2013 Esri International User Conference
July 8–12, 2013 | San Diego, California

Technical Workshop

Administering your Postgres Geodatabase

Jim Gough and Jim McAbee

Agenda

- Requirements and Configuration
- Administration and Connections
- Backup and Upgrade
- Performance and Advanced Topics

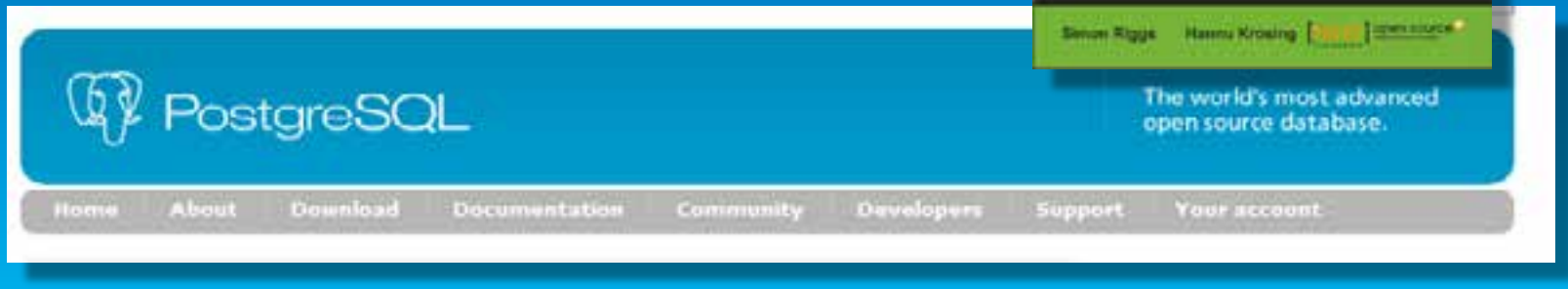
Agenda

- **Requirements and Configuration**
- Supported Versions
- Setup and Configuration
- Spatial Types

PostgreSQL

A Free Open Source Option

- Introduction to PostgreSQL
  - http://www.postgresql.org/
  - Open Source
  - Enterprise level RDBMS
  - Free, distributed with bsd license
  - Supported by an active online development community
- Learn more:
  - planet.postgresql.org,
  - PGCON: http://www.pgcon.org/2012/



Where to get Software?

- PostgreSQL.org and Customer Care Portal: customers.esri.com
  - PostgreSQL Installation
  - PostgreSQL Client Libraries
  - ArcSDE Installation – ESRI Customer Care Portal only
  - st_geometry library
    - In all ArcGIS clients

Postgres Versions and Supported OS

tested and certified: verify at support.esri.com

9.2.2	Linux: Red Hat 5, 6 Windows Server 2012 & 2008 R2
9.1.3	Linux: Red Hat 5 Windows Server 2008 R2
9.0.5	Linux: Red Hat 5, 6 & SUSE 11 Windows Server 2008 R2 & 2003

ArcGIS 10.2

64-bit DBMS and OS only

PostGIS 1.5.1 and 2.0

9.1.3	Linux: Red Hat 5 Windows Server 2008 R2
9.0.5	Linux: Red Hat 5, 6 & SUSE 11 Windows Server 2008 R2 & 2003

ArcGIS 10.1

64-bit DBMS and OS only

PostGIS 1.5.1 and 2.0 *

**Support begins at 10.1 SP1. See technical article 40553 for more information.*

8.4.1	Linux: Red Hat 5 & SUSE 10 Windows Server 2008 R2 & 2003
8.3.8	Linux: Red Hat 5 & SUSE 10 Windows Server 2008 R2 & 2003

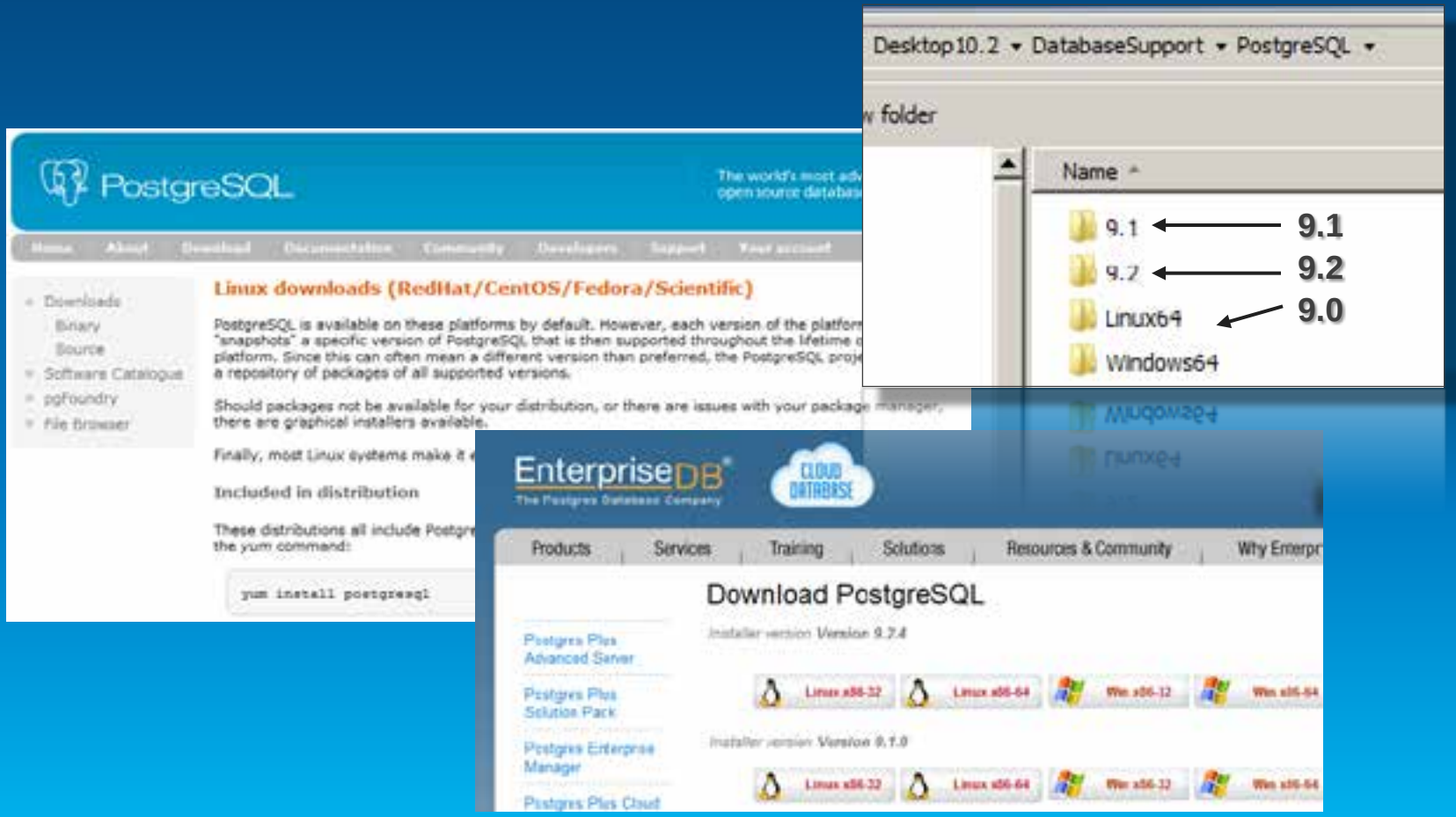
ArcGIS 10

32 & 64-bit DBMS and OS

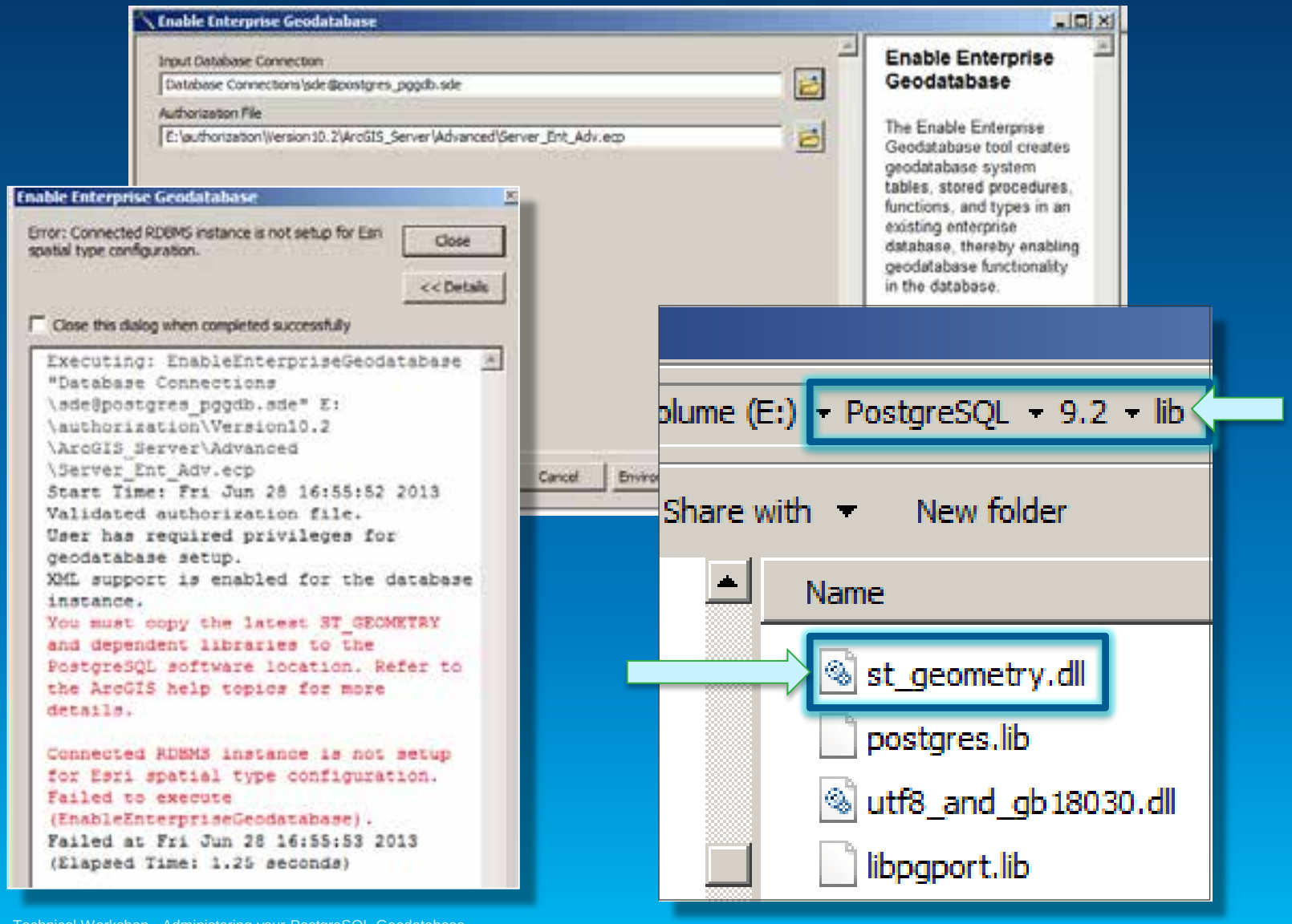
PostGIS 1.4.0 supported

Linux and Windows Notes

- Installation of PostgreSQL
  - one click Installer, yum install, etc…
  - st_geometry.so, st_geometry.dll distributed with ArcGIS

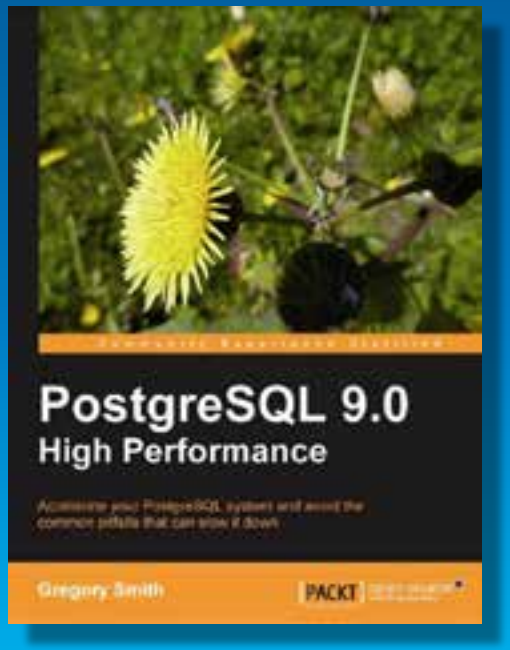


Configuring PostgreSQL



PostgreSQL Initialization Parameters

- postgresql.conf
  - most defaults ok, testing and monitoring should be done
- Memory
  - #shared_buffers=32MB …
  - Windows – best 64MB to 512MB, little benefit to set higher, use OS cache
  - Linux – 25% of Physical Memory to start and as possible (no. of instances)
  - PostgreSQL 9.0 High Performance – Gregory Smith
- Query Optimization
  - cursor_tuple_fraction - set to 1.0 vs. default of 0.1 (per 10.1 SP1 notes)



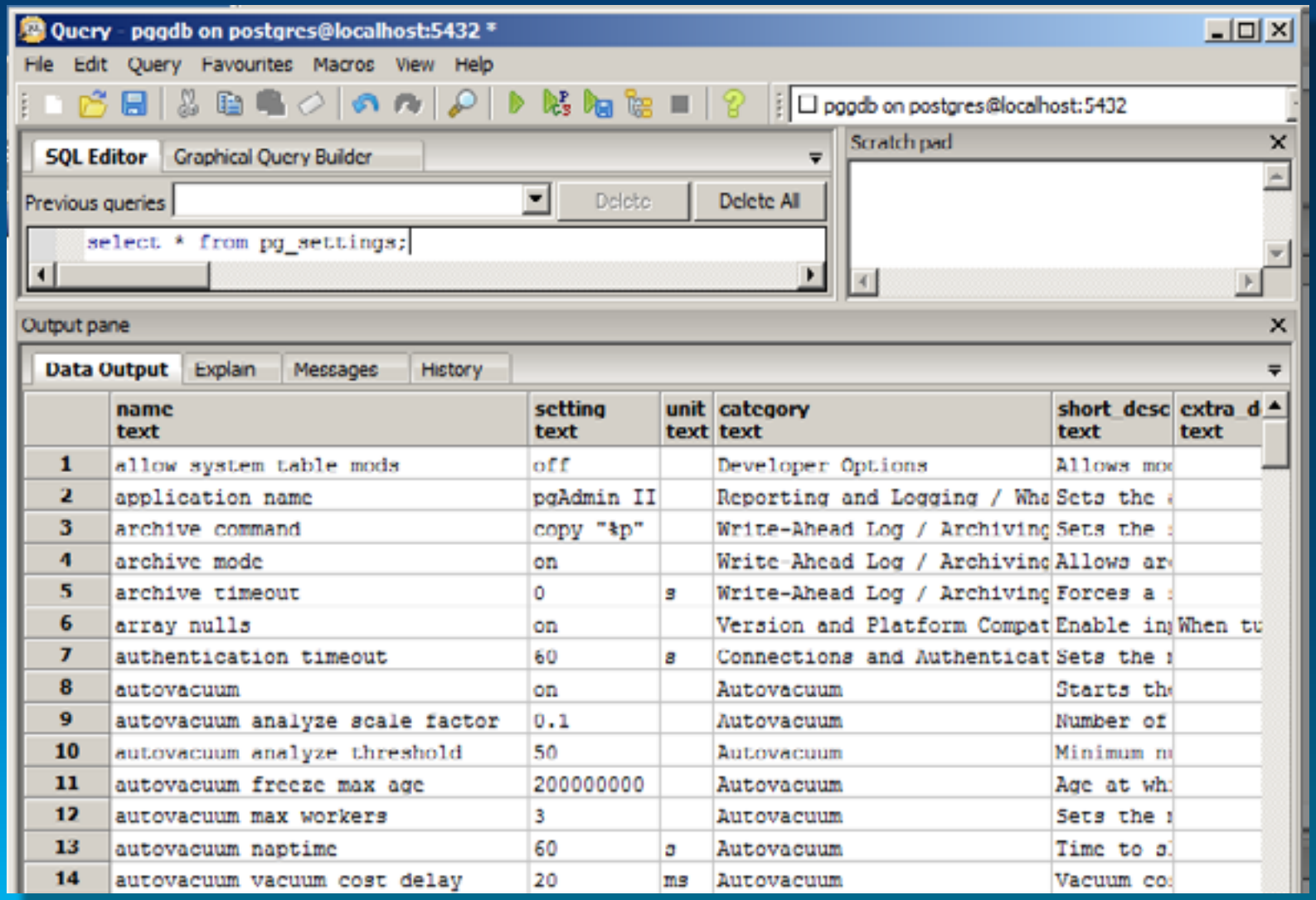
PostgreSQL Initialization Parameters

- Connections
  - #max_connections=100 (default)
  - one ArcGIS connection = multiple PostgreSQL connections
  - default max connections for Geodatabase (sde_server_config)
  - #superuser_reserved_connections=3
- Logging
  - #log_statement = 'none'
- Vacuum/ Analyze
  - #autovacuum = on

	prop_name [PK] character varying(32)	char_prop_value character varying(5	num_prop_value integer
1	ALLOWSESSIONLOGFILE		0
2	ATTRBUFSIZE		50000
3	AUTH_KEY	arcsdeserver,101	
4	AUTOCOMMIT		1000
5	BLOBMEM		1000000
6	CONNECTIONS		64
7	DEFAULTPRECISION		64

Reviewing Configuration Parameters

select * from pg_settings;



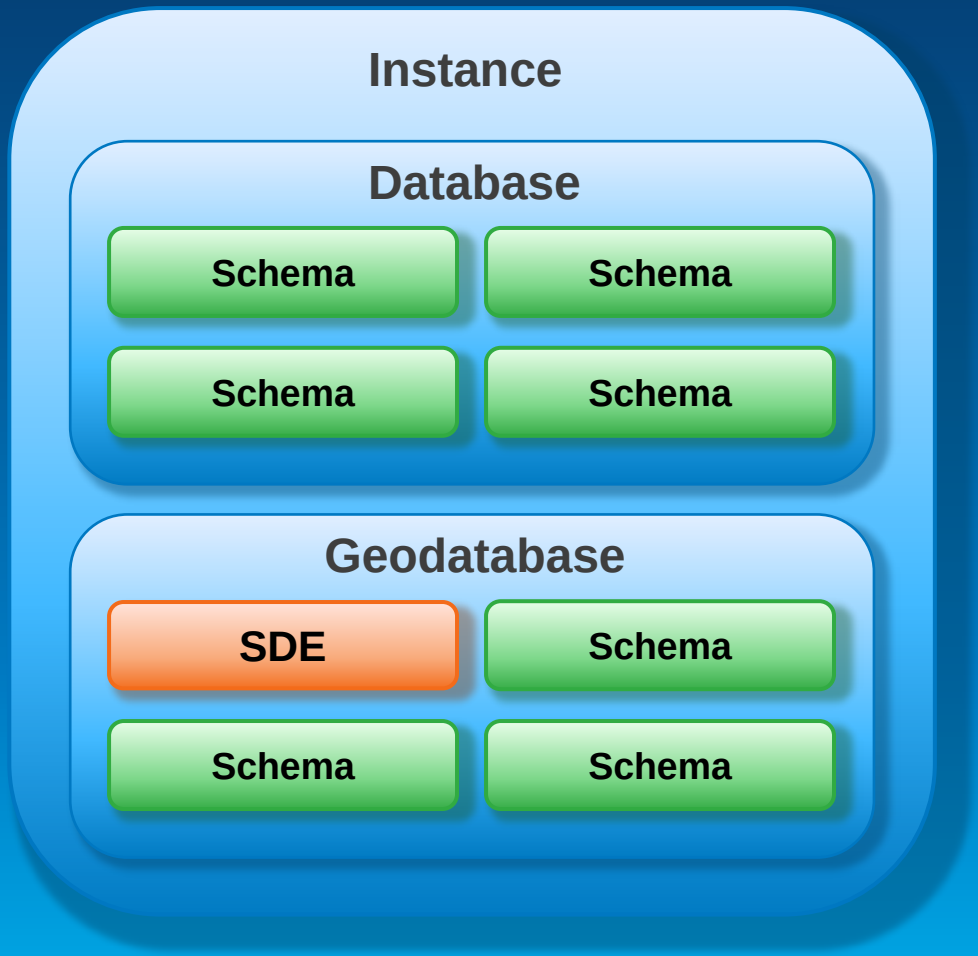
	name text	setting text	unit text	category text	short desc text	extra d text
1	allow system table mods	off		Developer Options	Allows mo	
2	application name	pgAdmin II		Reporting and Logging / Wha	Sets the	
3	archive command	copy "%p"		Write-Ahead Log / Archiving	Sets the	
4	archive mode	on		Write-Ahead Log / Archiving	Allows ar	
5	archive timeout	0	s	Write-Ahead Log / Archiving	Forces a	
6	array nulls	on		Version and Platform Compat	Enable in	When tu
7	authentication timeout	60	s	Connections and Authenticat	Sets the	
8	autovacuum	on		Autovacuum	Starts th	
9	autovacuum analyze scale factor	0.1		Autovacuum	Number of	
10	autovacuum analyze threshold	50		Autovacuum	Minimum n	
11	autovacuum freeze max age	200000000		Autovacuum	Age at wh	
12	autovacuum max workers	3		Autovacuum	Sets the	
13	autovacuum naptime	60	s	Autovacuum	Time to s	
14	autovacuum vacuum cost delay	20	ms	Autovacuum	Vacuum co	

Upper Case Database Identifiers

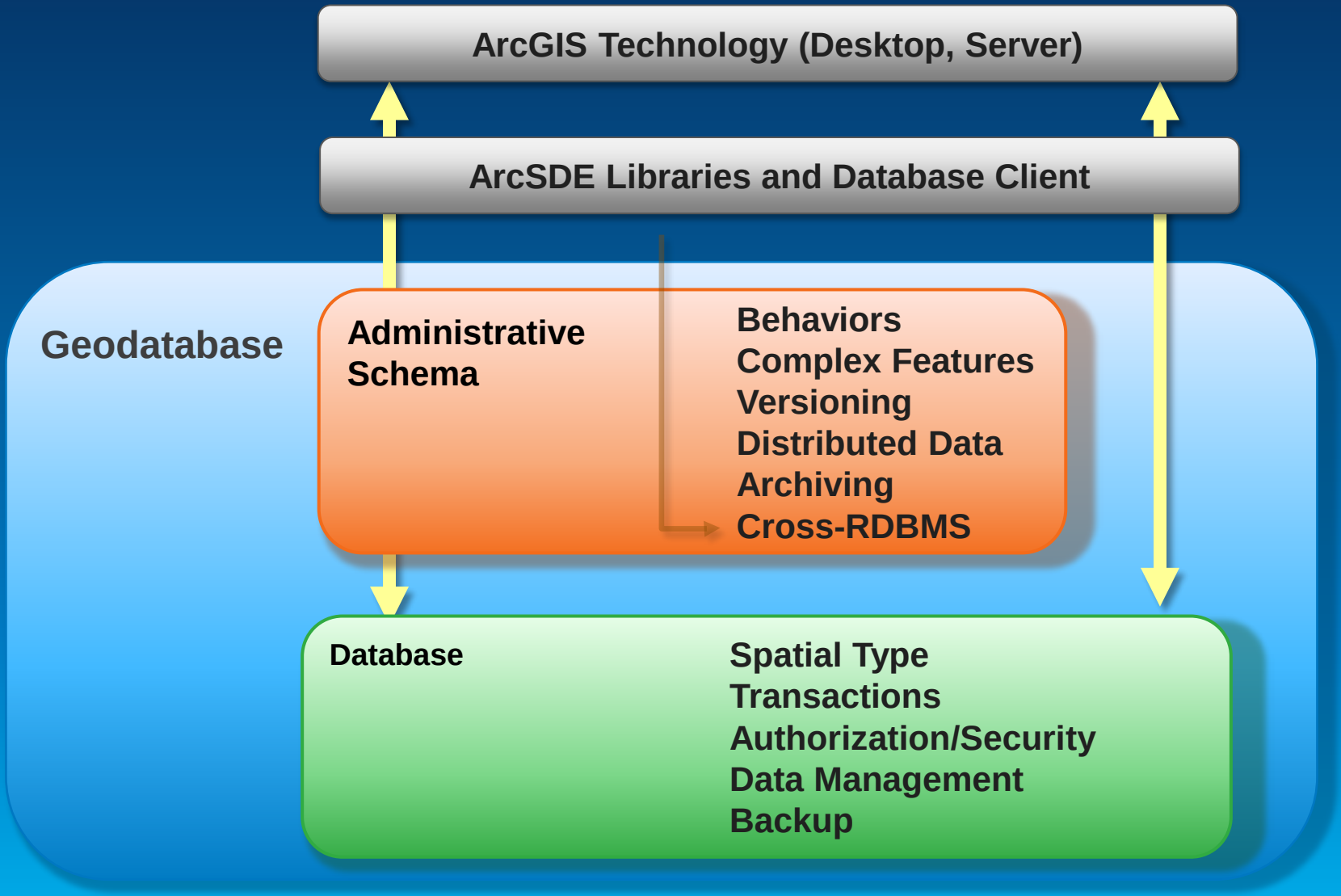
- PostgreSQL is case sensitive
  - It stores all of it's object identifies in lower case
    - Names of: Databases, Tables, Indexes, Column
    - SDE/GDB also stores all identifiers in lower case
    - User data can be in any case
  - To use identifiers in upper case, they need to be quoted
    - PgAdminIII quotes them automatically
  - ArcGIS does not look for quoted strings
    - Identifiers with upper case names not found

```
CREATE DATABASE "KASIA"
  WITH ENCODING='UTF8'
       OWNER=postgres
       CONNECTION LIMIT=-1;
```

Postgres: Instance, Database and Geodatabase



Geodatabase: Technology Components

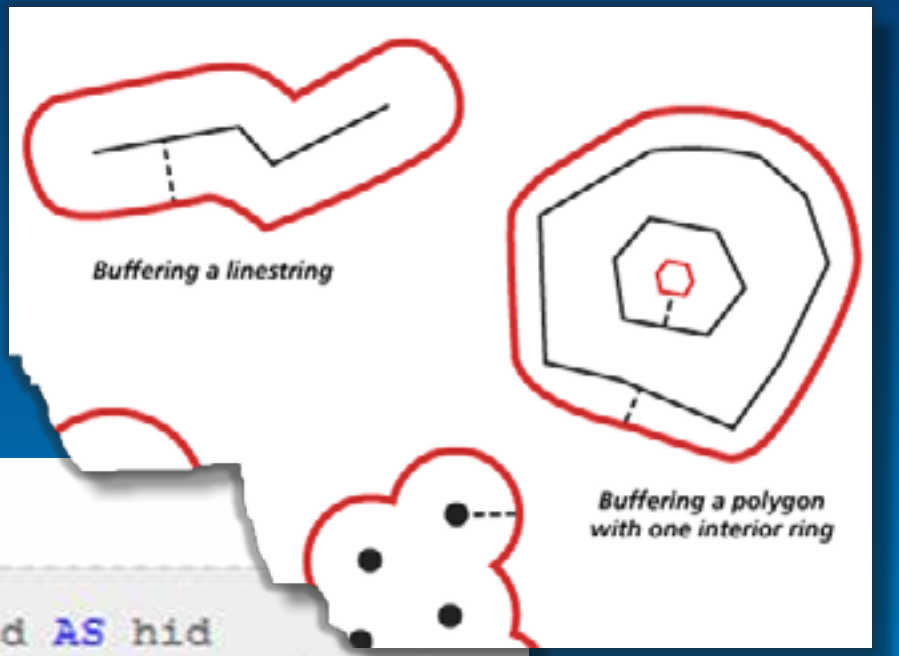


Multiple Geodatabase Configuration

- Multiple Geodatabases in PostgreSQL
  - In same instance
  - If using same name in multiple instances (e.g. clone prod to stage)
- Reasons for multiple geodatabases
  - Editing and publishing (web)
  - Production and Staging
  - Different application needs
  - Separate version management

Spatial Types and Functions

- Creation of Features through SQL
- Analysis through SQL



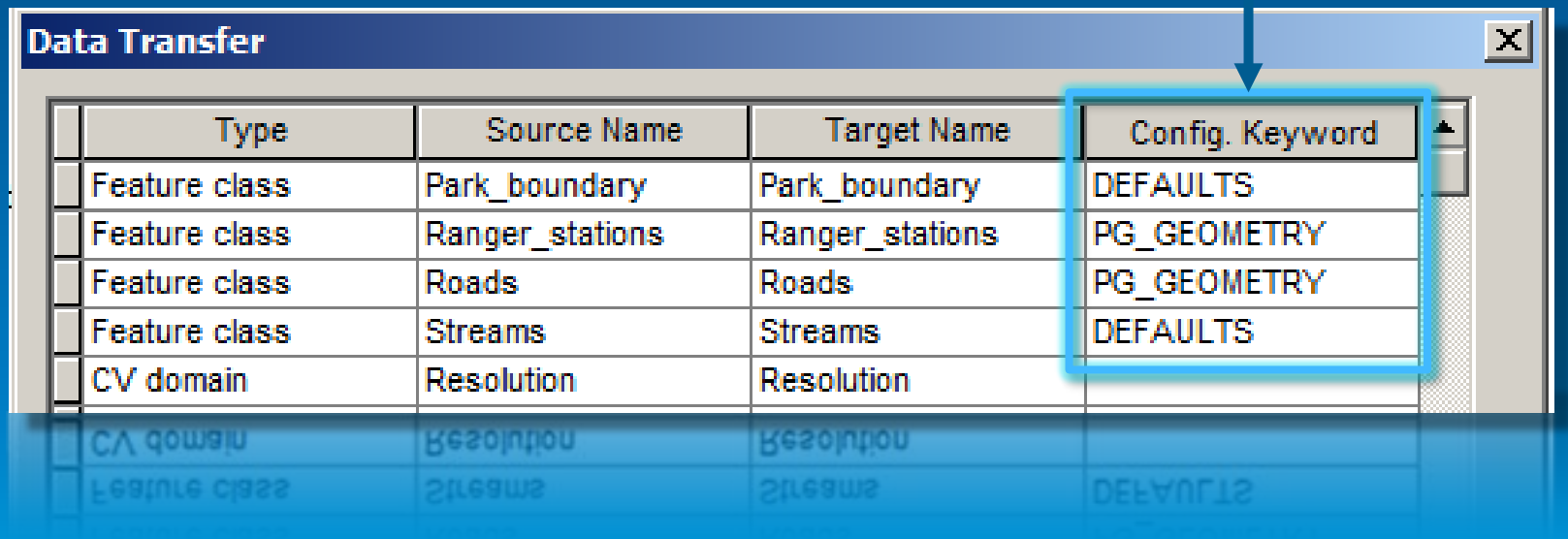
PostgreSQL

```
SELECT sa.id AS sid, hs.id AS hid
FROM sensitive_areas sa, hazardous_sites hs
WHERE sde.st intersects (sde.st buffer
(hs.site, .1), sa.shape) = 't'
ORDER BY sa.id;
```

sid	hid
1	5
2	5

Spatial Type Storage

SDE.SDE_DBTUNE
table for storing keywords and associated parameters

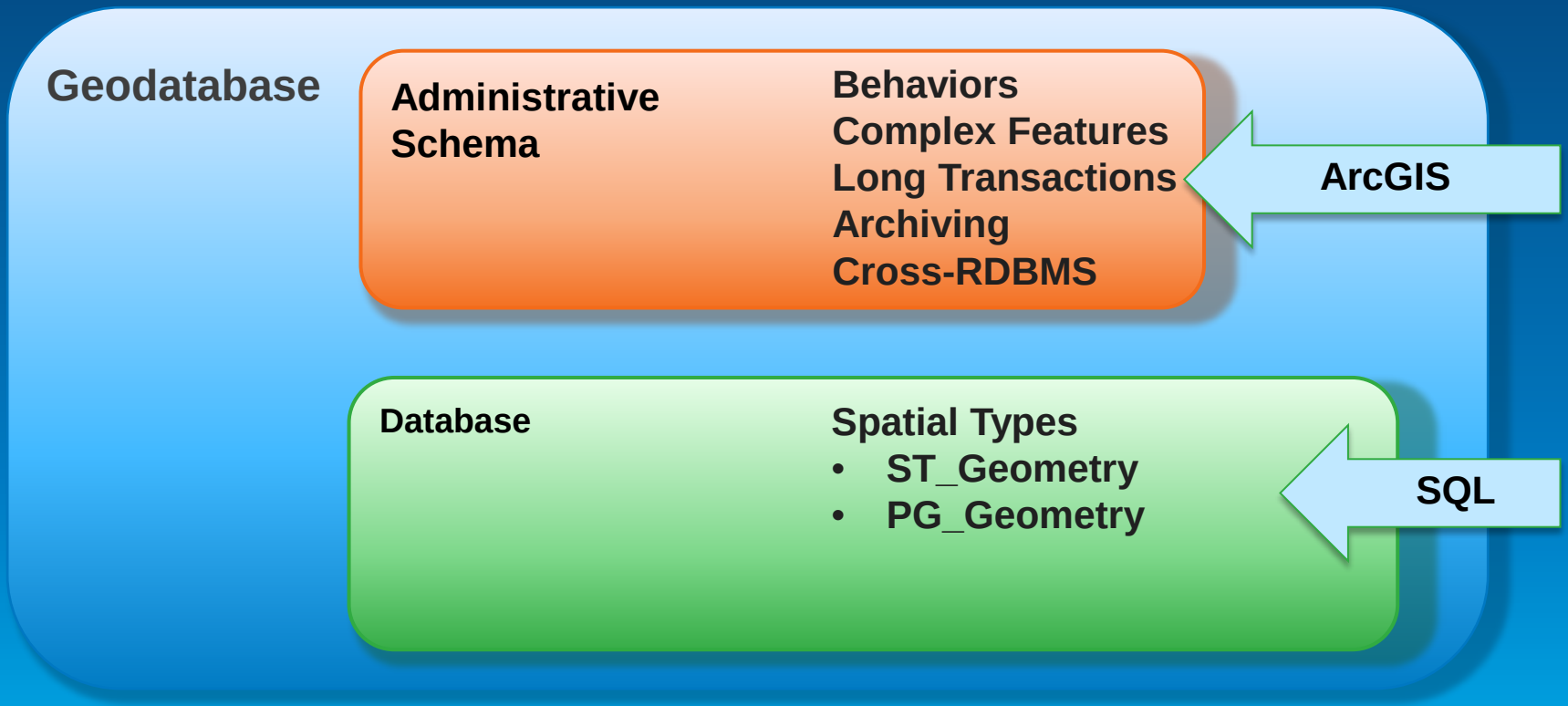


ST_Geometry: Default Geometry Datatype

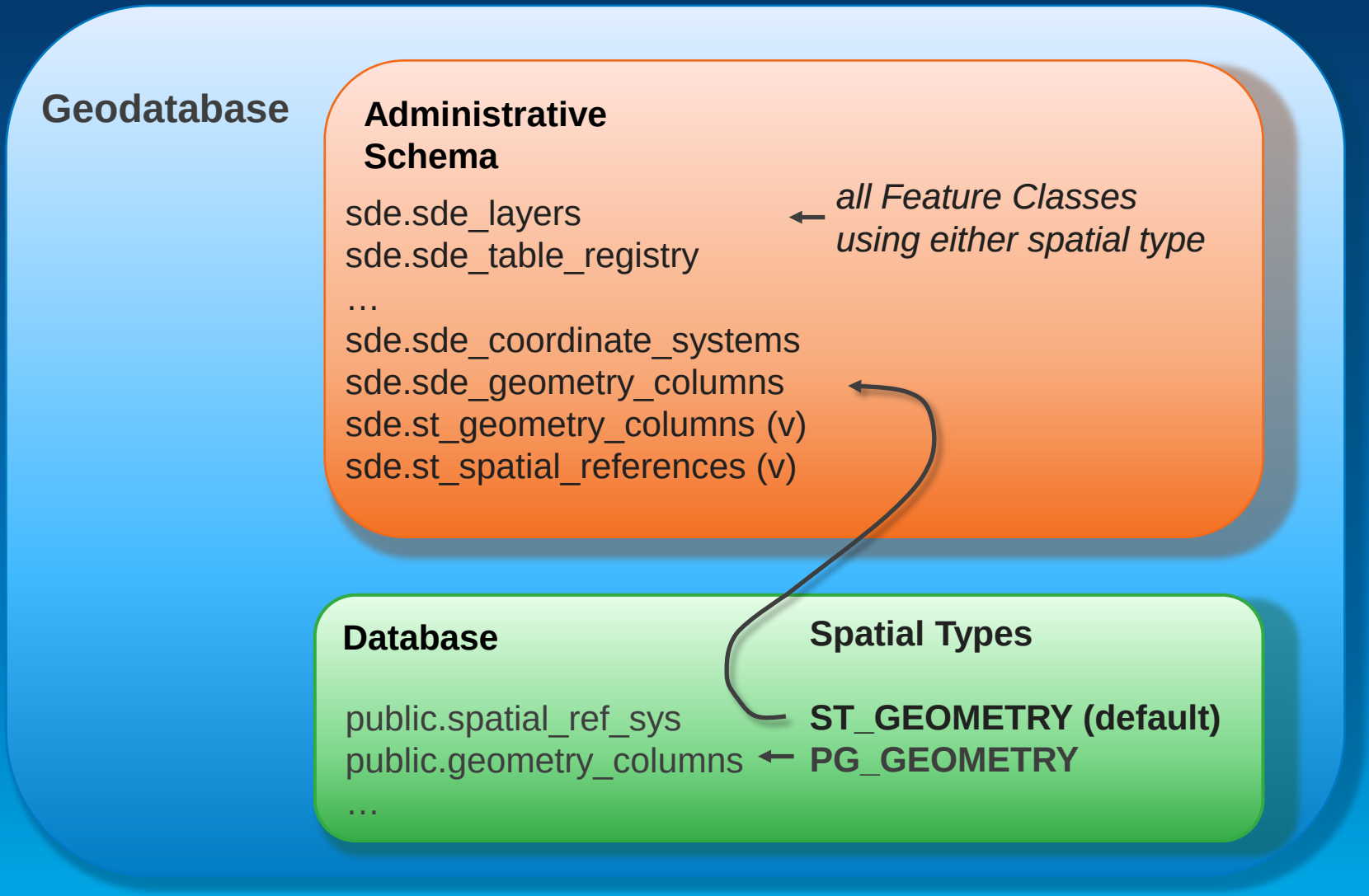
- SQL 3 specification of user-defined data types (UDTs) – ISO and OGC compliant
- Provides structured query language (SQL) access to the geodatabase and database.
- Can be used in PostgreSQL databases that contain a geodatabase and those that do not.

Geodatabase: Editing through SQL

- Geodatabase behavior not supported through SQL



Geodatabase: Spatial Types



PostGIS spatial type

Requirements and Limitations

- Refractions Research http://www.refractions.net/
- PostGIS 1.5.x, 2.0 (10.1 SP1 and 10.2)
  - must use PostGIS database template to create database
  - sde and data owners require permissions
  - must use spatial references in public.spatial_ref_sys table
  - must use PG_GEOMETRY keyword
  - cannot rename tables (public.geometry_columns not modified)
- 64-bit: linux build for 1.5 and 2.0, windows build only 2.0
- ArcGIS behavior vs. PostGIS behavior (e.g. topology)
- Do not plan to support: raster, 3d/4d indexing, etc…
- Support Geometry, not Geography

Agenda

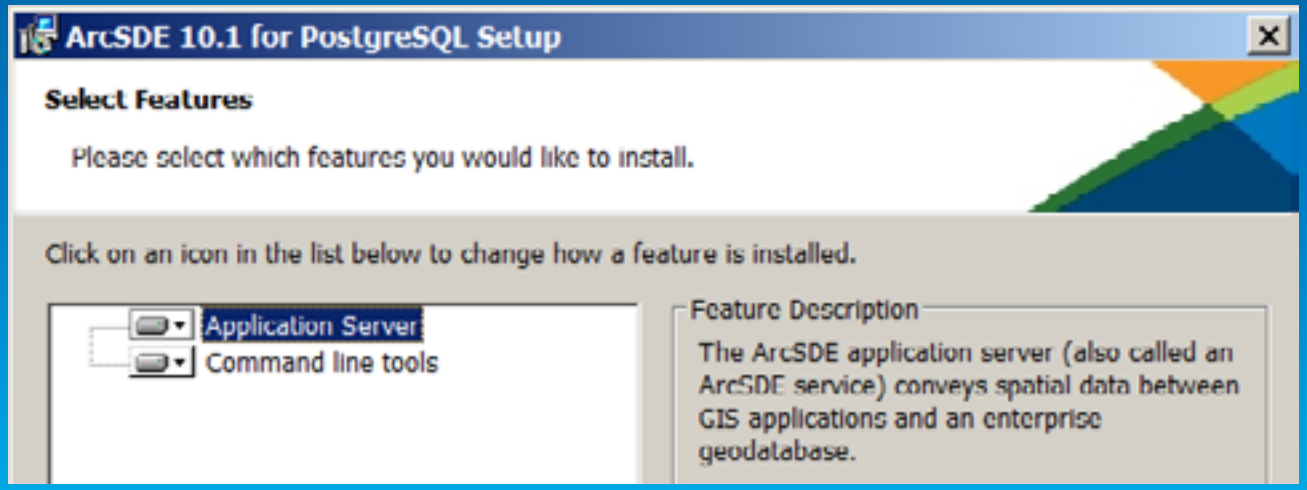
- Requirements and Configuration
- **Administration and Connections**
- Backup and Upgrade
- Performance and Advanced Topics

Agenda

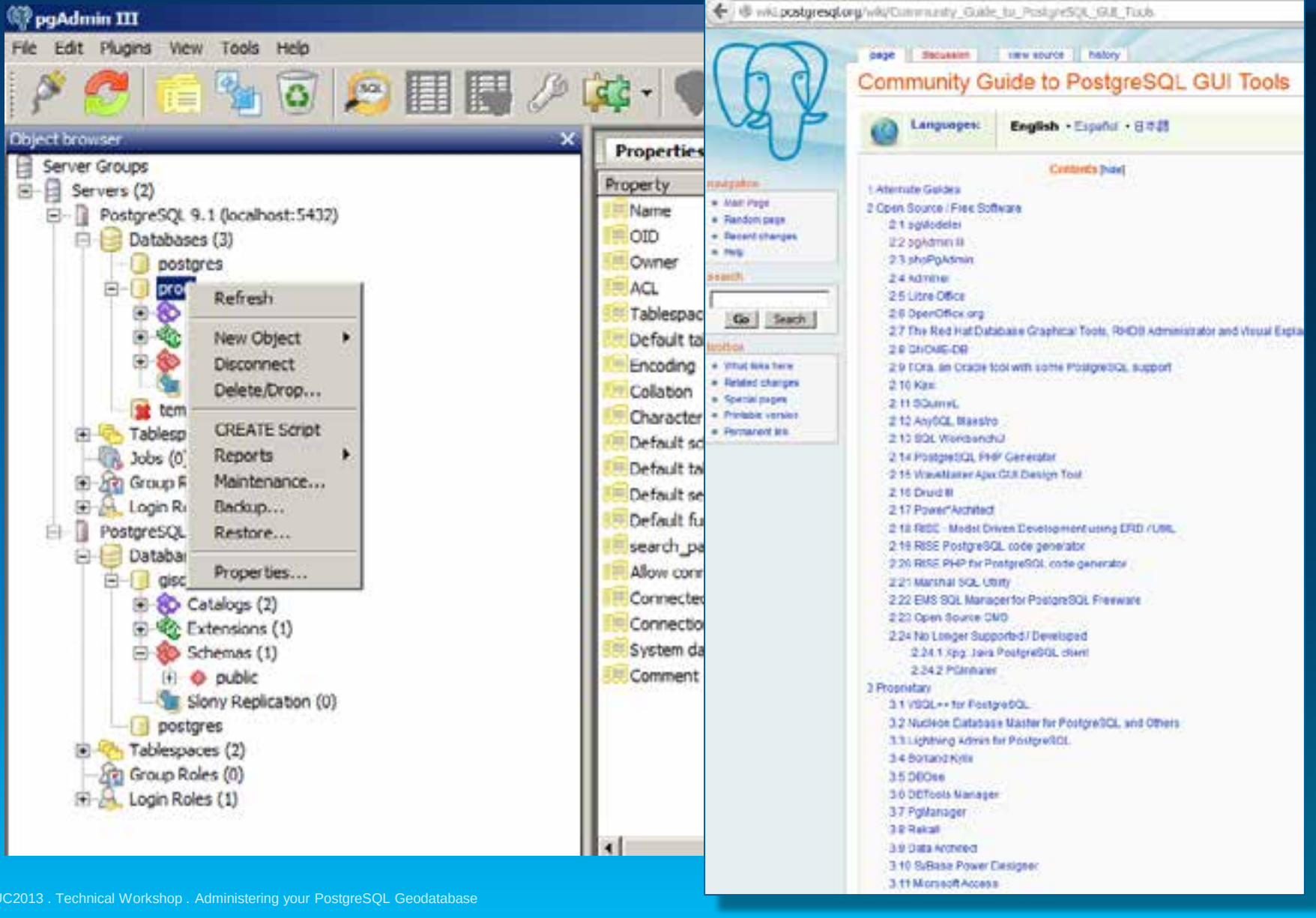
- **Administration and Connections**
- Administration Tools
- Connections
- User Permissions
- Client Compatibility

Administration Tools

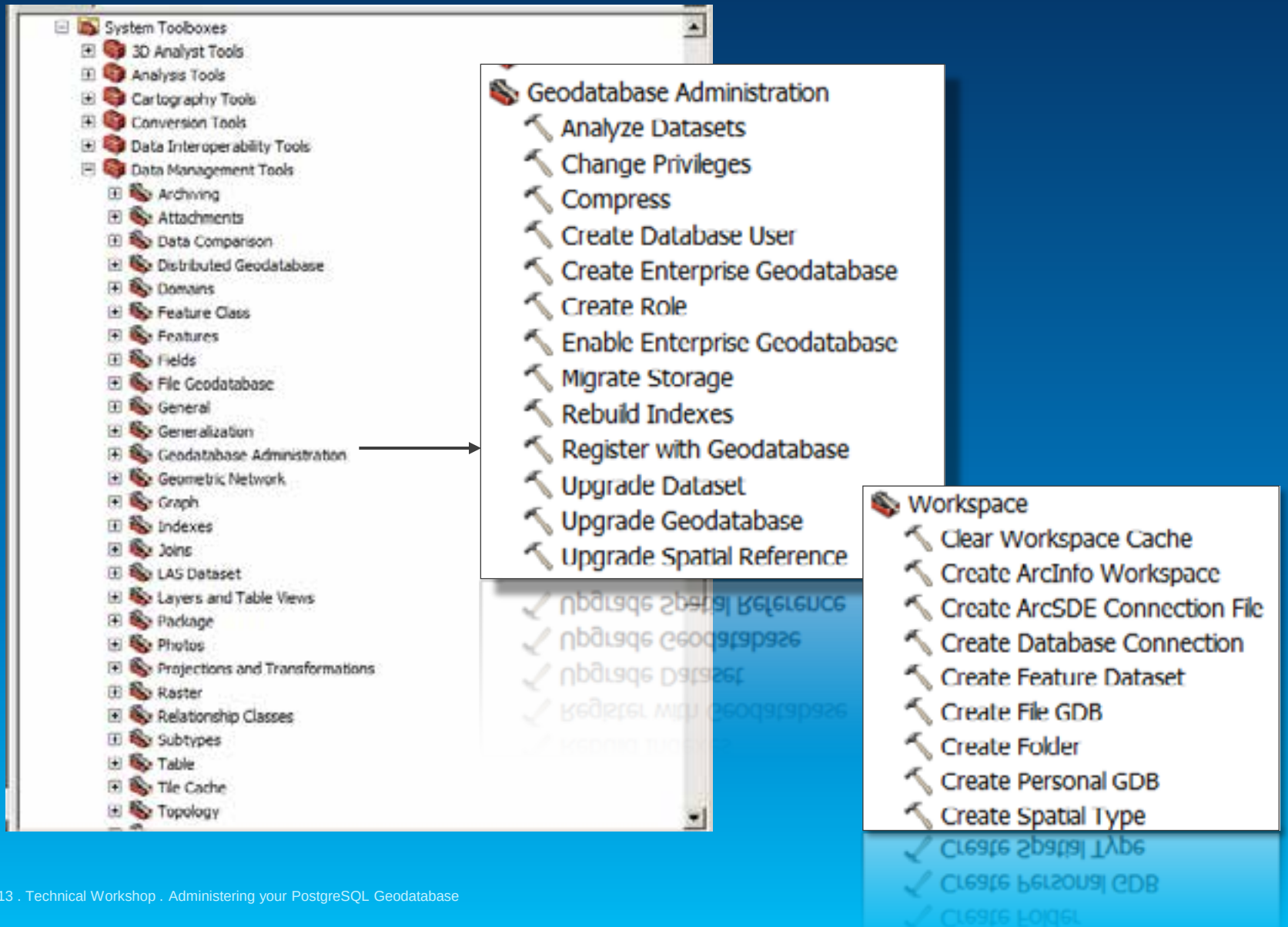
- PostgreSQL and ArcGIS
- ArcSDE Command Line Tools (10.2 last release)
- Some things are still only possible with the cmd
  - DBTUNE maintenance
  - Sde service creation and maintenance
    (if sde application server type connection used)



Administrative Tools: PostgreSQL

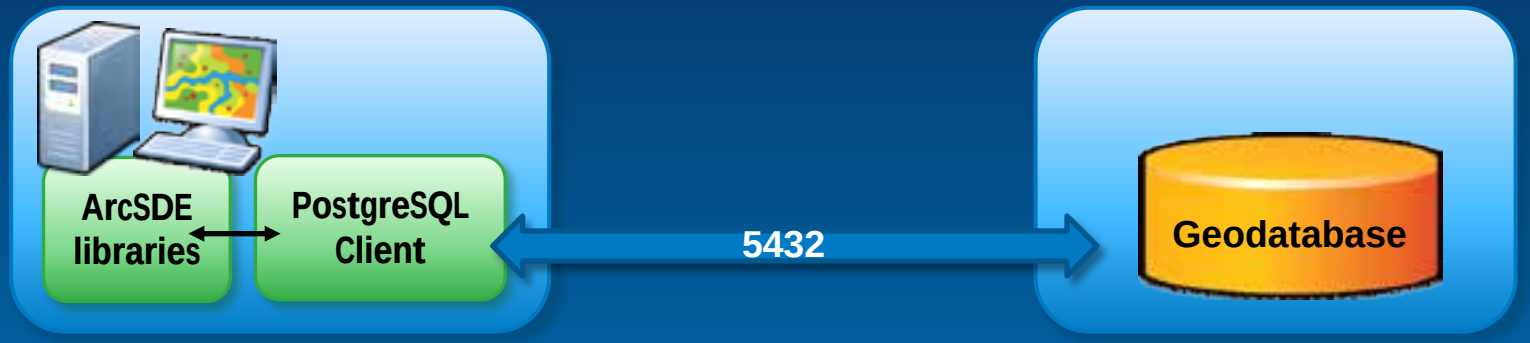


Administrative Tools: ArcGIS

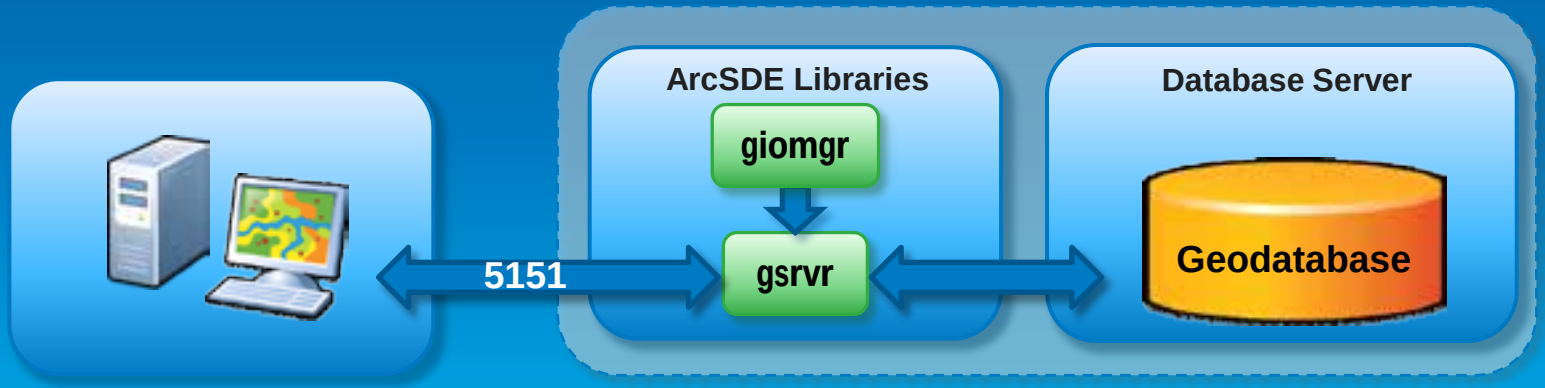


Connection Architectures

"Direct Connect" – recommended and future method post-10.2

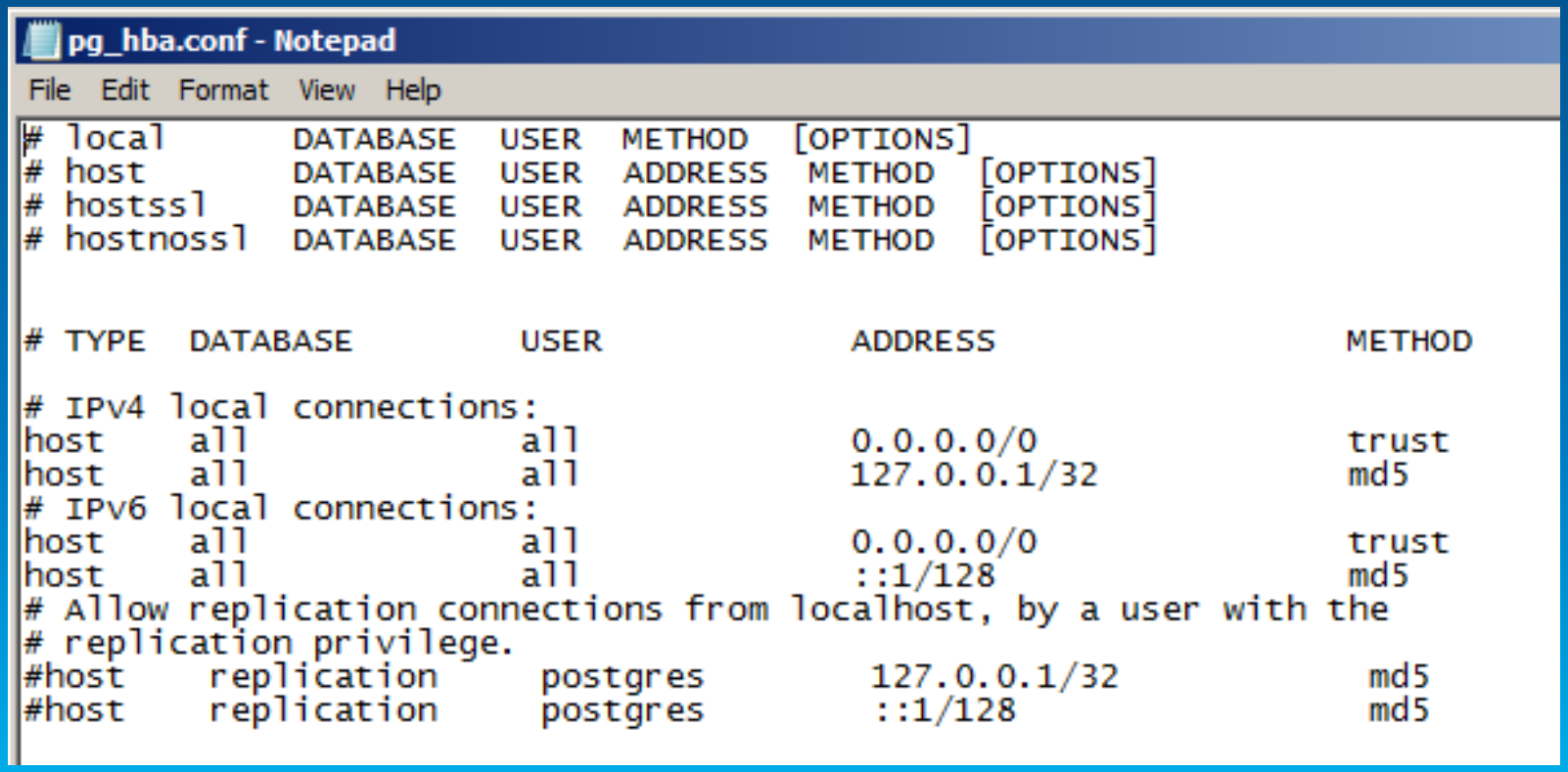


"Application Server" – legacy connection method, 10.2 last release supporting



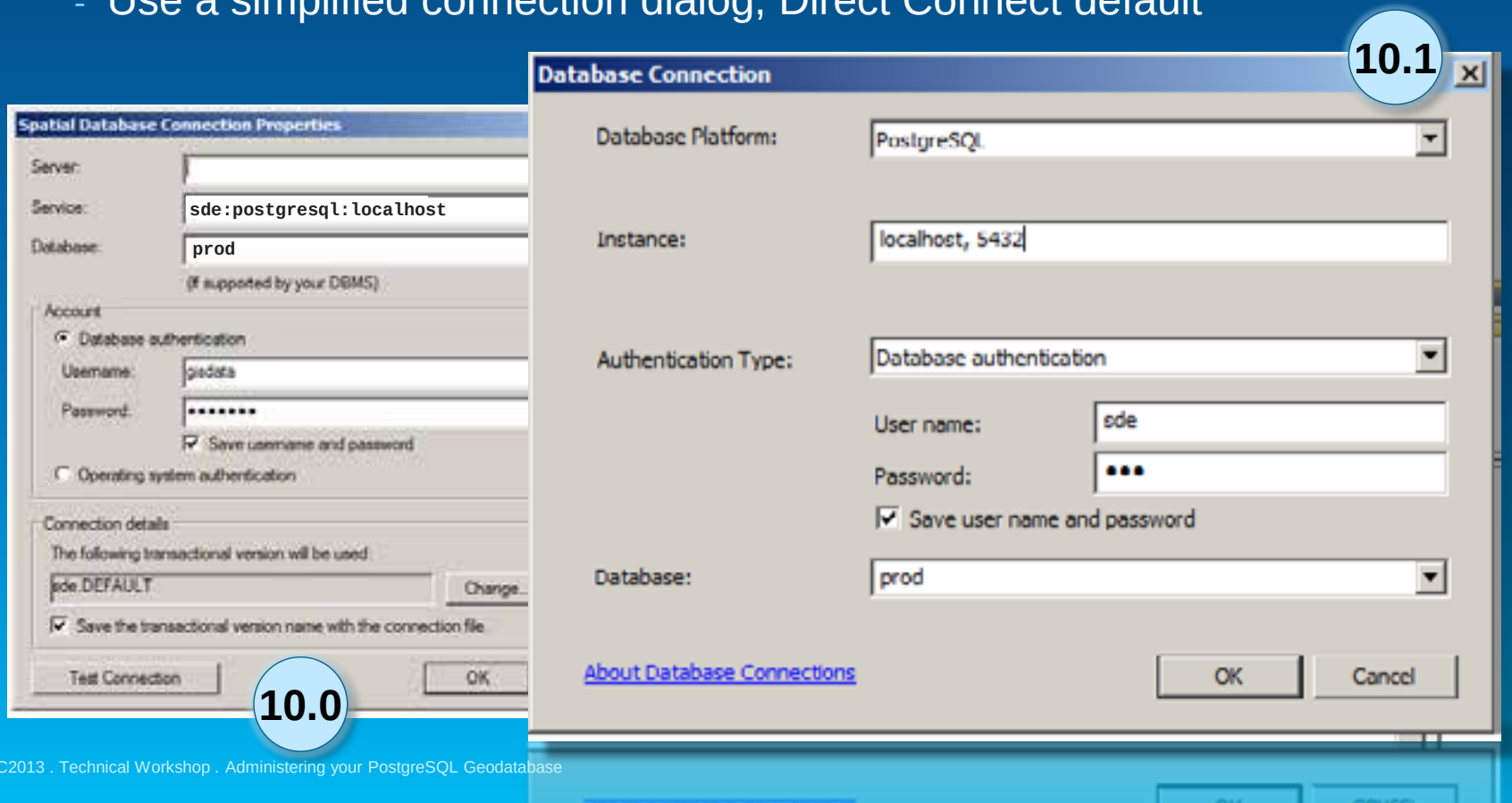
Connecting to PostgreSQL: pg_hba. conf

- PostgreSQL configuration file for connections
  - Depending on your network , entries for both types of addresses may be needed
- IPv4 and IPv6 Addresses



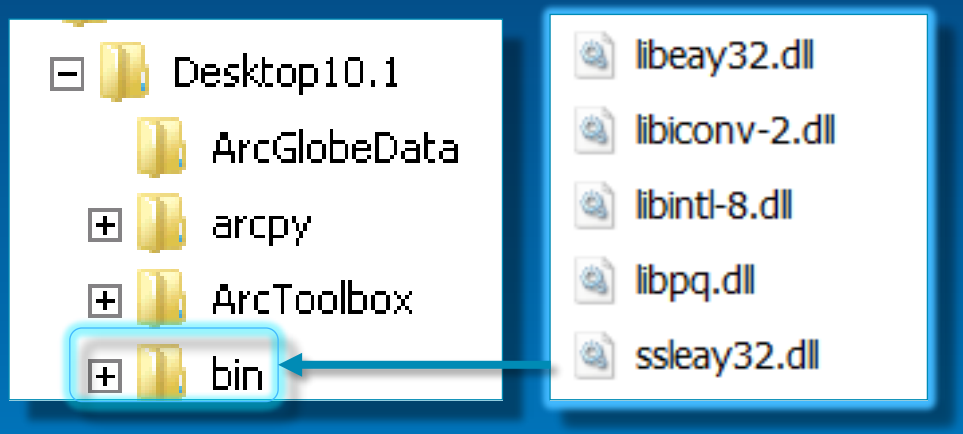
Connections: New Approach at ArcGIS 10.1

- A new approach to connecting to databases:
  - Connect to databases as well as Geodatabases, no SDE administrative schema required.
  - Populate the ArcGIS with database client libraries
  - Use a simplified connection dialog, Direct Connect default



PostgreSQL Client Libraries

- Copy the PostgreSQL client libraries into Desktop\bin or Server\bin. 32-bit for Desktop, 64-bit for Server



- available at Customer Care Portal or at PostgreSQL site.

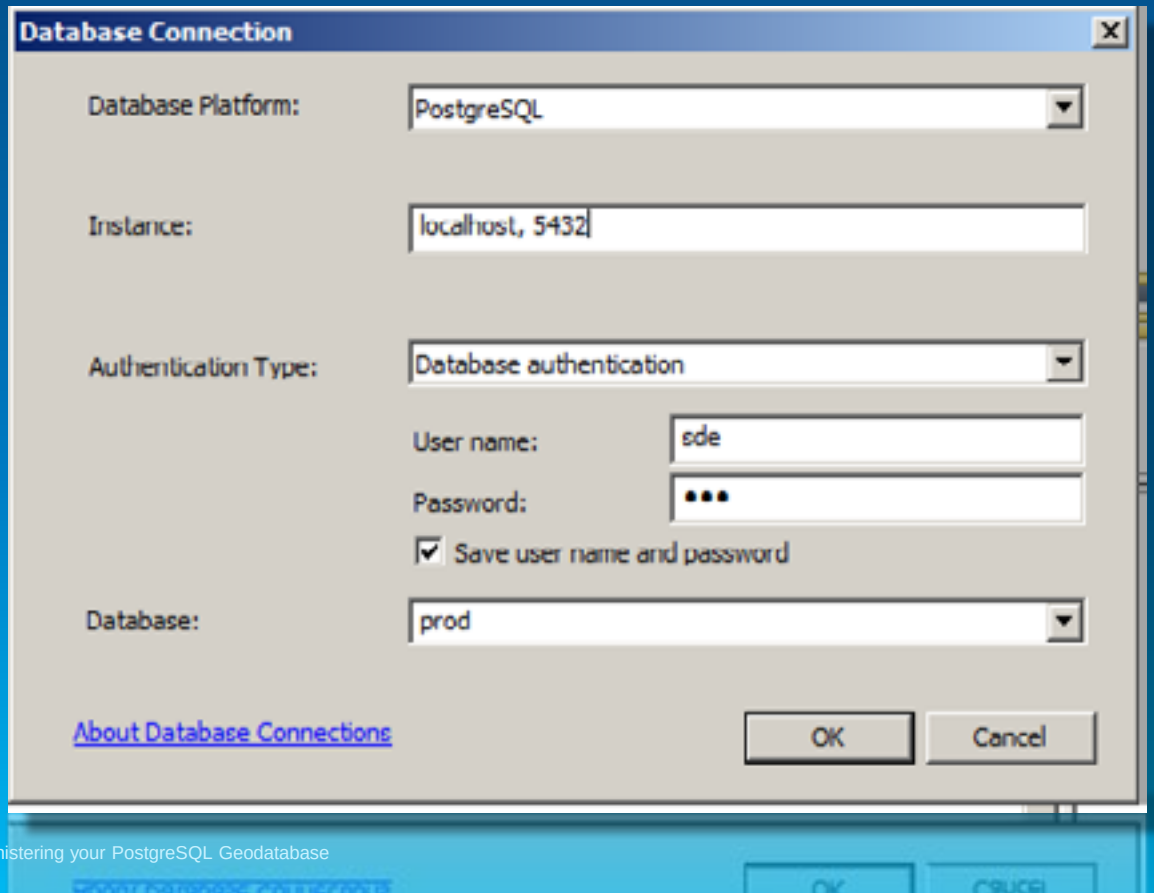
PostgreSQL Client Libraries: Linux Notes

- ArcGIS Server needs to be aware of the client libraries: setup environment variables
- /home/ags/arcgis/server/usr
- Init_user_param.sh
  - Has sections to specify environment variables for each db
  - PostgreSQL Section:

```
# For Direct Connect with PostgreSQL
#
export PGHOME=/opt/PostgreSQL/9.0
export PATH=$PGHOME/bin:$PATH
export LD_LIBRARY_PATH=$PGHOME/lib:$LD_LIBRARY_PATH
```

Connecting to PostgreSQL

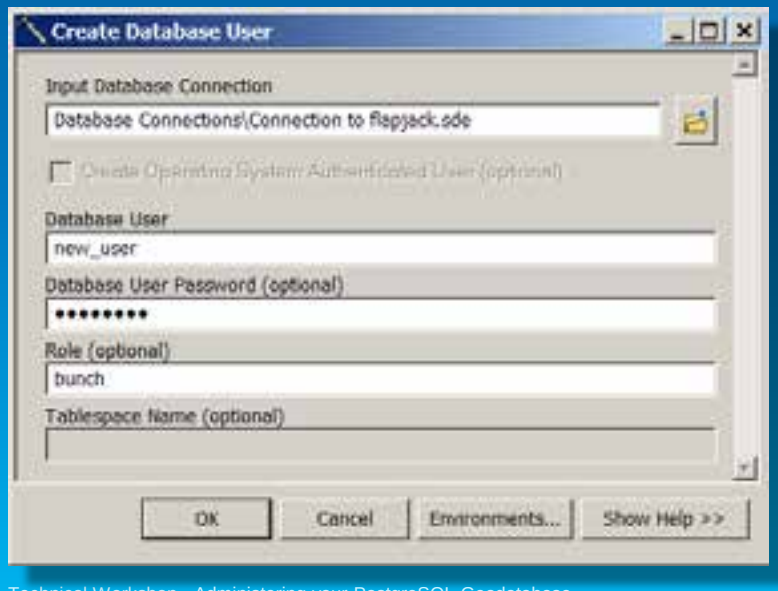
- Must specify an instance (name or IP address of server) & database.
- If instance is listening on a different port than the default (5432), include the port number in the instance. For example: gisprod4, 5435
- The database name is limited to 31 characters.



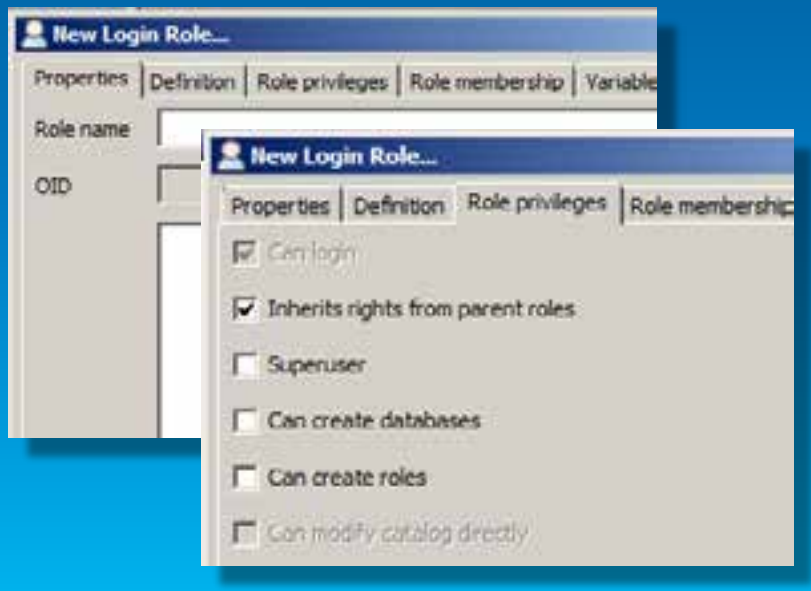
Database Users - Logins

- SDE Administrative user, Data Owners, Editors, Viewers
- login and schema must be same name for logins that will own objects in the geodatabase.
- You cannot create a schema for a group role.
- Can rerun tool to create a schema in a second database

ArcGIS Tools



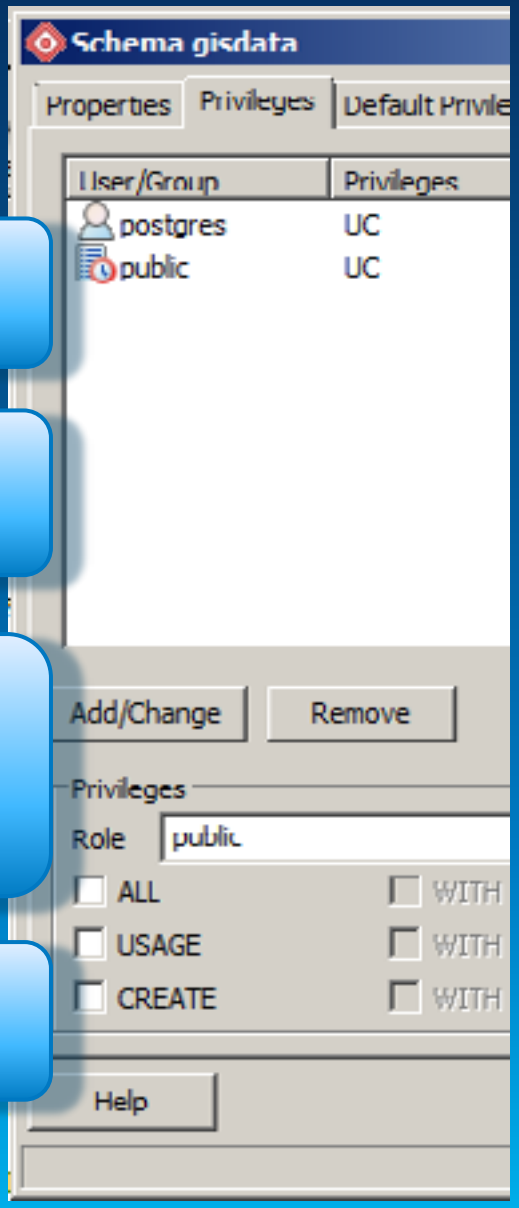
PostgreSQL Tools



PostgreSQL User Permissions

Login Roles (Users) and Group Roles (Groups)

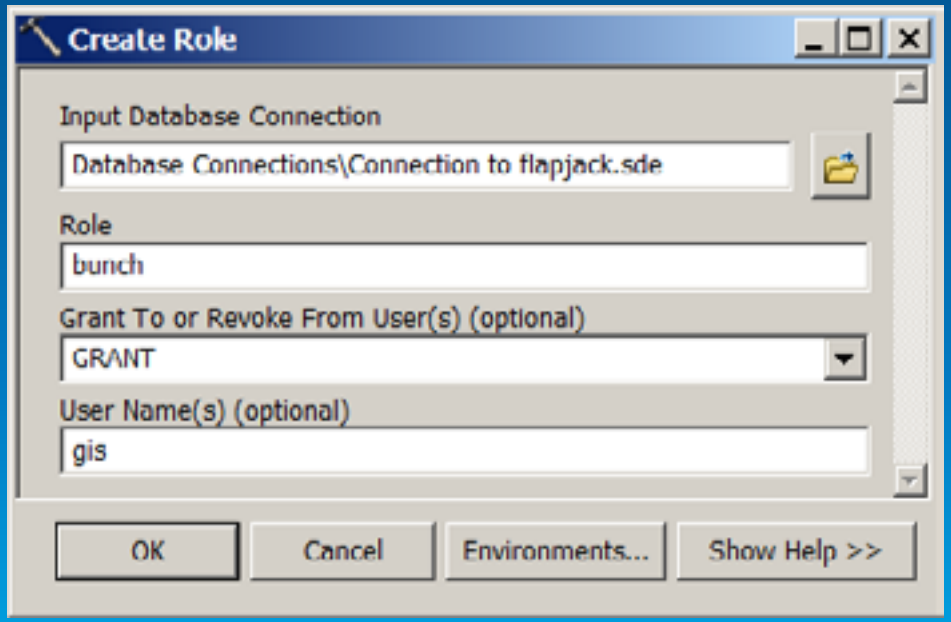
Data Viewer	Usage on SDE Schema Usage on data schemas to be viewed
Data Editor	Usage on SDE Schema Usage on data schemas to be edited
Data Creator	Usage on SDE Schema Authorization on user's own schema Usage on any other data schemas where access is required
GDB Admin (SDE)	Authorization on SDE schema Usage on all other user schemas



Create Role

- Creates a PostgreSQL Group role
- The GP tool does the same as the sql:

```
CREATE ROLE gis LOGIN ENCRYPTED PASSWORD ‘gis‘ INHERIT;
CREATE ROLE bunch VALID UNTIL ‘infinity’;
GRANT bunch to gis;
```



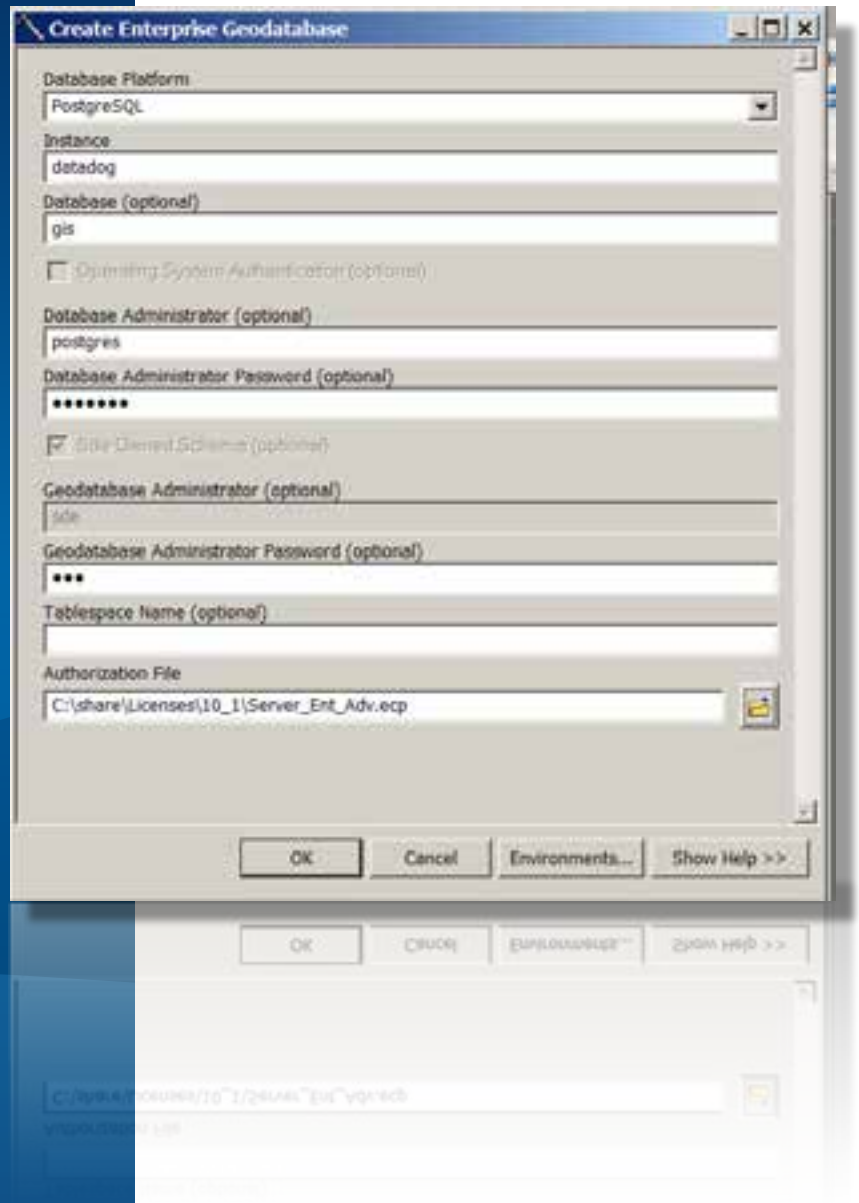
Client – Geodatabase Compatibility

- 10.x forwards and backwards compatibility
- Note: 10.2 will be last major release for:
  - ArcSDE command line tools
  - ArcSDE Application Server
  - ArcSDE C and Java SDK

Client release	Enterprise, workgroup, or desktop geodatabase using a direct connection	Enterprise geodatabase using an ArcSDE service connection
9.3	9.3 (if client has pre-9.3 geodatabase direct-connect files* installed, can connect to 9, 9.1, and 9.2)	9, 9.1, 9.2, 9.3
10	9.2, 9.3, 10, 10.1, 10.2	9.2, 9.3, 10, 10.1, 10.2
10.1	9.3, 10, 10.1, 10.2	9.3, 10, 10.1, 10.2
10.2	9.3, 10, 10.1, 10.2	9.3, 10, 10.1, 10.2

Demo: Setting up an Enterprise Geodatabase on PostgreSQL

James Gough



Agenda

- Requirements and Configuration
- Administration and Connections
- **Backup and Upgrade**
- Performance and Advanced Topics

Agenda

- **Backup and Upgrade**
- **Recovery Models**
- **Backup and Restore**
- **Upgrades**

PostgreSQL Recovery Models

- What are needs
  - how often does data change?
  - how long can application(s) be down?
  - how fast does recovery need to be?
- Weekly or nightly backups
  - recovery to most recent backup – most common
- Write-Ahead Logging (WAL)
  - point-in-time recovery
  - must test thoroughly to understand resource requirements (e.g. disk i/o)
- Other options
  - Standby or Failover configurations
  - High Availability configurations

Moving or Copying a PostgreSQL Database

ddd

- Purposes
  - cloning to staging or development environments
  - migrating to new VM environment or hardware
- Methods
  - via ArcGIS to new database – copy/paste, export/import, etc…
  - pg_dump/pg_restore commands
    - Restore PUBLIC schema first, then SDE, then data owner schemas
    - run as superuser
    - run ANALYZE after to update statistics
    - re-create Tablespaces
    - Text version of a table cannot be larger than 8GB if output to TAR
    - use –o option if object identifiers (OIDs) in user-defined data

Backup

pg_dump, pg_dumpall and other methods

- Typical Backup Methods
  - database – pg_dump (typical method)
  - instance – pg_dumpall (backs up logins and roles)

```
pg_dump -h localhost -p 5432 -U postgres  -F c -v -f
E:\backups\prod_050612.bak prod
```

- Some Other Backup Methods
  - file based (cold) backup
  - VM backup

Restore

pg_restore

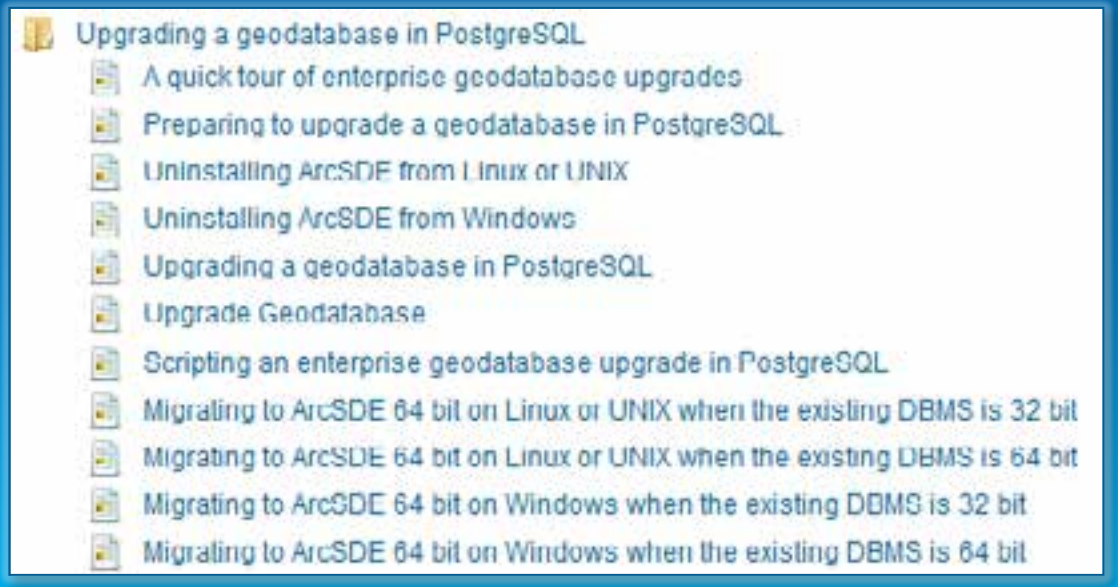
- Restoring – pg_restore
  - restore schemas in order – public, sde, data owners
  - if restoring pg_dumpall, restore public schema from a pg_dump first to avoid errors

```
pg_restore -n public  -p 5432 -U postgres -d
db_name –c -v E:\backups\db_050712.bak
```

  - rebuild spatial indexes and gather statistics once restored
- Some Notes
  - many dependent objects between schemas, may need to drop cascade in psql (sde and data owners) and recreate
  - may get errors if trying to drop via pgadminIII, use psql with cascade
  - drop in reverse order (data owners, then sde)

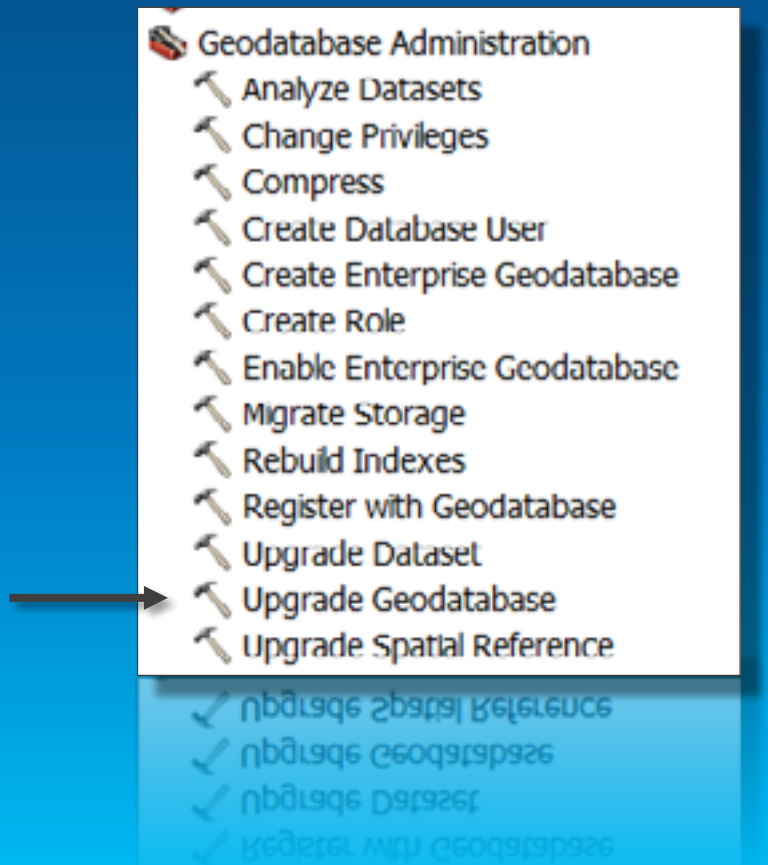
Upgrade Process

- Upgrading PostgreSQL? or only Geodatabase?
- Make sure to upgrade client libraries as needed
- Save configuration files for comparison purposes, don't copy old files back in, inspect and modify new ones.
  - pg_hba.conf
  - postgresql.conf
  - dbtune.sde



Upgrading the Geodatabase

- Test in a staging or test environment first
  (good opportunity to test your backup and recovery process)
- Place new st_geometry lib in postgres lib location
- Upgrade
  - With python script
  - With gp tool
  - Right click in ArcGIS



Agenda

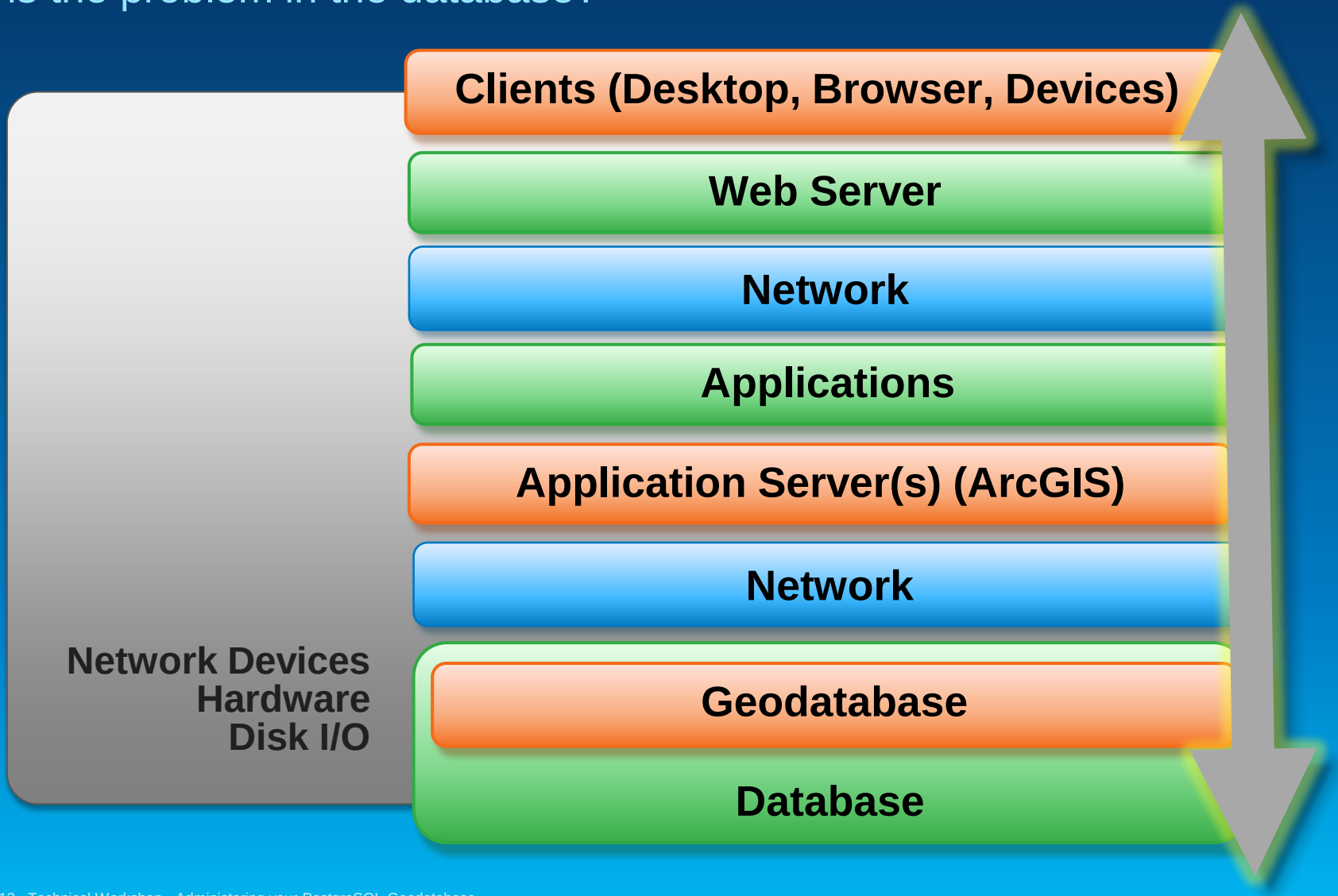
- Requirements and Configuration
- Administration and Connections
- Backup and Upgrade
- **Performance and Advanced Topics**

Agenda

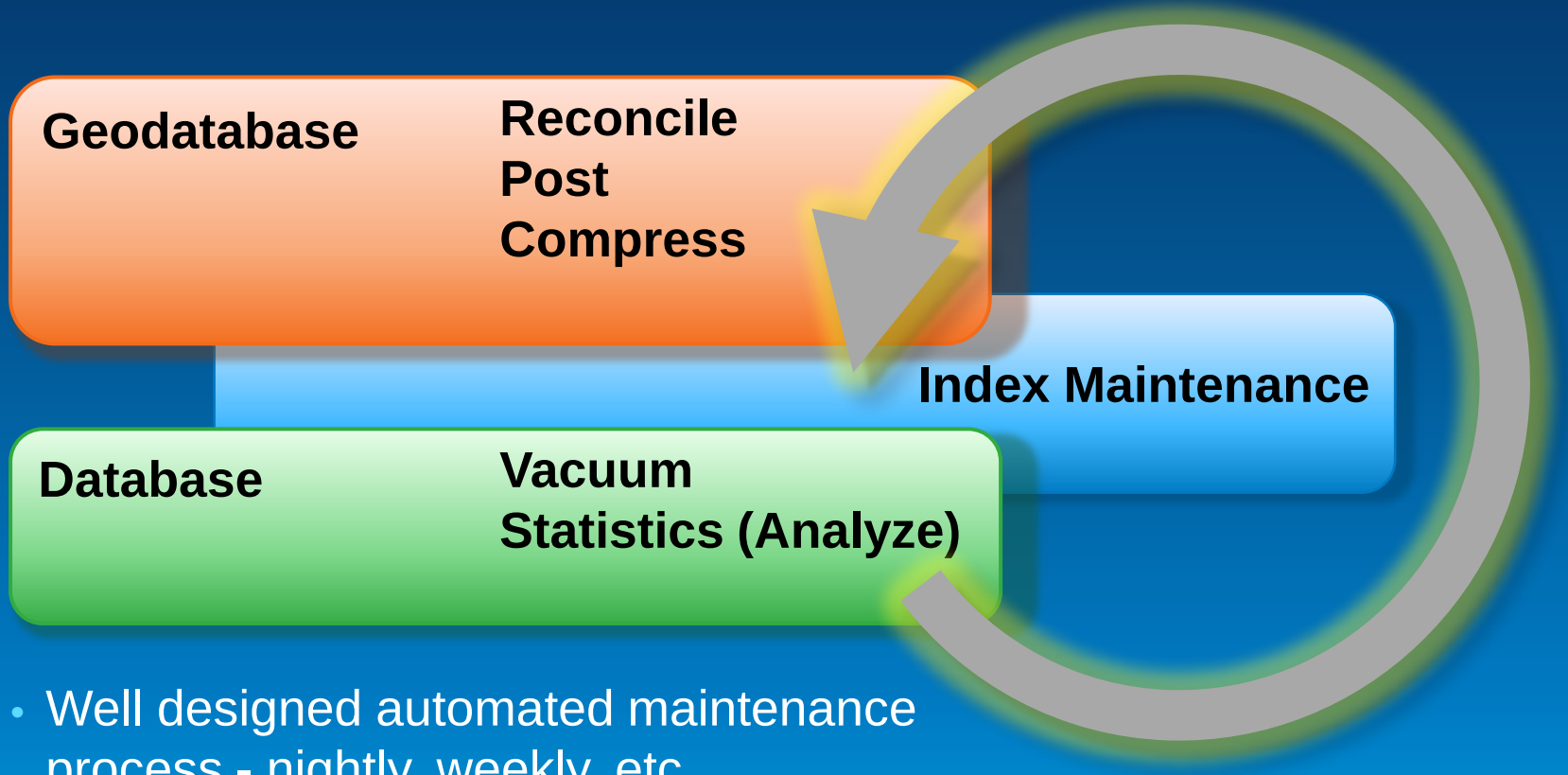
- **Performance and Advanced Topics**
- Performance
- Monitoring and Troubleshooting
- Advanced Topics
- Conclusion

Performance: Understand the Stack and Isolate

is the problem in the database?



Geodatabase: Proper Maintenance = Performance



- Well designed automated maintenance process - nightly, weekly, etc..
- Well designed and maintained Version and Replica architecture
- include Backup, ETL's, Reporting, etc…

Monitoring: Why monitor?

- Establish performance benchmarks to measure impacts:
  - upgrades and patches
  - new applications or workflows
  - new server resources or deployment patterns
- Assist in troubleshooting
  - assist in isolating a problem when one takes place

What to monitor?

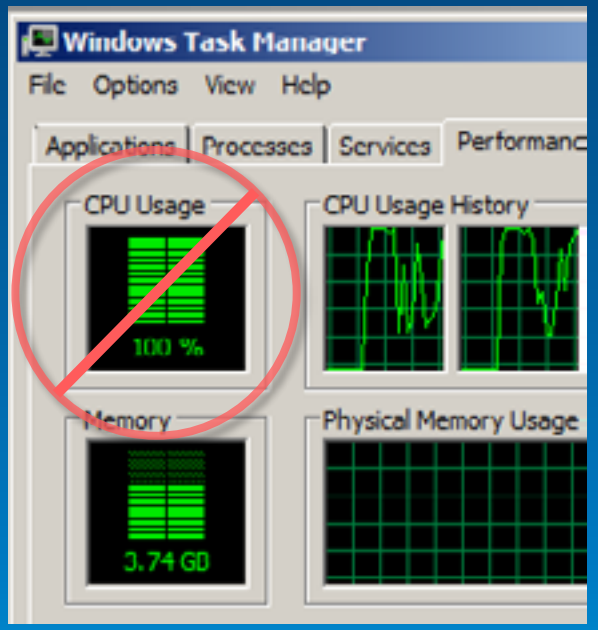
- Server Resources (cpu utilization, memory, storage i/o)
  - Windows and Linux tools (top, vmstat, iostat, etc..)
- Client Performance
  - various tools and logs (e.g. ArcGIS Server logs)
- PostgreSQL Performance
  - Performance views and Postgresql logs (located in …\Data\pg_log)

```
2013-05-21 13:00:43 PDT DETAIL:  parameters: $1 = '13580'
2013-05-21 13:00:43 PDT LOG:  duration: 0.000 ms  parse
sde_1369166443_0_793: SELECT lineage_id FROM
prod.sde.sde_state_lineages WHERE lineage_name = $1 AND lineage_id <=
$2 ORDER BY lineage_id
2013-05-21 13:00:43 PDT LOG:  duration: 0.000 ms  bind
sde_1369166443_0_793: SELECT lineage_id FROM
prod.sde.sde_state_lineages WHERE lineage_name = $1 AND lineage_id <=
$2 ORDER BY lineage_id
2013-05-21 13:00:43 PDT DETAIL:  parameters: $1 = '13579', $2 =
'13580'
```

Monitor Resources

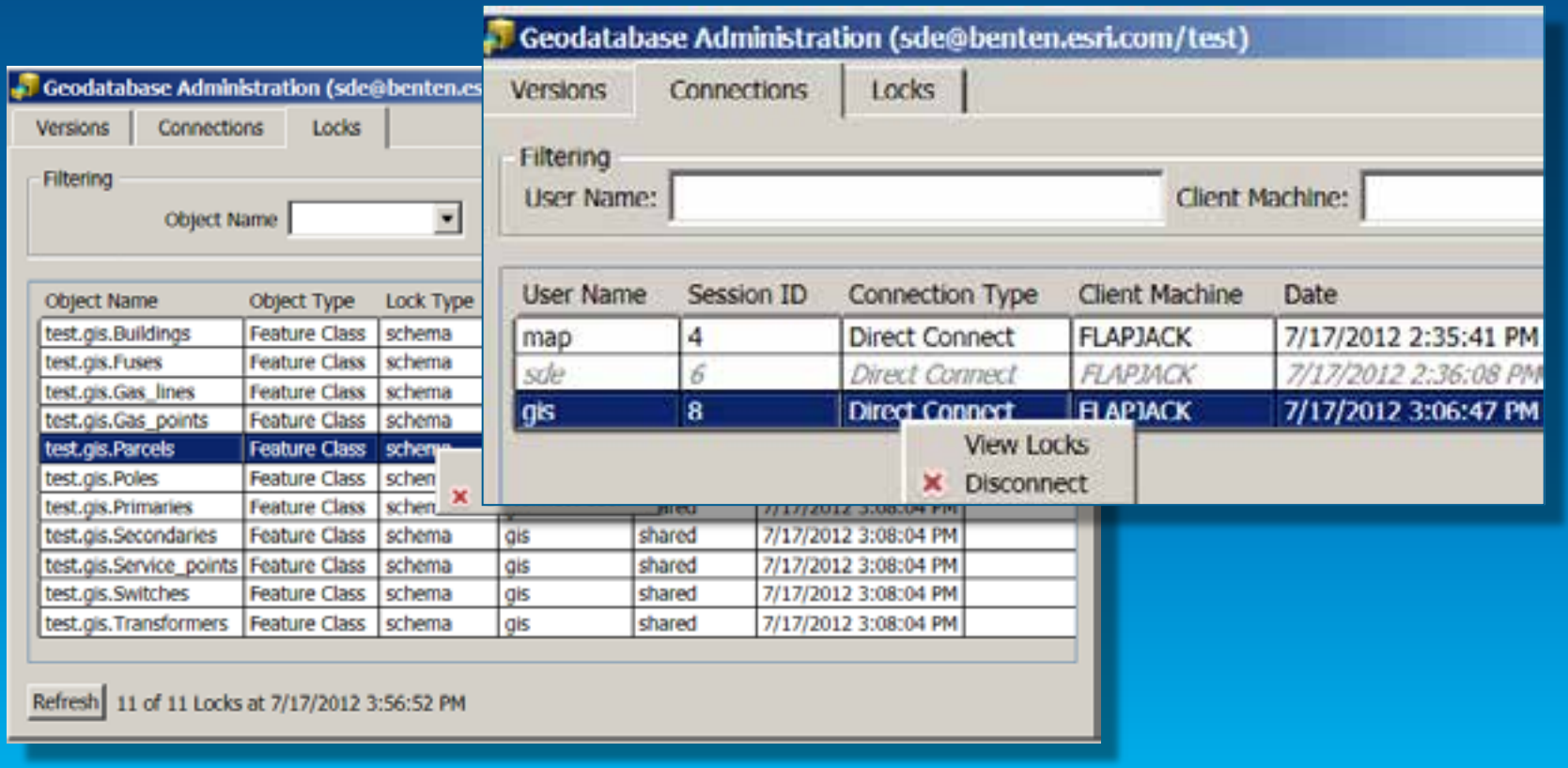
Physical and Virtual Environments

- Be careful of any of the following thresholds:
  - Processor utilization > 70%
  - Memory utilization > 80% of physical
  - Storage utilization > 80% of storage capacity
  - Average Disk Seconds / Read > 10ms
  - Average Disk Seconds / Write > 10ms
- If Cloud deployment
  - different locations may have different behavior (resources/equipment)
-



Monitor Connections and Locks in ArcGIS

- Monitor: Connections and Locks
  - “kill” connections - Superuser
  - Direct Connections and Application Server Conections
  - check lock type



Monitor Versions in ArcGIS

- Monitor: Versions
  - See all versions in the Geodatabase
    - Properties
    - Relationships
  - Manage Versions

Performance Considerations

- Excessive normalization
  - Too many indexes
  - No optimizer hints, index use can not be forced
  - Need not worry about the Spatial Index
    - GIST index used, self correcting
- Can change postgresql.conf initialization parameters
- Issue with long running ArcGIS edit sessions
  - The larger the number of states
  - The larger the bloat in indexing belonging to the Feature Class

PostgreSQL Geodatabase Performance

Proper Maintenance

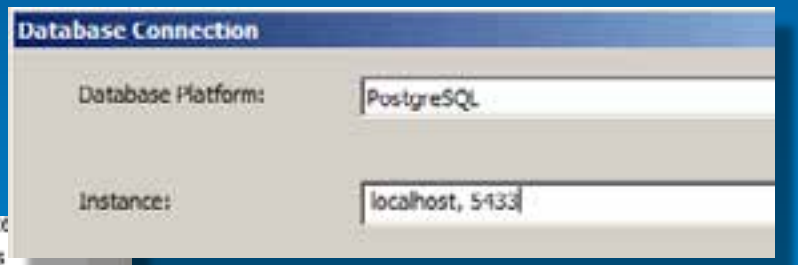
- Vacuum
  - removes dead tuples (rows)
  - Autovacuum – on by default, can do an analyze
- Statistics (Analyze)
  - Statistics – table size, cardinality of joins, distribution of indexes, etc…
  - pg_stat_statements (create extension pg_stat_statements)
  - module provides a means for tracking execution statistics of all SQL statements executed by server.
  - shared_preload_libraries = '"E:\\PostgreSQL\\9.2\\lib\\pg_stat_statements.dll"'
- Indexes (Rebuildx)

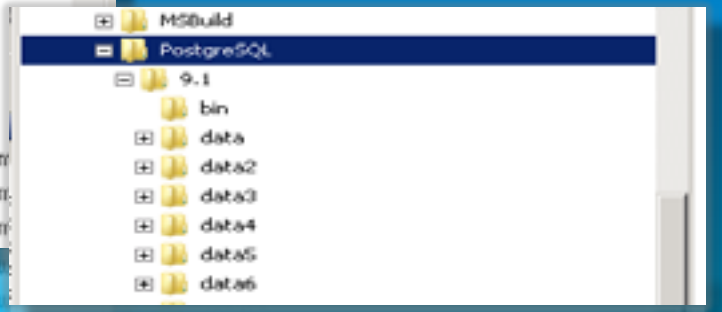
Multiple instances on same server

- Create Instance
  E:\PostgreSQL\9.2\installer\server>initcluster.vbs postgres postgres gisdata.101 “E:\PostgreSQL\9.2" “E:\PostgreSQL\9.2\pgdata3" 5434 DEFAULT
  Usage: initcluster.vbs <OSUsername> <SuperUsername> <Password> <Install dir> <Data dir> <Port> <Locale>

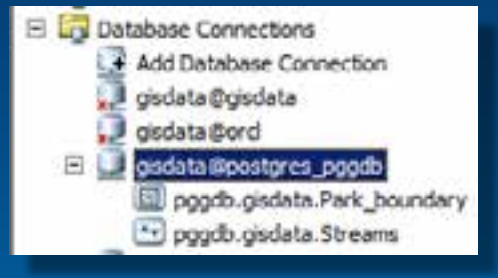
- Create Windows Service
  E:\PostgreSQL\9.2\installer\server>startupcfg.vbs 9.2 postgres gisdata.101 “E:\PostgreSQL\9.2" “E:\PostgreSQL\9.2\pgdata3" postgresql-9.2-3
  Usage: startupcfg.vbs <Major.Minor version> <Username> <Password> <Install dir> <Data dir> <ServerName>

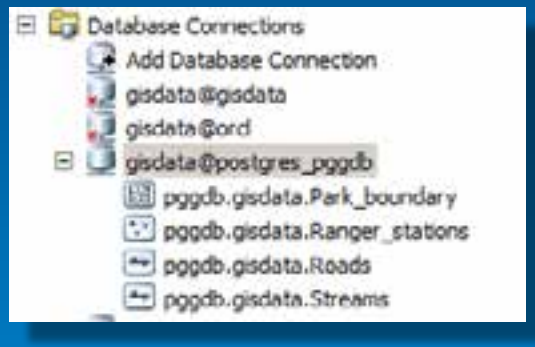
- Connect to one instance per ArcMap Session
- Environment Variables available
  - PGDATABASE
  - PGHOST
  - PGPORT
  - PGUSER



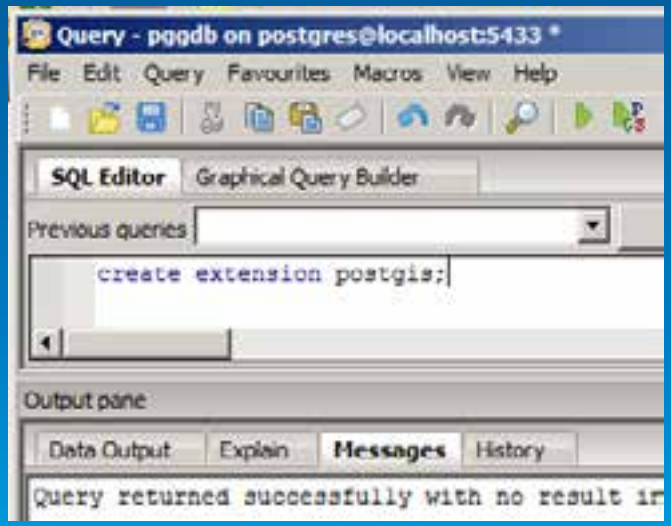


pg_restore, multiple instances and PostGIS



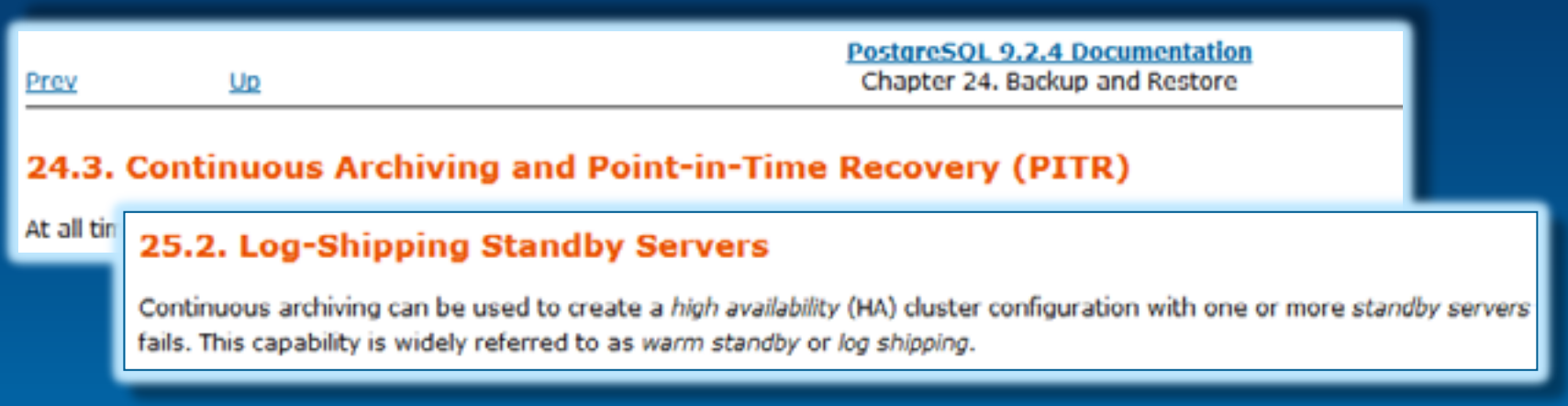


pg_restore: creating TABLE roads
pg_restore: [archiver (db)] Error from TOC entry 251; 1259 18626 TABLE roads gisdata
pg_restore: [archiver (db)] could not execute query:
ERROR: type "public.geometry" does not exist
LINE 18: shape public.geometry,
^



- Remember to enable PostGIS for each instance
  - install and enable
  - **`create extension postgis;`**
  - **`create extension postgis_topology;`**

Point-in-time-recovery (PITR) and Failover



- PostgreSQL configurations use WAL (write-ahead logging)
- PostgreSQL does not provide failure detection itself, add-ons or OS configurations can.
  - OpenSCG's pgHA (PITR and pgbouncer), also PITR and Slony replication
  - pgpool

High-Availability (HA) and Replication Options

from PostgreSQL 9.2 documentation

Table 25-1. High Availability, Load Balancing, and Replication Feature Matrix

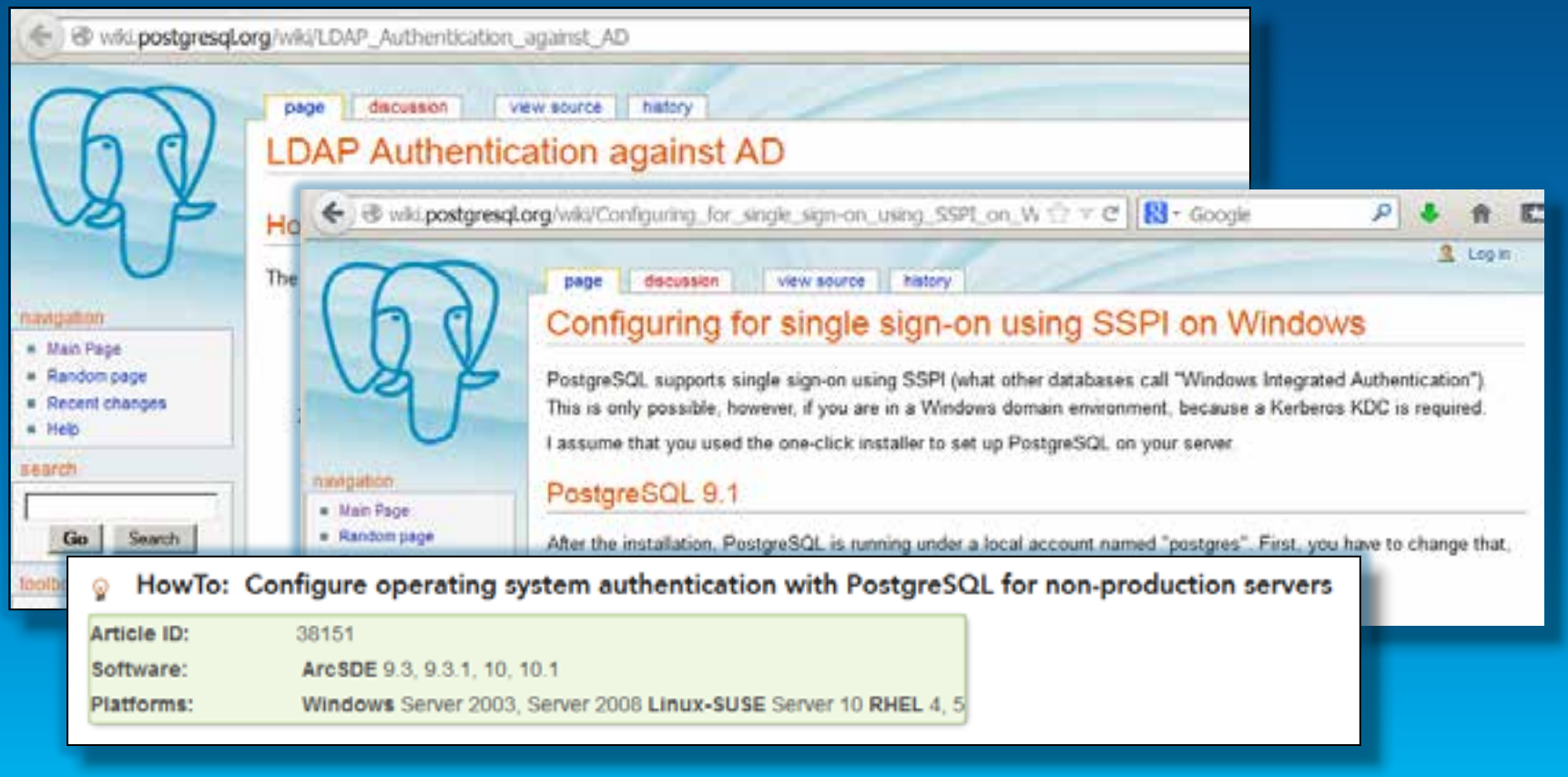
Feature	Shared Disk Failover	File System Replication	Transaction Log Shipping	Trigger-Based Master-Standby Replication	Statement-Based Replication Middleware	Asynchronous Multimaster Replication	Synchronous Multimaster Replication
Most Common Implementation	NAS	DRBD	Streaming Repl.	Slony	pgpool-II	Bucardo	
Communication Method	shared disk	disk blocks	WAL	table rows	SQL	table rows	table rows and row locks
No special hardware required		•	•	•	•	•	•
Allows multiple master servers					•	•	•
No master server overhead	•		•		•		
No waiting for multiple servers	•		with sync off	•		•	
Master failure will never lose data	•	•	with sync on		•		•
Standby accept read-only queries			with hot	•	•	•	•
Per-table granularity				•		•	•
No conflict resolution	•	•	•	•			•

Testing required. Must be entire Geodatabase and Standby Read-Only

Windows AD, LDAP and Single Sign-On

- Editors and Viewers

http://wiki.postgresql.org/wiki/LDAP_Authentication_against_AD
http://wiki.postgresql.org/wiki/Configuring_for_single_sign-on_using_SSPI_on_Windows
http://support.esri.com/en/knowledgebase/techarticles/detail/38151



Example LDAP pg_hba.conf

SDE and Data Owners login and schema name must match

PostgreSQL Authenticated Users
SDE Administrative User
Data Owner

```
# IPv4 local connections:
host    all     sde      0.0.0.0/0       md5
host    all     gisdata  0.0.0.0/0       md5
host    all     postgres 0.0.0.0/0       md5
host    all     all      127.0.0.1/32    md5
host    all     all       0.0.0.0/0       ldap
ldapserver=vmtester.bos.esri.com  ldapprefix="TESTING\"
# IPv6 local connections:
host    all     all        ::1/128        md5
host    all                all            ::/0
md5
```

Editors and Viewers via LDAP

Datatype Mapping

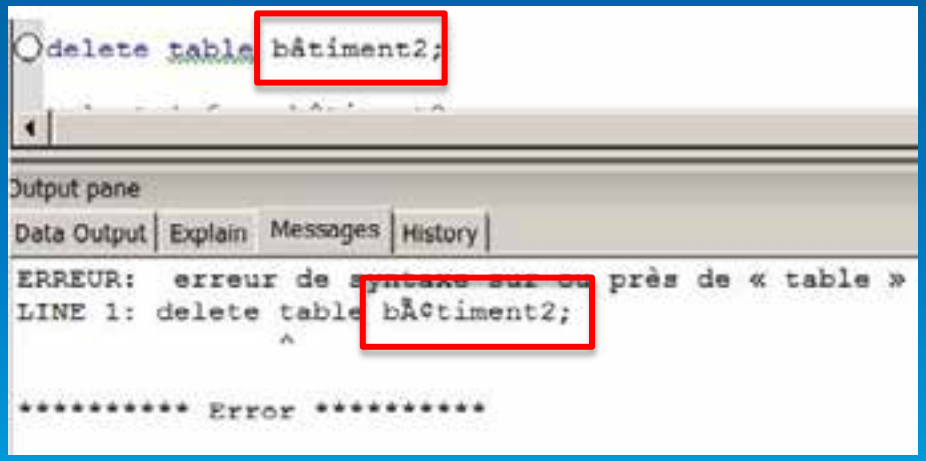
- PostgreSQL supports almost 100 datatypes
- ArcGIS has 8
- Some PostgreSQL datatypes are mapped to one ArcGIS datatype
- Some datatypes are not supported
  - Error: “invalid column datatype”
  - Documented in online help

PostgreSQL

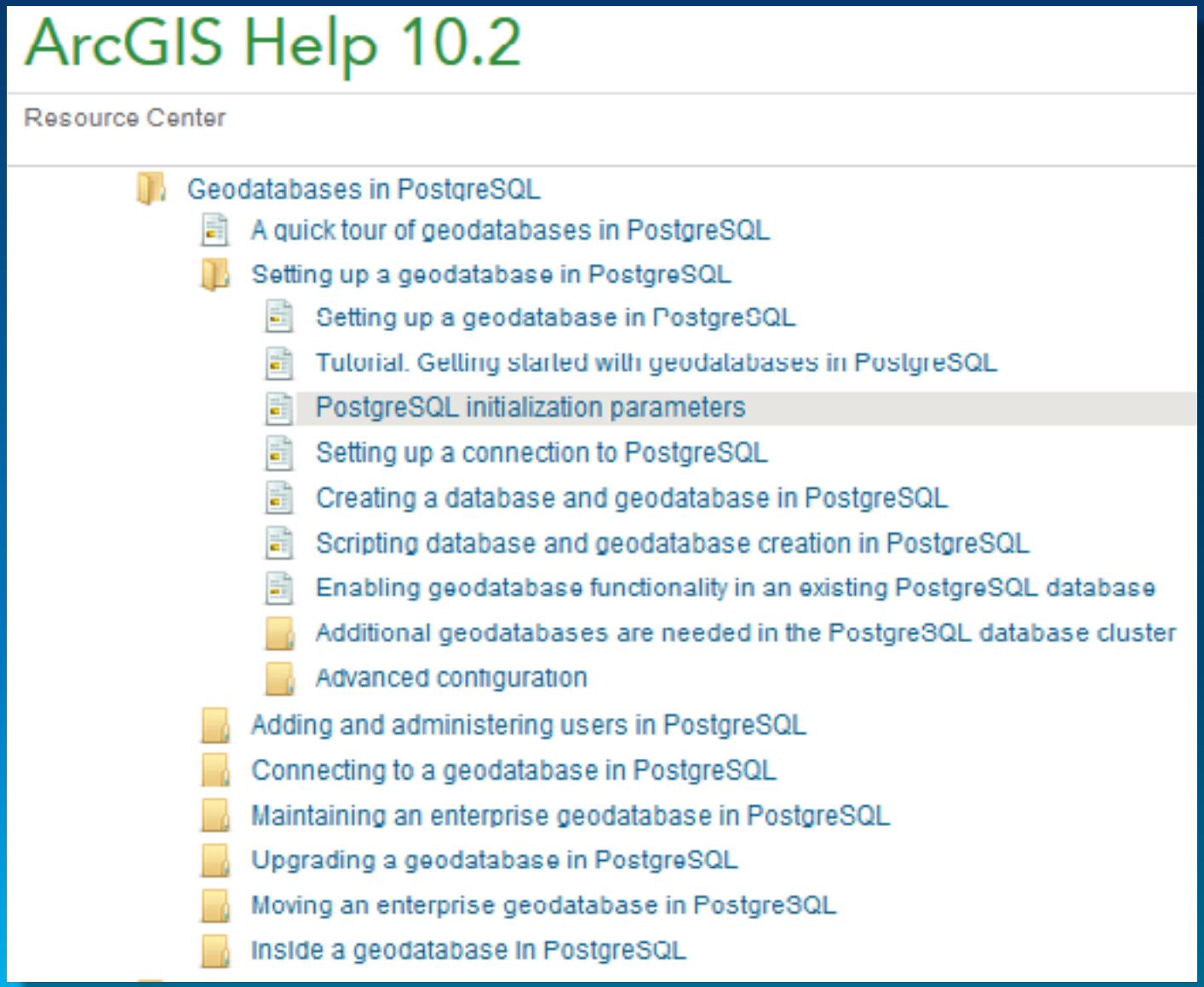
ArcGIS data types	PostgreSQL data types created	Other PostgreSQL data types that can be viewed	Notes
BLOB	BYTEA		
DATE	TIMESTAMP WITHOUT ZONE	TIMESTAMP	
DOUBLE	NUMERIC(p)	BIG SERIAL, DOUBLE PRECISION	The precisio
FLOAT	NUMERIC(p)		The precisio
GEOMETRY	ST_GEOMETRY or GEOMETRY		In a geodata created in th ST_Geomet the ST_Poin created in A subtype you To use ST_ To use the P PostGIS. Se
GLOBAL ID	VARCHAR(38)		Only suppor
GUID	VARCHAR(38)		

Observation: Multibyte characters

- Working with customers we have observed a difficulty working with certain letters.
- Not an issue officially recognized by Postgres
- Seen only on diacritic marks in Latin based alphabets
  - German umlauts: Ä, Ö, Ü
  - Spanish: Ñ
  - French: a, o, u



PostgreSQL Geodatabase Help Topics



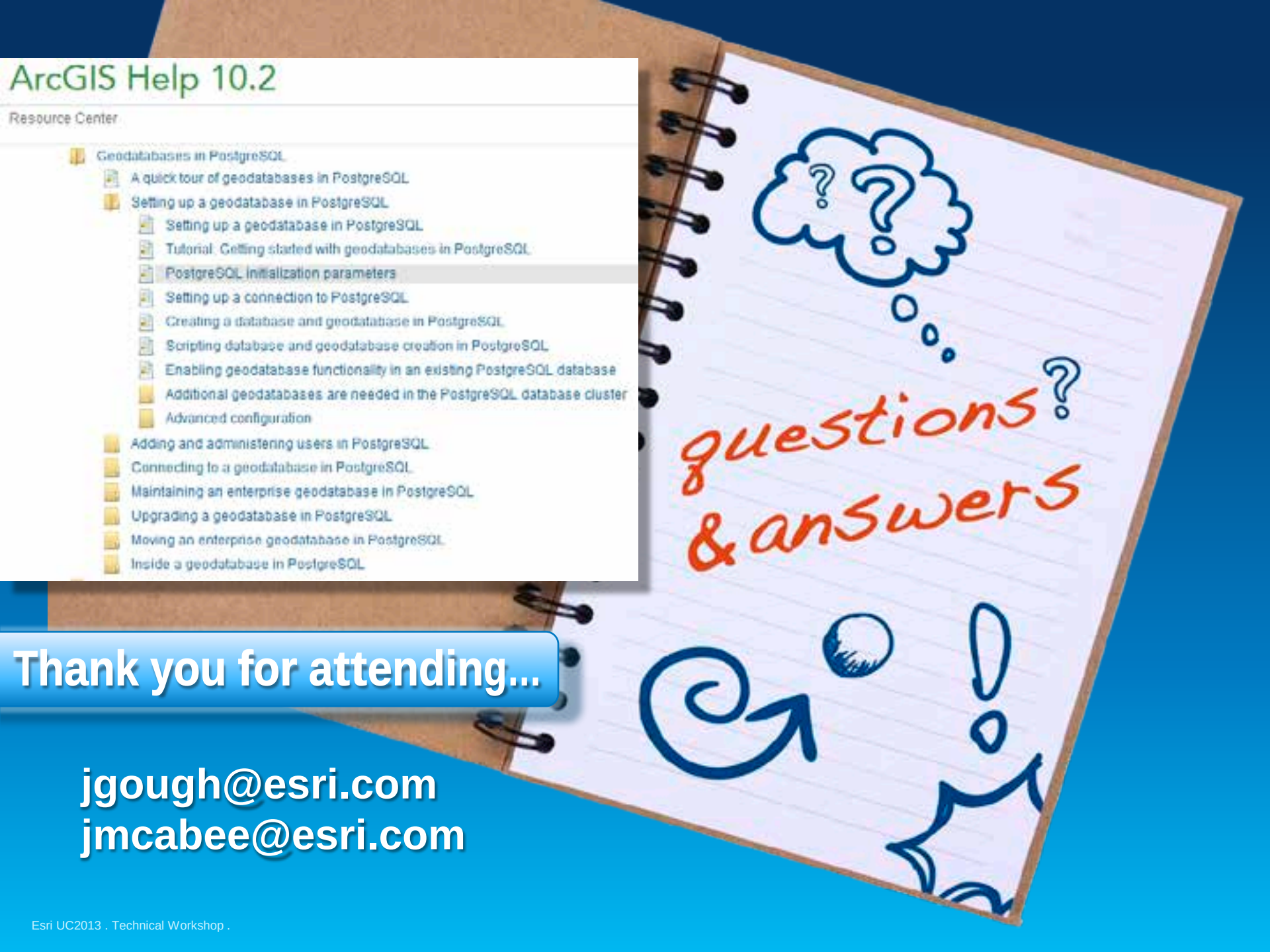
Thank you…

Please fill out the session evaluation

ID: 1193

Online – www.esri.com/ucsessionsurveys

Paper – pick up and put in drop box



Thank you for attending...

jgough@esri.com
jmcabee@esri.com