

GO-JEK's Data Journey*

Johanes Alexander

johanes@go-jek.com

*Business Intelligence (BI), GO-JEK



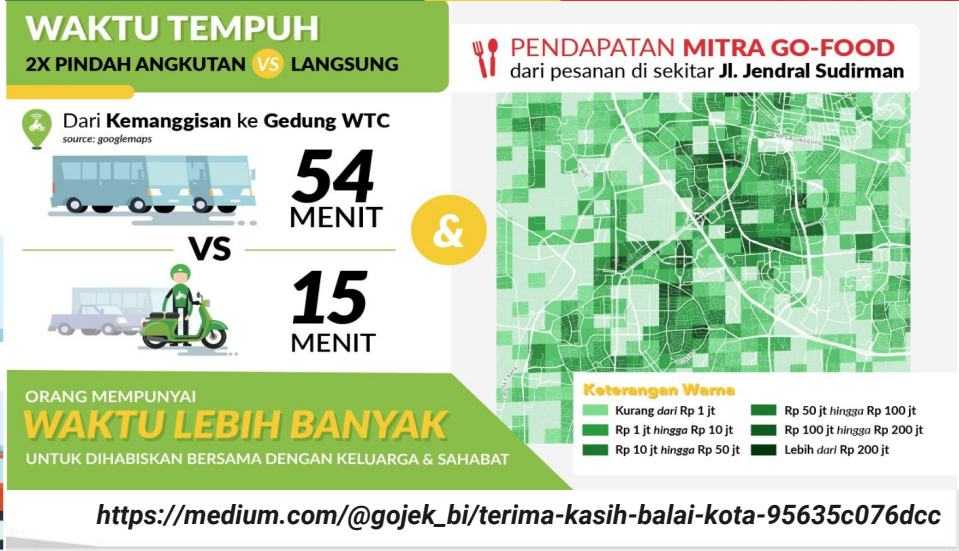
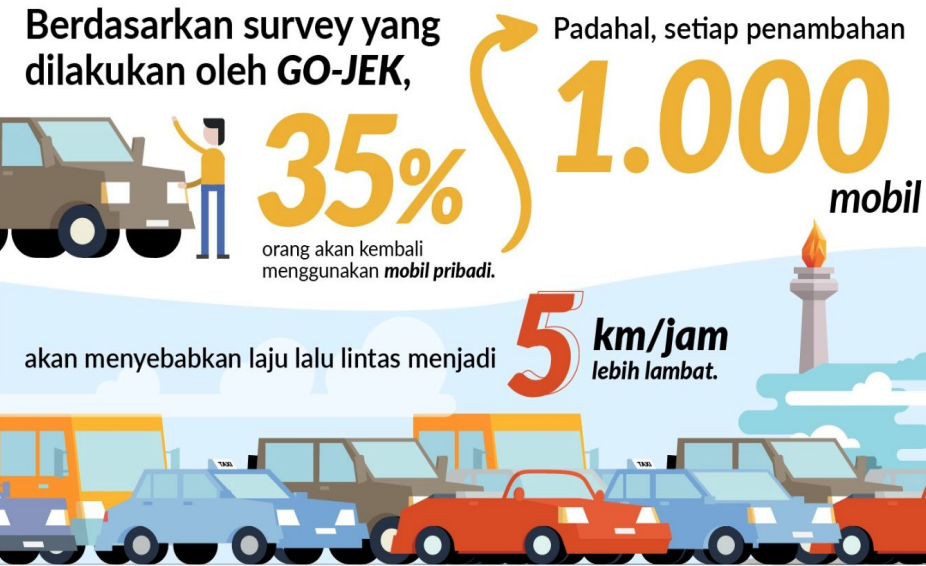
Today's discussion

- How it started
- (Over)simplified timeline
- Design Rationale
 - Audience of Data
 - Data Hierarchy
- BI Architecture
- Current Stats

How it started

By it's nature, GO-JEK have:

- **Wide array of products:** big product coverage, from transport, logistic to commerce
- **Wide array of business requirement:** different in business model, metrics understanding
- **Wide array of architecture:** variety of backend design, data structure



How it started - cont'd

In 2016:

- **Business Intelligence (BI) is the SVOT:** every business team rely on BI to help them make a decision
- **Most of data related activity is owned by BI:** generating report, creating insight, helping data-related production issues
- **There is no data repository:** every BI's (need to be) a rockstar

(Over)simplified timeline

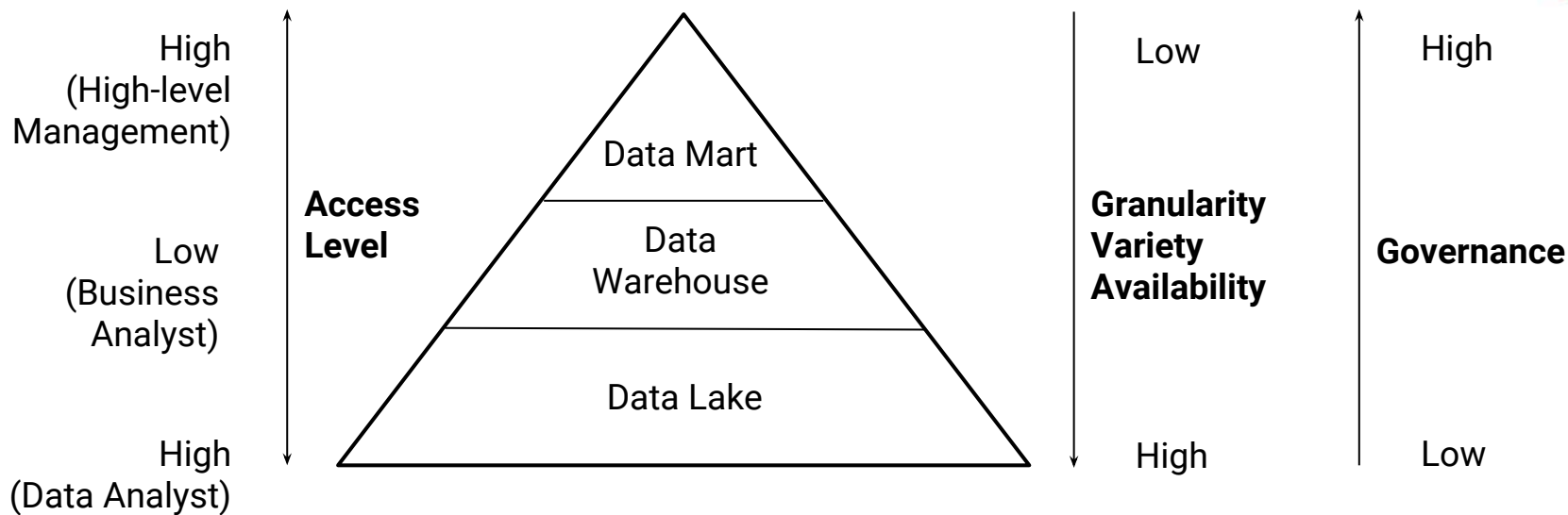


Design Rationale

Audience of Data

Who get access to data?	High Level Management	Business Analyst (and everyone else)	Data Analyst
What kind of data granularity ?	Low – business summary data	Medium – combination of summary and raw data	High – raw data
Why they need it?	Decision Support System	Hypothesis Analysis	Exploratory Analysis, System Creation
How they access it?	Worksheets or reports with simple slice and dice operations	Worksheets or reports with simple slice and dice operations	SQL query for complex analysis
		SQL query for more complex analysis	Custom code for statistical analysis, visualization, machine-based decision making, etc.

Data Hierarchy



BI Architecture

- ***What we want***

- High performance, scalable, minimum operation maintenance
- Full resolution dataset
- Easy data discovery process

- ***What do we built?***

- Leveraging cloud data warehouse, *BigQuery*
- Data modeling, creation of *denormalized tables*
- Creation of universal data dictionary, e.g. *Datadex*

COMPOSE QUERY

Query History

Job History

trips?

< sql New Query ?

Query Editor

UDF Editor

X

SQL

1

Standard SQL Dialect X

Ctrl + Enter: run query, Tab or Ctrl + Space: autocomplete.

RUN QUERY

Save Query

Save View

Format Query

Show Options

BQ Mate

- Public Datasets
- nyc-tlc:green (2)

nyc-tlc:yellow (1)

trips

Table Details: trips

Refresh

Query Table

Copy Table

Export Table

Delete Table

Schema

Details

Preview

Row	vendor_id	pickup_datetime	dropoff_datetime	pickup_longitude	pickup_latitude	dropoff_longitude	dropoff_latitude	rate_code	passenger_count	trip_distance	payme
1	CMT	2010-03-04 00:35:16 UTC	2010-03-04 00:35:47 UTC	-74.035201	40.721548	-74.035201	40.721548	1	1	0.0	Cas
2	CMT	2010-03-15 17:18:34 UTC	2010-03-15 17:18:35 UTC	0.0	0.0	0.0	0.0	1	1	0.0	Cas
3	VTS	2015-03-18 01:07:02 UTC	2015-03-18 01:07:07 UTC	0.0	0.0	0.0	0.0	1	5	0.0	CRD
4	VTS	2015-03-09 18:24:03 UTC	2015-03-09 18:25:37 UTC	-73.937248229980469	40.758201599121094	-73.937263488769531	40.7581901550293	1	1	0.0	CRD
5	CMT	2010-03-06 06:33:41 UTC	2010-03-06 06:36:06 UTC	-73.785514	40.6454	-73.784564	40.648681	1	2	4.1	Cas
6	CMT	2013-08-07 00:42:45 UTC	2013-08-07 00:58:43 UTC	-74.025817	40.763044	-74.046752	40.78324	5	1	4.8	CSH
7	VTS	2015-04-26 02:56:37 UTC	2015-04-26 03:00:01 UTC	-73.987655639648438	40.771656036376953	-73.987556457519531	40.771751403808594	1	1	0.0	CSH
8	VTS	2015-04-29 18:45:03 UTC	2015-04-29 18:49:01 UTC	0.0	0.0	0.0	0.0	1	1	1.0	CSH
9	CMT	2010-03-11 21:24:48 UTC	2010-03-11 21:46:51 UTC	-74.571511	40.9108	-74.628928	40.964321	1	1	68.4	Dis
10	CMT	2013-08-24 01:58:23 UTC	2013-08-24 01:58:23 UTC	-73.972171	40.759439	0.0	0.0	5	4	0.0	CSH
11	CMT	2015-01-27 18:24:31 UTC	2015-01-27 18:24:46 UTC	-73.983390808105469	40.684906005859375	-73.983390808105469	40.684906005859375	5	0	1.1	CRD
12	CMT	2015-02-07 00:05:40 UTC	2015-02-07 00:40:35 UTC	-73.9864730834961	40.745235443115234	-74.125823974609375	40.649356842041016	5	1	12.8	CRD
13	CMT	2015-01-11 15:52:20 UTC	2015-01-11 15:54:50 UTC	-73.990921020507812	40.771244049072266	-73.998123168945312	40.7611083984375	5	2	8.1	NOC
14	CMT	2015-01-25 02:13:45 UTC	2015-01-25 02:14:28 UTC	-73.879440307617188	40.828388214111328	-73.878921508789062	40.828487396240234	5	1	9.3	CSH
15	CMT	2015-06-12 04:44:54 UTC	2015-06-12 04:50:02 UTC	-73.9959945678711	40.738574981689453	-73.994544982910156	40.752754211425781	5	2	1.3	NOC

Table

JSON

First < Prev Rows 1 - 15 of 1108779463 Next > Last

Data Modeling for BigQuery

<https://www.youtube.com/watch?v=Vj6ksosHdhw>

GO  JEK



Google Cloud

1 hit

name:*merchant_name* AND table:*gofood*

[Uses lucene query syntax](#)

Add a filter +

data-dict* ▾

Selected Fields

t

 description

t


 name

t

 table

t

 type

Available Fields 

Popular

t

 annotation

t

 _id

t

 _index

#

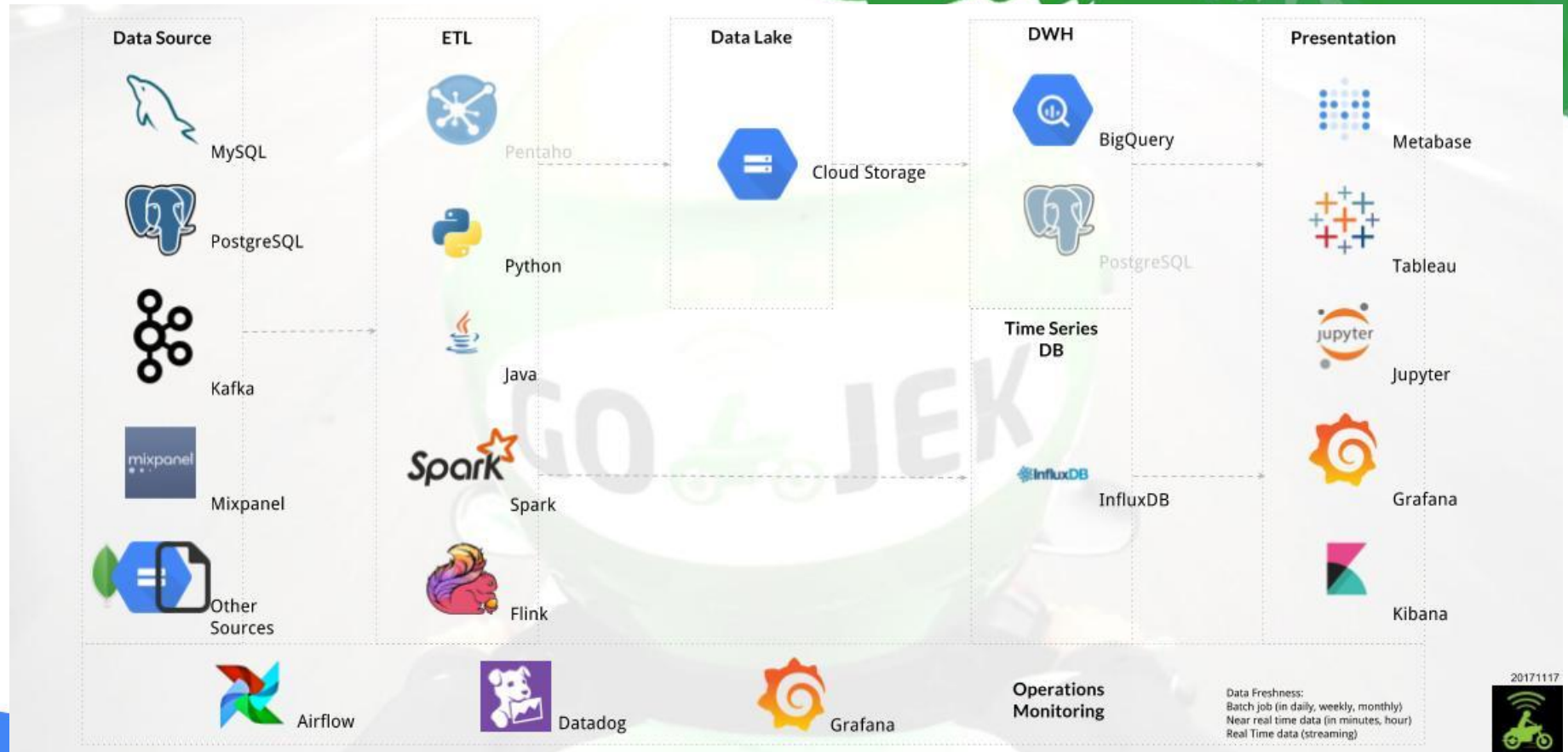
 _score

t

 _type



table	name	description	type
▶ access.sd_gofood_booking	merchant_name	Ordered gofood merchant name generated from merchant outlet name to have generic name for the merchant that have multiple branches. e.g: KFC, Hokben	STRING
access.sd_gokilatshop_booking			
access.sd_gomart_booking			



Current Stats

> 30%

Growing Data Volume per Month

**This is only business metrics data collected by BI.*

> 150

Multi Resolution Dataset

**Contains full resolution, reusable summary and roll-up dataset.*

> 6000

Metabase Cards & Tableau Dashboards

> 1000

Data Points

From over 35 internal and external data source

~ 500

Average Daily Metabase & Tableau Users

**Everyone just loves data!*

Thank You

