The Art of Separation of Concerns

Introduction

In software engineering, Separation of Concerns refers to the establishment of logical boundaries to define and delineate purpose. This improves the ability to manage and maintain systems by providing order and elegance to the form and function of system design.

The goal of this article is to promote the understanding of Separation of Concerns and to establish a set of foundational concepts useful to developers and architects toward its application.

Defining Separation of Concerns

Separation of Concerns is both principle and process. The principle of Separation of Concerns might be stated as the premise that things should be no more and no less than what they should be. That is to say, things should contain the essential attributes and behaviors inherent to their nature, but should be void of attributes and behaviors not inherent to their nature. The process of Separation of Concerns is comprised of techniques and constructs for constraining concerns within boundaries. Boundaries are often established through the use of methods, objects, components, and services to define core behavior within an application; projects, solutions, and folder hierarchies for source organization; application layers and tiers for processing organization; and
versioned libraries and installers for product release organization.

Separation of Concerns is not merely the division of systems into multiple parts, but rather the establishing of boundaries based on essence. The distinction here is that unguided division leads to chaos, while guided division leads to order. Separation of Concerns involves both division and cohesion. The goal of Separation of Concerns is the establishment of boundaries which contain inherent properties and processes, but are void of non-inherent properties and processes. Arbitrary divisions, such as the separation of processes based solely on number of lines of code, or excessive division, such as separating a task into the smallest possible steps without tangible benefits or consideration of the task as a whole, may actually be detrimental to the maintainability of a system. As Albert Einstein stated, "Make everything as simple as possible, but not simpler."

At its essence, Separation of Concerns is about order. The overall goal of Separation of Concerns is to establish a well organized system where each part fulfills a meaningful and intuitive role, while maximizing its ability to adapt to change. The achievement of separation of concerns is a goal guided more by principle than rules. Because of this, Separation of Concerns might be considered as much an art as it is a science.

**The Value of Separation**

Applying the principle of Separation of Concerns can increase the stability, maintainability, extensibility, reusability, adoption, and even the logical and aesthetic design and flow of a system. Many of these benefits also play a supporting role in the accomplishment of one another. For instance, strategies which facilitate the reusability of components within a system, such as ensuring singularity of purpose and decoupling, can have a correlation to the system’s maintainability. The maintainability of a system often has a direct correlation to its stability since process failure is often the result of fragile, overly complex, and convoluted designs. The logical and aesthetic design and flow of a system has a correlation to the system’s maintainability and adoption through its promotion of understandable and harmonious processes. The system’s maintainability and aesthetic design has a correlation to the facilitation and understanding of its extensibility which reciprocates in its contributions back to the system’s maintainability and stability. Though not all of these benefits may be specific goals of every endeavor, each can prove instrumental in the achievement of the overall goal of manageability.

When applied to development organizations, well conceived separation of concerns can result in more agile development by minimizing the coordination necessary between the teams and maximizing the potential of each team’s ability to deliver features pertinent to their center of competency. This is particularly true of groups responsible for common concerns such as overall infrastructure, service development, and centralized user interface needs due to their increased potential for domain specific encroachment.
The principle of Separation of Concerns also greatly enhances problem resolution in enterprise wide systems. When proper delineation of responsibilities is allocated across an enterprise, it becomes easier to identify a problem’s origin.

The principle of Separation of Concerns extends into almost every facet of software development. It is the foundational principle for good architecture and system management.

Terms of Separation

Discussions pertaining to Separation of Concerns require an understanding of key concepts and definitions. The following is a list of some of the terms related to the subject of Separation of Concerns.

**Concern** – An inherent or imposed characteristic. Concerns are the attributes and behaviors by which a thing may be defined.

**Boundary** – A physical or conceptual delineation around one or more related concerns. A boundary may be a readily perceivable construct which outlines the applicable scope of a set of concerns, as is the case with methods or component libraries, or it may merely be an agreed upon organization, such as the grouping of members based on scope accessibility. Motivations for the use of boundaries include defining entities according to their base ontological essence, aiding in the elimination of redundancy, delineating roles and relationships, correlating related concerns, and improving portability and reuse.

**Abstraction** – the representation of an entity in non-concrete terms. As within the study of philosophy, the process of abstraction within computer science allows one to deal with ideas and entities without being encumbered by unnecessary details. Abstraction is generally accomplished either through the use of language specific constructs, such as abstract classes and interfaces, or with conceptual layers of functionality which expose only the necessary elements of a lower level set of components. The goal of abstraction is generally to allow one or both sides of a relationship to vary independently from the other. For example, if a presentation layer accesses a database through a domain layer serving as the layer of abstraction, the database and method of access can be modified without affecting the presentation layer.

**Decoupling** – the process of disconnecting an association or awareness of two components or systems.

**Loose Coupling** – the state of possessing an association or awareness from one component or system to another by means of abstraction. Loose coupling provides the ability for one component to be associated with another while allowing independent variation. For example, if component A interacts with component B through an abstraction of component B, the implementation of component B can be changed without affecting component A.
**Tight Coupling** – the state of possessing a direct association or awareness between one component or system and another which affects the ability for the associating component or system to vary independently.

**One-way decoupling** – the process of disconnecting an association between two components or systems allowing only one to vary independently from the other. An example of one-way decoupling would be to use the Observer Pattern to enable a domain model object to raise event notifications to an external observer, with the external observer calling methods on the domain model once notification is received. In this scenario, the domain model is decoupled from the observer receiving the message, while the observer is coupled to the domain model.

**Two-way decoupling** – the process of disconnecting an association between two components or systems allowing both to vary independently. An example of two-way loose coupling would be two components which interact with one another through interfaces specific to each component. An example of complete two-way decoupling would be the use of a brokering mechanism which maintains the relationship between two or more components which exchange messages. In this case the components are coupled with an infrastructure level component which allows them to remain decoupled from other domain level components.

**Monolithic** – the attribute of lacking clear or any component separation. Monolithic applications often result in large, unwieldy, rigid systems and contain functionality which cannot easily be shared with other applications outside of the use of the entire application.

**Modular** – the attribute of having clear component separation. Modular applications are comprised of multiple components or sub-systems which work together to achieve the overall system goals.

**Add-In (or Plug-in)** – a component which adds new capabilities to systems designed for the consumption of dynamic or configurable modules. Add-Ins provide the ability for systems to delegate concerns using forms of loose-coupling.

**Provider** – a component which satisfies a contract for a specific set of abstract functionality. Providers are specialized Add-Ins which accomplish a specific set of concerns within an application.

**Extension** – a component which enhances the existing capabilities of a system designed to yield control or expose data and functionality at predetermined points within its internal processes. An extension allows for the delegation of concerns not directly inherent to, or practical for the core system’s processes. Extensions often take the form of a hosted add-in, but include any process designed specifically to add new capabilities to an existing system.

**Layered** – the attribute of having clear separation between groups of entities with
common processing goals within a system. Examples of layered separation would be the division between components concerned with the visual presentation of an application and components concerned with the data access of the application.

**Inversion of Control** – the process of moving a concern outside of an established boundary. Inversion of Control externalizes concerns formerly contained within a component and may be used for tasks such as minimizing non-essential concerns, or decoupling components from direct participation in specific abstraction strategies. An example of Inversion of Control would be externalizing the main control flow of a component.

**Dependency Injection** – the process of separating the concerns of acquiring dependencies from an entity. Dependency Injection is a specific instance of Inversion of Control where the process being inverted is the acquisition of external dependencies. An example of Dependency Injection would be a system which initializes references to a logging service for each instantiated component within the system. This would allow the system to vary the implementation assigned to all, or a subset of components within the system without the need of modifying the consuming components.

**Dependency Direction** – the route of association between one component or system to another. The dependency direction follows the path drawn from a consuming component or system to a consumed dependency. For example, if component A calls a method on component B then the dependency direction would be from A to B.

**Refactoring** – the process of changing a system such that its external interfaces and functionality remain the same, but the internal organization and processes are made more efficient and maintainable.

**Horizontal Separation**

Horizontal Separation of Concerns refers to the process of dividing an application into logical layers of functionally that fulfill the same role within the application.

One common division for graphical user interface applications is the separation of processes into the layers of Presentation, Business, and Resource Access. These categories encompass the main types of concerns for most application needs, and represent an organization of concerns which minimizes the level of dependencies within an application. Figure 1 depicts a typical three-layered application:
The Presentation Layer encompasses processes and components related to an application’s user interface. This includes components which define the visual display of an application, and may include advanced design concepts such as controllers, presenters, or a presentation model. The primary goal of the Presentation Layer is to encapsulate all user interface concerns in order to allow the application domain to be varied independently. The Presentation Layer should include all components and processes exclusively related to the visual display needs of an application, and should exclude all other components and processes. This allows other layers within the application to vary independently from its display concerns.

The Business Layer encompasses processes and components related to the application domain. This includes components which define the object model, govern business logic, and control the workflow of the system. The business layer may be represented through specialized components which represent the workflow, business processes, and entities used within the application, or through a traditional object-oriented domain model which encapsulates both data and behavior. The primary goal of the Business Layer is to encapsulate the core business concerns of an application exclusive of how data and behavior is exposed, or how data is specifically obtained. The Business Layer should include all components and processes exclusively related to the business domain of the application, and should exclude all other components and processes.

The Resource Access Layer encompasses processes and components related to the access of external information. This includes components which interface with a local data store or remote service. The goal of the Resource Access Layer is to provide a layer of abstraction around the details specific to data access. This includes tasks such as the establishing of database and service connections, maintaining knowledge about
database schemas or stored procedures, knowledge about service protocols, and the marshalling of data between service entities and business entities. The Resource Access Layer should include all components and processes exclusively related to accessing data external to the system, and should exclude all other components and processes.

Another common division used by Service Oriented Applications is the division of the application into the layers of Service Interface, Business, and Resource Access as depicted in Figure 2:

In this division, the Business and Resource Access layers serve the same purposes as previously discussed, with the service exposure concerns encapsulated into a Service Interface Layer. This layer encompasses service interface concerns such as the exposure of business processes through various protocols and the management of service specific contracts and data types. The primary goal of the Service Interface Layer is to encapsulate all service interface concerns in order to allow the application domain to be varied independently. The Service Interface Layer should include all components and processes exclusively related to the exposure of the application as a service, and should exclude all other components and processes.

By grouping processing concerns based on their role within the application, a number of benefits are gained which improve the overall system manageability. These benefits include ease of maintenance through consistent architecture and isolation of process, increased insulation from change impact, increased adaptability to change, and increased potential for reuse.
Vertical Separation

Vertical Separation of Concerns refers to the process of dividing an application into modules of functionality that relate to the same feature or sub-system within an application. Vertical separation divides the features of an application holistically, associating any interface, business, and resource access concerns within a single boundary. Figure 3 depicts an application separated into three modules:

![Figure 3](image)

Separating the features of an application into modules clarifies the responsibilities and dependencies of each feature which can aid in testing and overall maintenance. Boundaries may be defined logically to aid in organization, or physically to enable independent development and maintenance.

Logical boundaries imply the existence of modularity, though the methods used to denote separation may have no bearing on the actual deployment or runtime behavior of an application. This can be useful for improving the maintainability of an application as well easing any future efforts for physically separating features. Figure 4 depicts an application containing logical boundaries:
Physical boundaries are generally used in the context of developing add-ins or composite applications, and can enable features to be managed by disparate development teams. Applications supporting add-in modules often employ techniques such as auto-discovery, or initializing modules based on an external configuration source. Figure 5 depicts a hosting framework containing modules developed by separate development teams:
While vertical separation groups a set of concerns based on their relevance to the total fulfillment of a specific feature within an application, this does not preclude the use of other separation of concerns strategies. For example, each module may itself be designed using layers to delineate the role of components within the module. Figure 6 depicts an application using both horizontal and vertical separation of concerns strategies:

![Figure 6](image)

**Aspect Separation**

Aspect Separation of Concerns, better known as Aspect-Oriented Programming, refers to the process of segregating an application’s cross-cutting concerns from its core concerns. Cross-cutting concerns, or aspects, are concerns which are interspersed across multiple boundaries within an application. Logging is one example of an activity performed across many system components.

Figure 7 depicts an application with several cross-cutting concerns:
Cross-cutting concerns can be difficult to maintain due to their widespread placement throughout an application. Their mixture with core concerns also adds complexity, rendering the application more difficult to maintain. By separating these concerns, both core concerns and cross-cutting concerns are made easier to manage.

Aspect separation differs from other separation techniques in that it relies on a pre-processing, compile time, or run-time fusion of cross-cutting concerns with the application. The process of fusing the two concerns back together is referred to as “weaving”. By using various strategies, cross-cutting concerns can be separated and woven back into the application for run time cohesion.

Figure 8 depicts a conceptual weaving process where separately maintained cross-cutting concerns are joined with core concerns:
Aspect separation tools are available for a wide range of programming languages. For more information on this topic including a list of available tools for facilitating aspect separation, see the Wikipedia Web page for Aspect-Oriented Programming.

**Dependency Direction**

One characteristic of good Separation of Concerns is the ideal establishment of dependency direction. An ideal dependency direction establishes the roles of consumer and dependency such that the role of dependency is occupied by the entity possessing the highest potential for reuse.

A simple example illustrating the concept of dependency direction is the common relationship between a business component and a utility component. Consider a system which provides an order inquiry process which requires that frequently requested data be cached for greater efficiency. To facilitate the caching of order inquiries, a caching utility component may be developed to separate the concerns of caching from the rest of the order inquiry process. Figure 9 depicts this system with its two components:
Because the function of caching is a more generic behavior than an order inquiry process, the caching component has the higher potential for reuse among the two. Therefore, the best dependency direction in this case would be from the business component to the utility component. This dependency is depicted in Figure 10:

Such a relationship allows the caching component to remain agnostic of the business inquiry process enabling it to be reused by future processes.

In contrast, if the caching component is design to depend specifically on the type of data being cached, or on some other aspect of the business component, then the caching component would be incapable of being reused by future processes without modification. Such a relationship is depicted in Figure 11:
While extending this relationship to other components would be possible, this approach would eventually result in a caching component which is overly complex and unwieldy.

Another example demonstrating the concept of dependency direction is the relationship between layers within an application. Figure 12 depicts an application which separates the concerns of presentation, business processing, and resource access into separate layers:

As discussed in the section on Horizontal Separation, these layers are a common pattern of separation because they represent an organization of concerns which minimizes the level of dependencies within an application. By ensuring that each layer only establishes
dependencies on layers for which it possesses no mutual relationship, either directly or indirectly, each layer can be isolated from changes to direct and indirect consuming layers. Figure 13 depicts this prescribed relationship:

![Diagram](Image)

Figure 13

This improves the application’s ability to adapt to change and enables layers to be reused within a different context (e.g. multiple presentations of the same business processes).

In contrast, a violation of this prescription results in layers which are less insulated to change and less capable of being reused. Figure 14 depicts the resource access layer accessing the business layer.
The establishment of this dependency can be considered a violation of the principle of separation of concerns. This is because the resource access layer, by definition, is not inherently concerned with business processes. The boundary which comprises the resource access should only be concerned with obtaining resources such as the retrieval of data from a database or Web service. This violation results in a resource access layer which will require modification when the dependant portion of the business layer changes, and prevents the resource access layer from being reused by other applications without requiring an indirect dependency on this application’s business layer.

Another recommended best practice related to dependency direction within application layers is the prescription that layers limit their dependencies to only the layer immediately following itself in a dependency direction chain. That is to say, layers should never bypass logically adjacent layers. This restriction centralizes the access to each layer which insulates other layers from change impact, and allows consistent processes to be applied to layer access. An example of this would be restricting the Presentation Layer from directly accessing the Resource Access Layer. Such a restriction ensures that changes to the Resource Access Layer would only affect the Business Layer. Additionally, this would allow consistent processes to be applied to the access of data, such as caching, failover, auditing, error checking, etc., and the insulation of changes to these processes.

**Data Concerns**

Applying the principle of Separation of Concerns to data involves properly modeling
information managed by a system. While the aspects of data modeling apply to both object-oriented and database design, this discussion focuses on object-oriented data concerns since the primary needs of a database often require form to follow function.

When organizing data within an object-model, the exposed information should be inherent to the entity being represented. For example, given a system which sells products to customers, an object which defines the product should not contain customer related information. This is because products are not inherently concerned with who may be purchasing the product. A better approach would be the creation of a conceptual order object which composes both customer and product information. This enables the product object to be reused by other processes in the future which may not be concerned with customer information.

In addition to considering the potential reuse of a data model within differing contexts, the intuitive organization of data is also beneficial when maintaining highly complex systems. For example, if a new developer were tasked with accessing the serial number of a particular part composed by a product object already in memory, the developer would likely first attempt to locate the particular product part and then examine the part for a “SerialNumber” or similarly named property. The developer would probably not think to look for something such as a “product number to serial number” dictionary located on the product object because this would not be a natural representation of the data. This would be similar to considering a person to have a serial number by virtue of their wearing a wrist-watch that possessed a serial number.

There are times, however, when the natural organization of data does not present an efficient mechanism for information inquiry. For example, an exceptionally complex product object which needs to be inspected frequently to derive the total count of all copper elements may not be efficiently handled through the examination or cross-referencing of each of its composed elements. In cases where natural modeling is not itself sufficient, the integrity of an object-model can be maintained by supplementing conceptual types to satisfy the specialized need. For example, if the product’s composition remains static once assembled, a conceptual model representing the product’s precious metal information (e.g. “ProductPreciousMetalManifest”) might be composed at a peer level to the product information. If the product’s composition changes frequently and the composition processes are centralized then the precious metal information might be updated as part of this process. Otherwise, a specialized component could be conceived (e.g. “PreciousMetalDetector”) to dynamically return the product’s precious metal information. As with the first example, the benefits of separating conceptual needs from an otherwise natural model are that other processes may reuse the model without incurring the overhead of non-inherent characteristics, and the model is kept easily maintainable.

**Behavior Concerns**

Separating behavior involves the division of system processes into logical, manageable,
and reusable units of code. This represents the most fundamental type of Separation of Concerns. Within object-oriented systems, fine-grained behavior is separated using methods while course-grained behavior is separated using objects, components, applications, and services.

As with the separation of data, encapsulated behavior should be inherent to its containing boundaries. For instance, a method named CreateCustomer() would be expected to only contain behavior relevant to the creation of a new customer. It wouldn’t be expected to, for example, place orders for a new customer. Similarly, a component named ProductAssembler would be expected to contain data and behavior relevant to the assembly of products. Similarly, it wouldn’t be expected to contain data or behavior related to customers.

Achieving good separation of behavior is often an iterative process. The primary behavior of a system is generally conceived during a design phase, but the specific implementation of a system design often requires several iterations of refactoring as fine-grained concerns become more apparent.

When organizing behavior, the following goals should be sought:

- Eliminate the duplication of functionality.
- Restrict the scope of work to a maintainable size.
- Restrict the scope of work to the description of the containing boundary.
- Restrict the scope of work to the inherent behavior of the containing boundary.
- Minimize external dependencies.
- Maximize the potential for reuse.

### Extending Concerns

Extensions are a separation of concerns strategy which enables the addition of new behavior to an existing set of concerns. Extensions are used to enhance existing concerns where the desired behavior cannot be added to the targeted system, is not an inherent behavior of the system, or is otherwise impractical for inclusion as part of the system’s core set of features. Figure 15 depicts the dependency relationship of an extension to a target system:
Extensions interact with the target system in response to notifications of key events. Examples of behavior provided by extensions would include the display of new views or controls, the alteration of normal processing, or the alteration of data within the target system. Extensions generally take the form of a hosted add-in, and are typically registered through configuration or a dynamic discovery process.

When multiple client applications are maintained within an enterprise, it may be discovered that behavior provided by one client’s extension would be valuable to add to another client as well. If the extension is highly customized to the target application and provides a relatively simplistic behavior, the best course of action would be to reproduce the functionality in the new application. However, if the extension’s behavior is substantial in size and complexity then it may be more appropriate to provide the behavior generically through an enterprise service. If the behavior provided by the service is still inappropriate to include as a core dependency of the host system then a new extension can be developed to interact with the service on behalf of the application.

**Delegating Concerns**

Delegating concerns refers to the process of assigning the responsibility for fulfilling behavior to a subordinate component. This strategy separates the concerns of responsibility from execution, and is beneficial for designing components whose implementation details may vary depending on external conditions. Using this strategy, components proxy some or all data requests and method invocations to another component designed to fulfill the request in a specialized way. For example, a component designed to return a list of roles assigned to the current user may delegate the request to one or more subordinate components designed to retrieve roles from a local XML file, database, and/or a remote service. Figure 16 illustrates the delegation of authorization concerns to components specialized for differing data sources:
Delegation can be facilitated using the Strategy Pattern, the Plug-in Pattern, the Microsoft Provider Model, and other variant patterns which enable portions of a component’s behavior to be fulfilled externally.

**Inverting Concerns**

Inverting concerns, better known as Inversion of Control, refers to the process of moving a concern outside of an established boundary. Some uses for inversion of concerns include effecting aspect separation, minimizing non-essential processes, decoupling components from specific abstraction strategies, or relocating responsibilities to infrastructure components. Some specific applications would include alleviating responsibilities pertaining to hardware interaction, workflow management, registration processes, or obtaining dependencies.

Figure 17 depicts an inversion of concerns process where both a presentation component and a domain component have had concerns moved to infrastructure level components:
One example where inversion of concerns can be seen is in the use of the Template Method pattern. This pattern is used to generalize the behavior of a process in order to allow variation through inheritance. When applying this pattern to an existing component, the steps performed by the desired process would be generalized and encapsulated in a base class. The existing component would then inherit from the new base class, maintaining only specialized behavior. Other types would then be free to inherit from the base class to provide variant implementations. This would be an example of inverting the concerns of an algorithm sequence.

Another example of inverting concerns can be observed when converting an interactive console application to a GUI application. A typical interactive console application provides a main loop which prompts the user for information and waits for data to be entered. When converted to a GUI application, the main control of the program is generally provided by infrastructure components, and notification of user interaction is accomplished using the Observer Pattern. This inverts the main control of the user interaction process by relying on the application infrastructure or hosting environment for these concerns.

Dependency Injection is a term applied to the inversion of concerns which relate to obtaining external dependencies. This term was conceived to more clearly describe the behavior of a class of frameworks concerned with this form of inversion. The goal of dependency injection is to separate the concerns of how a dependency is obtained from the core concerns of a boundary. This improves reusability by enabling components to be supplied with dependencies which may vary depending on context.

The process of dependency injection generally involves the roles of container, dependency, and recipient.

The role of container is occupied by a component responsible for assigning dependencies to a recipient component. Containers are generally implemented using the
Factory Pattern to allow creation of the recipient and dependency components.

The role of dependency is occupied by the resource being provided to the recipient. Containers are often implemented to allow existing objects to be registered for use as a dependency, or to create new instances when required.

The role of recipient is occupied by the component receiving a dependency from the container. Recipients are required to declare dependencies using a strategy supported by the container. Strategies for determining dependency information generally employ the use of reflection to examine one or more of the recipient’s member types, and may require the use of attributes/annotations to implicitly or explicitly specify dependencies. A sequence diagram representing these roles is presented in Figure 18.

One example use of dependency injection would be to abstract caching functionality from a business component in lieu of using the Factory or Factory Method patterns. By supplying a recipient with a caching component using dependency injection, the recipient remains loosely coupled to the caching component without the added coupling to a specific factory.

Another example would be the use of dependency injection to control the number of caching components created in lieu of using the Singleton Pattern. This enables the application to ensure that only a single instance is used throughout the lifetime of an
application, or to control which recipients receive which instances of the component. This also increases the ability to test components in isolation of their dependencies by enabling mock dependencies to be supplied rather than requiring singletons to be tested along with their consumers.

Components using patterns such as Factory, Factory Method, Abstract Factory, Singleton, Plug-in, Service Locator to abstract the concerns of object creation or dependency retrieval are good candidates to consider for supplementing or replacement with dependency injection. By using dependency injection in lieu of, or in cooperation with these and other such patterns, consuming components can be completely abstracted from the particular abstraction strategy.

The Exaggeration Exercise

It is often the case that the negative consequences of a design choice, particularly those relating to scalability and reuse, do not become apparent until long after a system has become established. Problems with scalability and reuse are usually rooted in a lack of adherence to the principle of Separation of Concerns. One process that can aid in optimizing concerns is considering the impact of a design when applied under exaggerated circumstances. Through this exercise, the use of a system is hypothetically exaggerated beyond the system’s known expectations in order to reveal potential weaknesses in a design approach. For example, when designing an object model to be used by two existing systems, one might consider what the consequences would be if the object model were shared by fifty systems. By exaggerating the use of a design, poorly organized responsibilities will often become easier to identify.

To demonstrate this exercise, consider the following example concerning the creation of a new composite customer relationship management system:

Within a corporation, an IT department has been requested to create a custom CRM application which will allow disparate development teams to contribute specialized modules of functionality. The main customer related segments within the corporation will be sales, billing, and technical support, though there may be multiple development teams assigned to support the functions of each segment. It has been requested that the application present a main screen allowing customers to be queried by a number of different criteria including name, address, phone number, order number, and any registered product serial numbers. However, the resulting views and workflows should depend upon which business function is being carried out at a given time. For example, upon submitting customer search criteria, sales users may be displayed with a view pertaining to past purchasing trends, credit rating scores, and suggested up-selling scripts; while a billing user may be presented with a view displaying customer payment history, billing dispute history, and outstanding balances. An initial analysis reveals that there will be a total of three workflow variations resulting from a total of five different search criteria fields, and that there will be three backend systems involved in obtaining the information needed by all segments. Because the variations are low, the choice is made to centralize the main view, customer search functionality, and workflow initiation
within a search module. It is understood that the addition of new search criteria and workflow needs will require modifications by the search module development team, but it is believed that these needs will be infrequent.

Applying the exaggeration exercise to this design choice might entail considering what consequences might ensue if the number of business segments or number of backend systems were increased to fifty. As a result of centralizing the search functionality and workflow initiation, increasing the scope of these concerns will also increase the responsibility and workload of the search module development team. This would include the scoping, analysis, design, coding, and testing of all new features and change requests related to these concerns. In turn, this would likely result in the need to increase the number of development resources required for the search module team. It is also likely that the same initial assumptions that lead to the consolidation of these concerns also informed other design decisions made for the search module. This may have led to design choices that would not readily accommodate such an increase in responsibilities, requiring some level of internal redesign to handle new concerns.

An outstanding problem that can be observed through this exercise is that the amount of work required by the search team is proportionally related to the number of business segments or workflows supported by the system. The fact that the initial solution doesn’t scale as the need increases can be an indication that concerns may have been inappropriately distributed throughout the system. The decision for consolidating search functionality and the resulting workflow decisions within the search module was the result of recognizing similarities between each use case. What was not given equal consideration was the fact that each use case had to be treated in a specialized way, and the fact that the number of possibilities has no true boundaries. Because the search screen provides a central function for many modules, it can be said to possess inherent infrastructure concerns. It should, therefore, be expected to provide the behavior that is common to all modules. However, the details of each case are not common, and can be considered the inherent concerns of each respective module. Upon recognizing the inherent responsibilities, an alternate approach might include the development of a framework to consolidate the common concerns, but enabling the distribution of each domain specific concern. Such an approach could be accomplished by requiring each module to provide an add-in which registers available search criteria and provides handling capabilities to invoke the corresponding workflow. A framework would then be responsible for presenting a consolidated view of the search criteria, and providing a generic infrastructure for associating each search criteria to its corresponding workflow handler. By using this approach, the search module can be designed to accommodate an unlimited number of use cases.

While the exaggeration exercise can be useful in creating highly scalable designs, this is only a byproduct of its primary goal which is the achievement of optimal Separation of Concerns. The exaggeration exercise might be compared to the medical practice of taking a patient’s temperature to test for symptoms of a possible underlying problem. This exercise attempts to identify problems pertaining to separation of concerns by examining the scalability aspect of a potential design. This might be compared to using a magnifying glass to look at the details of a design as depicted in Figure 19.
It is through the exaggeration of a design’s actual scope that small problems are magnified for easier identification. Once identified, it may then be determined what course of action should be taken.

**Separation Anxiety**

Applying the principle of separation of concerns often involves advanced concepts and constructs which bring a certain level of complexity to the application beyond that of merely addressing the domain concerns of the application. For developers inexperienced in these programming techniques the reaction falls within the spectrum of enthusiastic excitement over the opportunity to add to one’s repertoire of design skills, and an adverse reaction to the additional amount of complexity involved in getting the job done in the most expedient way possible. These techniques often lead inexperienced or more tactical-minded developers to characterize such designs as “overly-complex” or “over-engineered” based upon their frustrations in first learning, and then maneuvering through such architectures on a day-to-day basis. Furthermore, there is often an ever-present pressure from project managers, product managers, upper level management, marketing, or end users for instant gratification which tends to encourage and reward expediency over thoughtful design. These conditions can present obstacles to developing good designs beyond merely solving the technical problems.

While designs which promote separation of concerns often add complexity to an
application, it should be pointed out that they also remove the complexity that is generally associated with a lack of separation of concerns. For many applications, the trade-off is often between ordered complexity and disordered complexity. Applications which do not exhibit an appropriate amount of separation of concerns are often difficult to learn due to the need to understand the whole before understanding the part, and difficult to maintain and extend. Highly complex, yet poorly modularized applications also have the effect of causing high turnover in development staff (which tends to further compound design and implementation problems), or attracting individuals who have an aversion to change and seek to build their career by being the indispensible “master of the maze” within their organization. Development teams should certainly not seek complexity for complexity’s sake (unless entering an obfuscation contest), but the notion that avoiding advanced design concepts equates to avoiding complexity should be dispelled.

**Conclusion**

In a sense, the goal of Separation of Concerns is one of artistic and mechanical harmony. A system comprised of cooperating components each of which performs only those tasks inherent to its nature is a unique form of artistic achievement which blends both form and function. It is an endeavor which can satisfy both the creative and analytical cravings of the human mind while at the same time helping one to meet and often far surpass external expectations.