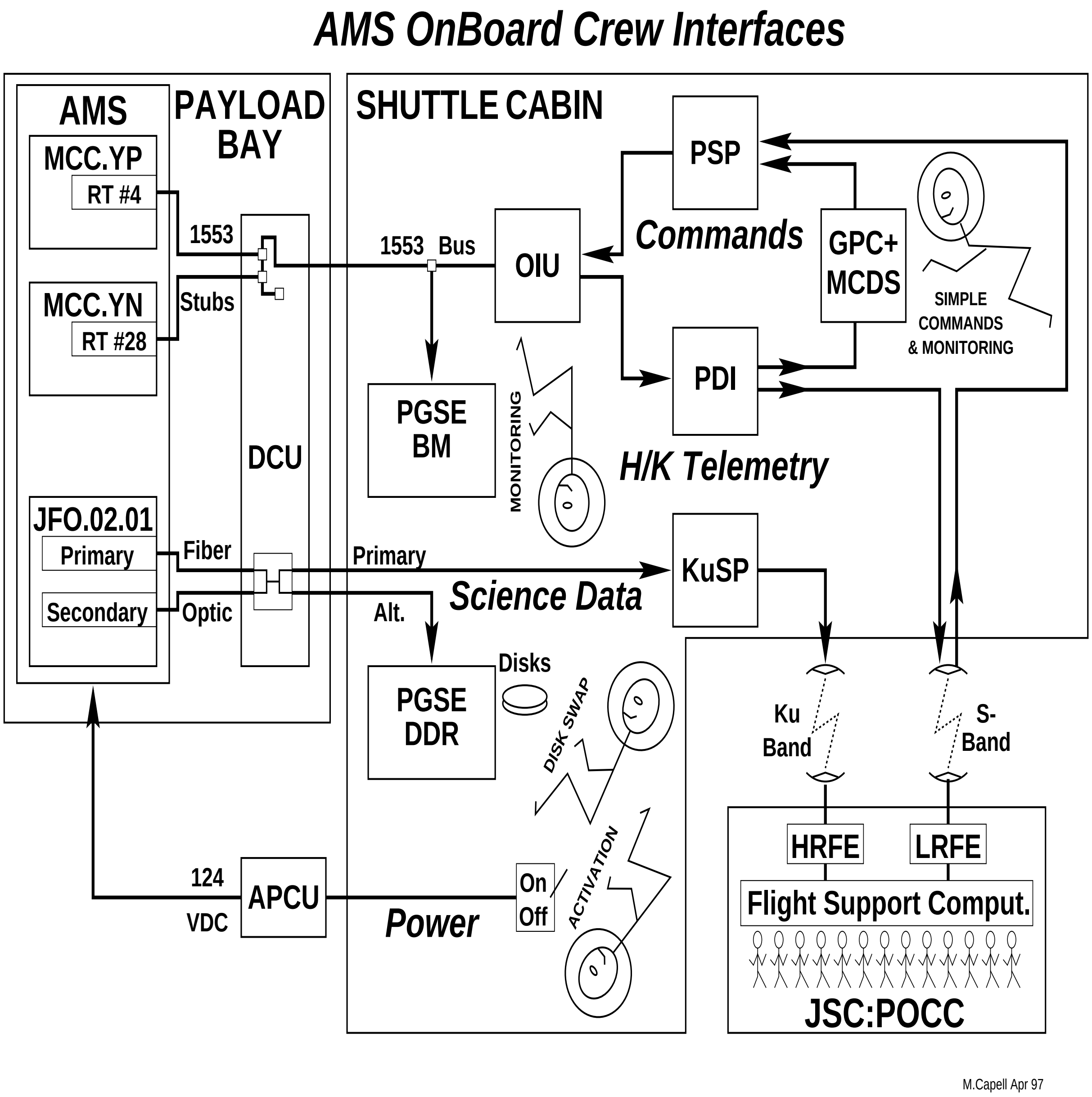


NASA-DOE MOU Energy related Civil Space Activities
July 9, 1992

Reviewed and approved by DOE
April 2-3, 1995 and March 15, 1999

CL99090R.Becker

STS91 Flight. AMS Data Flow



M.Capell Apr 97

ISS Flight. AMS Data Flow

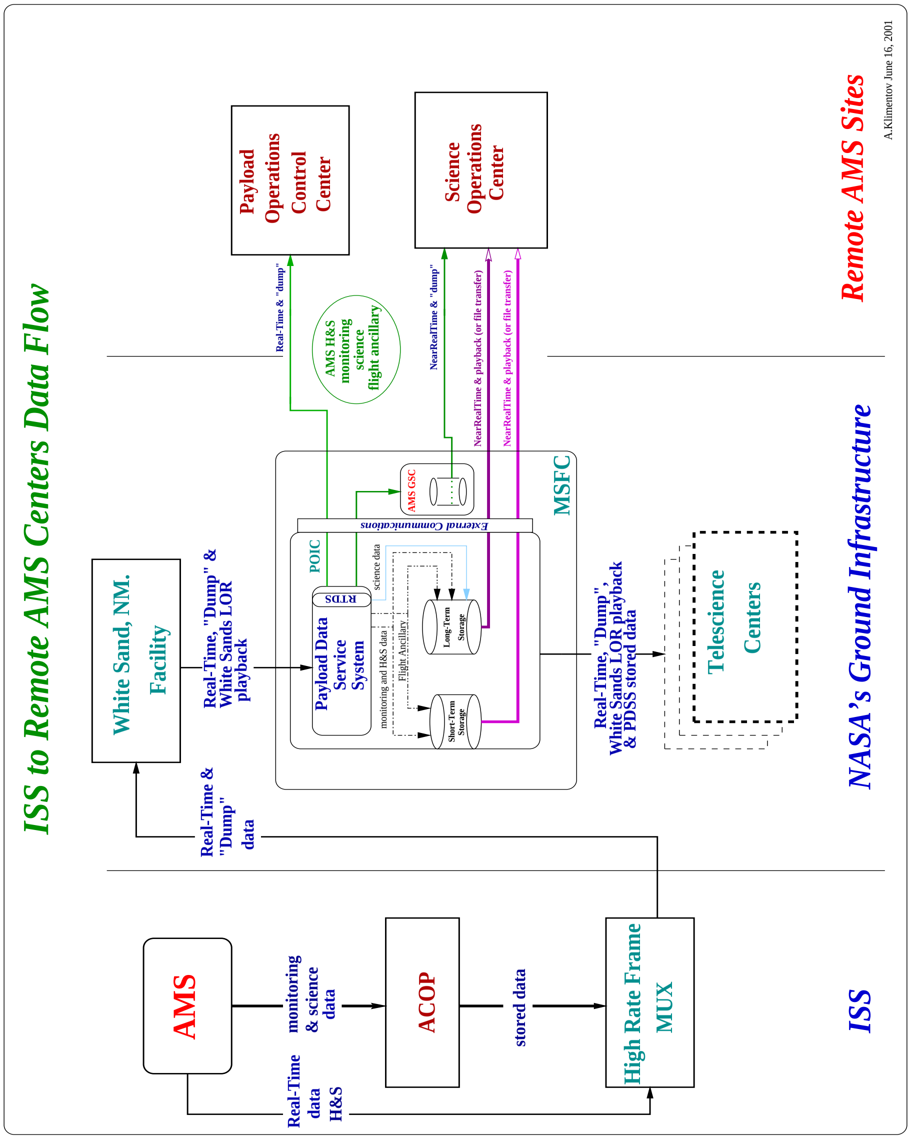

ISS to Remote AMS Centers Data Flow
AMS
Real-Time data H&S
monitoring & science data
ACOP
stored data
High Rate Frame MUX
Real-Time & "Dump" data
White Sand, NM. Facility
Real-Time, "Dump" & White Sands LOR playback
Payload Data Service System
RTDS
POIC
monitoring and H&S data
Flight Ancillary
science data
Short-Term Storage
Long-Term Storage
External Communications
AMS GSC
MSFC
Real-Time, "Dump", White Sands LOR playback & PDSS stored data
Telescience Centers
Real-Time & "dump"
AMS H&S monitoring science flight ancillary
NearRealTime & "dump"
NearRealTime & playback (or file transfer)
NearRealTime & playback (or file transfer)
Payload Operations Control Center
Science Operations Center
ISS
NASA's Ground Infrastructure
Remote AMS Sites
A.Klimentov June 16, 2001

Figure 1: ISS to remote AMS Centers Data Flow

ISS Flight: MSFC to AMS Groud Centers Data Flow

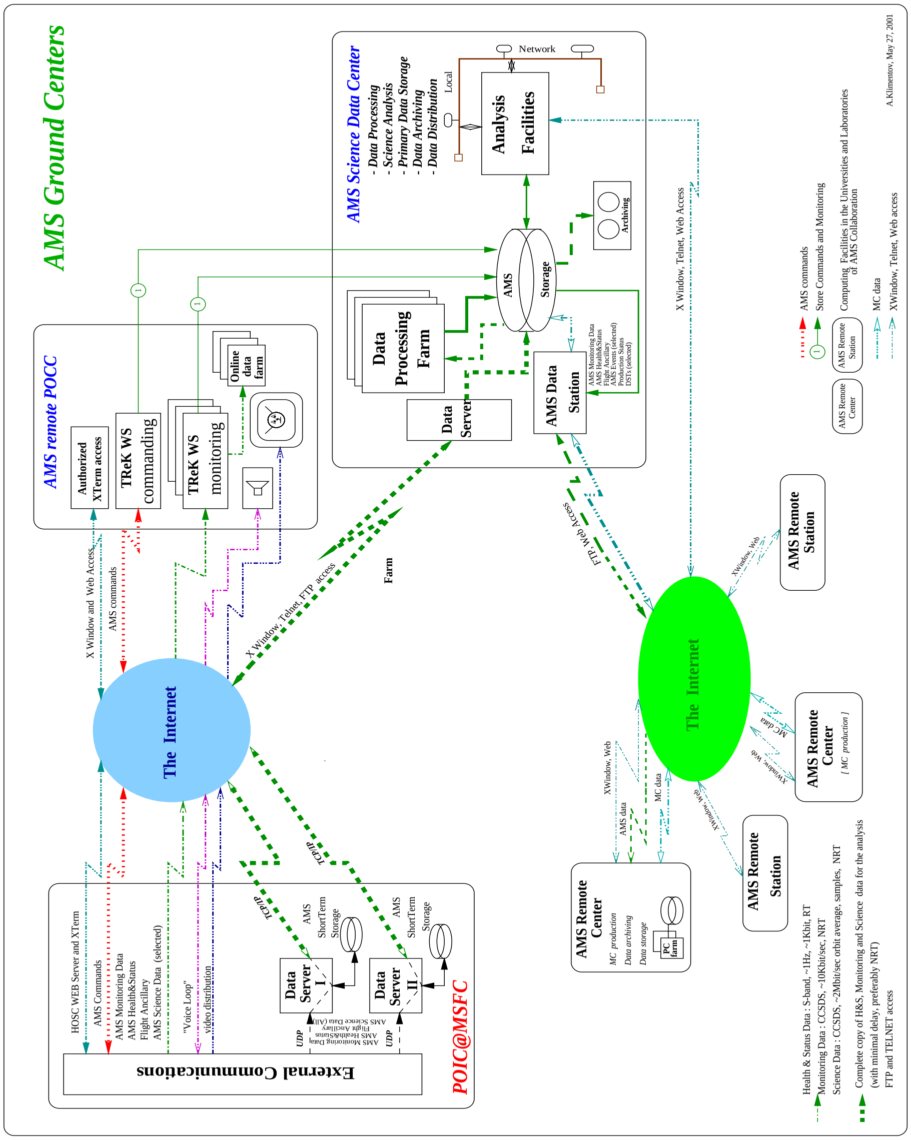


Figure 2: AMS Ground Centers

AMS Data Volumes

- **AMS-01. STS-91 Flight (Jun 2 - Jun 12, 1998)**
  Raw data volume 120 GB
  Total data volume 1.2 TB

- **AMS-02. AMS on ISS (Oct 2003 - 2006)**
  - **Total data volume 168 - 180 TB**
  - **50-60 dual-CPU 1.5+ GHz Pentiums (AMD) to process events in "quasi realtime" mode**

Table 1 Data Volumes.

Data Type	Total TByte per year	Data On Direct Access disks (TByte)	Data to be archived (TByte)
Raw Data	11-15	8 - 11	11-15
Event Summary Data	44	8.3	44
Event Tag	0.6	0.6	0.6
Total	56 - 60	17 - 20	56 - 60

Table 1: AMS-02 data volumes

AMS-01 Data Processing

September 1998 - May 1999.
STS-91 flight data processing at CERN

Sep. 1998. Setup processing/analysis center:

- **AlphaServer 4100 (4*600 MHz)**
- **4 AlphaServers 1000A (400 MHz)**
- **2 AlphaStations 500 (500 MHz)**
- **4 dual-processors Pentiums II (450 MHz)**
- **0.9 TB of disks attached to Pentiums and Alpha's, striped, NFS'ed via two private 100 MBit/s ethernet segments.**
- **OS - TruUNIX and Linux**
- **AMS software is ported on Linux, but not the database part**

No Objectivity version for heteregeneous environment (access to TruUnix hosted federation from Linux);

Use:

- **Root for Housekeeping data.**
- **Homemade database to store geometry and calibrations (TDVs).**
- **PAW NTuples for reconstructed data.**

Choice of Database

Relational Database + Root + flat files

– **Relational database to keep :**

  - ∗ *run, file, dataset catalogues;*
  - ∗ *event tags;*
  - ∗ *NASA ancillary data;*
  - ∗ *Health&Status, monitoring data*

Two main candidates : Oracle and MySQL;

Came up with Oracle in August 1999.

- ∗ *some experience in the past*
- ∗ *widely used at CERN especiallay for accelerators control, engineering and administrative applications;*
- ∗ *versions for all software/hardware platforms;*
- ∗ *supposed to live long enough;*
- ∗ *First tests in September 1999. All catalogues (STS91, MC datasets, access from Web, approx. 50 tables) stored in Oracle.*

Oracle to store Event Tags

- Environment :
  - AlphaServer 4100, quad-CPU (4*600MHz), 2GByte RAM, 0.3TB Raid array;
  - Digital UNIX (TruUnix 64) V4.0D (ams.cern.ch);
  - Oracle v8.1.7
  - Root v2.25 (TruUnix)
- AMS01 Reconstruction Status - 32 bits unsigned integer
  - charge, momentum sign, track pattern, β, geomagnetic latitude, ... total 16 parameters (1 to 5 bits per parameter).
- Flat files : 2424 files, tags : array of unsigned integers;
- Root files : 2424 files, Root Tree 1 per run;
- Oracle :
  - 10 partitions, 1 per day of flight
  - OracleN : tag is stored in one column (nimber(10)), no indices;
  - OracleI : tag is stored in one column (number(10)) + indices;
  - OracleS : tag is stored in 16 columns, no indices;
- 98.7M tags;

Oracle to store Event Tags

Indices for recostatus

- *charge : floor(mod(recostatus,256)/32)*
- β *: floor(floor(mod(recostatus,131072)/1024)/16)*
- ...

It takes 5-8 minutes to create indices for 98.7M tags on ams.cern.ch (8 Oracle processes are active simultaneously)

Queries :

(I) *charge == 1 AND tracker_quality* $>$ *0 AND* $\beta < 2$
(II) *charge == 1 AND tracker_quality* $>$ *0 AND* $\beta < 2$ *AND RUN* $=$ *Y*

takes 3.86 seconds
(1.38 M tags are matched the query (I)).

600 seconds for the version without indices

Oracle to store Event Tags

Method	Size (GB)	Query I time (sec)	Query II time (sec)	Time to populate dbase (sec)	Time per record (usec)
Flat Files	1.4	600	-	-	
Root (cl=1, sm=1)	0.9	700	8	2168	22
OracleN	3.4	1420	80	6467	66
OracleS	6.6	600	55	6467	66
OracleI	3.4	3.9	3.9	6467	66

Table 2: Space occupied to store AMS01 Event status

TDVs. Flat Files vs Oracle

- Time Dependent Values :
  - ∗ TDV name and ID;
  - ∗ validity time : begin / end
  - ∗ insert time
  - ∗ array of unsigned integers
  - ∗ size : 100 Byte - 8MByte
- – (a) TOF Temperature (156 Byte), 9835 files/records;
- – (b) Tracker Pedestals (101 KByte), 330 files/records;
- – TDV array defined as BLOB [1] in Oracle table

TDVname	Number of records	Flat Files (MB)	Oracle (MB)	msec per record to populate database
TOFTemperature	9835	1.9	2.8	17
TrackerPedestals (a)	330	36.3	45.2	75
TrackerPedestals (b)	330	36.3	44.9	103

Table 3: Space occupied to store AMS01 TDV

- – (a) BLOB array is stored inside the table
- – (b) BLOB array is stored outside of the table

The actual space overhead is less, than 20% (*TDV Id, insert, begin* and *end* time are duplicated in BLOB and table).

[1] *(BLOB - binary large object)*

Conclusion from tests

- For AMS01 event tags Oracle query time is 150 times better in comparison with search time for Root files or flat files

- it needs 2.4 and 3.8 times more disk space to store event tags in Oracle than in flat files or ROOT files respectively.

- it needs 20% more disk space to store TDVs in Oracle than in flat files

TDVs are stored in Oracle (TDV's size from 100 byte to 7.8 MB) using method "a".

AMS Data Production

- Oracle server[2] :
  - AlphaServer 4100, quad-CPU (4*600MHz), 2GByte RAM (ams.cern.ch);
  - Digital UNIX (TruUNIX 64) V4.0D;
  - Oracle 8.1.7
  - catalogues
  - TDVs
  - Tags
  - Production status
- Processing Nodes and Disk Severs
  - 1.5 GHz Pentium IV
  - 1.2 GHz AMD Athlon
  - dual-CPU 933 MHz Pentium III
  - dual-CPU 600 MHz Pentium III
  - 3 dual-CPU 450 MHz Pentium II
  - Linux RH 6.1

[2] ams.cern.ch is also used for interactive analysis and batch processing, average system loading 70%

AMS Data Production

- Raw data NFS'ed to Linux machines via 100MBit/s dedicated ethernet segment
- Magnetic Field Map, Calibrations, Slow Control parameters "read" from Oracle
- New calibrations, Tags, Catalogues, Status info "write" to Oracle
- ESD : Ntuples ("write" to local and NFS'ed disks)
- All production processes run on Linux machines (Oracle I/O via net)
  - 5 AMS server processes (Oracle clients)
  - 20 AMS producer processes (clients of AMS servers)
- CORBA client/server for interprocess communication and control (communication via private ethernet segment)
- LSF to start/stop/kill new servers, producers
- no user intervention is required for process control (except the starting of primary server, once at the begining of production)
- minor problem (misunderstanding) with MThreads in Oracle

AMS Data Production

Conclusions :

- **Oracle I/O performance and space overhead satisfy to AMS data store requirements.**
- **AMS TDVs, event tags and catalogues are stored in Oracle RDBMS**
- **AMS production software (Oracle+Corba) run 24h/day, 7 days/week, no crashes or memory leak is observed**
- **Running Oracle database server on Alpha with clients on Linux nodes doesn't cause any performance degradatation (93% efficiency with Oracle, 96% efficiency "pure" Corba and flat files),** *(reminder : no dedicated dbase machine yet, Oracle server runs on GPC)*
- **The performance of AMD based computers is approx. 30% better in comparison with Pentiums for AMS reconstruction code;**
- **Oracle, C++ compiler(s), Corba are compatible**
- **The choice of database for AMS02 is finalized**
- **but non-serialiazed MThread for Oracle applications needs deeper investigation**