# The F5 Cloud Bursting Reference Architecture

Extend your data center to the cloud, secure and accelerate connections, scale, and make use of PAYG resources with the F5 Cloud Bursting architecture.

# Introduction

Aggressive competition is forcing organizations to tighten budgets and extract more productivity from existing resources. But fluctuating demands on a leaner data center infrastructure bring a risk: tipping the scale from an acceptable user experience to catastrophic failure and service outage.

Solving this quandary lies in abstracting the processing and application delivery requirements from the static, underlying infrastructure and enabling those requirements to span physical locations as demand dictates. This can be achieved safely when an agile data center architecture is tightly coupled with clear insight into the application and the consumer experience, enabling workload fluidity across truly optimized resources.

Data center agility enables swift reactions to changes in resource demand, and the faster an organization can react to change, the leaner the data center can be run. Quick reactions are only possible, however, through the automation and orchestration of processes—including the harvest of real-time user experience metrics and the execution of orchestration workflows.

In response to demand for these capabilities, service providers have developed pay-asyou- go (PAYG) licensing models. Successfully using such subscription services, however, requires an architecture that can quickly adapt to consume the PAYG resources prior to reaching demand levels that cause service outages.

# Business Challenges

Demand fluctuations are a reality in business today, and due to the explosion of Internet- connected devices, the ferocity of demand spikes is ever increasing. Scaling data center resources to match the worst-case scenario, however, inhibits competitiveness by building in incremental resource waste and thus failing to deliver an acceptable return on investment (ROI).

The causes of significant fluctuations in consumer demand vary greatly between industries and content type, but common reasons include service uptake that exceeds planning and consumer responses to promotions.
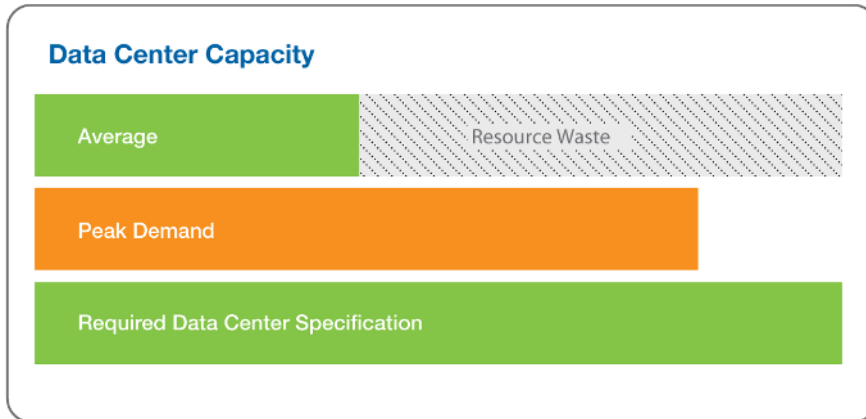
## Benefits of Automation

Reported benefits of automation to the IT process:

- 80% time savings
- 62% cost savings
- 54% improved SLA delivery to the business

"

79% of enterprises that have implemented process automation have experienced time savings; 69% claim improved productivity.
—Source: Redwood Software survey, October 2012

**Data Center Capacity**

| | |
|---|---|
| Average | Resource Waste |
| Peak Demand | |
| Required Data Center Specification | |

## Customer Uptake

Predicting capacity requirements is a dark art—often no more than an educated guess, in the case of a service that is open to all on the Internet. Occasionally there are exceptions to the rules of educated guessing and systems become overwhelmed by unforeseen success, causing a popularity-based DoS situation.

This phenomenon is regularly experienced both by new services and existing services delivering a one-time promotion. For example, consider a ticket sale for a popular concert tour. The resource demand is often only hours long, making it impossible to maintain a data center that would most of the time run only at minimum potential, consuming power and cooling resources, while also remaining able to scale to excessive demand. The ROI simply does not add up.

## Promotional Response

The Slashdot effect, also known as slashdotting, is a well-documented scenario that occurs when a popular website links to a smaller site, causing a massive increase in web traffic that overloads the smaller site, causing it to slow or even become temporarily unavailable. The name stems from the huge influx of web traffic that can result when the technology news site Slashdot links to other websites. So great is Slashdot's following that few web services can handle the concentrated surge.

The effects of slashdotting cannot be planned for, since the cause is outside of the control of any enterprise. However, the result is often public ridicule, with the response to a "server down" post on Slashdot causing even more hits from the "I want to see that myself" crowd.

Does the Siri outage reveal its success?

> Siri went down on Thursday for its first extended outage—around five hours, according to most counts. Five hours is hardly three days (like another noteworthy recent
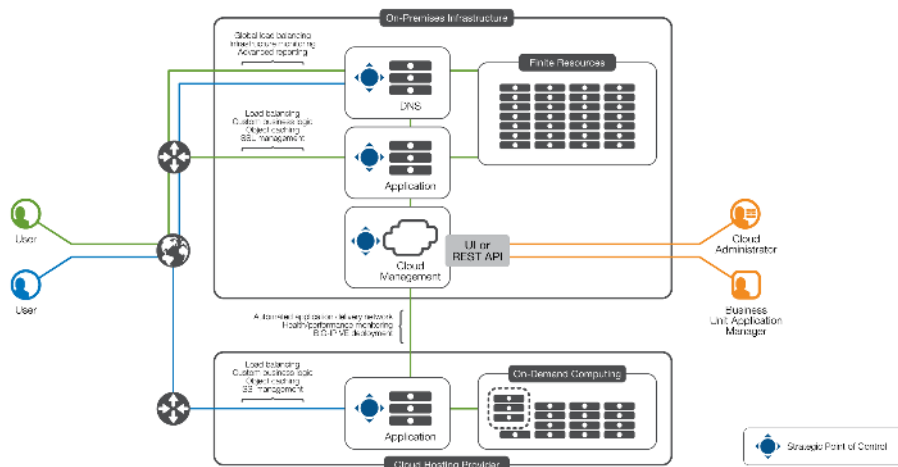
# Business Solution

Managing a more cost-effective data center is easy with predictable demand patterns like those experienced in test labs, but it's not so easy for real deployments. Running a leaner, more efficient data center is achieved first by not building architecture for the worst-case demand, leaving large amounts of equipment idle for the majority of the time.

mobile service blackout), but the reactions of media and users show Apple's personal assistant is making its presence felt.
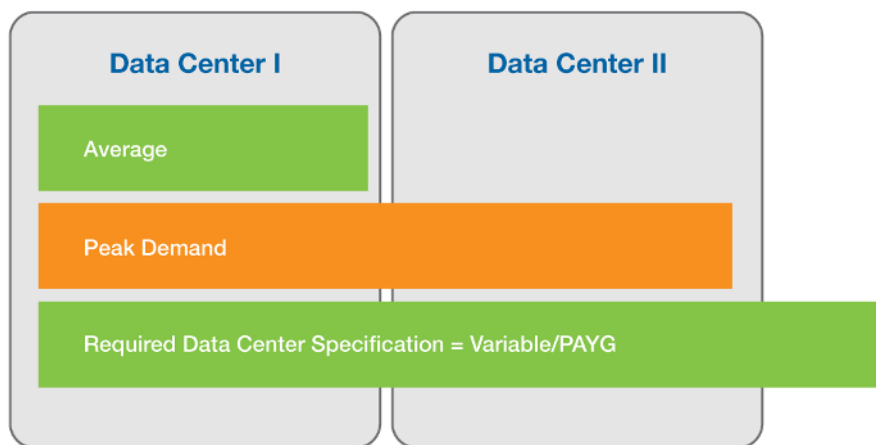
—Darrell Etherington, Gigaom



The F5 Cloud Bursting architecture

Cloud computing is today's obvious answer for achieving incremental capacity only when it's needed. The limits of servers and networks can be pushed safely with integration between the management tools and connectivity of public and private environments, creating a seamless experience across the two and delivering a transparent extension of the data center environment that avoids resource and management silos.



To fully exploit the benefits of an agile data center, however, organizations must have insight into when actions or workflows should be applied. This requires coherent visibility into user interactions with applications and the user experience itself—that is, user and application coherence. In addition to implementing the necessary agile infrastructure, the F5 Cloud Bursting architecture delivers this coherence by harvesting real-time service behavior metrics for infrastructure management and workflow systems. The collection of user experience data is crucial to implementing demand-based workflow routing, in and out of public and private data centers.

## Technology Solution

The F5 Cloud Bursting solution automates and orchestrates the deployment of application delivery services across both traditional and cloud infrastructures while also managing the dynamic redirection of workloads to the most suitable location. These application delivery services ensure predictable user experiences, replicated security policy, and workload agility.

F5 BIG-IQ Cloud federates management of F5 BIG-IP solutions across both traditional and cloud infrastructures, helping organizations deploy and manage application delivery services in a fast, consistent, and repeatable manner, regardless of the underlying infrastructure. In addition, BIG-IQ Cloud integrates or interfaces with existing cloud orchestration engines such as VMware vCloud Director to streamline the overall process of deploying applications.

BIG-IQ Cloud leverages F5 iApps Templates and a self-service model to rapidly provision application delivery services, enabling new applications to be made available to users in minutes instead of weeks. The IT organization can define a catalog of available application delivery services, including customized or multi-tiered offerings, from which administrators and application managers can quickly select as needed.

F5 BIG-IP Global Traffic Manager (GTM) directs users to the nearest data center that will provide the best application experience—based on both proximity and data center performance. This ensures that distributed applications are always available and fast. BIG-IP GTM employs a range of global load balancing methods and intelligent monitoring specific to each application and each user. It continuously monitors application availability and also routes traffic according to business policies and current network and user conditions.

## Business Benefits

The F5 Cloud Bursting architecture implements a rapid Application Delivery Network provisioning solution, drastically reducing the lead times for expanding application delivery capabilities across data centers, be they private or public. As a result, organizations can efficiently:

- Extend data centers to the cloud.
- Scale applications beyond the data center when required.
- Secure and accelerate connections to the cloud.
- Make use of PAYG resources.

## Conclusion

Organizations are looking to increase profitability and competitiveness by doing more with less. How far they can drive data center efficiency depends on the extent to which they can abstract the processing and service delivery requirements from the static, underlying infrastructure, enabling applications to span physical location as demand dictates. The F5 Cloud Bursting architecture delivers this abstraction, increasing workload fluidity and true resource commoditization while providing the necessary application and user coherence to ensure bursting that's transparent to users.

F5 Networks, Inc.
401 Elliott Avenue West, Seattle, WA 98119
888-882-4447 www.f5.com

Americas
info@f5.com

Asia-Pacific
apacinfo@f5.com

Europe/Middle-East/Africa
emeainfo@f5.com

Japan
f5j-info@f5.com