Management system of a vast number of operational jobs at JMA

Program Section,
Numerical Prediction Division,
Japan Meteorological Agency (JMA)

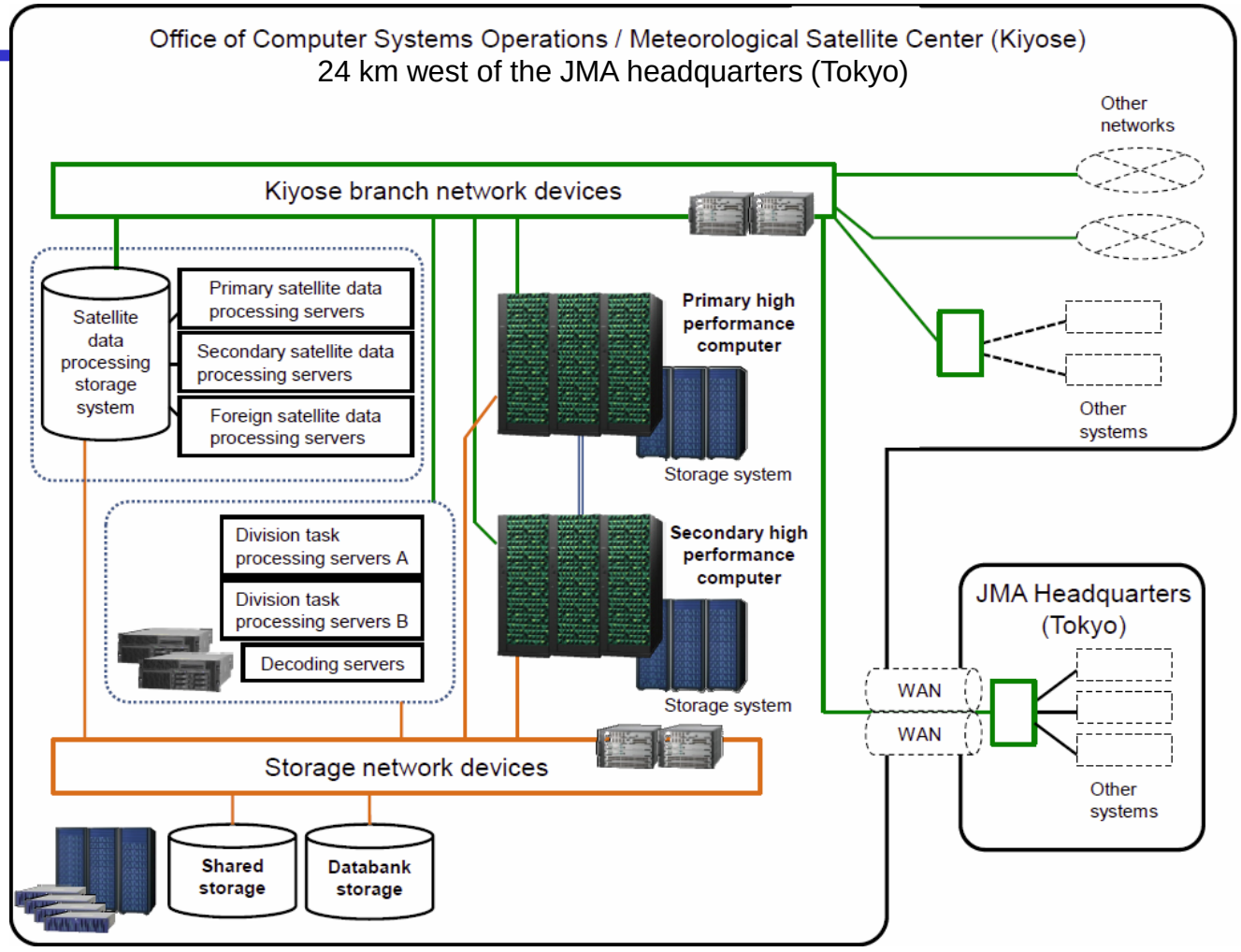
Contents

- Current computer system
- Operational suite
- Management system of operational jobs
- Summary

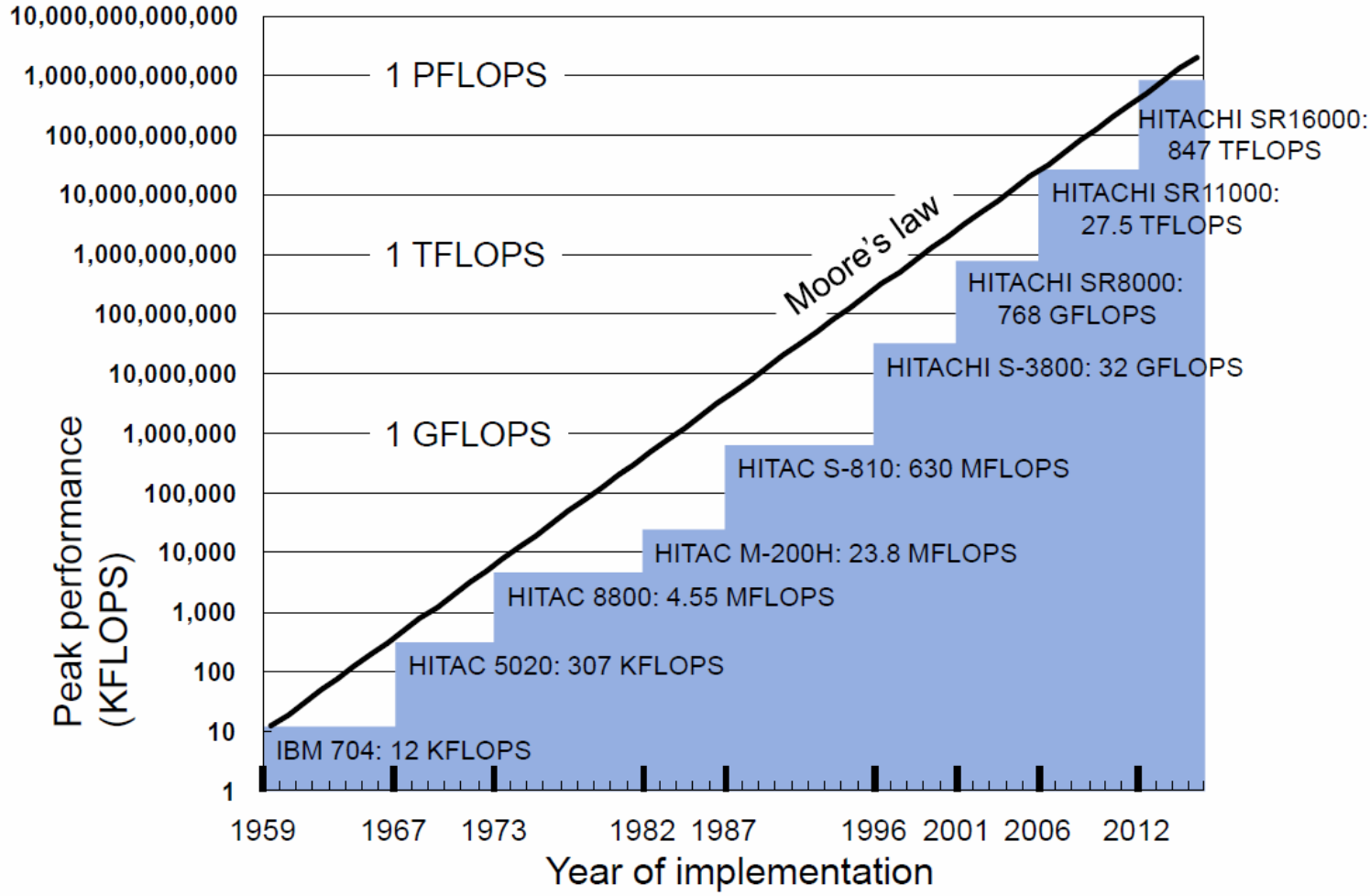
High performance computer

- HITACHI SR16000 model M1
  - Processor: IBM POWER7 (3.83 GHz), 8 cores per chip
  - One logical node: 4 processors (980.48 GFLOPS)
    - Each logical node runs AIX 7.1 operating system
  - One system: 432 logical nodes (423.5 TFLOPS)
    - 412 computational nodes (403.9 TFLOPS)
      + 10 I/O nodes + 4 service nodes + 6 spare nodes
  - Two independent systems with the same specifications
  - In operation since 5 June 2012

Computer system



History of computers at JMA



Contents

- Current computer system
- Operational suite
- Management system of operational jobs
- Summary

Operational suite, analysis (1)

	Global analysis (current)	Global analysis (near future)
Analysis scheme	4D-var	←
Horizontal resolution [inner model]	TL959 (~ 20 km) [TL319 (~ 60 km)]	←
Vertical layers	60 (up to ~ 0.1 hPa)	100 (up to ~ 0.01 hPa)
Analysis time	00, 06, 12, 18 UTC	←
Assimilation window	T–3 ~ T+3	←

Operational suite, analysis (2)

	Mesoscale analysis (current)	Mesoscale analysis (near future)
Analysis scheme	4D-var	←
Horizontal resolution [inner model]	5 km (721 × 577) [15 km (241 × 193)]	5 km (817 × 661) [10 km (409 × 331)]
Vertical layers	50 (up to ~ 22 km)	75 (up to ~ 30 km)
Analysis time	00, 03, 06, 09, 12, 15, 18, 21 UTC	←
Assimilation window	T–3 ~ T+0	←

Operational suite, analysis (3)

	Local analysis (current)	Local analysis (near future)
Analysis scheme	3D-var	←
Horizontal resolution	5 km (633 × 521)	←
Vertical layers	50 (up to ~ 22 km)	60 (up to ~ 20 km)
Analysis time	00, 03, 06, 09, 12, 15, 18, 21 UTC	Every hour

Operational suite, forecast (1)

	Global forecast (current)	Global forecast (near future)
Governing equation	Primitive equation	←
Horizontal resolution	TL959 (~ 20 km)	←
Vertical layers	60 (up to ~ 0.1 hPa)	100 (up to ~ 0.01 hPa)
Initial time	00, 06, 12, 18 UTC	←
Forecast time	84 hours (00, 06, 18 UTC) 216 hours (12 UTC)	84 hours (00, 06, 18 UTC) 264 hours (12 UTC)

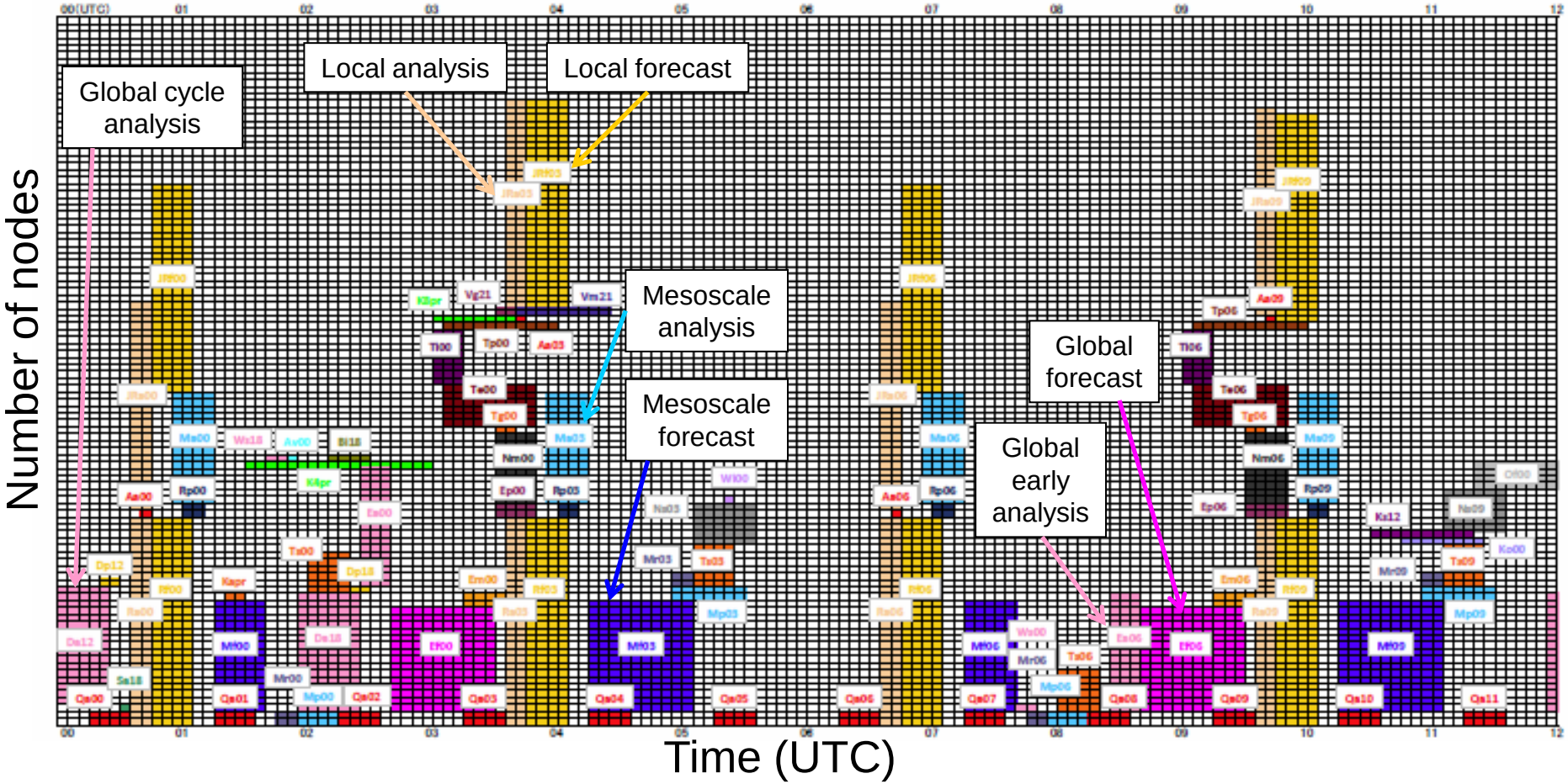
Operational suite, forecast (2)

	Mesoscale forecast (current)	Mesoscale forecast (near future)
Governing equation	Fully compressible nonhydrostatic equation	←
Horizontal resolution	5 km (721 × 577)	5 km (817 × 661)
Vertical layers	50 (up to ~ 22 km)	75 (up to ~ 30 km)
Initial time	00, 03, 06, 09, 12, 15, 18, 21 UTC	←
Forecast time	15 hours (00, 06, 12, 18 UTC) 33 hours (03, 09, 15, 21 UTC)	39 hours

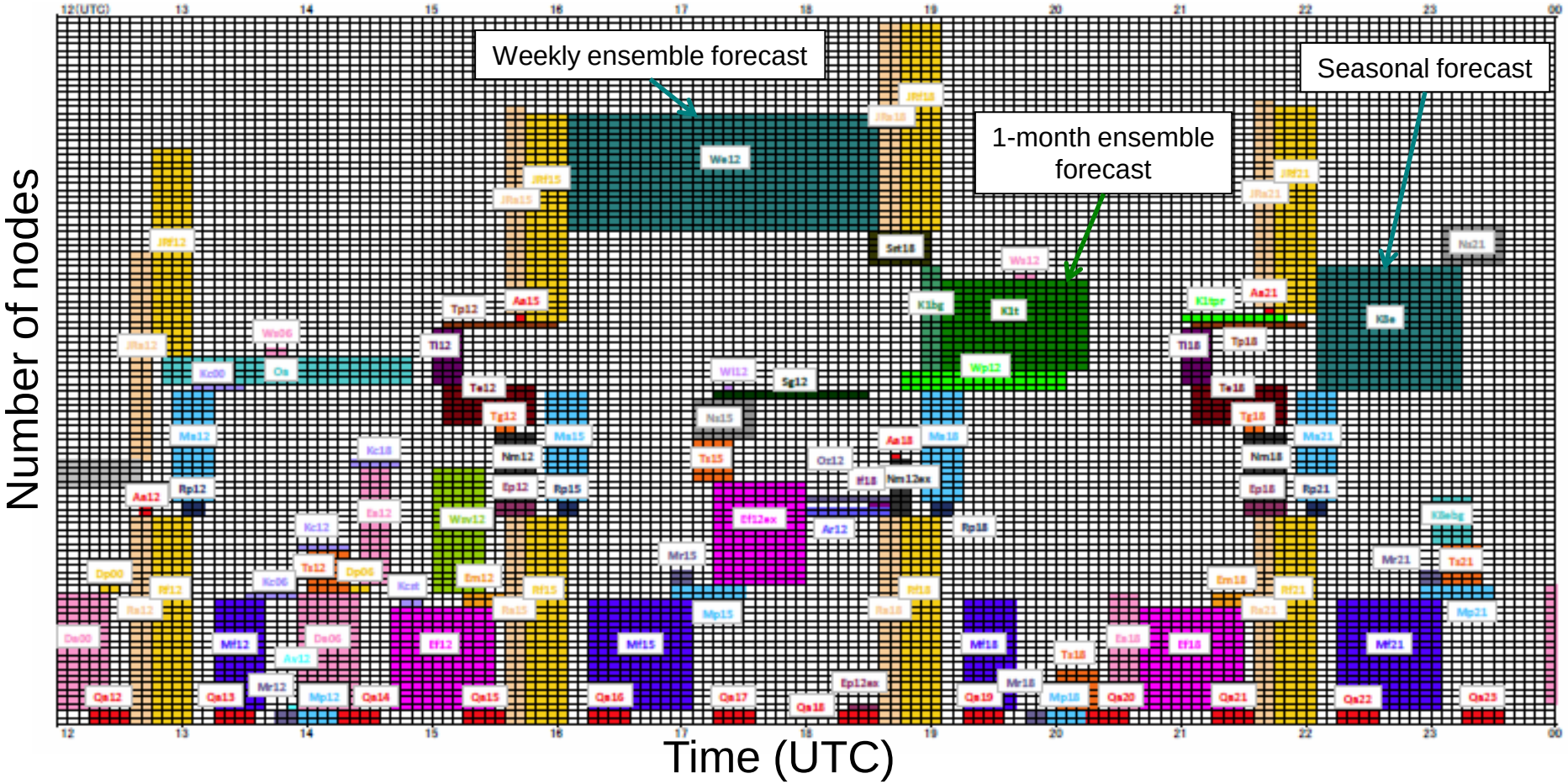
Operational suite, forecast (3)

	Local forecast (current)	Local forecast (near future)
Governing equation	Fully compressible nonhydrostatic equation	←
Horizontal resolution	2 km (551 × 801)	2 km (1581 × 1301)
Vertical layers	60 (up to ~ 20 km)	←
Initial time	00, 03, 06, 09, 12, 15, 18, 21 UTC	Every hour
Forecast time	9 hours	←

Daily schedule of operational suite: 00–12 UTC



Daily schedule of operational suite: 12–00 UTC



Contents

- Current computer system
- Operational suite
- Management system of operational jobs
- Summary

Step, job, job group

- Step
  - The smallest unit in a management system including only one executable
    - A shell script, an awk script, or a load module
- Job
  - A step or a group of several steps that run sequentially
- Job group
  - A group of several jobs

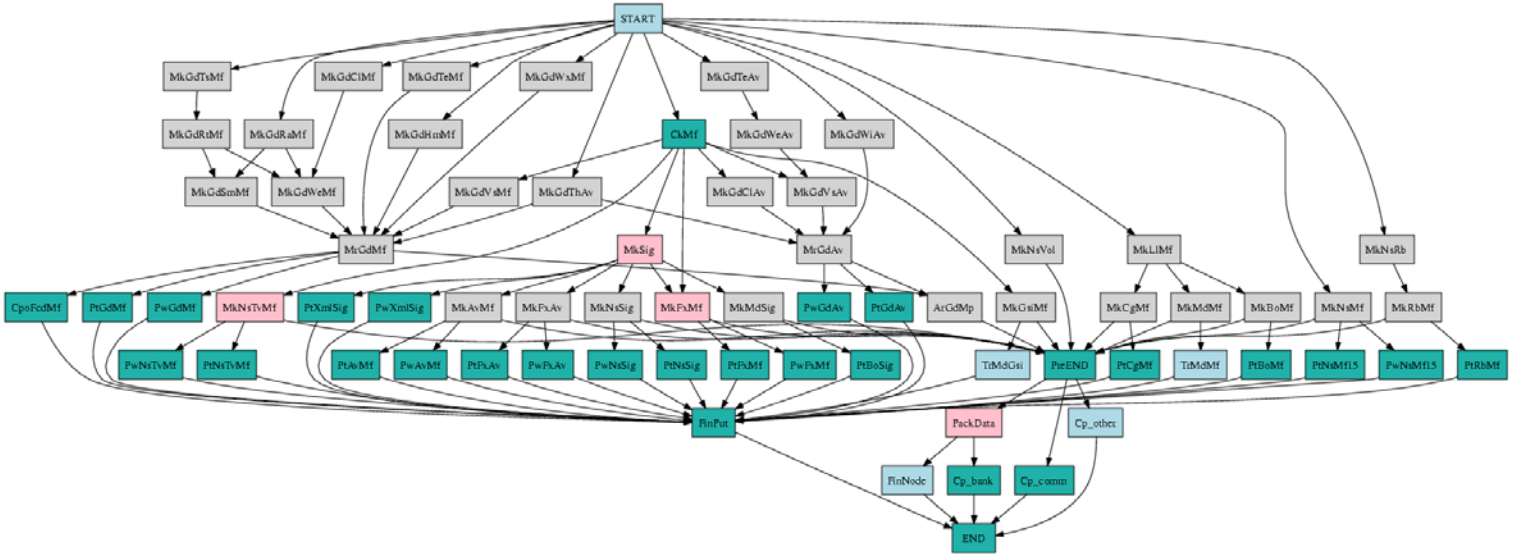
Operational job groups

- Number of operational job groups ~ 70
  - Global analysis, global forecast, mesoscale analysis, mesoscale forecast, local analysis, local forecast, etc.
- Dependencies between job groups
  - 06 UTC analysis before 06 UTC forecast
  - 06 UTC forecast before 09 UTC analysis
  - Parent forecast model giving lateral boundary conditions before child forecast model
  - etc.

Operational jobs

- Number of operational jobs ~ 10,600 per day
  - Complicated dependencies between jobs in a job group

e.g. Diagram of a job group (products of mesoscale forecast), each box indicates a job



Datasets

- Datasets consist of directories and files
- Number of datasets
  - Constant datasets
    - Parameters, namelists, climatic average data, etc.
    - Number of datasets ~ 3,000
  - Variable datasets
    - Output from analyses, forecasts and products, etc.
    - Number of datasets ~ 8,800

Lots of files (e.g. Mesoscale forecast)

- 10 input datasets
  - Namelists for model settings
  - Initial conditions for atmosphere, land surface, sea surface
  - Lateral boundary condition
  - Topography data, ozone data, aerosol data, etc.
- 15 output datasets
  - Forecast for surface, isobars, model layers
  - Physics monitors, etc.
- Complicated dependencies between lots of files

Management system of operational jobs (1)

- To assure operational jobs run correctly by eliminating man-made errors
- To manage operational jobs systematically
  - All the information about jobs, datasets, executables is stored in a database management system (DBMS)
- Started to develop in 2004
  - Started to develop some parts of the system in 2000
- Installed in the operational system in 2006

Management system of operational jobs (2)

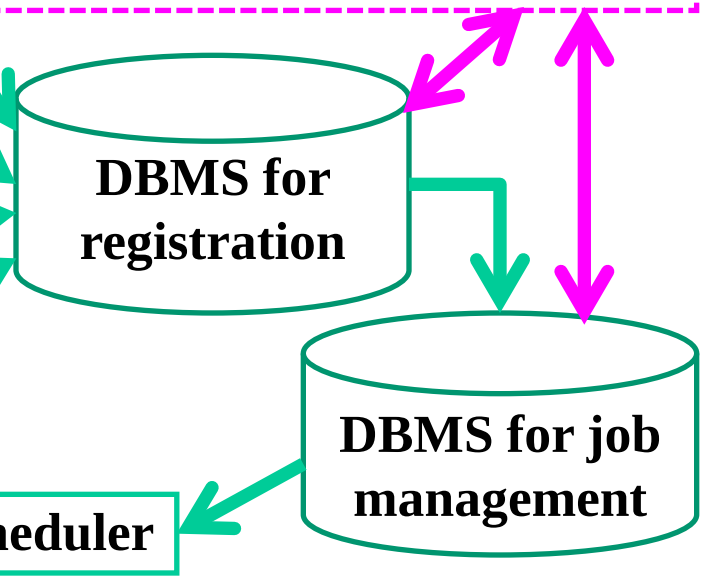
Registration form (Microsoft Excel sheets with VBA macros):
Information about job groups, jobs, datasets, executables, etc.
Submitted when a new job is added, or existing datasets or executables are updated

Job definition file (JDF):
Information about a job group and jobs including the schedule (time to run), the order (preceding job groups and jobs), computational resources required, etc.

Job control language (JCL):
Converted into a shell script

Program build file-format (PBF):
Converted into a makefile to compile load modules

Utility programs to register information and check the consistency, etc.



New job control language (JCL)

- Similar to a JCL used for mainframe computers
- In the job management system of JMA, a JCL is converted into a Unix shell script using a utility program and DMBS for job management
  - Treatment of directory names
    - In a JCL, abstract directory names are used
    - In a shell script, abstract directory names are converted into actual directory names defined in the DBMS
      - Actual directory names differ whether the job is operational or not

Sample of a simple JCL

```
job JGNAME:JOBNAME
step step_name1
  pgm jpp:/exe/Comm/program1
  dd in  name=INPUT  data=jpp:/jgdir/Test/input1.txt
  dd out name=OUTPUT data=jpp:/jgdir/Test/output1.txt
```

- Defining a job including only one step
  - One executable, one input file, one output file
- Number of lines in JCL = 5
- Statically defined action
- Written by developers

Sample of a shell script converted

Converted automatically from the JCL shown in the previous slide
Number of lines in the shell script = 148 (snipped here)

```
test -f ${JGDIR:?}/Test/input1.txt  || exit 199
test -d ${JGDIR:?}/Test             || exit 199
test -x mod/Comm/Exe/program1       || exit 199

PGM=mod/Comm/Exe/program1

touch jg.Test.output1.txt
ln -fs ${JGDIR:?}/Test/input1.txt  INPUT   # in
ln -fs jg.Test.output1.txt         OUTPUT  # out

set +e
${PGM:?}  < /dev/null >log.100_step_name1
RC=$?
set -e
rcchk ${RC:?} 0 0
RC_EXIT=${RC:?}
rm -f INPUT OUTPUT
chmod 444 jg.Test.output1.txt

rm -f ${JGDIR:?}/Test/output1.txt
mv -f  jg.Test.output1.txt  ${JGDIR:?}/Test/output1.txt

exit ${RC_EXIT:?}
```

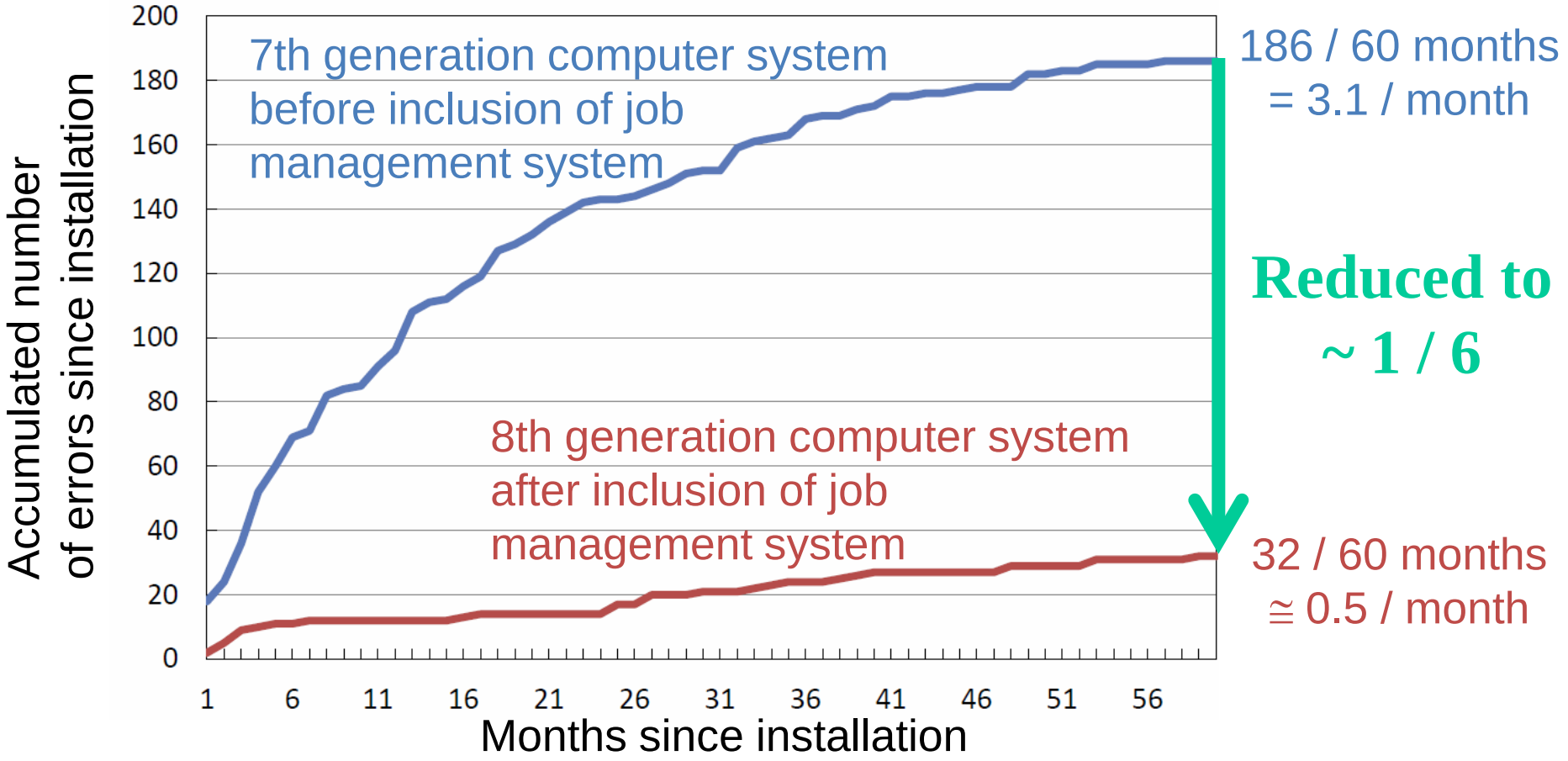
Added Unix `test` commands for an input file, directory, executable

`${JGDIR}` depends on whether this job is operational or not

Implemented 'atomic write':
The program writes to a temporary file `jg.Test.output1.txt` and rename it to a final output name `${JGDIR}/Test/output1.txt` after the program finishes

Added utility functions (not shown)

Reduced man-made errors in operation



Contents

- Current computer system
- Operational suite
- Management system of operational jobs
- Summary

Summary

- Installed a new computer system on 5 June 2012
  - HITACHI SR16000 model M1, 423.5 TFLOPS × 2
- Implemented a job management system
  - Operational suite consists of about 10,600 jobs per day
  - Total number of datasets is larger than 10,000
  - A comprehensive job management system using a database management system was implemented
  - Number of man-made errors was reduced to ~ 1 / 6

Thank you