Cloud Computing & Big Data

PARALLEL & SCALABLE MACHINE LEARNING & DEEP LEARNING

Prof. Dr. – Ing. Morris Riedel
Associated Professor
School of Engineering and Natural Sciences, University of Iceland, Reykjavik, Iceland
Research Group Leader, Juelich Supercomputing Centre, Forschungszentrum Juelich, Germany

@Morris Riedel @MorrisRiedel @MorrisRiedel

LECTURE 9

Platform-As-A-Service (PAAS)

November 03, 2020
Online Lecture

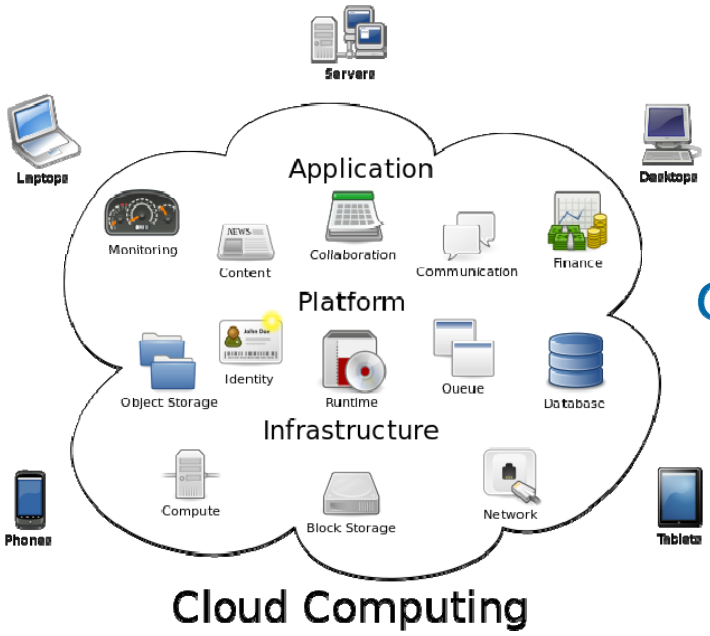
UNIVERSITY OF ICELAND
SCHOOL OF ENGINEERING AND NATURAL SCIENCES
FACULTY OF INDUSTRIAL ENGINEERING, MECHANICAL ENGINEERING AND COMPUTER SCIENCE

JÜLICH Forschungszentrum | JÜLICH SUPERCOMPUTING CENTRE

HELMHOLTZ AI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

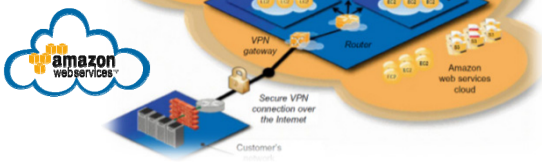
Review of Lecture 8 – Infrastructure-As-A-Service (IAAS)

- Cloud Service Models: *AAS

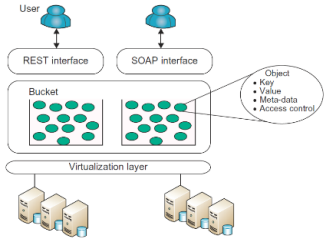


(no strict boundaries & public clouds offer services on all levels these days)

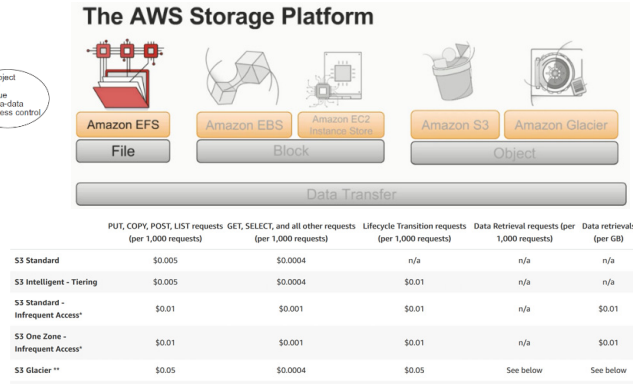
- IAAS Services with Application Examples
  - Application for IAAS means often 'some form of compute/storage/network'



(AWS EC2 Compute IAAS)



(AWS S3 Storage IAAS 'Object Storage')



On-premise full custom	Map-Reduce Appliance	Map-Reduce Hosting	Map-Reduce As-A-Service

Bare-metal ⟷ *Virtualized* ⟷ *Clouds*

[1] Wikipedia 'Cloud computing' [2] Distributed & Cloud Computing Book [3] AWS Web page modified from [4] Accenture Deployment Models [5] Amazon S3 Pricing Models

Outline of the Course

1. Cloud Computing & Big Data Introduction
2. Machine Learning Models in Clouds
3. Apache Spark for Cloud Applications
4. Virtualization & Data Center Design
5. Map-Reduce Computing Paradigm
6. Deep Learning driven by Big Data
7. Deep Learning Applications in Clouds
8. Infrastructure-As-A-Service (IAAS)
9. Platform-As-A-Service (PAAS)
10. Software-As-A-Service (SAAS)
11. Big Data Analytics & Cloud Data Mining
12. Docker & Container Management
13. OpenStack Cloud Operating System
14. Online Social Networking & Graph Databases
15. Big Data Streaming Tools & Applications
16. Epilogue

+ additional practical lectures & Webinars for our hands-on assignments in context

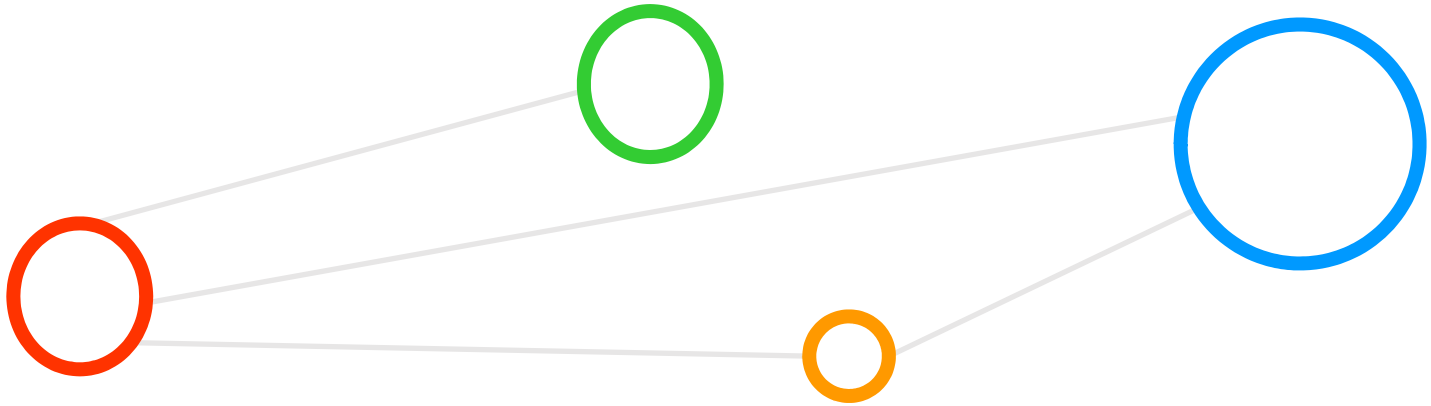
- Practical Topics
- Theoretical / Conceptual Topics

Outline

- Understanding PAAS Environments
  - Different Cloud Service Levels Reviewed & PAAS ‘Lego Bricks‘
  - Google Cloud Products & PAAS Building Blocks Examples
  - Google App Engine (GAE) & NoSQL Databases as ‘Key-Value Stores‘
  - Understanding the Differences between SQL & NoSQL Databases
  - Selected PAAS Application Examples & GAE Pricing Models
- Advanced PAAS Topics & Applications
  - The Role of Software Development Kits (SDKs) & GAE Example
  - PAAS Services Big Query & Machine Learning PAAS Services
  - Natural Language Processing (NLP) & Sentiment Analysis
  - Rovio Games Angry Birds, Twitter, IKEA Application Examples
  - Google Cloud TPU & Disadvantages of the PAAS Approach with Vendor-Locks

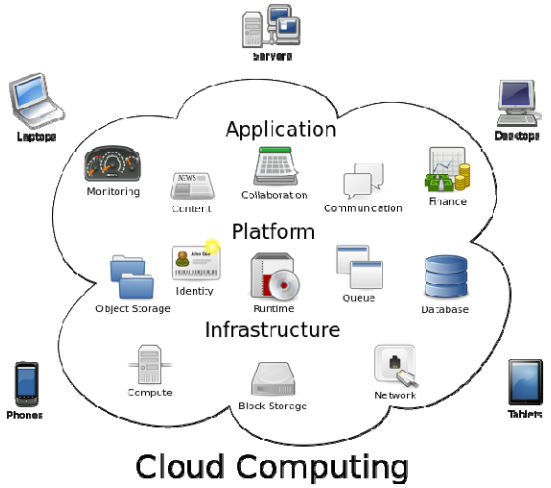
- **Promises from previous lecture(s):**
- ***Lecture 1:*** **Lecture 9 provides more details about Google Cloud services and its Platform-as-a-Service (PAAS) models & various cloud services**
- ***Lecture 4:*** **Lecture 8 & 9 & 10 offer more insights into concrete cloud systems and their use of virtualization on different levels of cloud services**
- ***Lecture 4:*** **Lecture 8 & 9 & 10 will clarify & compare cloud deployment models with the different cloud computing layers IAAS, PAAS, and SAAS**

Understanding PAAS Environments



Three Levels of Cloud Service Models: *AAS – Revisited

- Levels oriented towards different users
  - Full customization to direct usable applications
- Software as a Service (SAAS)
  - Provides specific 'ready-to-run applications'
  - Sometimes related to geographical location
- Platform as a Services (PAAS) **focus in this lecture**
  - Virtual images ready to deploy your software
  - Includes a 'platform for creation of your services'
- Infrastructure as a Service (IAAS)
  - Provides 'bare metal infrastructure' & virtual IT resources (cf. Lecture 4)
  - Use and tune infrastructure as needed (compute, storage, networking, …)



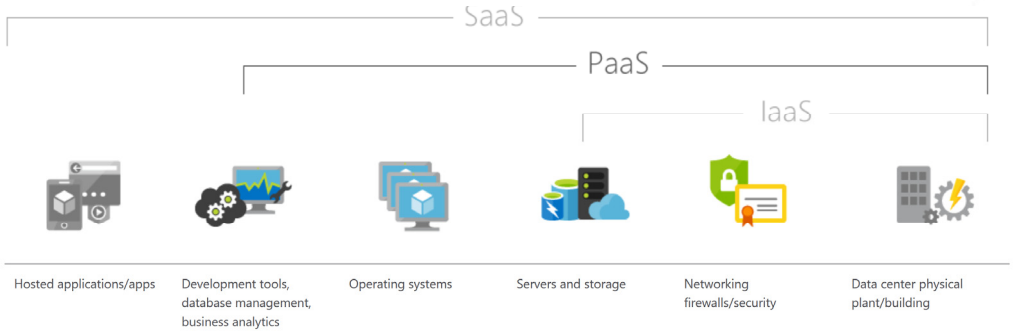
[1] Wikipedia 'Cloud computing'

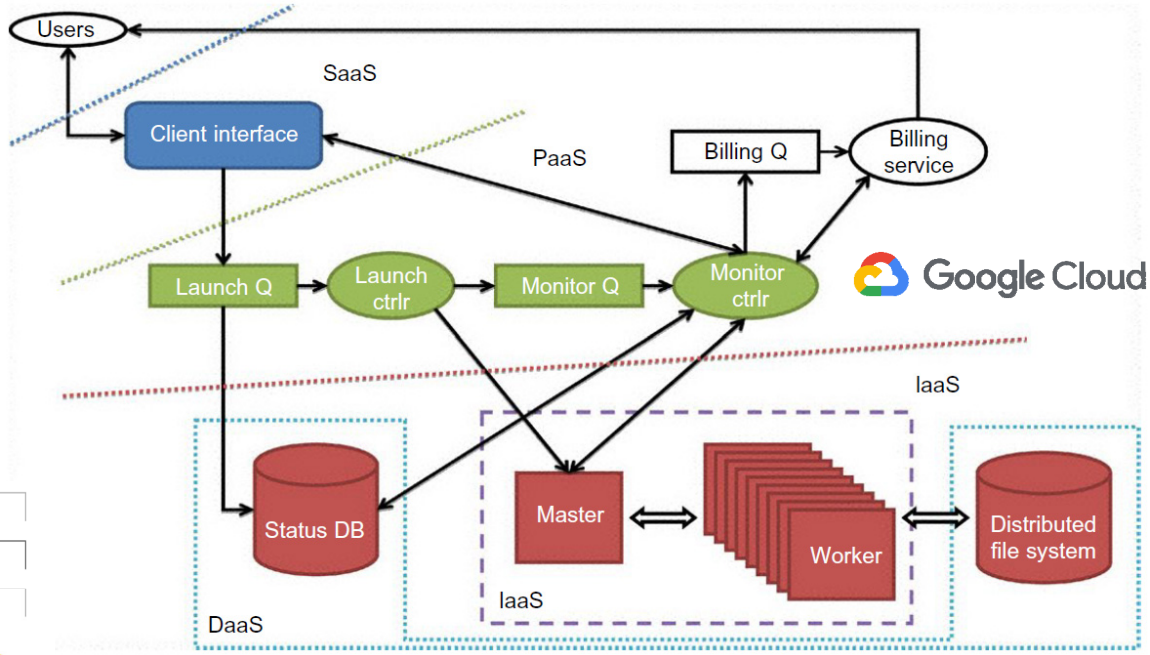
- **Cloud computing infrastructures typically offer services on three different levels: Infrastructure as a Service (IAAS), Platform as a Service (PAAS), and Software as a Service (SAAS) whereby also often some services build on one another (e.g., AWS EMR PAAS builds on top of AWS EC2 IAAS)**
- **Often Cloud computing service providers (i.e., AWS, MS Azure, Google Cloud) offer service on all these different levels today: IAAS, PAAS, and SAAS**

Different Cloud Service Models – PAAS – Revisited

- Platform-As-A-Service (PAAS)
  - E.g. used to provision billing services, handle compute job queing, launching of images, and monitoring to support application developers
  - E.g. Google Cloud

- **The Conceptual ideas and key usage of the PAAS cloud service model is building cloud applications with software development kits (SDKs) & application programming Interfaces (APIs) via basic services**
- **PAAS is based on known application frameworks similiar to ASP, J2EE, JSP, and languages like Python, Java, Ruby, etc.**





[22] MS Azure, What is PAAS? *[2] Distributed & Cloud Computing Book*

Google Cloud Example – Cloud IAAS Computing Products & Categories

Offering	Common uses	Industry
Compute Engine Scalable, high-performance and general purpose VMs.	• LOB apps • Web hosting • Enterprise apps • Databases • Most workloads	• Education • Energy • Financial services • Gaming • Government • Healthcare • Life sciences • Media and entertainment • Retail • Telecommunications
Migrate for Compute Engine Server and VM migration to Compute Engine (formerly Velostrata).	Migrate applications from on-premises, multiple data centers, or clouds to Google Cloud.	
Cloud GPUs GPUs for machine learning, scientific computing, and 3D visualization.	• Machine learning • Medical analysis • Seismic exploration • Video transcoding • Graphic visualization • Scientific simulations	• Gaming • Information technology • Life sciences • Media and entertainment
Preemptible VMs Affordable, short-lived compute instances suitable for batch jobs and fault-tolerant workloads.	• Short-lived or fault-tolerant workloads • Financial modeling • Rendering • Media transcoding • Manufacturing design • Hadoop and big data • Continuous integration • Web crawling	• Energy • Finance • Healthcare • Media and entertainment • Pharmaceuticals
Shielded VMs Hardened virtual machines.	• Defend against rootkits and bootkits • Protect enterprise workloads • Protect against remote attacks, privilege escalation, and malicious insiders • Enables UEFI Secure Boot capability	• Financial services • Logistics • Manufacturing • Media and entertainment • Retail • Supply chain
Sole-tenant nodes Dedicated hardware for your compliance, licensing, and management needs.	• BYOL • Dedicated compute for workloads • Help meet security and compliance needs • Meet per-socket or per-core licensing requirements • View physical core usage information	• Financial services • Logistics • Manufacturing • Media and entertainment • Retail • Supply chain

(cf. Lecture 4)

(cf. Lecture 4)

(cf. Lecture 6 & 7)

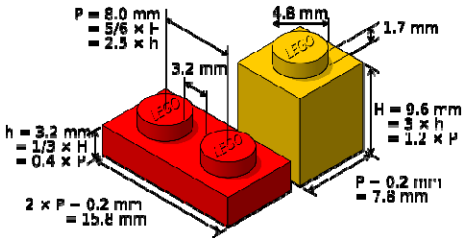
Google Cloud

[9] Google Cloud – Compute Products

- **The Google Cloud Platform offers a wide variety of IAAS/PAAS/SAAS technologies in the areas of compute, storage and databases, networking, big data, machine learning, management tools, developer tools, as well as identity and security – many of them can be combined for a wide set of applications**
- **The Google App Engine (GAE) PAAS is taking advantage of the IAAS Google infrastructure to build highly scalable applications on a managed platform**

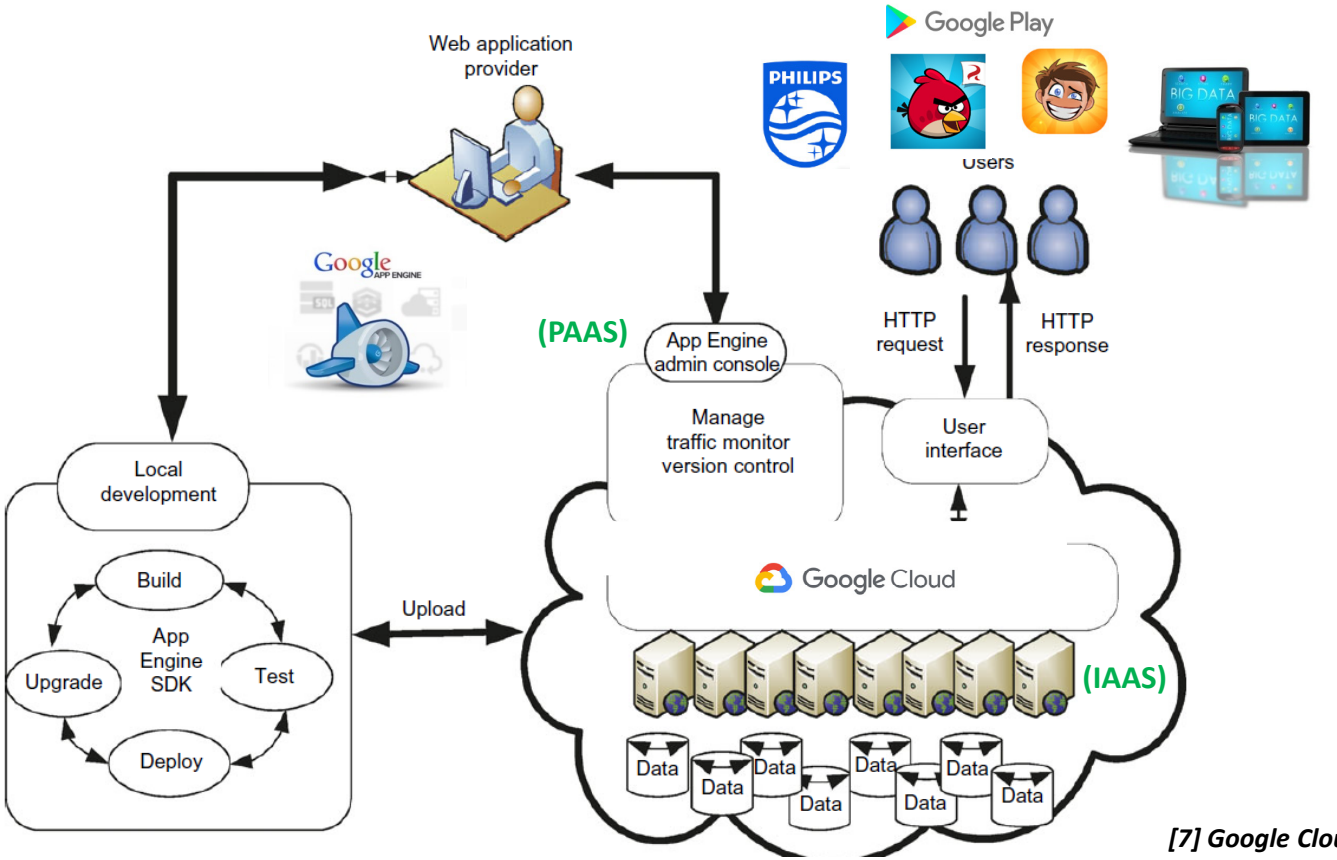
PAAS – Building Cloud Solutions Overview

- Simple idea
  - Abstract from underlying computing and storage infrastructure by using a platform, e.g. Google Cloud
  - Platform provides easy-to-use 'lego bricks' (aka 'service bricks') to create online services & Apps



[6] Lego Bricks

Modified from [2] Distributed & Cloud Computing Book



[7] Google Cloud

Google Cloud Example – Google App Engine as PAAS

Google Cloud | Why Google | Solutions | Products | Pricing | Getting Started | Docs | Support | English | Console

Serverless computing | Contact Sales

- App Engine
  - Benefits
  - Key features
  - Customers
- Documentation
- Use cases
  - Modern web applications
  - Scalable mobile back ends
- All features
- Pricing
- Take the next step

App Engine

Build highly scalable applications on a fully managed serverless platform.

New customers get $300 in free credits to spend on Google Cloud during the first 90 days. All Google Cloud customers get 28 instance hours per day free of charge.

Go to console

- ✓ Scale your applications from zero to planet scale without having to manage infrastructure
- ✓ Free up your developers with zero server management and zero configuration deployments
- ✓ Stay agile with support for popular development languages and a range of developer tools

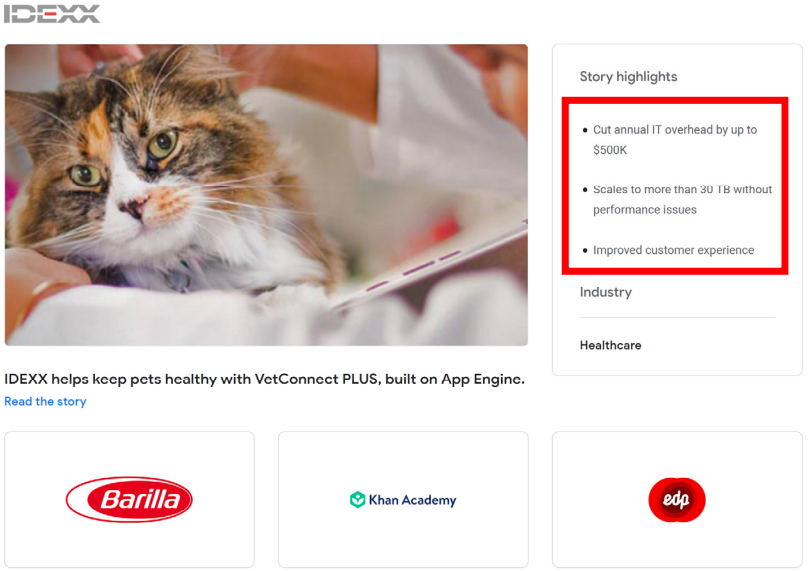
VIDEO

App Engine in 1 minute

[8] Google Cloud – GAE

PAAS Example – Google App Engine (GAE)

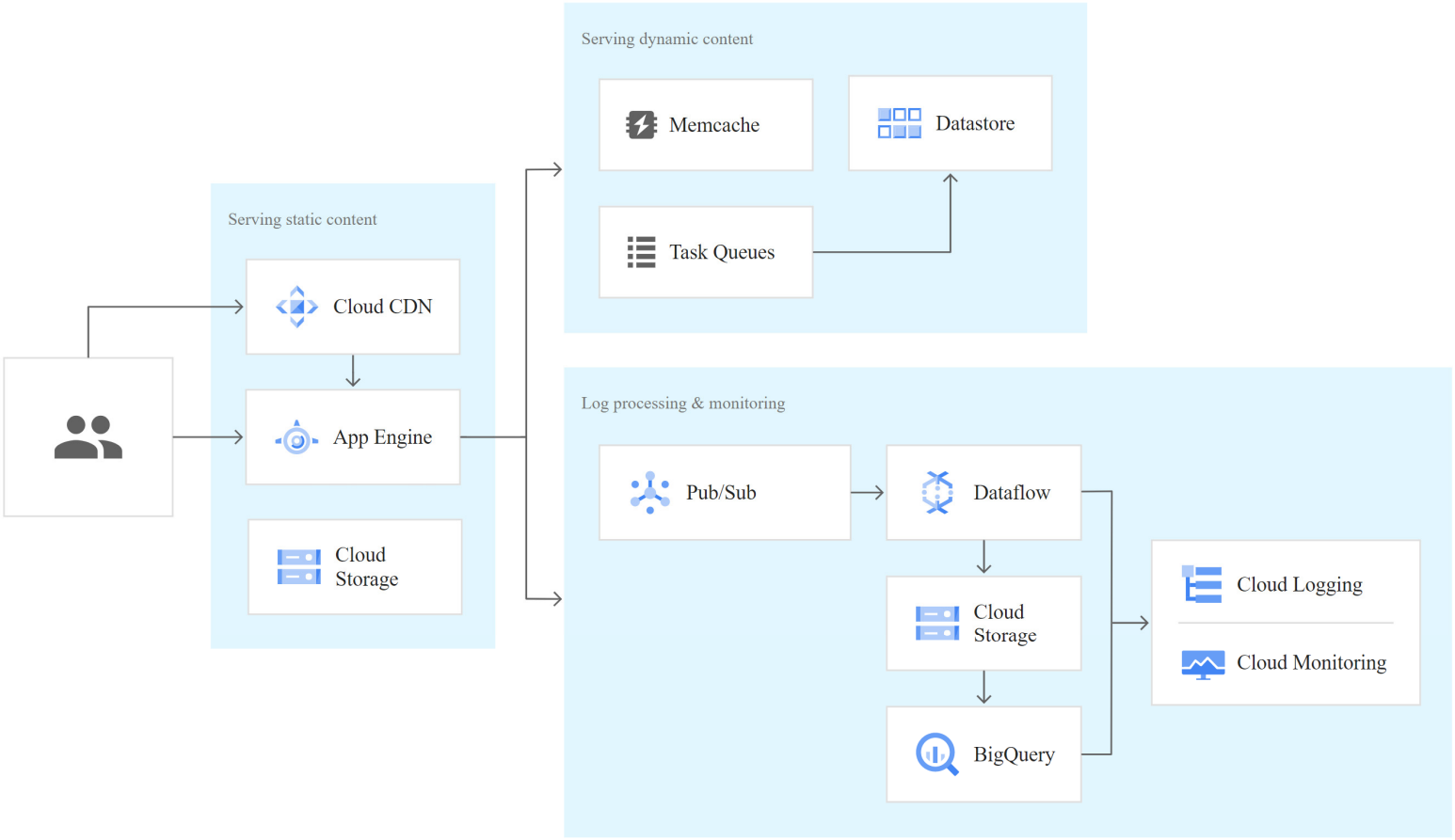
- Platform for building scalable Web applications
  - Based on the Google Cloud Platform – ‘Compute‘ area
  - Offers built-in services common to most applications
  - Upload app code and Google manages app availability
  - No need for servers to buy and/or maintain
  - Scales automatically from 1 to million users
  - Supports known programming languages like Java, Python, PHP or Go
- Selected Business Cases
  - Use powerful platform & existing services to manage apps for ‘uncertainty‘
  - Build server-based Apps that instantly launch as many servers as required



- **Google App Engine (GAE) is a PAAS-based cloud model for building scalable Web applications and mobile backends that scale automatically in response to the amount of traffic they receive**
- **GAE offers services of the Google infrastructure common to most applications: NoSQL datastore, user authentication API, memcache, Google cloud SQL, Google-like search, load balancing, health checks, application logging, security vulnerabilities scans, traffic splitting, task queues, etc.**

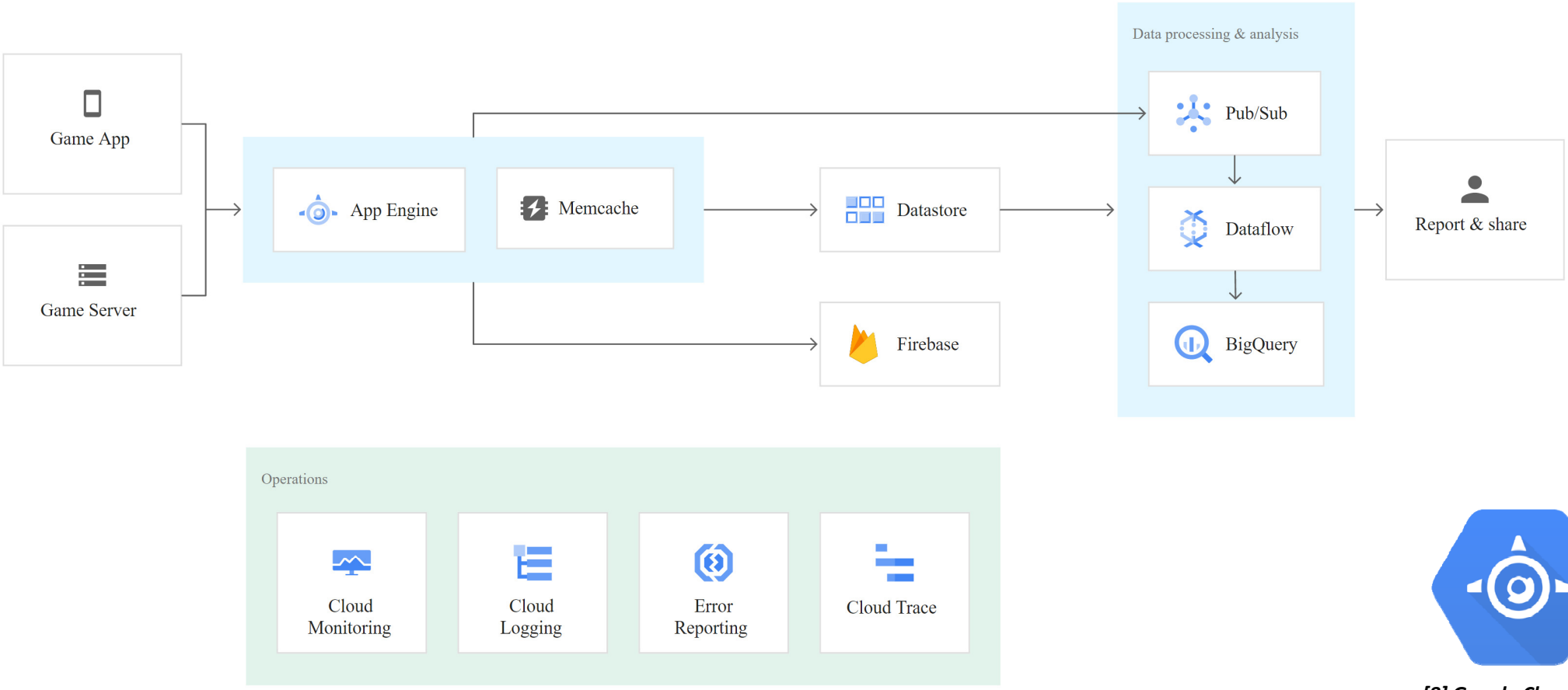
[8] Google Cloud – GAE

GAE PAAS Example – Creating a Web Application with Static/Dynamic Content



[8] Google Cloud – GAE

GAE PAAS Example – Creating Scalable Mobile Back Ends for Games



[8] Google Cloud – GAE

GAE based on Google Cloud – Review Pricing Options & Scalable Cloud Approach

Instance class	Cost per hour per instance
B1	$0.05
B2	$0.10
B4	$0.20
B4_1G	$0.30
B8	$0.40
F1	$0.05
F2	$0.10
F4	$0.20
F4_1G	$0.30

Resource	Unit	Unit cost (in US $)
Blobstore stored data*	Gigabytes per month	$0.026
Dedicated memcache	Gigabytes per hour	$0.06
Logs API	Gigabytes	$0.12
Search API**: Total storage (documents and indexes)	per GB per month	$0.18
Search API: Queries	per 10K queries	$0.50
Search API: Indexing searchable documents	per GB	$2.00
Sending email, shared memcache, cron, APIs (Task Queues, Image, Files, Users)		No Additional Charge

Pay only for what you use

Operate in a serverless environment without worrying about over or under provisioning. App Engine automatically scales depending on your app traffic and consumes resources only when your code is running.

Resource	Unit	Unit cost (in US $)
Outgoing network traffic*	Gigabytes	$0.12
Incoming network traffic	Gigabytes	Free

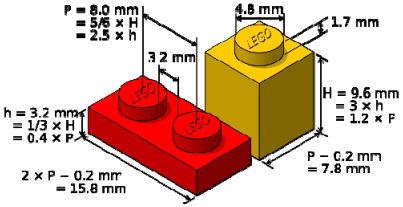
(encourage migration to the Google Cloud by having free incoming networking traffic)

Resource	Unit	Unit cost
vCPU	per core hour	$0.0526
Memory	per GB hour	$0.0071
Persistent disk	Priced as Compute Engine persistent disk, which is called "Storage PD Capacity" on your bill.	
Outgoing network traffic	Priced as Compute Engine internet egress.	
Incoming network traffic	Gigabytes	Free

[8] Google Cloud – GAE Pricing

Popular Modern PAAS Building Block Example – NoSQL Database

- Motivations
  - Simplicity of design, horizontal scaling, and finer control over availability
- 'Implementation concepts' of NoSQL Databases
  - Often highly optimized 'key–value stores'
  - Intended for 'simple retrieval and appending operations' (no update!)
  - Increasing 'latency' and 'throughput' through having less constrained consistency models
- Big Data – 'Accumulate data over time'
  - Data is indexed using 'row' and 'column names' (e.g. arbitrary strings)
  - Data are represented by 'uninterpreted strings' (but clients often serialize various forms of structured and semi-structured data into strings)



[6] Lego Bricks

[11] Wikipedia on 'NoSQL'

- **NoSQL databases provide mechanisms for storage and retrieval of data that employs less constrained consistency models than traditional relational databases (i.e. non-relational databases)**
- **NoSQL databases are also referred to as 'Not only SQL' to indicate that SQL-like queries still usable**

[8] Google Cloud – GAE

Philips Example – GAE & NoSQL Cloud Datastore (1)

- Business Case
  - Enhance interactice home lighting with Philipps Hue
  - Control home lighting from smartphone apps
  - Revenue: new product lines interconnected with phones
- Benefits
  - Frees engineers to work on product development rather than managing IT
  - Serve 200 million transactions per day
- Approach
  - Use Google Cloud Platform with GAE and NoSQL-based Google Cloud Datastore

[12] Philipps Case Study

- **Philipps connects light bulbs to the Internet and offers interactive app by using the GAE**
- **GAE provides a powerful backend PAAS solution that let Philipps apps securely access, monitor, and interact through the Internet connected Philipps Hue Bridge with the home lightning system**

[8] Google Cloud – GAE

Philips Example – GAE & NoSQL Cloud Datastore (2)

- Business Case
  - Enhance interactice home lighting with Philipps Hue
  - Control home lighting from smartphone apps with unanticipated number of users / day
  - Revenue: new product lines interconnected with phones and scale to number of users / day

[12] Philipps Case Study

- Facts & Statements

"With Philips Lighting and Philips Hue, we don't just sell light bulbs—we sell a way for lighting to change your home and life. We chose Google Cloud Platform to power Philips Hue's backend because it scales instantly, freeing engineers to work on product development rather than managing infrastructure."

—George Yianni, Head of Technology, Home Systems, Philips Lighting

- Runs a platform at 10 times the scale of other similar projects with one-tenth the workforce
- Routes 200 million transactions per day through a Google powered backend
- Integrates Philips Hue with other smart home devices and platforms, including Nest

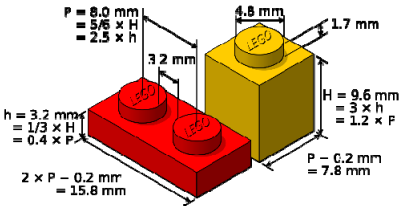
Processing 25 million remote lighting commands per day

- **Benefits for Philipps & other Cloud PAAS users is to free engineers to work on product development rather than managing infrastructure**

[8] Google Cloud – GAE

GAE – Using NoSQL-based Cloud Datastore

- NoSQL Features
  - Built for automatic scaling and high performance (including replication)
  - Easier application development (e.g. no database schema means changes to the underlying data structure possible as application evolve)
  - Interface similar like SQL traditional databases (API with REST calls)
  - Different way how relationships are described (compared to SQL)
  - Offers powerful query engine with search for data across properties (+sort)
- Philipps application example
  - E.g. store data like which light is on/off in which room of which building in which country…
  - E.g. flexible data structure, lamp moved from room A into room B
  - E.g. scaling the solution while growing with number of users



[6] Lego Bricks

Concept	Cloud Datastore	Relational database
Category of object	Kind	Table
One object	Entity	Row
Individual data for an object	Property	Field
Unique ID for an object	Key	Primary key

[13] Google Cloud Datastore

Highly scalable NoSQL database

Firestore is the next generation of Datastore. Learn more about upgrading to Firestore.

Datastore is a highly scalable NoSQL database for your applications. Datastore automatically handles sharding and replication, providing you with a highly available and durable database that scales automatically to handle your applications' load. Datastore provides a myriad of capabilities such as ACID transactions, SQL-like queries, indexes, and much more.

[8] Google Cloud – GAE

Google Cloud – NoSQL Databases with Specific Properties

Databases →

Cloud Bigtable
NoSQL wide-column database for storing big data with low latency.

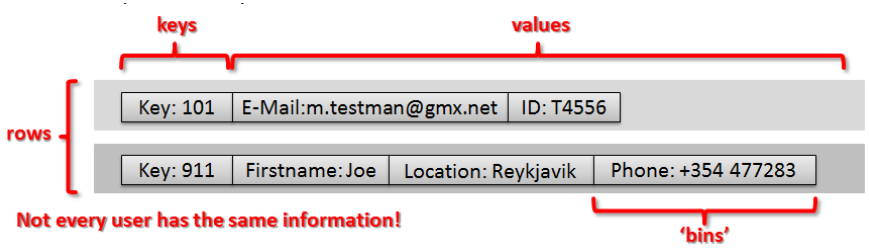
Firestore
NoSQL document database for mobile and web application data.

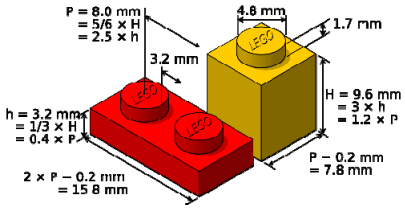
Memorystore
In-memory data store service for Redis for fast data processing.

Cloud Spanner
Relational database with unlimited scale and high availability.

Cloud SQL
Relational database services for MySQL, PostgreSQL, and SQL server.

Firebase Realtime Database
NoSQL cloud database for storing and syncing data in real time.





[6] Lego Bricks

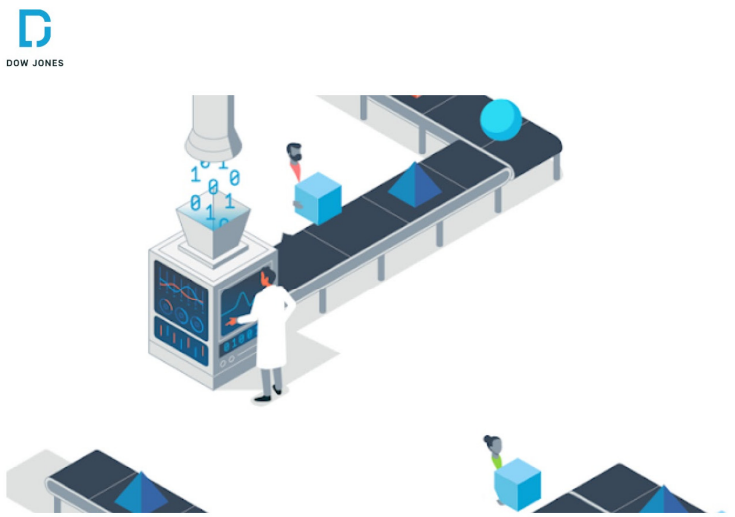
- **GAE services re-use other existing databases services of the Google Cloud Platform through well defined interfaces such as the Cloud SQL service or the NoSQL services Cloud BigTable & Firestore document database & Firebase Realtime Database**
- **One key property of NoSQL databases is that information stored for each user is not the same and a key differentiator to standard SQL databases that follow a table structure where each information / user is the same**

Google Cloud

[14] Google Cloud – Databases

Google Cloud BigTable – General Characteristica

- Examples
  - ‘Massive data stored’ for Web indexing, Google Earth, and Google Finance
- Characteristica
  - Scalability: Bigtable is designed to ‘reliably scale to PBs of data’ and thousands of machines for processing this data efficiently
  - ‘String-based approach’ NoSQL to support large quantities of texts
  - Shares many ‘implementation strategies’ with traditional databases
  - Possible deployments on top of distributed file systems, e.g. Google file system → ~HDFS (cf. Lecture 5)



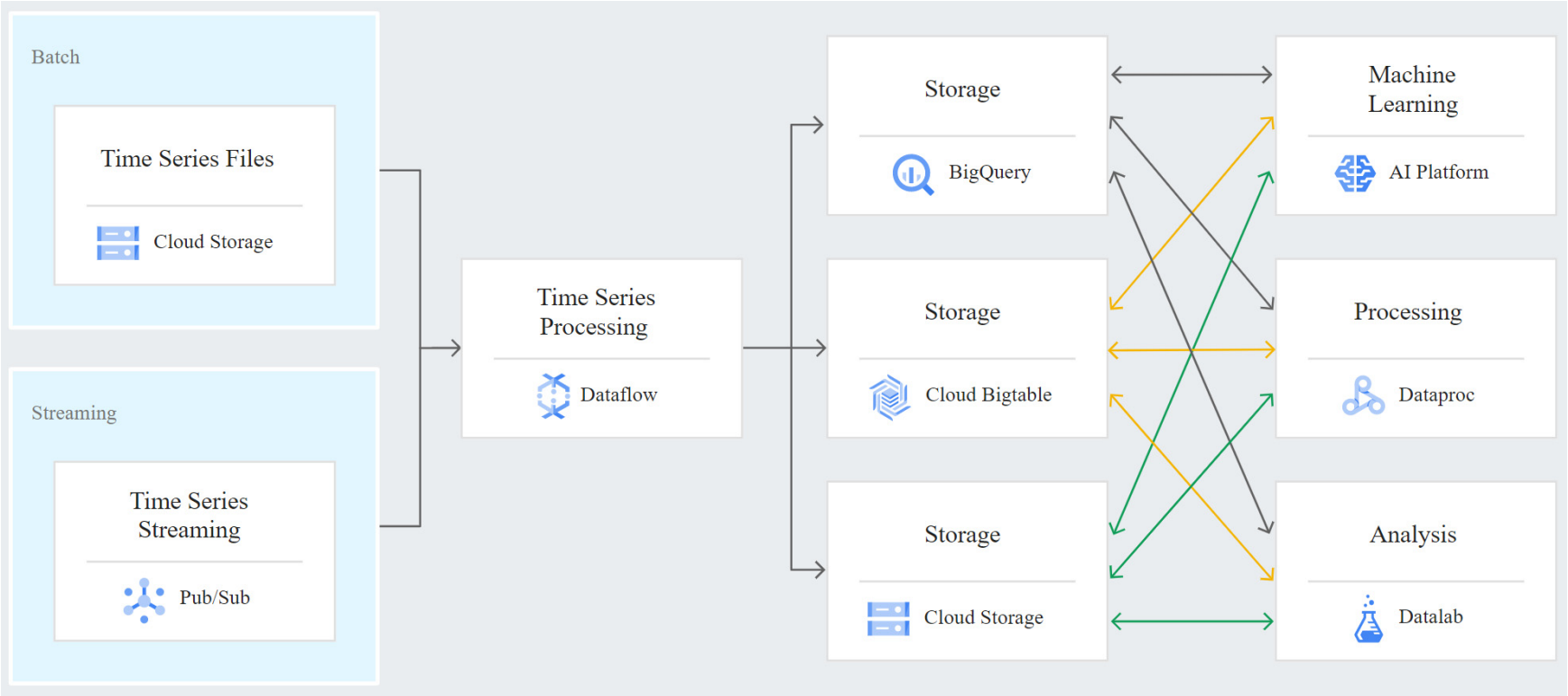
Dow Jones brings key historical events datasets to life with Cloud Bigtable.

- **BigTable is a distributed storage system for managing structured data**
- **Designed to scale to a very large size (e.g. PBs) of data across thousands of commodity servers**
- **BigTable is based on a flexible data model that enables dynamic control over data layout/format**

Google Cloud

[15] Google Cloud – BigTable

PAAS Example – Creating a Scalable System using Machine Learning for Finance

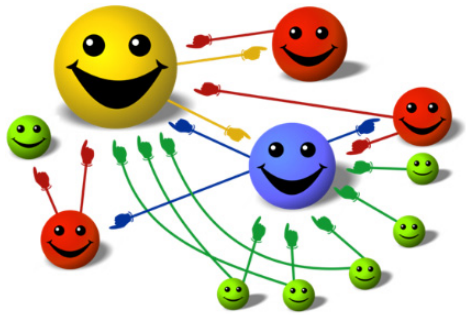


Google Cloud

[15] Google Cloud – BigTable

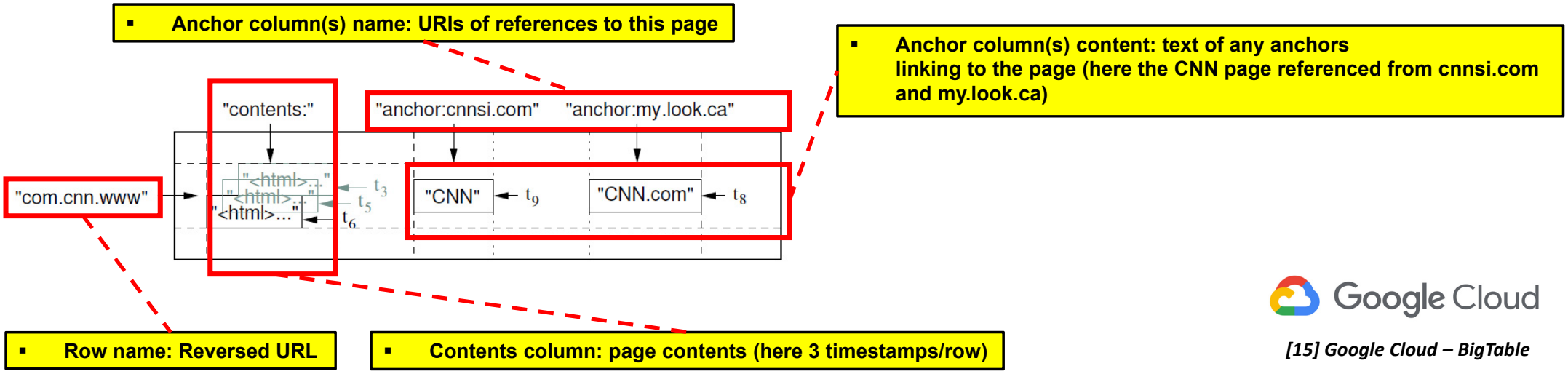
Google BigTable – Data Model

- Sparse, distributed, persistent ‘multidimensional sorted map’
  - Map is indexed by: ‘row key, column key, and a timestamp key’
    (row:string, column:string, time:int64) → string
  - Map values are ‘uninterpreted’ array of bytes
- Example: ‘Storing Web pages’
  - E.g. ‘PageRank needs links to this page’, cf. Lecture 3



[17] Apache Spark

[16] Parallel Programming with Spark

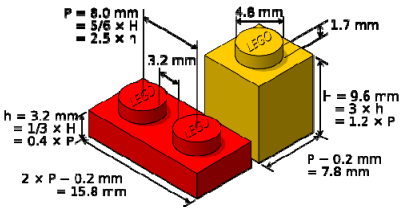


Google Cloud

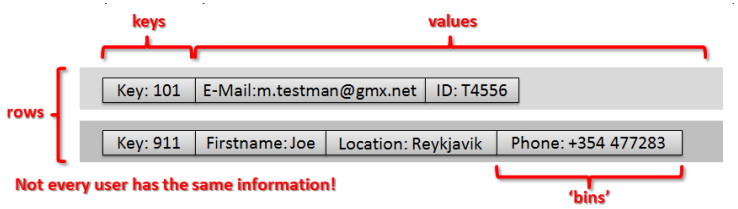
[15] Google Cloud – BigTable

PAAS Building Blocks – Understanding SQL vs. NoSQL

- Simple NoSQL example (e.g. Google Datastore & Google Bigtable)
  - E.g. store flexible key - value pairs vs. Structured Query Language (SQL) use



[6] Lego Bricks



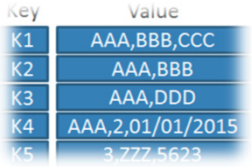
- SQL Databases
  - Good for 'structured data'
  - Relational Database Management Systems
  - SQL DB schema with well-defined columns/rows
- NoSQL Databases
  - Good for 'unstructured data'
  - Easy to deploy & implement
  - Often geographically distributed
  - Designed with 'no schemas'
  - Low data consistency requirements

- **SQL and NoSQL databases are two complementary database approaches and PAAS applications use both of them for different purposes**

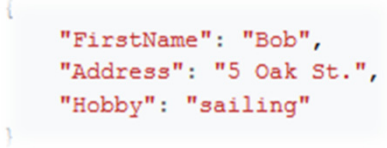
NoSQL Concepts for Databases

- Wide variety of emerging NoSQL databases for ‘specific data’
  - Most commonly agreed classification based on data models (below)
  - More and more NoSQL data-base implementation become available
  - Several closed source, many open source databases available

- Column-based
  - E.g. Apache HBase



- Document-based
  - E.g. MongoDB, CouchDB



- Key-Value-based
  - E.g. Apache Cassandra

- Graph-based
  - E.g. Neo4J

➢ **Lecture 14 provides more details about using graph-based databases as another complementary database to NoSQL and SQL databases**

Understanding NoSQL Databases – Apache HBase Example

- Data model design (cf. BigTable design earlier)
  - Columns can be added anytime 'w/o announcement' to 'Columnfamilies'
  - Allows seamless adding of more 'links to this page'
- Conceptual View

Row Key	Time Stamp	ColumnFamily contents	ColumnFamily anchor
"com.cnn.www"	t9		anchor:cnnsi.com = "CNN"
"com.cnn.www"	t8		anchor:my.look.ca = "CNN.com"
"com.cnn.www"	t6	contents:html = "<html>..."	
"com.cnn.www"	t5	contents:html = "<html>..."	
"com.cnn.www"	t3	contents:html = "<html>..."	

APACHE HBASE

[18] HBASE Web Page

- Physical View
  - No empty fields are stored, difference to Relational Database Management System (RDMS)

Row Key	Time Stamp	Column Family anchor
"com.cnn.www"	t9	anchor:cnnsi.com = "CNN"
"com.cnn.www"	t8	anchor:my.look.ca = "CNN.com"

Row Key	Time Stamp	ColumnFamily "contents:"
"com.cnn.www"	t6	contents:html = "<html>..."
"com.cnn.www"	t5	contents:html = "<html>..."
"com.cnn.www"	t3	contents:html = "<html>..."

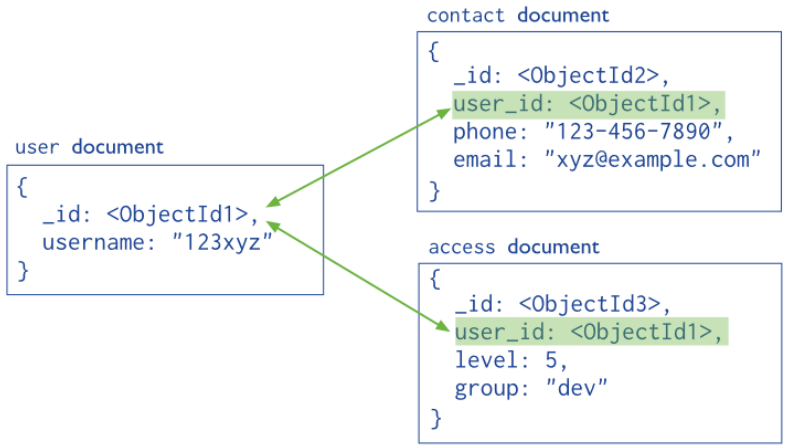
Understanding NoSQL Databases – Apache MongoDB Example

- Document-based Example: MongoDB
  - Selected features: index support, querying, open source, C++, …
  - Data Model: Simple & dynamic schemas with 'JSON-Style documents' (JSON = JavaScript Object Notation)

mongoDB

[19] MongoDB Web Page

```
{
  _id: <ObjectId1>,
  username: "123xyz",
  contact: {
             phone: "123-456-7890",
             email: "xyz@example.com"
           },
  access: {
             level: 5,
             group: "dev"
           }
}
```



Understanding NoSQL Databases – Apache Cassandra NoSQL Database

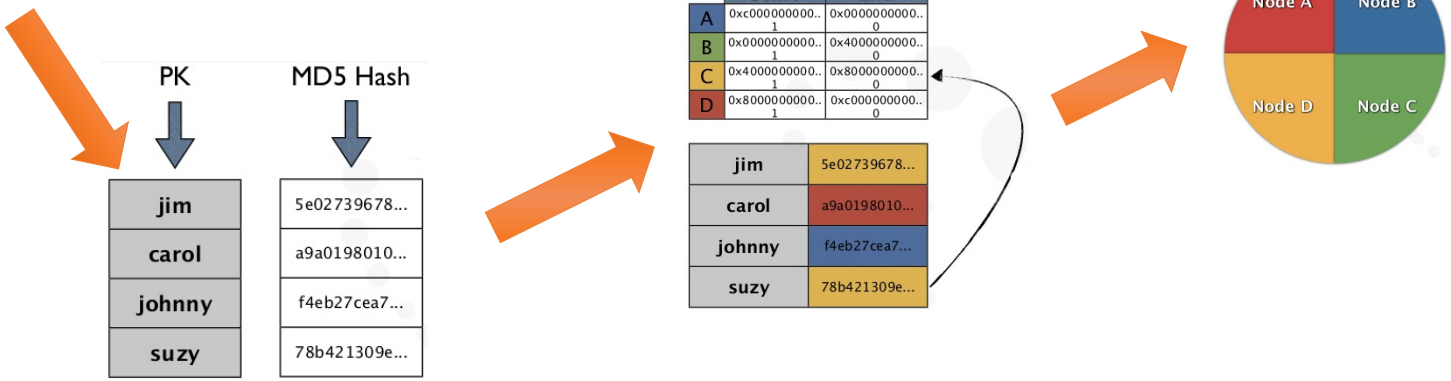
- Key-Value-based Example: Cassandra
  - (in use at Twitter, eBay, Netflix, etc.)
  - Selected features: decentralized – every node equal, elastic, replication
  - Approach: Using primary keys and 'hash values' to determine placement

Cassandra

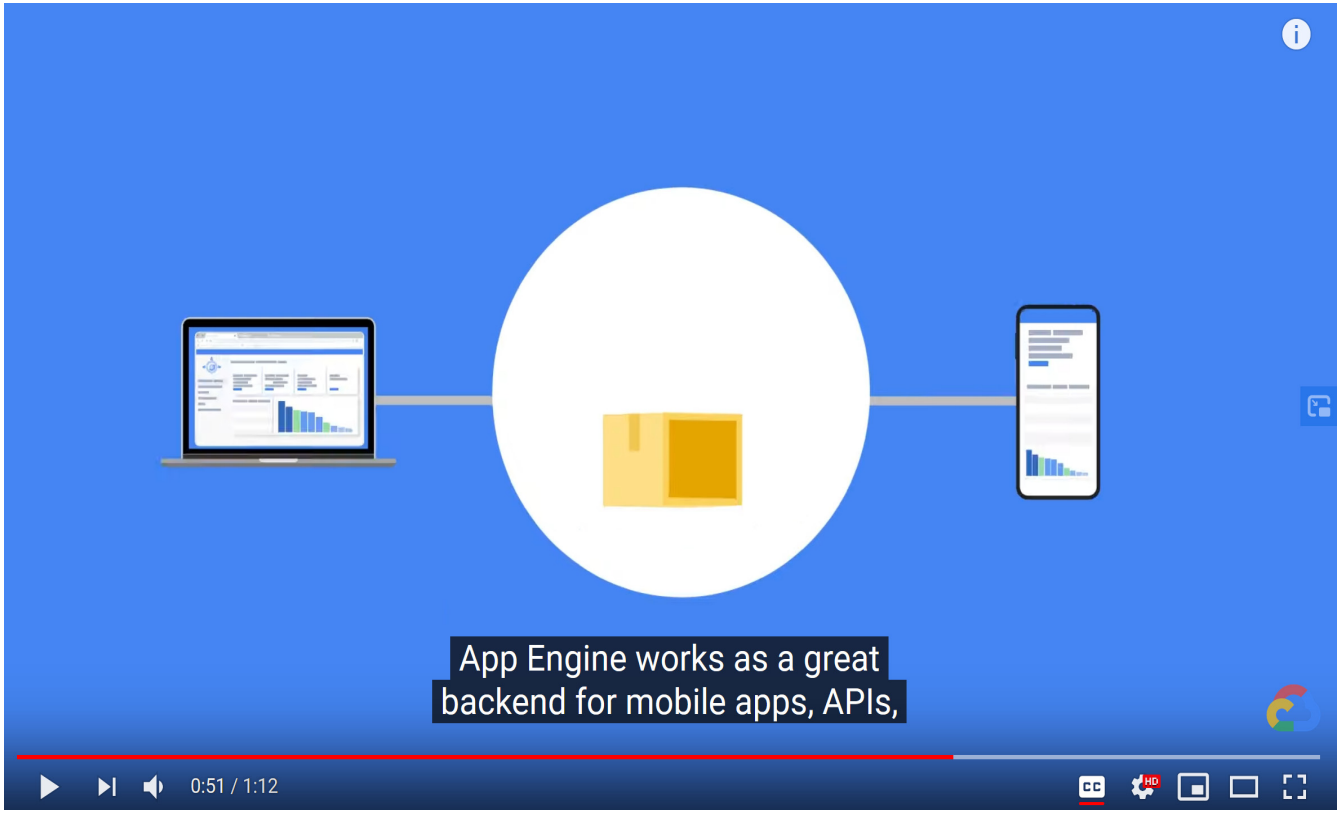
[20] Cassandra Web Page

jim	age: 36	car: camaro	gender: M
carol	age: 37	car: subaru	gender: F
johnny	age:12	gender: M	
suzy	age:10	gender: F	

(compare to relational database that would have lots of missing entries, essentially being especially for social media applications a sparse table)

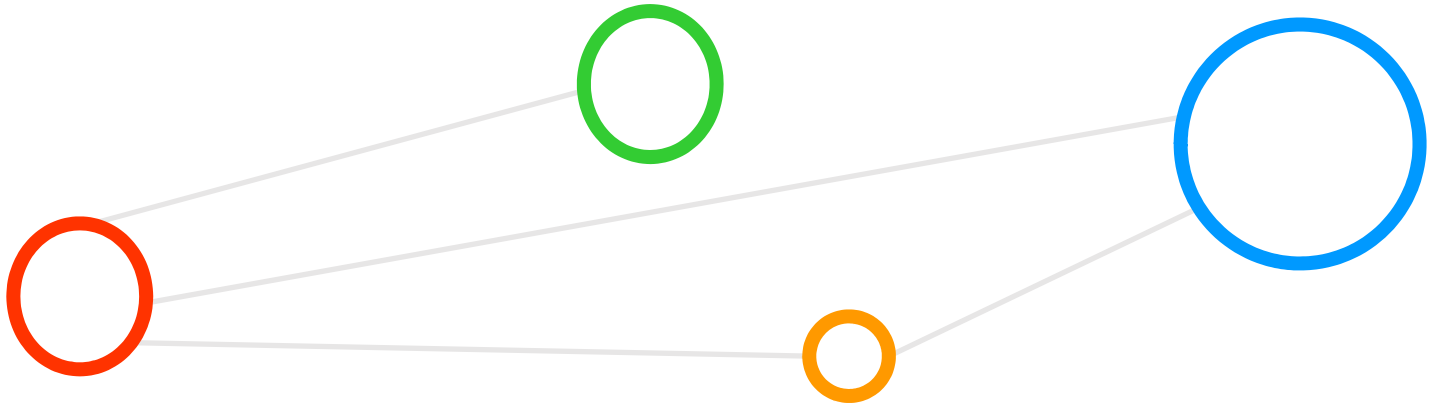


[Video] Google App Engine Summary



[23] YouTube video, Google App Engine in a Minute

Advanced PAAS Topics & Applications



Selected Benefits from PAAS Clouds & Usage for Development of Applications

- Cut coding time
  - On top of IAAS infrastructure: scalable middleware, development tools, business tools
- Add development capabilities without adding staff & Application Lifecycle
  - Selected PAAS services give development teams new capabilities (e.g., machine learning)
  - Efficiently manage the application lifecycle & better maintenance options (e.g., automation)
- Develop for multiple platforms—including mobile—more easily
  - Development options for multiple platforms, such as computers, mobile devices, and browsers making cross-platform apps quicker & easier to develop
- Use sophisticated tools affordably
  - Pay-as-you-go model makes it possible for individuals or organizations to use sophisticated development software & business intelligence & analytics tools
- Support geographically distributed development teams
  - Development environment is via Internet; remote development teams easier

- **The common service areas for PAAS are in the realm of development frameworks, analytics or business intelligence, or any other form of additional supporting services that enhance applications, such as workflows, directory, security, and task scheduling**

[22] MS Azure, What is PAAS?

The Role of PAAS Software Development Kits (SDKs) – GAE Example

Serverless computing > App Engine > Documentation > Standard Environment > Java 8

Rate and review

Setting Up Your Development Environment

Python 2.7/3 | Java 8/11 | PHP 5/7 | Ruby | Go 1.11 / 1.12+ | Node.js

To set up your environment for developing on Java 8:

1. Install the latest release of Java 8.

   See Java 8 Runtime Environment for a list of the supported versions.

2. Install and initialize the Cloud SDK. If you already have the Cloud SDK installed and initialized, run the `gcloud components update` command to update to the latest release.

   The Cloud SDK provides you the `gcloud` command-line tooling for deploying and managing your apps.

   By downloading, you agree to be bound by the Terms that govern use of the Cloud SDK for App Engine.

   Install and initialize the Cloud SDK

3. Install the gcloud component that includes the App Engine extension for Java 8.

   If you used the `apt` or `yum` package managers to install the Cloud SDK, use those same package managers to install the gcloud component.

   Otherwise, use the following command:

```
gcloud components install app-engine-java
```

Developer Tools →

Cloud SDK
Command line tools and libraries for Google Cloud.

Container Registry
Private Docker storage for container images on Google Cloud.

Cloud Build
Continuous integration and continuous delivery platform.

Cloud Source Repositories
Private Git repository to store, manage, and track code.

Cloud Scheduler
Cron job scheduler for task automation and management.

Tekton
Kubernetes-native resources for declaring CI/CD pipelines.

Cloud Tasks
Task management service for asynchronous task execution.

Cloud Code
IDE support to write, run, and debug Kubernetes applications.

Tools for Visual Studio
Tools to enable development in Visual Studio on Google Cloud.

Tools for Eclipse
Plugin for Google Cloud development inside the Eclipse IDE.

Cloud Code for IntelliJ
IDE support for debugging production cloud apps inside IntelliJ.

Not seeing what you're looking for?

See all developer tools

Give permission to Cloud Build

When you deploy your app, App Engine uses Cloud Build to build the app into a container and deploy the container to the runtime. Cloud Build does not have permission to deploy Java 8 apps by default, so you need to give Cloud Build permission to deploy apps in your project.

[24] Google Cloud – Tools

[21] Google Cloud – GAE SDK

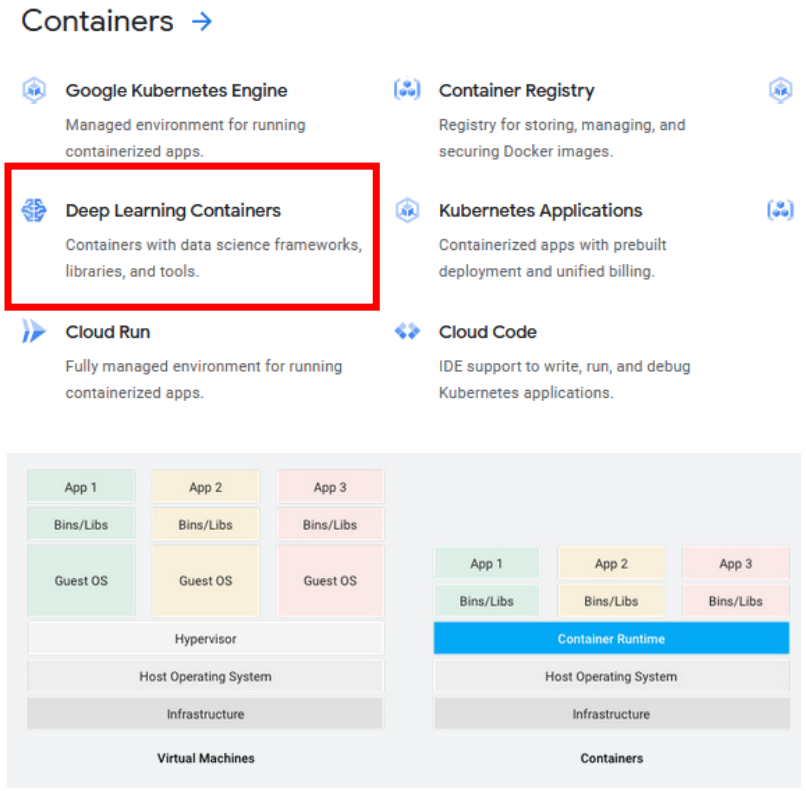
➢ **Lecture 12 provides more details about using containers in conjunction with Cloud computing for various different application areas**

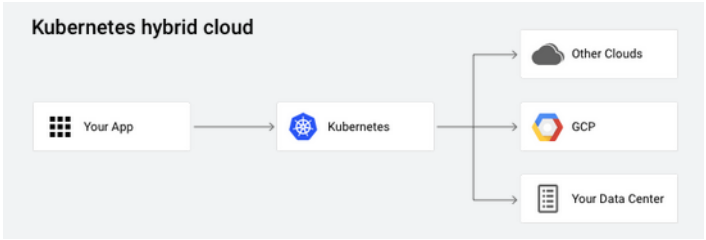
Public Clouds using Virtualization in Large Data Centers – Revisited

Provider	AWS	Microsoft Azure	GAE
Compute cloud with virtual cluster of servers	x86 instruction set, Xen VMs, resource elasticity allows scalability through virtual cluster, or a third party such as RightScale must provide the cluster	Common language runtime VMs provisioned by declarative descriptions	Predefined application framework handlers written in Python, automatic scaling up and down, server failover inconsistent with the web applications
Storage cloud with virtual storage	Models for block store (EBS) and augmented key/blob store (SimpleDB), automatic scaling varies from EBS to fully automatic (SimpleDB, S3)	SQL Data Services (restricted view of SQL Server), Azure storage service	MegaStore/BigTable
Network cloud services	Declarative IP-level topology; placement details hidden, security groups restricting communication, availability zones isolate network failure, elastic IP applied	Automatic with user's declarative descriptions or roles of app. components	Fixed topology to accommodate three-tier web app. structure, scaling up and down is automatic and programmer-invisible

[2] Distributed & Cloud Computing Book

Google Cloud Example – Containerized Applications & Libraries for Developers





- **An additional PAAS element is not only the wide variety of cloud services required for building applications, but also the environment to build those applications and to deploy and maintain them in a scalable way**
- **Containers help tremendously to organize PAAS software elements (e.g., deep learning containers to manage the massive amount of versions of multiple relevant deep learning software frameworks & underlying libraries)**
- **The Google Cloud offers Deep Learning Containers with data science frameworks, libraries, and tools**

Google Cloud

[25] Google Cloud – Containers

➢ **Lecture 12 provides more details about using containers in conjunction with Cloud computing for various different application areas**

Google Cloud Example – Data Analytics (aka ‘Smart Analytics‘)

Data Analytics →

BigQuery
Data warehouse for business agility and insights.

Looker
Platform for BI, data applications, and embedded analytics.

Dataflow
Streaming analytics for stream and batch processing.

Pub/Sub
Messaging service for event ingestion and delivery.

Dataproc (cf. Lecture 3&5)
Service for running Apache Spark and Apache Hadoop clusters.

Cloud Data Fusion
Data integration for building and managing data pipelines.

Cloud Composer
Workflow orchestration service built on Apache Airflow.

Data Catalog
Metadata service for discovering, understanding and managing data.

Dataprep
Service to prepare data for analysis and machine learning.

Google Data Studio
Interactive data suite for dashboarding, reporting, and analytics.

Google Marketing Platform
Marketing platform unifying advertising and analytics.

Cloud Life Sciences
Tools for managing, processing, and transforming biomedical data.

- **The GAE services re-use other existing Big Data services of the Google Cloud Platform through well defined interfaces such as the Google BigQuery or Google Dataproc for Apache Spark & Hadoop**
- **Google Dataproc enables scalable & automated cluster management for Apache Spark with quick deployment, logging, and monitoring in order to focus on data analysis and not on infrastructure**

vodafone

Vodafone Group moves 600 on-premises Apache Hadoop servers to the cloud.

Story highlights

- Migrated on-premises Apache Hadoop to Google Cloud
- 226 models running in production
- Two months to roll out Google Cloud to first country

Industry

Telecommunications

49:57

Twitter moved from on-premises Hadoop to Google Cloud to more cost-effectively store and query tweets, users, impressions, and more.

pandora®

50:51

Pandora migrated 7 PB+ of data from their on-premises Hadoop data lake to Google Cloud to unlock processing scale and help lower costs.

METRO

00:00

Spinning up and down Dataproc clusters helped METRO reduce infrastructure costs by 30% to 50%.

Google Cloud

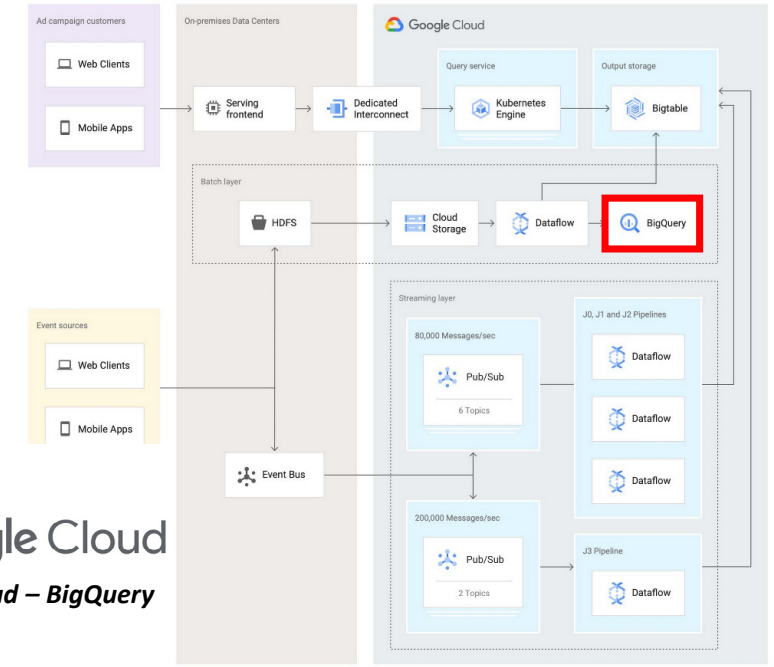
[26] Google Cloud – Data Analytics

Google Cloud Example – BigQuery Use Case Application from Twitter

- Business Case
  - Twitter advertising platform serves billions of ad engagement events
  - Each event affects hundreds of downstream aggregate metrics
- Challenges
  - Enable/Measure for advertisers the user engagement and track ad campaigns efficiency
  - Twitter Dashboards that can aggregate millions of metrics per second in near-real time
- Approach
  - Twitter Revenue Data Platform migrated on-premise architecture to Google Cloud

CASE STUDY

Twitter modernizes ad engagement platform; combines millions of metrics/second.



Google Cloud

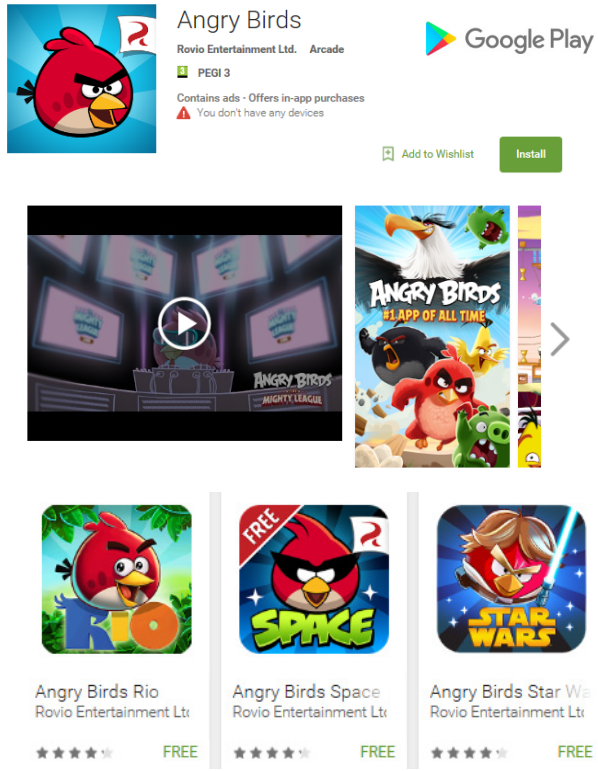
[29] Google Cloud – BigQuery

- **Google Big Query is a fast solution for large-scale big data analytics to find meaningful insights**
- **Google Big Query enables users to take advantage of the power of Google search engine for data**
- **Twitter uses BigQuery serverless and highly scalable data warehouse, to support ad-hoc and batch queries**
- **Twitter uses Cloud Bigtable low-latency, fully managed NoSQL database, to serve as a back end for online dashboards & consumer APIs.**

Rovio Games Example – Angry Birds Game App

- Business Case
  - Angry Birds destroy different pigs
  - Each bird has unique talents
  - Revenue: In-app purchases & ads
- Challenges
  - More than 140 million downloads & users
  - Robust capabilities and scalability to deliver a superior user experience for various apps
- Approach
  - Own infrastructure would be to costly
  - Services that are required already implemented in Google Cloud
  - Rovio Web games tend to be popular immediately (no scaling over time)
  - Use(d) Google Cloud Platform: GAE, Memcache API, Google Cloud Datastore, …

ROVIO



[27] Google Play Angry Birds

[28] Rovio Case Study

Google Cloud Platform MemCache – Angry Birds Application Example

- Rovio application use case
  - Apps often face ‘similiar situations‘ that needs the same data
  - Boost performance by providing tempory high-speed data access to same query result data in the Angry Birds app
  - High performance scalable Web applications such as Angry Birds could use a distributed in-memory data cache in front persistent storage

- **High-Replication Datastore** for scalable, long-term storage of game data
- **Memcache API** to boost performance by providing temporary, high-speed data access through a high-performance memory cache
- **Task queues** to run certain complex operations in the background, improving game responsiveness for users
- **Users API** to authenticate users with their Google usernames and passwords, which provides a seamless experience when accessing a game

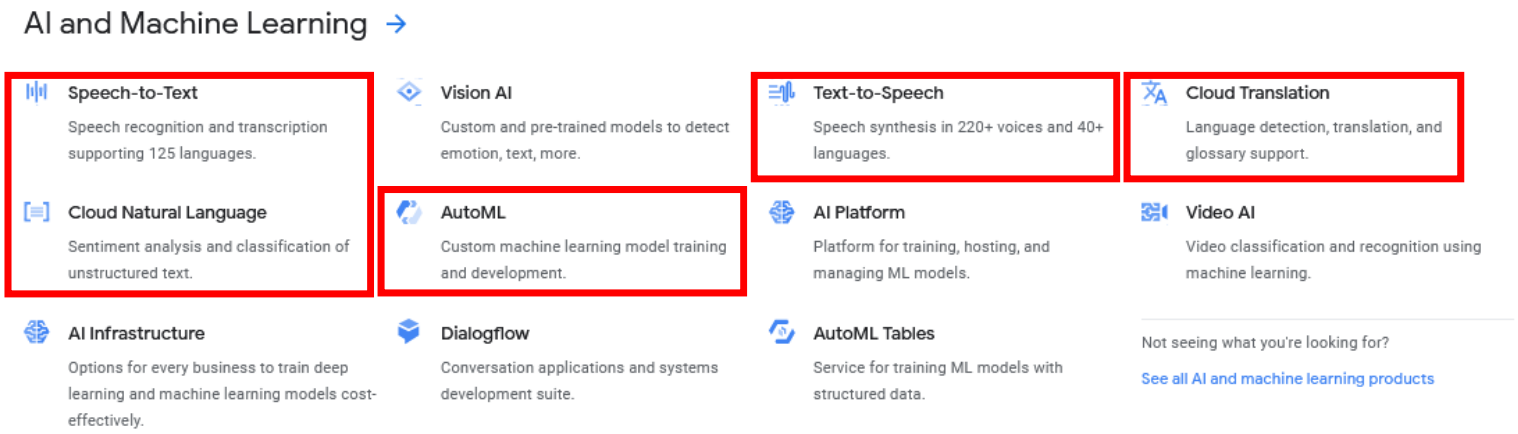
ROVIO

[27] Google Play Angry Birds

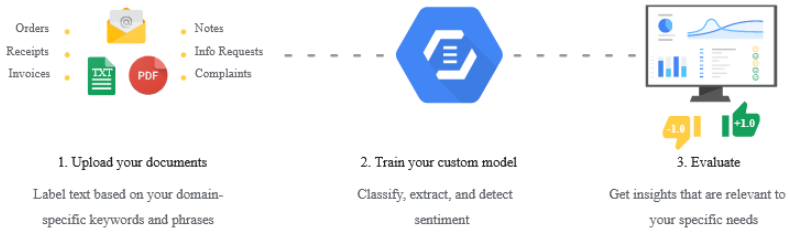
- GAE Memcache
  - Idea: check the memcache first for results of older queries
  - Only perform the datastore query if the results are absent or expired
  - Session data, user preferences, and other data returned by queries for Apps are good candidates for caching and can benefit from Memcache
  - Other example are temporary values (e.g. rain yes/no, might disappear)

- **GAE MemCache enables speed up of common datastore queries by using a global cache approach**
- **GAE MemCache can cache the results if many requests make same query with same parameters**

Google Cloud Example – Machine Learning & Natural Language Processing (NLP)



- Example:
  - Sentiment analysis



- **Sentiment analysis (aka opinion mining or emotion AI) refers to the use of natural language processing (NLP), text analysis, computational linguistics to systematically identify, extract, quantify, and study affective states and subjective information (e.g., my product is bad or good in customer experience)**
- **Sentiment analysis is widely applied to voice of the customer materials such as reviews & survey responses, online & social media, & healthcare materials for applications that range from marketing to customer service to clinical medicine**

modified from [30] Wikipedia

Google Cloud

[34] Google Cloud – Solutions

Google Cloud Example – ‘Solutions View‘ for a wide range of Industries

Industry Solutions →

Reduce cost, increase operational agility, and capture new market opportunities.

Retail
Analytics and collaboration tools for the retail value chain.

Financial Services
Computing, data management, and analytics tools for financial services.

Healthcare and Life Sciences
Health-specific solutions to enhance the patient experience.

Media and Entertainment
Solutions for content production and distribution operations.

Telecommunications
Hybrid and multi-cloud services to deploy and monetize 5G.

Gaming
AI-driven solutions to build and scale games faster.

Manufacturing
Migration and AI tools to optimize the manufacturing value chain.

Energy
Multi-cloud and hybrid solutions for energy companies.

Government
Data storage, AI, and analytics solutions for government agencies.

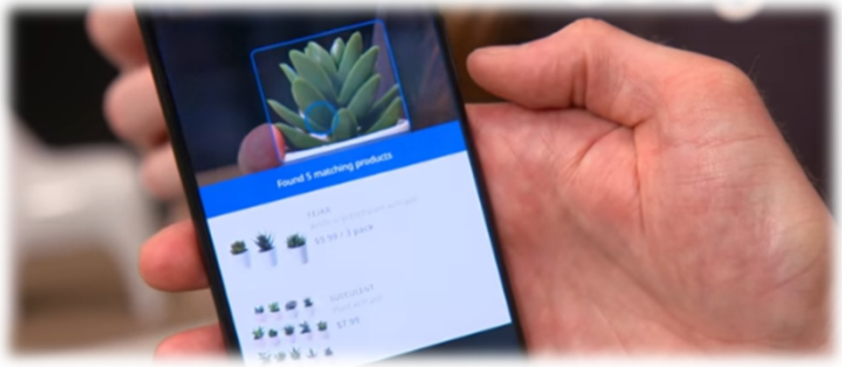
Education
Teaching tools to provide more engaging learning experiences.

Small and Medium Business
Explore SMB solutions for web hosting, app development, AI, analytics, and more.

Cloud Natives
Resources and solutions for cloud-native organizations.

Carrefour · GANT · ULTA · wayfair · KOHL'S · DSW · Loblaws · Etsy · woolworths · shopify

- Example:
  - Image Analysis and directly point to product



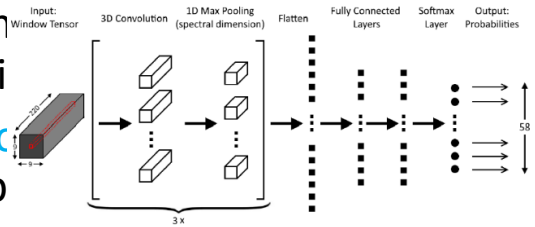
Google Cloud

[34] Google Cloud – Solutions

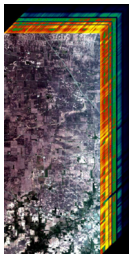
➢ **Lecture 12 provides more details about using containers in conjunction with Cloud computing for various different application areas**

Google Cloud Example – Cloud Tensor Processing Unit

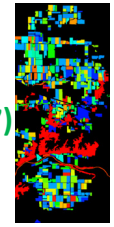
- Hardware-driven PAAS Service
  - Offers accelerator technology to speed up machine learning workloads on Google Clouds
  - Enables TensorFlow (cf. Lecture 6) deep learnin on a speciali
  - Provides 'Cl with high-sp network interconnects



(convolutional neural network deep learning model using tensors, cf. Lecture 6)

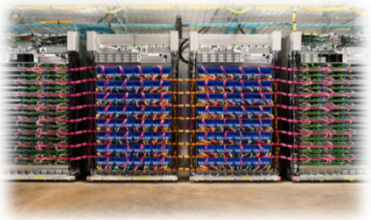


(land cover detection, cf. Lecture 6 & 7)

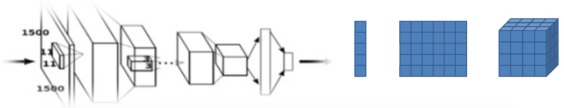


TensorFlow

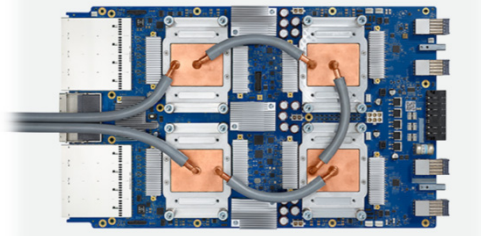
[37] Tensorflow



[35] Google Cloud TPU



[36] Big Data Tips, What is a Tensor?



- **The concept of the Google Cloud Tensor Processing Unit (TPU) is to train & run machine learning models faster using specialized hardware chip as Application-specific Integrated Circuit (ASIC)**
- **Google Cloud TPU is very useful to train & run computationally demanding deep learning models with TensorFlow in clouds**

Other PAAS Providers & Disadvantages

- **PAAS disadvantages are that application developers are being locked in to a certain platform (aka 'vendor lock') and that they may not be able to use known tools (e.g. ORACLE DB) with services**

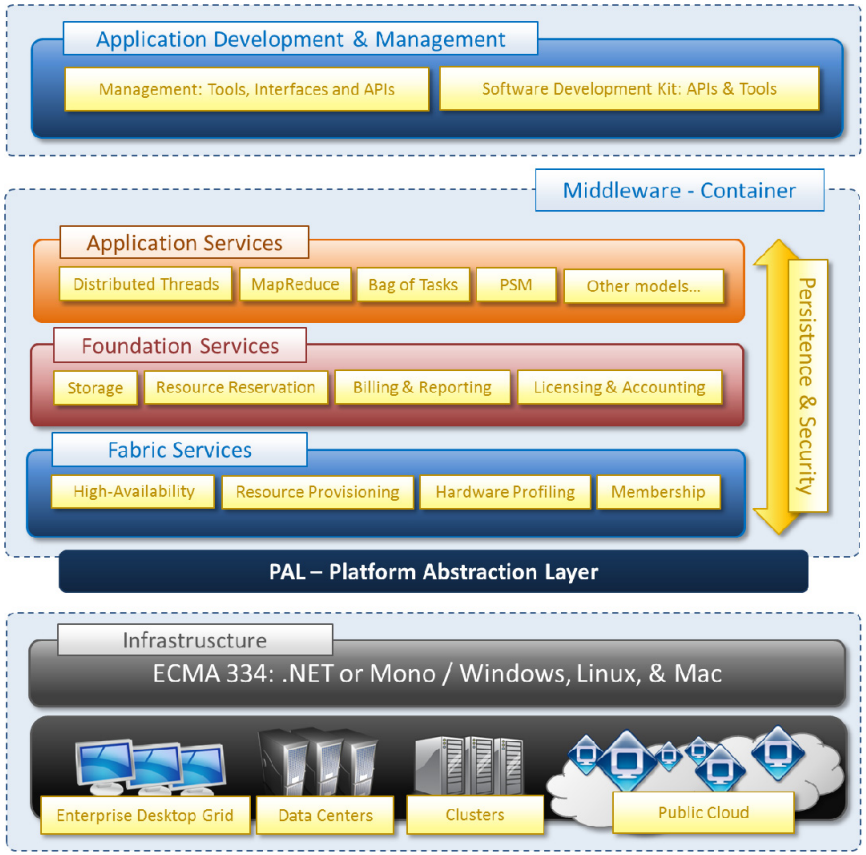
Cloud Name	Languages and Developer Tools	Programming Models Supported by Provider	Target Applications and Storage Option
Google App Engine	Python, Java, and Eclipse-based IDE	MapReduce, web programming on demand	Web applications and BigTable storage
Salesforce.com's Force.com	Apex, Eclipse-based IDE, web-based Wizard	Workflow, Excel-like formula, Web programming on demand	Business applications such as CRM
Microsoft Azure	.NET, Azure tools for MS Visual Studio	Unrestricted model	Enterprise and web applications
Amazon Elastic MapReduce	Hive, Pig, Cascading, Java, Ruby, Perl, Python, PHP, R, C++	MapReduce	Data processing and e-commerce
Aneka	.NET, stand-alone SDK	Threads, task, MapReduce	.NET enterprise applications, HPC

[2] Distributed & Cloud Computing Book

PAAS Provider: Manjrasoft Aneka – Standards Example & Vendor Dominance

[17] Manjrasoft Web page

(neutral key player from Australia)



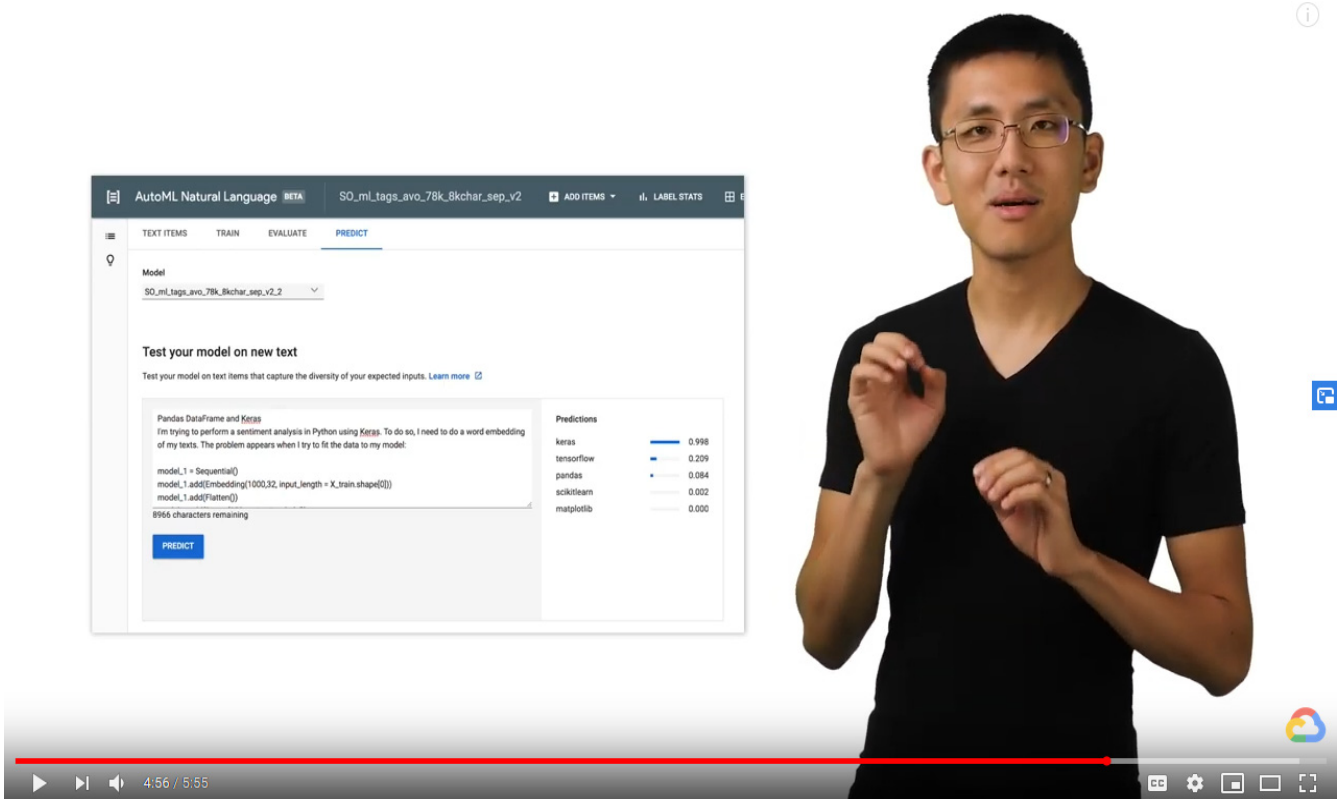
(Approach good, but not extremely succesful because of vendor dominance of key organizations in Cloud markets)

(Standard ECMA-334: is C# language specification)

(ECMA International: European Association for standardizing information and communication systems)

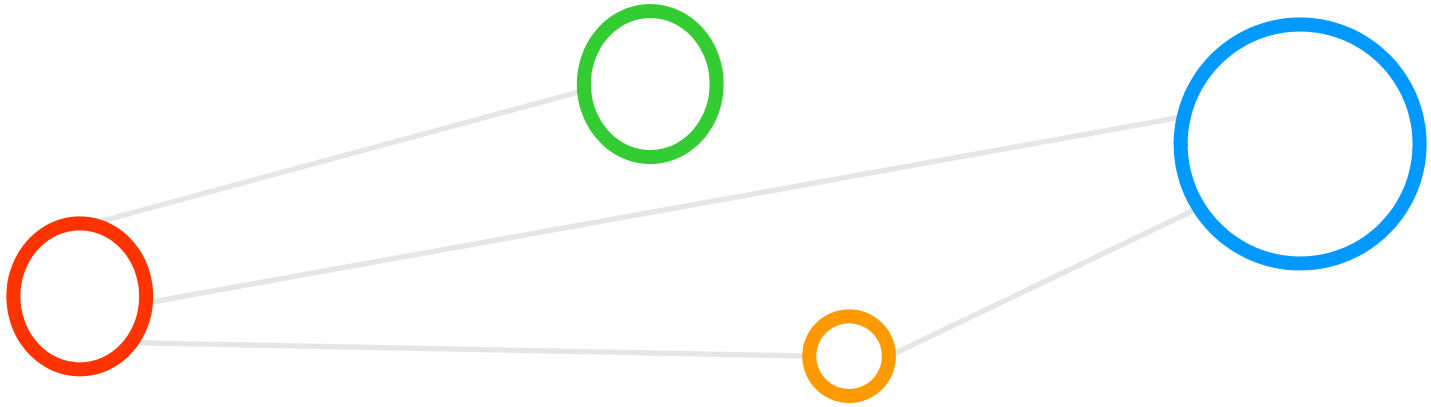
- **Use of standard programming languages and interface standards avoids or reduces vendor-locks in Cloud Computing but standard developments and standard adoptions are often given through de-facto standards from strong vendors (e.g., Amazon S3 interface)**

[Video] Using AutoML Natural Language for custom text classification



[32] YouTube video, Using AutoML Natural Language for custom text classification

Lecture Bibliography



Lecture Bibliography (1)

- [1] Wikipedia, ‚Cloud Computing‘, Online: http://de.wikipedia.org/wiki/Cloud-Computing
- [2] K. Hwang, G. C. Fox, J. J. Dongarra, ‘Distributed and Cloud Computing’, Book, Online: http://store.elsevier.com/product.jsp?locale=en_EU&isbn=9780128002049
- [3] Amazon Web Services Web Page, Online: https://aws.amazon.com
- [4] Accenture, Hadoop Deployment Comparison Study, Online: http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-Hadoop-Deployment-Comparison-Study.pdf
- [5] AWS S3 Pricing Models, Online: https://aws.amazon.com/s3/pricing
- [6] Lego Bricks, Online: https://commons.wikimedia.org/wiki/File:Lego_dimensions.svg
- [7] Google Cloud, Online: https://cloud.google.com/
- [8] Google Cloud – GAE, Online: https://cloud.google.com/appengine
- [9] Google Cloud – Compute Products, Online: https://cloud.google.com/products/compute/
- [10] Google Cloud – GAE Pricing, Online: https://cloud.google.com/appengine/pricing
- [11] Wikipedia, ’NoSQL’, Online: http://en.wikipedia.org/wiki/NoSQL

Lecture Bibliography (2)

- [12] Google Cloud Platform Case Study, ‘Philipps‘, Online: https://cloud.google.com/customers/philips/
- [13] Google Cloud Datastore, Online: https://cloud.google.com/datastore
- [14] Google Cloud Databases, Online: https://cloud.google.com/products/databases
- [15] Google Cloud BigTable, Online: https://cloud.google.com/bigtable
- [16] M. Zaharia, ‘Parallel Programming with Spark‘, 2013, Online: http://ampcamp.berkeley.edu/wp-content/uploads/2013/02/Parallel-Programming-With-Spark-Matei-Zaharia-Strata-2013.pdf
- [17] Apache Spark Web page, Online: http://spark.apache.org/
- [18] Apache HBase, Online: http://hbase.apache.org/
- [19] MongoDB, Online: http://www.mongodb.org/
- [20] Apache Cassandra, Online: http://cassandra.apache.org/
- [21] Google App Engine Java SDK Download, Online: https://cloud.google.com/appengine/docs/standard/java/setting-up-environment
- [22] MS Azure, ‘What is PAAS‘, Online: https://azure.microsoft.com/en-us/overview/what-is-paas/

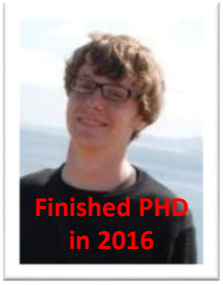
Lecture Bibliography (3)

- [23] YouTube Video, ‘Google App Engine in a Minute‘, Online: https://www.youtube.com/watch?v=Xuf3J6SKVV0
- [24] Google Cloud, Tools, Online: https://cloud.google.com/products/tools
- [25] Google Cloud, Containers, Online: https://cloud.google.com/containers
- [26] Google Cloud, Smart / Data Analytics, Online: https://cloud.google.com/solutions/smart-analytics
- [27] Google Play Angry Birds, Online: https://play.google.com/store/apps/details?id=com.rovio.angrybirds&hl=en
- [28] Google Cloud Platform Case Study, ‘Rovio‘, Online: https://cloud.google.com/files/Rovio.pdf
- [29] Google Cloud BigQuery, Online: https://cloud.google.com/bigquery
- [30] Google Cloud BigQuery Twitter Use Case, Online: https://cloud.google.com/blog/products/data-analytics/modernizing-twitters-ad-engagement-analytics-platform
- [31] Wikipedia, ‘sentiment analysis‘, Online: https://en.wikipedia.org/wiki/Sentiment_analysis
- [32] YouTube Video, ‘‘, Online: https://www.youtube.com/watch?v=ieaqfU1BwJ8
- [33] Manjrasoft Web page, Online: http://www.manjrasoft.com/aneka_architecture.html

Lecture Bibliography (4)

- [34] Google Cloud, ‘Solutions‘, Online: https://cloud.google.com/solutions
- [35] Google Cloud TPU Products, Online: https://cloud.google.com/tpu/
- [36] Big Data Tips, ‘What is a Tensor?‘, Online: http://www.big-data.tips/what-is-a-tensor
- [37] Tensorflow Deep Learning Framework, Online: https://www.tensorflow.org/

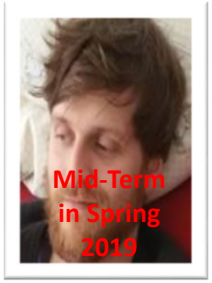
Acknowledgements – High Productivity Data Processing Research Group



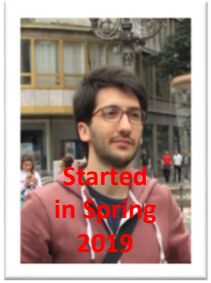
PD Dr.
G. Cavallaro

Senior PhD
Student A.S. Memon

Senior PhD
Student M.S. Memon



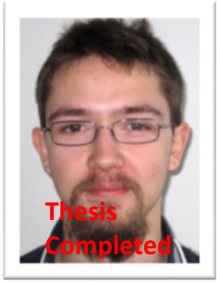
PhD Student
E. Erlingsson



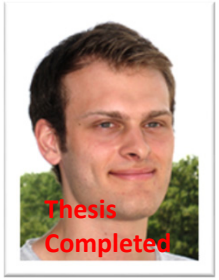
PhD Student
S. Bakarat

PhD Student
R. Sedona

Dr. M. Goetz
(now KIT)



MSc M.
Richerzhagen
(now other division)



MSc
P. Glock
(now INM-1)

MSc
C. Bodenstein
(now
Soccerwatch.tv)

MSc Student
G.S. Guðmundsson
(Landsverkjun)

This research group has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 763558 (DEEP-EST EU Project)


Data
Science
Big
Research
science
analysis
methods
big
simulations
processing
technologies
computing
European
TB
Scientific
computational
research
Computational
Climate
modelling
structures
databases
HPC
Juelich
datasets
performance
Supercomputing
access
Centre
climate
Computing
nodes
centers
storage
ENES
EUDAT
management