RFC Operations in a Linux Environment
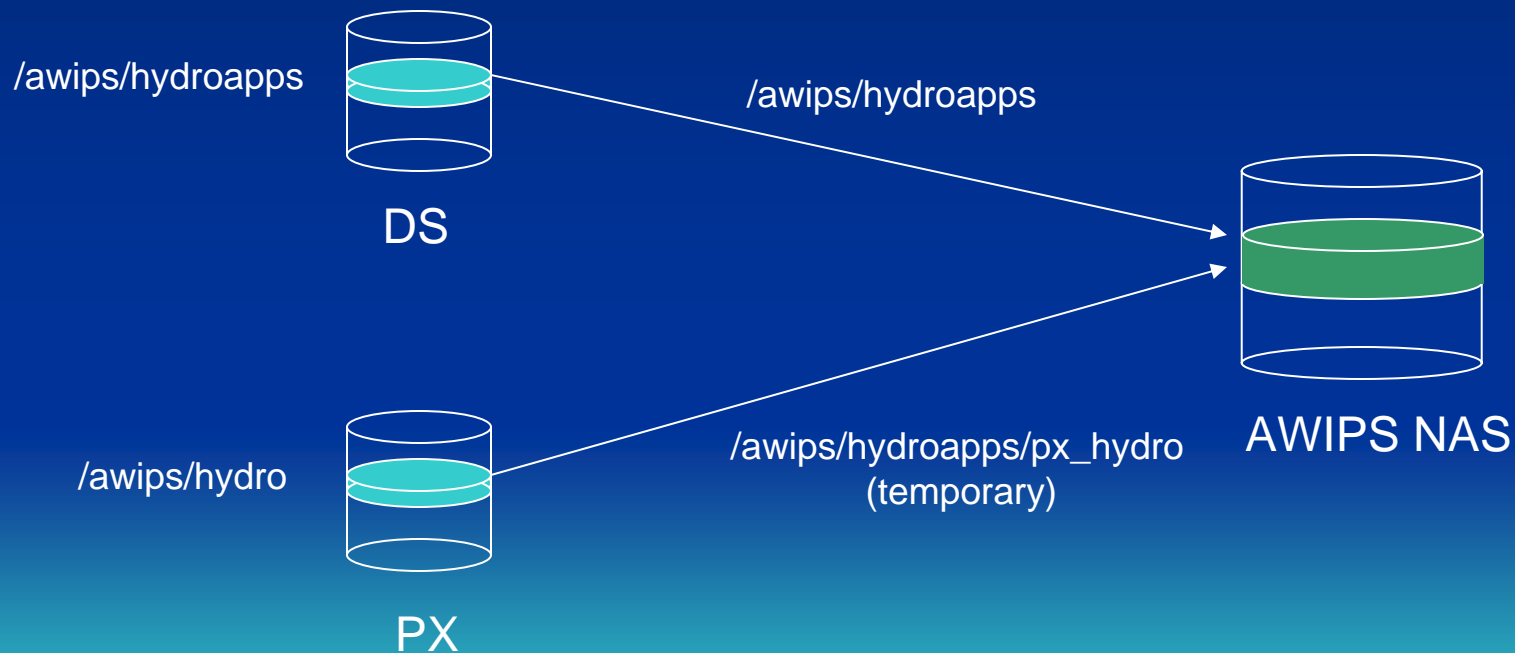
DOH Science Steering Team Conference Call

9 November 2004

Moving Hydro Files to New NAS

- Done via a script as part of the DX/NAS installation

/awips/hydroapps

DS

/awips/hydroapps

/awips/hydro

PX

/awips/hydroapps/px_hydro
(temporary)

AWIPS NAS

Transition to PostgreSQL

DOH Teleconference
Nov 9, 2004

Prepared by Paul Tilles

Transition to PostgreSQL/Schedule

- AWIPS will officially move to PostgreSQL for OB6
- Delivery to field in Fall 2005
- Scripts to convert from OB5/Informix to OB6/PostgreSQL to be provided with install
- New OB6 executables and scripts installed

Transition to PostgreSQL/Schedule

- What can be done "early" at OB4 sites?

Transition to PostgreSQL/OB4

- Sites should download copy of PostgreSQL 7.4.6
- Sites should install on a Linux box via RPM

- Convert Informix IHFS OB4 db to PostgreSQL
  - Sites export Informix db on UNIX machine (OB4 snapshot)
  - Sites run OHD provided scripts on Linux machines to create PostgreSQL db, create tables, primary keys, etc., edit .unl files, load data into tables
  - OHD offers limited support to POCs

Transition to PostgreSQL/OB4

- Sites begin converting code and testing local apps
  - ecpg replaces esqlc
  - psql replaces dbaccess/sqlcmd

Transition to PostgreSQL/OB5

- AWIPS installs OB5 (Spring, 2005)
  - Informix db converted to OB5
  - PostgreSQL version 7.4.5 delivered

- Convert Informix IHFS OB5 db to PostgreSQL
  - Sites export Informix db on UNIX machine (OB5 snapshot)
  - Sites run OHD provided scripts on Linux machine to create PostgreSQL db, create tables, primary keys, etc,edit .unl files, load into tables
  - OHD offers limited support to POCs

Transition to PostgreSQL/OB5

- Sites continue work converting code and testing local apps using AWIPS installed PostgreSQL environment

Transition to PostgreSQL/Lessons Learned

- See HSEB website

  www.nws.noaa.gov/oh/hrl/hseb/index.htm

Transition to PostgreSQL/Current Status at OHD

- Currently testing OB6 versions of

  - shefdecode
  - db_purge
  - DPA decoder
  - HydroBase

- Other apps to follow in next few months

Transition to PostgreSQL Training

- Most requested RFC workshop for FY05
- Will be patterned after the one held in Silver Spring for AWIPS development organizations
- Probably held at the NWS Training Center
- 2nd quarter FY05 (Jan. – Mar. 2005)
- Jeff Zimmerman working out details

Postgresql Performance

Compared to Informix (on Linux)
At CBRFC

Steve Shumate, DSST conf. call 9Nov2004

Methods

- I made a reasonable effort to level the playing field, but in the end it was not perfectly level and differences are described in the next two slides about configuration and conditions

Configurations

- Informix
  - Version 7.31(Linux)
  - Machine is a dual Xeon 2.2Ghz w/1.7G ram
  - Modified configuration
- Postgresql
  - Version 7.4.6 built from source code (not RPM)
  - Machine is a dual Xeon 2.8Ghz w/6G ram
 - Modified configuration

Conditions of the test

- Informix machine was operating with normal operational load

- Postgresql machine was otherwise idle

- So the postgresql machine has the advantage here, being faster with more ram and less load

Performance – mass loading

- 1 million rows into one table
  - Informix, dbload, 1000 rows per transaction
  - Postgresql, psql, \copy

minutes to load 1 million rows

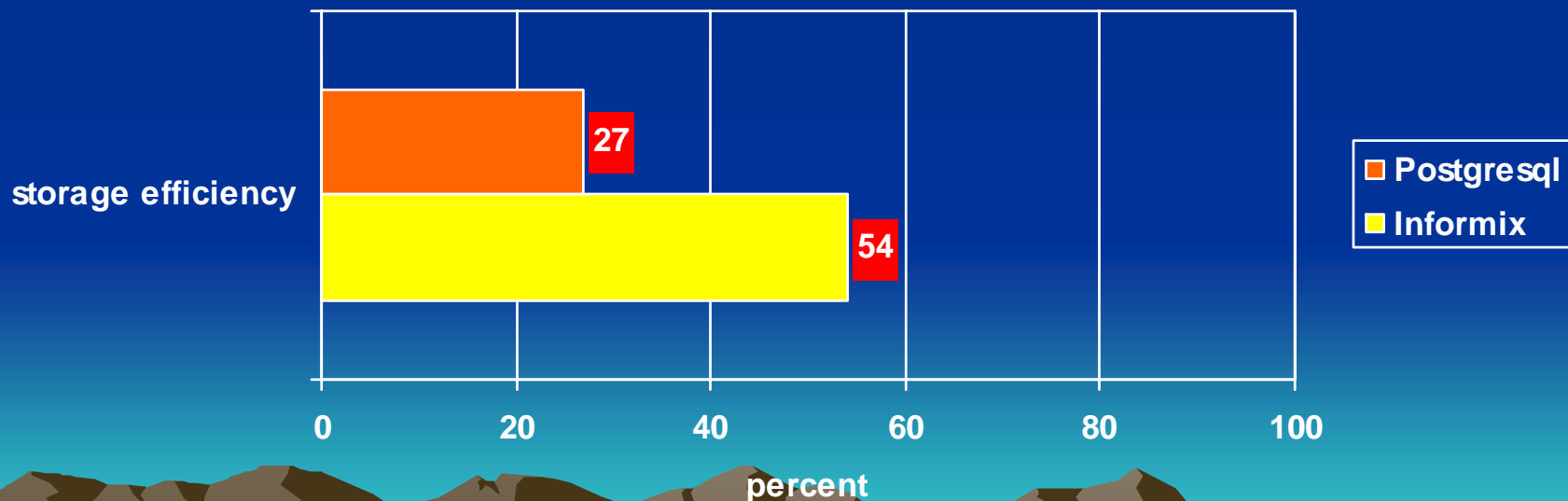| | 24 | Postgresql |
| | 34 | Informix |

0 5 10 15 20 25 30 35 40
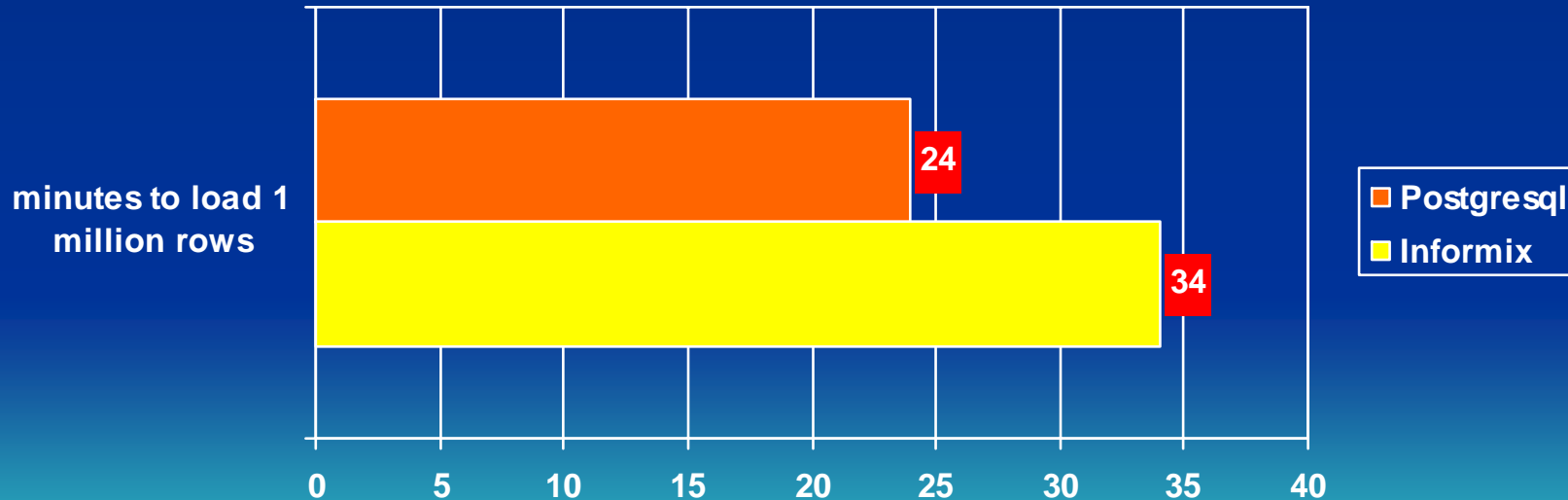
Performance - queries

- Time series query from a 5 million row table with no index – ID and data-type specified without date information
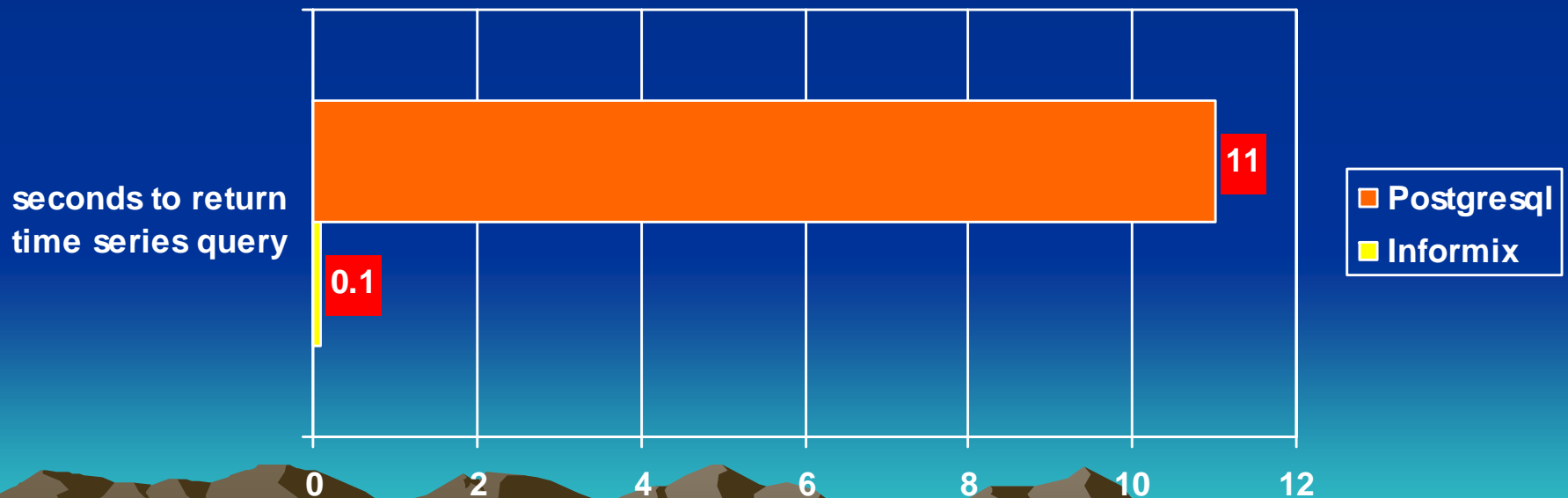
Performance – storage efficiency

- Informix – cooked file space – 37G ascii data took 69G of disk space
- Postgresql – 4.8G ascii data took 18G of disk space

Conclusions

- Mass loading performance is on par with Informix

- Storage efficiency loss can be solved with more disk space (may need 2x the disk space to store the same amount of data)

- Query performance may be similar to Informix (was ~equal when limit 1 was applied to select statement)

Conclusions, cont.

- Better Performance? That depends on where you are coming from
  - Moving from HP-Informix it will be an improvement
  - From Linux-Informix it may be a step back
- Better overall?
  - Open source, free
  - Arrays!

Conclusions, cont.

- Did I mention Arrays? !!!
  - A better, less awkward way (than CBRFC-fastetc method) to store time series data
  - May increase performance
- Maintenance & Fragmentation (extents)
  - Minimal!
  - No inter-table fragmentation with pg
    - Not more than one table per file-system file
  - Vacuum commands will have to be ran routinely. This is essential.

RFC Ensemble Processors

Where we want to go

- Purpose of REP: support Ensemble Forecasting
- Multi-Threaded applications
  - Ensembles, Grids, Data assimilation, iterative model solutions
- Services oriented architecture centered on the REP
- Additional CPU as we need them
- Hydro only DB engine on REP
  - Multiple DB engines

Current Considerations

- Two Network Attached Storage (NAS) devices
  - AWIPS NAS best choice for daily operational data
    - 4 hour support
  - REP NAS best choice for GIS, Calibration data
    - 24 hour support
- Network should be fast enough to support using AWIPS NAS for data and REP for computations
  - Each device on 1 gigabit LAN connected by 1 gigabit switch
  - REP cpu should see /awips/hydroapps
- RFC solutions/suggestions

GIS for RFC Operations

- April 2004 – OST agreed to consider ESRI on MS Windows as part of AWIPS baseline
- May 2004 – presentation by NGIT
  - Offered GRASS as alternative to ESRI
- May-June 2004 - Richards collected GIS requirements
  - Delay because of other priorities (Richards)
  - Revisit in Jan. 2005
  - Have third party evaluate GRASS against requirements?

GIS for RFC Operations

- Is there sufficient consensus to establish requirements for a single baseline solution?

- Are GRASS and ESRI the only options?

- Given budget constraints, importance of a strong, unified approach

GIS for RFC Operations

- FOP

  – National, but support not centralized

  – Use requirements process to:

    (1) Convert to Linux

    (2) Establish formal support

- CAP

  – Not nationally supported

  – Consider submitting as a requirement

GIS Issues

- Enterprise License Agreement with ESRI
  - Although not 'dead,' no reason to expect agreement in near future

- Postgress
  - PostGIS [postgis.refractions.net]
    - 'Middleware' to store/access geospatial info
    - Connectivity with:
      - (1) Minnesota Map Server [mapserver.gis.umn.edu/]
      - (2) GRASS (PostGRASS)
      - (3) GeoTools Java GIS Toolkit
      - (4) ESRI products [www.esri.com/software/arcgis/extensions/datainteroperability/about/features1.html]