ArcGIS Technology Workshop Summary and Recommendations

Prepared for:

VCGI

Date: November 2012

Prepared by:

Environmental Systems Research Institute, Inc. (ESRI)
380 New York Street
Redlands, California 92373-8100
Phone: (909) 793-2853

This document contains commercial or financial information or trade secrets of ESRI, which are confidential and exempt from disclosure to the public under the Freedom of Information Act 5 USC 522(b)(4), and unlawful disclosure thereof is a violation of the Trade Secret Act 18 USC 1905.

Copyright © 2009 ESRI
All rights reserved.
Printed in the United States of America.

The information contained in this document is the exclusive property of ESRI. This work is protected under United States copyright law and other international copyright treaties and conventions. No part of this work may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, or by any information storage or retrieval system, except as expressly permitted in writing by ESRI. All requests should be sent to Attention: Contracts and Legal Services Manager, ESRI, 380 New York Street, Redlands, CA 92373-8100 USA.

The information contained in this document is subject to change without notice.

U.S. GOVERNMENT RESTRICTED/LIMITED RIGHTS

Any software, documentation, and/or data delivered hereunder is subject to the terms of the License Agreement. In no event shall the U.S. Government acquire greater than RESTRICTED/LIMITED RIGHTS. At a minimum, use, duplication, or disclosure by the U.S. Government is subject to restrictions as set forth in FAR §52.227-14 Alternates I, II, and III (JUN 1987); FAR §52.227-19 (JUN 1987) and/or FAR §12.211/12.212 (Commercial Technical Data/Computer Software); and DFARS §252.227-7015 (NOV 1995) (Technical Data) and/or DFARS §227.7202 (Computer Software), as applicable. Contractor/Manufacturer is ESRI, 380 New York Street, Redlands, CA 92373-8100 USA.

@esri.com, 3D Analyst, ACORN, ADF, AML, ArcAtlas, ArcCAD, ArcCatalog, ArcCOGO, ArcData, ArcDoc, ArcEdit, ArcEditor, ArcEurope, ArcExplorer, ArcExpress, ArcGIS, ArcGlobe, ArcGrid, ArcIMS, ARC/INFO, ArcInfo, ArcInfo Librarian, ArcInfo—Professional GIS, ArcInfo—The World's GIS, ArcLessons, ArcLocation, ArcLogistics, ArcMap, ArcNetwork, *ArcNews,* ArcObjects, ArcOpen, ArcPad, ArcPlot, ArcPress, ArcQuest, ArcReader, ArcScan, ArcScene, ArcSchool, ArcScripts, ArcSDE, ArcSdl, ArcSketch, ArcStorm, ArcSurvey, ArcTIN, ArcToolbox, ArcTools, ArcUSA, *ArcUser,* ArcView, ArcVoyager, *ArcWatch,* ArcWeb, ArcWorld, ArcXML, Atlas GIS, AtlasWare, Avenue, Business Analyst Online, BusinessMAP, Community, CommunityInfo, Data Automation Kit, Database Integrator, DBI Kit, EDN, ESRI, ESRI BIS, ESRI—Team GIS, ESRI—*The GIS Company*, ESRI—The GIS People, ESRI—The GIS Software Leader, FormEdit, GeoCollector, Geographic Design System, Geography Matters, Geography Network, GIS by ESRI, GIS Data ReViewer, GIS Day, GIS for Everyone, GISData Server, JTX, MapBeans, MapCafé, MapData, MapObjects, Maplex, MapStudio, ModelBuilder, MOLE, MPS-Atlas, NetEngine, PC ARC/INFO, PC ARCPLOT, PC ARCSHELL, PC DATA CONVERSION, PC STARTER KIT, PC TABLES, PC ARCEDIT, PC NETWORK, PC OVERLAY, PLTS, Rent-a-Tech, RouteMAP, SDE, Site·Reporter, SML, Sourcebook·America, Spatial Database Engine, StreetEditor, StreetMap, Tapestry, the ARC/INFO logo, the ArcAtlas logo, the ArcCAD logo, the ArcCAD WorkBench logo, the ArcCOGO logo, the ArcData logo, the ArcData Online logo, the ArcEdit logo, the ArcEurope logo, the ArcExplorer logo, the ArcExpress logo, the ArcGIS logo, the ArcGIS Explorer logo, the ArcGrid logo, the ArcIMS logo, the ArcInfo logo, the ArcLogistics Route logo, the ArcNetwork logo, the ArcPad logo, the ArcPlot logo, the ArcPress for ArcView logo, the ArcPress logo, the ArcScan logo, the ArcScene logo, the ArcSDE CAD Client logo, the ArcSDE logo, the ArcStorm logo, the ArcTIN logo, the ArcTools logo, the ArcUSA logo, the ArcView 3D Analyst logo, the ArcView Data Publisher logo, the ArcView GIS logo, the ArcView Image Analysis logo, the ArcView Internet Map Server logo, the ArcView logo, the ArcView Network Analyst logo, the ArcView Spatial Analyst logo, the ArcView StreetMap 2000 logo, the ArcView StreetMap logo, the ArcView Tracking Analyst logo, the ArcWorld logo, the Atlas GIS logo, the Avenue logo, the BusinessMAP logo, the Community logo, the Data Automation Kit logo, the Digital Chart of the World logo, the ESRI Data logo, the ESRI globe logo, the ESRI Press logo, the Geography Network logo, the MapCafé logo, the MapObjects Internet Map Server logo, the MapObjects logo, the MOLE logo, the NetEngine logo, the PC ARC/INFO logo, the Production Line Tool Set logo, the RouteMAP IMS logo, the RouteMAP logo, the SDE logo, The Geographic Advantage, The Geographic Approach, The World's Leading Desktop GIS, *Water Writes,* www.esri.com, www.esribis.com, www.geographynetwork.com, www.gis.com, www.gisday.com, and Your Personal Geographic Information System are trademarks, registered trademarks, or service marks of ESRI in the United States, the European Community, or certain other jurisdictions.

Other companies and products mentioned herein may be trademarks or registered trademarks of their respective trademark owners.

1.0 Executive Summary

1.1 Background and Goals

The following is a summarization of notes and recommendations resulting from a 2-day workshop held on June 11th & 12th, 2012, to discuss the deployment and use of ArcGIS technology at Vermont Center for Geographic Information (VCGI). The goal of the workshop was to provide database security and management support through knowledge transfer of ArcSDE and ArcGIS Server best practices. The following topics were discussed:

- Review best practices for Geodatabase and ArcGIS Server security and Windows authentication
- Review best practices for ArcSDE data management
- Review data sharing and synchronization best practices
- Support towards implementing data sharing/replication workflows

1.2 Summary Recommendations

- Establish a publishing and editing Geodatabase architecture
  - Create publishing geodatabases as specified in the form of the GDB_ANR, GDB_VCGI, GDB_VTRANS, GDB_<Agency> as articulated in the SDE and Data Exchange protocol
  - Latest versions of ArcGIS technology being implemented by state agencies.
  - Ensure proper test and staging environments are available.
- Each Agency within the state participating in sharing data should strive to implement an Enterprise Geodatabase where a “publication/publishing/sharing” Geodatabase can be created which VCGI can setup "synchronization" processes against.
- Rework and refine ArcSDE configuration and management guidelines in the VGIS Handbook drafted in December 2006 based upon:
  - Newly developed state-wide data management, sharing and synchronization architecture and management practices.
  - Latest versions of ArcGIS and RDBMS technology being implemented by state agencies.
- Compile and maintain a list of map services provided by VCGI and other state agencies to include
  - Information about audience, security requirements, data layers used, data currency requirements
  - Information about high-availability(HA) needs and Mean-Time-To-Recovery (MTTR) needs (how long can service be down)
- Redesign the VCGI's Data Warehouse (which is used to compile and maintain a list of feature classes/layers provided by VCGI and other state agencies) to include
  - Information about audience, security requirements, projection, data update timing
- This could be a document but may be more easily maintained if it is integrated into the Warehouse as a “metadata/dictionary tables” within each GDB_<Agency> Geodatabase, possibly something to consider for the next revision of the Warehouse design if it is not possible to implement in the current. Establish performance measuring mechanisms to establish performance baseline characteristics as new virtual environment is implemented and put into production
  - Capture connection characteristics (process_information table in Geodatabase), connection counts, which services, which server etc…
  - Capture cpu, memory, disk i/o, network usage – snapshots at specific times of day.
    Where possible establish a recommended set of performance metrics to capture and snapshot time for all agencies (same information at same snapshot times – e.g. Mondays at 3pm). This can be captured from the Windows Task manager/Resource monitor, or jobs can be setup and automated to capture this information. Capture this information at the same time as the Geodatabase connection information (a simple query on

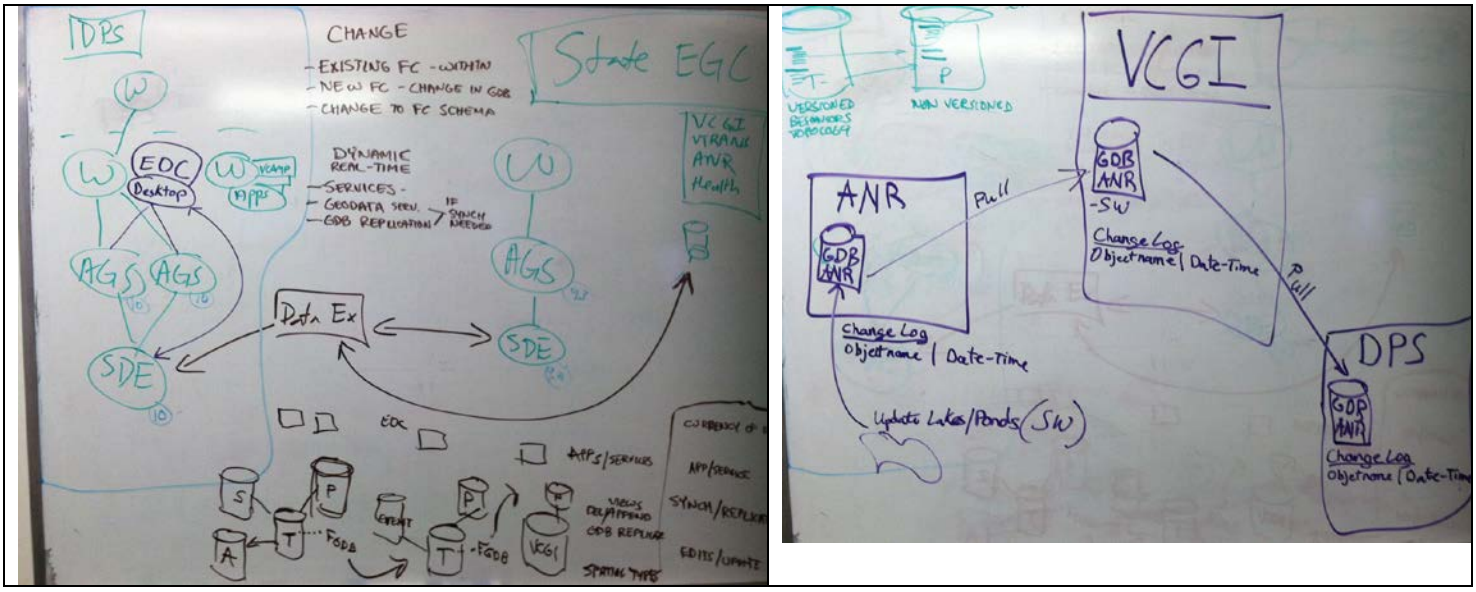
the sde.process_information table to understand number of connections and from which desktops or services they are coming from).

2.0 ArcGIS Technology Workshop Discussion and Recommendations

The workshop involved interactive whiteboard discussions over the two days. The first day primarily focused on Geodatabase architecture, data sharing, synchronization and security topics. The second day continued these discussions and added further discussion regarding ArcGIS Server architecture and security topics. Most of the following topics specified by VCGI, were discussed during the workshop:

- Reviewed current Geodatabase and ArcGIS Server architecture within state government agencies
  Discussed some sample application workflows with specific security requirements (primarily from the Department of Public Safety)
  Geodatabase data sharing and synchronization architecture, methods and best practices
    - Heterogeneous environments (different RDBMS, SDE and AGS versions)
    - EGDB -> fGDB, EGDB->EGDB, with and without GDB replication (different versions)
    - Geodata services and data exchange/sharing (Pros/Cons/Issues)
    - Handling table locks caused by map services ("is versioning only option?")
    - Automation for copy/paste because it keeps object level metadata like domains, subtypes, etc…Reviewed current EGC SDE guidelines to provide context
    - With the basic concept that each Agency has its own Geodatabase(s).
    - SQL Logins and Roles
  Best practices for Windows AD security integration with MSSQL and SDE – multiple domains? Trusted networks?
  ArcGIS Server Security
    - HTTPS, Windows Domain Authentication, Multiple AGS Services admins/publishers – VCGI, E911, etc.. share an instance
    - Publishing services via Desktop – Access levels requires (admin vs publishers)
  Imagery Caching
    - Best fit for Image Services and cache creation
- Compressed vs. uncompressed
- Mosaic Dataset vs. image catalogs
- Is cache create from the image service or does it dial back to the base imagery?
- Performance issues/tuning
- Coordinate systems (state plane vs. web Mercator) – pre-process or reproject on-the-fly
    - AWS and elastic block storage configurations

Most of these topics were discussed, the notes summarization focuses on those topics discussed in the most detail.



Data Management, Sharing and Synchronization

The current scenario in the state was discussed. Various agencies have deployed ArcGIS/ArcSDE technology and use it to manage their GIS data. Most agencies acquire basemap data and data from other agencies via VCGI's Data Warehouse, either by copying the data or more recently by taking advantage of ArcGIS Server map services some of the state agencies and VCGI have published. VCGI, in the context of the strategy developed by the Enterprise GIS Consortium (EGC), is working on establishing a federated framework for the sharing and synchronization of data across state agencies. One of the concerns with the current architecture and environment is the creation of "old" copies or silos of data, which become "disconnected" from their original authoritative data sources. The Data Exchange Protocol does define a manual process for "synchronizing" or copying the data which allows for the use of File Geodatabases (FGDB) which EGC partners can use to update their repositories. The process of updating data via copying (or even delete and append, bulk update processes) typically does not have a synchronization tracking component. It then becomes the responsibility of the individual or organization to manually keep track of this. There were discussions during the workshop about how to maintain a record of the copying or updating of the data (e.g. when it was last done) to ensure where partners are keeping their own copies of the data, that it is in fact the most current data available. This can be something as simple as a database table they update when they last did a manual copy of the data this way. The copy-paste process can be automated by creating an delete-append geoprocessing tool/script which is then automated via the window's scheduler to run at a specific time each week or month against whatever File Geodatabase may be in a specific location. The feature classes in those file geodatabases could likewise be updated via similar scripts upon an automated basis. This automation could help to remove some risk that an update cycle might be missed, and can be considered. Multiple options for keeping agencies/partners in synch will be the most flexible and allow a "ramping" of the technique they use as they mature in the use of the various ArcGIS technologies. EGC partners also have the option to directly connect to VCGI and ANR's SDE databases directly removing the need to retain internal copies.

The EGC in May 2011 approved a Geospatial Data Exchange Protocol for the following purpose:

"This data exchange protocol addresses the exchange of file-based and ArcSDE datasets. It addresses the sharing and exchange of data in the public domain, as well as data that may have access constraints (confidential or sensitive data exempt from Federal and State public record law)." – EGC Geospatial Data Exchange Protocol, May 2011

The guiding principles behind this protocol were the following (from the EGC Geospatial Data Exchange protocol, May 2011):

1. Ensure consistency and predictability.
2. Be based on established standards.
3. Ensure data integrity through the use of appropriate data management practices and exchange mechanisms.
4. Allow for the exchange of information between heterogeneous environments.
5. Control access to "secure and confidential" data.

Proposed Data Exchange mechanisms:

1. Portable Data Packages (via File Geodatabases or Personal Geodatabases)
2. Direct Data Connections
3. ArcGIS for Server Geodata Services

Naming conventions (detailed notes below*):

Database	Steward/Source
GDB_VCGI	VCGI
GDB_ANR	ANR
GDB_VTrans	VTrans
GDB_ACCD	ACCD

* EGC Proposed Database Naming Convention: The following naming convention will be used.

Naming Convention: GDS_<AGENCY_ACRONYM>
Example: GDS_VCGI
The naming convention also supports the notion of application specific geodata services which agencies want to share with other EGC members.
Naming Convention (application specific): GDS_<AGENCY_ACRONYM>_<APP>
Example: GDS_VCGI_VirtualVT
Naming Convention (secure and confidential data): GDS_SECURE_<AGENCY_ACRONYM>
Example: GDS_SECURE_VCGI
Naming Convention (secure and confidential data - application specific): GDS_SECURE_<AGENCY_ACRONYM>_<APP>
Example: GDS_SECURE_VCGI_VirtualVT (*to support VirtualVT application*)

This naming convention is for "Geodata Services". 2 additional naming conventions have been defined for the 2 other data exchange "methods" defined in the protocol.

Critical Factors in Determining Sharing and Synchronization Architectures

The critical factors in determining the best mechanisms and architecture for the sharing and synchronization of data are:

1. The editing workflows producing the data.
2. The update cycle or frequency of data updates.
3. The types of data stores (file based or Geodatabase).
4. The architecture (single or multiple databases, feature classes, feature datasets, etc..) of the Geodatabase(s) involved.
5. The transport mechanisms and performance between sites (network).
6. The data characteristics and requirements (e.g. secure vs. non-secure) and stability of the schema (fields) and behaviors.

Summary of Primary Editing Workflows

There are many different editing workflows that can be used to manage data, which may be secure or non-secure in nature. These workflows fall into two primary categories:

1) Transactional
   a) ArcGIS Versioned editing (long-term transactions) – providing support for
      i) Long-term editing with undo/redo and conflict resolution
      ii) Use of versions for QA/QC, project phases and security
      iii) Geodatabase archiving
      iv) Geodatabase replication and synchronization
      v) Required for complex features such as Geometric Networks

b) ArcGIS non-Versioned editing(short-term transactions) – provide support for
   i) Less administration, but no undo/redo or conflict resolution

2) Non-Transactional – batch or bulk data updates

The choice of what type of editing workflow and its architecture is dependent upon the requirements of the organization and characteristics of their data. Some of the following requirements should be considered:

Functional Data Requirements	• simple or complex features • behaviors (GDB domains, RDBMS constraints) • schema or behavior modifications required
Application Requirements	• currency of data (dynamic, weekly, annually)
Workflow Requirements	• single user • multiple user – single office • multiple users – multiple locations • multiple ArcGIS based products • multiple ArcGIS and 3rd party
Deployment Pattern Requirements	• Desktop • Web Editing • Mobile (Mobile, iOS, Android….) • Survey Devices • Other Devices
Infrastructure Requirements	• High Availability (HA) • Mean Time to Recovery (MTTR) • Disaster Recovery (DR)

Where possible each agency should have some type of mechanism for managing metadata/information regarding the authoritative layers/feature classes and/or tables in their database. VCGI has already built a distributed data cataloging systems as defined here http://www.vcgi.org/dataware/?page=./search_tools/linked_servers.cfm. There is a planned revision to this system. Any new revisions should ensure that some of the suggested characteristics discussed in this document are included.

Feature class	Complex data	Non-ESRI access	Isolate edit sessions (QA)	Geodatabase replication	Archive
CAL_LANDCOVER	No	No	Yes	Yes	No
CELL_TOWERS	No	No	Yes	Yes	No
CITY_LIMITS	No	No	Yes	No	Yes
FREEWAYS	No	Yes	Yes	No	No
MAJOR_ROADS	No	Yes	Yes	No	No
PARCELS	No	Yes	Yes	Yes	Yes
PARKS	No	No	Yes	No	No

E	F	G	H	I	J	K	L	M	N	O	
Projection	Geographic Coordinate	Table	Relationship Class	Update Interval	Where Updated	In Other Database	Update Method	Add To Publishing	Replication to Notebooks	Replication Type	
Tranverse_Mercator	GCS_North_American_198		No	Weekly	Office/Field	Yes	FC Replication		Yes		May be
Tranverse_Mercator	GCS_North_American_198		No	Yearly	Office	Yes			No		Maintai
Tranverse_Mercator	GCS_North_American_198		No	Yearly	Office/Field				No		Update
Tranverse_Mercator	GCS_North_American_198		No	Yearly	Office/Field				No		Maintai
Tranverse_Mercator	GCS_North_American_198		No	Yearly	Office	Yes	Delete/Append-GP Tool				Drawn
NAD_1983_StatePlane	GCS_North_American_198		No	Twice A Yea	Office/Field				No		Update
			No	Daily	Office/Field	Yes	FC Replication		Yes		Also pa
		Yes	No	Yearly	Office/Field	Yes	Delete/Append-GP Tool		No		OEP
Tranverse_Mercator	GCS_North_American_198		No	Yearly	Office/Field	Yes	Delete/Append-GP Tool		No		Update
Tranverse_Mercator	GCS_North_American_198		No	Yearly	Office	Yes	Delete/Append-GP Tool		No		Drawn
Tranverse_Mercator	GCS_North_American_198		No	Daily	Office/Field	Yes	FC Replication		Yes		Drawn

Locks and Feature Dataset Visibility

Two specific concerns came up during the workshop discussions relating to the management of the data.

1. Shared table locks – how to deal with the shared table locks that ArcGIS Server and Desktop place on feature classes served up by map services, and
2. Visibility of feature datasets when user(s) may not have permission to any of the underlying feature classes.

The creation of publishing vs. editing Geodatabases was suggested. This may require some modifications to the EGC's Geodatabase naming conventions. The publishing Geodatabase that supports the ArcGIS services could be called GDB_VCGI_web for example. Naming conventions are in place, some discussion should take place with the partners to see if they are sufficient to differentiate between publishing and editing geodatabases. The simplest may be to keep the publishing geodatabases named for the agency (e.g. GDB_VCGI, GDB_ANR, etc..) as defined. Then make the production editing Geodatabase for agencies that may have these, something more descriptive such as GDB_ANR_EDIT.

Several methods are available which can allow for the update of feature classes being served out through map services, depending upon whether the features are simple or complex, these include:

1. Use of publication geodatabases
2. Use of "covering" ArcSDE views for simple feature classes, or a simple feature class in a feature dataset or base tables of versioned feature classes, these views could use a specific naming convention so they are understood to be views (e.g. V_<SourceFCname>).
3. Adjust SchemaLockingEnabled parameter for ArcGIS Server
4. Geodatabase replication and synchronization workflows

While the data can be updated, the feature class cannot have its privileges, schema (fields), or other "definition" characteristics modified, it can only be manipulated when the shared table lock is present due to its being published

through a map service or being used by a Desktop user. Data manipulation (e.g. SELECT, INSERT, UPDATE, DELETE), also know as DML is ok, DDL (Data Definition Language) can only be done when shared table locks are not present. DML and DDL are the two primary categories of SQL Statements used in databases. To manipulate the "definition" characteristics or behaviors of tables or feature classes, the locks protecting the referential integrity of the data need to be released. This requires any ArcGIS for Desktop or ArcGIS for Server applications and services to be stopped to release those locks. Many methods can be implemented for this, but they require an understanding of the applications or services high-availability (HA) requirements regarding downtime. How long can an application be down? (*comment from Ivan - There was a time when I proactively established a weekly window of opportunity for geodatabase tasks that require down time. There was a designated hour in the week for the down time (whether the hour was needed or not). The organization was aware that the geodatabase can be unavailable on day x of each week during hour x.)*
This can determine whether the services can be temporarily stopped for specific update operations or not. The more complex the behaviors and features being served the more thought has to be given to how the data updates will be done. Various methods include

1. Direct update of data with services published against "covering" ArcSDE views (simple features – stand alone or in a feature dataset, or base tables of versioned feature classes only posted data would be seen in view) – 7x24
2. Direct update of data with delete and append geoprocessing tools (7x24 depending upon characterstics of data being published)
3. One-way Geodatabase Replication synchronization. This can be between versions (e.g. 10 to 9.3) – 7x24
4. Round-robin updates, where duplicate map services are maintained point to two different geodatabases and one is updated while the other is serving up the data through its map services. There are variations on this where this could incorporate not only multiple geodatabases, but also multiple arcgis servers which are load balanced, allowing for servers to be shutdown, updated and subsequently turned back on after update – this is useful when the services and/or map documents themselves have to also be updated due to more comprehensive schema changes. – very brief interruption of few seconds to few minutes.
5. Shutdown of services while data is updated – interruption of couple minutes to hours depending upon amount of data being updated

Where ArcGIS services are required to be 7x24, ArcSDE views can be used for simple feature classes. *Simple Feature Classes are stand-alone feature classes, in comparison to the more complex "composite" structures that are found in the Geodatabase, such as Geometric Networks which are comprised of multiple simple feature classes. Simple feature classes can live in a feature dataset. A database view only understands the table it is built upon, the RDBMS does not know about geometric networks, topologies and other constructs that ArcGIS assembles at the application or desktop client.* Services can then be built on top of the views, this then allows the underlying data to be updated as needed without having to stop the server.

<u>Enterprise Geodatabase Schema Lock Behaviors</u>

ArcGIS technology when working with data in the Geodatabase must work on the principle that the geodatabase schema remains fixed and is not changing at any time that you are working with the contents. ArcGIS automatically acquires shared locks on individual datasets when they are in use. This information is maintained in the geodatabase's administrative tables (primarily table_locks, but 3 other lock table are present for other types of locks – state, object, layer) and relates back to connection information maintained in the process_information table. Two types of locks exist – shared and exclusive. Shared schema locks allow multiple users to select and edit the data in a table but not make changes to its schema or behaviors. There can be any number of shared schema locks on an individual feature class or table, however, a shared schema lock will prevent an exclusive schema lock from being established. An exclusive schema lock is placed on a feature class or table when modifications to its schema or behaviors are taking place. Ver They are typically very brief in duration, but the feature class or table cannot be in use by any other users or server connections. More detailed information on this topic is available in the ArcGIS Resource Center - http://help.arcgis.com/en/arcgisserver/10.0/help/arcgis_server_dotnet_help/index.html#//0093000000sv000000

The behavior of schema locking for Geodatabase Feature datasets should also be noted. Feature datasets are a Geodatabase object, managed by ArcObjects, that allow you to group together related feature classes so they can participate in specific functionality. Typically some type of spatial relationship such as a geometric network, network dataset, parcel fabric, etc… Because of this, certain behaviors and functionality implemented by the Geodatabase, are done feature dataset-wide, to all the feature classes in the feature dataset. If one feature class in a feature dataset is being used in a map document in ArcGIS for Desktop or being published through ArcGIS for Server, then shared table locks are placed on all of the other feature classes in the feature dataset. This behavior remains the same in version 10.1. If finer granularity is need for security purposes then stand-alone feature classes should be considered or possibly the use of views where appropriate. These views can be spatial which would include the geometry field and appear as a layer to the Geodatabase either through registering the view, creating it using the available GP tools in 10.1, creating through the sdetable –o create_view command or use of query layers to expose the view. The views are comprised of SQL which can be modified to present only specific fields where security is required, as well as more advanced logic where needed.

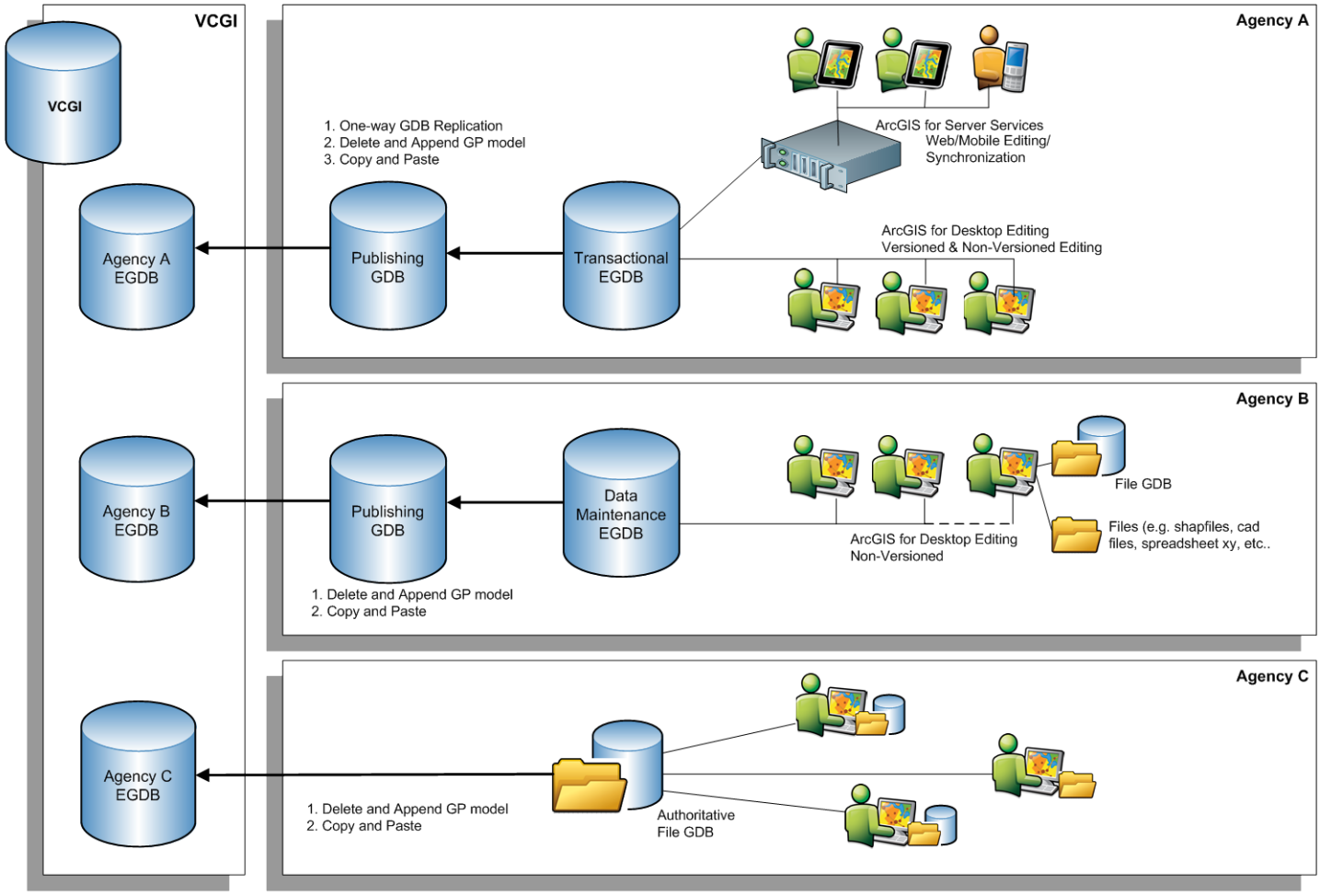
Connection Architecture Recommendations

Most users should be using layer files or mxd templates to connect and work with the data. Ideally those connections should be made to SDE "covering views" to eliminate table locks. If they are only viewing (read-only) they should be directed to ArcGIS Server Map or Feature Services depending upon the functionality required. It is recommended that where possible, all connections be migrated to direct connect architecture.

Other Architecture Recommendations

Generally test/staging and development environments are a recommended best practice.
VCGI traditionally has not had resources to configure test/staging and development environments, only in some ad hoc cases have these been configured as resources were available.

Managing Change within feature classes



Where versioned editing is used the modifications can be done through Geodatabase Replication synchronization processes. These can be within an agency, setup in a one-way replica to a publishing Geodatabase that is used for synchronizing with the statewide publishing Geodatabase maintained for each agency within VCGI's publishing SQL Server instance. However, in most cases, editing workflows may remain non-versioned in nature. In these cases change is not being tracked and a bulk update of some type must be made when the feature are modified. As shown in the diagram above there are several methods (not all are shown, for example SQL could be used in some cases) for the updating and synchronization of data. The easiest method for bulk updating data that has the potential for lowest impact to applications is combining the delete and append geoprocessing tools into a model and automating this. To provide finer granularity this could be done for each feature class or feature dataset and run based upon a synchronization table established in the database, which would describe which feature classes have been modified. In this way, only those feature classes that have been modified would be truncated and updated. These synchronization processes however only discuss updates to the data within the feature class or table (tables could be handled in a similar manner). The more complex the data becomes – geographic networks and feature-linked annotation for example – the more likely versioned editing workflows will be necessary. However, even with the versioned editing workflows, it is possible for the agency to publish the underlying feature classes out through one-way replication to stand alone feature classes, if the geometric network or topology does not need to be published out.

How the copying or synchronization of data is automated can be done through several different methods. There could be a DATA_CATALOG_<Agency> table within each GDB_<Agency> describing all of the objects in the database and how they are handled including flags for sync/update procedures. These catalog tables could be further versioned and synced to master DB used for data discovery, however this could add further complexity. This was an area of discussion that

would likely require further elaboration on the design and mechanisms (scripts) that might be put in place to automate it. Further discussion on this topic is noted in the VCGI-Agency Synchronization table section.

Managing Schema and Behavior Changes

A significant discussion during the workshop revolved around change in a federated spatial data sharing environment. This was whether there was an ability for VCGI to know whether other state agencies have added new data (feature classes or tables) or made schema or behavioral changes to existing feature classes or tables. Some queries could be run against the Geodatabase administrative schema. Establishing data characteristics, such as those feature classes or tables with stable schemas vs. dynamic schemas, where fields will remain stable, vs. those which are more likely to have fields change.

The impacts of schema changes can include the following, these focus on schema changes within the context of Geodatabase Replication but some of these are present with all forms of synchronization (*extracted from ArcGIS resource center):*

- *Edits that don't synchronize—Data synchronization only imports changes for tables and fields that are common to both replicas. If an edit is made to a field that is not in the relative replica, it will not be applied when importing changes. Another example is when a geometric network is dropped on one replica but not the other. When this happens, the orphan junction class is removed and, therefore, changes from the relative orphan junction class are not applied.*

- *Invalid values—Changes that violate domains, subtypes, connectivity rules, and relationship rules are applied when synchronizing changes. The validation tools on the editor can be used to check the newly imported values.*

- *Data synchronization errors—These can happen when you manually make a schema change to both replicas. For example, you may want to add a field to a table. If you do this, be sure to make the exact same schema change in all cases. If there is a difference (for example, a field is a string on one replica but an integer in the other), a data synchronization error will occur.*

- *Unsupported changes—Some types of schema changes can cause synchronization to fail, but no warning is displayed if you make the change. These changes are not detectable by the geodatabase replication system. They include database-level operations like changing permissions on tables in the database. If permissions are changed to read-only for replicated data, a failure will happen when trying to import changes from the relative replica.*

For those instances where versioned editing workflows are in use the following ESRI KB article may be useful – FAQ: What schema changes can be made to versioned data?
http://support.esri.com/en/knowledgebase/techarticles/detail/29885

Geoprocessing tools for comparing schemas are available but these are intended for use with Geodatabase Replica pairs used in Geodatabase Replication based synchronization. Synchronization methods outside of the transactional versioned editing workflows participating in Geodatabase replication, such as bulk update synchronization methods using the delete and append gp tools would need to use some other mechanism. Schema changes that are only focused on either addition of new feature classes or modification of fields within feature classes can be reviewed by checking the following Geodatabase administrative tables between the 2 geodatabases – column_registry, table_registry and layers. This could be implemented through the use of SQL and python. Modifications of behaviors such as topology rules, domains or geometric network rules will require more sophisticated custom detection methods using ArcObjects. If schema changes (e.g. fields added or removed) or new feature classes are created then in those cases deleting the feature class and replacing it, or creating the new feature class would be necessary in a bulk

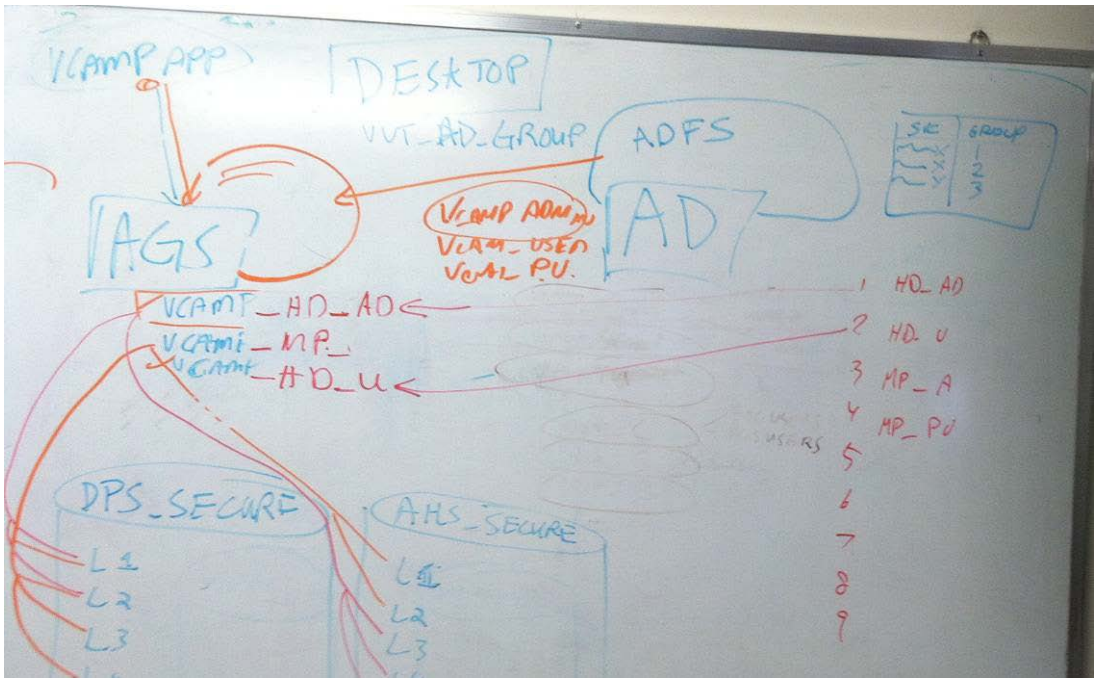
update/synchronization scenario vs. Geodatabase replication where schema changes can be done to both sides of the replica pair. New feature classes however, can not be added at v. 10 without the recreation of the replica pair.

VCGI-Agency Synchronization table

Methods for keeping the copying and manual synchronization processes current were discussed. This discussion revolved around whether ECG/VCGI would push/pull updates to agency publishing gdb's (eg: GDB_ANR, GDB_VTrans, etc) from central instance – this is likely the easiest to manage and automate but requires coordination regarding the automation of the GP models or scripts involved. It is likely however that some of the Agency DBA's may not want VCGI pushing data into their databases, so it was suggested that some type of a simple Synchronization Table be created that would record the date of the last time a feature class or table was updated, that way an agency or the script could check and see when the last time the update/synchronization model or scripts were run and only update/synchronize those feature classes or tables that require updating. It was suggested that the scripts be created on a feature class by feature class basis to allow for the incorporation of new feature classes as an agency may have the need to create one. These scripts/models could be done on an agency by agency basis, with the ultimate goal being a standard automated synchronization process taking place as needed.

Some discussion of whether to remain using feature datasets took place. Feature datasets should primarily be used when some type of spatial behavior is being established/enforced across a set of feature classes (e.g. a Geometric Network). There are cases where customers have implemented feature datasets for organizational purposes, but they do carry some overhead (e.g. shared table locks placed on all feature classes even if only 1 is "touched", privileges granted to all, etc…). This is a decision that can likely be made on an as needed basis, as synchronization and editing workflows are scoped and implemented.

ArcGIS Server Application and Security Notes



VCGI has the following Services currently configured:

- basemap_services
  - basemap_services/VGIS_BASEMAP_OS_SP (MapServer)
  - basemap_services/VGIS_BASEMAP_OS_WM (MapServer)
  - basemap_services/VIL_CACHE_COLOR_SP (MapServer)

basemap_services/VIL_CACHE_COLOR_WM (MapServer)
basemap_services/VIL_CACHE_PAN_SP (MapServer)
basemap_services/VIL_CACHE_PAN_SP_test (MapServer)
basemap_services/VIL_CACHE_PAN_WM (MapServer)
basemap_services/VIL_NAIP2003_COLOR_SP_BETA (MapServer)
basemap_services/VIL_NAIP2008_CIR_SP_BETA (MapServer)
basemap_services/VIL_NAIP2008_COLOR_SP_BETA (MapServer)
basemap_services/VIL_NAIP2009_COLOR_SP_BETA (MapServer)
EGIS_geocoding_services
EGIS_geocoding_services/GEOCODE_COMPOSITE (GeocodeServer)
EGIS_geocoding_services/GEOCODE_COMPOSITE_WM (GeocodeServer)
EGIS_geocoding_services/GEOCODE_ESITE (GeocodeServer)
EGIS_geocoding_services/GEOCODE_ESITE_WM (GeocodeServer)
EGIS_geocoding_services/GEOCODE_RDSE911 (GeocodeServer)
EGIS_geocoding_services/GEOCODE_RDSTIGER (GeocodeServer)
EGIS_geocoding_services/GEOCODE_STREETMAPUS (GeocodeServer)
geocoding_services
geocoding_services/GEOCODE_ESITE (GeocodeServer)
geodata_services
geoprocessing_services
globe_services
image_services
map_services
map_services/CRS_STRUCTDENS (MapServer)
map_services/NHD_BASEMAP (MapServer)
map_services/NHD_GNIS (MapServer)
map_services/NHD_HYDRO (MapServer)
map_services/VTREA_Demo (MapServer)
Other_Services
Test_Services
Test_Services/TOPO_TEST_SP (MapServer)
Test_Services/VGIS_BASEMAP_OS_SP_AWS_SDE (MapServer)

The following help topics discuss version 10 Internet Security –
http://help.arcgis.com/en/arcgisserver/10.0/help/arcgis_server_dotnet_help/index.html#/Internet_security_checklist/0093000000pq000000/
http://help.arcgis.com/en/arcgisserver/10.0/help/arcgis_server_dotnet_help/index.html#//0093000000q1000000
http://resources.arcgis.com/content/enterprisegis/10.0/authentication_mechanism
http://help.arcgis.com/en/arcgisserver/10.0/help/arcgis_server_dotnet_help/index.html#//0093000000pp000000

Notes on Security

Typically for ArcGIS Server, a best practice would be to put the public (non-secure) server into a DMZ, or at a minimum into its own server, whether physical or virtual. Then the internal (secure)/intranet server is placed on internal networks with security applied. However, many different configurations and security requirements can exist both externally and internally. The following are some tips related to helping prevent unauthorized access or use. Unauthorized use is any violation of law or license.

Most common examples:

- Deleting data

- Service being accessed by unauthorized client
- Manager exposed to Internet
- Opening up "services directory" to cross-site scripting attack
- People can "sniff" my data

Some precautions to take:

- Every feature service with the edit capability should be secured.

- Unauthorized client: use long-term token combined with monitoring, monitor and change as necessary.

- Do not let manager be exposed to the Internet: Normally administration actions through DCOM are blocked by firewalls. Manager secured by username/password could be subject to brute force attack. Java Manager on port 8399, is usually safe and can be turned off. .Net Manager is usually installed on port 80 (same as services). Instead don't install Manager on default instance - Run AddInstance.exe to add Manager to second instance – http://bit.ly/nj69gP

- Services Directory Cross-Site Scripting: Upgrade to 10 SP2, Turn off services directory when exposed to Internet. Help for Java and .Net here: http://bit.ly/ngX0BD

- Precaution against data sniffing: Make your folders encrypted in Manager, this will require https and won't allow http requests. If you don't check though, encrypted people can accidentally use http. See Doc: http://bit.ly/pzflyc

User Security

A common pattern for security on users – is to have sde administrative user and data owners be database authenticated, and all other users externally authenticated (e.g. Active Directory). There may be some cases where there are also some generic department/agency users used for web services that are also database authenticated.

Appendices – Supporting Notes and Related Documents

ArcGIS Server Architecture, Deployment and Performance

ArcGIS Server at 10.1

ArcGIS for Server has been rearchitected at 10.1 to work in 64-bit architectures and use a more cloud-friendly, web services-oriented design. The following architectural changes have been made:

- ArcGIS for Server runs exclusively as a native 64-bit application.
- The SOM and SOC have been replaced with a single component, the GIS server known as the ArcGIS Server site. A site is a deployment of one or more machines (GIS servers) that have ArcGIS for Server installed and work together. This provides a more robust architecture reducing chances for failures and simplifying the provisioning and recovery of new machines.
- ArcGIS Web Adaptor, an application that runs in your existing web site and forwards requests to your GIS servers. It polls your site at a regular interval to learn which machines have been added or removed. It then forwards traffic to only the currently participating machines. It is designed to integrate with your existing web server. It is compatible with IIS, and Java EE servers such as WebSphere and WebLogic. This component is recommended when you configure a site with multiple machines, and/or when you want to tighten security on your GIS server.

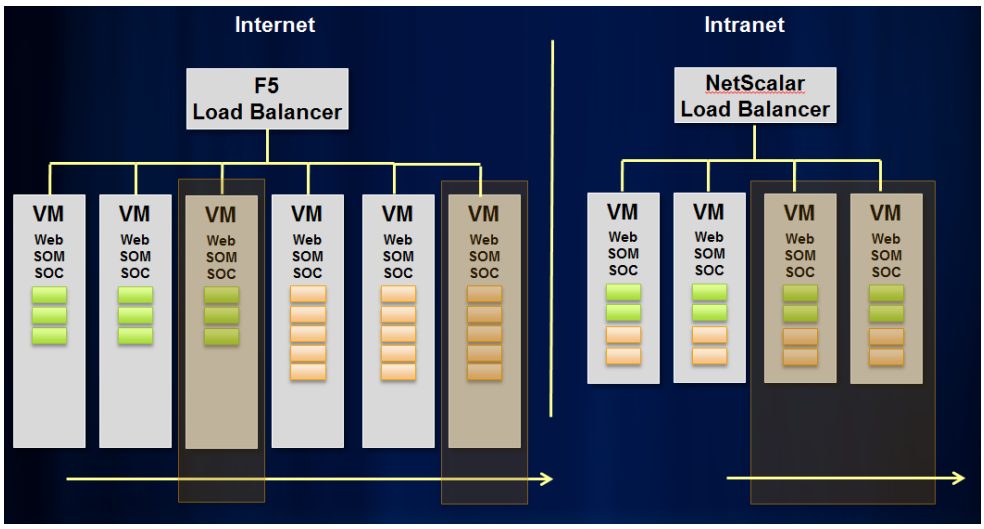
- ArcGIS for Server local (DCOM) connections are no longer supported. All communication with GIS services is through HTTP using SOAP or REST.
- ArcGIS for Server can host services out of the box and does not require a web server on installation. You can optionally connect it to your own web server using a new component called Web Adapter.
- The new REST-based ArcGIS Server Administrator API allows you to administer your ArcGIS Server site through scripting. With this API you can automate tasks such as creating a site, adding machines, publishing services, querying the logs, and starting and stopping services.

The installation process and configuration of multiple machine configurations is easier as well (this is also the reason for horizontal scaling discussed in the next section). The following improvements have been made at 10.1 for its installation and configuration. The ArcGIS for Server installation experience has been greatly simplified to eliminate many dependencies and steps that were required in previous releases.

- You are no longer required to install the .NET framework or a particular Java runtime.
- Integrating with a separate web server (Microsoft IIS, WebSphere, and so on) is no longer a requirement. Installing ArcGIS for Server provides you with a ready-to-use web services-based GIS server, which includes all web management tools and applications.
- During the setup you provide you'll be asked to provide just one account that will access your GIS resources, data, and run the ArcGIS Server service. This account is called the ArcGIS Server account. At 10.1, there is no postinstallation to complete, and there are no SOM, SOC, or ArcGIS web services accounts to configure.
- There are no dependencies on DCOM, simplifying the configuration of firewalls within your network.

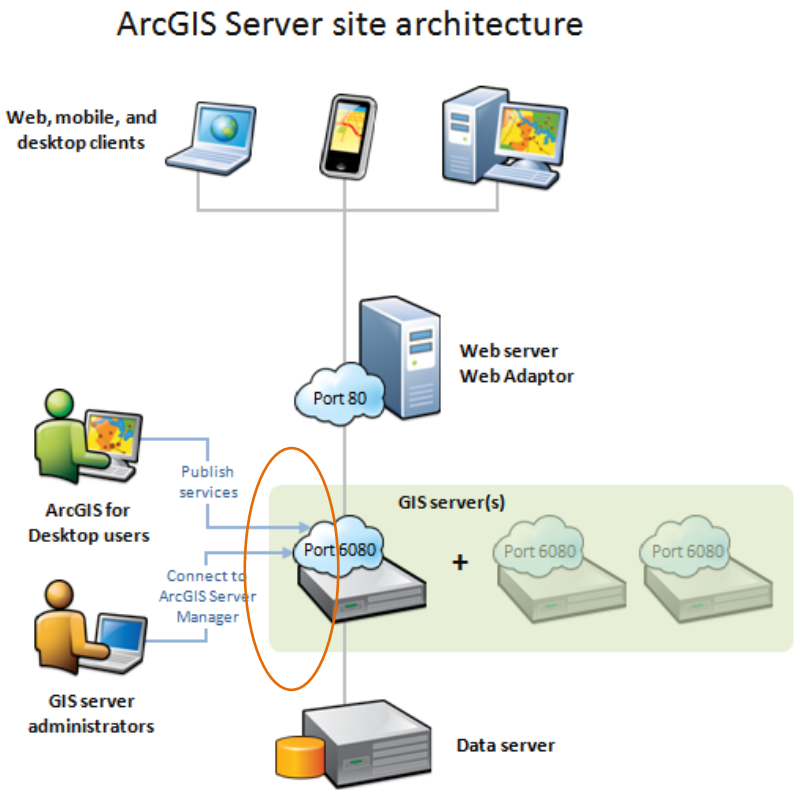
The process of creating a multiple-machine deployment of ArcGIS Server has also been simplified. You run the same installation on each machine and connect the machines using ArcGIS Server Manager.

<u>ArcGIS Server Architecture: Horizontal Scaling</u>



The graphic above is an example of horizontal scaling. Horizontal scaling as defined here, is the spreading of applications/services across multiple servers, compared to the addition of additional resources (e.g. cpu, memory) to a single server (physical or virtual). Traditionally, in the past, physical servers have been deployed and used vertical scaling where more resources are added as needed. With the advent of and maturing of virtualization environments dynamic scaling through the addition of duplicate load-balanced servers is possible, providing for a more elastic environment. This can be done in ArcGIS Server version 10.0 by creating a single server deployment of ArcGIS Server with both the SOM and SOC on the same server, as well as the web server if desired. The services can be configured and then the server can be cloned. Starting with a 1 or 2 virtual cpu (core) server and a minimum of 4GB of memory, and

organization can deploy ArcGIS Server across multiple load-balanced virtual servers. ArcGIS Server 10.1 is designed for the single server or multiple server deployment as shown in the diagram below.



Virtualization Best Practices

ArcGIS Technology has been increasingly deployed in virtual environments over the last 5 years. A great deal of experience internally at ESRI (e.g. deployment of online training, virtual machines for classroom training, internal configuration and performance testing, etc…) and with customer implementations has resulted in some best practice recommendations. The following recommendations can be found at resources.arcgis.com - http://blogs.esri.com/dev/blogs/arcgisserver/archive/2011/02/17/arcgis-server-virtualization_3a00_-dos-and-don_2700_ts.aspx

- Consider physical host resources. Without exception, the most important consideration when working with virtual machines is to ensure the physical host can provide the resources allocated to the virtualized platform.

- Disk I/O - ArcGIS Server is a disk intensive application, so plan accordingly. Consider using RAID (Redundant Array of Independent Disks) configurations that support both high read/write performance and fault tolerance, such as RAID 5 or 10.

- Do I have to use RAID/SAN solutions if I just want to get started? No.

- Should I use RAID/SAN solutions in production environments? Yes.

- Regarding memory/RAM. Ensure your physical host can support the upper limit of the RAM you intend to use for your virtual machine.

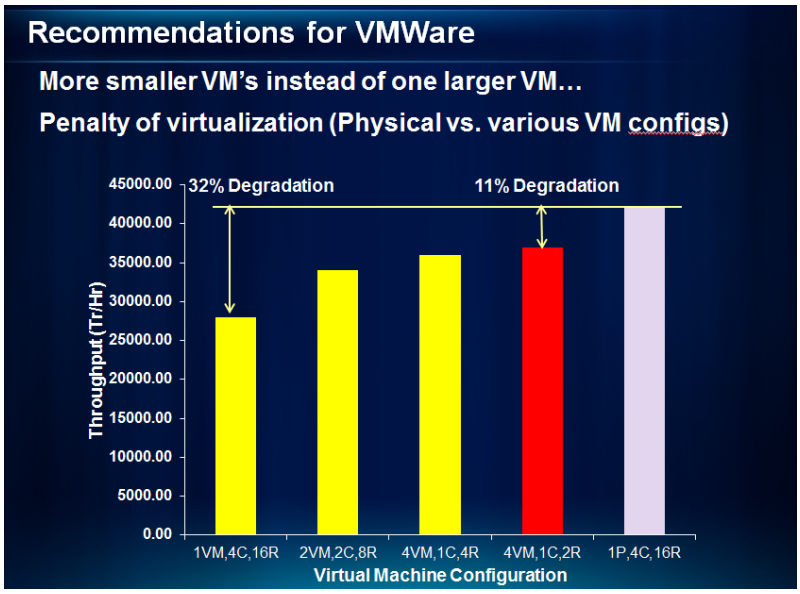
- Regarding CPU - As a general rule, provision 2-3GB of RAM per CPU. For example:

1-2 virtual CPU ("core") | 3-6 GB RAM

- Do: Scale virtualized ArcGIS Servers horizontally - When choosing between one powerful ArcGIS Server (4+ cores / single machine deployment) vs. multiple less powerful ArcGIS Servers (2-cores / multiple machine deployment); multiple 2-core ArcGIS Servers will run more efficiently than a single 4+ core system in a virtualized environment.

- Don't: Run multiple virtual machines on a single physical disk.

Virtualization tests performed by Esri indicate that intensive disk I/O operations, such as dynamic mapping and map caching.



Recent improvement in CPU architectures have improved the processing side of virtual environments, but care should be taken with providing

Some Notes on loading Data having Geodatabase Domains and Behaviors

When an application uses an ETL process via sql it is coming in below ArcObjects and assuming some responsibility to not load data in violation of any coded value domains which have been defined for ArcObjects to maintain. Loading data using SQL is fine, as long as this responsibility is understood.

There are 3 ways to load the data (there are more, but these are the 3 major ones) within ArcGIS Desktop:

- The Append GP tool
- The Simple Data Loader
- Object Loader (would provide for validation of the Coded Value Domains during the import of the data)

Object Loader versus Simple Data Loader

While the Object Loader and Simple Data Loader are similar loading wizards, the Object Loader provides the following functionality that the Simple Data Loader does not have:

- Because the Object Loader loads data during an edit session in ArcMap, once you've finished loading, you can undo the changes if needed.
- If the feature coordinates you're loading are not precisely located, you can choose to honor the current snapping environment, snapping coordinates as they load.
- If you're loading into a feature class that has validation rules, such as attribute domain or geometric network connectivity rules, you can validate the features added and create a selection of the loaded features that are in violation of these rules.
- With the Object Loader, you can load into feature classes in a geometric network, feature classes in a relationship with messaging, or feature classes that have feature-linked annotation. You cannot load into these types of feature classes with the Simple Data Loader.

Training

Consider the following classes:

New 10.1 Courses

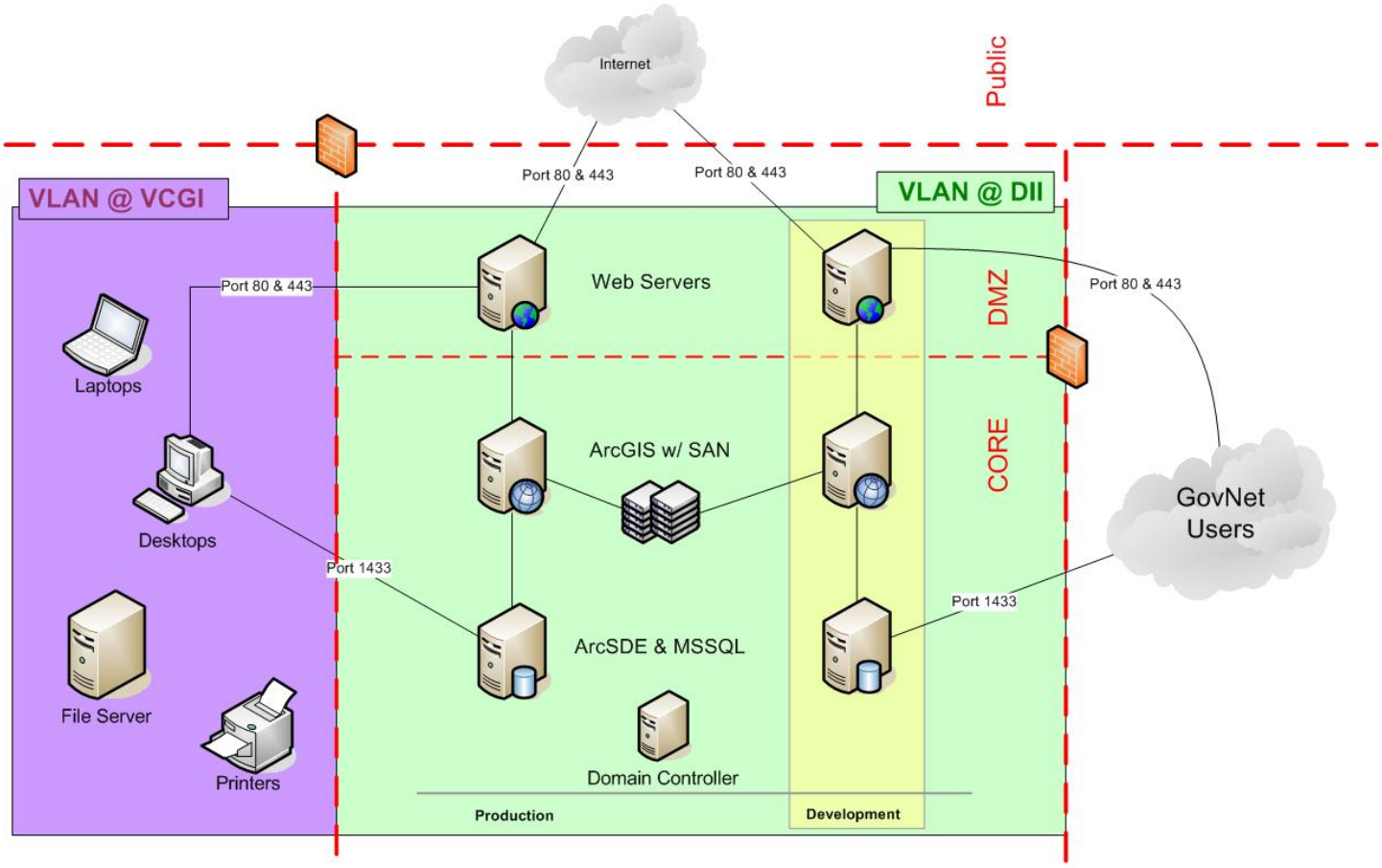
- Migrating to ArcGIS 10.1 for Server
- Implementing Versioned Workflows in a Multiuser Geodatabase
- Future class on Geodatabase Replication

The following help documentation was also referenced in some sections of the note summary:

Desktop 10 Help - http://help.arcgis.com/en/arcgisdesktop/10.0/help/
ArcGIS Server 10 Help - http://help.arcgis.com/en/arcgisserver/10.0/help/arcgis_server_dotnet_help/index.html
Whats new at 10.1 -
http://resources.arcgis.com/en/help/main/10.1/#/A_quick_tour_of_what_s_new_in_ArcGIS_10_1/016w0000002r000000/
ArcGIS Server 10.1 Help -
http://resources.arcgis.com/en/help/main/10.1/#/Welcome_to_the_ArcGIS_10_1_for_Server_Windows_Help/0154000002np000000/

Supporting Notes regarding VCGI's Architecture

Projected VCGI Domain and GIS Server Infrastructure



Note: Port numbers stated above are not exhaustive. Please refer to VCGI's firewall rule config document for details.

DG 3/30/12