Google’s effort is concentrated in the area of Software-as-a-Service (SaaS). It is estimated that the

number of servers used by Google was close to 1.8 million in January 2012 and was expected to reach

close to 2.4 million in early 2013 [289].

Services such as Gmail, Google Drive, Google Calendar, Picasa, and Google Groups are free of

charge for individual users and available for a fee for organizations. These services are running on a cloud

and can be invoked from a broad spectrum of devices, including mobile ones such as iPhones, iPads,

Black-Berrys, and laptops and tablets. The data for these services is stored in data centers on the cloud.

The Gmail service hosts emails on Google servers and, provides aWeb interface to access them and

tools for migrating from Lotus Notes and Microsoft Exchange. Google Docs is Web-based software

for building text documents, spreadsheets, and presentations. It supports features such as tables, bullet

points, basic fonts, and text size; it allows multiple users to edit and update the same document and view

the history of document changes; and it provides a spell checker. The service allows users to import and

export files in several formats, including Microsoft Office, PDF, text, and OpenOffice extensions.

Google Calendar is a browser-based scheduler; it supports multiple calendars for a user, the ability

to share a calendar with other users, the display of daily/weekly/monthly views, and the ability to search

events and synchronize with the Outlook Calendar. Google Calendar is accessible from mobile devices.

Event reminders can be received via SMS, desktop popups, or emails. It is also possible to share your

calendar with other Google Calendar users. Picasa is a tool to upload, share, and edit images; it provides

1 GB of disk space per user free of charge. Users can add tags to images and attach locations to photos using Google Maps. Google Groups allows users to host discussion forums to create messages online

or via email.

Google is also a leader in the Platform-as-a-Service (PaaS) space. AppEngine is a developer platform

hosted on the cloud. Initially it supported only Python, but support for Java was added later and detailed

documentation for Java is available. The database for code development can be accessed with Google

Query Language (GQL) with a SQL-like syntax.

The concept of structured data is important to Google’s service strategy. The change of search

philosophy reflects the transition from unstructured Web content to structured data, data that contains

additional information, such as the place where a photograph was taken, information about the singer

of a digital recording of a song, the local services at a geographic location, and so on [227].

Search engine crawlers rely on hyperlinks to discover new content. The deep Web is content stored

in databases and served as pages created dynamically by querying HTML forms. Such content is

unavailable to crawlers that are unable to fill out such forms. Examples of deep Web sources are sites

with geographic-specific information, such as local stores, services, and businesses; sites that report

statistics and analysis produced by governmental and nongovernmental organizations; art collections;

photo galleries; bus, train, and airline schedules; and so on. Structured content is created by labeling;

Flickr and Google Co-op are examples of structures where labels and annotations are added to objects,

images, and pages stored on the Web.

Google Co-op allows users to create customized search engines based on a set of facets or categories.

For example, the facets for a search engine for the database research community available at

http://data.cs.washington.edu/coop/dbresearch/index.html are professor, project,

publication, jobs.

Google Base is a service allowing users to load structured data from different sources to a central

repository that is a very large, self-describing, semi-structured, heterogeneous database. It is selfdescribing

because each item follows a simple schema: (item type, attribute names). Few users are

aware of this service. Google Base is accessed in response to keyword queries posed on Google.com,

provided that there is relevant data in the database. To fully integrate Google Base, the results should

be ranked across properties. In addition, the service needs to propose appropriate refinements with

candidate values in select menus; this is done by computing histograms on attributes and their values

during query time.

Google Drive is an online service for data storage that has been available since April 2012. It gives

users 5 GB of free storage and charges $4 per month for 20 GB. It is available for PCs, MacBooks,

iPhones, iPads, and Android devices and allows organizations to purchase up to 16 TB of storage.

Specialized structure-aware search engines for several interest areas, including travel, weather, and

local services, have already been implemented. However, the data available on theWeb covers a wealth

of human knowledge; it is not feasible to define all the possible domains and it is nearly impossible to

decide where one domain ends and another begins.

Google has also redefined the laptop with the introduction of the Chromebook, a purelyWeb-centric

device running Chrome OS. Cloud-based applications, extreme portability, built-in 3G connectivity,

almost instant-on, and all-day battery life are the main attractions of this device with a keyboard.

Google adheres to a bottom-up, engineer-driven, liberal licensing and user application development

philosophy, whereas Apple, a recent entry in cloud computing, tightly controls the technology stack, builds its own hardware, and requires application developers to follow strict rules. Apple products,

including the iPhone, the iOS, the iTunes Store, Mac OS X, and iCloud, offer unparalleled polish and

effortless interoperability, but the flexibility of Google results in more cumbersome user interfaces for

the broad spectrum of devices running the Android OS.

Google as well as the other cloud service providers must manage vast amounts of data. In a world

where users would most likely desire to use multiple cloud services from independent providers, the

question of whether the traditional data base management services (DBMSs) are sufficient to ensure

interoperability comes to mind. A DBMS efficiently supports data manipulations and query processing

but operates in a single administrative domain and uses well-defined schema. The interoperability of data

management services requires semantic integration of services based on different schemas. An answer

to the limitations of traditional DBMS is the so-called dataspaces introduced in [127]; dataspaces do

not aim at data integration but rather at data coexistence.