



IBM Cognos Forum

Ignite knowledge, ideas, connections

Cognos8 Deployment Best Practices for Performance/Scalability

Barnaby Cole

Practice Lead, Technical Services

Information Management

Cognos.
software

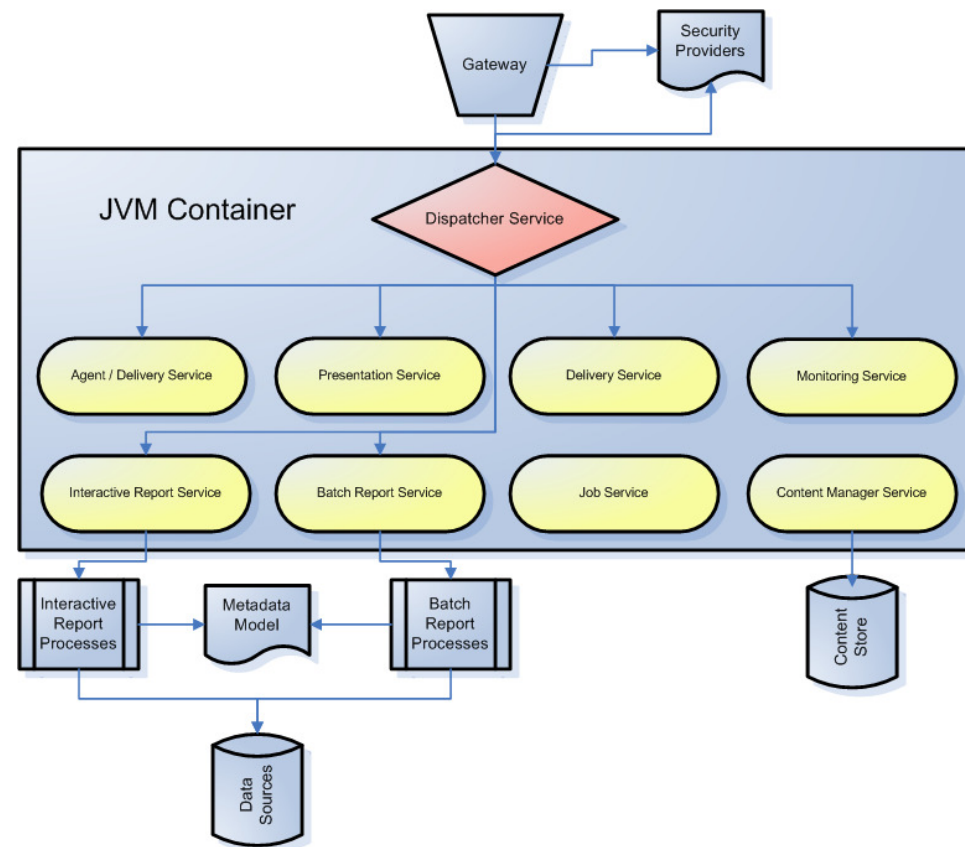
Agenda

- > Cognos 8 Architecture Overview**
- > Cognos 8 Components**
- > Load Balancing**
- > Deployment Options**
- > Cognos 8 Report Processing**
- > Basic Tuning Suggestions**
- > System Monitoring**

Architecture Overview

> Cognos 8 is made of 5 major install components

- Application Tier
- Gateway
- Content Manager
- Cognos Content Database (optional)
- Framework Manager

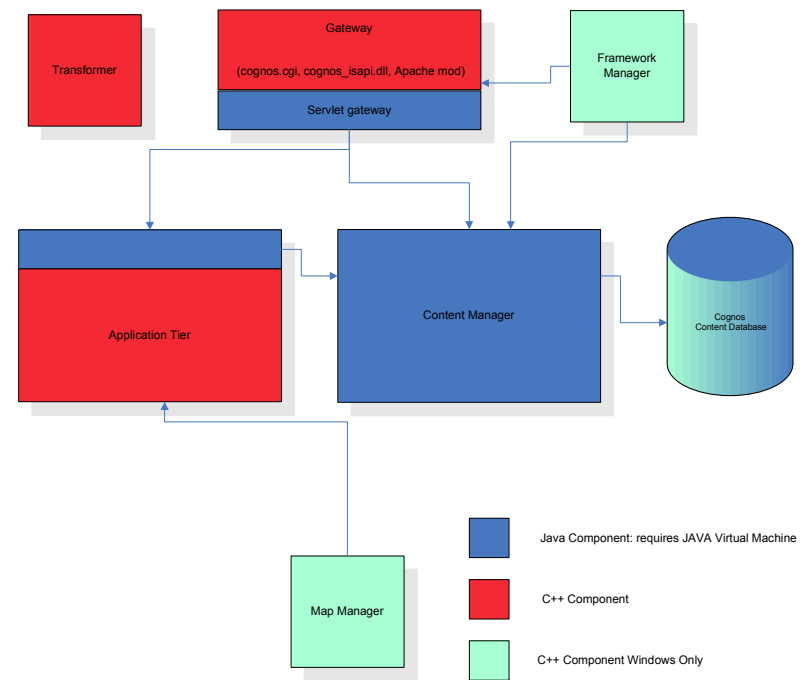


- Cognos 8 Business Intelligence Server
 - Application Tier Components
 - Gateway
 - Content Manager
 - Cognos Content Database

- Cognos 8 Business Intelligence Modeling
 - Framework Manager
 - Map Manager

Basic Cognos 8 Components

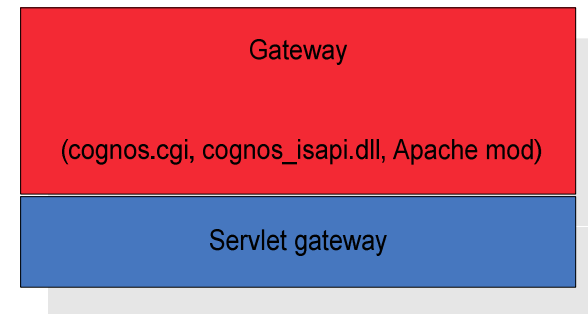
- > **Web communication in Cognos 8 is typically via Gateways.**
 - Gateways typically reside on Web Servers (IIS or an implementation of Apache)
- > **Content Manager**
 - Manages the storage of application data
 - Communicates with the Content Store Database directly by JDBC
 - Supports failover
- > **Dispatcher**
 - Provide application tier communications
 - Dispatcher for Gateway
 - Peer to peer application tier server communications
 - Request routing and load balancing
 - Java based services
 - Self-registering with Content Manager



- Cognos 8 is a mixture of c++ and java components
- Some components require the installation of Application Servers

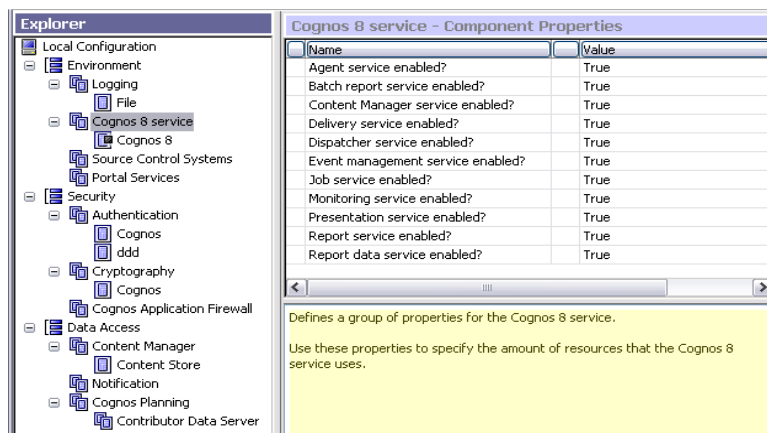
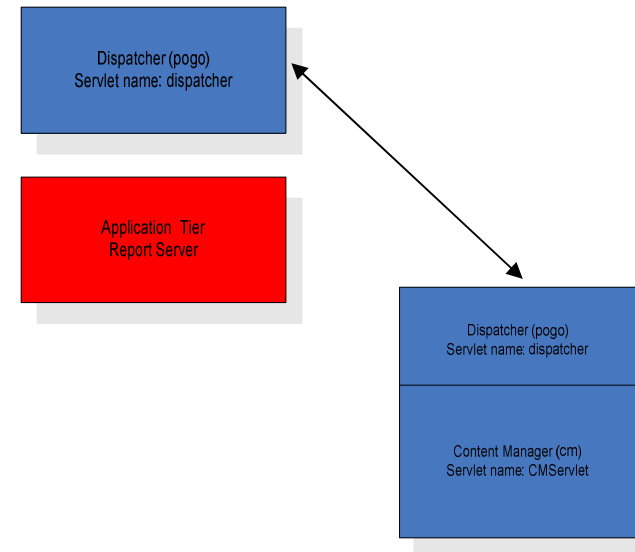
Gateways

- > Web communication in Cognos 8 is typically through a gateway, which resides on one or more Web servers.
- > A gateway is an extension of a Web server program that transfers information from the Web server to another server.
- > Four possible types:
 - CGI (common gateway interface)
 - ISAPI
 - apache_mod (Apache 1.x & Apache 2)
 - servlet
- > Enable HTTP compression and Content Caching for performance



Cognos 8 Dispatcher (Java Servlet)

- > Available on each installed instance by default (except gateway)
- > Starts all Cognos 8 services configured and enabled on an install
- > Routes requests to a local service, such as: report, presentation, job, or monitoring service etc



- > Routes requests to a remote dispatcher to run a given request.
- > Routes requests to specific dispatchers based on load-balancing needs
- > If a dispatcher fails or is unavailable, requests for that dispatcher are routed to the next available dispatcher.

Scaling C8 – Horizontally or Vertically

Horizontally

Pros:

- > Better failover for hardware issues
- > Higher performance with multiple OS instances and dedicated hardware
- > Added flexibility for with C8 routing sets

Cons:

- > Higher administration overhead

Best For:

- > Entry level to mid-sized servers (2-8 CPUs)

Vertically

Pros:

- > Lower administration overhead

Cons:

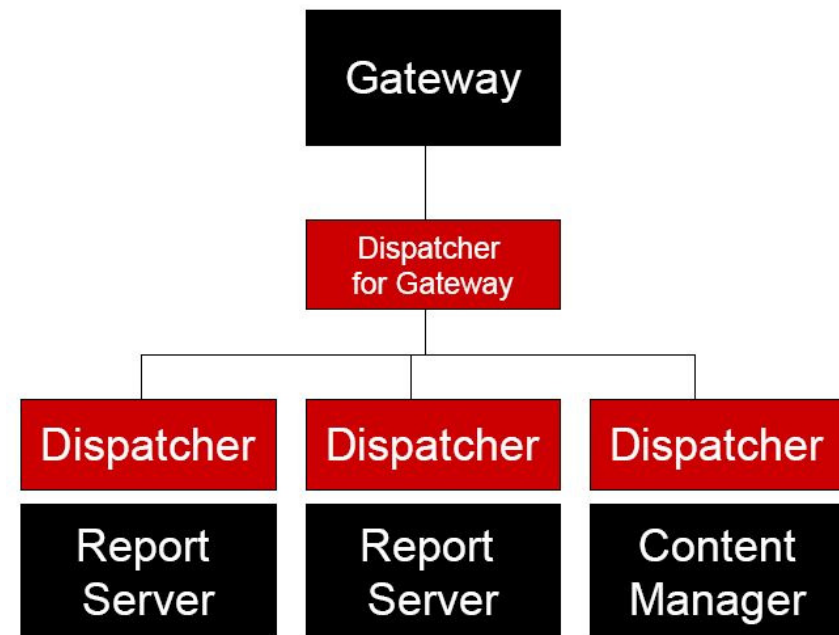
- > When the server goes down, the whole solution is down
- > Single OS instance may introduce extra overhead, JVM addressing space limitations

Best For:

- > High-end servers (8+ CPUs)

Dispatcher Load Balancing

- > Dispatcher is a multithreaded Java process
- > Requests can be routed to specific dispatchers based on load-balancing needs
- > Load Balancing is based on a weighted round-robin algorithm
- > Routing is configurable by Administrators
- > Processing is configurable by Administrators

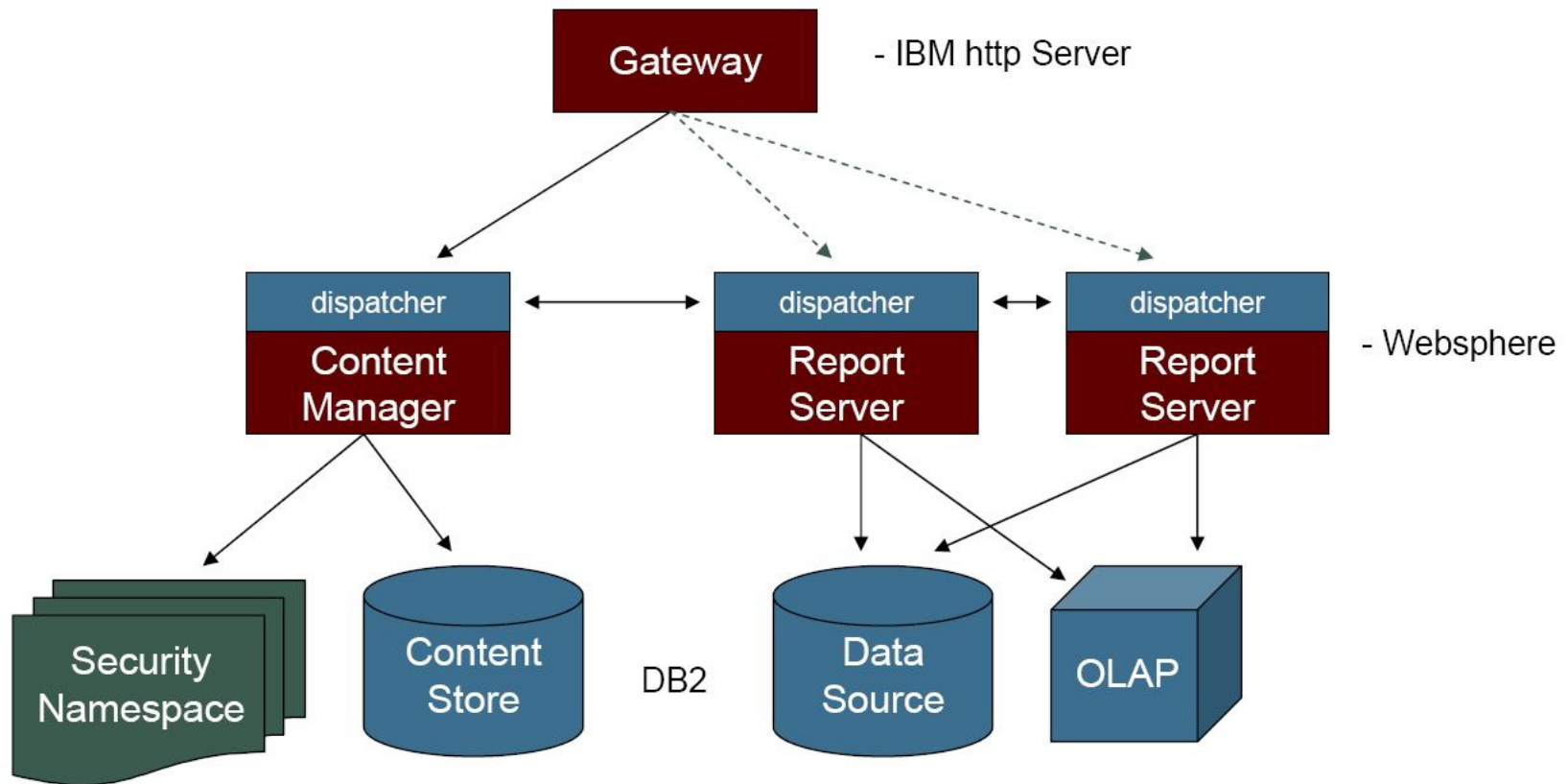


Topology Requirements

- > Topologies for shared infrastructure can be driven by two overall requirements**
 - **Pure Combined Load**
 - Assume all users are the same (same user needs)
 - Deploy hardware to meet overall utilization & performance requirements

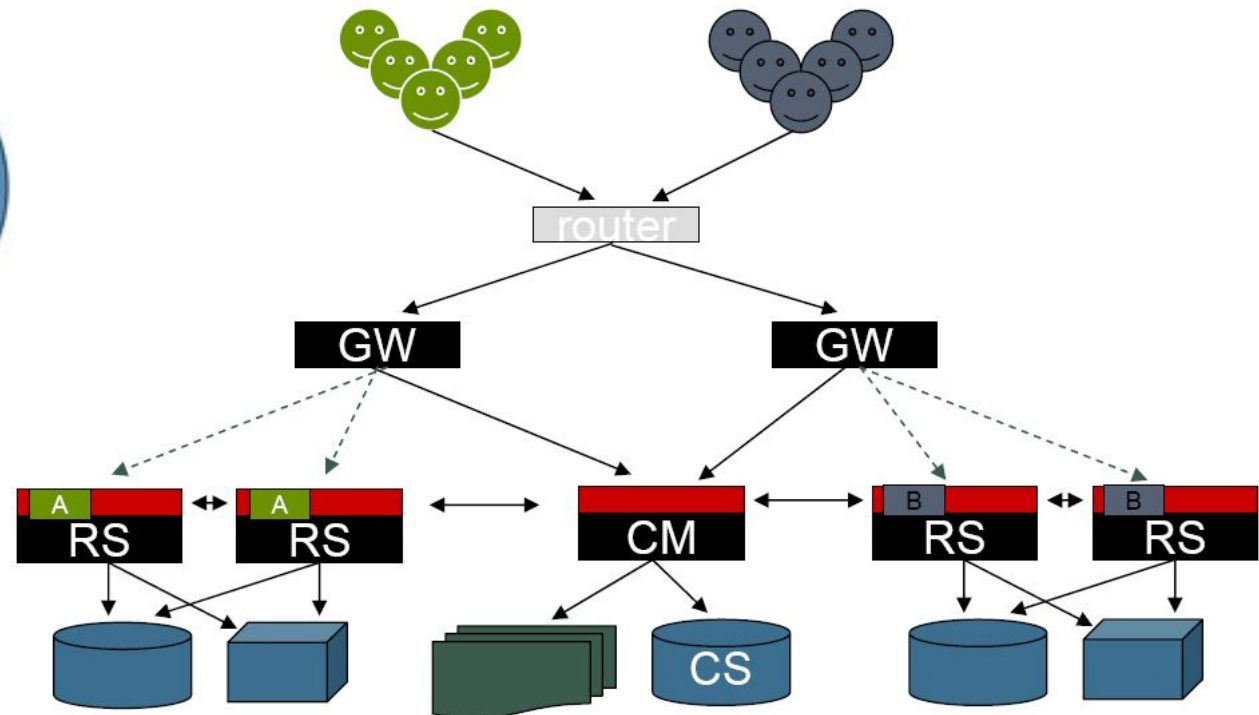
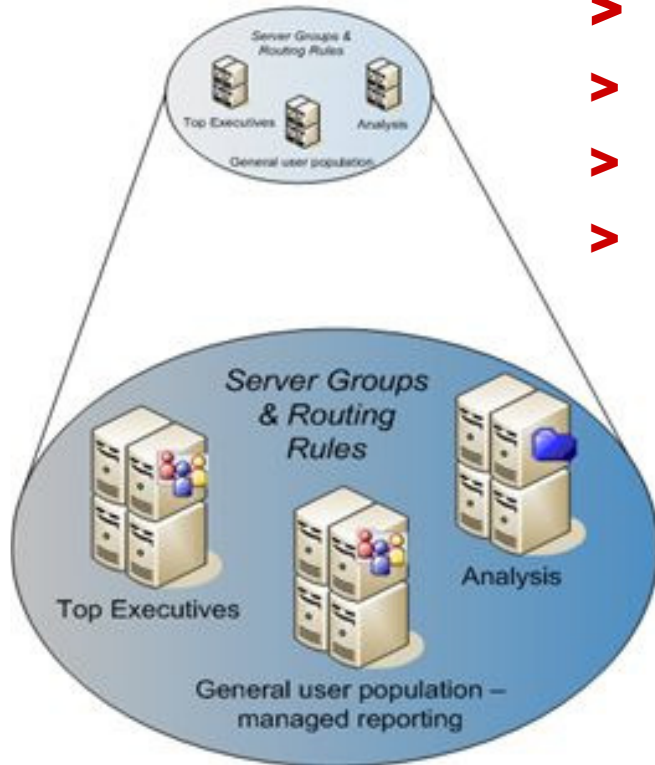
 - **Load Segregation**
 - Assume that load must be managed by user group
 - Deploy hardware to meet the requirements of individual groups of users (Business Units for example)

Scenarios –Single Platform



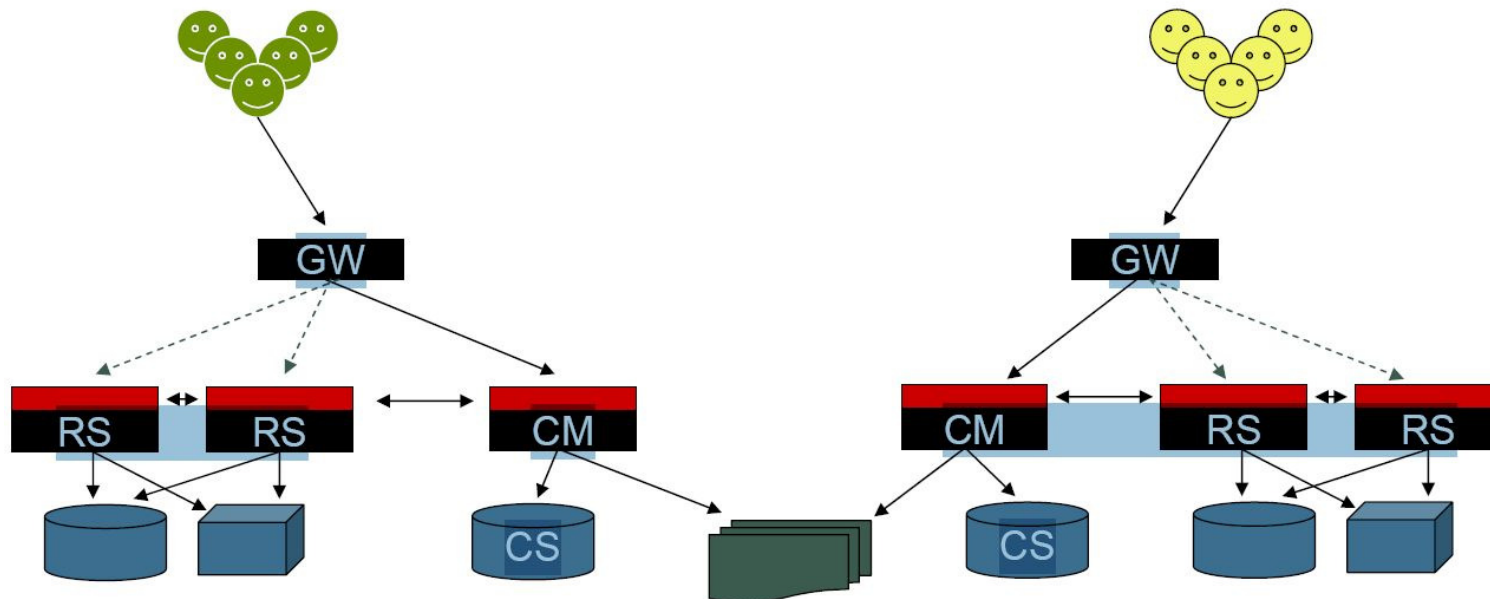
Scenarios –Single Platform + Server Groups

- > Dispatchers can be arranged in Server Groups
- > Load can only be balanced within a server group
- > The Content Manager is universally reachable.
- > Shared gateways



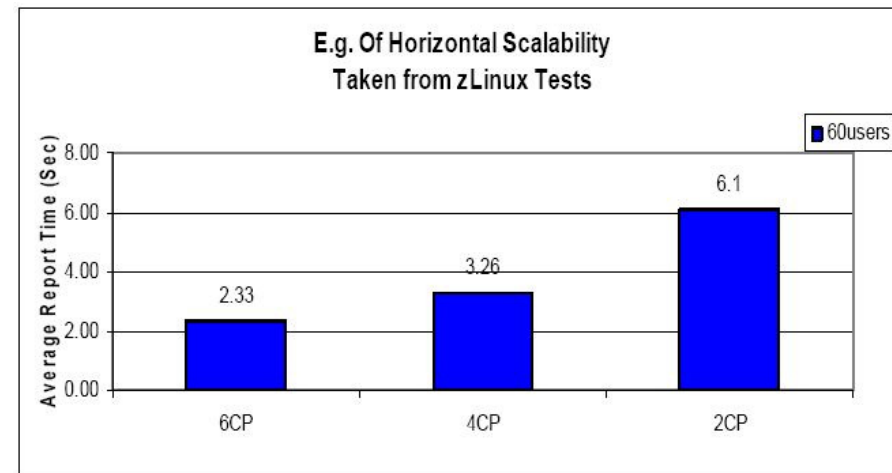
Scenarios – Multiple Platforms

- Load can only be balanced within a dedicated environment
- Each user community has their own Content Manager
- Each user community has their own Gateway



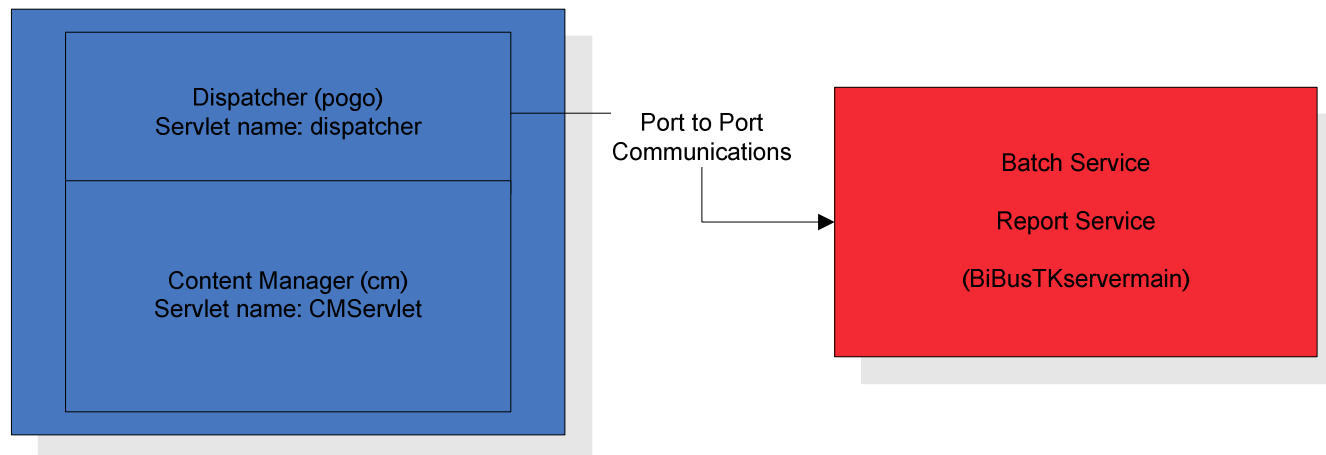
Scalability Characteristics

- > Our Cognos8 products are tested to evaluate the horizontal, vertical scalability characteristics
- > By scaling out our Report Server/Presentation layer and if need be web tier we have demonstrated very predictable horizontal and vertical scalability profiles.
- > One would consider scaling out if response times are not meeting expectations and system resource are heavily utilized or growth in user load is expected.



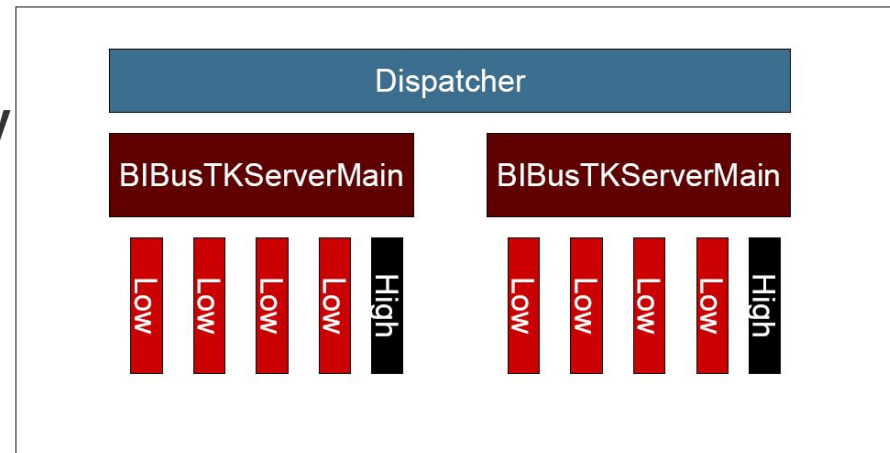
Cognos 8 Report Processing (BIBUSTKSERVERMAIN)

- Report Server Process (Batch and Interactive)
- Spawned by Dispatcher (child of the java process)
- Controlled by Administrator configuration settings
- Processes report requests
- Maintains connection pools



Threads = Units of Work (Affinity Connections)

- > The dispatcher manages a queue of requests
- > A queue develops if there are no affinity connections available for processing
- > The pools increase respectively
 - Default low=4: a new BiBus is spawned on every 2 long running reports
- > Queuing can be caused by
 - Time intensive requests (i.e. long running database queries)
 - Capacity and process allocation across dispatchers is not tuned appropriately





<input type="checkbox"/>	Tuning	Number of high affinity connections for the batch report service	<input type="text" value="1"/>
<input type="checkbox"/>	Tuning	Number of low affinity connections for the batch report service	<input type="text" value="2"/>
<input type="checkbox"/>	Tuning	Maximum number of processes for the batch report service	<input type="text" value="2"/>

The setting of High and Low affinity connections is set on the **Server Administration Configuration** page in **Cognos Connection**.

Dispatcher Request Queue

- > Queue time is called 'request latency'
 - This can be monitored through Cognos Connection

<input type="checkbox"/>	...> Name ▲	Service	Status ⇅	Processes	Latency	Seconds per request	Requests per minute	Update interval	Actions
<input type="checkbox"/>	...> http://wottcs-taylorclp:9300/p2pd > ReportService	Report service		1/2	0/240	0	0	--	 More...

Tuning Suggestions

- > Common Performance Questions**
 - What can be tweaked to improve performance?
 - When should I scale up and/or out?
 - How to build fail over and redundancy?
 - What Logging is Required?

Java Tuning

- > Setting `jvmheap` to 768 MB or 1024MB will often improve performance

- > Websphere: Increase the Thread Pool `WebContainermax` from default 50 to 500 in your `jvm` will vastly improve scalability
 - WAS 6.1: `com.ibm.ws.webcontainer.channelwritetype=sync` will eliminate a large memory growth due to how it handles large SOAP bodies
 - On 64bit increase Min and Max heap settings quite high to minimize garbage collection improves pure java call significantly

- > Typically run with the latest and greatest `jre` for you `jvm`, often many perceived leaks can be avoided

Java Process – Memory management

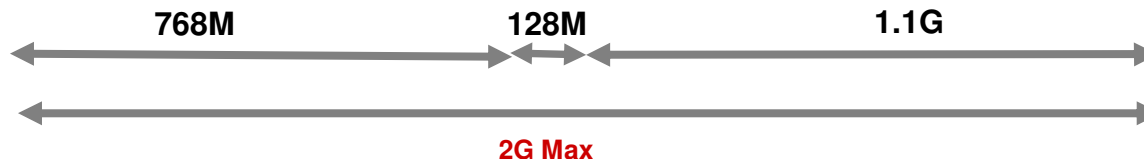
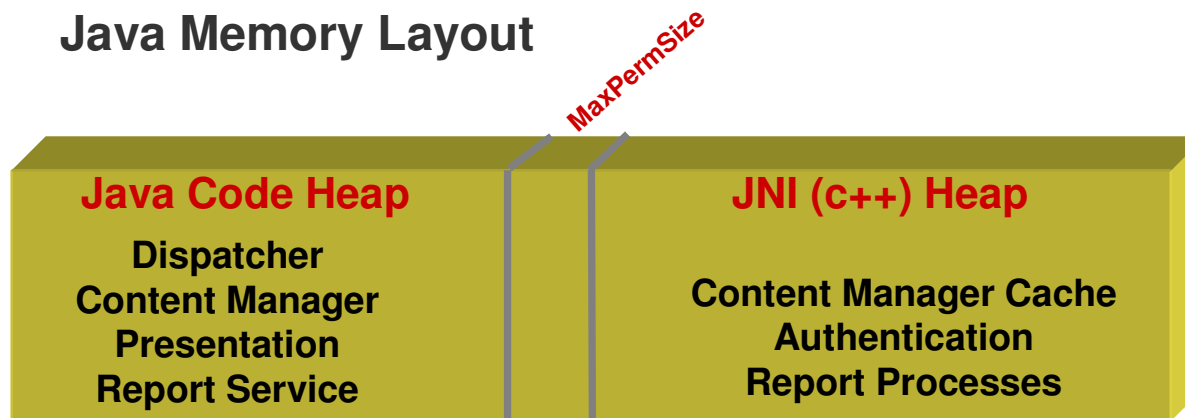
Java Memory



Note: Java Heap set to 768M (for this example)
 JNI = Java Native Interface
 MaxPerm = Java classes and method objects



Java Memory Layout



Windows Suggestions

- > Connections are established, used and closed quickly by Cognos8 hence some slight modifications to the windows registry setting can optimize performance and scalability**
 - Update the TCP/IP Registry settings: The MaxUserPort must be set to 65534 and TcpTimedWaitDelay must be lowered to 30 on every server**

Batch Report Tuning

- > Maximum Jobs During Non-Peak Period**
 - Based on Cognos 8 testing, it may be advantageous to reset the "maximum" number of jobs to "maximum number of batch report service processes" * "number of low affinity connections (Batch Report service)"

- > Maximum Jobs During Peak Period**
 - This setting is important to relate batch and scheduled activities to interactive usage. If scheduled activities are performed during non-peak periods, then this setting can be skipped.

- > For large overnight batch runs, create a master job to spawn sequential multiple jobs that equal the number of Batch Report Services**

Report Tuning

- > **Maximum Execution Time**

- > **Increase the “Soft Buffer Size”**
 - Memory allocated for every process that requires local sorting

- > **Specify Queue Limit Times**

- > **PDF File Settings**
 - ‘VirtualMemoryDiagnostics’
 - PDF Character Encoding
 - Font Embedding
 - Content Compression Type
 - Compression Algorithm Level

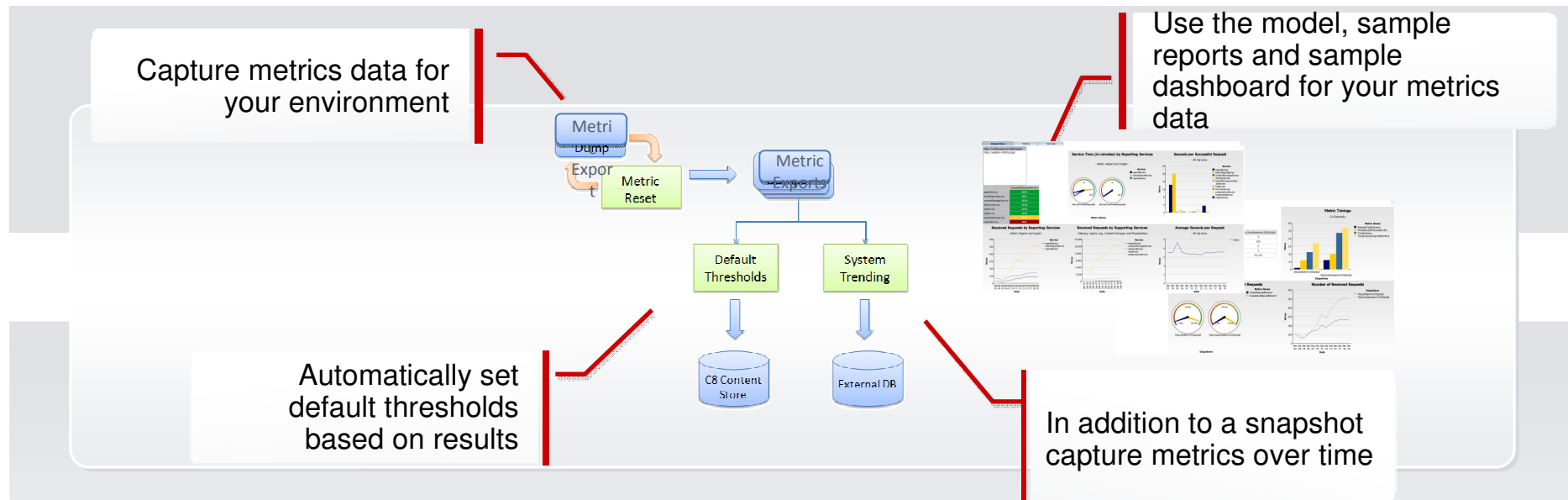
File System

- > Move temporary files location to a separate physical disk, preferably a high speed SAN.**
 - This maximizes parallel disk access and avoids the unnecessary sequential access that is common when only a single disk device is used.**

- > Reduce the logging levels unless troubleshooting**

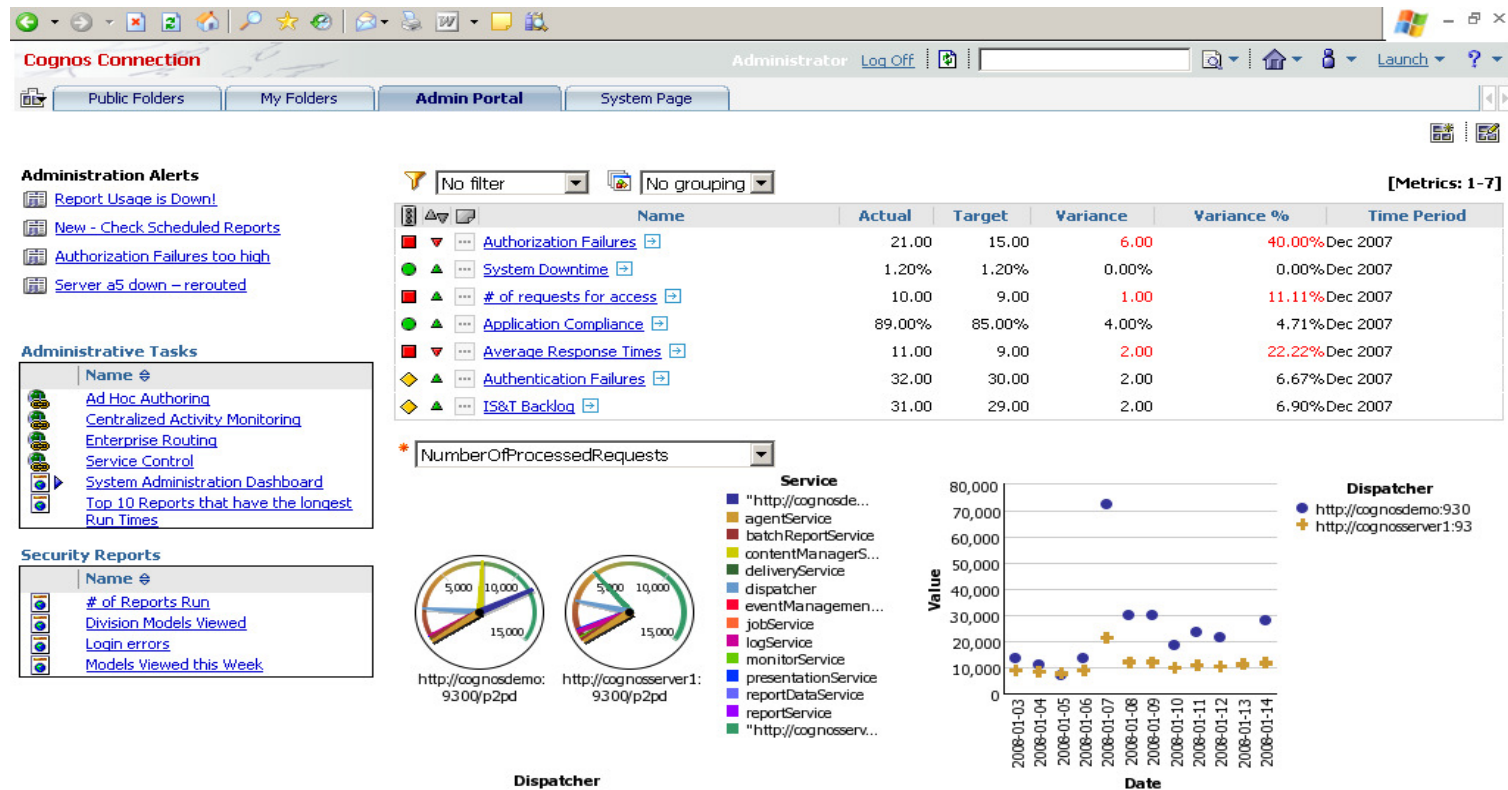
- > Ensure Virus Scanning is not degrading performance**

Cognos8 Environment Monitoring



- > Administrators can quickly leverage thresholds to proactively identify issues before business impact
- > Administrators are informed of system management trends to optimize their deployment environment over time.
- > SMM = Proven Practices, sample reports, Framework Manager models

Cognos8 system thresholds via dashboard



Summary

- > Your deployment is based upon your business needs often with cost constraints**
- > Report Server is the easiest tier to tune and often with best results**
- > There are some environment modifications per operating system that may affect performance and scalability**
- > Scaling out usually are determined by a balance of response times, system resources and user loads**

Cognos Professional Services – The Enablers



When it comes time to implement your software project, Cognos Consulting offers three levels of service based on the requirements of your business.

Project Services

Comprehensive services that encompass the entire implementation from beginning to end

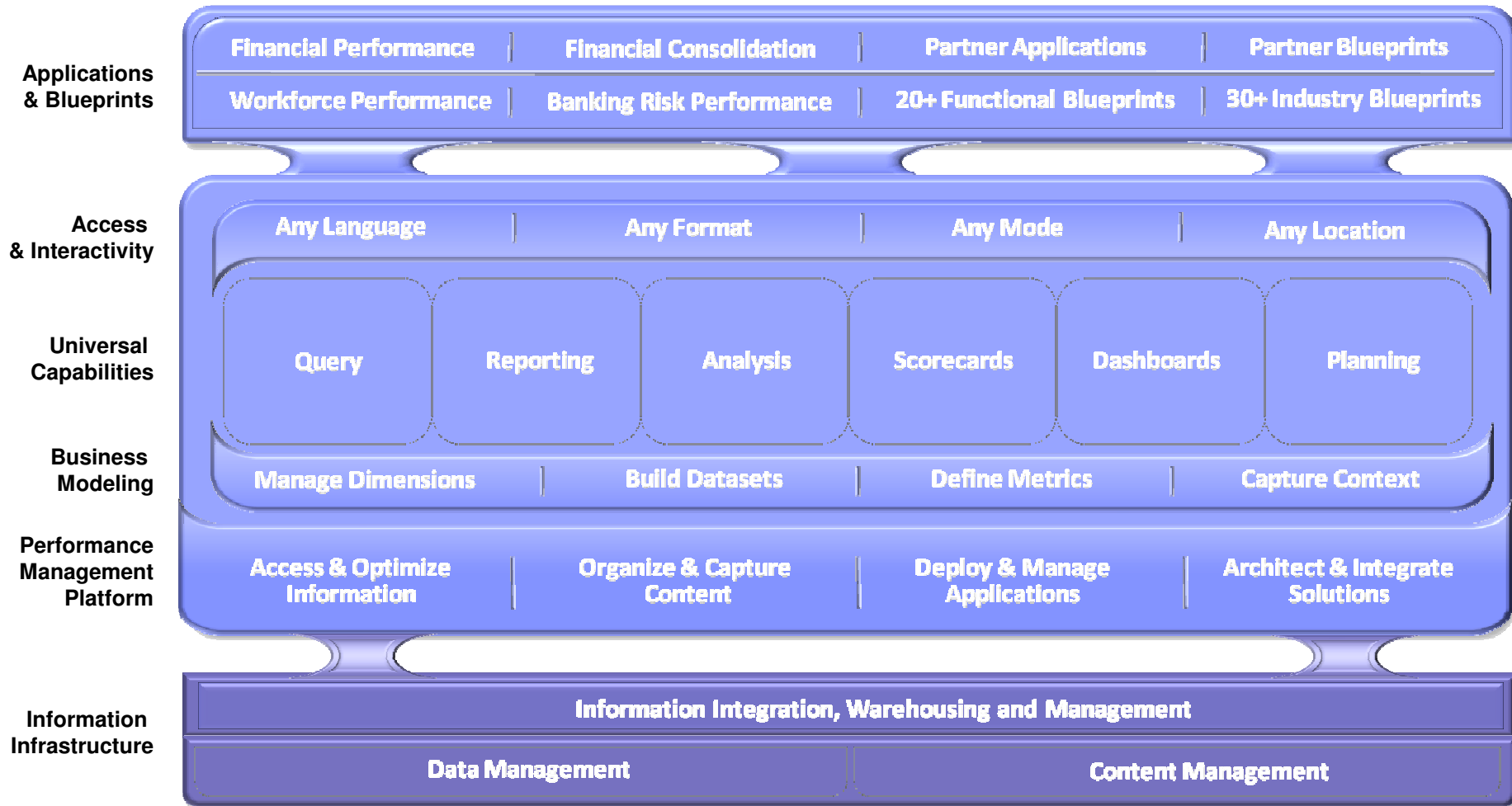
Guardian Services

When a **PARTNER** or the Customer's internal team is leading the project

Assist Services

When you want to own your software project for its entire lifecycle, leverage specific Cognos expertise at the right time to ensure success

Enablers of IBM Cognos Performance Management System



IBM Cognos Software Group Service – Enablement Offerings

- > **Project Services**
 - Cognos Series 7 (Reportnet, Powerplay, Impromptu)
 - Cognos 8 (Enterprise Planning, Business Intelligence, Data Warehousing, Project Management)
- > **Review Service (Health Checks, Application Review)**
- > **Specialist Services (Technical Architects)**
- > **Guardian Service (Mentoring & Best Practice)**
- > **Migration Service**
- > **Conversion Service**
- > **Off Shore Services**
- > **End User Adoption Service**
- > **Business Intelligence Competency Centres**
- > **Education Services**

Evaluation Forms

SHARE YOUR FEEDBACK AND WIN

Visit IBM Cognos Central to fill out your session evaluations online.

- > **Each completed evaluation** qualifies you to win one of five \$100 daily prizes.
- > **Complete evaluations for every session you attend** and qualify to win an additional \$500! An overall conference survey will be available at IBM Cognos Central on Friday morning, and will also be emailed to you.