

Kom i gang med DAPLA

Øyvind Bruer-Skarsbø

10/9/2022

Innhold

Velkommen	5
Forord	6
I Introduksjon	7
1 Hva er Dapla?	9
2 Hvorfor Dapla?	10
3 Arkitektur	11
4 Innlogging	12
5 Jupyterlab	13
6 Bakke vs. sky	14
II Opprette Dapla-team	15
7 Hva er Dapla-team?	16
8 Opprette Dapla-team	17
9 Google Cloud Console	18
10 Lagre data	19
11 Hente data	20
12 Fra bakke til sky	21
13 Administrasjon av team	22

III Beste-praksis for koding	23
14 SSB-project	24
14.1 Opprett GitHub-bruker	24
14.2 Koble seg til SSB	24
14.3 Autentifisering	24
14.4 ssb-project-cli	24
15 Git og Github	25
16 Virtuelle miljøer	26
16.1 Python	26
16.2 R	26
17 Jupyter-kernels	27
18 Installere pakker	28
18.1 Python	28
18.1.1 Installering	28
18.1.2 Avinstallering	28
18.1.3 Oppgradere pakker	28
18.2 R	28
18.2.1 Installering	28
18.2.2 Avinstallering	28
18.2.3 Oppgradere pakker	28
19 Samarbeid	29
20 Vedlikehold	30
IV Jupyterlab på bakken	31
21 Lese inn filer	32
21.1 sas7bdat	32
21.2 Oracle	32
21.3 Fame	32
21.4 Tekstfiler	32
21.5 Parquet	32
V Avansert	33
22 IDE'er	34
22.1 RStudio	34

22.2 VSCode	34
22.3 Pycharm	34
23 Scheduling	35
24 Databases	36
24.1 BigQuery	36
24.2 CloudSQL	36
Referanser	37

Velkommen

DAPLA står for dataplattform og er SSBs nye plattform for statistikkproduksjon. Arbeidet startet som et utviklingsprosjekt i 2018 i sammenheng med Skatteetatens prosjekt *Sirius*. Idag er plattformen mer moden og klar for å ta imot flere statistikker. Denne boken er ment som

DAPLA står for dataplattform og er SSBs nye plattform for statistikkproduksjon. Arbeidet startet som et utviklingsprosjekt i 2018 i sammenheng med Skatteetatens prosjekt *Sirius*. Idag er plattformen mer moden og klar for å ta imot flere statistikker. Denne boken er ment som

i Denne boken er skrevet med [Quarto](#) og er publisert på <https://statisticsnorway.github.io/dapla-manual/>. Alle ansatte i SSB kan bidra til boken ved klone [dette repoet](#), gjøre endringer i en branch, og sende en pull request til administratorene av repoet (Team Statistiktjenester).

Forord

Denne boken vil la SSB-ansatte ta i bruk grunnleggende funksjonalitet på DAPLA uten hjelp fra andre.

Part I

Introduksjon

Målet med dette kapitlet er å gi en grunnleggende innføring i hva som legges i ordet **Dapla**. I tillegg gis en forklaring på hvorfor disse valgene er tatt.

1 Hva er Dapla?

2 Hvorfor Dapla?

3 Arkitektur

Hvilke komponenter er plattformen bygd opp på? Forklart på lettest mulig måte.

4 Innlogging

5 Jupyterlab

6 Bakke vs. sky

Part II

Opprette Dapla-team

7 Hva er Dapla-team?

Mer kommer

8 Opprette Dapla-team

9 Google Cloud Console

10 Lagre data

11 Hente data

12 Fra bakke til sky

13 Administrasjon av team

Part III

Beste-praksis for koding

14 SSB-project

Mer kommer.

14.1 Opprett GitHub-bruker

14.2 Koble seg til SSB

14.3 Autentifisering

14.4 ssb-project-cli

15 Git og Github

16 Virtuelle miljøer

16.1 Python

Poetry bør benyttes for python. Mer kommer

16.2 R

17 Jupyter-kernels

18 Installere pakker

18.1 Python

Mer kommer.

18.1.1 Installering

18.1.2 Avinstallering

18.1.3 Oppgradere pakker

18.2 R

18.2.1 Installering

18.2.2 Avinstallering

18.2.3 Oppgradere pakker

19 Samarbeid

Noen har opprettet et ssb-project og pushet til Github. Hvordan skal kollegaer gå frem for å bidra inn i koden?

20 Vedlikehold

Part IV

Jupyterlab på bakken

21 Lese inn filer

Mer kommer.

21.1 sas7bdat

21.2 Oracle

21.3 Fame

21.4 Tekstfiler

21.5 Parquet

Part V

Avansert

22 IDE'er

Forklare situasjonen nå. Kun Jupyterlab. Kan kjøre remote session med Rstudio, Pycharm og VSCode.

22.1 RStudio

22.2 VSCode

22.3 Pycharm

23 Scheduling

24 Databaser

24.1 BigQuery

24.2 CloudSQL

Referanser