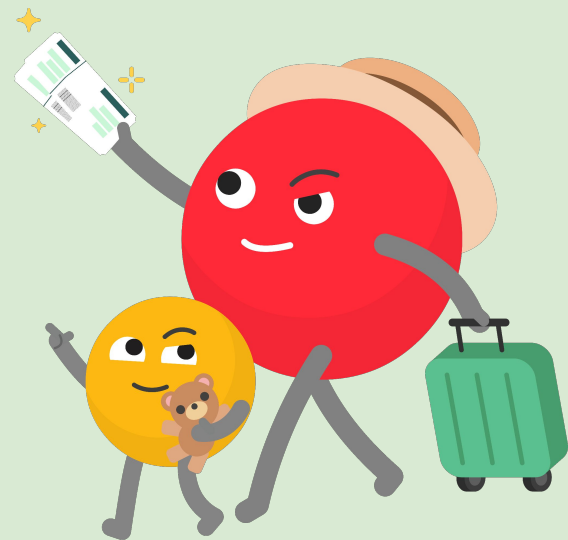


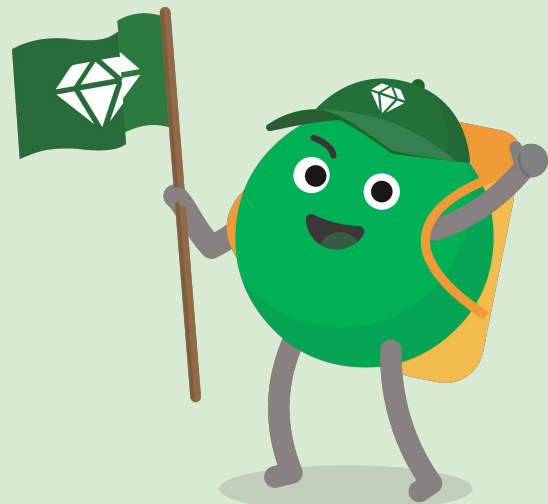
Knowledge Sharing Session

070153709 Big Data | **Agoda x KMUTNB**



Agenda

1. Quick Introduction
2. Big Data in Agoda: Our Tech Stack 101
3. PySpark Workshop
 - a. Discussion Time!
4. Q&A + Closing Words





Slogan

See The World For Less

Purpose

Bridging the World
Through Travel

Agoda, a digital travel platform, helps anyone **see the world for less** with its **great value deals** on a global network of 4.5M hotels and holiday properties worldwide, plus flights, activities, and more... Agoda is part of Booking Holdings (Nasdaq: **BKNG**) and employs more than 7,000 staff in 27 markets, dedicated to leveraging best-in-class technology to make travel even easier.




About Me

- Graduated from International School of Engineering (ISE), Chulalongkorn University
- Worked as a Data Analyst at Agoda for almost 2 years now under the **Pricing Optimization** team
- My 1st time giving a seminar!

Welcome! You can **enter dates** at the search box to continue booking at HOTEL MYSTAYS

Ueno-Inaricho: Lowest price includes Free cancellation



HOTEL MYSTAYS Ueno-Inaricho
 ★★☆☆ Ueno, Tokyo - 0.2 km to center

This property offers:
Free WiFi

7.9 Very good
 2,831 reviews

2 ฿ 4620 applied
 Per night before taxes and fees
~~5,594~~ -45%
~~฿ 3,422~~
 Subject to Cashback Terms
฿ 3,055
 + FREE CANCELLATION

Coupon applicable Cashback Rewards: ฿ 1,835

[See all](#)



Hotel Name

Check In

Check Out

Wednesday

Monday

1 room

Overview

Rooms

Trip recommendations

Facilities

Reviews

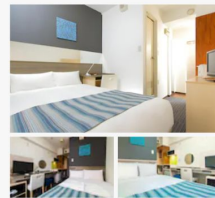
Location

Policies

from **฿ 3,055**

Room Offer

Standard Small Double Room - Smoking, Optional Cleaning for Additional Cost

Excellent 8.2
Room comfort & quality

Room photos and details

- 1 semi double bed
- Room size: 12 m²/129 ft²
- City view
- Balcony/terrace
- Smoking allowed
- Shower and bathtub
- Kitchenette
- [See all room facilities](#)

- ✓ Free cancellation before February 26, 2025
- ✓ Pay nothing until February 24, 2025
- ✓ Free WiFi

Rewards & discounts

Coupon applicable

Cashback Rewards: ฿ 1,835



฿ 4620 applied

~~฿ 5,594~~ -45%
฿ 3,422

Subject to Cashback Terms

฿ 3,055

Per night before taxes and fees

1

Book now

No payment
today

Our last room!

- ✓ Free cancellation before March 25, 2025
- ✓ Pay nothing until March 23, 2025
- ✓ Free WiFi

Rewards & discounts

Coupon applicable

Cashback Rewards: ฿ 2,033



฿ 5120 applied

~~฿ 5,594~~ -39%
฿ 3,792

Subject to Cashback Terms

฿ 3,386

Per night before taxes and fees

1

Book now

No payment
today

Price may increase!

Our last room!

- ✓ Non-refundable (Low price!)
- ✓ Book and pay now
- ✓ Free WiFi

Rewards & discounts

Coupon applicable

Cashback Rewards: ฿ 5



฿ 2939 applied

~~฿ 5,594~~ -36%
฿ 3,605

Subject to Cashback Terms

฿ 3,604

Per night before taxes and fees

1

Book now

Price may increase!

Our last room!

Benefits and
Extra Discounts

Prices per day

Competitive marketplace and price comparison tools educate customers to search for best price

Centara Watergate Pavilion Hotel Bangkok

[Close](#)
[Overview](#) [Prices](#) [Reviews](#) [Photos](#) [About](#)

4.4 ★★★★★ (3.6K) • 4-star hotel • Eco-certified
567 Ratchaprarop Rd, Makkasan, Ratchathewi, Bangkok 10400 • 02 625 1234

DEAL THB 2,676
16% less than usual

[Website](#) [Directions](#) [Save](#) [Share](#) [Book a room](#)

[View more photos](#)

Check-in Check-out

Nightly total

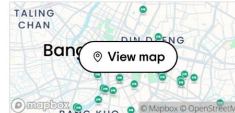
Sponsored • Featured options

Centara Watergate Pavilion Hotel Bangkok <small>Official site</small> THB 3,413 THB 2,901 Visit site <small>Member rate, save 15%</small>	
	Superior City View Twin 1 single bed - 2 guests THB 2,901 Visit site
	Deluxe City View King 1 king bed - 2 guests THB 3,602 Visit site
	Agoda 2 guests THB 2,676 Visit site
	Booking.com 2 guests THB 3,072 Visit site
	Priceline 2 guests THB 2,948 Visit site

[Discover](#) [Trips](#) [Review](#) [USD](#) [Sign in](#)
[Bangkok](#) [Hotels](#) [Things to Do](#) [Restaurants](#) [Flights](#) [Vacation Rentals](#) [Cruises](#) [Rental Cars](#) [Forums](#)

Asia > Thailand > Bangkok > Bangkok Hotels > Top Hotel Deals in Bangkok, Thailand

Best Bangkok, Thailand Hotel Specials & Deals



Top Hotel Deals in Bangkok, Thailand

Check In Check Out Guests

Deals

- ☐ Fully refundable
- ☐ No prepayment needed
- ☐ Properties with Special Offers

Price

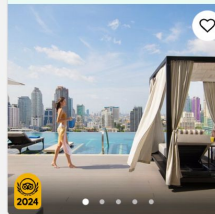


5,548 properties in Bangkok

Sort by: **Featured**

Marriott Bonvoy member rate available

[Sign in to view](#)



1. Bangkok Marriott Hotel The Surawongse

\$192

Fully refundable
No prepayment needed

[View deal](#)

Booking.com
\$193
Trip.com
\$192
Vio.com
\$192

[View all deals from \\$192](#)

2,289 reviews

[Special offer](#)
[Visit hotel website](#)

To make profit, Agoda must optimize price according to the market elasticity

Price	Profit per Room	Room Sold	Total Profits
\$120	\$30	15	\$450
\$100	\$10	50	\$500
\$95	\$5	60	\$300

An (Over)simplified Example of Optimization Function

1

Assume **Demand function** $Q(P)$ is linear in relation to price (P) . $Q(P) = a - bP$ where

- (a) is the intercept, representing demand when the price is zero.
- (b) is the slope, indicating how sensitive demand is to changes in price.

2

Assume **Cost Function** is also linear $C(Q) = cQ + F$ where

- (c) is the variable cost per room sold.
- (F) is the fixed costs regardless of room sales.

3

$$\begin{aligned}
 \text{Profit Function } Pr &= \text{Revenue} - \text{Cost} \\
 &= P \cdot Q(P) - C(Q(P)) \\
 &= P(a - bP) - (cQ(P) + F) \\
 &= P(a - bP) - c(a - bP) - F \\
 &= aP - bP^2 - ca + cbP - F
 \end{aligned}$$

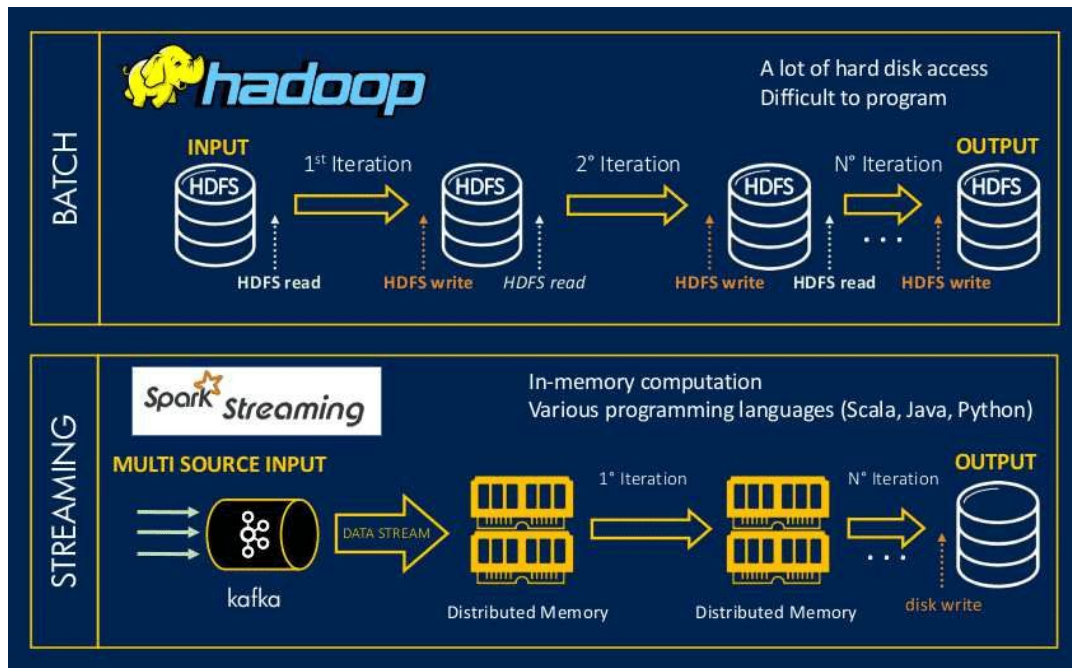
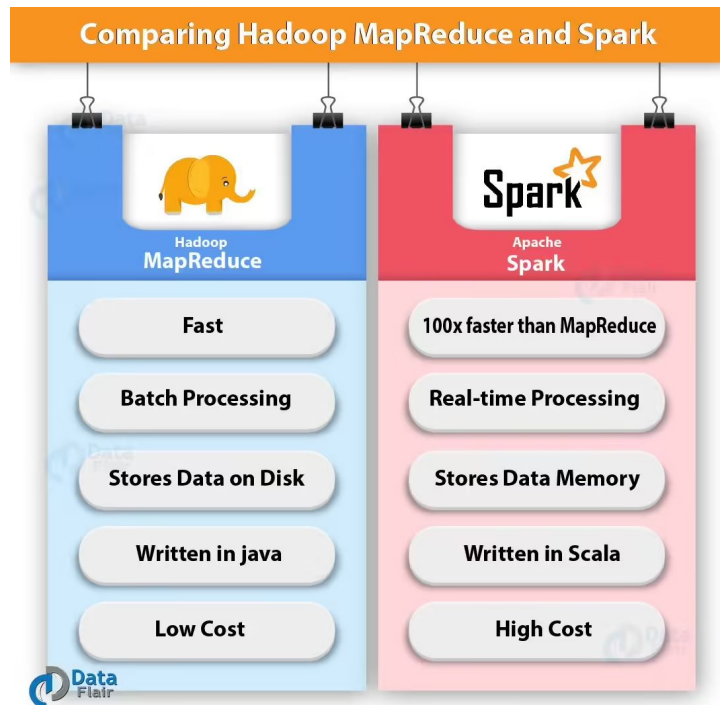
4

To Optimize Pr w.r.t P ,

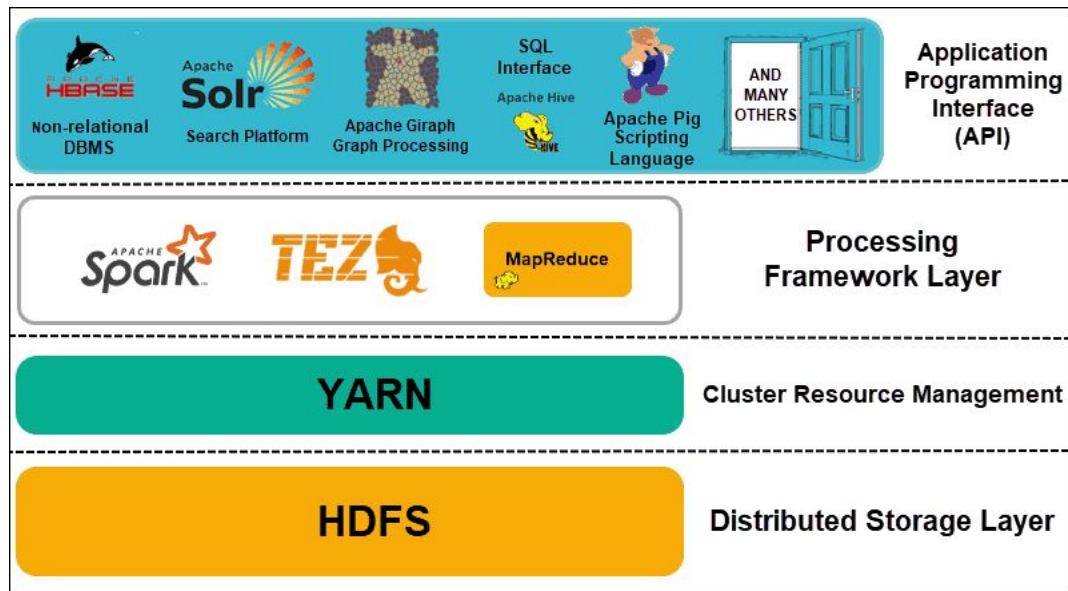
$$dPr/dP = a - 2bP + cb = 0$$

$$P = (a + cb) / 2b$$

Hadoop and Spark are two open-source frameworks that you can use to manage and process large amounts of data.



Like many other companies, Agoda also uses Spark for our large-scale data processing



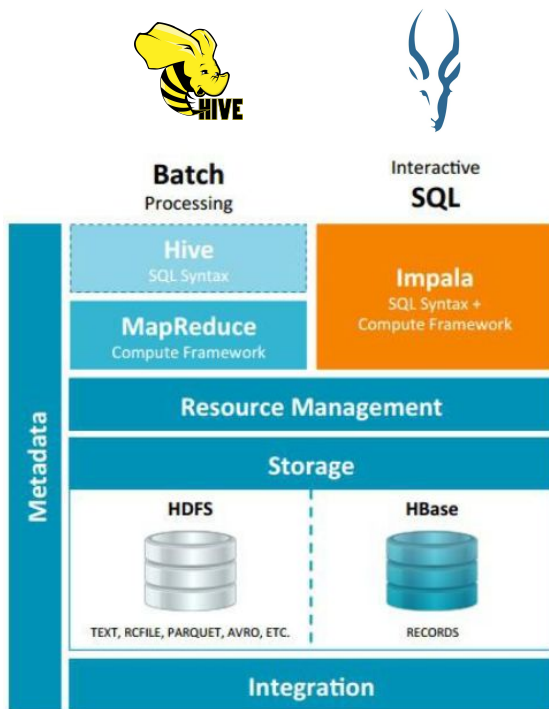
Spark SQL



Fun Facts

- Agoda stores ~30+ PB (10^{15}) of data, with ~600 TB (10^{12}) of new data written daily.
- Our platform processes approximately ~2.8 trillion events through Kafka and runs around 100K Spark jobs every day to handle data processing at scale.

Although related, Hive and Impala are completely different Application Interface level tools



Hive:

- Uses MapReduce as an execution engine (batch query)
- Fault-tolerance based on MapReduce and YARN framework
- High runtime overhead

Impala:

- Uses process-to-process data exchange (ad-hoc query)
 - Multi Parallel Processing (MPP)
- No fault tolerance
- Low runtime overhead

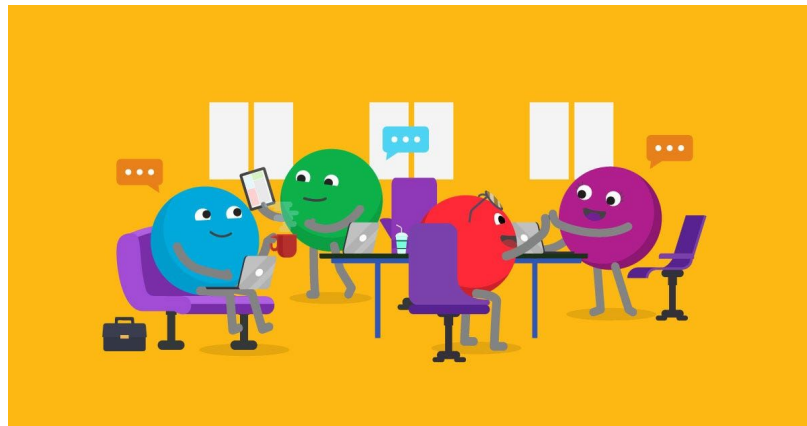
We have various Big Data engines for different use-cases to democratize our data, ensuring everyone can make data-driven decisions



	Apache Hive	Apache Impala	Apache Spark SQL
Users	ETL Developers	Business Analysts	Data Engineers & Data Scientists
Strengths	<ul style="list-style-type: none"> Built for very long running ETL, data processing or batch processing Supports custom file formats Handles massive ETL sorts with joins 	<ul style="list-style-type: none"> Scales to high-concurrency Supports high-performance interactive SQL Compatible with BI tools & skills Hadoop integration and usability 	<ul style="list-style-type: none"> Easily embed SQL into Java, Scala or Python applications Simple language for common operations Seamlessly mix SQL & Spark code within a single applications Automatic performance optimizations

Let's get your hands on Spark!

1. Prerequisite
 - a. Introducing PySpark, the Python API for Spark
 - b. Introducing the dataset
 - c. Example PySpark codes
2. Data Exploration time (40 mins)
3. Workshop Discussion (30 mins)





All the best for the rest of your academic journey & career!!!

JOIN US! :D

<https://careersatagoda.com/vacancies/>

